

MATEMÁTICAS
T.D
42

DISEÑO BAYESIANO DE MUESTRAS, POSIBLEMENTE
ESTRATIFICADAS, CUANDO LA UTILIDAD TERMINAL
ES FUNCION DE LA INFORMACION CONSEGUIDA

Por
ERNESTO VERES FERRER

Tesis presentada para aspirar
al grado de Doctor en Ciencias Exactas
en la
Universidad de Valencia

Departamento de Bioestadística de la Facultad de Medicina
de la Universidad de Valencia

Marzo 1981



UMI Number: U607782

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U607782

Published by ProQuest LLC 2014. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

DISEÑO BAYESIANO DE MUESTRAS, POSIBLEMENTE
ESTRATIFICADAS, CUANDO LA UTILIDAD TERMINAL
ES FUNCION DE LA INFORMACION CONSEGUIDA

Por E. VERES FERRER

Marzo, 1981

BIBLIOTECA
FACULTAD C. MATEMATICAS
VALENCIA

UNIVERSIDAD DE VALENCIA	
FACULTAD DE CIENCIAS MATEMATICAS	
BIBLIOTECA	
N.º Registro	<u>1011</u>
<u>MATEMATICAS</u>	
SIGNATURA	<u>TESIS D./42</u>
C. D. U. 519.2 (043.2)	

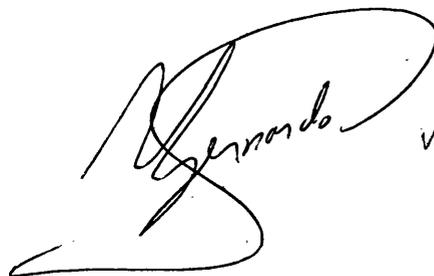
i19095600
b16837393

CERTIFICADO

JOSE, M. BERNARDO HERRANZ, Profesor Agregado de Bioestadística y Director del Departamento de Bioestadística de la Facultad de Medicina de la Universidad de Valencia

CERTIFICA: Que la presente Memoria: "Diseño Bayesiano de muestras, posiblemente estratificadas, cuando la Utilidad terminal es función de la Información conseguida" ha sido realizada bajo mi dirección, en el Departamento de Bioestadística de la Facultad de Medicina de la Universidad de Valencia, por el Sr. D. Ernesto Veres Ferrer, y constituye su tesis para optar al grado de Doctor en Ciencias, sección de Matemáticas.

Y para que conste, en cumplimiento de la legislación vigente, presento ante la Facultad de Ciencias de la Universidad de Valencia, a 23 de febrero de 1.981 .

A handwritten signature in black ink, appearing to read 'Bernardo', with a large, sweeping flourish at the end.

Prof. Dr. D. José M. Bernardo Herranz

RESUMEN

La Inferencia Estadística sobre el valor de cierto parámetro θ desconocido es considerada aquí como un problema de decisión particular en el que el espacio de decisiones es la clase de distribuciones posteriores de probabilidad. De hecho, en este trabajo se considera que la realización de un experimento - consistente fundamentalmente en la obtención de una muestra de unidades relacionadas, de alguna manera especificada, con la cantidad de interés desconocida - tiene como objetivo adquirir conocimiento sobre θ , viniendo expresada la utilidad asociada a su realización en función de la información que proporciona. En definitiva, este planteamiento implica considerar a la información esperada como una adecuada medida de la utilidad que debe definirse en todo problema de decisión.

Definidas unas funciones adecuadas para expresar la utilidad asociada a un experimento, se aplican a resolver el problema de diseño experimental consistente en la determinación del tamaño muestral. En este sentido, la maximización de estas funciones - que son funciones, a su vez, de la información esperada - como un procedimiento de diseño de experimentos no es sino una aplicación particular del principio Bayesiano general de maximización de la utilidad esperada.

Más adelante, la situación anterior se generaliza a procedimientos más complejos, considerando experimentos que llevan asociados procesos de muestreo estratificados. El concepto de información esperada debe generalizarse a la nueva situación impuesta por la existencia de una muestra estratificada que, a su vez, incide directamente sobre la estructura de la cantidad de interés. Con esta generalización se abren nuevas posibilidades en el tratamiento de otros muchos problemas sugeridos por las distintas medidas de información.

Interesantes subproductos de la investigación anterior consisten, por una parte, en el modelo resuelto para una población uniforme y densidad inicial de Pareto, que ofrece una alternativa al ya tan estudiado modelo normal. Dicho ejemplo acompaña los resultados teóricos de la

tesis, ilustrando su caracterización práctica. Por otra, se resuelve el comportamiento asintótico de la información esperada proporcionada por una muestra estratificada ante transformaciones aproximadamente normales de la cantidad de interés. Esta aplicación pone de manifiesto para la situación generalizada de la información esperada unos resultados en línea con los obtenidos para procesos no estratificados. Todas estas cuestiones son expresadas en términos conducentes a resolver el problema planteado por la determinación de tamaños muestrales óptimos.

Palabras clave: Comportamiento asintótico; Diseño de Experimentos; Estadística Bayesiana; Inferencia Estadística; Medidas de Información; Muestreo Estratificado; Tamaño muestral óptimo; Utilidad esperada.

Clasificación AMS (1980):

Primaria: 62A15, 62B10, 62F15

Secundaria: 62B15, 62C10, 62D05, 62K05

PREFACIO

La presente tesis tuvo su motivación inicial en la búsqueda de aplicaciones concretas que demostraran que las medidas de información representan un lenguaje lo suficientemente fértil y amplio como para describir con ellas variadas situaciones planteadas por la Estadística Clásica. La determinación del tamaño muestral podía ser una de estas situaciones y en su estudio están orientados los primeros cuatro capítulos del trabajo que ahora presentamos. La segunda de las situaciones de diseño muestral tratada en los últimos dos capítulos responde a una estructura de mayor sofisticación: extender el concepto, ya clásico, de información esperada a experimentos asociados con procesos de muestreo estratificados.

La tesis está dividida en seis capítulos y cada uno de éstos en varias secciones. A veces, se distinguen también en las secciones distintos apartados que agrupan conceptos, técnicas de trabajo o aplicaciones diferenciadas. La numeración empleada para denotar los Teoremas, Lemmas, Ecuaciones y Ejemplos es la decimal y siempre referida, en sus dos primeros dígitos, a la sección a la que el correspondiente concepto pertenece. Así, por ejemplo, con Teorema 4.1.2 indicaremos el Teorema 2 de la Sección 1 del cuarto capítulo. Por otra parte, cada uno de éstos comienza con un breve resumen y comentario sobre su contenido.

Algunas de las ideas abarcadas en la tesis han sido presentadas en artículos del autor publicados en los números 80-81 y 82-83 de la revista Estadística Española, editada por el Instituto Nacional de Estadística.

La presente investigación ha sido realizada bajo la dirección del Profesor Dr. D. José-Miguel Bernardo Herranz, al que agradezco vivamente el interés prestado a este trabajo así como a la ayuda que, en todo momento, he encontrado en sus continuas explicaciones. También quisiera dejar constancia de mi agradecimiento a los miembros del Departamento de Estadística Matemática de la Facultad de Ciencias de la Universidad de Valencia, al frente de los cuales está mi primer educador en la Ciencia

Estadística, el Profesor Dr. D. Segundo Gutiérrez Cabria.

Finalmente, quisiera dedicar este trabajo a mi padre por la gran ilusión que puso en esta tesis, a pesar de representar una investigación tan alejada a las disciplinas de su carrera docente.

E.V.F.

VIII

INDICE

pág

RESUMEN

IV

PREFACIO

VI

INDICE

VIII

CAPITULO I

INTRODUCCION

- 1.1 La determinación del tamaño muestral como un problema de decisión 1
- 1.2 Notación y definiciones 3
- 1.3 Medida de información 5
- 1.4 Transformaciones de la cantidad de interés 11
- 1.5 Utilidad y valor de un experimento 13

CAPITULO II

DOS MODELOS CONCRETOS. SUS DIFERENCIAS

- 2.1 El modelo normal 18
- 2.2 Información esperada para una población uniforme y opiniones iniciales descritas por la distribución de Pareto 20
- 2.3 Aproximación a ambos modelos. Sus diferencias. 29

CAPITULO III

FUNCIONES PARA LA UTILIDAD DE UN EXPERIMENTO. SUS PROPIEDADES

- 3.1 Distintas expresiones para la utilidad de un experimento 37
- 3.2 Propiedades satisfechas por las anteriores funciones de utilidad 45

I. Utilidades positivas	45
II. Aditividad de las funciones de utilidad	47
III. La utilidad esperada como función del tamaño muestral	58
IV. Comportamiento de las funciones de utilidad como funcionales de la densidad inicial	64
3.3 La utilidad ante transformaciones de la cantidad de interés	77

CAPITULO IV

DETERMINACION DEL TAMAÑO MUESTRAL OPTIMO

4.1 Tamaño muestral óptimo	80
4.2 Aplicación	93
4.2.1 Aplicación al modelo normal	93
4.2.2 Aplicación a una población uniforme y densidad inicial de Pareto	100

CAPITULO V

EXTENSION AL MUESTREO CON ESTRATOS

5.1 Objetivos y notaciones. Hipótesis de trabajo comunes	113
5.2 Hipótesis específicas	122
5.3 Información esperada $I^{\theta} \{E^S(\underline{n}), p(\underline{g})\}$	136
5.4 Información esperada $I^{\theta} \{E^S(\underline{n}), p(\underline{g})\}$	149
5.5 Información esperada $I^{\theta} \{E^S(\underline{n}), p(\underline{g})\}$	195

CAPITULO VI

CALCULO DE LA INFORMACION ESPERADA PROPORCIONADA POR EL EXPERIMENTO $E^S(\underline{n})$. RESULTADOS PRACTICOS

6.1 Dos métodos para la obtención exacta de $I^{\theta} \{E^S(\underline{n}), p(\underline{g})\}$	212
I. Obtención de $I^{\theta} \{E^S(\underline{n}), p(\underline{g})\}$ por extensión de la cantidad de interés	213



II.	Obtención de $I^{\vartheta} \{ E^{\vartheta}(\underline{n}), p(\underline{\vartheta}) \}$ por consideración de estadísticos suficientes en todos los estratos	219
6.2	El modelo normal	222
I.	Obtención de la información esperada en el modelo normal por extensión de la cantidad de interés	225
II.	Obtención de la información esperada en el modelo normal por consideración de estadísticos suficientes en todos los estratos	228
III.	Obtención de las informaciones esperadas sobre ϑ y ϑ_i	235
6.3	Comportamiento asintótico de $I^{\vartheta} \{ E^{\vartheta}(\underline{n}), p(\underline{\vartheta}) \}$	240
6.4	Ejemplo de cálculo aproximado: modelo multinomial	254
	CONCLUSIONES Y AREAS PARA INVESTIGACIONES ADICIONALES	266
	BIBLIOGRAFIA	268

CAPITULO 1

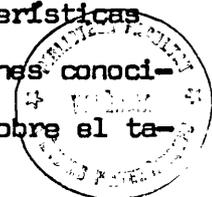
INTRODUCCION

En este primer capítulo presentaremos las definiciones así como la notación básica empleadas. Así mismo, centraremos el primero de los dos objetivos fundamentales que se pretenden en la presente tesis: desarrollar una teoría normativa de fácil comprensión para la determinación del tamaño muestral óptimo a partir de la instrumentación proporcionada por la Teoría de la Información y bajo una visión eminentemente Bayesiana.

1.1 LA DETERMINACION DEL TAMAÑO MUESTRAL COMO UN PROBLEMA DE DECISION

Tanto desde un punto de vista teórico como en el diseño práctico de una encuesta por muestreo probabilístico nos encontramos, tarde o temprano, con el problema presentado por la determinación del tamaño muestral. Fundamentalmente, se trata de un problema de conflicto de intereses y de asignación óptima de recursos. En efecto, el hablar de tamaño muestral presupone considerar la existencia de un costo asociado a la obtención de dicha muestra. Es evidente que un tamaño muestral excesivamente pequeño puede dar lugar a inferencias no representativas de la realidad, mientras que un tamaño excesivamente grande puede resultar prohibitivo a nuestro presupuesto, al tiempo que puede implicar una dilapidación de recursos comparativamente a la precisión exigida a nuestros propósitos. El tamaño muestral óptimo será, pues, la solución de compromiso entre la maximización de la precisión y la minimización del costo.

El investigador que debe realizar cierta inferencia sobre el valor de cierta magnitud desconocida puede apoyarse en la información y conocimientos proporcionados por una muestra de alguna población cuyas características estén relacionadas con la magnitud desconocida a través de expresiones conocidas de densidades de probabilidad. Al mismo tiempo deberá decidir sobre el ta-



maño de dicha muestra. Presentado así el problema no se trata sino de un especial problema de decisión en el que el espacio de posibles decisiones queda reducido al conjunto de los números naturales, $D = N$ (ó bien, en un sentido amplio, puede suponerse que $D = Z$, de forma que la decisión asociada al subconjunto Z^- sería, evidentemente, la de no obtener muestra alguna). En lo que sigue, sin embargo, supondremos siempre que se realiza una experimentación consistente en la extracción de una muestra de, al menos, una unidad muestral a fin de recabar información sobre cierta magnitud desconocida. Por tanto, el investigador deberá ser capaz de determinar hasta qué punto debe llevar el muestreo para que éste le resulte rentable, esto es, deberá determinar el tamaño muestral.

Precisamente, la Teoría de la Información, a través del concepto de información esperada proporcionada por un experimento, nos proporciona a su vez el instrumento adecuado para resolver el problema. Concretamente, el investigador confrontará la utilidad que espera recabar con la obtención de una muestra de tamaño n con su costo esperado. Mientras que aquélla, la utilidad del experimento, será una función de la información esperada como tendremos ocasión de justificar, es evidente que el costo tendrá una expresión más sencilla y concreta como función del tamaño muestral y que, en situaciones prácticas, suele llevarnos a expresiones muy manejables del mismo. Así pues, el investigador dispondrá de una función dependiente, en último término, del tamaño muestral que debidamente optimizada nos proporcionará el valor idóneo para n .

Los demás elementos que intervienen en todo problema de decisión con experimentación - aparte del espacio de decisiones que, como hemos dicho, se trata en este caso de un espacio numerable - son: el espacio paramétrico o espacio de estados de la naturaleza, esto es, un conjunto en el que toma valores un parámetro desconocido cuyo valor concreto pretendemos inferir con la realización de cierto experimento; una función de pérdida o de ganancia que llamaremos más adelante el valor esperado del experimento (definido como diferencia entre su utilidad y costo esperados, expresados ambos en las mismas unidades), y que será función del tamaño muestral; una familia de distribuciones de probabilidad, cada una de ellas identificadas por los valores del parámetro, de una cierta cantidad aleatoria que toma valores en un cierto espacio muestral y sobre el que realizamos el experimento; y, finalmente, situándonos dentro de una visión Bayesiana, supondremos que, a priori, el investigador es capaz de asignar una distribución de probabilidad para describir sus opiniones iniciales sobre

el valor del parámetro.

Con todos los elementos anteriores vamos a desarrollar una sencilla normativa que permitirá obtener el tamaño muestral óptimo por maximización del valor esperado de un experimento. Este es, pues, el objetivo fundamental de esta primera parte de la tesis que vamos a desarrollar.

1.2 NOTACION Y DEFINICIONES

A lo largo de todo este trabajo supondremos que el objetivo pretendido por el investigador al recabar información procedente de una muestra es el de mejorar su conocimiento sobre el valor de una cierta cantidad aleatoria θ que toma valores en un determinado conjunto Θ . La incógnita θ recibe los nombres de cantidad de interés y/o parámetro. Básicamente, el investigador proyectará un experimento de forma que la distribución de sus resultados dependa de θ de alguna manera especificada, proporcionando así una cierta cantidad de información sobre su valor. Tal como demuestra Bernardo (1975) el concepto de información esperada proporciona un poderoso instrumento para analizar el contenido de la información que se obtiene a partir de una muestra.

El modelo matemático básico que vamos a considerar para el ulterior desarrollo consistirá de un espacio muestral X , de elementos x indistintamente llamados muestra, dato y/o resultado experimental, provisto de una adecuada σ -álgebra de conjuntos sobre los que existe una familia dada de medidas de probabilidad. Estas medidas están identificadas por el parámetro θ con valores en el espacio paramétrico Θ . Esta es, pues, la relación existente entre la cantidad de interés desconocida con la muestra, y de cuya investigación se pretende obtener por inferencia nuestros resultados.

Por regla general, y con objeto de simplificar la notación, no intentaremos especificar la descripción de las funciones de densidad. Así, $p_x(\cdot)$ denotará la función de densidad de la cantidad aleatoria x en un punto genérico perteneciente a X , mientras que $p_x(x_0)$ no será sino su valor en la abscisa particular $x = x_0$. Análogamente, con $p_\theta(\cdot)$ denotaremos la función de densidad de la cantidad aleatoria θ , sin que ello nos permita asegurar que las cantidades aleatorias x y θ tengan la misma distribución. Sin embargo,

las densidades especiales utilizadas en ejemplos concretos se denotarán por símbolos propios. Concretamente, dos serán los ejemplos básicos que servirán para el desarrollo posterior de la tesis y sobre los que aplicaremos los resultados encontrados. Y así, siendo x una variable aleatoria descrita por una densidad normal de media μ y varianza σ^2 , y si ϑ se distribuye según una densidad de Pareto con parámetros ϑ_0 y k , las respectivas funciones de densidad vendrán expresadas con la notación $N_x(\cdot/\mu, \sigma^2)$ y $P_\vartheta(\cdot/\vartheta_0, k)$, donde

$$N_x(x/\mu, \sigma^2) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}\right) \quad x \in \mathbb{R}$$

$$P_\vartheta(\vartheta/\vartheta_0, k) = \frac{k \cdot \vartheta_0^k}{\vartheta^{k+1}} \quad \vartheta > \vartheta_0$$

Denotaremos por $E = \{X, \mathcal{G}, p_x(\cdot/\vartheta)\}$ el experimento que consiste en obtener una observación de la cantidad aleatoria $x \in X$, la cual se distribuye, para un cierto $\vartheta \in \mathcal{G}$ dado, de acuerdo con la densidad de probabilidad $p_x(\cdot/\vartheta)$ y con respecto a alguna medida dominante σ -finita definida sobre X . Por otra parte, supondremos que si ϑ_i y ϑ_j son valores diferentes de ϑ , las distribuciones correspondientes para x expresadas a través de las densidades $p_x(\cdot/\vartheta_i)$ y $p_x(\cdot/\vartheta_j)$ difieren salvo un conjunto de medida dominante nula. Es evidente que muchas situaciones experimentales pueden describirse con estos elementos. El experimento consistente en n repeticiones del anterior lo denotaremos por $E(n)$. Este experimento - donde las repeticiones se suponen independientes - resulta, por tanto, función del tamaño muestral n a través de la correspondiente densidad $p_{x(n)}(\cdot/\vartheta)$, siendo $x(n) = (x_1, x_2, \dots, x_n)$ el dato obtenido.

Anteriormente se indicó que el problema de determinación del tamaño muestral iba a ser situado bajo una visión Bayesiana. A tal fin, dicho argumento extiende el modelo básico anterior al suponer que \mathcal{G} es soporte de alguna σ -álgebra, existiendo una medida de probabilidad sobre ella que denotaremos por $p_\vartheta(\cdot)$. Esta medida de probabilidad puede interpretarse (de Finetti, 1937; Villegas, 1967) como descripción de las opiniones iniciales y personales del científico sobre el valor del parámetro ϑ antes de que se realice el experimento que proporciona la muestra. Se trata, por tanto, de una densidad "inicial",

en contraposición con las densidades finales que describen las opiniones que el investigador tiene sobre ϑ una vez que conoce y ha estudiado la información proporcionada por el resultado experimental y dispone, por ello, de más elementos de juicio para afinar su descripción sobre ϑ .

Por simplicidad en la notación, la integración respecto la medida dominante \mathcal{G} -finita definida sobre el conjunto Ω de posibles valores de una cantidad aleatoria w , se denotará simplemente por

$$\int \dots dw \quad ,$$

puesto que siempre supondremos, salvo indicación expresa al contrario, que el rango de integración en las fórmulas usadas en la presente tesis está siempre extendido a todo el espacio, por lo que lo omitiremos. En caso contrario, y en ejemplos concretos, haremos mención expresa de los límites de integración.

Si la densidad inicial para ϑ es $p_{\vartheta}(\cdot)$, definiremos como densidad "predictiva" de los resultados experimentales proporcionados por el experimento $E = \{X, \mathcal{C}, p_x(\cdot/\vartheta)\}$ a la integral

$$p_x(x) = \int p_x(x/\vartheta) \cdot p_{\vartheta}(\vartheta) \cdot d\vartheta \quad .$$

Además, una vez realizado el experimento E y obtenido el resultado x , las opiniones del investigador sobre ϑ vendrán expresadas por la densidad "posterior" de Bayes, definida por el cociente entre densidades de probabilidad

$$p_{\vartheta}(\vartheta/x) = p_x(x/\vartheta) \cdot p_{\vartheta}(\vartheta) / p_x(x) \quad .$$

Con todos los elementos anteriores estamos ya en condiciones de definir el concepto de información esperada proporcionada por un experimento, concepto fundamental que, a su vez, permitirá llegar al del valor esperado de un experimento como expresión que engloba su utilidad y costo esperados y que nos permitirá resolver el problema planteado por la determinación del tamaño muestral óptimo.

1.3 MEDIDA DE INFORMACION

En muchos sectores de la experimentación científica las personas

encargadas de realizarla parecen no considerar una acción específica, moviéndose de una manera más o menos intuitiva por lo que consideran su utilidad. Así pues, ésta parece describirse aproximadamente por la información que sobre la cantidad de interés θ en estudio obtienen a través del resultado experimental x . Precisamente por ello y a fin de introducir el concepto de información esperada, convendrá definir previamente a la entropía, como componente que interviene en la definición de dicha información esperada.

Con la notación expresada en la sección anterior y para una función de densidad $p_{\theta}(\cdot)$, definimos:

DEFINICION. Dada la densidad de probabilidad $p_{\theta}(\cdot)$, llamamos operador entropía y lo denotamos por $H(p_{\theta}(\cdot))$, a la integral

$$H(p_{\theta}(\cdot)) = - \int p_{\theta}(\vartheta) \cdot \log p_{\theta}(\vartheta) \cdot d\vartheta \quad .$$

Esta expresión, que puede encontrarse en Lindley (1956), no es necesariamente finita y, por tanto, no tiene asegurada su existencia. Sin embargo, en todo lo que sigue, supondremos que la integral anterior es convergente, dejando a un lado el problema abierto de encontrar las condiciones necesarias y suficientes para la existencia de la entropía. Precisamente el nombre de entropía procede de su utilización en estadística física. Consideraremos, además, que para un cierto ϑ tal que $p_{\theta}(\vartheta) = 0$, entonces $p_{\theta}(\vartheta) \cdot \log p_{\theta}(\vartheta) = 0$, para así completar la definición.

Realizado un experimento $E = \{X, \mathcal{W}, p_x(\cdot/\theta)\}$ y obtenido el resultado muestral x las opiniones sobre el parámetro θ se modificarán por la información que, sobre él, está contenida en x . Así, las opiniones sobre θ vendrán a describirse por la densidad posterior $p_{\theta}(\cdot/x)$, resultando entonces la expresión

$$- \int p_{\theta}(\vartheta/x) \cdot \log p_{\theta}(\vartheta/x) \cdot d\vartheta$$

una función de x . Su valor esperado (suponiendo su existencia) respecto a x proporciona la siguiente definición:

DEFINICION. Dada una densidad inicial $p_{\theta}(\cdot)$, y realizado el experimento $E = \{X, \mathcal{W}, p_x(\cdot/\theta)\}$, obtenido el resultado experimental x , definimos como entropía esperada de la densidad posterior $p_{\theta}(\cdot/x)$ a la esperanza

$$E_x H(p_{\theta}(\cdot/x)) = - \int p_x(x) \left\{ \int p_{\theta}(\vartheta/x) \cdot \log p_{\theta}(\vartheta/x) \cdot d\vartheta \right\} dx \quad .$$

Nuevamente, en lo que sigue supondremos que la doble integral anterior es convergente. Por otra parte, teniendo en cuenta la igualdad

$$p_{\theta}(\theta/x) \cdot p_x(x) = p_{\theta x}(\theta, x)$$

la entropía esperada admite como nueva expresión

$$E_x H(p_{\theta}(\cdot/x)) = - \iint p_{\theta x}(\theta, x) \cdot \log p_{\theta}(\theta/x) \cdot d\theta \cdot dx \quad .$$

Ya se ha indicado que el objetivo declarado que motiva la realización de un experimento es el de proporcionar conocimiento sobre la cantidad de interés. En efecto, cualquier experimento E proporciona información sobre θ , de forma que para diseñar adecuadamente dicho experimento es importante considerar el acopio de información que se espera proporcione. Pues bien, Shannon (1948) propone una adecuada medida de la información que se espera proporcione el resultado muestral x , y Lindley (1956) traslada esto al lenguaje estadístico considerando la información esperada proporcionada por un experimento como diferencia entre la entropía de la densidad a priori (que es la que describe las opiniones iniciales del investigador) y la entropía esperada sobre la densidad posterior una vez realizado el experimento (esto es, cuando las antiguas opiniones se modifican por la información contenida en el resultado muestral). Resulta, pues, la siguiente

DEFINICION. Dada una densidad inicial $p_{\theta}(\cdot)$, la información total esperada proporcionada por el experimento $E = \{X, \mathcal{E}, p_x(\cdot/\theta)\}$ cuando el resultado muestral obtenido es x , viene dada por

$$I^{\theta} \{E, p_{\theta}(\cdot)\} = H(p_{\theta}(\cdot)) - E_x H(p_{\theta}(\cdot/x)) \quad .$$

Desarrollemos la diferencia anterior. Resulta:

$$\begin{aligned} I^{\theta} \{E, p_{\theta}(\cdot)\} &= H(p_{\theta}(\cdot)) - E_x H(p_{\theta}(\cdot/x)) = - \int p_{\theta}(\theta) \cdot \log p_{\theta}(\theta) \cdot d\theta + \\ &+ \int p_x(x) \left\{ \int p_{\theta}(\theta/x) \cdot \log p_{\theta}(\theta/x) \cdot d\theta \right\} dx = - \int \log p_{\theta}(\theta) \left\{ \int p_{\theta x}(\theta, x) \cdot dx \right\} d\theta + \\ &+ \int \left\{ \int p_{\theta x}(\theta, x) \cdot \log p_{\theta}(\theta/x) \cdot d\theta \right\} dx = - \int \left\{ \int p_{\theta x}(\theta, x) \cdot \log p_{\theta}(\theta) \cdot dx \right\} d\theta + \\ &+ \int \left\{ \int p_{\theta x}(\theta, x) \cdot \log p_{\theta}(\theta/x) \cdot dx \right\} d\theta \end{aligned}$$

de donde, al ser

$$p_{\theta x}(\theta, x) = p_x(x) \cdot p_{\theta}(\theta/x)$$

resulta finalmente

$$I^{\theta} \{E, p_{\theta}(\cdot)\} = \iint p_{\theta x}(\theta, x) \cdot \log \frac{p_{\theta}(\theta/x)}{p_{\theta}(\theta)} \cdot d\theta \cdot dx$$

como expresión más operativa para la información esperada proporcionada por un experimento y en la que se aprecia claramente su interpretación como comparación entre las opiniones iniciales y finales del investigador, comparación expresada a través del cociente

$$p_{\theta}(\theta/x) / p_{\theta}(\theta) = p_x(x/\theta) / p_x(x) \quad .$$

La base de la función logaritmo utilizada en la definición es irrelevante. A menudo suele elegirse la base 2 en cuanto que se deduce de forma natural de una interpretación física de $I^{\theta} \{E, p_{\theta}(\cdot)\}$: como el número esperado de preguntas binarias sobre θ que razonablemente debe esperarse realizar para obtener igual información sobre θ que la proporcionada por el experimento E (Renyi, 1970, p.564). Sin embargo, y por conveniencia matemática en los desarrollos de los teoremas posteriores, utilizaremos siempre la base e de los logaritmos neperianos.

Queda como problema abierto la determinación de las condiciones necesarias y suficientes bajo las cuales $I^{\theta} \{E, p_{\theta}(\cdot)\}$ es finita y, por tanto, existe, en íntima relación con el problema análogo mencionado para la entropía. Una condición necesaria (Osteyee & Good, 1974, p. 32) es que la medida conjunta de x y θ sea absolutamente continua con respecto a la medida producto. Generalmente, con densidades iniciales suavemente positivas suelen encontrarse siempre informaciones finitas. También, y en lo que sigue, supondremos siempre la existencia de la información esperada.

Realizado el experimento $E(n)$ (consistente, según ya definimos, en n repeticiones independientes del experimento simple E) y obtenido el resultado $x(n)$, es inmediata la siguiente extensión del concepto de información esperada proporcionada por el experimento $E(n)$, como

$$I^{\theta} \{E(n), p_{\theta}(\cdot)\} = H(p_{\theta}(\cdot)) - E_{x(n)} H(p_{\theta}(\cdot/x(n))) =$$

$$= \iint p_{x(n), \theta}(x(n), \theta) \cdot \log \frac{p_{\theta}(\theta/x(n))}{p_{\theta}(\theta)} \cdot d\theta \cdot dx(n)$$

donde el espacio muestral resulta ser $X \times X \times \dots \times X^{(n)} = X^{(n)}$ y la integral correspondiente está extendida a dicho espacio. Para simplificar el desarrollo posterior no haremos distinción, de momento, entre los resultados muestrales $x^{(n)}$ y x , esto es, supondremos que extraemos una muestra genérica sin especificar su tamaño. No obstante, es evidente que la anterior expresión para la información esperada sobre θ proporcionada por $E(n)$ es función del tamaño muestral n , circunstancia ésta fundamental ya que permitirá optimizarla como función del tamaño muestral.

Admitiendo su existencia, la información esperada cumple una serie de importantes propiedades que, demostradas por Lindley(1956) y Stone (1958), le confieren precisamente su gran carácter operativo y su justificación empírica. Dichas propiedades responden al estudio de la información esperada como función del tamaño muestral, por una parte, y, por otra, como funcional del experimento E y de la densidad inicial $p_\theta(\cdot)$. Es conveniente mencionarlas dado que más adelante comprobaremos cuáles de ellas se mantienen en las expresiones de la utilidad del experimento que, como funciones de la información esperada, se desarrollaran en el capítulo tercero. También y dentro de esta línea, se estudiarán dichas propiedades -al extender el concepto de información esperada a un proceso de muestreo estratificado.

1ª PROPIEDAD. Lindley (1956) demuestra que la información esperada $I^\theta \{E, p_\theta(\cdot)\}$ es no negativa y que solamente se anula si y sólo si no existe variación en las opiniones del investigador por el hecho de realizar la experimentación, esto es, si $p_\theta(\cdot/x) = p_\theta(\cdot)$. Es evidente que la información esperada proporcionada por un experimento puede ser, en un momento dado, negativa, sobre todo cuando se obtengan resultados muestrales inesperados que no están previstos en las opiniones iniciales. Sin embargo la anterior propiedad asegura que la información esperada es no negativa. Esto permite afirmar que un experimento resulta, por término medio, útilmente informativo, a no ser que no se modifiquen las opiniones iniciales en cuyo caso, evidentemente, es irrelevante efectuar o no el experimento.

2ª PROPIEDAD. Considerada como funcional del experimento, la información esperada verifica cierta aditividad. Así, suponiendo que el dato x del experimento $E = \{X, \mathcal{D}, p_x(\cdot/\theta)\}$ consta de un par de observaciones $x_1 \in X$ y $x_2 \in X$, el par $E_i = \{X, \mathcal{D}_i, p_{x_i}(\cdot/\theta)\}$ ($i=1,2$) son dos experimentos y E se llama suma de E_1 y E_2 . También definimos el experimento $E_2(x_1) = \{X, \mathcal{D}, p_{x_2}(\cdot/\theta, x_1)\}$. Si consideramos $I^\theta \{E_2/E_1, p_\theta(\cdot)\}$ como la información esperada proporcionada por E_2 después de que E_1 se ha realizado y obtenido el dato

x_1 , su valor esperado sobre x_1 , $I^\theta \{E_2/E_1, p_\theta(\cdot)\}$ se define como la información esperada proporcionada por E_2 después de que se haya realizado E_1 . Bajo estas hipótesis Lindley (1956) demuestra la siguiente propiedad aditiva de la información esperada: "La información esperada sobre ϑ proporcionada por el experimento suma $E = (E_1, E_2)$ puede escribirse de la forma $I^\theta \{E_1, p_\theta(\cdot)\} + I^\theta \{E_2/E_1, p_\theta(\cdot)\}$, donde el segundo término es la información esperada sobre ϑ proporcionada por E_2 una vez que E_1 ha sido realizado".

La importancia de esta segunda propiedad es evidente al asegurar que la información esperada puede calcularse en etapas - al extraer muestras de tamaño superior a uno - sin más que ir sumando las informaciones residuales que, en cada momento, se obtienen suponiendo conocidos el proceso y resultados anteriores. Tal como demuestra Lee (1964), si esta segunda propiedad se exige en una formulación axiomática de medidas de información solamente son necesarias débiles condiciones de regularidad para probar la unicidad de la expresión logarítmica.

Cuando al pasar de una de estas etapas a la siguiente no exista variación en nuestras opiniones, esto es, si al pasar del experimento E_1 al E_2/E_1 resulta ser que $p_\theta(\cdot/x_1, x_2) = p_\theta(\cdot/x_1)$ independientemente del valor de x_2 , es evidente que, por la primera propiedad, la información esperada $I^\theta \{E_2/E_1, p_\theta(\cdot)\}$ es cero. De esta forma no existe ganancia de información al realizar esa segunda etapa, coincidiendo las informaciones $I^\theta \{E_1, p_\theta(\cdot)\}$ y $I^\theta \{E, p_\theta(\cdot)\}$, con $E = (E_1, E_2)$. La formulación de este hecho constituye un corolario de la segunda propiedad, demostrado también por Lindley (1956). Resulta evidente que ambos resultados, propiedad y corolario, pueden generalizarse de forma inmediata a un número finito de etapas. Por otra parte, el corolario establece que no existe pérdida de información si ésta se ciñe a la observación de un estadístico suficiente con respecto a la cantidad de interés.

3ª PROPIEDAD. Para un experimento $E(n)$ y una densidad inicial $p_\theta(\cdot)$, la información esperada considerada como función del tamaño muestral es cóncava y creciente. Esta propiedad (Lindley, 1956) no es sino una confirmación del hecho empírico de que la información marginal de sucesivas observaciones independientes y equidistribuidas es decreciente.

4ª PROPIEDAD. Por último, Lindley (1956) demuestra que la información esperada $I^\theta \{E, p_\theta(\cdot)\}$, considerada como funcional de la densidad inicial $p_\theta(\cdot)$, es cóncava. Esta propiedad, de difícil interpretación práctica,

se utiliza, en cambio, en ciertas demostraciones para desarrollar diseños óptimos por maximización de una función de información.

Estas cuatro propiedades que aseguran una formulación coherente para cualquier medida de información se comprobarán posteriormente para determinadas expresiones definidas en función de la información esperada. Y así, en el capítulo tercero al definir ciertas funciones concretas de utilidad para poder determinar el valor de un experimento aparecerán expresiones que, como funciones de $I^{\vartheta}\{E, p_{\vartheta}(\cdot)\}$, verificarán total o parcialmente dichas propiedades.

En el siguiente capítulo encontraremos y desarrollaremos dos ejemplos concretos para así trabajar con expresiones conocidas de $I^{\vartheta}\{E, p_{\vartheta}(\cdot)\}$ y que, elegidos por su contenido ilustrativo, nos acompañarán a lo largo de esta tesis.

1.4 TRANSFORMACIONES DE LA CANTIDAD DE INTERES

Un interesante problema, que permitirá posteriormente hablar de aproximaciones para informaciones esperadas desconocidas, consiste en conocer lo que ocurre al considerar como cantidad de interés una nueva cantidad aleatoria ψ , relacionada con ϑ a través de una función conocida. Comentemos, pues, brevemente algunos resultados obtenidos por Bernardo (1975) al extender el concepto de información esperada ante transformaciones de la cantidad de interés.

Efectivamente, si el investigador está más interesado en conocer el valor de $\psi = \psi(\vartheta)$ más bien que el de ϑ , entonces resultará de mayor interés conocer la distribución posterior de ψ que la de ϑ . Dado que ψ debe ser una función conocida de ϑ , puede definirse una transformación T_{ψ} de forma que

$$p_{\psi}(\cdot) = T_{\psi}(p_{\vartheta}(\cdot))$$

será la densidad de probabilidad de ψ respecto la medida definida en $\mathcal{Y} = \psi(\mathcal{X})$ e implicada por $p_{\vartheta}(\cdot)$. La interpretación para $p_{\psi}(\cdot)$ es, para la nueva cantidad de interés ψ , idéntica a la ya conocida para ϑ : $p_{\psi}(\cdot)$ describe las opiniones iniciales del investigador sobre ψ . Análogamente, una vez que el resultado experimental x es conocido, puede deducirse la densidad posterior $p_{\psi}(\cdot/x)$ a través, nuevamente, de la transformación T_{ψ} ,

$$p_{\psi}(\cdot/x) = T_{\psi}(p_{\vartheta}(\cdot/x))$$

Notemos que la densidad final $p_{\psi}(\cdot/x)$ depende obviamente del experimento E

a través de su resultado x y de la distribución inicial $p_{\theta}(\cdot)$.

Así pues y con los elementos anteriores, Bernardo (1978) define el concepto de información esperada útil, como extensión a la información esperada total definida en la sección anterior, y cuando la cantidad de interés es φ y no θ , a través de la análoga doble integral siguiente:

DEFINICION. La información esperada útil sobre $\varphi = \varphi(\theta)$ proporcionada por $E = \{X, \mathcal{D}, p_x(\cdot/\theta)\}$ y cuando la densidad inicial es $p_{\theta}(\cdot)$, es

$$I^{\varphi}\{E, p_{\theta}(\cdot)\} = \iint p_x(x) \cdot p_{\varphi}(\varphi/x) \cdot \log \frac{p_{\varphi}(\varphi/x)}{p_{\varphi}(\varphi)} d\varphi \cdot dx$$

Evidentemente que ambas informaciones esperadas, la útil y la total, coinciden cuando la transformación considerada es la trivial $\varphi(\theta) = \theta$. Una forma alternativa de expresar la definición anterior y que pone claramente de manifiesto que $I^{\varphi}\{E, p_{\theta}(\cdot)\}$ es invariante bajo transformaciones biyectivas tanto del espacio muestral X como del espacio paramétrico \mathcal{D} , es la expresada por la doble integral (Bernardo, 1978):

$$\iint p_{\theta x}(\theta, x) \cdot \log \frac{p_{\varphi}(\varphi(\theta)/x)}{p_{\varphi}(\varphi(\theta))} d\theta \cdot dx$$

Así pues, si la transformación $\varphi = \varphi(\theta)$ es biyectiva la información esperada útil sobre φ proporcionada por $E = \{X, \mathcal{D}, p_x(\cdot/\theta)\}$ es igual a la información esperada total sobre θ proporcionada por el mismo experimento, esto es,

$$I^{\theta}\{E, p_{\theta}(\cdot)\} = I^{\varphi}\{E, p_{\theta}(\cdot)\} \quad \text{con } \varphi = \varphi(\theta) \text{ biyección.}$$

También la información esperada útil verifica las tres primeras de las propiedades mencionadas para la información esperada total y cuyas demostraciones pueden encontrarse en Bernardo (1978). Sin embargo, $I^{\varphi}\{E, p_{\theta}(\cdot)\}$ no es una funcional cóncava de la densidad inicial $p_{\theta}(\cdot)$. No obstante, si es cierto que manteniendo fijo $p_{\theta}(\theta/\varphi)$, entonces $I^{\varphi}\{E, p_{\theta}(\cdot)\}$ sí es una funcional cóncava de $p_{\theta}(\cdot)$.

La diferencia entre las informaciones esperadas útil y total, esto es, la diferencia

$$I^{\theta/\varphi}\{E, p_{\theta}(\cdot)\} = I^{\theta}\{E, p_{\theta}(\cdot)\} - I^{\varphi}\{E, p_{\theta}(\cdot)\}$$

recibe el nombre de información esperada residual respecto θ , proporcionada por el experimento E y cuando la cantidad de interés es $\varphi = \varphi(\theta)$ y la densidad inicial es $p_{\theta}(\cdot)$. Se demuestra que $I^{\theta/\varphi}$ es no negativa, con lo que

queda puesto de manifiesto el carácter comparativo entre las dos informaciones esperadas definidas: la información esperada útil no puede exceder a la información esperada total.

Por último, aclaremos que siempre que hablemos de información esperada nos referiremos, salvo especificación en contra, a la información esperada total.

1.5 UTILIDAD Y VALOR DE UN EXPERIMENTO

Recordemos nuevamente que el objetivo propuesto es el de resolver el problema de diseño consistente en determinar un tamaño muestral óptimo a fin de que la información contenida en la muestra de dicho tamaño sea, a su vez, la óptima para realizar inferencia sobre la cantidad de interés θ . Necesitaremos, pues, una función de utilidad para completar la descripción de este problema de decisión.

De acuerdo con los argumentos de coherencia para una aproximación Bayesiana cuando se toma una decisión bajo incertidumbre (DeFinetti, 1937; Ramsey, 1926; Savage, 1954; Pratt et al., 1964; Lindley, 1971) y admitiendo que las consecuencias de tomar una elección particular dependen solamente de la elección y del valor verdadero (pero desconocido) de la cantidad de interés, existen unos pocos principios básicos para asegurar que el procedimiento de elección es razonable. De esta forma, un procedimiento coherente es el que no viola estos principios. Se ha demostrado que para asegurar esta coherencia es necesario estimar una distribución de probabilidad $p_{\theta}(\cdot)$ que describa las opiniones iniciales sobre θ , construir una función que represente la utilidad de todas las consecuencias posibles y seleccionar la decisión procurando la mayor utilidad esperada.

Por otra parte, dado el especial problema en el que nos hemos situado y en el que suponemos que al menos extraemos una muestra de tamaño unitario, resulta evidente que debido a la información contenida en dicha muestra las primitivas opiniones descritas por una densidad inicial pasan a describirse por una densidad posterior. De esta forma se aumenta la cantidad de información sobre θ , disminuyendo así el riesgo de tomar una decisión no acorde con el nivel de información exigido para realizar una buena inferencia sobre la cantidad de interés.

Algunos autores (Cox, 1958; Fisher, 1959; Robinson, 1975) argumentan que una estructura teórica de decisión puede resultar inapropiada para problemas de inferencia científica en los que, aparentemente, no existe una decisión a tomar. En contra de esta opinión, el punto de vista de Bernardo (1979) considera que dichos argumentos suponen una pérdida efectiva en la definición de un problema de inferencia, toda vez que la inferencia estadística puede considerarse como un caso particular o un problema de decisión particular en el que se aplica con estricta exactitud el argumento Bayesiano. El científico que obtiene una muestra de tamaño n para realizar inferencia sobre ϑ deberá describir sus opiniones finales a través de una densidad posterior. Además, la utilidad de estas conclusiones finales deberá, por tanto, medirse por la información que el proceso de muestreo proporciona. Expresado nuestro problema de esta forma, no se trata sino de realizar una experimentación científica consistente en un proceso de muestreo y del que la determinación del tamaño muestral no es sino una última consecuencia obtenida al maximizar la utilidad de todo el proceso. Podemos, pues, afirmar con Bernardo (1979) que el problema de determinación de un tamaño muestral no es sino una consecuencia del más general consistente en realizar inferencia estadística y donde ya se presupone la realización del experimento que consiste en la extracción de una muestra. Y como tal problema de inferencia estadística puede considerarse como un problema de decisión, en presencia de incertidumbre sobre ϑ , en el cual el espacio de decisiones es la clase de distribuciones finales de ϑ y la función de utilidad es una adecuada medida de información.

Denotemos, pues, por D al conjunto de todas las densidades de probabilidad que describen a la cantidad de interés ϑ , esto es, definidas sobre el conjunto \mathcal{D} y tomando valores sobre la recta real positiva, verificando que

$$\int p_{\vartheta}(\vartheta) \cdot d\vartheta = 1$$

El investigador, inicialmente, describirá sus opiniones sobre ϑ a partir de una de estas densidades. Sea, también, una función de utilidad

$$u: D \times \mathcal{D} \longrightarrow R$$

que exprese la utilidad de describir a la cantidad de interés por la densidad de probabilidad $p_{\vartheta}(\cdot)$ cuando el valor verdadero y desconocido de dicha cantidad sea ϑ . Evidentemente, siendo $p_{\vartheta}(\cdot)$ la densidad inicial, el investigador

deberá buscar una densidad de probabilidad que maximice su utilidad esperada, esto es, que verifique

$$\sup_{p_{\theta}^+(\cdot) \in D} \int u(p_{\theta}^+(\cdot), \vartheta) \cdot p_{\theta}(\vartheta) \cdot d\vartheta \quad .$$

Evidentemente, y a fin de ser coherente y honesto, sus opiniones deberán maximizar a la integral anterior cuando coincidan con las expresadas en un principio. En definitiva, si el investigador es coherente, la integral anterior debe maximizarse precisamente con la función de densidad inicial $p_{\theta}(\cdot)$. Este es el sentido de la definición de una utilidad exacta. Supongamos, pues, que trabajamos con utilidades exactas. El investigador puede desear recabar mayor información sobre θ . Para ello realizará el experimento $E = \{X, \mathcal{D}, p_x(\cdot/\vartheta)\}$ de forma que sus opiniones iniciales expresadas por $p_{\theta}(\cdot)$ pasarán a describirse a través de la densidad posterior $p_{\theta}(\cdot/x)$, de forma que la utilidad (insistimos en que la suponemos exacta) que se esperará de la elección coherente vendrá dada por

$$\int u(p_{\theta}(\vartheta/x), \vartheta) \cdot p_{\theta}(\vartheta/x) \cdot d\vartheta \quad .$$

Esta integral es función del resultado muestral x . Su valor esperado respecto a la densidad predictiva $p_x(\cdot)$ proporciona, a su vez, la utilidad esperada conocida antes de que el valor concreto del resultado muestral se obtenga. Resultaría:

$$\int p_x(x) \left\{ \int u(p_{\theta}(\vartheta/x), \vartheta) \cdot p_{\theta}(\vartheta/x) \cdot d\vartheta \right\} dx \quad .$$

Así pues, y por el hecho de realizar el experimento E , se obtendría una ganancia en la utilidad dada por la diferencia

$$\int p_x(x) \int u(p_{\theta}(\vartheta/x), \vartheta) \cdot p_{\theta}(\vartheta/x) \cdot d\vartheta \cdot dx - \int u(p_{\theta}(\vartheta), \vartheta) \cdot p_{\theta}(\vartheta) \cdot d\vartheta \quad .$$

Dicha diferencia recibe el nombre de utilidad esperada del experimento

$$E = \{X, \mathcal{D}, p_x(\cdot/\vartheta)\} \quad .$$

Por otra parte, y con objeto de decidir sobre el valor o mérito de un experimento, es imprescindible oponer su utilidad a su costo. Supondremos en todo lo que sigue que tanto la utilidad como el costo esperados de un experimento vienen expresados en idénticas unidades para así garantizar su comparabilidad. Si expresamos por $c(E, x, \vartheta)$ el costo de realizar el experimento E y obtener el resultado x cuando el valor verdadero del parámetro es ϑ , entonces antes de su realización esperaremos un costo para E expresado a través

de la doble integral

$$\iint c(E, x, \theta) \cdot p_x(x/\theta) \cdot p_\theta(\theta) \cdot d\theta \cdot dx = \iint c(E, x, \theta) \cdot p_{\theta x}(\theta, x) \cdot d\theta \cdot dx \quad .$$

Conocidos la utilidad y costo esperados del experimento E , y denotados por $U(E)$ y $C(E)$ respectivamente, puede ya definirse el valor o mérito esperado del experimento $E = \{X, \mathcal{D}, p_x(\cdot/\theta)\}$ como:

DEFINICION. El valor o mérito esperado del experimento $E = \{X, \mathcal{D}, p_x(\cdot/\theta)\}$ viene dado por la diferencia entre su utilidad y costo esperados, esto es,

$$W(E) = U(E) - C(E) \quad .$$

Es claro que la coherencia del investigador exige que un experimento deberá realizarse siempre y cuando su valor esperado $W(E)$ sea positivo.

Notemos que la expresión de $W(E)$ es función del tamaño muestral n . En efecto, al realizar el experimento $E(n) = \{X(n), \mathcal{D}, p_{x(n)}(\cdot/\theta)\}$ y obtener el resultado $x(n)$, éste influirá sobre la expresión de la densidad posterior, apareciendo como incógnita en las dobles integrales que definen a $U(E)$ y a $C(E)$. Esta circunstancia es la que permitirá posteriormente determinar el tamaño muestral óptimo.

Por último, hemos caracterizado nuestro especial problema como uno de decisión en el que la función de utilidad debe ser una adecuada medida de la información. En este sentido es interesante destacar los resultados encontrados por Bernardo (1979) en los que pone en relación ambos conceptos, utilidad e información. Efectivamente, cuando las preferencias del investigador vienen descritas a través de la función de utilidad $u(p_\theta(\cdot), \theta)$ local, exacta y regular se demuestra que dicha función es la logarítmica para, posteriormente, demostrar que exigiendo su invarianza, la utilidad del experimento $E = \{X, \mathcal{D}, p_x(\cdot/\theta)\}$ propuesto para realizar inferencia sobre θ es proporcional a la información esperada tal como la hemos definido con anterioridad. Este resultado es fundamental y pone claramente de manifiesto cómo en los problemas de inferencia sobre el valor de una determinada cantidad aleatoria θ desconocida, la utilidad del experimento consistente en obtener una muestra para investigar su información es, precisamente, proporcional a ésta. La exigencia de que las preferencias del investigador vengan descritas por utilidades locales, regulares, exactas e invariantes es bastante suave. En efecto, ya se ha indicado que el que la utilidad sea exacta responde a la necesidad de coherencia y honradez por parte del investigador, a fin de que sus opiniones finales vengan

descritas por aquella densidad que maximiza su utilidad esperada. La exigencia de que la utilidad sea, a su vez, invariante no es sino confirmación de que ante transformaciones biyectivas de la cantidad de interés - y que, por tanto, ni enriquecen ni empobrecen para nada las hipótesis de cada problema concreto - su utilidad permanece constante. La localidad supone que la utilidad de tomar una descripción final depende únicamente de la densidad de probabilidad relacionada con el valor verdadero de θ . Y, por último, la condición de regularidad no es sino una exigencia operativa: la de existencia de segundas derivadas parciales.

En el capítulo tercero desarrollaremos tres tipos de funciones de utilidad que lo son, a su vez, de la información esperada y comprobaremos sus propiedades a fin de determinar qué propiedades de las satisfechas por la información esperada siguen siendo verificadas por ellas. Estas funciones de utilidad permitirán, aplicándolas a los dos ejemplos concretos desarrollados en el próximo capítulo, obtener tamaños muestrales óptimos al maximizar las correspondientes expresiones del valor del experimento $E(n)$.

CAPITULO 2

DOS MODELOS CONCRETOS. SUS DIFERENCIAS

En este capítulo vamos a presentar dos ejemplos elegidos por su carácter ilustrativo. El primero de ellos, estudiado ya por Lindley (1956), será mencionado sólo en sus resultados; el segundo, en cambio, se desarrollará con exhaustividad. Ambos permitirán poner de manifiesto el carácter práctico de la información esperada en esos dos casos concretos. Nos acompañarán a lo largo de la tesis y servirán de comprobación práctica a cuantas afirmaciones realicemos con posterioridad. En el capítulo cuarto calcularemos para ellos las expresiones particulares satisfechas por el tamaño muestral óptimo y de las que podrá calcularse dicho tamaño. Por otra parte, al comparar ambas informaciones esperadas deduciremos de forma práctica la no aproximatividad que, respecto al modelo normal, tiene el ejemplo desarrollado en la sección segunda.

2.1 EL MODELO NORMAL

El modelo normal - desarrollado ya por Lindley (1956), por lo que omitiremos cualquier cálculo, limitándonos a dar simplemente el resultado - es de gran importancia en cualquier desarrollo práctico en el que intervenga el valor esperado de un experimento como función de la información que proporciona. Ello es debido a que en muchas ocasiones sus resultados pueden extenderse a otros sistemas en los que las densidades de probabilidad son distintas, bastando aproximar la densidad inicial que describe las opiniones iniciales del investigador sobre el parámetro por una densidad aproximadamente normal, siempre que esto sea factible.

Supongamos, pues, una población normal $N(\vartheta, \sigma^2)$ de media ϑ y desviación típica σ conocida. La media ϑ actúa como cantidad de interés y a fin de aumentar la información que a priori se dispone de ella, se obtiene

una muestra de la población anterior de tamaño n , esto es, $x(n)$. Supongamos, a su vez, que las opiniones iniciales que el investigador tiene sobre θ vienen descritas por otra normal $N(\theta_0, \sigma_0^2)$. Bajo estas hipótesis la distribución posterior de θ es $N(\theta_n, \sigma_n^2)$, donde

$$\bar{x} = \sum_{i=1}^n x_i / n \quad , \text{ esto es, la media muestral ,}$$

$$\theta_n = (n\bar{x}\sigma_0^2 + \sigma_0^2\theta_0) / (n\sigma_0^2 + \sigma^2)$$

$$\sigma_n^2 = \sigma_0^2\sigma^2 / (n\sigma_0^2 + \sigma^2) \quad .$$

Teniendo en cuenta las definiciones dadas para la entropía y la información esperada, puede encontrarse en Lindley (1956) las siguientes igualdades:

$$H(N(\theta_0, \sigma_0^2)) = \frac{1}{2} \log(2\pi e \sigma_0^2)$$

$$E_{x(n)} H(N(\theta_n, \sigma_n^2)) = \frac{1}{2} \log(2\pi e \sigma_n^2)$$

$$I^\theta \{E(n), N(\theta_0, \sigma_0^2)\} = \frac{1}{2} \log\left(1 + n \frac{\sigma_0^2}{\sigma^2}\right)$$

Notemos que, efectivamente, la información esperada proporcionada por el experimento $E(n)$ verifica, en el modelo normal, las propiedades mencionadas para la información esperada en general. Y así, es estrictamente positiva, verifica la misma aditividad para el experimento suma que la indicada para la información esperada, y, considerada como función del tamaño muestral es creciente, al serlo la función logarítmica, y cóncava, por idéntica razón.

Acabamos de afirmar, para justificar la importancia del modelo normal, que muchas situaciones pueden reducirse a la especificada por dicho modelo aproximando la densidad inicial por otra aproximadamente normal. En efecto, mencionemos el resultado encontrado por Bernardo (1979)¹⁶ y que posteriormente aplicaremos para comparar los dos modelos citados en el presente capítulo. Es conocido que si $\psi = \psi(\theta)$ es una transformación biyectiva de la cantidad de interés se verifica

$$I^\psi \{E(n), p_\theta(\cdot)\} = I^\theta \{E(n), p_\theta(\cdot)\} \quad .$$

Pues bien, si la transformación $\psi = \psi(\theta)$ es además tal que la densidad de probabilidad a la que da lugar es aproximadamente normal de precisión h_0 (definiendo como precisión la inversa de la varianza) y bajo ciertas hipótesis adicionales sobre las densidades $p_x(\cdot/\theta)$ y $p_\theta(\cdot)$ y expresiones rela-

cionadas con ellas, entonces

$$I^{\theta} \{ E(n), p_{\theta}(\cdot) \} = \frac{1}{2} \log \left(1 + \frac{nh}{h_0} \right) + \varepsilon \quad (1)$$

donde

$$\lim_{n \rightarrow \infty} \varepsilon = 0$$

$$nh \simeq n \int_{\psi} p_{\psi}(\psi) \cdot i(\psi) \cdot d\psi \quad \Psi = \psi(\theta)$$

$$i(\psi) = - \int_{X(n)} p(x(n)/\psi) \cdot \frac{\partial^2}{\partial \psi^2} \log p(x(n)/\psi) \cdot dx(n)$$

Es evidente la importancia de este resultado. Cuando resulta difícil el cálculo de la información esperada para una cierta densidad a priori $p_{\theta}(\cdot)$, podemos conocer su comportamiento en muestras grandes, en definitiva, su comportamiento asintótico, a través de la expresión (1) y a partir de una transformación biyectiva $\psi = \psi(\theta)$, con densidad $p_{\psi}(\cdot) = T_{\psi}(p_{\theta}(\cdot))$ aproximadamente normal. De esta forma, en aquellos casos en los que sea imposible conocer la información esperada exacta, siempre se podrá disponer de una idea aproximativa a través de la información aproximada (1), siempre y cuando pueda presuponerse que la densidad $p_{\theta}(\cdot)$ pueda aproximarse, con cierto grado de bondad, por una densidad de probabilidad normal.

2.2 INFORMACION ESPERADA PARA UNA POBLACION UNIFORME Y OPINIONES INICIALES DESCRITAS POR LA DISTRIBUCION DE PARETO

La distribución de Pareto ha sido estudiada con exhaustividad (Alcaide, 1966) en Econometría, como medida de la distribución de la renta en una población de renta mínima θ_0 . Este límite mínimo puede fijarse, por ejemplo, por disposiciones sobre impuestos (Hagstroem, 1925 y 1944). Por otra parte, la distribución uniforme es representativa de todos aquellos casos en los que no existen puntos más o menos probables que los otros en un intervalo real dado. La utilización combinada de las dos distribuciones anteriores conduce a considerar el siguiente e interesante modelo: una población uniforme en el intervalo $(0, \theta)$, donde (como puede apreciarse) su límite superior es variable y constituye el parámetro sobre cuyo valor el investigador desea realizar inferencia. Precisamente sus opiniones iniciales vienen expresadas a través de la densidad de Pareto, con parámetros $\theta_0 > 0$ y $k > 0$. A fin de

disponer de mayor información, el investigador realiza el experimento $E(n) = \{X(n), \theta_0, + \infty [, 1/\theta]\}$ consistente en obtener una muestra de tamaño genérico n de la población uniforme. Es conocido (DeGroot, 1970) que bajo las hipótesis anteriores la distribución a posteriori de la inicial de Pareto es siempre otra distribución de Pareto (se trata, por tanto, de la familia conjugada de distribuciones de probabilidad para este modelo y en el sentido de Raiffa & Schlaifer, 1961), lo que hace que este sistema tenga una acusada operatividad.

Con las hipótesis anteriores demostramos el siguiente

TEOREMA 2.2.1

Siendo $x(n) = (x_1, x_2, \dots, x_n) \in X(n)$ una muestra aleatoria de una población uniforme en $(0, \theta)$; si θ se distribuye inicialmente según una Pareto, $P_{\theta}(\cdot/\theta_0, k)$, entonces

$$I^{\theta} \{E(n), P_{\theta}(\cdot/\theta_0, k)\} = \log(1 + \frac{n}{k})$$

y donde $E(n) = \{X(n), \theta_0, + \infty [, 1/\theta]\}$ es el experimento repetición n veces del consistente en obtener un dato de la población uniforme $U(0, \theta)$

Demostración.— Es conocido (DeGroot, 1970) que si obtenemos una muestra de tamaño n de la población anterior, la distribución posterior de θ , una vez obtenida dicha muestra, es una distribución de Pareto con parámetros θ_1 y $k+n$, donde

$$\theta_1 = \max(\theta_0, x_1, x_2, \dots, x_n) > 0$$

Con la notación anterior resultará, pues,

1. $p_{\theta}(\theta) = k \theta_0^k / \theta^{k+1}$ $\theta > \theta_0$
2. $p_x(x/\theta) = 1/\theta$ $0 < x < \theta$
3. $p_{x(n)}(x(n)/\theta) = 1/\theta^n$ $x(n) \in (0, \theta)^n$
4. $p_{x(n), \theta}(x(n), \theta) = k \theta_0^k / \theta^{k+n+1}$ $\theta > \theta_0, x(n) \in (0, \theta)^n$
5. $p_{\theta}(\theta/x(n)) = (k+n) \theta_1^{k+n} / \theta^{k+n+1}$ $\theta > \theta_1$

De ahí se tiene que

$$p_{x(n), \theta}(x(n), \theta) \cdot \log \frac{p_{\theta}(\theta/x(n))}{p_{\theta}(\theta)} = (k \theta_0^k / \theta^{k+n+1}) \cdot \log \frac{k+n}{k \theta_0^k} +$$

$$+ (k(k+n) \theta_0^k / \theta^{k+n+1}) \cdot \log \theta_1 - (nk \theta_0^k / \theta^{k+n+1}) \cdot \log \theta$$

De donde

$$\begin{aligned}
I^\vartheta \{E(n), P_\vartheta(\cdot/\vartheta_0, k)\} &= \int_{\vartheta_0}^{\infty} \int_{(0, \vartheta)^n} (k \vartheta_0^k / \vartheta^{k+n+1}) \cdot \log((k+n)/k \vartheta_0^k) \cdot dx(n) \cdot d\vartheta + \\
&+ \int_{\vartheta_0}^{\infty} \int_{(0, \vartheta)^n} (k \vartheta_0^k (k+n) / \vartheta^{k+n+1}) \cdot \log \vartheta_1 \cdot dx(n) \cdot d\vartheta + \\
&- \int_{\vartheta_0}^{\infty} \int_{(0, \vartheta)^n} (nk \vartheta_0^k / \vartheta^{k+n+1}) \cdot \log \vartheta \cdot dx(n) \cdot d\vartheta
\end{aligned}$$

siendo

$$\begin{aligned}
\int_{\vartheta_0}^{\infty} \int_{(0, \vartheta)^n} (k \vartheta_0^k / \vartheta^{k+n+1}) \cdot \log((k+n)/k \vartheta_0^k) \cdot dx(n) \cdot d\vartheta &= \log(1 + \frac{n}{k}) - k \cdot \log \vartheta_0, \\
\int_{\vartheta_0}^{\infty} \int_{(0, \vartheta)^n} (nk \vartheta_0^k / \vartheta^{k+n+1}) \cdot \log \vartheta \cdot dx(n) \cdot d\vartheta &= \frac{n}{k} + n \cdot \log \vartheta_0.
\end{aligned}$$

Calculemos ahora

$$\begin{aligned}
\int_{(0, \vartheta_0)^n} \log \vartheta_1 \cdot dx(n) &= \int_{\vartheta_0}^{\vartheta_0} \dots \int_{\vartheta_0}^{\vartheta_0} \log \max(\vartheta_0, x_1, x_2, \dots, x_n) \cdot dx_1 \cdot dx_2 \dots dx_n + \\
+ \sum_{p=1}^n \binom{n}{p} \int_{\vartheta_0}^{\vartheta_0} \dots \int_{\vartheta_0}^{\vartheta_0} \int_{\vartheta_0}^{\vartheta_0} \dots \int_{\vartheta_0}^{\vartheta_0} \log \max(\vartheta_0, x_1, x_2, \dots, x_n) \cdot dx_1 \cdot dx_2 \dots dx_n &= \\
&= \int_{\vartheta_0}^{\vartheta_0} \dots \int_{\vartheta_0}^{\vartheta_0} \log \vartheta_0 \cdot dx_1 \cdot dx_2 \dots dx_n + \\
+ \sum_{p=1}^n \binom{n}{p} \int_{\vartheta_0}^{\vartheta_0} \dots \int_{\vartheta_0}^{\vartheta_0} \int_{\vartheta_0}^{\vartheta_0} \dots \int_{\vartheta_0}^{\vartheta_0} \log \max(x_1, x_2, \dots, x_n) \cdot dx_1 \cdot dx_2 \dots dx_n &= \\
= \vartheta_0^n \cdot \log \vartheta_0 + \sum_{p=1}^n \binom{n}{p} \vartheta_0^{n-p} \cdot p \int_{\vartheta_0}^{\vartheta_0} \int_{\vartheta_0}^{\vartheta_0} \dots \int_{\vartheta_0}^{\vartheta_0} \log x_1 \cdot dx_1 \cdot dx_2 \dots dx_p &= \\
= \vartheta_0^n \cdot \log \vartheta_0 + \sum_{p=1}^n \binom{n}{p} \vartheta_0^{n-p} \int_{\vartheta_0}^{\vartheta_0} p(x - \vartheta_0)^{p-1} \cdot \log x \cdot dx &= \\
= \vartheta_0^n \cdot \log \vartheta_0 + \int_{\vartheta_0}^{\vartheta_0} \left\{ \sum_{p=1}^n \binom{n}{p} \vartheta_0^{n-p} \cdot p(x - \vartheta_0)^{p-1} \right\} \log x \cdot dx.
\end{aligned}$$

Calculemos el sumatorio que aparece en la última integral:

$$\sum_{p=1}^n \binom{n}{p} \vartheta_0^{n-p} \cdot p(x - \vartheta_0)^{p-1} = \sum_{p=1}^n p \cdot \binom{n}{p} \sum_{i=0}^{p-1} (-1)^i \vartheta_0^{n-p+i} \cdot \binom{p-1}{i} \cdot x^{p-1-i} \quad (2)$$

Desarrollando el doble sumatorio resultaría:

$$p=1, i=0 \quad \binom{n}{1} \binom{0}{0} \vartheta_0^{n-1}$$

$$p=2, i=0,1 \quad 2\binom{n}{2} \left\{ \binom{1}{0} \vartheta_0^{n-2} x - \binom{1}{1} \vartheta_0^{n-1} \right\}$$

$$p=3, i=0,1,2 \quad 3\binom{n}{3} \left\{ \binom{2}{0} \vartheta_0^{n-3} x^2 - \binom{2}{1} \vartheta_0^{n-2} x + \binom{2}{2} \vartheta_0^{n-1} \right\}$$

y así sucesivamente hasta

$$p=n-1, i=0,1,2,\dots,n-2 \quad (n-1)\binom{n}{n-1} \left\{ \binom{n-2}{0} \vartheta_0 x^{n-2} - \binom{n-2}{1} \vartheta_0^2 x^{n-3} + \dots + (-1)^{n-2} \binom{n-2}{n-2} \vartheta_0^{n-1} \right\}$$

$$p=n, i=0,1,2,\dots,n-1 \quad n\binom{n}{n} \left\{ \binom{n-1}{0} x^{n-1} - \binom{n-1}{1} \vartheta_0 x^{n-2} + \dots + (-1)^{n-1} \binom{n-1}{n-1} \vartheta_0^{n-1} \right\}$$

Sumando diagonalmente, puede sacarse factor común a los términos de la forma $x^{m-1} \vartheta_0^{n-m}$, resultando:

$$\begin{aligned} & \vartheta_0^{n-1} \left\{ \binom{n}{1} \binom{0}{0} - 2\binom{n}{2} \binom{1}{1} + 3\binom{n}{3} \binom{2}{2} + \dots + (-1)^{n-1} n \binom{n-1}{n-1} \binom{n}{n} \right\} + \\ & + \vartheta_0^{n-2} x \left\{ 2\binom{n}{2} \binom{1}{0} - 3\binom{n}{3} \binom{2}{1} + 4\binom{n}{4} \binom{3}{2} + \dots + (-1)^{n-2} n \binom{n-1}{n-2} \binom{n}{n} \right\} + \dots + \\ & + \vartheta_0 x^{n-2} \left\{ (n-1) \binom{n}{n-1} \binom{n-2}{0} - n \binom{n}{n} \binom{n-1}{1} \right\} + x^{n-1} n \binom{n}{n} \binom{n-1}{0} \end{aligned}$$

Salvo para el último término, la representación genérica de la suma de los números combinatorios entre corchetes admite como expresión la:

$$m \binom{n}{m} \binom{m-1}{0} - (m+1) \binom{n}{m+1} \binom{m}{1} + (m+2) \binom{n}{m+2} \binom{m+1}{2} + \dots + (-1)^{n-1} n \binom{n}{n} \binom{n-1}{n-m}$$

donde m la suponemos constante en la anterior expresión. La anterior suma admite la siguiente representación abreviada:

$$\begin{aligned} & \sum_{i=0}^{n-m} (-1)^i \binom{m+i}{m+i} \binom{n}{m+i} \binom{m+i-1}{i} = \sum_{i=0}^{n-m} (-1)^i n \binom{n-1}{m+i-1} \binom{m+i-1}{i} = \\ & = n \sum_{i=0}^{n-m} (-1)^i \frac{(n-1)!(m+i-1)!}{(m+i-1)!(n-m-i)!i!(m-1)!} = n \frac{(n-1)!}{(m-1)!} \sum_{i=0}^{n-m} (-1)^i \frac{1}{i!(n-m-i)!} = \\ & = n \frac{(n-1)!}{(m-1)!(n-m)!} \sum_{i=0}^{n-m} (-1)^i \frac{(n-m)!}{(n-m-i)!i!} = n \binom{n-1}{m-1} \sum_{i=0}^{n-m} (-1)^i \binom{n-m}{i} = 0 \end{aligned}$$

Así pues, el doble sumatorio (2) se reduce al término correspondiente a $p=n$, $i=0$, esto es, al término

$$n \binom{n}{n} \binom{n-1}{0} x^{n-1} = n x^{n-1} \quad .$$

De ahí que

$$\int_{(\theta, \theta)^n} \log \vartheta_1 \cdot dx(n) = \vartheta_0^n \cdot \log \vartheta_0 + \int_{\vartheta_0}^{\vartheta} n x^{n-1} \cdot \log x \cdot dx = \vartheta^n \cdot \log \vartheta - \frac{1}{n} (\vartheta^n - \vartheta_0^n) .$$

Finalmente,

$$\begin{aligned} & \int_{\vartheta_0}^{\infty} \frac{k(k+n) \vartheta_0^k}{\vartheta^{k+n+1}} \left\{ \int_{(\theta, \theta)^n} \log \vartheta_1 \cdot dx(n) \right\} d\vartheta = \\ & = \int_{\vartheta_0}^{\infty} \frac{k(k+n) \vartheta_0^k}{\vartheta^{k+n+1}} \left(\vartheta^n \log \vartheta - \frac{1}{n} (\vartheta^n - \vartheta_0^n) \right) \cdot d\vartheta = n \left(\log \vartheta_0 + \frac{1}{k} \right) + k \cdot \log \vartheta_0 \end{aligned}$$

Uniendo ya todos los resultados parciales obtenidos, resulta en último término:

$$\begin{aligned} I^{\vartheta} \{ E(n), P_{\vartheta}(\cdot / \vartheta_0, k) \} &= \log \left(1 + \frac{n}{k} \right) - k \cdot \log \vartheta_0 - \frac{n}{k} - n \cdot \log \vartheta_0 + \\ &+ n \left(\log \vartheta_0 + \frac{1}{k} \right) + k \cdot \log \vartheta_0 = \log \left(1 + \frac{n}{k} \right) \end{aligned}$$

(c.q.d.)

Puede representarse gráficamente a $I^{\vartheta} \{ E(n), P_{\vartheta}(\cdot / \vartheta_0, k) \}$ en función de k y de n , y para distintos valores de n y k , respectivamente. Es evidente que la información esperada proporcionada por $E(n)$ debe ser mayor, para cada k , que la dada por $E(n-1)$, como queda recogido en la figura 1.

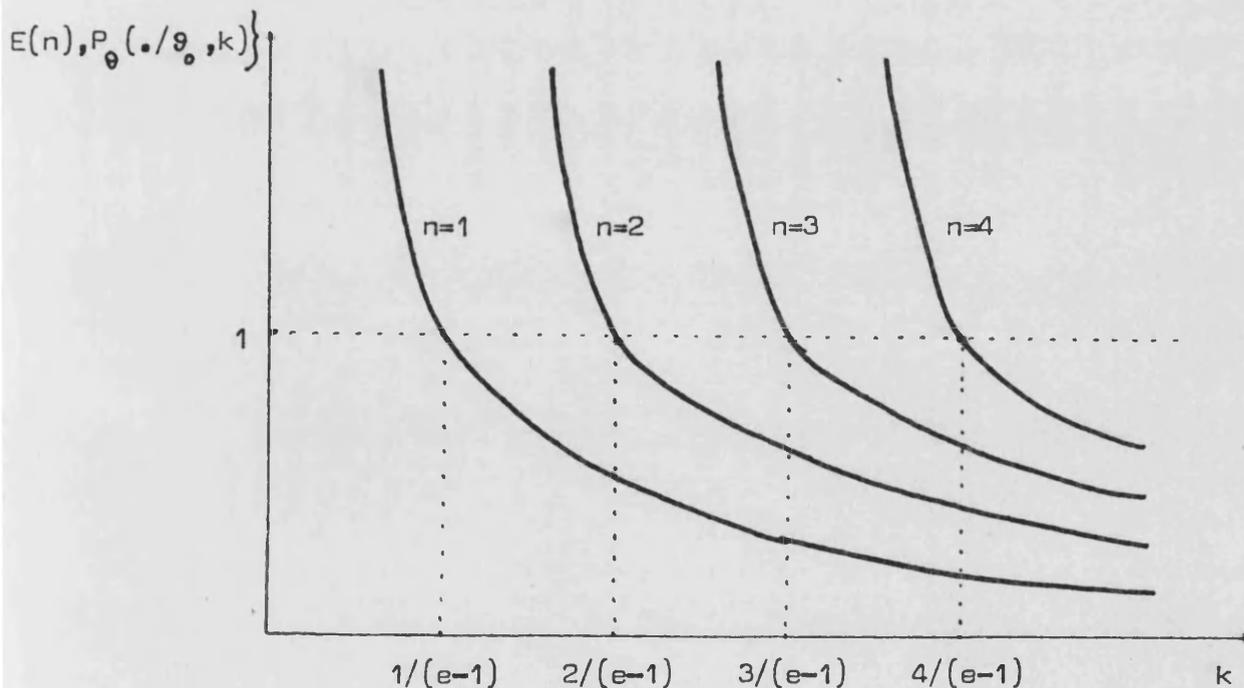


Figura 1.- Representación de $I^{\vartheta} \{ E(n), P_{\vartheta}(\cdot / \vartheta_0, k) \}$ para distintos valores de n , como función del parámetro k de la distribución a priori.

El carácter creciente y cóncavo de la información esperada como función del tamaño muestral (supuestamente continuo) queda recogido en la figura 2.

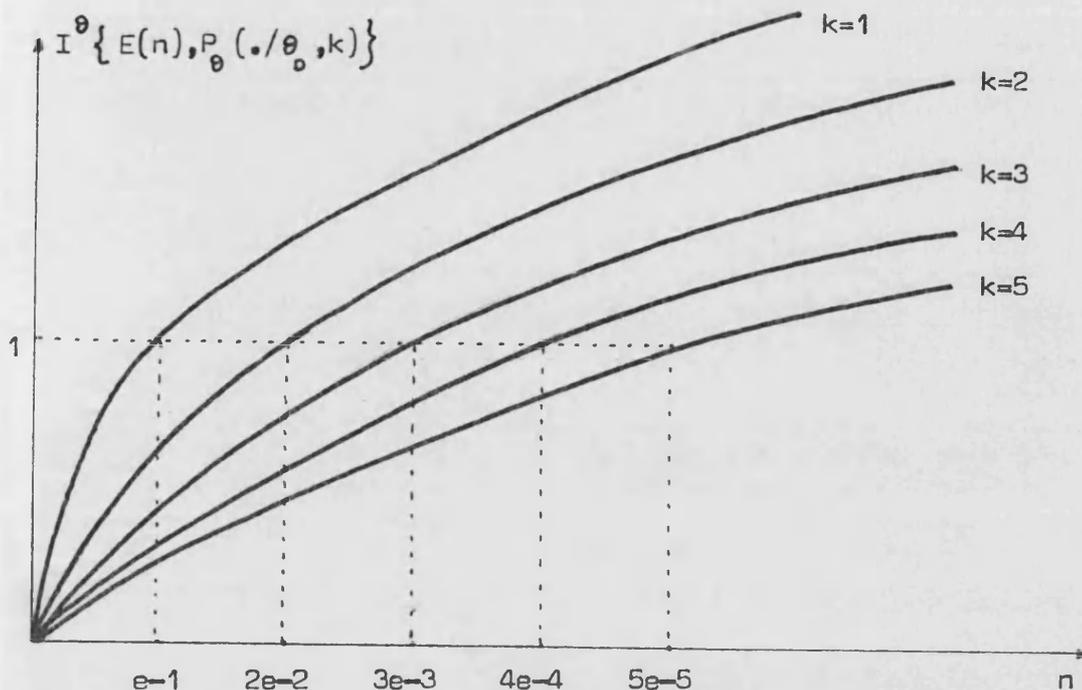
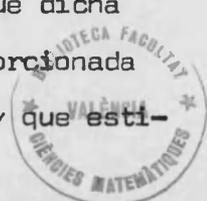


Figura 2.- Representación de $I^{\vartheta} \{ E(n), P_{\vartheta}(\cdot/\vartheta_0, k) \}$ para distintos valores de k , como función del tamaño muestral n .

La figura 3, que no es sino la superposición, en tres dimensiones, de las dos primeras figuras, es la representación gráfica de la superficie información esperada en función de las dos variables n y k que intervienen en su definición.

Un hecho interesante a destacar, deducido del anterior teorema, consiste en que la información esperada total obtenida tras realizar el experimento $E(n)$ sobre la población $U(0, \vartheta)$ y cuando ϑ se distribuye inicialmente según una densidad de Pareto con parámetros ϑ_0 y k , no depende del parámetro ϑ_0 o límite inferior de los valores de ϑ . Este hecho, análogo al ocurrido en el modelo normal respecto la media de la distribución inicial, supone que dicho parámetro es irrelevante a la hora del posible diseño de un experimento que venga expresado en términos de dependencia única de la información esperada. Precisamente, en estudios sobre la distribución de la renta y cuando ésta viene representada a partir de la distribución de Pareto, existe cierta dificultad en determinar esta cota inferior. Por ello, el que dicha cota sea irrelevante en el cálculo de la información esperada proporcionada por la muestra uniforme es interesante, toda vez que únicamente hay que estimar el valor del parámetro k .



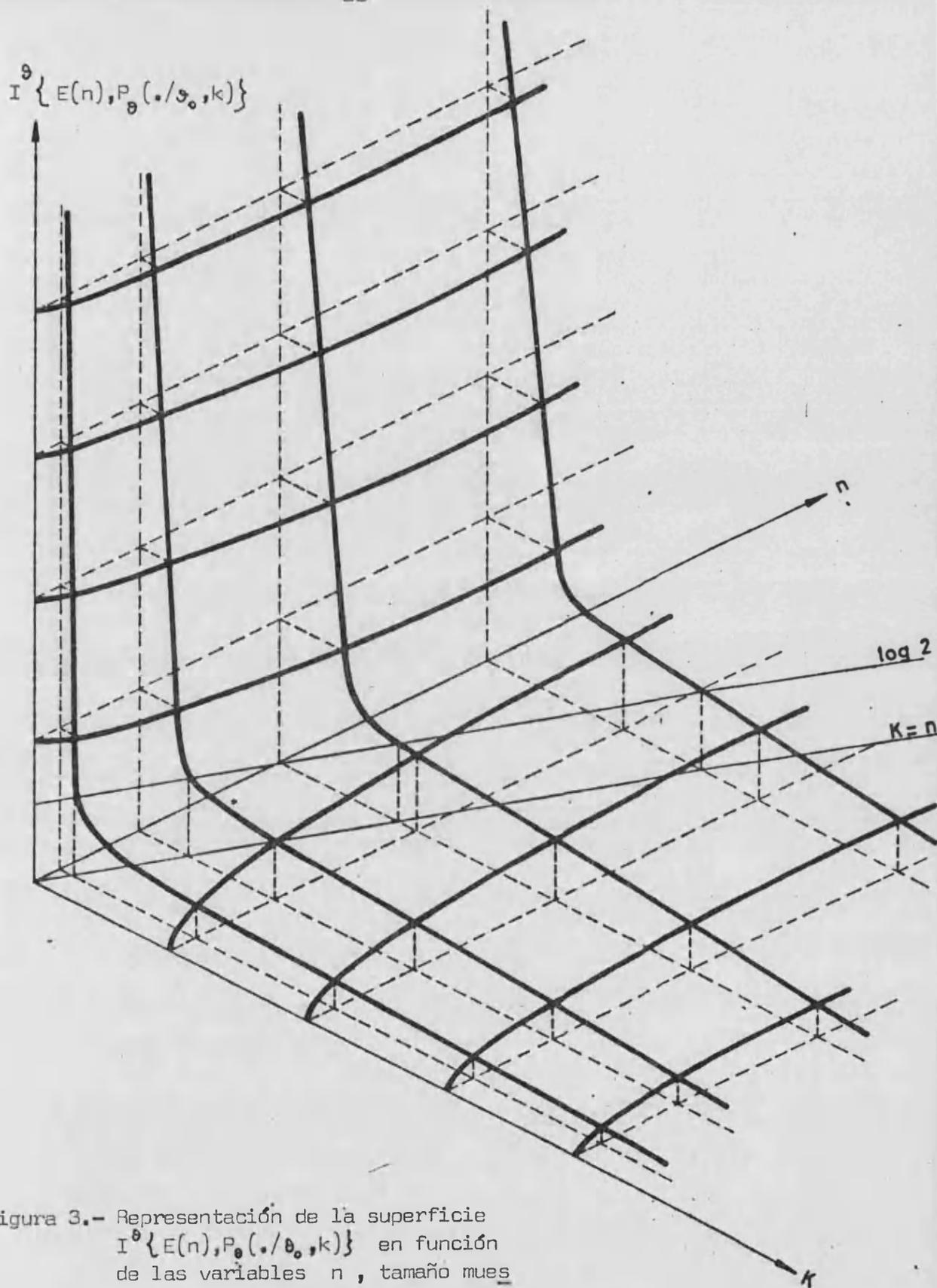


Figura 3.- Representación de la superficie $I^{\theta} \{E(n), P_{\theta}(\cdot/\theta_0, k)\}$ en función de las variables n , tamaño muestral, y k , parámetro de la distribución inicial.

Resulta evidente, pues, el siguiente

COROLARIO 2.2.1

Si $x(n) = (x_1, x_2, \dots, x_n)$ es una muestra aleatoria de una población uniforme en $(0, \theta)$, con θ desconocido; si las opiniones iniciales sobre ϑ se describen por una distribución perteneciente a la familia de distribuciones de Pareto, con parámetro fijo $k > 0$, y siendo ϑ_0 cualesquiera, entonces la información esperada proporcionada por el experimento $E(n) = \{X(n), P_{\vartheta_0}(\cdot/\cdot, k)\}$, viene dada por

$$I^{\vartheta} \{E(n), P_{\vartheta_0}(\cdot/\cdot, k)\} = \log\left(1 + \frac{n}{k}\right)$$

Teniendo en cuenta que la entropía para una distribución de Pareto viene dada por

$$H(P_{\vartheta_0}(\cdot/\vartheta_0, k)) = \int_{\vartheta_0}^{\infty} \frac{k \vartheta_0^k}{\vartheta^{k+1}} \log \frac{k \vartheta_0^k}{\vartheta^{k+1}} \cdot d\vartheta = \log(\vartheta_0/k) + \frac{1+k}{k}$$

obtenemos por diferencia entre la información esperada y la entropía que la entropía esperada para una densidad de Pareto, con parámetros ϑ_0 y k , admite como formulación la siguiente:

$$E_{x(n)} H(P_{\vartheta_0}(\cdot/\vartheta_0, k)) = \log(\vartheta_0/(k+n)) + \frac{1+k}{k}$$

y donde la muestra $x(n)$ es extraída de una población uniforme en $(0, \theta)$. No tenemos que, efectivamente, la información esperada proporcionada por $E(n)$ verifica en este modelo de población uniforme y densidad inicial de Pareto, las propiedades citadas para una información esperada en general. Y así, la información $I^{\vartheta} \{E(n), P_{\vartheta_0}(\cdot/\vartheta_0, k)\}$ es estrictamente positiva, verifica idéntica aditividad para el experimento suma que la indicada para la información esperada y, considerada como función del tamaño muestral n , es creciente y cóncava al serlo la función logaritmo.

Finalmente, la tabla de la siguiente página contiene una serie de valores de $I^{\vartheta} \{E(n), P_{\vartheta_0}(\cdot/\vartheta_0, k)\}$ para valores concretos de n y k .

VALORES DE $I^{\theta} \{E(n), P_{\theta}(\cdot/\theta, k)\}$ PARA CIERTOS VALORES DE n Y k

n \ k	0'1	0'2	0'3	0'4	0'5	0'75	1	2'5	5	7'5	10	25	50	75	100
1	2'3979	1'7918	1'4663	1'2528	1'0986	0'8473	0'6931	0'3364	0'1823	0'1252	0'0953	0'0392	0'0198	0'0132	0'0100
2	3'0445	2'3979	2'0369	1'7918	1'6094	1'2993	1'0986	0'5878	0'3364	0'2364	0'1823	0'0770	0'0392	0'0263	0'0198
3	3'4340	2'7726	2'3979	2'1401	1'9459	1'6094	1'3863	0'7865	0'4700	0'3365	0'2624	0'1133	0'0583	0'0392	0'0296
4	3'7136	3'0445	2'6626	2'3979	2'1972	1'8458	1'6094	0'9555	0'5878	0'4274	0'3365	0'1484	0'0770	0'0520	0'0392
5	3'9318	3'2581	2'8717	2'6027	2'3979	2'0369	1'7918	1'0986	0'6931	0'5108	0'4055	0'1823	0'0953	0'0645	0'0488
6	4'1109	3'4340	3'0445	2'7726	2'5649	2'1972	1'9459	1'2238	0'7885	0'5878	0'4700	0'2151	0'1133	0'0770	0'0583
7	4'2627	3'5835	3'1918	2'9178	2'7081	2'3354	2'0794	1'3350	0'8755	0'6592	0'5306	0'2469	0'1310	0'0892	0'0677
8	4'3944	3'7136	3'3202	3'0445	2'8332	2'4567	2'1972	1'4351	0'9555	0'7259	0'5878	0'2776	0'1484	0'1014	0'0770
9	4'5109	3'8286	3'4340	3'1570	2'9444	2'5649	2'3026	1'5261	1'0296	0'7885	0'6419	0'3075	0'1655	0'1133	0'0862
10	4'6151	3'9318	3'5361	3'2581	3'0445	2'6626	2'3979	1'6094	1'0986	0'8473	0'6931	0'3364	0'1823	0'1252	0'0953
20	5'3033	4'6151	4'2146	3'9318	3'7136	3'3202	3'0445	2'1972	1'6094	1'2993	1'0986	0'5878	0'3364	0'2364	0'1823
30	5'7071	5'0173	4'6151	4'3307	4'1109	3'7136	3'4340	2'5649	1'9459	1'6094	1'3863	0'7885	0'4700	0'3365	0'2624
40	5'9940	5'3033	4'9003	4'6151	4'3944	3'9951	3'7136	2'8332	2'1972	1'8458	1'6094	0'9555	0'5878	0'4274	0'3365
50	6'2166	5'5255	5'1220	4'8363	4'6151	4'2146	3'9318	3'0445	2'3979	2'0369	1'7918	1'0986	0'6931	0'5108	0'4055
60	6'3986	5'7071	5'3033	5'0173	4'7958	4'3944	4'1109	3'2189	2'5649	2'1972	1'9459	1'2238	0'7885	0'5878	0'4700
70	6'5525	5'8608	5'4567	5'1704	4'9488	4'5468	4'2627	3'3673	2'7081	2'3354	2'0794	1'3350	0'8755	0'6592	0'5306
80	6'6859	5'9940	5'5897	5'3033	5'0814	4'6790	4'3944	3'4965	2'8322	2'4567	2'1972	1'4351	0'9555	0'7259	0'5878
90	6'8035	6'1115	5'7071	5'4205	5'1985	4'7958	4'5109	3'6109	2'9444	2'5649	2'3026	1'5261	1'0296	0'7885	0'6419
100	6'9088	6'2166	5'8121	5'5255	5'3033	4'9003	4'6151	3'7136	3'0445	2'6626	2'3979	1'6094	1'0986	0'8473	0'6931
200	7'6014	6'9088	6'5038	6'2166	5'9940	5'5897	5'3033	4'3944	3'7136	3'3202	3'0445	2'1972	1'6094	1'2993	1'0986
300	8'0067	7'3139	6'9088	6'6214	6'3986	5'9940	5'7071	4'7958	4'1109	3'7136	3'4340	2'5649	1'9459	1'6094	1'3863
400	8'2943	7'6014	7'1962	6'9088	6'6859	6'2810	5'9940	5'0814	4'3944	3'9951	3'7136	2'8332	2'1972	1'8458	1'6094
500	8'5174	7'8224	7'4192	7'1317	6'9088	6'5038	6'2166	5'3033	4'6151	4'2146	3'9318	3'0445	2'3979	2'0369	1'7918
600	8'6997	8'0067	7'6014	7'3139	7'0909	6'6859	6'3986	5'4848	4'7958	4'3944	4'1109	3'2189	2'5649	2'1972	1'9459
700	8'8538	8'1608	7'7555	7'4679	7'2449	6'8398	6'5525	5'6384	4'9488	4'5468	4'2627	3'3673	2'7081	2'3354	2'0794
800	8'9873	8'2943	7'8890	7'6014	7'3784	6'9732	6'6859	5'7714	5'0814	4'6790	4'3944	3'4965	2'8332	2'4567	2'1972
900	9'1051	8'4121	8'0067	7'7191	7'4961	7'0909	6'8035	5'8889	5'1985	4'7958	4'5109	3'6109	2'9444	2'5649	2'3026
1000	9'2104	8'5174	8'1120	7'8224	7'6014	7'1962	6'9088	5'9940	5'3033	4'9003	4'6151	3'7136	3'0445	2'6626	2'3979

2.3 APROXIMACION A AMBOS MODELOS. SUS DIFERENCIAS

Desarrollado el modelo anterior, al compararlo con el modelo normal destaca claramente el que las expresiones obtenidas para las informaciones esperadas son de gran sencillez en ambos, reduciéndose a simples funciones logarítmicas. Esta circunstancia es tanto más patente cuando comparamos estos resultados con los de otros ejemplos, como puede ser el modelo binomial desarrollado por Lindley (1956), Bernardo (1975) y Basulto & Bernardo (1978). En los trabajos de los autores citados pueden encontrarse expresiones mucho más complejas - y, por ello, menos operativas - para la información esperada.

El problema queda patentizado con gran énfasis cuando nos encontramos poblaciones y densidades iniciales para las que el cálculo de la información esperada se hace extraordinariamente difícil. Para resolver esta circunstancia hay que buscar procedimientos indirectos, como pueden ser, por ejemplo, el estudio de una información esperada ante una transformación biyectiva de la cantidad de interés o aproximar la información esperada por la obtenida mediante transformaciones biyectivas de la cantidad de interés que sean aproximadamente normales.

El primero de estos caminos permite la obtención exacta y no aproximada de la información esperada. Supongamos, por tanto, que dada la población que se distribuye con densidad $p_x(./\theta)$, el cálculo de la información esperada para una densidad inicial $p_\theta(.)$ resulta extraordinariamente complejo. Sin embargo, si se logra encontrar una transformación biyectiva de la cantidad de interés, la $\psi = \psi(\theta)$, tal que la correspondiente densidad inicial transformada

$$p_\psi(.) = T_\psi(p_\theta(.))$$

haga que el cálculo de $I^\psi\{E(n), p_\theta(.)\}$ sea conocido, el problema estaría resuelto al verificarse, en este caso de transformación biyectiva, que las informaciones esperadas útil y total son iguales. De esta forma podría conocerse con total exactitud la información esperada que, calculada directamente, era de problemática obtención.

Por ejemplo, supongamos una población que se distribuye uniformemente en el intervalo $(0, \psi_0^\theta)$, siendo θ la cantidad de interés. Si θ se distribuye exponencialmente con parámetro k , esto es, si

$$p_{\theta}(\theta) = k \cdot \exp(-k\theta) \quad \theta > 0$$

anulándose la densidad para valores negativos del parámetro, es conocido (De Groot, 1970) que la transformación biyectiva $\varphi = \varphi_0 e^{\theta}$ proporciona una densidad inicial para φ dada por

$$p_{\varphi}(\varphi) = k \varphi_0^k / \varphi^{k+1}, \quad \varphi > \varphi_0$$

esto es, una densidad de Pareto con parámetros φ_0 y k . Denotando por $E(k)$ a la distribución exponencial de parámetro k , siendo $p_x(\cdot/\theta)$ una densidad de probabilidad uniforme en $(0, \varphi_0 e^{\theta})$, realizado el experimento $E(n)$:

$$E(n) = \left\{ (0, \varphi_0 e^{\theta}), R^+, E(k) \right\},$$

puede afirmarse que

$$I^{\theta} \left\{ E(n), E(k) \right\} = \log \left(1 + \frac{n}{k} \right)$$

que no es sino la información esperada deducida en la sección anterior.

Este procedimiento de obtención exacta de informaciones esperadas a partir de otras conocidas es de una gran elegancia, aunque a veces resulta de difícil aplicación dado que debe conjugarse el comportamiento tanto de la densidad inicial como el de la población sobre la que se extrae la muestra. De ahí que haya que buscar procedimientos alternativos que, si bien no proporcionan una información esperada exacta, sí pueden admitirse como aproximaciones asintóticas en muestras grandes. Efectivamente, cuando resulta difícil, sino imposible, el cálculo de la información esperada para una cierta densidad a priori $p_{\theta}(\cdot)$, será interesante conocer, al menos, su comportamiento aproximado en muestras grandes. En definitiva, su comportamiento asintótico. Por ejemplo, esta es la idea expresada en las relaciones (1), donde puede apreciarse la similitud de dicho resultado

$$\frac{1}{2} \log \left(1 + n \frac{h}{h_0} \right)$$

con la información esperada proporcionada por el modelo normal

$$\frac{1}{2} \log \left(1 + n \frac{\sigma_0^2}{\sigma^2} \right)$$

toda vez que h_0 no es sino la precisión ($h_0 = 1/\sigma_0^2$) de la densidad aproximadamente normal.

Ibragimov & Has'Minsky (1973) obtienen un importante resultado a fin de aproximar una información esperada mediante expresiones asintóticas cuan

do el tamaño muestral es grande, exigiendo solamente ciertas condiciones de regularidad a ciertas expresiones relacionadas con las densidades $p_{\theta}(\cdot)$ y $p_x(\cdot/\theta)$. Suponiendo además transformaciones biyectivas aproximadamente normales de la cantidad de interés, Bernardo (1979)⁸ obtiene la aproximación asintótica expresada por el conjunto de fórmulas (1), mencionadas ya al hablar de la importancia del modelo normal en la primera sección de este capítulo.

De la comparación de los dos modelos desarrollados observamos la existencia de una diferencia fundamental entre ellos que, además, vulnera las condiciones necesarias exigidas por Ibragimov & Has'Minsky (1973) para poder aproximar en muestras grandes la información esperada $I^{\theta}\{E(n), P_{\theta}(\cdot/\theta_0, k)\}$ mediante alguna de las fórmulas deducidas por dichos autores. En efecto, a diferencia de, por ejemplo, el modelo normal, en el segundo de los ejemplos el recinto en el que toma valores el dato x depende del parámetro θ , estando, por ello, extendidas las integrales correspondientes a límites de integración que son funciones de θ , tal como puede apreciarse de la sección anterior. La especial definición del modelo uniforme con densidad inicial de Pareto no verifica, por ejemplo, las condiciones necesarias de continuidad uniforme exigidas para la aproximatividad de las expresiones encontradas por Ibragimov & Has'Minsky (1973). Este resultado queda comprobado de forma efectiva mediante la demostración de los teoremas siguientes, donde se intenta aplicar las fórmulas (1) al modelo de la sección segunda. De esta forma quedará comprobado que aproximando la información $I^{\theta}\{E(n), P_{\theta}(\cdot/\theta_0, k)\}$ mediante (1) se obtiene, en el límite (esto es, con muestras extraordinariamente grandes) una información esperada aproximada que es mitad de la obtenida directamente.

TEOREMA 2.3.1

En una población uniforme en $(0, \theta)$ se describen las opiniones iniciales sobre el parámetro θ a través de una distribución de Pareto, de parámetros θ_0 y k . Para cualquier transformación biyectiva $\psi = \psi(\theta)$ de la cantidad de interés, la constante h que aparece en la expresión de la aproximación (1) es:

$$h \simeq -\frac{1}{k} \int \frac{p_{\psi}(\psi) \cdot p'_{\psi}(\psi)}{\int_{-\infty}^{\psi} p_{\psi}(t) dt} \cdot d\psi + \frac{1}{k} \int \frac{p_{\psi}(\psi)^3}{\left(\int_{-\infty}^{\psi} p_{\psi}(t) dt\right)^2} \cdot d\psi \quad (3)$$

donde: $p_{\psi}(\cdot)$ es la coorespondiente densidad inicial de la nueva canti-

dad de interés ψ , deducida de la transformación $\psi = \psi(\theta)$ aplicada sobre $p_{\theta}(\cdot)$; las integrales anteriores están extendidas a $\mathcal{Y} = \psi(\mathcal{D})$, y

$$p'_{\psi}(\psi) = \frac{d}{d\psi} (p_{\psi}(\psi)) \quad .$$

Demostración.- Denotemos por $\theta = f(\psi)$ a la inversa de la transformación biyectiva $\psi = \psi(\theta)$. Supongamos que la densidad inicial para la nueva cantidad de interés ψ es $p_{\psi}(\cdot)$, donde

$$p_{\psi}(\psi) = p_{\theta}(f(\psi)) \cdot \left| \frac{df(\psi)}{d\psi} \right| = \frac{k \theta_0^k}{(f(\psi))^{k+1}} \cdot \frac{df(\psi)}{d\psi}$$

de donde

$$-\theta_0^k / f(\psi)^k = \int_{-\infty}^{\psi} p_{\psi}(t) \cdot dt$$

de forma que la transformación biyectiva $\theta = f(\psi)$ debe verificar que

$$f(\psi) = -\theta_0 / \left(\int_{-\infty}^{\psi} p_{\psi}(t) \cdot dt \right)^{1/k} \quad .$$

Dado que

$$i(\psi) = - \int p(x/\psi) \cdot \frac{\partial^2}{\partial \psi^2} \log p(x/\psi) \cdot dx \quad , y$$

$$p_x(x/\theta) = 1/\theta \quad \text{en } 0 < x < \theta$$

se deduce que

$$p(x/\psi) = - \left(\int_{-\infty}^{\psi} p_{\psi}(t) \cdot dt \right)^{1/k} / \theta_0$$

y que, por tanto, $p(x/\psi)$ no depende de x . Por ello,

$$\begin{aligned} i(\psi) &= - \int p_x(x/\psi) \cdot \frac{\partial^2}{\partial \psi^2} \log \left(- \frac{\left(\int_{-\infty}^{\psi} p_{\psi}(t) \cdot dt \right)^{1/k}}{\theta_0} \right) \cdot dx = \\ &= \frac{\partial^2}{\partial \psi^2} \log \left(- \frac{\left(\int_{-\infty}^{\psi} p_{\psi}(t) \cdot dt \right)^{1/k}}{\theta_0} \right) = - \frac{\partial}{\partial \psi} \left(\frac{1}{k} \cdot \frac{p_{\psi}(\psi)}{\int_{-\infty}^{\psi} p_{\psi}(t) \cdot dt} \right) = \\ &= - \frac{1}{k} \cdot \frac{p'_{\psi}(\psi) \int_{-\infty}^{\psi} p_{\psi}(t) \cdot dt - (p_{\psi}(\psi))^2}{\left(\int_{-\infty}^{\psi} p_{\psi}(t) \cdot dt \right)^2} \end{aligned}$$

de donde

$$h \approx \int p_{\psi}(\psi) \cdot i(\psi) \cdot d\psi = -\frac{1}{k} \int \frac{p_{\psi}(\psi) \cdot p_{\psi}'(\psi)}{\int_{-\infty}^{\psi} p_{\psi}(t) \cdot dt} d\psi + \frac{1}{k} \int \frac{p_{\psi}(\psi)^3}{\left(\int_{-\infty}^{\psi} p_{\psi}(t) \cdot dt\right)^2} d\psi$$

(c.q.d.)

Suponiendo válida la aproximatividad expresada por (1), siempre que h fuese finito la información esperada $I^{\psi} \{E(n), P_{\psi}(\cdot/\psi, k)\}$ podría aproximarse por la expresión

$$\frac{1}{2} \log\left(1 + n \frac{h}{h_0}\right)$$

en la que h_0 , que es la precisión de la densidad $p_{\psi}(\cdot)$ aproximadamente normal, es también una cantidad finita. Sin embargo, la bondad de esta aproximación - como era de esperar al no verificarse todas las condiciones necesarias exigidas - es realmente burda, como queda puesto de manifiesto al compararla con la información esperada exacta en el límite

$$\lim_{n \rightarrow \infty} \frac{\log\left(1 + \frac{n}{k}\right)}{\frac{1}{2} \log\left(1 + n \frac{h}{h_0}\right)} = 2$$

independientemente de cualquier conjunto de valores finitos para k , h y h_0 (como comprobaremos ahora mismo, la finitud de h quedará garantizada por el teorema 2.3.2). Por otra parte, el límite anterior indica que la información exacta en el modelo de población uniforme y densidad inicial de Pareto es, con la exactitud que se desee sin más que considerar tamaños muestrales suficientemente grandes, doble que la información proporcionada por la aproximación (1). Este resultado confirma la imposibilidad de aproximar la información esperada del modelo de la sección segunda mediante las fórmulas de aproximación (1).

TEOREMA 2.3.2

Bajo las hipótesis del Teorema 2.3.1, y siendo $p_{\psi}(\cdot)$ una densidad aproximadamente normal, se deduce que la constante h de la expresión (3) es finita.

Demostración.- Dado que $p_{\psi}(\cdot)$ es aproximadamente normal, puede suponerse que lo es según una $N(0, 1)$, bastando para ello normalizar la

nueva cantidad de interés, ψ . Por tanto, puede escribirse:

$$p_{\psi}(\psi) \approx \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\psi^2} \quad F(\psi) = \int_{-\infty}^{\psi} p_{\psi}(t) \cdot dt \approx \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\psi} e^{-\frac{1}{2}t^2} \cdot dt$$

Son evidentes las igualdades:

$$F'(\psi) = \frac{d}{d\psi} F(\psi) = p_{\psi}(\psi) \quad \text{y} \quad p'_{\psi}(\psi) = \frac{d}{d\psi} p_{\psi}(\psi) = -\psi \cdot p_{\psi}(\psi)$$

Así pues, la expresión (3) admite la siguiente formulación:

$$\begin{aligned} & -\frac{1}{k} \left(\int_{-\infty}^{\infty} \frac{p_{\psi}(\psi) \cdot p'_{\psi}(\psi)}{F(\psi)} d\psi - \int_{-\infty}^{\infty} \frac{p_{\psi}(\psi)^3}{F(\psi)^2} d\psi \right) = \\ & = -\frac{1}{k} \int_{-\infty}^{\infty} p_{\psi}(\psi) \cdot \frac{d}{d\psi} \left(\frac{p_{\psi}(\psi)}{F(\psi)} \right) \cdot d\psi \end{aligned}$$

Calculemos por partes la integral anterior:

$$\begin{aligned} & \int_{-\infty}^{\infty} p_{\psi}(\psi) \cdot \frac{d}{d\psi} \left(\frac{p_{\psi}(\psi)}{F(\psi)} \right) \cdot d\psi = \lim_{a \rightarrow \infty} \int_{-a}^a p_{\psi}(\psi) \cdot \frac{d}{d\psi} \left(\frac{p_{\psi}(\psi)}{F(\psi)} \right) \cdot d\psi = \\ & = \lim_{a \rightarrow \infty} \left(\frac{p_{\psi}(\psi)^2}{F(\psi)} \right) \Big|_{-a}^{+a} - \lim_{a \rightarrow \infty} \int_{-a}^a p'_{\psi}(\psi) \cdot \frac{p_{\psi}(\psi)}{F(\psi)} d\psi \end{aligned}$$

Estudiemos el límite de la expresión entre corchetes:

$$\lim_{a \rightarrow \infty} \left(\frac{p_{\psi}(a)^2}{F(a)} \right) = 0$$

$$\lim_{a \rightarrow -\infty} \frac{p_{\psi}(a)^2}{F(a)} = \lim_{a \rightarrow -\infty} (-2a \cdot p_{\psi}(a)) = \lim_{a \rightarrow -\infty} \left(-\frac{1}{\sqrt{2\pi}} 2a \cdot e^{-\frac{1}{2}a^2} \right) = 0$$

De ahí que

$$\begin{aligned} & -\frac{1}{k} \left(\int_{-\infty}^{\infty} \frac{p_{\psi}(\psi) \cdot p'_{\psi}(\psi)}{F(\psi)} d\psi - \int_{-\infty}^{\infty} \frac{p_{\psi}(\psi)^3}{F(\psi)^2} d\psi \right) = \\ & = \frac{1}{k} \int_{-\infty}^{\infty} \frac{p_{\psi}(\psi) \cdot p'_{\psi}(\psi)}{F(\psi)} d\psi \end{aligned}$$

esto es,

$$\int_{-\infty}^{\infty} \frac{p_{\psi}(\psi)^3}{F(\psi)^2} d\psi = 2 \int_{-\infty}^{\infty} \frac{p_{\psi}(\psi) \cdot p'_{\psi}(\psi)}{F(\psi)} d\psi .$$

Mostrando que una cualquiera de las integrales anteriores es finita (por ejemplo, la primera), quedará demostrada la tesis del teorema.

$$\begin{aligned} \lim_{\psi \rightarrow \infty} \frac{p_{\psi}(\psi)^3}{(1/\psi^2)F(\psi)^2} &= \lim_{\psi \rightarrow \infty} \frac{\psi^2 p_{\psi}(\psi)^3}{F(\psi)^2} = \\ &= \lim_{\psi \rightarrow \infty} \frac{2\psi p_{\psi}(\psi)^3 - 3\psi^3 p_{\psi}(\psi)^3}{2p_{\psi}(\psi)F(\psi)} = \lim_{\psi \rightarrow \infty} \frac{p_{\psi}(\psi)^2(2\psi - 3\psi^3)}{2F(\psi)} = \\ &= \lim_{\psi \rightarrow \infty} \frac{-2\psi(p_{\psi}(\psi))^2(2\psi - 3\psi^3) + (2 - 9\psi^2)(p_{\psi}(\psi))^2}{2p_{\psi}(\psi)} = \\ &= \lim_{\psi \rightarrow \infty} p_{\psi}(\psi) \cdot Q(\psi) = 0 \end{aligned}$$

dado que $Q(\psi)$ es un polinomio real de cuarto grado. Entonces, dado un cierto $\varepsilon = 1$, puede afirmarse que existe un $M > 0$ tal que:

$$\left| \psi^2 \frac{p_{\psi}(\psi)^3}{F(\psi)^2} \right| < 1 \quad \text{si} \quad |\psi| \geq M ;$$

y puesto que tanto $p_{\psi}(\psi)$ como $F(\psi)$ son positivos en toda la recta real, se deduce que

$$\frac{p_{\psi}(\psi)^3}{F(\psi)^2} < \frac{1}{\psi^2} \quad \text{si} \quad |\psi| \geq M .$$

Como a su vez la integral

$$\int_M^{\infty} \frac{d\psi}{\psi^2}$$

es convergente, al ser

$$\lim_{a \rightarrow +\infty} \int_M^a \frac{d\psi}{\psi^2} = \lim_{a \rightarrow +\infty} \left(\frac{1}{M} - \frac{1}{a} \right) = \frac{1}{M}$$

deducimos que $p_{\psi}(\psi)^3 / F(\psi)^2$ es integrable en $[M, +\infty[$.

Análogamente se demuestra que el anterior cociente es integrable en $] -\infty, -M]$. El teorema queda, pues, demostrada sin más que considerar que

la expresión $p_{\psi}(\psi)^3 / F(\psi)^2$ es continua en $[-M, +M]$, tomando en dicho intervalo valores distintos de cero y positivos, por lo que es integrable en él. Por tanto, existen las integrales

$$\int_{-\infty}^{\infty} \frac{p_{\psi}(\psi)^3}{F(\psi)^2} d\psi \quad \text{y} \quad \int_{-\infty}^{\infty} \frac{p_{\psi}(\psi) \cdot p'_{\psi}(\psi)}{F(\psi)} d\psi$$

y, por ello, la constante h que aparece en la aproximación (1) es finita.

(c.q.d.)

Análogamente al resultado encontrado por Bernardo (1979)^{**}, y ante la sencillez de la expresión de $I^{\theta} \{E(n), P_{\theta}(\cdot/\theta), k\}$, queda como posible problema abierto la consideración del segundo modelo de este capítulo en un papel semejante al representado por el modelo normal para todos aquellos casos que, cumpliendo las hipótesis de regularidad exigidas por Ibragimov & Has'Minsky (1973), tienen definidos un valor muestral x cuyo recinto de variación es independiente del valor paramétrico θ . En este sentido, el modelo de población uniforme y densidad inicial de Pareto podría ser representativo de todos aquellos modelos cuya constante C_0 - de la aproximación contenida en el artículo de Ibragimov & Has'Minsky tantas veces ya citado - toma el valor 1, y cuyos recintos de definición para los valores muestrales dependen, precisamente, del parámetro θ . La idea sería, por tanto, aproximar en muestras grandes las informaciones esperadas de estos modelos mediante expresiones tan sencillas como la

$$\log(1 + t \cdot n)$$

siendo t una constante a determinar. Queda, pues, abierto el problema de encontrar las condiciones necesarias y las expresiones de aproximación para este tipo de modelos y cuya forma representativa puede ser la del modelo resuelto en este capítulo.

CAPITULO 3

FUNCIONES PARA LA UTILIDAD DEL EXPERIMENTO. SUS PROPIEDADES

Todo problema de decisión lleva consigo la consideración de una cierta función de utilidad. Desde el punto de vista práctico las dificultades nacen al determinar una función de utilidad específica en cada caso concreto. En este capítulo van a desarrollarse tres tipos fundamentales de funciones de la información esperada y que pueden ser consideradas como las funciones de utilidad del experimento $E(n)$. Una vez definidas estudiaremos qué propiedades de las enumeradas para la información esperada satisfacen.

3.1 DISTINTAS EXPRESIONES PARA LA UTILIDAD DE UN EXPERIMENTO

Ya se ha indicado en el primer capítulo que la inferencia estadística realizada sobre un parámetro desconocido θ puede considerarse como un caso particular o un problema de decisión particular en el que el argumento Bayes se aplica con exactitud. Dado que el objetivo declarado de la experimentación científica es el de obtener información con objeto de adquirir conocimiento sobre θ , el científico, una vez ha realizado su investigación, deberá describir sus conclusiones finales sobre el valor de la cantidad investigada, juntamente con los argumentos que le han permitido obtener esas conclusiones. Por otra parte, la utilidad de su descripción viene dada por la nueva información que proporciona. Así pues, tal como indica Bernardo (1979), la inferencia estadística sobre el valor de una cantidad o parámetro de interés puede enfocarse como un problema de decisión, en presencia de incertidumbre, en el que el espacio de decisiones es la clase de distribuciones finales sobre el parámetro y la función de utilidad una medida apropiada de la información.

También en el primer capítulo se ha definido como valor de un ex-

perimento $E(n)$ a expresiones que engloban a su utilidad y costo esperados. Así pues, deberán considerarse expresiones de la forma

$$u(E(n), x(n), \theta) = u_t(x(n), \theta) - c_s(x(n)) \quad ,$$

es decir, una utilidad aditiva compuesta de una utilidad final u_t y un costo muestral c_s , ambos expresados en idénticas unidades. En este capítulo dejaremos a un lado la consideración del costo muestral. Por tanto, nos centraremos en encontrar una expresión adecuada para el valor esperado de la utilidad del experimento $E(n) = \{X(n), \mathcal{D}, p_{X(n)}(\cdot/\theta)\}$, la $U(E(n))$, con objeto de poder resolver así cualquier problema de diseño experimental que venga expresado en términos informacionales.

La primera de las expresiones para la utilidad de un experimento es la de aparición más natural e intuitiva. Efectivamente, una utilidad directamente proporcional a la información esperada indica claramente que la utilidad crece con la información. Es evidente que conforme crezca la información esperada proporcionada por el experimento $E(n)$ debe también crecer su utilidad, y si cuando tanto crezca aquélla crece ésta con un factor de proporcionalidad, es claro que la expresión que se ajusta a estas circunstancias viene dada por

$$U_1(E(n)) \propto I^\theta \{E(n), p_\theta(\cdot)\} \quad .$$

Por otra parte, la introducción de esta primera función de la utilidad esperada puede establecerse en términos completamente teóricos. Tal como se indicaba en la última sección del capítulo primero, cuando las preferencias del investigador vengan descritas a través de una función de utilidad local, regular, exacta e invariante, la utilidad del experimento $E(n) = \{X(n), \mathcal{D}, p_{X(n)}(\cdot/\theta)\}$ realizado para efectuar inferencia sobre el parámetro θ , viene dada por la expresión (Bernardo, 1979) :

$$U_1(E(n)) = g_1 \cdot I^\theta \{E(n), p_\theta(\cdot)\}$$

donde $I^\theta \{E(n), p_\theta(\cdot)\}$ es, evidentemente, la información esperada proporcionada por el experimento $E(n)$, cuando las opiniones iniciales sobre θ se describen a través de la densidad inicial $p_\theta(\cdot)$. Esta introducción teórica en la forma clara con la intuitiva expresada con anterioridad y, en último término, pone de manifiesto la relación utilidad-información que antes apuntábamos.

Evidentemente, la constante g_1 en la ecuación anterior es posi-

tiva. Su interpretación queda precisada al considerar la estructura de la relación entre $U(E(n))$ e $I^{\theta} \{E(n), p_{\theta}(\cdot)\}$. En efecto, g_1 no es sino la utilidad esperada de una unidad informante sobre el valor del parámetro θ . Ello se deduce al despejar g_1 y quedar como cociente entre la utilidad proporcionada por el experimento $E(n)$ consistente en obtener una muestra de tamaño n y la información esperada contenida en dicho experimento. La base 2 empleada para los logaritmos de la definición de la información esperada proporciona una idea intuitiva para el significado de la constante de proporcionalidad g_1 y que permite volver a formular como sigue la afirmación precedente: la constante g_1 es la utilidad esperada proporcionada a priori por el respondiente a la pregunta binaria sobre la cantidad de interés (Renyi, 1970, p. 564).

Una segunda función para la utilidad del experimento $E(n)$ puede introducirse al considerar que la probabilidad p de un determinado suceso puede ser una medida de nuestra capacidad de predicción. Por ello, $-\log_2 p$ es una medida cuantitativa de la incertidumbre asociada a su ocurrencia. Inversamente, si $I(\theta)$ denota a una medida de la incertidumbre a priori sobre el parámetro θ , a cada incertidumbre $I(\theta)$ puede asociársele un número $P(\theta) = 2^{-I(\theta)}$, para interpretarse como una medida en $[0, 1]$ de nuestra capacidad de predicción sobre θ . Por unicidad con la notación utilizada en el resto de la tesis trabajaremos con logaritmos neperianos. De ahí que pueda decirse que cada entropía puede asociarse, en caso finito, a un número $P(H(p_{\theta}(\cdot))) = \exp(-H(p_{\theta}(\cdot)))$ para interpretarse nuevamente como una medida en $[0, 1]$ de nuestra capacidad de predicción sobre θ . En particular, la capacidad de predicción es cero para una incertidumbre infinita y 1 cuando la incertidumbre sea cero.

Teniendo en cuenta las ideas anteriores es lógico suponer que, realizado el experimento $E(n) = \{X(n), \mathcal{U}, p_{X(n)}(\cdot/\theta)\}$ a fin de recabar información sobre la cantidad de interés θ , la utilidad de dicho experimento venga expresada en términos de su proporcionalidad al crecimiento relativo producido en la capacidad de predicción sobre θ . Esto es, siendo

$$\Delta = \exp(-E_{X(n)} H(p_{\theta}(\cdot/X(n)))) - \exp(-H(p_{\theta}(\cdot)))$$

el crecimiento absoluto producido en la capacidad de predicción sobre θ por el hecho de realizar el experimento $E(n)$, entonces la utilidad esperada de dicho experimento puede venir expresada por

$$U_2(E(n)) \propto \Delta / \exp(-E_{x(n)} H(p_{\theta}(\cdot/x(n)))) , \text{ esto es,}$$

$$U_2(E(n)) = g_2 \cdot (1 - \exp(-H(p_{\theta}(\cdot)) + E_{x(n)} H(p_{\theta}(\cdot/x(n)))))$$

utilidad esperada que, expresada en función de la información esperada, admite como formulación más operativa la siguiente:

$$U_2(E(n)) = g_2 \cdot (1 - e^{-I^{\theta}\{E(n), p_{\theta}(\cdot)\}}) .$$

Una vez más la ecuación anterior pone de manifiesto la relación entre la información y la utilidad esperadas del experimento $E(n)$.

Hay que destacar que cuando la información esperada sea arbitrariamente grande, esto es, en el caso de información total infinita, $U_2(E(n))$ toma el valor de la constante g_2 . Así pues, la interpretación práctica para esta constante es, en tales circunstancias, la de la utilidad proporcionada por la información total y, por ello, se trata de una magnitud estrictamente positiva. En general, puede considerarse que g_2 representa la utilidad final de investigar con información completa ó, abreviadamente, la ganancia final posible. Por su misma interpretación, g_2 admite una estimación fácil en casos prácticos.

La utilidad $U_2(E(n))$ se ha introducido al considerar el crecimiento relativo en la capacidad de predicción del investigador sobre θ . Si ahora se considera el crecimiento absoluto en la capacidad de predicción sobre la cantidad de interés θ , resulta una tercera función de utilidad para el experimento $E(n) = \{X(n), p_{x(n)}(\cdot/\theta)\}$ definida como aquella que es proporcional a dicho incremento absoluto. Se tendrá, pues, que:

$$U_3(E(n)) \propto e^{-E_{x(n)} H(p_{\theta}(\cdot/x(n)))} - e^{-H(p_{\theta}(\cdot))} , \text{ esto es,}$$

$$U_3/E(n) = g_3 \cdot (e^{-E_{x(n)} H(p_{\theta}(\cdot/x(n)))} - e^{-H(p_{\theta}(\cdot))}) .$$

Sacando factor común a $\exp(-H(p_{\theta}(\cdot)))$ la definición anterior resulta:

$$U_3(E(n)) = g_3 \cdot e^{-H(p_{\theta}(\cdot))} (e^{I^{\theta}\{E(n), p_{\theta}(\cdot)\}} - 1)$$

como tercera expresión de la utilidad esperada del experimento $E(n)$, y en donde nuevamente se aprecia la relación entre la información y la utilidad esperadas.

La interpretación de la constante g_3 es más forzada que en casos anteriores y, por ello, de estimación más difícil. En efecto, expresada la utilidad esperada a través de $U_3(E(n))$, dicha utilidad y la constante g_3 quedan igualadas para una información igual al valor $\log(1 + \exp(H(p_\theta(.))))$. Así pues, g_3 no es sino la utilidad esperada proporcionada por una información igual al $\log(1 + e^{H(p_\theta(.))})$, siendo, por tanto, positiva. Es de resaltar que el valor de la constante g_3 dependerá de cada modelo concreto puesto que está definida en función de la entropía de la densidad inicial, aunque también es cierto por su misma definición que no depende del particular resultado muestral obtenido tras la realización del experimento. Este último hecho asegura la coherencia al utilizar la función $U_3(E(n))$ en la práctica, al ser su estimación independiente del resultado muestral particular obtenido. Cuando la entropía de la densidad inicial sea muy pequeña, la constante g_3 puede tomar valores muy cercanos al cero, cosa que no ocurre para valores de $H(p_\theta(.))$ altamente positivos. Por otra parte, cuando dicha entropía es cero, g_3 no es sino la utilidad esperada de la información igual al $\log 2$. De ahí que si se utilizan logaritmos de base 2, g_3 resultaría ser la utilidad unitaria siempre que la entropía de la densidad inicial fuese cero.

Una de las propiedades satisfechas por la información esperada y en la que se basa gran parte de su operatividad es la invarianza respecto a transformaciones biyectivas de la cantidad de interés, problema estudiado por Bernardo (1975). Evidentemente, esta propiedad no se ve alterada por la presencia del factor de proporcionalidad g_1 , de forma que también la función de la información esperada $U_1(E(n))$ es invariante ante tales transformaciones. Idéntico comportamiento sigue la segunda de las funciones de utilidad, $U_2(E(n))$, toda vez que su definición depende exclusivamente de la invariante información esperada. Sin embargo, este no es el comportamiento seguido por la tercera de las funciones, $U_3(E(n))$. En efecto, dicha función no solamente lo es de la información esperada $I^\theta\{E(n), p_\theta(.)\}$, sino que también está definida dependiendo de la entropía de la densidad inicial a través del factor $\exp(-H(p_\theta(.)))$. La entropía no es invariante ante transformaciones biyectivas de la cantidad de interés, como puede apreciarse comparando las correspondientes entropías de la densidad de Pareto con la de la densidad exponencial en el ejemplo propuesto en la sección 2.3,

al verificarse que

$$- \int k e^{-k\theta} \log k e^{-k\theta} \cdot d\theta = 1 - \log k \quad .$$

Este resultado disminuye, evidentemente y de forma considerable, la operatividad de la tercera de las funciones de utilidad consideradas. Sin embargo, y a pesar de esta situación adversa, seguiremos idéntico tratamiento con las tres, a fin de comprobar qué propiedades de las satisfechas por la información esperada siguen verificándose por dichas funciones de la utilidad esperada proporcionada por el experimento $E(n)$.

Anteriormente, haciendo referencia al capítulo primero, mencionábamos que bajo ciertas condiciones generales de regularidad se demuestra la unicidad de la función $U_1(E(n))$ como aquella que cumple una serie de aconsejables condiciones (Bernardo, 1979). No obstante, la utilización de las funciones $U_2(E(n))$ y $U_3(E(n))$ puede justificarse, en ejemplos prácticos, por su mismo nacimiento empírico como funciones directamente proporcionales a los crecimientos relativo y absoluto, respectivamente, de la capacidad de predicción del investigador por el hecho de realizar una experimentación. Recordemos, sin embargo, que si bien la anterior afirmación es válida en toda su amplitud para $U_2(E(n))$, la tercera de las funciones, la $U_3(E(n))$, está sujeta por la limitación de su no invarianza ante transformaciones biyectivas de la cantidad de interés, circunstancia ésta que, según acabamos de comentar, frena su operatividad en la práctica.

Veamos qué expresiones toman las distintas funciones de la utilidad esperada en los dos ejemplos contenidos en el capítulo segundo. Bajo las hipótesis del modelo normal resultaría:

$$U_1(E(n)) = \frac{1}{2} g_1 \cdot \log\left(1 + n \frac{\sigma_0^2}{\sigma^2}\right) \quad ,$$

$$U_2(E(n)) = g_2 \cdot \left(1 - \frac{1}{\sqrt{1 + n \frac{\sigma_0^2}{\sigma^2}}}\right) \quad , \quad y$$

$$U_3(E(n)) = g_3 \cdot \frac{\sqrt{n \sigma_0^2 + \sigma^2} - \sigma}{\sigma \sigma_0 \sqrt{2\pi e}}$$

Para el caso de una población uniforme y opiniones iniciales descriptas por una distribución de Pareto, resultaría:

$$U_1(E(n)) = g_1 \cdot \log\left(1 + \frac{n}{k}\right) ,$$

$$U_2(E(n)) = g_2 \cdot n / (k + n) , \quad y$$

$$U_3(E(n)) = g_3 \cdot n / \theta_0 \cdot \exp\left(1 + \frac{1}{k}\right) .$$

Notemos que la última de las funciones de la utilidad esperada anteriores sí depende del parámetro θ_0 de la distribución inicial de Pareto, introducido por la presencia de la entropía de la densidad inicial en el factor $e^{-H(P_\theta(\cdot/\theta_0, k))}$ de la definición general de $U_3(E(n))$. Nuevamente, esta circunstancia no es sino confirmación para este modelo concreto de la menor aplicabilidad práctica de la tercera de las funciones de la utilidad esperada consideradas.

Por último, estudiemos las relaciones que ligan a las constantes g_i ($i=1,2,3$) entre sí.

TEOREMA 3.1.1

La relación entre las distintas constantes g_i ($i=1,2,3$) viene expresada a partir de las siguientes igualdades:

$$\frac{U_1(E(n))}{g_1} = \log \frac{g_2}{g_2 - U_2(E(n))} = \log \frac{1}{1 - (U_2(E(n)) / g_2)}$$

$$\frac{U_1(E(n))}{g_1} = \log\left(1 + \frac{U_3(E(n)) \cdot e^{H(P_\theta(\cdot))}}{g_3}\right)$$

$$\frac{g_2}{g_2 - U_2(E(n))} = 1 + \frac{U_3(E(n)) \cdot e^{H(P_\theta(\cdot))}}{g_3}$$

Demostración.- Tomando las definiciones de las distintas $U_i(E(n))$ ($i=1,2,3$) expresadas como funciones de la información esperada, despejando dicha información e igualando, resulta:

$$I^\theta\{E(n), P_\theta(\cdot)\} = \frac{U_1(E(n))}{g_1}$$

$$I^\theta\{E(n), P_\theta(\cdot)\} = \log \frac{g_2}{g_2 - U_2(E(n))} , \quad e^{I^\theta\{E(n), P_\theta(\cdot)\}} = \frac{g_2}{g_2 - U_2(E(n))}$$

$$I^{\theta} \{E(n), p_{\theta}(\cdot)\} = \log \left(1 + \frac{U_3(E(n)) \cdot e^{H(p_{\theta}(\cdot))}}{g_3} \right) ,$$

$$e^{I^{\theta} \{E(n), p_{\theta}(\cdot)\}} = \left(1 + \frac{U_3(E(n)) \cdot e^{H(p_{\theta}(\cdot))}}{g_3} \right)$$

deduciéndose finalmente el resultado al igualar las correspondientes expresiones.

(c.q.d.)

Denotando por $\mu_i = U_i(E(n)) / g_i$ ($i=1,2,3$), las relaciones anteriores podrían reescribirse como sigue:

$$\mu_1 = \log \frac{1}{1 - \mu_2} = -\log(1 - \mu_2) \quad 0 \leq \mu_2 \leq 1, \quad \mu_1 \geq 0$$

$$\mu_1 = \log(1 + \mu_3 \cdot e^{H(p_{\theta}(\cdot))}) \quad \mu_3 \geq 0, \quad \mu_1 \geq 0$$

$$\frac{1}{1 - \mu_2} = 1 + \mu_3 \cdot e^{H(p_{\theta}(\cdot))} \quad 0 \leq \mu_2 \leq 1, \quad \mu_3 \geq 0$$

y en las que puede apreciarse con claridad que el crecimiento de una μ_i en particular implica el crecimiento de las demás. Por otra parte, resulta evidente que mientras que μ_1 y μ_3 pueden tomar cualquier valor real positivo, μ_2 por su misma definición toma valores en el intervalo $[0,1]$. Además, si una determinada función $U_i(E(n))$ se anula para el experimento $E(n) = \{X(n), p_{X(n)}(\cdot/\theta)\}$, las otras dos también toman el valor 0. Análogamente, una μ_i infinita implica (si $i=1,3$) que μ_3 ó μ_1 también lo sea, mientras que μ_2 tome el máximo valor de 1. Todas estas afirmaciones pueden deducirse fácilmente observando las sencillas representaciones de las expresiones que ligan a las constantes g_i entre sí, al tiempo que aseguran la coherencia en el empleo de cualquiera de las tres funciones de utilidad.

En lo que resta, vamos a centrar la discusión en el estudio de cuáles propiedades de las satisfechas por $I^{\theta} \{E(n), p_{\theta}(\cdot)\}$ lo son, a su vez, por las distintas $U_i(E(n))$, $i=1,2,3$. Y así, podrá comprobarse que $U_1(E(n))$ las satisface todas; $U_2(E(n))$ no satisface la aditividad como funcional del experimento $E(n)$ en idénticos términos que la información esperada; y, $U_3(E(n))$, finalmente, incumple tanto la aditividad como la concavidad, ya respecto el tamaño muestral, como funcional de la densidad inicial $p_{\theta}(\cdot)$. Respecto al em-

pleo de una u otra de las funciones de la utilidad esperada, y como se comentará con posterioridad, dependerá de las condiciones particulares de cada problema concreto.

3.2 PROPIEDADES SATISFECHAS POR LAS ANTERIORES FUNCIONES DE UTILIDAD

La información esperada proporcionada por el experimento $E(n) = \{X(n), \mathcal{D}, p_{X(n)}(\cdot/\mathcal{D})\}$ verifica una serie de propiedades, enunciadas ya en el capítulo primero. Dichas propiedades se refieren tanto a su comportamiento como función del tamaño muestral n , como a su carácter como polinomial de la densidad inicial $p_{\mathcal{D}}(\cdot)$, ya respecto al experimento $E(n)$, así como a los valores que $I^{\mathcal{D}}\{E(n), p_{\mathcal{D}}(\cdot)\}$ toma para densidades y tamaños concretos. Las tres funciones de utilidad definidas en la sección 3.1 como funciones del experimento $E(n)$ lo son, a su vez, de la información esperada. Es interesante, pues, comprobar qué propiedades de ésta se "trasladan" a aquéllas, dado que precisamente son estas propiedades las que les confieren un acusado carácter operativo y refuerzan el concepto de utilidad aplicado a las tres funciones anteriores.

Para la primera de las funciones, la $U_1(E(n))$, las distintas propiedades de $I^{\mathcal{D}}\{E(n), p_{\mathcal{D}}(\cdot)\}$ se cumplen de forma evidente por su misma definición como directamente proporcional a dicha información esperada y sin más que sacar factor común en todas las demostraciones a la constante de proporcionalidad positiva g_1 . Por ello, solamente las mencionaremos remitiendo sus demostraciones al artículo de Lindley (1956) citado en la bibliografía.

I. UTILIDADES POSITIVAS

En Lindley (1956) puede encontrarse la demostración de que la información esperada $I^{\mathcal{D}}\{E(n), p_{\mathcal{D}}(\cdot)\}$ sobre \mathcal{D} proporcionada por el experimento $E(n)$ y cuando las opiniones iniciales se describen a partir de una densidad $p_{\mathcal{D}}(\cdot)$ toma valores positivos y se anula si y sólo si no existe modificación en las opiniones iniciales al estudiar la información contenida en la muestra $x(n)$, esto es, cuando $p_{\mathcal{D}}(\cdot/x(n)) = p_{\mathcal{D}}(\cdot)$. Las dos densidades, inicial y posterior, coinciden.

Considerando, pues, la primera propiedad de la información esperada tenemos el siguiente conjunto de resultados:

TEOREMA 3.2.I.1

$$U_1(E(n)) \geq 0 \quad \text{y} \quad U_1(E(n)) = 0 \quad \text{sii} \quad p_{\theta}(\cdot/x(n)) = p_{\theta}(\cdot)$$

TEOREMA 3.2.I.2

$$U_2(E(n)) \geq 0 \quad \text{y} \quad U_2(E(n)) = 0 \quad \text{sii} \quad p_{\theta}(\cdot/x(n)) = p_{\theta}(\cdot)$$

Demostración.- La constante g_2 es positiva y la información esperada no negativa. Así pues,

$$0 < \exp(-I^{\theta}\{E(n), p_{\theta}(\cdot)\}) \leq 1$$

esto es,

$$0 \leq 1 - e^{-I^{\theta}\{E(n), p_{\theta}(\cdot)\}} < 1$$

de donde se deduce la primera parte del teorema.

Para la doble implicación hay que tener en cuenta que

$$U_2(E(n)) = 0 \quad \text{sii} \quad 1 = e^{-I^{\theta}\{E(n), p_{\theta}(\cdot)\}} \quad \text{sii} \quad I^{\theta}\{E(n), p_{\theta}(\cdot)\} = 0$$

$$\text{sii} \quad p_{\theta}(\cdot/x(n)) = p_{\theta}(\cdot)$$

(c.q.d.)

TEOREMA 3.2.I.3

$$U_3(E(n)) \geq 0 \quad \text{y} \quad U_3(E(n)) = 0 \quad \text{sii} \quad p_{\theta}(\cdot/x(n)) = p_{\theta}(\cdot)$$

Demostración.- Como la información esperada es no negativa se deduce que

$$H(p_{\theta}(\cdot)) \geq E_{x(n)} H(p_{\theta}(\cdot/x(n)))$$

desigualdad que implica la primera parte del teorema sin mas que considerar que la constante g_3 es positiva.

La segunda parte es inmediata al considerar que

$$U_3(E(n)) = 0 \quad \text{sii} \quad H(p_{\theta}(\cdot)) = E_{x(n)} H(p_{\theta}(\cdot/x(n))) \quad \text{sii} \quad I^{\theta}\{E(n), p_{\theta}(\cdot)\} = 0$$

$$\text{sii} \quad p_{\theta}(\cdot/x(n)) = p_{\theta}(\cdot)$$

(c.q.d.)

Las distintas funciones de utilidad calculadas para los dos ejemplos desarrollados anteriormente sirven para ilustrar su no negatividad de forma práctica. En efecto, en el modelo normal resultan las desigualdades estrictas:

$$1 + n \frac{\sigma_0^2}{\sigma^2} > 1, \text{ esto es, } \log\left(1 + n \frac{\sigma_0^2}{\sigma^2}\right) > 0$$

$$\sqrt{1 + n \frac{\sigma_0^2}{\sigma^2}} > 1, \text{ esto es, } 1 - \frac{1}{\sqrt{1 + n \frac{\sigma_0^2}{\sigma^2}}} > 0$$

$$\sqrt{\sigma^2 + n\sigma_0^2} \geq \sigma, \text{ esto es, } (\sqrt{n\sigma_0^2 + \sigma^2} - \sigma) / \sigma\sigma_0\sqrt{2\pi e} > 0$$

Y para una población uniforme con densidad inicial de Pareto resulta:

$$1 + \frac{n}{k} > 1, \text{ esto es, } \log\left(1 + \frac{n}{k}\right) > 0$$

$$n / (k+n) > 0$$

$$n / \theta_0 \cdot e^{((1+k)/k)} > 0$$

Así pues, puede apreciarse como era de esperar que para ambos modelos las tres funciones de utilidad consideradas son no negativas, pues basta multiplicar las expresiones anteriores por las correspondientes constantes positivas g_i .

Esta primera propiedad demostrada para las funciones de utilidad $U_i(E(n))$ ($i=1,2,3$) pone de manifiesto que las utilidades así definidas son positivas y que, si no hay variación en las opiniones que el investigador posee sobre la cantidad de interés θ (expresada esta circunstancia con la igualdad entre las densidades inicial y posterior) al observar el resultado muestral, no se incrementa la utilidad al realizar la experimentación. Todo ello expresado, evidentemente, en términos esperados. Así pues, en este último caso, toda la utilidad es la que ya se disponía a priori, no aumentando por la observación del resultado muestral. Esta circunstancia asegura la coherencia en el empleo de cualquiera de las tres funciones de utilidad. Inversamente, cuando la densidad posterior de θ varía con $x(n)$, el grupo de teoremas anterior indica que, por término medio, un experimento resulta útilmente informativo.

II. ADITIVIDAD DE LAS FUNCIONES DE UTILIDAD

La segunda de las propiedades satisfechas por la información esperada tiene un marcado carácter aditivo que permite, en situaciones prácticas, descomponer el proceso de muestreo en una serie de etapas, cada una con la correspondiente información adicional, y cuya suma total es la información de todo el proceso. Para desarrollar esta segunda propiedad para las funciones de utilidad esperada es conveniente, previamente, introducir la siguiente notación.

Supongamos, pues, que el resultado x de un experimento E consiste en dos observaciones $x_1 \in X_1$ y $x_2 \in X_2$. Entonces, $E_1 = \{X_1, \Theta, p_{x_1}(\cdot/\vartheta)\}$ ($i=1,2$) son dos experimentos con espacio paramétrico Θ común, y $E = (E_1, E_2)$ se llama experimento "suma" de E_1 y E_2 . Se define el experimento $E_2(x_1) = \{X_2, \Theta, p_{x_2}(\cdot/\vartheta, x_1)\}$ como aquel que consiste en obtener la segunda observación x_2 cuando ya se ha obtenido y estudiado convenientemente la primera observación x_1 (y, por tanto, las opiniones iniciales se ven modificadas oportunamente). Notemos que, para el experimento $E_2(x_1)$, la densidad de probabilidad $p_{\vartheta}(\cdot/x_1)$ actúa como una auténtica densidad inicial. Sobre este experimento puede definirse la correspondiente $I^{\vartheta}\{E_2(x_1), p_{\vartheta}(\cdot/x_1)\}$ como la información esperada proporcionada por el experimento consistente en obtener una segunda observación después de que se ha realizado E_1 y obtenido e investigado el dato x_1 . La información esperada así definida es función, evidentemente, del resultado muestral x_1 debido a la modificación que introduce su estudio sobre la densidad inicial $p_{\vartheta}(\cdot)$, que pasa a convertirse en la $p_{\vartheta}(\cdot/x_1)$. El valor esperado respecto a x_1 de la información anterior, que lo denotaremos por $I^{\vartheta}\{E_2/E_1, p_{\vartheta}(\cdot)\}$, se define como la información esperada sobre ϑ proporcionada por E_2 después de que se haya realizado E_1 . Con esta notación, Lindley (1956) demuestra para la información esperada la igualdad:

$$I^{\vartheta}\{E, p_{\vartheta}(\cdot)\} = I^{\vartheta}\{E_1, p_{\vartheta}(\cdot)\} + I^{\vartheta}\{E_2/E_1, p_{\vartheta}(\cdot)\} .$$

Resulta evidente que todo el proceso anterior es iterativo y puede generalizarse a un número finito de etapas (Lindley, 1956). En cada una de ellas se considera la información adicional producida al realizar una observación, suponiendo conocida la anterior. De esta forma siempre podrá decomponerse la información esperada proporcionada por el experimento $E(n)$ en una suma de las informaciones esperadas obtenidas en cada observación, supuestamente conocida e investigada la información contenida en la anterior observación.

Reflejo de esta segunda propiedad de la información esperada se tienen las siguientes versiones para la aditividad de la utilidad esperada del experimento $E(n)$:

TEOREMA 3.2.II.1

Bajo las condiciones anteriores, esto es, para un par de experimentos

$E_i = \{X_i, \Theta, p_{X_i}(\cdot/\theta)\}$ ($i=1,2$) con espacio paramétrico Θ común, y

una densidad inicial $p_0(\cdot)$, entonces

$$U_1(E) = U_1(E_1) + U_1(E_2/E_1)$$

donde $E = (E_1, E_2)$ es el experimento suma de los E_i ($i=1,2$).

La generalización a un número finito de experimentos con espacio paramétrico Θ común admite la siguiente formulación:

$$U_1(E(n)) = U_1(E_1) + \sum_{i=2}^n U_1(E_i/E_{i-1})$$

Las correspondientes expresiones que toma el teorema en los dos ejemplos propuestos muestran el mecanismo concreto de su aplicación práctica.

Es conocido (Lindley, 1956) que bajo las hipótesis del modelo normal:

$$p_{\theta}(\cdot/x_1) = N\left(\frac{x_1 \sigma_0^2 + \vartheta_0 \sigma^2}{\sigma^2 + \sigma_0^2}, \sqrt{\frac{\sigma^2 \sigma_0^2}{\sigma^2 + \sigma_0^2}}\right)$$

De ahí que

$$U_1(E) = g_1 \cdot I^{\vartheta} \left\{ E, N(\vartheta_0, \sigma_0^2) \right\} = \frac{1}{2} g_1 \cdot \log\left(1 + 2 \frac{\sigma_0^2}{\sigma^2}\right)$$

$$U_1(E_1) = g_1 \cdot I^{\vartheta} \left\{ E_1, N(\vartheta_0, \sigma_0^2) \right\} = \frac{1}{2} g_1 \cdot \log\left(1 + \frac{\sigma_0^2}{\sigma^2}\right)$$

$$U_1(E_2/E_1) = g_1 \cdot I^{\vartheta} \left\{ E_2/E_1, N\left(\frac{x_1 \sigma_0^2 + \vartheta_0 \sigma^2}{\sigma^2 + \sigma_0^2}, \sqrt{\frac{\sigma^2 \sigma_0^2}{\sigma^2 + \sigma_0^2}}\right) \right\} =$$

$$= \frac{1}{2} g_1 \cdot \log\left(1 + \frac{\sigma_0^2}{\sigma^2 + \sigma_0^2}\right)$$

cumpléndose que

$$\frac{1}{2} g_1 \cdot \log\left(1 + 2 \frac{\sigma_0^2}{\sigma^2}\right) = \frac{1}{2} g_1 \cdot \log\left(1 + \frac{\sigma_0^2}{\sigma^2}\right) + \frac{1}{2} g_1 \cdot \log\left(1 + \frac{\sigma_0^2}{\sigma^2 + \sigma_0^2}\right)$$

tal como indica el teorema 3.2.II.1.

En el caso de una población uniforme con opiniones iniciales descritas por una densidad de Pareto, es conocido (DeGroot, 1970) que

$$p_{\vartheta}(\cdot/x_1) = P_{\vartheta}(\cdot/\vartheta_1, 1+k) \quad \text{con} \quad \vartheta_1 = \max(\vartheta_0, x_1)$$

Denotando por $\vartheta_2 = \max(\vartheta_0, x_1, x_2)$ (aunque resulte irrelevante el valor de ϑ_2 ,

en el cálculo de la información esperada, en virtud del Corolario 2.2.1)
resultará:

$$U_1(E) = g_1 \cdot I^\theta \{E, p_\theta(\cdot/\theta, k)\} = g_1 \cdot \log\left(1 + \frac{2}{k}\right)$$

$$U_1(E_1) = g_1 \cdot I^\theta \{E_1, p_\theta(\cdot/\theta, k)\} = g_1 \cdot \log\left(1 + \frac{1}{k}\right)$$

$$U_1(E_2/E_1) = g_1 \cdot I^\theta \{E_2/E_1, p_\theta(\cdot/\theta, 1+k)\} = g_1 \cdot \log\left(1 + \frac{1}{1+k}\right)$$

verificándose efectivamente que

$$g_1 \cdot \log\left(1 + \frac{2}{k}\right) = g_1 \cdot \log\left(1 + \frac{1}{k}\right) + g_1 \cdot \log\left(1 + \frac{1}{1+k}\right) .$$

La interpretación del teorema 3.2.II.1 es la dual de la mencionada para la información esperada. Y así, la utilidad proporcionada por el experimento consistente en la obtención de dos unidades muestrales, cuando esta utilidad viene expresada a partir de la función $U_1(E)$, es suma de la utilidad esperada proporcionada por el experimento consistente en la obtención de la primera unidad muestral y de una utilidad esperada residual que corresponde a la del experimento consistente en obtener la segunda observación conocido ya el resultado de la primera. Esta aditividad es fundamental y, junto a la no negatividad, constituye una pieza básica para cualquier desarrollo axiomático de la utilidad esperada. Por otra parte, la no negatividad de la utilidad esperada implica que, realizado un determinado experimento $E(n)$, es decir, llegado el momento de la extracción de la unidad muestral n -ésima, la extracción de la $(n+1)$ -ésima observación es, por término medio, útilmente informativa. La inmediata pregunta a la afirmación anterior consistirá en saber si, en ausencia de un costo asociado al experimento, existe un momento en el que no valga la pena seguir obteniendo nuevas observaciones muestrales puesto que desaparezca la utilidad esperada marginal. Este es el sentido del corolario siguiente:

COROLARIO 3.2.II.1

Si x_1 es suficiente para $x = (x_1, x_2)$ con respecto a θ , en el sentido de que $p_\theta(\cdot/x) = p_\theta(\cdot/x_1)$, entonces

$$U_1(E) = U_1(E_1) .$$

La demostración es evidente sin más que considerar que ya Lindley (1956) demuestra que si x es suficiente según el sentido expresado en el corolario, entonces $I^\theta \{E, p_\theta(\cdot)\} = I^\theta \{E_1, p_\theta(\cdot)\}$.

Este corolario establece que no hay pérdida de utilidad esperada cuando la información se ciñe a la observación de un estadístico suficiente respecto al parámetro objeto de estudio. Inversamente, si un estadístico es insuficiente en el sentido descrito, entonces existe una pérdida de utilidad puesto que en tal caso la información esperada (y por ello, la utilidad correspondiente) $I^{\theta}\{E_2/E_1, p_{\theta}(\cdot)\}$ es estrictamente positiva (Bernardo, 1978). Por otra parte, llegada a una etapa en la que se produce la igualdad entre las densidades expresadas en el corolario, si no se tiene en cuenta el costo asociado al experimento, deberá pararse la realización del muestreo toda vez que no es de esperar ganar utilidad con la obtención de una nueva observación. Evidentemente, en tales circunstancias y con el costo tenido en cuenta, puede que el proceso quedara interrumpido con anterioridad.

El comportamiento seguido en situaciones como la descrita en el teorema 3.2.II.1 por la segunda de las funciones de la utilidad esperada de un experimento, la $U_2(E)$, es ligeramente distinto debido a la presencia de un nuevo factor. No obstante, dicho comportamiento está en el contexto de interpretación realizado para $U_1(E)$. Demostremos el siguiente

TEOREMA 3.2.II.2

Para un par de experimentos $E_i = \{X_i, \mathcal{D}, p_{X_i}(\cdot/\theta)\}$ ($i=1,2$) con espacio paramétrico \mathcal{D} común, se verifica

$$U_2(E) = U_2(E_1) + U_2(E_2/E_1) \cdot e^{-I^{\theta}\{E_1, p_{\theta}(\cdot)\}}$$

siendo $E = (E_1, E_2)$ el experimento suma de los E_i ($i=1,2$).

Demostración.- Dado que

$$\begin{aligned} U_2(E) &= g_2 \cdot (1 - e^{-I^{\theta}\{E, p_{\theta}(\cdot)\}}) = g_2 \cdot (1 - e^{-I^{\theta}\{E_1, p_{\theta}(\cdot)\}} \cdot e^{-I^{\theta}\{E_2/E_1, p_{\theta}(\cdot)\}} - \\ &- e^{-I^{\theta}\{E_1, p_{\theta}(\cdot)\}} + e^{-I^{\theta}\{E_1, p_{\theta}(\cdot)\}}) = g_2 \cdot (1 - e^{-I^{\theta}\{E_1, p_{\theta}(\cdot)\}}) + \\ &+ g_2 \cdot e^{-I^{\theta}\{E_1, p_{\theta}(\cdot)\}} (1 - e^{-I^{\theta}\{E_2/E_1, p_{\theta}(\cdot)\}}) = U_2(E_1) + U_2(E_2/E_1) \cdot \\ &\cdot e^{-I^{\theta}\{E_1, p_{\theta}(\cdot)\}} \end{aligned}$$

(c.q.d.)

Del resultado anterior puede apreciarse que no se verifica la aditividad en el sentido descrito para la segunda propiedad de la información

esperada, debido a la presencia del factor $\exp(-I^\theta \{E_1, p_\theta(\cdot)\})$ que multiplica a la utilidad residual $U_2(E_2/E_1)$. Este resultado era previsible ya que Bernardo (1975) demuestra que, bajo condiciones muy generales de regularidad, la utilidad esperada $U_1(E)$ es la única que verifica la mencionada aditividad. Sin embargo, la interpretación del resultado es idéntica a la realizada para la primera de las funciones de la utilidad esperada. En efecto, basta considerar como utilidad residual obtenida tras la investigación de la segunda observación una vez que la primera ya ha sido estudiada al producto

$$U_2(E_2/E_1) \cdot e^{-I^\theta \{E_1, p_\theta(\cdot)\}}$$

Una formulación distinta para el anterior teorema viene expresada de la forma siguiente:

TEOREMA 3.2.II.3

Bajo las hipótesis del teorema 3.2.II.2 ,

$$U_2(E) = U_2(E_1) \cdot e^{-I^\theta \{E_2/E_1, p_\theta(\cdot)\}} + U_2(E_2/E_1)$$

Demostración.- Se verifica la siguiente cadena de igualdades:

$$\begin{aligned} U_2(E) &= g_2 \cdot (1 - e^{-I^\theta \{E, p_\theta(\cdot)\}}) = g_2 \cdot (1 - e^{-I^\theta \{E_1, p_\theta(\cdot)\}} \cdot e^{-I^\theta \{E_2/E_1, p_\theta(\cdot)\}} - \\ &- e^{-I^\theta \{E_2/E_1, p_\theta(\cdot)\}} + e^{-I^\theta \{E_2/E_1, p_\theta(\cdot)\}}) = g_2 \cdot (1 - e^{-I^\theta \{E_2/E_1, p_\theta(\cdot)\}}) + \\ &+ g_2 \cdot e^{-I^\theta \{E_2/E_1, p_\theta(\cdot)\}} (1 - e^{-I^\theta \{E_1, p_\theta(\cdot)\}}) = U_2(E_1) \cdot e^{-I^\theta \{E_2/E_1, p_\theta(\cdot)\}} + \\ &+ U_2(E_2/E_1) \end{aligned}$$

(c.q.d.)

De los dos últimos teoremas se deduce el evidente corolario:

COROLARIO 3.2.II.2

Bajo las hipótesis del teorema 3.2.II.2 ,

$$U_2(E_1) + U_2(E_2/E_1) \cdot e^{-I^\theta \{E_1, p_\theta(\cdot)\}} = U_2(E_2/E_1) + U_2(E_1) \cdot e^{-I^\theta \{E_2/E_1, p_\theta(\cdot)\}}$$

Aplicaremos estos resultados a los modelos del capítulo segundo.

Y así, en el modelo normal, teniendo en cuenta las igualdades:

$$U_2(E) = g_2 \cdot \left(1 - \frac{1}{\sqrt{1 + 2 \frac{\sigma_0^2}{\sigma^2}}}\right)$$

$$U_2(E_1) = g_2 \cdot \left(1 - \frac{1}{\sqrt{1 + \frac{\sigma_0^2}{\sigma^2}}}\right)$$

$$U_2(E_2/E_1) = g_2 \cdot \left(1 - \frac{1}{\sqrt{1 + \frac{\sigma_0^2}{\sigma^2 + \sigma_0^2}}}\right)$$

$$\exp(-I_{E_1, P}(\cdot)) = \frac{1}{\sqrt{1 + \frac{\sigma_0^2}{\sigma^2}}}$$

se verifica la igualdad descrita en el teorema 3.2.II.2, esto es,

$$g_2 \cdot \left(1 - \frac{1}{\left(1 + 2 \frac{\sigma_0^2}{\sigma^2}\right)^{\frac{1}{2}}}\right) = g_2 \cdot \left(1 - \frac{1}{\left(1 + \frac{\sigma_0^2}{\sigma^2}\right)^{\frac{1}{2}}}\right) + \\ + g_2 \cdot \left(1 - \frac{1}{\left(1 + \frac{\sigma_0^2}{\sigma^2 + \sigma_0^2}\right)}\right) \cdot \left(1 + \frac{\sigma_0^2}{\sigma^2}\right)^{-\frac{1}{2}}.$$

En el caso de una población uniforme y unas opiniones iniciales descritas por una densidad de probabilidad de Pareto la comprobación de los teoremas anteriores se realiza teniendo en cuenta las expresiones siguientes:

$$U_2(E) = g_2 \cdot \frac{2}{k+2}$$

$$U_2(E_1) = g_2 \cdot \frac{1}{k+1}$$

$$U_2(E_2/E_1) = g_2 \cdot \frac{1}{k+2}$$

$$\exp(-I_{E_1, P_{\theta}}(\cdot/\theta_0, k)) = \frac{k}{k+1}$$

Análogamente a lo ocurrido con $U_1(E)$, cuando se trabaje con un estadístico suficiente ya en la primera observación resulta irrelevante obte

ner un segundo elemento del espacio muestral. Ello es debido a que la utilidad marginal de esa segunda etapa de muestreo es nula, de forma que este resultado si está en la línea del dual para $U_1(E)$ y para la información esperada $I^\theta\{E, p_\theta(\cdot)\}$. Su interpretación es, pues, completamente análoga a la mencionada para el caso anterior. Este resultado está contenido en el siguiente

COROLARIO 3.2.II.3

Si x_1 es suficiente para $x = (x_1, x_2)$ con respecto a θ , en el sentido de que $p_\theta(\cdot/x) = p_\theta(\cdot/x_1)$, entonces

$$U_2(E) = U_2(E_1) \quad .$$

Demostración.-

$$U_2(E) = g_2 \cdot (1 - e^{-I^\theta\{E, p_\theta(\cdot)\}}) = g_2 \cdot (1 - e^{-I^\theta\{E_1, p_\theta(\cdot)\}}) = U_2(E_1)$$

(c.q.d.)

Para la tercera de las funciones de la utilidad esperada, $U_3(E)$, se obtienen resultados en la línea de los demostrados para $U_2(E)$, apareciendo una aditividad entre la utilidad proporcionada por la primera etapa y una utilidad residual, definida considerando la densidad $p_\theta(\cdot/x_1)$ como densidad inicial. Una vez más, si bien $U_3(E)$ no verifica la aditividad satisfecha por $U_1(E)$, sí cumple cierta aditividad en la línea de la demostrada para $U_2(E)$. Sea, pues, el siguiente

TEOREMA 3.2.II.4

Para un par de experimentos $E_i = \{X_i, \Theta, p_{X_i}(\cdot/\theta)\}$ ($i=1,2$) con el mismo espacio paramétrico Θ , y una densidad inicial $p_\theta(\cdot)$, se verifica que

$$U_3(E) = U_3(E_1) + U_3^*(E_2/E_1)$$

Donde $E = (E_1, E_2)$ es el experimento suma de los E_i ($i=1,2$) y el término $U_3^*(E_2/E_1)$ considera en su definición como densidad inicial la $p_\theta(\cdot/x_1)$.

Demostración.- La tesis del teorema se deduce de la siguiente cadena de igualdades:

$$\begin{aligned}
 U_3(E) &= g_3 \cdot \exp(-H(p_\theta(\cdot))) \cdot (e^{I^\theta\{E, p_\theta(\cdot)\}} - 1) = g_3 \cdot e^{-H(p_\theta(\cdot))} \cdot \\
 &(e^{I^\theta\{E_1, p_\theta(\cdot)\}} \cdot e^{I^\theta\{E_2/E_1, p_\theta(\cdot)\}} - e^{I^\theta\{E_1, p_\theta(\cdot)\}} + e^{I^\theta\{E_1, p_\theta(\cdot)\}} - 1) = \\
 &= g_3 \cdot e^{-H(p_\theta(\cdot))} (e^{I^\theta\{E_1, p_\theta(\cdot)\}} - 1) + g_3 \cdot e^{-H(p_\theta(\cdot))} \cdot e^{I^\theta\{E_1, p_\theta(\cdot)\}} \cdot (\\
 &(e^{I^\theta\{E_2/E_1, p_\theta(\cdot)\}} - 1) = U_3(E_1) + g_3 \cdot e^{-E_{x_1} H(p_\theta(\cdot/x_1))} \cdot (e^{I^\theta\{E_2/E_1, p_\theta(\cdot)\}} - 1) = \\
 &= U_3(E_1) + U'_3(E_2/E_1)
 \end{aligned}$$

al actuar $p_\theta(\cdot/x_1)$ como densidad inicial en la obtención de la segunda observación.

(c.q.d.)

Aclarando la notación introducida en el anterior teorema hay que resaltar que, dado que en la definición de $U_3(E)$ aparece el factor exponencial de la entropía de la densidad inicial, en una segunda etapa, cuando las opiniones pasan a describirse por la densidad posterior $p_\theta(\cdot/x_1)$, dicho factor debe considerarse definido sobre la entropía esperada $E_{x_1} H(p_\theta(\cdot/x_1))$, toda vez que la densidad anterior depende evidentemente del particular valor obtenido al realizar la extracción de una unidad muestral. Por ello, al escribir $U'_3(E_2/E_1)$ debe entenderse que se trata de la expresión

$$e^{-E_{x_1} H(p_\theta(\cdot/x_1))} \cdot (e^{I^\theta\{E_2/E_1, p_\theta(\cdot)\}} - 1)$$

que aparece en el transcurso de la demostración.

Una demostración ligeramente distinta da lugar a una nueva expresión para $U_3(E)$. En efecto:

TEOREMA 3.2.II.5

Bajo las condiciones del teorema 3.2.II.4 ,

$$U_3(E) = U_3(E_1) \cdot e^{I^\theta\{E_2/E_1, p_\theta(\cdot)\}} + U'_3(E_2/E_1) \cdot e^{-I^\theta\{E_1, p_\theta(\cdot)\}}$$

Demostración.-

$$\begin{aligned}
 U_3(E) &= g_3 \cdot e^{-H(p_\theta(\cdot))} (e^{I^\theta\{E, p_\theta(\cdot)\}} - 1) = g_3 \cdot e^{-H(p_\theta(\cdot))} \cdot (\\
 &(e^{I^\theta\{E_1, p_\theta(\cdot)\}} \cdot e^{I^\theta\{E_2/E_1, p_\theta(\cdot)\}} - e^{I^\theta\{E_2/E_1, p_\theta(\cdot)\}} + e^{I^\theta\{E_2/E_1, p_\theta(\cdot)\}} - 1) =
 \end{aligned}$$

$$\begin{aligned}
&= g_3 \cdot e^{-H(p_\theta(\cdot))} \cdot e^{I^\theta\{E_2/E_1, p_\theta(\cdot)\}} \cdot (e^{I^\theta\{E_1, p_\theta(\cdot)\}} - 1) + \\
&+ g_3 \cdot e^{-H(p_\theta(\cdot))} \cdot (e^{I^\theta\{E_2/E_1, p_\theta(\cdot)\}} - 1) = U_3(E_1) \cdot e^{I^\theta\{E_2/E_1, p_\theta(\cdot)\}} + \\
&+ g_3 \cdot e^{-H(p_\theta(\cdot)) + I^\theta\{E_1, p_\theta(\cdot)\}} \cdot e^{-I^\theta\{E_1, p_\theta(\cdot)\}} \cdot (e^{I^\theta\{E_2/E_1, p_\theta(\cdot)\}} - 1) = \\
&= U_3(E_1) \cdot e^{I^\theta\{E_2/E_1, p_\theta(\cdot)\}} + g_3 \cdot e^{-E_{x_1} H(p_\theta(\cdot/x_1))} \cdot (\\
&(e^{I^\theta\{E_2/E_1, p_\theta(\cdot)\}} - 1) \cdot e^{-I^\theta\{E_1, p_\theta(\cdot)\}} = U_3(E_1) \cdot e^{I^\theta\{E_2/E_1, p_\theta(\cdot)\}} + \\
&+ U_3'(E_2/E_1) \cdot e^{-I^\theta\{E_1, p_\theta(\cdot)\}}
\end{aligned}$$

(c.q.d.)

De los dos teoremas inmediatos anteriores se deduce inmediatamente la siguiente igualdad:

COROLARIO 3.2.II.4

Bajo las hipótesis del teorema 3.2.II.4 ,

$$\begin{aligned}
U_3(E_1) + U_3'(E_2/E_1) &= U_3(E_1) \cdot e^{I^\theta\{E_2/E_1, p_\theta(\cdot)\}} + \\
&+ U_3'(E_2/E_1) \cdot e^{-I^\theta\{E_1, p_\theta(\cdot)\}}
\end{aligned}$$

Nuevamente puede apreciarse en todos estos resultados que bajo $U_3(E)$ la utilidad obtenida en un proceso de muestreo bietápico es la proporcionada por la primera observación más una utilidad esperada remanente proporcionada por el experimento E_2/E_1 , esto es, proporcionada por la obtención de una segunda observación suponiendo conocido el resultado de la primera. Sin embargo, también como en casos anteriores, con estadísticos suficientes ya en la primera etapa no se incrementa la utilidad esperada al obtener una segunda observación. Este es, pues, el sentido del corolario siguiente:

COROLARIO 3.2.II.5

Si x_1 es suficiente para $x = (x_1, x_2)$ con respecto a θ , en el sentido de que $p_\theta(\cdot/x_1) = p_\theta(\cdot/x)$, entonces

$$U_3(E) = U_3(E_1)$$

Demostración.- En efecto,

$$\begin{aligned}
 U_3(E) &= g_3 \cdot \exp(-H(p_\theta(\cdot))) \cdot (e^{I^\theta\{E, p_\theta(\cdot)\}} - 1) = \\
 &= g_3 \cdot e^{-H(p_\theta(\cdot))} \cdot (e^{I^\theta\{E_1, p_\theta(\cdot)\}} - 1) = U_3(E_1) \\
 &\hspace{15em} (\text{c.q.d.})
 \end{aligned}$$

El anterior corolario también podría haberse demostrado sin más que considerar que cuando las densidades inicial y final coinciden al no haber variación en las opiniones del investigador ante la información proporcionada por la muestra, la entropía esperada para el experimento E coincide con la obtenida para el experimento E_1 , esto es,

$$E_x H(p_\theta(\cdot/x)) = E_{x_1} H(p_\theta(\cdot/x_1))$$

con $p_\theta(\cdot/x) = p_\theta(\cdot/x_1)$. En efecto, empleando la notación $X = X_1 \times X_2$, resulta:

$$\begin{aligned}
 E_x H(p_\theta(\cdot/x)) &= \int_{\Theta} \int_X p_{\theta x}(\theta, x) \cdot \log \frac{p_\theta(\theta/x)}{p_\theta(\theta)} \cdot dx \cdot d\theta = \\
 &= \int_{\Theta} \int_X p_{\theta x}(\theta, x) \cdot \log \frac{p_\theta(\theta/x_1)}{p_\theta(\theta)} \cdot dx \cdot d\theta = \int_{\Theta} \int_{X_1} p_{\theta x}(\theta, x_1) \cdot \log \frac{p_\theta(\theta/x_1)}{p_\theta(\theta)} \cdot dx_1 \cdot d\theta = \\
 &= E_{x_1} H(p_\theta(\cdot/x_1)) .
 \end{aligned}$$

De ahí que definida $U_3(E)$ a partir del incremento absoluto en la capacidad de predicción:

$$\begin{aligned}
 U_3(E) &= g_3 \cdot (e^{-E_x H(p_\theta(\cdot/x))} - e^{-H(p_\theta(\cdot))}) = \\
 &= g_3 \cdot (e^{-E_{x_1} H(p_\theta(\cdot/x_1))} - e^{-H(p_\theta(\cdot))}) = U_3(E_1)
 \end{aligned}$$

que no es sino la confirmación del resultado del corolario 3.2.II.5 .

Apliquemos nuevamente todos estos resultados al modelo normal. Se tendría:

$$\begin{aligned}
 U_3(E) &= g_3 \cdot \frac{\sqrt{2\sigma_0^2 + \sigma^2} - \sigma}{\sigma \sigma_0 \sqrt{2\pi e}} \\
 U_3(E) &= g_3 \cdot \frac{\sqrt{\sigma_0^2 + \sigma^2} - \sigma}{\sigma \sigma_0 \sqrt{2\pi e}}
 \end{aligned}$$

$$U_3'(E_2/E_1) = g_3 \cdot \frac{\sqrt{G^2 + 2G_0^2} - \sqrt{G^2 + G_0^2}}{GG_0 \sqrt{2\pi e}}$$

verificándose la igualdad expresada en el Teorema 3.2.II.4 :

$$g_3 \cdot \frac{\sqrt{2G_0^2 + G^2} - G}{GG_0 \sqrt{2\pi e}} = g_3 \cdot \frac{\sqrt{G_0^2 + G^2} - G}{GG_0 \sqrt{2\pi e}} + g_3 \cdot \frac{\sqrt{G^2 + 2G_0^2} - \sqrt{G^2 + G_0^2}}{GG_0 \sqrt{2\pi e}}$$

En el caso de una población uniforme y densidades iniciales descritas por la densidad de Pareto la aplicación del correspondiente comportamiento de la tercera de las funciones de la utilidad esperada ante el experimento suma se realizaría de acuerdo con las expresiones:

$$U_3(E) = 2g_3 / \vartheta_0 \exp((1+k)/k)$$

$$U_3(E_1) = g_3 / \vartheta_0 \exp((1+k)/k)$$

$$\begin{aligned} U_3'(E_2/E_1) &= g_3 \cdot e^{-E_{x_1} H(p_\vartheta(\cdot/x_1))} \cdot (e^{I^\vartheta\{E_2/E_1, p_\vartheta(\cdot)\}} - 1) = \\ &= g_3 \cdot \frac{k+1}{\vartheta_0 \cdot e^k} \cdot \left(\frac{1}{k+1} \right) = g_3 / \vartheta_0 \exp((1+k)/k) \end{aligned}$$

cumpliéndose la evidente igualdad:

$$2g_3 / \vartheta_0 \exp((1+k)/k) = g_3 / \vartheta_0 \exp((1+k)/k) + g_3 / \vartheta_0 \exp((1+k)/k) .$$

III. LA UTILIDAD ESPERADA COMO FUNCION DEL TAMAÑO MUESTRAL

La información esperada sobre ϑ proporcionada por un experimento $E(n)$ y considerada como función del tamaño muestral, es cóncava y creciente (Lindley, 1956) . Verifica, por tanto, para cualquier valor de $n \geq 2$, las siguientes desigualdades:

$$\begin{aligned} I^\vartheta\{E(n), p_\vartheta(\cdot)\} - I^\vartheta\{E(n-1), p_\vartheta(\cdot)\} &\geq 0 \\ 2 \cdot I^\vartheta\{E(n), p_\vartheta(\cdot)\} - I^\vartheta\{E(n+1), p_\vartheta(\cdot)\} - I^\vartheta\{E(n-1), p_\vartheta(\cdot)\} &\geq 0 . \end{aligned}$$

Este comportamiento de la información esperada no es sino confirmación de la creencia común de que debe ganarse información cuando se realizan experimentos de mayor tamaño muestral, al tiempo que la ganancia en la información marginal disminuye ante observaciones independientes y equidistribuidas. Veamos ahora qué funciones de la utilidad esperada $U_i(E(n))$ ($i=1,2,3$) siguen

teniendo este carácter creciente y cóncavo como funciones del tamaño muestral.

Dado que, por definición, la constante g_1 es positiva, las propiedades de la información esperada se verifican trivialmente para $U_1(E(n))$. De ahí que (Lindley, 1956) se siga de forma inmediata el siguiente resultado:

TEOREMA 3.2.III.1

$U_1(E(n))$ es una función creciente y cóncava de n .

La interpretación de este resultado es idéntica a la establecida para la información esperada. Supone una doble confirmación para $U_1(E(n))$: por una parte, confirma su empleo como función de la utilidad esperada; por otra, confirma el hecho empírico, de aceptación común, de la disminución marginal en las utilidades esperadas conforme aumenta el tamaño muestral, y para observaciones independientes y equidistribuidas. Puede, por ello, afirmarse que, expresada la utilidad esperada a través de $U_1(E(n))$, el incremento de utilidad al pasar de la muestra $x(n)$ a la $x(n+1)$ es menor, pero nunca negativo, que el experimentado al pasar del resultado $x(n-1)$ al $x(n)$.

La aplicación de este resultado al modelo normal conduce a las desigualdades:

$$U_1(E(n)) - U_1(E(n-1)) = \frac{1}{2}g_1 \cdot \log\left(\frac{\sigma^2 + n\sigma_0^2}{\sigma^2 + (n-1)\sigma_0^2}\right) > 0$$

$$2U_1(E(n)) - U_1(E(n+1)) - U_1(E(n-1)) = g_1 \cdot \log\left(\frac{\sigma^2 + n\sigma_0^2}{\sqrt{(\sigma^2 + n\sigma_0^2)^2 - \sigma_0^4}}\right) > 0$$

Para una población uniforme y una densidad inicial de Pareto, sería:

$$U_1(E(n)) - U_1(E(n-1)) = g_1 \cdot \log\left(\frac{n+k}{n-1+k}\right) > 0$$

$$2U_1(E(n)) - U_1(E(n+1)) - U_1(E(n-1)) = g_1 \cdot \log\left(\frac{(k+n)^2}{(k+n)^2 - 1}\right) > 0$$

El comportamiento seguido por $U_2(E(n))$ como función de n es idéntico al señalado para $U_1(E(n))$. Efectivamente:

TEOREMA 3.2.III.2

$U_2(E(n))$ es una función creciente y cóncava de n .

Demostración.- Si es creciente, pues

$$U_2(E(n)) - U_2(E(n-1)) = g_2 \cdot (e^{-I^\theta\{E(n-1), p_\theta(\cdot)\}} - e^{-I^\theta\{E(n), p_\theta(\cdot)\}}) \geq 0$$

dado que la información esperada es creciente y la constante g_2 es positiva.

Para demostrar la concavidad basta comprobar que

$$\Delta = (U_2(E(n)) - U_2(E(n-1))) - (U_2(E(n+1)) - U_2(E(n)))$$

es positivo. Y, según acaba de verse,

$$U_2(E(n)) - U_2(E(n-1)) = g_2 \cdot (e^{-I^\theta\{E(n-1), p_\theta(\cdot)\}} - e^{-I^\theta\{E(n), p_\theta(\cdot)\}})$$

$$U_2(E(n+1)) - U_2(E(n)) = g_2 \cdot (e^{-I^\theta\{E(n), p_\theta(\cdot)\}} - e^{-I^\theta\{E(n+1), p_\theta(\cdot)\}})$$

con lo que su diferencia resulta:

$$\begin{aligned} \Delta &= g_2 \cdot (e^{-I^\theta\{E(n+1), p_\theta(\cdot)\}} + e^{-I^\theta\{E(n-1), p_\theta(\cdot)\}} - 2e^{-I^\theta\{E(n), p_\theta(\cdot)\}}) = \\ &= g_2 \cdot e^{-I^\theta\{E(n), p_\theta(\cdot)\}} (e^{I^\theta\{E(n), p_\theta(\cdot)\}} - e^{I^\theta\{E(n+1), p_\theta(\cdot)\}} + \\ &+ e^{I^\theta\{E(n), p_\theta(\cdot)\}} - e^{I^\theta\{E(n-1), p_\theta(\cdot)\}} - 2) . \end{aligned}$$

Dado que $e^x \geq 1 + x$, se deduce que

$$\begin{aligned} &g_2 \cdot e^{-I^\theta\{E(n), p_\theta(\cdot)\}} (e^{I^\theta\{E(n), p_\theta(\cdot)\}} - e^{I^\theta\{E(n+1), p_\theta(\cdot)\}} + 1 + \\ &+ e^{I^\theta\{E(n), p_\theta(\cdot)\}} - e^{I^\theta\{E(n-1), p_\theta(\cdot)\}} + 1 - 2) = g_2 \cdot e^{-I^\theta\{E(n), p_\theta(\cdot)\}} (\\ &(2e^{I^\theta\{E(n), p_\theta(\cdot)\}} - e^{I^\theta\{E(n+1), p_\theta(\cdot)\}} - e^{I^\theta\{E(n-1), p_\theta(\cdot)\}}) \geq 0 \end{aligned}$$

puesto que la información esperada es, como ya se ha indicado, cóncava respecto n .

(c.q.d.)

Nuevamente el teorema anterior pone de manifiesto la doble confirmación a la que aludíamos para $U_1(E(n))$ y que corroboran el uso de

$U_2(E(n))$ como una efectiva función de la utilidad esperada. Este comportamiento en el modelo normal queda patentizado con las siguientes desigualdades:

$$U_2(E(n)) - U_2(E(n-1)) = g_2 \cdot \left(\frac{\sigma}{\sqrt{\sigma^2 + (n-1)\sigma_0^2}} - \frac{\sigma}{\sqrt{\sigma^2 + n\sigma_0^2}} \right) > 0$$

$$2U_2(E(n)) - U_2(E(n+1)) - U_2(E(n-1)) = \sigma g_2 \cdot \left(\frac{1}{\sqrt{\sigma^2 + (n-1)\sigma_0^2}} + \frac{1}{\sqrt{\sigma^2 + (n+1)\sigma_0^2}} - \frac{2}{\sqrt{\sigma^2 + n\sigma_0^2}} \right) > 0$$

Y en el caso de una población uniforme con densidad inicial de Pareto, resultarían las siguientes desigualdades:

$$U_2(E(n)) - U_2(E(n-1)) = g_2 \cdot \frac{k}{(k+n)(k+n-1)} > 0$$

$$2U_2(E(n)) - U_2(E(n+1)) - U_2(E(n-1)) = g_2 \cdot \frac{2k}{(k+n)(k+n+1)(k+n-1)} > 0$$

La última de las funciones de la utilidad esperada consideradas no verifica en toda su extensión la tercera de las propiedades de la información esperada. Ello es debido a la especial definición de $U_3(E(n))$ en función de $I^\theta\{E(n), p_\theta(\cdot)\}$ como:

$$U_3(E(n)) = g_3 \cdot e^{-H(p_\theta(\cdot))} \left(e^{I^\theta\{E(n), p_\theta(\cdot)\}} - 1 \right)$$

expresión en la que $e^{-H(p_\theta(\cdot))}$ no depende de n y, por tanto, no influye en el comportamiento de $U_3(E(n))$ como función del tamaño muestral. Si bien la información esperada es cóncava, la función exponencial es convexa. De esta forma la posible concavidad de $U_3(E(n))$ dependerá, en cada caso concreto, de la dominancia existente entre la concavidad de la información esperada y la convexidad de la función exponencial. Por consiguiente, $U_3(E(n))$ no respeta, en general, el decrecimiento de la utilidad marginal conforme aumenta el tamaño muestral. No obstante, la inversa tampoco es cierta como se pondrá de manifiesto para el modelo normal. En definitiva, el comportamiento cóncavo o convexo de $U_3(E(n))$ como función de n dependerá de las especiales hipótesis de partida sobre las densidades de probabilidad de la pobla-

ción y de la cantidad aleatoria θ de cada problema particular.

Precisamente el modelo propuesto en la Sección 2.2 pone de manifiesto el carácter no estrictamente cóncavo de $U_3(E(n))$, puesto que bajo las hipótesis de dicho modelo la tercera de las funciones propuestas para representar la utilidad esperada es una función lineal del tamaño muestral. En efecto,

$$U_3(E(n)) = g_3 \cdot n / \vartheta_0 \exp((k+1)/k)$$

de forma que la desigualdad empleada en esta sección

$$2U_3(E(n)) - U_3(E(n+1)) - U_3(E(n-1))$$

toma siempre el valor cero.

Sin embargo, $U_3(E(n))$ sí respeta el hecho empírico de la ganancia de utilidad conforme aumenta el tamaño del experimento. El siguiente teorema demuestra que $U_3(E(n))$ sigue siendo una función creciente del tamaño muestral:

TEOREMA 3.2.III.3

$$U_3(E(n)) \text{ es una función creciente de } n .$$

Demostración.- Es inmediata, pues

$$U_3(E(n)) - U_3(E(n-1)) = g_3 \cdot e^{-H(p_\theta(\cdot))} (e^{I^\vartheta\{E(n), p_\theta(\cdot)\}} - e^{I^\vartheta\{E(n-1), p_\theta(\cdot)\}}) \geq 0$$

dado que la información esperada es creciente y tanto la constante g_3 como la función exponencial son positivas.

(c.q.d.)

La aplicación de la desigualdad anterior al modelo normal tomaría la forma:

$$U_3(E(n)) - U_3(E(n-1)) = g_3 \cdot \frac{\sqrt{\sigma^2 + n\sigma_0^2} - \sqrt{\sigma^2 + (n-1)\sigma_0^2}}{\sigma\sigma_0\sqrt{2ne}} > 0$$

Y para el segundo modelo de los considerados:

$$U_3(E(n)) - U_3(E(n-1)) = g_3 / \vartheta_0 e^{\frac{1+k}{k}} > 0 .$$

Nuevamente la tercera de las funciones de la utilidad esperada falla en una importante propiedad que respeta la común creencia de la disminución en la utilidad marginal conforme va creciendo el tamaño muestral.

Esta circunstancia implica, evidentemente, una nueva disminución en la aplicabilidad de $U_3(E(n))$. No obstante, y por dos motivos fundamentales que vamos a indicar, la función de utilidad $U_3(E(n))$ puede admitirse como correcta para describir ciertas situaciones concretas. En primer lugar, aquellas aplicaciones prácticas en las que se dispongan de datos acerca del incremento absoluto producido en la capacidad de predicción por el hecho de realizar el experimento $E(n)$ pueden necesitar describirse por utilidades donde existan marcados crecimientos marginales en todo el proceso o en parte de él. Por ejemplo, esto es posible en circunstancias en las que, a priori, el desconocimiento sobre la cantidad de interés sea total, necesitándose llegar a un determinado tamaño muestral n_0 a partir del cual las sucesivas informaciones disminuyan en importancia. En segundo lugar, tal como queda recogido en el teorema siguiente, en el modelo normal sí existe un comportamiento en la línea demostrada para las otras dos funciones de la utilidad esperada, y esto a pesar de que la concavidad respecto n no sea una propiedad general para $U_3(E(n))$. Este hecho es interesante en cuanto que, con frecuencia, se utilizan en la práctica aproximaciones al modelo normal. Por tanto, con tamaños muestrales no excesivamente grandes y para ciertas aplicaciones, podrá aceptarse con gran credibilidad que también la utilidad esperada $U_3(E(n))$ sigue un comportamiento cóncavo como función de n , respetándose de esta forma la creencia - no exclusiva para todas las aplicaciones - de la disminución de las utilidades marginales conforme crezca el tamaño muestral. El teorema al que se ha aludido es el siguiente:

TEOREMA 3.2.III.4

En el modelo normal, $U_3(E(n))$ es una función cóncava de n .

Demostración.- Bajo las hipótesis del modelo normal

$$U_3(E(n)) = g_3 \cdot \left(\sqrt{1 + n \frac{\sigma_0^2}{\sigma^2}} - 1 \right) / \sigma_0 \sqrt{2\pi e}$$

Considerando la función de variable real que admite la misma expresión que $U_3(E(n))$, esto es,

$$y(x) = g_3 \cdot \left(\sqrt{1 + x \frac{\sigma_0^2}{\sigma^2}} - 1 \right) / \sigma_0 \sqrt{2\pi e}$$

tiene como segunda derivada

$$y''(x) = -g_3 \cdot \left(\frac{G_0}{G}\right)^3 \cdot \left(1 + x \frac{G_0^2}{G^2}\right)^{-3/2} / 4 \sqrt{2\pi e}$$

expresión que es menor que cero para cualquier x positivo. Y dado que n es combinación lineal de $n-1$ y $n+1$,

$$n = \frac{1}{2}(n-1) + \frac{1}{2}(n+1)$$

y que $y(x)$ es cóncava en las abscisas positivas, resultará

$$y(n) \geq \frac{1}{2}y(n-1) + \frac{1}{2}y(n+1)$$

esto es,

$$2U_3(E(n)) \geq U_3(E(n+1)) + U_3(E(n-1))$$

(c.q.d.)

IV. COMPORTAMIENTO DE LAS FUNCIONES DE UTILIDAD COMO FUNCIONALES DE LA DENSIDAD INICIAL

La concavidad como funcional de la densidad inicial es la última de las propiedades características de la información esperada. Se trata de una propiedad de difícil interpretación práctica, admitiendo representación gráfica solamente en el caso de que el espacio paramétrico se reduzca a dos valores, circunstancia en la que cualquier densidad inicial queda expresada por un valor real en el intervalo $[0,1]$. Sin embargo se trata de un instrumento para desarrollar la teoría del diseño de experimentos óptimos por maximización de una función de información. Veamos, pues, qué expresiones íntimamente relacionadas con la información esperada verifican esta propiedad y, en particular, qué funciones $U_i(E)$ ($i=1,2,3$) son cóncavas como funcionales de la densidad inicial $p_\theta(\cdot)$.

Consideremos el siguiente experimento: con probabilidad λ (para todos los valores de θ) realizamos el experimento $E_1 = \{X_1, \mathcal{G}_1, p_{X_1}(\cdot/\theta_1)\}$; con probabilidad $1-\lambda$ (para todos los valores de θ) se realiza el experimento $E_2 = \{X_2, \mathcal{G}_2, p_{X_2}(\cdot/\theta_2)\}$, suponiendo que tanto para E_1 como para E_2 el espacio paramétrico es idéntico, $\mathcal{G}_1 = \mathcal{G}_2 = \mathcal{G}$. La observación consistirá en la extracción de una unidad muestral de acuerdo con alguno de los dos experimentos que, a su vez, son elegidos de acuerdo con las probabilidades asignadas de λ ó $1-\lambda$, respectivamente. Así pues, obtenido un determinado dato existe una doble incertidumbre: la implicada por el parámetro desconocido θ y la derivada por el desconocimiento de qué experimento E_i ($i=1,2$) es el realizado para obtener la observación. Este experimento compuesto se denotará por la expresión $(\lambda E_1 + (1-\lambda)E_2)$. En términos matemáticos sería

$(\lambda E_1 + (1-\lambda)E_2) = \{X = X_1 \cup X_2, \Theta, P\}$ donde P es el conjunto de funciones de densidad $p_x(x/\theta)$ definidas como sigue: si $x \in X_1$, entonces $p_x(x/\theta) = \lambda p_{x_1}(x_1/\theta)$ con $x=x_1$; si $x \in X_2$, entonces $p_x(x/\theta) = (1-\lambda) \cdot p_{x_2}(x_2/\theta)$ con $x=x_2$. Así pues, $p_x(x/\theta)$ se trata de una combinación lineal convexa de las densidades $p_{x_1}(x_1/\theta)$ y $p_{x_2}(x_2/\theta)$. Lindley (1956) demuestra los resultados siguientes:

$$I^\theta \{ \lambda E_1 + (1-\lambda)E_2, p_\theta(\cdot) \} = \lambda I^\theta \{ E_1, p_{1\theta}(\cdot) \} + (1-\lambda) I^\theta \{ E_2, p_{2\theta}(\cdot) \}$$

y, manteniendo un experimento fijo E , $I^\theta \{ E, p_\theta(\cdot) \}$ es una funcional cóncava de la densidad inicial $p_\theta(\cdot)$, esto es, siendo $p_{1\theta}(\cdot)$, $p_{2\theta}(\cdot)$ y $p_\theta(\cdot)$ densidades de probabilidad tales que

$$p_\theta(\cdot) = \lambda p_{1\theta}(\cdot) + (1-\lambda) p_{2\theta}(\cdot)$$

con $0 \leq \lambda \leq 1$, resulta ser

$$I^\theta \{ E, p_\theta(\cdot) \} \geq \lambda I^\theta \{ E, p_{1\theta}(\cdot) \} + (1-\lambda) \cdot I^\theta \{ E, p_{2\theta}(\cdot) \} .$$

Notemos que siendo $p_{1\theta}(\cdot)$ y $p_{2\theta}(\cdot)$ funciones de densidad, cualquier combinación lineal convexa de ellas sigue siendo función de densidad, lo que implica que la densidad anterior esté bien definida. La desigualdad en sentido inverso define a una funcional convexa.

A fin de comprobar los resultados posteriores es conveniente desarrollar previamente un sencillo modelo en el que el espacio paramétrico es bipuntual y que servirá para proporcionar una representación gráfica del carácter cóncavo de la información esperada y de otras expresiones relacionadas con ella consideradas como funcionales de la densidad inicial.

Consideremos, pues, un espacio muestral X en el que toma valores cierta cantidad aleatoria x que se distribuye uniformemente en $(0, \theta)$. La cantidad de interés desconocida admite dos posibles valores, el 1 y el 2, siendo por tanto el espacio paramétrico el $\Theta = \{\theta_1 = 1, \theta_2 = 2\}$. La densidad inicial vendrá dada, pues, por el par $(p, 1-p)$ que describe las opiniones iniciales que el investigador posee sobre la ocurrencia de los dos posibles estados de la naturaleza θ_1 y θ_2 , respectivamente. Bajo estas hipótesis se demuestra el siguiente

TEOREMA 3.2.IV.1

Si x es una cantidad aleatoria uniforme en $(0, \theta)$, y θ toma los

valores 1 y 2 con probabilidades p y $1-p$, respectivamente ($p \in (0,1)$), resulta ser

$$I^{\vartheta}\{E,p\} = p \cdot \log 2 - \frac{1}{2}(1+p) \cdot \log(1+p) - \frac{1}{2}(1-p) \cdot \log(1-p)$$

Demostración.- Se tienen las siguientes densidades de probabilidad:

$$p_x(\cdot/\vartheta) = 1/\vartheta \quad \text{en } (0, \vartheta), \text{ esto es, } p_x(\cdot/\vartheta) = \begin{cases} 1 & x \in (0,1) \text{ y } \vartheta = 1 \\ \frac{1}{2} & x \in (0,2) \text{ y } \vartheta = 2 \end{cases}$$

$$p_{\vartheta}(\cdot) = \begin{cases} p & \text{si } \vartheta = 1 \\ 1-p & \text{si } \vartheta = 2 \end{cases}$$

$$p_{\vartheta x}(\cdot, \cdot) = \begin{cases} p & \text{en } (0,1) \times 1 \\ \frac{1}{2}(1-p) & \text{en } (0,2) \times 2 \end{cases}$$

$$p_x(\cdot) = \begin{cases} \frac{1}{2}(1+p) & \text{en } (0,1) \\ \frac{1}{2}(1-p) & \text{en } (1,2) \end{cases}$$

$$p_{\vartheta}(\cdot/x) = \begin{cases} \frac{2p}{1+p} & \text{en } \vartheta = 1 \quad \text{y} \quad \frac{1-p}{1+p} & \text{en } \vartheta = 2 \quad \text{cuando } x \in (0,1) \\ 0 & \text{en } \vartheta = 1 \quad \text{y} \quad 1 & \text{en } \vartheta = 2 \quad \text{cuando } x \in (1,2) \end{cases}$$

La entropía y la entropía esperada admiten las siguientes expresiones:

$$H(p) = - \sum_{\vartheta_i \in \mathcal{O}} p_{\vartheta}(\vartheta_i) \cdot \log p_{\vartheta}(\vartheta_i) = -p \cdot \log p - (1-p) \cdot \log(1-p)$$

$$E_x H(p) = - \int \left\{ \sum_{\vartheta_i \in \mathcal{O}} p_x(x/\vartheta_i) \cdot p_{\vartheta}(\vartheta_i) \cdot \log p_{\vartheta}(\vartheta_i/x) \right\} dx =$$

$$= - \int_0^1 \left\{ p \cdot \log \frac{2p}{1+p} + \frac{1}{2}(1-p) \cdot \log \frac{1-p}{1+p} \right\} dx = -p \cdot \log 2 - p \cdot \log p -$$

$$- \frac{1}{2}(1-p) \cdot \log(1-p) + \frac{1}{2}(1+p) \cdot \log(1+p)$$

Como ya quedó indicado en su momento, en los cálculos anteriores se ha supuesto que la expresión $0 \cdot \log 0$ toma el valor cero. Finalmente, restando los dos resultados anteriores

$$I^{\vartheta}\{E,p\} = H(p) - E_x H(p) = p \cdot \log 2 - \frac{1}{2}(1+p) \cdot \log(1+p) - \frac{1}{2}(1-p) \cdot \log(1-p)$$

(c.q.d.)

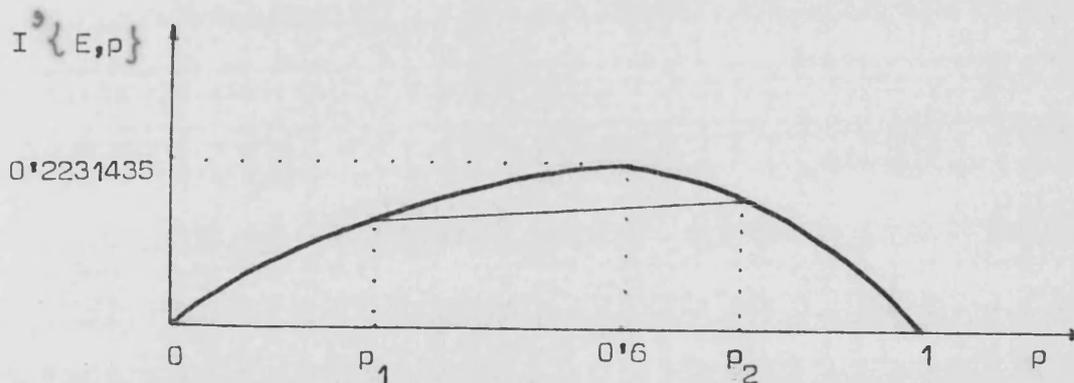
Este resultado permite comprobar prácticamente la concavidad de la información esperada como funcional de la densidad inicial. En efecto, con una variable real la desigualdad

$$I^\vartheta \{E, p_\vartheta(\cdot)\} \geq \lambda I^\vartheta \{E, p_{1\vartheta}(\cdot)\} + (1-\lambda) \cdot I^\vartheta \{E, p_{2\vartheta}(\cdot)\}$$

que define la concavidad de la información esperada como funcional de $p_\vartheta(\cdot)$ es equivalente a que la segunda derivada de $I(E, p)$, cuando exista, con respecto a p sea no positiva (Fleming, 1969). En el sencillo modelo anterior

$$\frac{d^2 I^\vartheta \{E, p\}}{dp^2} = - \frac{1}{1-p^2}$$

expresión que es menor que cero para cualquier $p \in]0, 1[$, confirmando así el resultado encontrado por Lindley (1956) sobre la concavidad de la información esperada. El mismo ejemplo proporciona, además, una representación gráfica de la información esperada en función de p . Dicha gráfica sería la siguiente:



Representación gráfica de $I^\vartheta \{E, p\}$ como función de p .

Del examen de la anterior figura puede asegurarse que cuando se admita a priori que la probabilidad de ocurrencia para el valor ϑ_1 del parámetro ϑ está comprendido entre los valores p_1 y p_2 , al obtener el dato muestral x a fin de adquirir información sobre el verdadero valor de ϑ se obtiene una información esperada de, al menos, la de la recta que pasa por los puntos $(p_1, I^\vartheta \{E, p_1\})$ y $(p_2, I^\vartheta \{E, p_2\})$, esto es, la recta de ecuación

$$I^\vartheta \{E, p\} = \frac{p-p_1}{p_2-p_1} \cdot (I^\vartheta \{E, p_2\} - I^\vartheta \{E, p_1\}) + I^\vartheta \{E, p_1\}$$

De esta forma puede quedar establecida una cota inferior para la información esperada de este caso concreto de espacio paramétrico bipuntual,

permitiendo adquirir una cierta idea de dicha información cuando, ante la im-
 posibilidad de precisar una probabilidad inicial, el investigador sí es ca-
 paz de incluirla dentro de un intervalo de posibilidades. Generalizando este
 comentario puede afirmarse que, para aquellas situaciones en las que el cál-
 culo de la información esperada es de difícil obtención, su concavidad
 como funcional de la densidad inicial puede proporcionar una adecuada cota in-
 ferior. En efecto, si para una cierta densidad inicial $p_{\theta}(\cdot)$ la obtención
 de $I^{\theta}\{E, p_{\theta}(\cdot)\}$ resulta complicada, conocida dicha información para otras
 dos densidades iniciales $p_{1\theta}(\cdot)$ y $p_{2\theta}(\cdot)$ relacionadas con la primera a
 través de la combinación lineal convexa $p_{\theta}(\cdot) = \lambda p_{1\theta}(\cdot) + (1-\lambda)p_{2\theta}(\cdot)$,
 podrá asegurarse que, al menos, la información esperada proporcionada por el
 experimento $E = \{X, \mathcal{G}, p_x(\cdot/\theta)\}$ cuando las opiniones iniciales vienen des-
 critas por $p_{\theta}(\cdot)$, será

$$\lambda I^{\theta}\{E, p_{1\theta}(\cdot)\} + (1-\lambda) \cdot I^{\theta}\{E, p_{2\theta}(\cdot)\} .$$

Estudieemos ahora el comportamiento seguido al respecto por la
 entropía y la entropía esperada.

TEOREMA 3.2.IV.2

$H(p_{\theta}(\cdot))$ es una funcional cóncava de $p_{\theta}(\cdot)$.

Demostración.- Denotando por

$$\Delta H(p_{\theta}(\cdot)) = H(p_{\theta}(\cdot)) - \lambda H(p_{1\theta}(\cdot)) - (1-\lambda)H(p_{2\theta}(\cdot))$$

y siendo $p_{\theta}(\cdot) = \lambda p_{1\theta}(\cdot) + (1-\lambda) \cdot p_{2\theta}(\cdot)$, bastará demostrar que $\Delta H(p_{\theta}(\cdot))$
 es no negativo. Y así:

$$\begin{aligned} \Delta H(p_{\theta}(\cdot)) &= - \int p_{\theta}(\vartheta) \cdot \log p_{\theta}(\vartheta) \cdot d\vartheta + \lambda \int p_{1\theta}(\vartheta) \cdot \log p_{1\theta}(\vartheta) \cdot d\vartheta + \\ &+ (1-\lambda) \int p_{2\theta}(\vartheta) \cdot \log p_{2\theta}(\vartheta) \cdot d\vartheta = \lambda \int p_{1\theta}(\vartheta) \cdot \log \frac{p_{1\theta}(\vartheta)}{p_{\theta}(\vartheta)} \cdot d\vartheta + \\ &+ (1-\lambda) \int p_{2\theta}(\vartheta) \cdot \log \frac{p_{2\theta}(\vartheta)}{p_{\theta}(\vartheta)} \cdot d\vartheta \geq 0 \end{aligned}$$

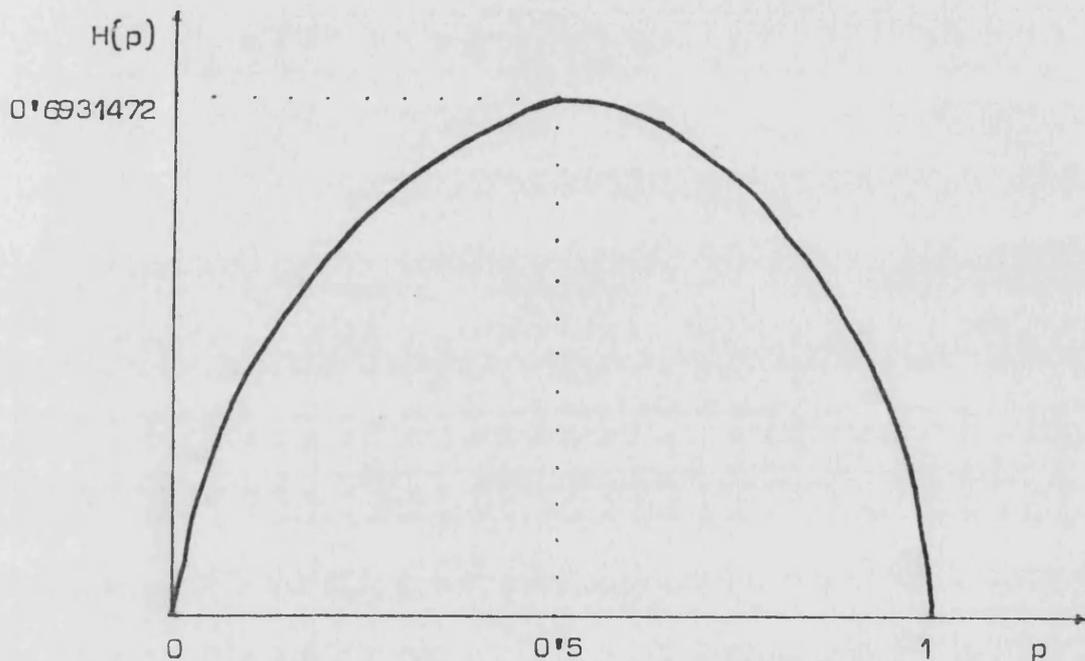
pues las integrales anteriores son positivas, sin más que considerar la desi-
 gualdad de Hardy, Littlewood & Pólya (1952), con $p_{1\theta}(\vartheta) / p_{\theta}(\vartheta)$ como
 función $u(\vartheta)$.

(c.q.d.)

En el ejemplo desarrollado resultaría:

$$\frac{d^2 H(p)}{dp^2} = \frac{-1}{p(1-p)}$$

cociente negativo para cualquier $p \in]0,1[$ (en los extremos la segunda derivada anterior tiende a $-\infty$), verificándose, pues, el teorema 3.2.IV.2. La representación gráfica de $H(p)$ pone de manifiesto su carácter cóncavo en el modelo que sirve de exposición práctica en el estudio de la concavidad como funcional de la densidad inicial.



Representación de $H(p)$ como función de p .

TEOREMA 3.2.IV.3

$E_x H(p_\theta(. / x))$ es una funcional cóncava de $p_\theta(.)$

Demostración.- Denotando por

$$\Delta E_x H(p_\theta(. / x)) = E_x H(p_\theta(. / x)) - \lambda E_x H(p_{1\theta}(. / x)) - (1-\lambda) E_x H(p_{2\theta}(. / x))$$

y siendo $p_\theta(.) = \lambda p_{1\theta}(. / x) + (1-\lambda) p_{2\theta}(. / x)$ $\lambda \in [0,1]$ habrá que demostrar que $\Delta E_x H(p_\theta(. / x))$ es no negativo. En efecto,

$$\Delta E_x H(p_\theta(. / x)) = - \iint p_x(x/\theta) \cdot (\lambda p_{1\theta}(\theta) + (1-\lambda) p_{2\theta}(\theta)) \cdot \log$$

$$\begin{aligned} & \cdot \log \frac{p_x(x/\vartheta) \cdot p_\vartheta(\vartheta)}{p_x(x)} \cdot d\vartheta \cdot dx + \lambda \iint p_x(x/\vartheta) \cdot p_{1\vartheta}(\vartheta) \cdot \log \\ & \cdot \log \frac{p_x(x/\vartheta) \cdot p_{1\vartheta}(\vartheta)}{p_{1x}(x)} \cdot d\vartheta \cdot dx + (1-\lambda) \iint p_x(x/\vartheta) \cdot p_{2\vartheta}(\vartheta) \cdot \log \\ & \cdot \log \frac{p_x(x/\vartheta) \cdot p_{2\vartheta}(\vartheta)}{p_{2x}(x)} \cdot d\vartheta \cdot dx \end{aligned}$$

donde

$$p_\vartheta(\cdot) = \lambda p_{1\vartheta}(\cdot) + (1-\lambda) \cdot p_{2\vartheta}(\cdot)$$

$$p_{ix}(x) = \int p_x(x/\vartheta) \cdot p_{i\vartheta}(\vartheta) \cdot d\vartheta \quad (i=1,2)$$

$$p_x(\cdot) = \lambda p_{1x}(\cdot) + (1-\lambda) \cdot p_{2x}(\cdot)$$

El teorema quedará demostrado comprobando que cada una de las dos integrales siguientes

$$\begin{aligned} \Delta E_x H(p(\cdot/x)) &= \lambda \iint p_x(x/\vartheta) \cdot p_{1\vartheta}(\vartheta) \cdot \log \frac{p_{1\vartheta}(\vartheta) \cdot p_x(x)}{p_{1x}(x) \cdot p_\vartheta(\vartheta)} \cdot d\vartheta \cdot dx + \\ &+ (1-\lambda) \iint p_x(x/\vartheta) \cdot p_{2\vartheta}(\vartheta) \cdot \log \frac{p_{2\vartheta}(\vartheta) \cdot p_x(x)}{p_{2x}(x) \cdot p_\vartheta(\vartheta)} \cdot d\vartheta \cdot dx \end{aligned}$$

es no negativa. En efecto,

$$\begin{aligned} & \iint p_x(x/\vartheta) \cdot p_{i\vartheta}(\vartheta) \cdot \log \frac{p_x(x) \cdot p_{i\vartheta}(\vartheta)}{p_{ix}(x) \cdot p_\vartheta(\vartheta)} \cdot d\vartheta \cdot dx = \\ &= - \iint p_x(x/\vartheta) \cdot p_{i\vartheta}(\vartheta) \cdot \log \frac{p_{ix}(x) \cdot p_\vartheta(\vartheta)}{p_x(x) \cdot p_{i\vartheta}(\vartheta)} \cdot d\vartheta \cdot dx \end{aligned}$$

y puesto que para cualquier función positiva resulta ser $-\log f(x) \geq 1-f(x)$, desarrollando la última integral y enlazando con el cálculo del incremento

$$\Delta E_x H(p_\vartheta(\cdot/x)),$$

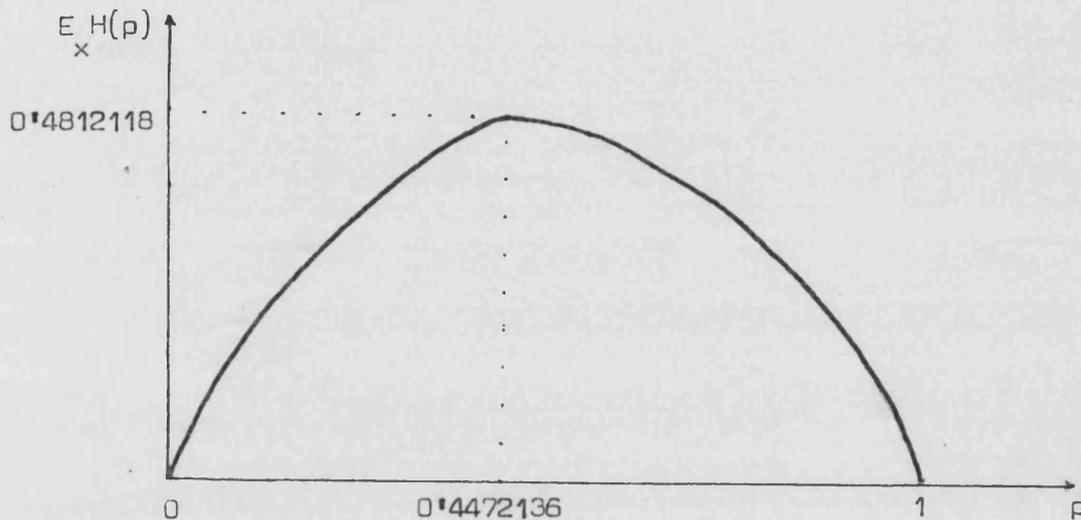
$$\begin{aligned} \Delta E_x H(p_\vartheta(\cdot/x)) &\geq - \iint (p_x(x/\vartheta) \cdot p_\vartheta(\vartheta) \cdot p_{ix}(x) / p_x(x)) \cdot d\vartheta \cdot dx + \\ &+ \iint p_x(x/\vartheta) \cdot p_{i\vartheta}(\vartheta) \cdot d\vartheta \cdot dx = \int p_{ix}(x) \cdot dx - \int p_x(x) \cdot (p_{ix}(x)/p_x(x)) \cdot dx = 0 \end{aligned}$$

(c.q.d.)

En el ejemplo propuesto se verifica que

$$\frac{d^2}{dp^2} E_x H(p_\theta(. / x)) = - \frac{1}{p(1-p)^2}$$

expresión que es menor que cero para $p \in]0, 1[$. La representación gráfica de $E_x H(p)$ muestra su carácter cóncavo:



Representación gráfica de $E_x H(p)$ en función de p .

Siendo $p_\theta(.) = \lambda p_{1\theta}(.) + (1-\lambda) \cdot p_{2\theta}(.)$ $0 \leq \lambda \leq 1$, consideremos las expresiones siguientes:

$$\begin{aligned} H(p_\theta(.)) - \lambda H(p_{1\theta}(.)) - (1-\lambda) \cdot H(p_{2\theta}(.)) &= \Delta H(p_\theta(.)) = \\ &= \lambda \int p_{1\theta}(\vartheta) \cdot \log \frac{p_{1\theta}(\vartheta)}{p_\theta(\vartheta)} \cdot d\vartheta + (1-\lambda) \int p_{2\theta}(\vartheta) \cdot \log \frac{p_{2\theta}(\vartheta)}{p_\theta(\vartheta)} \cdot d\vartheta \geq 0 \end{aligned}$$

$$\begin{aligned} \Delta E_x H(p_\theta(. / x)) &= E_x H(p_\theta(. / x)) - \lambda E_x H(p_{1\theta}(. / x)) - (1-\lambda) \cdot E_x H(p_{2\theta}(. / x)) = \\ &= \lambda \iint p_x(x/\vartheta) \cdot p_{1\theta}(\vartheta) \cdot \log \frac{p_{1\theta}(\vartheta/x)}{p_\theta(\vartheta/x)} \cdot d\vartheta \cdot dx + \\ &\quad + (1-\lambda) \iint p_x(x/\vartheta) \cdot p_{2\theta}(\vartheta) \cdot \log \frac{p_{2\theta}(\vartheta/x)}{p_\theta(\vartheta/x)} \cdot d\vartheta \cdot dx \geq 0 \end{aligned}$$

$$\Delta I^\vartheta \{E, p_\theta(.)\} = I^\vartheta \{E, p_\theta(.)\} - \lambda I^\vartheta \{E, p_{1\theta}(.)\} - (1-\lambda) \cdot I^\vartheta \{E, p_{2\theta}(.)\} =$$

$$= \lambda \iint p_x(x/\theta) \cdot p_{1\theta}(\theta) \cdot \log \frac{p_{1\theta}(\theta) \cdot p_{\theta}(x/\theta)}{p_{1\theta}(\theta/x) \cdot p_{\theta}(\theta)} \cdot d\theta \cdot dx +$$

$$+ (1-\lambda) \cdot \iint p_x(x/\theta) \cdot p_{2\theta}(\theta) \cdot \log \frac{p_{2\theta}(\theta/x) \cdot p_{\theta}(x/\theta)}{p_{2\theta}(\theta/x) \cdot p_{\theta}(\theta)} \cdot d\theta \cdot dx \geq 0$$

verificándose trivialmente que

$$\Delta I^{\theta} \{E, p_{\theta}(\cdot)\} = \Delta H(p_{\theta}(\cdot)) - \Delta E_x H(p_{\theta}(\cdot/x))$$

Con la notación anterior se deducen los dos corolarios siguientes:

COROLARIO 3.2.IV.1

$$\Delta H(p_{\theta}(\cdot)) \geq \Delta E_x H(p_{\theta}(\cdot/x))$$

Demostración.-

$$\Delta H(p_{\theta}(\cdot)) - \Delta E_x H(p_{\theta}(\cdot/x)) = \Delta I^{\theta} \{E, p_{\theta}(\cdot)\} \geq 0$$

(c.q.d.)

COROLARIO 3.2.IV.2

$$\Delta H(p_{\theta}(\cdot)) \geq \Delta I^{\theta} \{E, p_{\theta}(\cdot)\}$$

Demostración.-

$$\Delta H(p_{\theta}(\cdot)) - \Delta I^{\theta} \{E, p_{\theta}(\cdot)\} = \Delta E_x H(p_{\theta}(\cdot/x)) \geq 0$$

(c.q.d.)

Pasemos ahora a estudiar el carácter que, como funcionales de la densidad inicial $p_{\theta}(\cdot)$, tienen las funciones de la utilidad esperada $U_1(E)$ definidas con anterioridad. Concretamente, por su misma definición, $U_1(E)$ es también con evidencia cóncava como funcional de $p_{\theta}(\cdot)$.

TEOREMA 3.2.IV.4

$$U_1(E) \text{ es una funcional cóncava de la densidad inicial } p_{\theta}(\cdot)$$

El comportamiento seguido por las otras dos funciones difiere entre sí. Mientras que para la segunda de dichas funciones se respeta la misma propiedad que para la información esperada, nuevamente $U_3(E)$ tiene un com-

discordante con las otras dos utilidades esperadas. En el modelo que estamos aplicando, de espacio paramétrico bipuntual, las distintas funciones $U_i(E)$ toman las siguientes expresiones concretas:

$$U_1(E) = g_1 \cdot (p \cdot \log 2 - \frac{1}{2}(1+p) \cdot \log(1+p) - \frac{1}{2}(1-p) \cdot \log(1-p))$$

$$U_2(E) = g_2 \cdot \left(1 - \frac{\sqrt{(1+p)^{1+p} \cdot (1-p)^{1-p}}}{2^p} \right)$$

$$U_3(E) = g_3 \cdot \left(2^p \cdot p^p \cdot \sqrt{\frac{(1-p)^{1-p}}{(1+p)^{1+p}}} - p^p (1-p)^p \right)$$

El próximo resultado, demostrado para funcionales, se corresponde con el resultado análogo para funciones de variable real (Fleming, 1969) :

TEOREMA 3.2.IV.5

Si $F(p_\theta(\cdot))$ es una funcional cóncava de $p_\theta(\cdot)$, entonces $\exp(-F(p_\theta(\cdot)))$ es una funcional convexa.

Demostración.- Siendo $p_\theta(\cdot) = \lambda p_{1\theta}(\cdot) + (1-\lambda) \cdot p_{2\theta}(\cdot)$ con $0 \leq \lambda \leq 1$, demostrar que

$$e^{-F(p_\theta(\cdot))} \leq \lambda e^{-F(p_{1\theta}(\cdot))} + (1-\lambda) \cdot e^{-F(p_{2\theta}(\cdot))}$$

es equivalente a ver que

$$1 \leq \lambda e^{F(p_\theta(\cdot)) - F(p_{1\theta}(\cdot))} + (1-\lambda) \cdot e^{F(p_\theta(\cdot)) - F(p_{2\theta}(\cdot))}$$

Dado que $F(p_\theta(\cdot))$ es cóncava,

$$F(p_\theta(\cdot)) \geq \lambda F(p_{1\theta}(\cdot)) + (1-\lambda) \cdot F(p_{2\theta}(\cdot))$$

de donde

$$\begin{aligned} & \lambda e^{F(p_\theta(\cdot)) - F(p_{1\theta}(\cdot))} + (1-\lambda) e^{F(p_\theta(\cdot)) - F(p_{2\theta}(\cdot))} \geq \\ & \geq \lambda e^{(1-\lambda) \cdot (F(p_{2\theta}(\cdot)) - F(p_{1\theta}(\cdot)))} + (1-\lambda) e^{\lambda (F(p_{1\theta}(\cdot)) - F(p_{2\theta}(\cdot)))} \geq \end{aligned}$$

$$\geq \lambda (1 + (1-\lambda) \cdot (F(p_{2\theta}(\cdot)) - F(p_{1\theta}(\cdot)))) + (1-\lambda) \cdot (1 + \lambda ($$

$$(F(p_{1\theta}(\cdot)) - F(p_{2\theta}(\cdot)))) = \lambda + (1-\lambda) = 1$$

(c.q.d.)

El resultado anterior permite demostrar para la utilidad esperada $U_2(E)$ su concavidad como funcional de la densidad inicial $p_\theta(\cdot)$.

TEOREMA 3.2.IV.6

$U_2(E)$ es una funcional cóncava de la densidad inicial $p_\theta(\cdot)$

Demostración.- Dado que $I^\theta\{E, p_\theta(\cdot)\} = H(p_\theta(\cdot)) - E_x H(p_\theta(\cdot/x))$ es cóncava respecto $p_\theta(\cdot)$, resultará ser, con $0 \leq \lambda \leq 1$ y $p_\theta(\cdot) = \lambda p_{1\theta}(\cdot) + (1-\lambda) \cdot p_{2\theta}(\cdot)$, que

$$e^{-H(p_\theta(\cdot)) + E_x H(p_\theta(\cdot/x))}$$

es una funcional convexa de $p_\theta(\cdot)$, en virtud del teorema 3.2.IV.5. De

donde

$$\frac{e^{-H(p_\theta(\cdot))}}{e^{-E_x H(p_\theta(\cdot/x))}} \leq \lambda \frac{e^{-H(p_{1\theta}(\cdot))}}{e^{-E_x H(p_{1\theta}(\cdot/x))}} + (1-\lambda) \frac{e^{-H(p_{2\theta}(\cdot))}}{e^{-E_x H(p_{2\theta}(\cdot/x))}}$$

Así pues, con $g_2 > 0$,

$$g_2 \cdot \left(\frac{e^{-H(p_\theta(\cdot))}}{e^{-E_x H(p_\theta(\cdot/x))}} - 1 \right) \leq \lambda g_2 \cdot \left(\frac{e^{-H(p_{1\theta}(\cdot))}}{e^{-E_x H(p_{1\theta}(\cdot/x))}} - 1 \right) + (1-\lambda) \cdot g_2 \cdot \left(\frac{e^{-H(p_{2\theta}(\cdot))}}{e^{-E_x H(p_{2\theta}(\cdot/x))}} - 1 \right)$$

esto es,

$$g_2 \cdot (1 - e^{-I^\theta\{E, p_\theta(\cdot)\}}) \geq \lambda g_2 \cdot (1 - e^{-I^\theta\{E, p_{1\theta}(\cdot)\}}) + (1-\lambda) g_2 \cdot (1 - e^{-I^\theta\{E, p_{2\theta}(\cdot)\}})$$

(c.q.d.)

Nuevamente el ejemplo propuesto sirve de comprobación al resultado obtenido. En efecto, al ser

$$U_2(E) = g_2 \cdot (1 - e^{-I^\theta\{E, p\}})$$

una función de la variable real p , resulta:

$$\frac{d^2 U_2(E)}{dp^2} = g_2 \cdot e^{-I^\theta\{E,p\}} \left(\frac{d^2 I^\theta\{E,p\}}{dp^2} - \left(\frac{d I^\theta\{E,p\}}{dp} \right)^2 \right)$$

expresión no negativa para valores de p en el intervalo $]0,1[$ al ser

$$\frac{d^2 I^\theta\{E,p\}}{dp^2} = -\frac{1}{1-p^2}.$$

La interpretación de esta propiedad de concavidad para las funciones $U_1(E)$ y $U_2(E)$ es idéntica, en términos de la utilidad esperada, que la efectuada para la información esperada: si la obtención de la utilidad proporcionada por un experimento $E = \{X, \mathcal{D}, p_x(\cdot/\theta)\}$ es difícil para una cierta densidad a priori pero, en cambio, es conocida la utilidad esperada proporcionada por otras dos densidades $p_{1\theta}(\cdot)$ y $p_{2\theta}(\cdot)$ relacionadas con la primera a través de una combinación lineal convexa, la utilidad esperada en el primer caso será, al menos, igual a la idéntica combinación lineal convexa de las utilidades esperadas obtenidas cuando $p_{1\theta}(\cdot)$ y $p_{2\theta}(\cdot)$ son las densidades iniciales.

La tercera función $U_3(E)$ no es necesariamente cóncava como polinomial de $p_\theta(\cdot)$. Efectivamente, dado que la función exponencial es convexa, la posible concavidad o convexidad de $U_3(E)$ dependerá, en cada caso concreto, de la "dominancia" que exista entre las dos exponenciales que definen a $U_3(E)$, aún teniendo en cuenta que, por el corolario 3.2.IV.1, $H(p_\theta(\cdot))$ es más cóncava que $E_x H(p_\theta(\cdot/x))$. Existen, no obstante, cotas inferior y superior para las exponenciales $\exp(-H(p_\theta(\cdot)))$ y $\exp(-E_x H(p_\theta(\cdot/x)))$ que intervienen en la definición de $U_3(E)$.

TEOREMA 3.2.IV.7

Siendo $p_\theta(\cdot) = \lambda p_{1\theta}(\cdot) + (1-\lambda) \cdot p_{2\theta}(\cdot)$, con $0 \leq \lambda \leq 1$, se verifica

$$\Delta e^{-H(p_\theta(\cdot))} \leq e^{-H(p_\theta(\cdot))} \cdot (1 - e^{\Delta H(p_\theta(\cdot))})$$

y donde

$$\Delta e^{-H(p_\theta(\cdot))} = e^{-H(p_\theta(\cdot))} - \lambda e^{-H(p_{1\theta}(\cdot))} - (1-\lambda) \cdot e^{-H(p_{2\theta}(\cdot))}$$

Demostración.- Al ser $H(p_\theta(\cdot))$ cóncava, por el Teorema 3.2.IV.5,

$$\lambda e^{(1-\lambda)(H(p_{2\theta}(\cdot)) - H(p_{1\theta}(\cdot)))} + (1-\lambda) \cdot e^{-\lambda(H(p_{2\theta}(\cdot)) - H(p_{1\theta}(\cdot)))}$$


Dado que

$$\begin{aligned} \Delta e^{-H(p_\theta(\cdot))} &= \Delta e^{-H(p_\theta(\cdot))} \cdot \frac{e^{\lambda H(p_{1\theta}(\cdot)) + (1-\lambda) \cdot H(p_{2\theta}(\cdot))}}{e^{\lambda H(p_{1\theta}(\cdot)) + (1-\lambda) \cdot H(p_{2\theta}(\cdot))}} = \\ &= \frac{e^{-\Delta H(p_\theta(\cdot))} - \lambda e^{(1-\lambda)(H(p_{2\theta}(\cdot)) - H(p_{1\theta}(\cdot)))} - (1-\lambda)e^{-\lambda(H(p_{2\theta}(\cdot)) - H(p_{1\theta}(\cdot)))}}{e^{\lambda H(p_{1\theta}(\cdot)) + (1-\lambda) \cdot H(p_{2\theta}(\cdot))}} \end{aligned}$$

Y aplicando el resultado del teorema 3.2.IV.5 :

$$\Delta e^{-H(p_\theta(\cdot))} \leq \frac{e^{-\Delta H(p_\theta(\cdot))} - 1}{e^{H(p_\theta(\cdot))} - \Delta H(p_\theta(\cdot))} = e^{-H(p_\theta(\cdot))} (1 - e^{\Delta H(p_\theta(\cdot))})$$

(c.q.d.)

Siguiendo idéntica técnica de demostración se obtiene la siguiente cota superior para el incremento de la exponencial $\exp(-E_x H(p_\theta(\cdot)))$:

TEOREMA 3.2.IV.8

Siendo $p_\theta(\cdot) = \lambda p_{1\theta}(\cdot) + (1-\lambda) \cdot p_{2\theta}(\cdot)$, con $0 \leq \lambda \leq 1$, se verifica

$$\Delta e^{-E_x H(p_\theta(\cdot/x))} \leq e^{-E_x H(p_\theta(\cdot/x))} (1 - e^{\Delta E_x H(p_\theta(\cdot/x))})$$

donde

$$\begin{aligned} \Delta e^{-E_x H(p_\theta(\cdot/x))} &= e^{-E_x H(p_\theta(\cdot/x))} - \\ &- \lambda e^{-E_x H(p_{1\theta}(\cdot/x))} - (1-\lambda) \cdot e^{-E_x H(p_{2\theta}(\cdot/x))} \end{aligned}$$

Las correspondientes cotas inferiores vienen dadas por los próximos dos resultados:

TEOREMA 3.2.IV.9

Con la notación del teorema 3.2.IV.7, y siendo

$$\begin{aligned} \Delta H(p_\theta(\cdot)) \cdot e^{-H(p_\theta(\cdot))} &= H(p_\theta(\cdot)) \cdot e^{-H(p_\theta(\cdot))} - \lambda H(p_{1\theta}(\cdot)) \cdot e^{-H(p_{1\theta}(\cdot))} - \\ &- (1-\lambda) \cdot H(p_{2\theta}(\cdot)) \cdot e^{-H(p_{2\theta}(\cdot))}, \text{ se verifica que} \\ \Delta e^{-H(p_\theta(\cdot))} &\geq H(p_\theta(\cdot)) \cdot e^{-H(p_\theta(\cdot))} - \Delta H(p_\theta(\cdot)) \cdot e^{-H(p_\theta(\cdot))} + \end{aligned}$$

$$+ H(p_{\theta}(\cdot)) \cdot (\Delta H(p_{\theta}(\cdot)) - H(p_{\theta}(\cdot)))$$

Demostración.-

$$\begin{aligned} \Delta e^{-H(p_{\theta}(\cdot))} &= \lambda (e^{-H(p_{\theta}(\cdot))} - e^{-H(p_{1\theta}(\cdot))}) + (1-\lambda) \cdot (e^{-H(p_{\theta}(\cdot))} - e^{-H(p_{2\theta}(\cdot))}) = \\ &= \lambda e^{-H(p_{1\theta}(\cdot))} (e^{-H(p_{\theta}(\cdot))} + H(p_{1\theta}(\cdot)) - 1) + (1-\lambda) \cdot e^{-H(p_{2\theta}(\cdot))} \cdot (\\ &(e^{-H(p_{\theta}(\cdot))} + H(p_{2\theta}(\cdot)) - 1) \geq \lambda e^{-H(p_{1\theta}(\cdot))} \cdot H(p_{1\theta}(\cdot)) - \lambda e^{-H(p_{1\theta}(\cdot))} \cdot H(p_{\theta}(\cdot)) + \\ &+ (1-\lambda) \cdot e^{-H(p_{2\theta}(\cdot))} \cdot H(p_{2\theta}(\cdot)) - (1-\lambda) \cdot e^{-H(p_{2\theta}(\cdot))} \cdot H(p_{\theta}(\cdot)) \end{aligned}$$

expresión de la que se deduce inmediatamente la tesis del teorema.

(c.q.d.)

Y siguiendo idéntica técnica de demostración se tiene el corries pondiente resultado:

TEOREMA 3.2.IV.10

Con la notación del teorema 3.2.IV.8, y siendo

$$\begin{aligned} \Delta E_x H(p_{\theta}(\cdot/x)) \cdot e^{-E_x H(p_{\theta}(\cdot/x))} &= E_x H(p_{\theta}(\cdot/x)) \cdot e^{-E_x H(p_{\theta}(\cdot/x))} - \\ &- \lambda E_x H(p_{1\theta}(\cdot/x)) \cdot e^{-E_x H(p_{1\theta}(\cdot/x))} - (1-\lambda) E_x H(p_{2\theta}(\cdot/x)) \cdot e^{-E_x H(p_{2\theta}(\cdot/x))}, \end{aligned}$$

se verifica que

$$\begin{aligned} \Delta e^{-E_x H(p_{\theta}(\cdot/x))} &\geq E_x H(p_{\theta}(\cdot/x)) \cdot e^{-E_x H(p_{\theta}(\cdot/x))} - \\ &- \Delta E_x H(p_{\theta}(\cdot/x)) \cdot e^{-E_x H(p_{\theta}(\cdot/x))} + E_x H(p_{\theta}(\cdot/x)) (\Delta E_x H(p_{\theta}(\cdot/x)) - \\ &- E_x H(p_{\theta}(\cdot/x))) \end{aligned}$$

3.3. LA UTILIDAD ESPERADA ANTE TRANSFORMACIONES DE LA CANTIDAD DE INTERES

Quando el investigador está más interesado en conocer la cantidad aleatoria φ más bien que el parámetro θ , relacionados ambos por la trans-

formación $\varphi = \varphi(\vartheta)$, se define la correspondiente información útil esperada (Bernardo, 1978) como ya se indicó en la sección 1.4. De ahí que las utilidades esperadas admitan idénticas expresiones formales que las usadas en la sección 3.1 pero definidas en función de la información esperada útil y no en función de la información esperada total, resultando, por tanto, las expresiones siguientes:

$$\begin{aligned}
 U_{1\varphi}(E) &= g_{1\varphi} \cdot I^{\varphi}\{E, p_{\vartheta}(\cdot)\} \\
 U_{2\varphi}(E) &= g_{2\varphi} \cdot (1 - e^{-I^{\varphi}\{E, p_{\vartheta}(\cdot)\}}) \\
 U_{3\varphi}(E) &= g_{3\varphi} \cdot e^{-H(p_{\varphi}(\cdot))} \cdot (e^{I^{\varphi}\{E, p_{\vartheta}(\cdot)\}} - 1)
 \end{aligned}$$

y en donde $p_{\varphi}(\cdot) = T_{\varphi}(p_{\vartheta}(\cdot))$, siendo T_{φ} la correspondiente transformación que liga a las cantidades desconocidas ϑ y φ . La interpretación de las nuevas constantes $g_{i\varphi}$ es idéntica a la realizada para las g_i definidas en las primitivas funciones $U_i(E)$ ($i=1,2,3$), pero referidas evidentemente a la nueva situación introducida por la consideración del nuevo parámetro φ . Y así, $g_{1\varphi}$ representa la utilidad esperada de la información proporcionada por una unidad informante sobre el valor de φ ; $g_{2\varphi}$ representa la utilidad esperada final de investigar con información completa sobre el valor de φ ; y $g_{3\varphi}$ será la utilidad esperada de aquella información esperada sobre φ que toma el valor $\log(1 + \exp(H(p_{\varphi}(\cdot))))$. Estas nuevas constantes no tienen por qué coincidir con las primitivas g_i , toda vez que, por ejemplo, la transformación $\varphi = \varphi(\vartheta)$ puede haber restringido el campo de variabilidad del nuevo parámetro φ a un cierto subconjunto del espacio paramétrico primitivo \mathcal{W} , de forma que una unidad de información puede proporcionar más cantidad de ella sobre φ que sobre ϑ . Lo cierto, evidentemente, es que ante transformaciones biyectivas de la cantidad de interés, las constantes g_i y g_i coinciden ($i=1,2$), pues la información esperada es invariante ante tales transformaciones. No es el caso, como era de esperar, de $g_{3\varphi}$ puesto que está definida en función de la entropía y ésta no es invariante ante las transformaciones biyectivas de la cantidad de interés. Queda como problema abierto la comparación entre las utilidades esperadas definidas bien como funciones de la información esperada total, bien como funciones de la información esperada útil, y dentro del contexto del re

sultado encontrado por Bernardo (1978) para ambas informaciones esperadas. Sin embargo, si puede afirmarse que, teniendo en cuenta el resultado aludido que indica que la información esperada útil no es mayor que la información esperada total, se verifican las desigualdades:

$$\frac{U_{i\psi}(E)}{g_{i\psi}} \leq \frac{U_i(E)}{g_i} \quad (i=1,2)$$

$$\frac{U_{3\psi}(E) \cdot e^{H(p_\psi(\cdot))}}{g_{3\psi}} < \frac{U_3(E) \cdot e^{H(p_\theta(\cdot))}}{g_3}$$

Dado que la información esperada útil sigue idéntico comportamiento que la información esperada total como función del tamaño muestral y como funcional del experimento $E = (E_1, E_2)$, seguirán verificándose para las utilidades esperadas $U_{i\psi}(E)$ ($i=1,2,3$) análogos resultados que los expresados en los grupos de teoremas 3.I, 3.II y 3.III. En cambio, consideradas las $U_{i\psi}(E)$ como funcionales de la densidad inicial $p_\theta(\cdot)$, no puede asegurarse su concavidad puesto que tampoco la información esperada útil la verifica. Para ésta, Bernardo (1978) demuestra que fijo $p_\theta(\theta/\psi)$, la información $I^\psi\{E, p_\theta(\cdot)\}$ es una funcional cóncava de $p_\psi(\cdot)$. Basándonos en este resultado y siguiendo la misma técnica de demostración que la utilizada para demostrar los teoremas 3.2.IV.4 y 3.2.IV.6, pueden demostrarse los siguientes dos resultados correspondientes:

TEOREMA 3.3.1

Para un experimento $E = \{X, \mathcal{D}, p_x(\cdot/\theta)\}$, función $\psi = \psi(\theta)$ y fijo $p_\theta(\cdot/\psi)$, $U_{1\psi}(E)$ es una funcional cóncava de $p_\psi(\cdot)$.

TEOREMA 3.3.2

Para un experimento $E = \{X, \mathcal{D}, p_x(\cdot/\theta)\}$, función $\psi = \psi(\theta)$ y fijo $p_\theta(\cdot/\psi)$, $U_{2\psi}(E)$ es una funcional cóncava de $p_\psi(\cdot)$.

CAPITULO 4

DETERMINACION DEL TAMAÑO MUESTRAL OPTIMO

Este capítulo es consecuencia de todo el desarrollo anterior. En efecto, en él presentaremos una normativa sencilla y general para determinar un tamaño muestral por maximización de la expresión del valor esperado de un experimento para, posteriormente, aplicarla a los dos modelos concretos que nos acompañan a lo largo de toda la tesis.

4.1 TAMAÑO MUESTRAL OPTIMO

La determinación de un tamaño muestral es uno de los problemas de diseño más sencillo y de aplicación más frecuente. En todos estos problemas, y con gran generalidad, el objetivo último que se persigue al realizar un experimento muestral consiste en obtener información a través de la muestra sobre una cierta cantidad de interés θ , de forma que esta información adicional al conocimiento que sobre θ dispone a priori el investigador es representativa de la utilidad que el proceso de muestreo reporta. Esta dualidad información-utilidad ya ha sido convenientemente expresada en los capítulos anteriores por lo que para acabar de especificar el problema debemos contrastar dicha utilidad con el costo que lógicamente lleva aparejado el proceso de muestreo. Resulta evidente que, enfrentados de esta forma utilidad y costo, el tamaño muestral óptimo deberá representar una solución de compromiso entre la maximización de la utilidad y la minimización del costo. Precisamente, la diferencia entre la utilidad y el costo esperados recibe el nombre de valor esperado del experimento, de forma que el tamaño muestral óptimo debe encontrarse por maximización de dicho valor. Se establece como hipótesis fundamental la de la aditividad de la utilidad y el costo esperados. Ambos vendrán expresados, pues, en idénticas unidades.

Dado que desde el comienzo el problema queda enfocado desde el prisma Bayesiano debe suponerse que el investigador que debe decidir sobre el tamaño muestral a considerar es, a su vez, capaz de describir sus opiniones iniciales sobre θ a través de la densidad inicial $p_{\theta}(\cdot)$. Además, también deberá ser capaz de expresar mediante una función conocida la utilidad que le reporta realizar la experimentación y obtener una muestra determinada.

Denotando por

$$u(E(n), x(n), \theta) = u_t(x(n), \theta) - c_s(E(n))$$

a una utilidad aditiva compuesta por una utilidad terminal (u_t) y un costo muestral (c_s) - y que según se acaba de indicar vienen expresados en idénticas unidades - es conocido (Raiffa & Schlaifer, 1961) que la mejor acción (esto es, el mejor tamaño muestral en nuestro caso) debe obtenerse de la expresión

$$\max_n \iint u(E(n), x(n), \theta) \cdot p_{\theta}(\theta/x(n)) \cdot d\theta \cdot dx(n)$$

donde $p_{\theta}(\cdot/x(n))$ es la correspondiente densidad posterior Bayes.

En el capítulo tercero se han definido tres funciones para representar la utilidad esperada y que pueden emplearse según las características de cada problema concreto, al tiempo que se justificó oportunamente su dependencia como funciones de la información esperada. Los datos de los que se disponga a priori o la facilidad de acceder a los mismos serán los determinantes para la posible elección de una u otra función de la utilidad esperada. Concretamente, la facilidad en la estimación de las constantes g_1 y la disponibilidad de actuación bien respecto a informaciones globales, bien en crecimientos de ésta en términos absolutos o relativos, conducirá a considerar las funciones $U_1(E(n))$, $U_3(E(n))$ ó $U_2(E(n))$, respectivamente. Otro de los factores a tener en cuenta para la elección de una u otra de las funciones de la utilidad esperada consiste en considerar las exigencias sobre dichas utilidades. Y así, y según se demostró en el capítulo anterior, las distintas funciones $U_1(E(n))$ verifican una serie de propiedades íntimamente relacionadas con las satisfechas por la información esperada. Por ejemplo, la exigencia, en casos extremos, de utilidades no decrecientes marginalmente

conforme aumenta el tamaño muestral, ya hasta alcanzar un tamaño determinado, ya durante todo el proceso, conduciría a considerar como más idónea a la función $U_3(E(n))$ dado que, para ella, no se respeta el decrecimiento de las utilidades marginales conforme aumenta el tamaño muestral. Así pues, en definitiva la elección entre las distintas funciones de la utilidad esperada proporcionada por el experimento $E = \{X, \Theta, p_x(\cdot/\theta)\}$ dependerá del particular problema que se pretenda resolver y de la facilidad de estimación de las constantes g_i , así como de la estructura inicial de los datos del problema. Conjugando todas estas exigencias es como puede decidirse sobre la conveniencia en la utilización de una función $U_i(E(n))$ determinada. La comparación entre las condiciones del problema con las propiedades, carácter y definiciones concretas de cada una de las funciones de la utilidad esperada proporcionará la solución para su elección.

Una vez decididas por parte del investigador qué densidad inicial y qué función de utilidad son las que representan convenientemente el problema deberá terminar considerando el costo esperado asociado con la realización del experimento. Ya se ha indicado que en el caso de que exista una función $c(E(n), x(n), \theta)$ que mida el costo de realizar el experimento $E(n)$ cuando el resultado muestral obtenido haya sido el $x(n)$, siendo θ el verdadero valor de la cantidad desconocida o parámetro, el correspondiente costo esperado vendrá dado por la integral

$$C(E(n)) = \int \int c(E(n), x(n), \theta) \cdot p_{x(n)}(x(n)/\theta) \cdot p_{\theta}(\theta) \cdot d\theta \cdot dx(n) \quad .$$

Sin embargo, en la práctica no suelen utilizarse funciones de la forma anterior pues no suele disponerse de ellas. Por ello, a fin de determinar el costo esperado asociado a una experimentación, suele procederse a diversas estimaciones. En efecto, primeramente se considera la existencia de un costo fijo, c_0 , que existe de forma automática una vez que se decide realizar una investigación. Dicho costo fijo engloba a una serie de conceptos que, o no tienen relación con el tamaño muestral o, por su complejidad, se consideran sin tener en cuenta a éste (propaganda, infraestructura, ordenador, amortización de equipos, etc.). La correcta determinación de este costo c_0 no tiene incidencia sobre la decisión final del tamaño muestral a considerar como tendremos ocasión de comprobar y en ausencia de limitaciones introducidas por presupuestos fijos dados. Por todo ello, la estimación de c_0 puede reali-

zarse a posteriori, una vez decidido el tamaño muestral concreto a considerar. Su valor solamente es necesario a efectos de la determinación del costo cuantitativo de la investigación.

En otras muchas ocasiones puede estimarse el costo derivado de la extracción de una unidad muestral a través de una cierta función de su tamaño y que denotaremos por $c(n)$. Este es el caso, por ejemplo, de expresiones del costo medidas en términos de ganancia o pérdida de la precisión de las estimaciones. En este tipo de funciones $c(n)$ aparezcan una serie de parámetros cuyo valor es necesario estimar o conocer previamente. Por otra parte, también cada caso concreto puede exigir una u otra expresión para $c(n)$. Por regla general suele apreciarse una disminución en el costo marginal conforme aumenta el tamaño muestral, esto es, existe una especie de "prima" a la obtención de muestras grandes. En tal caso, las funciones que mejor describirían esta situación serían cóncavas como funciones de n , tomando expresiones de la forma \sqrt{n} , $\log n$, etc. Si, por el contrario, los costos marginales aumentan con el tamaño muestral, serían las funciones convexas de n las que mejor interpretarían esta situación. Sería, por tanto, lógico considerar funciones como las n^2 , n^3 , e^n , etc. Sin embargo, en otras muchas aplicaciones, una vez estimado el costo fijo c_0 , suele calcularse el costo estimado de una unidad de observación que denotaremos por c_1 , suponiéndose que el costo total aumenta siempre en proporción c_1 , esto es, que la descripción del costo esperado puede realizarse a través de una función lineal de pendiente c_1 . Debido a que la función de costo lineal es la más utilizada en la práctica, será la que empleemos para desarrollar la aplicación concreta a los dos modelos propuestos. Así pues y salvo indicación al contrario, en el posterior desarrollo se considerará como función de costo esperado, la:

$$C(E(n)) = c_0 + c_1 \cdot n$$

donde c_1 (que es una cantidad positiva) representa el costo unitario por observación muestral. Su estimación suele resultar fácil en la práctica y comprende conceptos como el costo asociado a los desplazamientos necesarios para obtener la respuesta (en el caso de encuestas por entrevista), gastos de correo en el envío de los cuestionarios, importe unitario de éstos, etc. Finalmente, otras situaciones pueden exigir la consideración de costos escalonados. Cuando el costo unitario de la unidad muestreada toma distintos va-

lores según unos determinados intervalos para el tamaño muestral, deben estimarse unos costos unitarios c_i tales que representen el costo unitario asociado a la obtención de unidades muestrales cuyo orden de extracción esté comprendido dentro de intervalos de la forma $[n_{i-1}, n_i]$ ($i=1,2,\dots$). Sin embargo, es de destacar nuevamente que las mismas condiciones y datos disponibles de cada caso concreto serán las que determinan la utilización de una u otra función de costo esperado concreta.

Una vez que el investigador dispone de las correspondientes expresiones de la utilidad $U(E(n))$ y el costo $C(E(n))$ esperados para la realización del experimento $E(n)$ y tal como quedó definido en la sección 1.5, considerará el valor esperado del experimento $E(n)$ como diferencia entre ambos, esto es,

$$W(E(n)) = U(E(n)) - C(E(n))$$

expresión que evidentemente es función del tamaño muestral. Ello permite maximizarla respecto n . Hay que destacar que si el investigador es coherente la realización del experimento $E(n)$ sólo tendrá lugar con $W(E(n))$ positivo. Según lo indicado siempre se supondrá que las unidades en las que están medidas la utilidad y el costo esperados son las mismas, para así asegurar su comparabilidad. Por otra parte, en la expresión de $U(E(n))$ deberá estimarse previamente la constante g_i que aparezca en la definición de la respectiva función de la utilidad esperada considerada. No obstante, existe una importante situación en la que no es necesario especificar la constante g_i ni la función de costo esperado $C(E(n))$ para realizar comparabilidades entre experimentos: es aquella situación en la que se puede suponer que todos los posibles experimentos variables tienen idéntico costo.

Obtenido el valor esperado $W(E(n))$ del experimento como función de n se está ya en condiciones de obtener el tamaño muestral óptimo. Para ello, y sin pérdida de generalidad, supondremos un comportamiento continuo para las expresiones en las que intervenga n , como si éste fuese una variable real con valores en R^+ , y a fin de asegurar la existencia de la inmediata condición, que no es sino la condición satisfecha por el tamaño muestral óptimo. Efectivamente, conocido $W(E(n))$, interesará aquél valor de n que lo maximiza. Así pues, el tamaño muestral óptimo deberá elegirse de la condición de maximización en los valores positivos de $W(E(n))$, esto es, a

través de la ecuación

$$\frac{d W(E(n))}{dn} = \frac{d U(E(n))}{dn} - \frac{d C(E(n))}{dn} = 0 \quad (1)$$

Como ya se indicó en su momento, si el investigador es coherente solamente elegirá como óptimos valores de n que verifiquen la positividad del valor esperado $W(E(n))$. Así pues, resulta evidente que en aquellas situaciones en las que

$$U(E(n)) \leq C(E(n)) \quad \text{para todo } n \geq 1 \quad (2)$$

la decisión óptima resultaría la de decidir sobre el valor del parámetro sin realizar experimentación alguna. El caso opuesto sería aquel en el que conforme aumenta el tamaño muestral la diferencia entre la utilidad y el costo esperados es cada vez mayor, esto es, el caso en el que el valor esperado del experimento $E(n)$ sea una función estrictamente creciente del tamaño muestral. En tal caso es también evidente que el comportamiento óptimo a seguir por el investigador - comportamiento que tiene que ser el opuesto al comentado para la situación anterior expresada por (2) - sería el de obtener la muestra de mayor tamaño que permitiera un presupuesto dado. En este sentido es interesante destacar que en las aplicaciones prácticas siempre deben manejarse las limitaciones presupuestarias.

Sin considerar ya en lo que sigue estos dos casos límite, para cuya solución no se requiere obtener la ecuación (1) y para los que el comportamiento óptimo del investigador ya ha quedado definido, el tamaño muestral óptimo deberá definirse de entre los enteros no negativos más próximos a la solución de la ecuación (1). Precisamente, la no-negatividad de dicho tamaño muestral óptimo queda garantizada al no verificarse la condición (2) por hipótesis. Por otra parte, la resolución de la ecuación (1) puede resultar compleja en determinados casos, debiendo recurrir entonces a la obtención de la solución mediante procedimientos de aproximación. Sin embargo, lo importante es la existencia de una condición satisfecha por el tamaño muestral óptimo.

Bajo ciertas condiciones puede garantizarse la existencia y la unicidad de la solución de (1). En efecto, admitida la hipótesis de la no consideración en lo que sigue de los dos casos extremos a los que hemos aludido ahora mismo, es conocido ya que la utilidad esperada es una función de

positiva y creciente del tamaño muestral. Además, tanto $U_1(E(n))$ como $U_2(E(n))$ son funciones cóncavas de n , esto es, mantienen idéntico signo en su segunda derivada para cualquier valor de n . Por otra parte, es evidente - evidencia que será considerada como hipótesis de trabajo - que el costo esperado $C(E(n))$ debe ser también una función positiva y creciente del tamaño muestral.

LEMA 4.1

Si consideradas como funciones del tamaño muestral la utilidad y el costo esperados son diferenciables y sus segundas derivadas (supuestamente existentes) mantienen su signo constante (sea el que sea) a lo largo de todo el campo de definición de n , queda garantizada la existencia de solución positiva para (1).

Demostración.- Excluido el caso expresado por la desigualdad $[a,b]$ (2), supongamos la existencia de un cierto intervalo cerrado $[a,b]$ en el que $U(E(n)) \geq C(E(n))$, con desigualdad estricta en uno de sus puntos (el límite inferior de este intervalo puede tomar el valor 1). En este caso, el teorema de Bolzano-Weierstrass asegura la existencia de, al menos, un máximo en dicho intervalo que, por otra parte, es único. En efecto, la existencia de un segundo intervalo $[c,d]$ en el que la utilidad esperada no fuese menor que el costo esperado implicaría el cambio de signo de la segunda derivada de una o de las dos funciones componentes de $W(E(n))$ en un punto interior del intervalo $]c,d[$. También esta razón, junto a que tanto $U(E(n))$ como $C(E(n))$ son crecientes por hipótesis aceptada, es la que impide la existencia de máximos aislados dentro del intervalo $[a,b]$. Así pues, queda garantizada la existencia de solución para (1) que puede venir expresada bien a través de un único punto $x_0 \in [a,b]$, ó a través de un subintervalo de máximos $[a_1, b_1] \subset [a,b]$ tal que:

$$\begin{aligned}
 W(E(x)) &= W(E(y)) && \text{para todo } x, y \in [a_1, b_1], \text{ y} \\
 W(E(x)) &> W(E(z)) && \text{para todo } x \in [a_1, b_1], \text{ para todo } y \\
 &&& y \in [a, b] - [a_1, b_1]
 \end{aligned}$$

En todo caso, pues, existe solución para (1).

Cuando el conjunto de valores de n en los que $U(E(n)) \geq C(E(n))$ (existiendo al menos un punto en el que la desigualdad es es-

tricta) sea el intervalo infinito $[a, +\infty[$ - nuevamente el límite inferior puede tomar el valor 1 - el signo de la segunda derivada de las funciones utilidad y costo esperados debe coincidir pues al ser crecientes ambas funciones el caso contrario implicaría un nuevo punto de corte entre ambas funciones y el intervalo dejaría de ser infinito. En estas circunstancias, excluidos los caso límite expresados por (2) y por experimentos con valor esperado estrictamente creciente a partir de un cierto tamaño muestral, solamente son dos los casos posibles que, en definitiva, se reducen al caso anterior de intervalo de valores de utilidad esperada mayor que el costo esperado finito: ó bien $C(E(n))$ tiende a confundirse asintóticamente con $U(E(n))$ en el infinito; ó existe un n_0 a partir del cual el valor esperado $W(E(n))$ permanece constante. En el primero de los casos basta considerar un n_1 lo suficientemente grande para el que se verifique que $W(E(n_1-1)) > W(E(n_1))$, y el problema queda resuelto considerando que hay que determinar un máximo en el intervalo cerrado $[a, n_1]$. Análogamente, para el segundo de los casos basta tener en cuenta el máximo en el intervalo $[a, n_0]$.

(c.q.d.)

Notemos que las hipótesis exigidas a la utilidad y costo esperados en el anterior teorema no son excesivamente exigentes en cuanto que es lógico pensar en unos comportamientos homogéneos para $U(E(n))$ y $C(E(n))$ a lo largo de todo el proceso de muestreo, en cuanto al aumento o disminución de las utilidades y los costos marginales se refiere. En este sentido, las utilidades esperadas $U_1(E(n))$ y $U_2(E(n))$ cumplen ya con este requisito sin necesidad de hipótesis previa. La diferenciabilidad exigida al valor esperado $W(E(n))$ es necesaria para que la expresión (1) tenga sentido y pueda así calcularse. Respecto a la continuidad implicada es satisfecha, por ejemplo, por las distintas funciones $U_i(E(n))$ ($i=1,2,3$) obtenidas para los dos modelos propuestos en el capítulo segundo al tiempo que dicha condición también es verificada por las funciones de costo empleadas en la práctica totalidad de los casos.

Obtenida la solución óptima de (1) puede ya hablarse del tamaño muestral óptimo según la siguiente definición:

DEFINICION. El tamaño muestral óptimo, n_{opt} , para un experimento $E(n)$ realizado con objeto de adquirir conocimiento sobre el parámetro θ es

un entero no negativo obtenido de las siguientes condiciones:

- 1ª n_{opt} maximiza al valor esperado de $E(n)$, esto es, a $W(E(n))$.
- 2ª a) Cuando existe una única solución x de (1), n_{opt} es $\xi(x)$ ó $\xi(x) + 1$, donde ξ es la función "parte entera de".
- b) Si existe un intervalo cerrado $[a_1, b_1]$ de óptimos para (1) que contienen al menos un entero, el menor de ellos es la solución óptima para el tamaño muestral del experimento. Si, por el contrario, no existe ningún entero en dicho intervalo, n_{opt} será $\xi(a_1)$ ó $\xi(a_1) + 1$, según cuál de ellos verifique la condición 1ª.

Destaquemos que se ha definido el tamaño muestral óptimo como aquel entero no negativo que verifica una cierta serie de condiciones, aún cuando el lema 4.1 asegura que dicho tamaño es positivo. Sin embargo, hemos preferido hacerlo así para englobar aquella situación especial expresada por (2), para la que el tamaño óptimo es 0.

La solución óptima que maximiza a la utilidad esperada $U_i(E(n))$ ($i=1,2,3$) es, independientemente de la función de utilidad esperada particular considerada, la que también optimiza la información $I^\theta\{E(n), p_\theta(\cdot)\}$. Este resultado era necesario para asegurar la coherencia en el tratamiento de la dualidad información-utilidad esperadas. En efecto, basta considerar que la solución de las ecuaciones:

$$\frac{d U_1(E(n))}{dn} = g_1 \cdot \frac{d I^\theta\{E(n), p_\theta(\cdot)\}}{dn} = 0$$

$$\frac{d U_2(E(n))}{dn} = g_2 \cdot e^{-I^\theta\{E(n), p_\theta(\cdot)\}} \cdot \frac{d I^\theta\{E(n), p_\theta(\cdot)\}}{dn} = 0$$

$$\frac{d U_3(E(n))}{dn} = g_3 \cdot e^{-E_{x(n)} H(p_\theta(\cdot/x(n)))} \cdot \frac{d I^\theta\{E(n), p_\theta(\cdot)\}}{dn} = 0$$

son también solución de la ecuación

$$\frac{d I^\theta\{E(n), p_\theta(\cdot)\}}{dn} = 0$$

y viceversa. De ahí que la condición (1) sea efectivamente una solución de compromiso entre la maximización de la información proporcionada por una muestra y su costo asociado.

En el caso de que el valor esperado $W(E(n))$ sea una función estrictamente cóncava del tamaño muestral no solamente se asegura la existencia de solución para (1) sino también su unicidad, simplificándose notablemente la definición de tamaño muestral óptimo dada con anterioridad.

TEOREMA 4.1.1

Bajo las hipótesis del Lema 4.1 y siendo $W(E(n))$ una función estrictamente cóncava, existe una única solución positiva a la condición (1)

Demostración.— Una función diferenciable y estrictamente cóncava tiene a lo sumo un punto crítico (Fleming, 1969). El teorema se sigue inmediatamente del Lema 4.1.

(c.q.d.)

Notemos que la hipótesis adicional del teorema se verifica automáticamente cuando $U(E(n))$ es estrictamente cóncava y $C(E(n))$ convexa.

Cuando el problema venga expresado en términos de crecimientos marginales la expresión (1) carece de operatividad. Evidentemente que en estos casos será interesante utilizar un criterio de comparabilidad entre los incrementos de utilidad y de costo experimentados al pasar de la unidad n -ésima a la $(n+1)$ -ésima. Recordemos que se admite como cierto el hecho empírico de que tanto la utilidad como el costo esperados son funciones crecientes del tamaño muestral, circunstancia ésta que ya se demostró para las distintas funciones $U_i(E(n))$ ($i=1,2,3$) definidas en la sección 3.1. En estas situaciones en las que se dispone de información sobre los crecimientos marginales de la utilidad y el costo esperados parece lógico, pues, estudiar dónde se produce el cambio de signo en la expresión

$$\Delta W(E(n)) = \Delta U(E(n)) - \Delta C(E(n)) \quad (3)$$

como alternativa al procedimiento expresado en la ecuación (1). Notemos que

$$\begin{aligned} \Delta W(E(n)) &= U(E(n-1+\Delta n)) - U(E(n-1)) - C(E(n-1+\Delta n)) + C(E(n-1)) = \\ &= U(E(n)) - U(E(n-1)) - C(E(n)) + C(E(n-1)) \end{aligned}$$

donde $\Delta n = 1$. Claramente se aprecia que el proceso de muestreo no debería interrumpirse (en ausencia de limitaciones presupuestarias) siempre que $\Delta W(E(n)) > 0$; y, por el contrario, la extracción de una nueva unidad muestral carecería de sentido cuando $\Delta W(E(n)) < 0$. Así pues, deberá tomarse como criterio para la elección del tamaño muestral óptimo el de escoger aquel entero no negativo más próximo, en la zona donde $\Delta W(E(n)) \geq 0$, al cambio de signo de dicho incremento $\Delta W(E(n))$. La práctica coincidencia en los ejemplos de aplicación de ambos criterios para la determinación del tamaño muestral óptimo queda asegurada teóricamente sin más que considerar que el criterio expresado por el estudio de la expresión (3) no es sino trasladar al lenguaje de incrementos la condición expresada por la ecuación (1).

De todo lo anterior se deduce que ante un problema expresado en términos de los incrementos marginales de la utilidad y el costo esperados, el tamaño muestral óptimo admitirá la siguiente formulación:

DEFINICION. El tamaño muestral óptimo n_{opt} para un experimento $E(n)$ realizado con objeto de adquirir conocimiento sobre el parámetro ϑ , es el entero no negativo situado en la zona de valores de n en la que el incremento (3) es no negativo y más próximo al cambio de signo de dicho incremento.

Notemos que la no consideración de los casos expresados por la condición (2) implica que el cambio de signo deberá estudiarse para aquellos valores de n en los que $U(E(n)) \geq C(E(n))$. Por otra parte, si $W(E(n))$ es estrictamente creciente, es evidente que no existe cambio de signo para el incremento $\Delta W(E(n))$. Como en este caso dicho incremento es positivo, el comportamiento óptimo del investigador coincidirá, según este segundo criterio, con el expresado anteriormente.

Bajo las misma hipótesis del Lema 4.1 los tamaños muestrales obtenidos de ambas definiciones coinciden. Este es el sentido del siguiente

TEOREMA 4.1.2

Bajo las hipótesis del Lema 4.1, las dos definiciones dadas para el tamaño muestral óptimo coinciden.

Demostración.- Sea n_{opt} el tamaño muestral óptimo obtenido según la definición segunda. Por tanto, se produce el cambio de signo del incremento $\Delta W(E(n))$ en un punto del intervalo $[n_{opt}, n_{opt} + 1[$. Si para n_{opt} resulta ser $\Delta W(E(n_{opt})) > 0$, consideraremos el intervalo cerrado $[n_{opt}, n_{opt} + 1]$. Aplicando en él el teorema de Bolzano para funciones continuas, existe un punto x_0 interior al intervalo $]n_{opt}, n_{opt} + 1[$ en el que $\Delta W(E(x_0)) = 0$. En definitiva, puede afirmarse que en un entorno de x_0 se verifica que

$$U(E(x_0 + \Delta x)) - U(E(x_0)) - C(E(x_0 + \Delta x)) + C(E(x_0)) = O(\Delta x)$$

esto es, en dicho entorno

$$\frac{W(E(x_0 + \Delta x)) - W(E(x_0))}{\Delta x} = 0$$

lo que implica, en definitiva, que en x_0 la derivada de $W(E(n))$ es nula y, por tanto, que x_0 es solución de (1). Por otra parte, $\phi(x_0) = n_{opt}$ y dado que, por hipótesis, los incrementos $\Delta W(E(n_{opt} - 1)) > 0$ y que $\Delta W(E(n_{opt})) < 0$, se deduce que $W(E(\phi(x_0))) > W(E(\phi(x_0) - 1))$ y que $W(E(\phi(x_0))) > W(E(\phi(x_0) + 1))$, de donde resulta finalmente que n_{opt} es también el tamaño muestral óptimo en el sentido de la primera definición. Si para n_{opt} resulta ser $\Delta W(E(n_{opt})) = 0$, la demostración anterior sigue siendo válida sin más que considerar a $[n_{opt} - 1, n_{opt} + 1]$ como el intervalo en el que se aplica el Teorema de Bolzano, con un x_0 que trivialmente es el mismo n_{opt} .

Inversamente, sea ahora un n_{opt} el tamaño muestral óptimo obtenido según la primera definición. Existe entonces al menos un x_0 en el intervalo $]n_{opt} - 1, n_{opt} + 1[$ que es solución de (1), verificándose además que $W(E(n_{opt})) > W(E(n_{opt} - 1))$ y que $W(E(n_{opt})) \geq W(E(n_{opt} + 1))$. Por tanto, existe un entorno de x_0 en el que, con la exactitud que se desee, se verifica que $\Delta W(E(x)) = 0$ en cualquier punto de él. Como además se verifican las relaciones:

$$W(E(n_{opt} + 1)) - W(E(n_{opt})) \leq 0$$

$$W(E(n_{opt})) - W(E(n_{opt} - 1)) > 0$$

resulta finalmente que el punto x_0 implica un cambio de signo del incremen

to $\Delta W(E(n))$. Y dado que para n_{opt} , $\Delta W(E(n_{opt}-1)) > 0$, aplicando la segunda definición resulta finalmente la coincidencia de ambos tamaños muestrales óptimos.

(c.q.d.)

Teniendo en cuenta los resultados obtenidos en el apartado II de la sección 3.2 y según que utilicemos una u otra de las funciones de la utilidad esperada $U_i(E(n))$ ($i=1,2,3$), las expresiones que toma $\Delta W(E(n))$ son, respectivamente, las siguientes:

$$\Delta W_1(E(n)) = U_1(E_n/E_{n-1}) - \Delta C(E(n))$$

$$\Delta W_2(E(n)) = U_2(E_n/E_{n-1}) \cdot \exp(-I^\theta \{E(n-1), p_\theta(\cdot)\}) - \Delta C(E(n))$$

$$\Delta W_3(E(n)) = U_3^*(E_n/E_{n-1}) - \Delta C(E(n))$$

donde $U_i(E_n/E_{n-1})$ ($i=1,2$) expresa la utilidad esperada obtenida al realizar la n -ésima observación, supuesto realizado el experimento $E(n-1)$ y por tanto conocido y estudiado el resultado $x(n-1)$ y la distribución a posteriori $p_\theta(\cdot/x(n-1))$; y $U_3^*(E_n/E_{n-1})$ tiene una interpretación parecida a la anterior sin más que considerar en la respectiva definición de $U_3(E)$ la sustitución de la entropía por la entropía esperada, toda vez que es la densidad a posteriori $p_\theta(\theta/x(n-1))$ la que hace el papel de densidad inicial conocida y tal como se indicó en su momento en la sección 3.2.

Cuando la expresión (3) sea siempre positiva, esto es, cuando para cualquier valor del tamaño muestral se obtenga un incremento mayor en la utilidad que en el costo por el hecho de obtener una nueva unidad muestral, no hay límites para el tamaño muestral óptimo a no ser la introducida por la exigencia de ajustarse a un presupuesto fijo, en cuyo caso deberá obtenerse la muestra de mayor tamaño que permitan las posibilidades económicas. Inversamente, si (3) es siempre negativa, esto es, si para cualquier valor del tamaño muestral y ante la expectativa de extracción de una nueva unidad de muestreo es de esperar que el costo sea mayor que la utilidad adicional reportada, entonces no deberá muestrearse y la decisión sobre la cantidad de interés deberá realizarse directamente a partir del conocimiento inicial expresado a través de la densidad a priori. Ambos casos, tal como se ha dicho, han quedado excluidos de las consideraciones generales y no han sido tenidos en cuenta

en los razonamientos de las demostraciones de los teoremas anteriores.

La utilización de uno u otro criterio de los expresados en las condiciones (1) y (3) dependerá de la especial naturaleza de cada problema y de cómo venga expresado. Tal como se ha indicado, la aplicación de (3) es de gran utilidad en aquellos casos en los que el problema venga expresado en términos de crecimientos marginales del valor esperado y del costo, mientras que la expresión (1) tiene como interpretación la propia de cualquier cálculo de máximos.

4.2 APLICACION

Desarrollemos, para los dos modelos propuestos en el capítulo segundo, la teoría expresada en la sección anterior. La aplicación al modelo normal es importante, toda vez que otras muchas situaciones con densidades distintas pueden reducirse a este modelo siempre que el tamaño muestral n se presuponga razonablemente grande. La función de costo que utilizaremos, salvo especificación en contra, será la lineal con costo fijo c_0 y costo unitario por observación c_1 .

4.2.1 APLICACION AL MODELO NORMAL

Según acabamos de indicar, el modelo normal es de gran importancia en cualquier desarrollo práctico en el que intervenga el valor esperado de un experimento como función de la información que proporciona. Ello es debido a que en muchas ocasiones sus resultados pueden extenderse a otros sistemas en los que las densidades de probabilidad son distintas. E, incluso, mediante aproximaciones debidamente justificadas, pueden utilizarse resultados del modelo normal para el cálculo de informaciones esperadas y expresiones relacionadas. De ahí que el tamaño muestral óptimo deducido bajo las hipótesis del modelo considerado cobra su gran valor sin más que considerarlo como indicativo en bastantes situaciones prácticas.

Bajo las hipótesis del modelo normal y suponiendo un costo esperado lineal como función del tamaño muestral, según utilicemos las distintas funciones de la utilidad esperada $U_i(E(n))$ ($i=1,2,3$), se tendrán las siguientes expresiones respectivas para el valor esperado del experimento $E(n)$:

$$W_1(E(n)) = \frac{1}{2}g_1 \cdot \log\left(1 + n \frac{\sigma_0^2}{\sigma^2}\right) - c_0 - c_1 \cdot n$$

$$W_2(E(n)) = g_2 \cdot \left(1 - \frac{1}{\sqrt{1 + n \frac{\sigma_0^2}{\sigma^2}}}\right) - c_0 - c_1 \cdot n$$

$$W_3(E(n)) = g_3 \cdot \frac{\sqrt{n \cdot \sigma_0^2 + \sigma^2} - \sigma}{\sigma \sigma_0 \sqrt{2\pi e}} - c_0 - c_1 \cdot n$$

Para estas tres expresiones del valor esperado del experimento $E(n)$ y tal como puede apreciarse en los resultados manifestados en el capítulo tercero, $\Delta U(E(n))$ tiende a cero conforme aumenta el tamaño muestral. Como c_1 es una cantidad positiva y $\Delta C(E(n)) = c_1$ queda garantizado que para el modelo normal los tres valores esperados del experimento, $W_i(E(n))$ ($i=1,2,3$), que se manejan a lo largo de toda la tesis, no son crecientes. Por otra parte, las tres funciones $W_i(E(n))$ ($i=1,2,3$) son en el modelo considerado estrictamente cóncavas como funciones del tamaño muestral. Por el teorema 4.1.1 existe para ellas un máximo único, independientemente de la definición utilizada en su obtención (teorema 4.1.2).

I) Empleando el criterio (1) para la maximización del valor esperado Bernardo (1975) encuentra su solución x_0 cuando el valor esperado del experimento viene expresado a través de $W_1(E(n))$. Dicha solución es:

$$x_0 = \frac{g_1}{2c_1} - \left(\frac{\sigma}{\sigma_0}\right)^2$$

de forma que el tamaño muestral óptimo resultará

$$n_{opt}^1 = \left\lfloor \left(\frac{g_1}{2c_1} - \frac{\sigma^2}{\sigma_0^2}\right) \right\rfloor \quad \text{ó} \quad n_{opt}^1 = \left\lfloor \left(\frac{g_1}{2c_1} - \frac{\sigma^2}{\sigma_0^2}\right) \right\rfloor + 1$$

dependiendo de cuál de esos dos valores maximice $W_1(E(n))$.

Cuando el problema venga expresado en términos de incrementos para la utilidad y el costo esperados, esto es, cuando se conozcan los incrementos

$$\Delta U_1(E(n)) = \frac{1}{2}g_1 \cdot \log \frac{\sigma^2 + n \cdot \sigma_0^2}{\sigma^2 + (n-1) \cdot \sigma_0^2} \quad y$$

$$\Delta C(E(n)) = c_1$$

el criterio (3) indica que deberá estudiarse el cambio de signo en la expresión

$$\begin{aligned} \Delta W_1(E(n)) &= \frac{1}{2}g_1 \cdot \log \frac{\sigma^2 + n \cdot \sigma_0^2}{\sigma^2 + (n-1) \cdot \sigma_0^2} - c_1 = \\ &= \frac{1}{2}g_1 \cdot \log \frac{\sigma^2 + n \cdot \sigma_0^2}{e^{(2c_1/g_1)} \cdot (\sigma^2 + (n-1) \cdot \sigma_0^2)} \end{aligned}$$

que se produce cuando

$$\frac{\sigma^2 + n \cdot \sigma_0^2}{e^{(2c_1/g_1)} \cdot (\sigma^2 + (n-1) \cdot \sigma_0^2)} = 1$$

ecuación que, suponiendo un comportamiento continuo para n , proporciona como solución la:

$$- \left(\frac{\sigma}{\sigma_0} \right)^2 + \frac{1}{1 - e^{-(2c_1/g_1)}}$$

Teniendo en cuenta que la investigación se centra sobre un máximo, verificándose que

$$\lim_{n \rightarrow \infty} \Delta U(E(n)) = 0 \quad y \quad \Delta C(E(n)) = c_1$$

resulta evidente que el tamaño muestral óptimo aplicando el criterio del estudio del cambio de signo de (3) es:

$$n_{opt}^2 = \frac{1}{1 - e^{-(2c_1/g_1)}} - \left(\frac{\sigma}{\sigma_0} \right)^2$$

Si bien ya es conocido, por el teorema 4.1.1 que los tamaños n_{opt}^1 y n_{opt}^2

son iguales, comprobémoslo para esta situación práctica.

LEMA 4.2.1

Para cualesquiera valores positivos de c_1 y g_1 , se verifica que

$$\frac{1}{2} < \frac{1}{1 - e^{-(2c_1/g_1)}} - \frac{g_1}{2c_1} < 1$$

Demostración.- Es conocido que para todo $x \neq 0$, $e^x > 1 + x$.

Así pues,

$$e^{-(2c_1/g_1)} > 1 - \frac{2c_1}{g_1} \longrightarrow 0 < \frac{1}{1 - e^{-(2c_1/g_1)}} - \frac{g_1}{2c_1}$$

Sea la función definida en $]0, +\infty[$

$$y = \frac{1}{1 - e^{-x}} - \frac{1}{x}$$

Se trata de una función positiva y continua para todo $x > 0$. Verifica las propiedades siguientes:

$$1. \lim_{x \rightarrow 0^+} y = \lim_{x \rightarrow 0^+} \frac{x - 1 + e^{-x}}{x(1 - e^{-x})} = \frac{1}{2} \quad \text{esto es, } y(0^+) = \frac{1}{2}$$

$$2. \lim_{x \rightarrow \infty} (y/x) = 0 \quad \text{y} \quad \lim_{x \rightarrow \infty} y = 1$$

siendo la recta $Y = 1$ una asíntota.

$$3. y' = \frac{(1 - e^{-x} + x \cdot e^{-\frac{1}{2}x})(1 - e^{-x} - x \cdot e^{-\frac{1}{2}x})}{x^2(1 - e^{-x})^2} > 0$$

pues $1 - e^{-x} + x \cdot e^{-\frac{1}{2}x} > 0$ para todo $x > 0$, y

$z = 1 - e^{-x} - x \cdot e^{-\frac{1}{2}x} > 0$ para todo $x > 0$, pues se trata de una función continua y creciente en $]0, +\infty[$ y tal que

$$\lim_{x \rightarrow 0} z = 0 \quad \text{y} \quad \lim_{x \rightarrow \infty} z = 1.$$

Así pues, la función y es estrictamente creciente. De las propiedades anteriores se deduce, finalmente, el lema.

(c.q.d.)

Utilizando este resultado demostraremos, finalmente, la coincidencia entre n_{opt}^1 y n_{opt}^2 . Notemos que, trivialmente, se verifica que:

$$\frac{1}{2} < \frac{1}{1 - e^{-(2c_1/g_1)}} - \left(\frac{\sigma}{\sigma_0}\right)^2 - \left(\frac{g_1}{2c_1} - \left(\frac{\sigma}{\sigma_0}\right)^2\right) < 1$$

TEOREMA 4.2.1

En el modelo normal, expresado el valor del experimento $E(n)$ a través de $W_1(E(n))$, se verifica que $n_{opt}^1 = n_{opt}^2$.

Demostración.- Supongamos que $n_{opt}^1 = \int \left(\frac{g_1}{2c_1} - \left(\frac{\sigma}{\sigma_0}\right)^2 \right)$.

Así pues, $W(E(n_{opt}^1)) \geq W(E(n_{opt}^1 + 1))$ y $W(E(n_{opt}^1)) > W(E(n_{opt}^1 - 1))$.

De ahí que el cambio de signo se produzca más adelante de n_{opt}^1 . Por el lema 4.2.1 :

$$n_{opt}^1 \leq \frac{1}{1 - e^{-(2c_1/g_1)}} - \left(\frac{\sigma}{\sigma_0}\right)^2 < n_{opt}^1 + 1$$

esto es,

$$n_{opt}^1 \leq \int \left(\frac{g_1}{2c_1} - \left(\frac{\sigma}{\sigma_0}\right)^2 \right) < \int \left(\frac{1}{1 - e^{-(2c_1/g_1)}} - \left(\frac{\sigma}{\sigma_0}\right)^2 \right) < n_{opt}^1 + 1$$

y, por ello,

$$n_{opt}^1 = n_{opt}^2$$

Supongamos ahora que $n_{opt}^1 = \int \left(\frac{g_1}{2c_1} - \left(\frac{\sigma}{\sigma_0}\right)^2 \right) + 1$. Así

pues, $W(E(n_{opt}^1)) > W(E(n_{opt}^1 - 1))$ y $W(E(n_{opt}^1)) \geq W(E(n_{opt}^1 + 1))$. Por

tanto, el cambio de signo se produce más adelante de n_{opt}^1 . Por el lema 4.2.1 :

$$\int \left(\frac{g_1}{2c_1} - \left(\frac{\sigma}{\sigma_0}\right)^2 \right) + 1 \leq \frac{1}{1 - e^{-(2c_1/g_1)}} - \left(\frac{\sigma}{\sigma_0}\right)^2 <$$

$$< \left\lceil \left(\frac{g_1}{2c_1} - \left(\frac{\sigma}{\sigma_0} \right)^2 \right) + 2 \right.$$

Y considerando la parte entera, resulta finalmente

$$n_{opt}^1 = \left\lceil \left(\frac{g_1}{2c_1} - \left(\frac{\sigma}{\sigma_0} \right)^2 \right) + 1 \right. = \left\lceil \left(\frac{1}{1 - e^{-(2c_1/g_1)}} - \left(\frac{\sigma}{\sigma_0} \right)^2 \right) \right. = n_{opt}^2$$

toda vez que n_{opt}^1 es, por definición, un número entero.

(c.q.d.)

II. Cuando en el modelo normal el valor esperado del experimento viene expresado a través de $W_2(E(n))$, la condición de optimización (1) conduce a la ecuación

$$\frac{g_2 \cdot \sigma_0^2}{2(\sigma^2 + n \cdot \sigma_0^2) \sqrt{1 + n \frac{\sigma_0^2}{\sigma^2}}} = c_1$$

de solución igual a

$$x_0 = \left(\frac{g_2 \sigma}{2c_1 \sigma_0} \right)^{2/3} - \left(\frac{\sigma}{\sigma_0} \right)^2 = \left(\frac{g_2}{2c_1} \right)^{2/3} \cdot \left(\frac{\sigma}{\sigma_0} \right)^{2/3} - \left(\frac{\sigma}{\sigma_0} \right)^2$$

de forma que el tamaño muestral óptimo resultará ser

$$n_{opt} = \left\lceil \left(\left(\frac{g_2 \sigma}{2c_1 \sigma_0} \right)^{2/3} - \left(\frac{\sigma}{\sigma_0} \right)^2 \right) \right. \delta$$

$$n_{opt} = \left\lceil \left(\left(\frac{g_2 \sigma}{2c_1 \sigma_0} \right)^{2/3} - \left(\frac{\sigma}{\sigma_0} \right)^2 \right) + 1 \right.$$

dependiendo de cuál de estos dos valores maximiza a $W_2(E(n))$.

III. Para la tercera de las expresiones del valor esperado del experimento $E(n)$, la condición

$$\frac{d W_3(E(n))}{dn} = 0$$

conduce a la ecuación:

$$\frac{g_3 \cdot \sigma_0}{2\sigma \sqrt{2\pi e} \sqrt{n_0 \sigma_0^2 + \sigma^2}} - c_1 = 0$$

de solución igual a

$$x_0 = \frac{1}{2\pi e \sigma^2} \cdot \left(\frac{g_3}{2c_1} \right)^2 - \left(\frac{\sigma}{\sigma_0} \right)^2$$

de forma que el tamaño muestral óptimo resultará ser

$$n_{opt} = \left\lfloor \left(\frac{1}{2\pi e \sigma^2} \left(\frac{g_3}{2c_1} \right)^2 - \left(\frac{\sigma}{\sigma_0} \right)^2 \right) \right\rfloor \quad \text{ó} \quad n_{opt} = \left\lfloor \left(\frac{1}{2\pi e \sigma^2} \left(\frac{g_3}{2c_1} \right)^2 - \left(\frac{\sigma}{\sigma_0} \right)^2 \right) \right\rfloor + 1$$

dependiendo de cuál de esos dos valores maximiza a $W_3(E(n))$.

Es de destacar que la estructura de las soluciones obtenidas es similar en los tres casos descritos, respondiendo a la forma general expresada por la fórmula

$$h \cdot F\left(\frac{g_i}{2c_1} \right) - \left(\frac{\sigma}{\sigma_0} \right)^2$$

donde h representa un factor de proporcionalidad y F una función positiva de la relación $g_i/2c_1$. En definitiva, de la anterior expresión general se deduce que en el modelo normal y utilizando expresiones para el valor esperado del experimento dadas por las funciones $W_i(E(n))$ ($i=1,2,3$) definidas con anterioridad, el tamaño muestral óptimo se obtiene comparando la relación $g_i/2c_1$ con la ganancia relativa de precisión al pasar de la densidad inicial a la posterior, ganancia expresada a través del cociente σ/σ_0 .

Conforme el proceso de muestreo sea más positivo - lo que se traduce en una disminución de σ frente a σ_0 - es evidente que valdrá más la pena, a igualdad de costo, obtener muestras de tamaño grande. El razonamiento es inverso, claramente, para mayores costos unitarios. Por otra parte, resulta también necesario suponer que

$$h \cdot F\left(\frac{g_i}{2c_1} \right) > \left(\frac{\sigma}{\sigma_0} \right)^2$$

a fin de que el tamaño muestral óptimo sea una cantidad positiva.

4.2.2 APLICACION A UNA POBLACION UNIFORME Y DENSIDAD INICIAL DE PARETO

Trabajemos ahora con una población uniforme en $(0, \theta)$. Teniendo en cuenta las distintas funciones de la utilidad esperada $U_i(E(n))$ ($i=1,2,3$) definidas en la sección 3.1 y suponiendo un costo lineal, se obtendrán sendas expresiones para el valor esperado del experimento $E(n)$ dadas por:

$$W_1(E(n)) = g_1 \cdot \log\left(1 + \frac{n}{k}\right) - (c_0 + c_1 \cdot n)$$

$$W_2(E(n)) = g_2 \cdot \frac{n}{n+k} - (c_0 + c_1 \cdot n)$$

$$W_3(E(n)) = g_3 \cdot \frac{n}{\vartheta_0 \cdot \exp\left(\frac{1+k}{k}\right)} - (c_0 + c_1 \cdot n)$$

obtenidas suponiendo que las opiniones iniciales del investigador vienen descritas mediante una distribución de Pareto, de parámetros ϑ_0 y k .

En este modelo y tal como puede apreciarse de las expresiones estudiadas en su momento en el capítulo tercero, para las dos primeras funciones $W_i(E(n))$ ($i=1,2$) el incremento de la utilidad esperada $\Delta U(E(n))$ tiende a cero conforme aumenta el tamaño muestral. Como c_1 es una cantidad positiva y $\Delta C(E(n)) = c_1$ queda garantizado el no crecimiento de $W_i(E(n))$ ($i=1,2$) en el modelo de población uniforme con densidad inicial de Pareto. Por otra parte, tanto $W_1(E(n))$ como $W_2(E(n))$ son estrictamente cóncavas como funciones del tamaño muestral. Por el teorema 4.1.1 existe para ellas un máximo único, independientemente de la definición utilizada en su obtención (teorema 4.1.2). En cambio, estos resultados no pueden garantizarse para la tercera de las expresiones de $W(E(n))$ ya que, al ser lineal como función de n , fallan las hipótesis de partida para la validez de los teoremas 4.1.1 y 4.1.2. Este caso está incluido en los casos que comentábamos en la sección 4.1 y de los que ya describíamos el comportamiento óptimo por parte del investigador.

I. Supongamos que se utiliza la primera de las expresiones para el valor esperado de $E(n)$, esto es, la $W_1(E(n))$. La condición (1) se traduce en la siguiente condición satisfecha por el tamaño muestral óptimo:

$$g_1 / (k+n) - c_1 = 0$$

de solución igual a

$$x_0 = \frac{g_1}{c_1} - k$$

de forma que el tamaño muestral óptimo resultará ser

$$n_{\text{opt}}^1 = \left\lfloor \left(\frac{g_1}{c_1} - k \right) \right\rfloor \quad \text{ó} \quad n_{\text{opt}}^1 = \left\lfloor \left(\frac{g_1}{c_1} - k \right) \right\rfloor + 1$$

dependiendo de cuál de esos dos valores maximiza $W_1(E(n))$.

Cuando la utilidad esperada viene expresada a través de una función directamente proporcional a la información, el tamaño muestral óptimo decrece conforme aumenta el costo unitario por unidad muestreada, tal como puede observarse del anterior resultado. Este hecho es completamente lógico y pone de manifiesto el carácter de equilibrio entre el costo y la utilidad que debe guardar el tamaño óptimo. Por otra parte, el tamaño muestral óptimo aumenta conforme es mayor la utilidad proporcionada por una unidad de información. Notemos que la proporción g_1/c_1 representa la relación entre la utilidad de la información proporcionada por una unidad muestral y su costo asociado. Dicha relación se compara, mediante sustracción, con el parámetro k de la densidad inicial.

Si se supone un especial problema en el que la estimación de g_1 como la utilidad de la información proporcionada por una unidad respondiente es fácil y en el que tanto la información esperada como el costo esperado vengan expresados en función de los respectivos incrementos al pasar de una unidad de muestreo a la siguiente, deberá utilizarse el criterio (3) para la obtención del tamaño muestral óptimo. Bajo las hipótesis del modelo concreto en el que nos movemos, los incrementos a utilizar en la expresión (3) admitirán las expresiones siguientes:

$$\Delta C(E(n)) = c_1 > 0$$

$$\Delta U_1(E(n)) = g_1 \cdot \log \frac{k+n}{k+n-1} > 0$$

Así pues y según se especifica en el criterio dado por la condición (3), deberá estudiarse el cambio de signo de la expresión

$$\Delta U_1(E(n)) - \Delta C(E(n))$$

que en este caso resulta ser

$$W_1(E(n)) = g_1 \cdot \log \frac{k+n}{k+n-1} - c_1 = g_1 \cdot \log \frac{k+n}{e^{(c_1/g_1)^{(k+n-1)}}}$$

donde el logaritmo anterior existe al ser positivas todas las cantidades que intervienen en él. El cambio de signo se producirá cuando

$$\frac{k+n}{e^{(c_1/g_1)^{(k+n-1)}}} = 1$$

donde se supone un comportamiento continuo para n . Resolviendo la ecuación anterior se obtendrá para n el valor

$$-k + \frac{1}{1 - e^{-(c_1/g_1)}}$$

Teniendo en cuenta que la investigación se centra sobre un máximo, verificándose que

$$\lim_{n \rightarrow \infty} \Delta U_1(E(n)) = 0 \quad \text{y} \quad \Delta C(E(n)) = c_1$$

resulta evidente que el tamaño muestral óptimo aplicando el criterio del estudio del cambio de signo de (3) es

$$n_{opt}^2 = \int \left(\frac{1}{1 - e^{-(c_1/g_1)}} - k \right)$$

Notemos que la estructura de las soluciones n_{opt}^1 y n_{opt}^2 son plenamente coincidentes con las respectivas soluciones del caso I del modelo normal.

Así pues, el lema 4.2.1 y el teorema 4.2.1 son de inmediata aplicación en estas circunstancias, permitiendo afirmar la igualdad entre n_{opt}^1 y n_{opt}^2 . Este mismo resultado estaba asegurado por el teorema 4.1.2. Resulta, pues, el siguiente

TEOREMA 4.2.2

En el modelo de población uniforme y densidad inicial de Pareto, expresado el valor del experimento a través de $W_1(E(n))$, se verifica que

$$n_{opt}^1 = n_{opt}^2$$

II. Para la segunda de las expresiones del valor esperado del experimento $E(n)$, la condición

$$\frac{d W_2(E(n))}{dn} = 0$$

conduce a la expresión

$$d \left(\frac{g_2 \cdot n}{(k+n)} - c_0 - c_1 \cdot n \right) / dn = 0$$

que proporciona la ecuación

$$\frac{g_2 \cdot k}{(k+n)^2} = c_1$$

de solución igual a

$$x_0 = (k)^{\frac{1}{2}} \cdot \left(\sqrt{\frac{g_2}{c_1}} - k \right)$$

de forma que el tamaño muestral óptimo resultará ser

$$n_{opt}^1 = \mathcal{L}(\sqrt{k}) \cdot \left(\sqrt{\frac{g_2}{c_1}} - k \right) \quad \text{ó} \quad n_{opt}^1 = \mathcal{L}(\sqrt{k}) \cdot \left(\sqrt{\frac{g_2}{c_1}} - k \right) + 1$$

dependiendo de cuál de esos dos valores maximice $W_2(E(n))$. La existencia de las raíces anteriores queda garantizada por el hecho de que las constantes g_2 , c_1 y k son positivas.

Notemos que, nuevamente, el tamaño muestral óptimo n_2 se obtiene al comparar la relación g_2/c_1 con el parámetro k , de forma que, a semejanza de lo ocurrido al utilizar la expresión $W_1(E(n))$, el tamaño muestral crece, en este caso, conforme aumenta g_2 y disminuye conforme aumenta el costo unitario c_1 .

Cuando los datos de un problema concreto obliguen a emplear el segundo criterio para la obtención del tamaño muestral óptimo y bajo las hipótesis particulares de este caso, resultarían los siguientes incrementos:

$$\Delta C(E(n)) = c_1$$

$$\Delta U_2(E(n)) = \frac{g_2 \cdot k}{(k+n)^2 - (k+n)}$$

Según queda especificado en el criterio dado por la condición (3) deberá estudiarse el cambio de signo de la expresión

$$\Delta u_2(E(n)) - \Delta c(E(n))$$

esto es, el incremento del valor esperado del experimento,

$$\Delta w_2(E(n)) = \frac{g_2 \cdot k}{(k+n)^2 - (k+n)} - c_1$$

o, lo que es equivalente, estudiar el cambio de signo en

$$\frac{g_2 \cdot k}{c_1} - (k+n)^2 + (k+n)$$

Llamando a $K+n = t$, la expresión anterior se reduce a un trinomio cuadrado que, igualado a cero, proporciona dos raíces. En el intervalo determinado por ellas la anterior expresión es positiva. Así pues, la condición de óptimo resultará positiva entre los valores de t comprendidos entre

$$\frac{1}{2} \left(1 - \sqrt{1 + \frac{4g_2 k}{c_1}} \right) \quad \text{y} \quad \frac{1}{2} \left(1 + \sqrt{1 + \frac{4g_2 k}{c_1}} \right)$$

esto es, en los valores de n (suponiéndole un comportamiento continuo) comprendidos entre

$$\frac{1}{2} \left(1 - \sqrt{1 + \frac{4g_2 k}{c_1}} \right) - k \quad \text{y} \quad \frac{1}{2} \left(1 + \sqrt{1 + \frac{4g_2 k}{c_1}} \right) - k$$

Las raíces cuadradas existen al ser positivas todas las constantes bajo el signo radical. Por esta misma razón el extremo izquierdo del anterior intervalo es negativo. Teniendo en cuenta que el tamaño muestral óptimo es no negativo y que la investigación se centra sobre un máximo, verificándose que

$$\lim_{n \rightarrow \infty} \Delta u_2(E(n)) = 0 \quad \text{y} \quad \Delta c(E(n)) = c_1$$

resulta evidente que el tamaño muestral óptimo aplicando el segundo criterio es

$$n_{\text{opt}}^2 = \frac{1}{2} \left(1 + \sqrt{1 + \frac{4g_2 k}{c_1}} \right) - k$$

Si bien por el teorema 4.1.2 es conocido que los tamaños muestrales n_{opt}^1 y n_{opt}^2 son iguales, vamos a comprobarlo para esta situación práctica. Para ello, demostremos previamente el siguiente resultado:

LEMA 4.2.2

Para cualesquiera valores positivos de c_1 , g_2 y k , se verifica que

$$\frac{1}{2} < \frac{1}{2} \left(1 + \sqrt{1 + \frac{4g_2 k}{c_1}} \right) - \sqrt{k} \sqrt{\frac{g_2}{c_1}} < 1$$

Demostración.- Llamando a $4g_2 k/c_1 = x$, el resultado se sigue inmediatamente de considerar la doble desigualdad estricta

$$\sqrt{x} < \sqrt{1+x} < 1 + \sqrt{x} \quad \text{para todo } x > 0$$

(c.q.d.)

A partir del lema 4.2.2 resulta trivialmente que:

$$0 < \frac{1}{2} \left(1 + \sqrt{1 + \frac{4g_2 k}{c_1}} \right) - k - \left(\sqrt{k} \sqrt{\frac{g_2}{c_1}} - k \right) < 1$$

desigualdad que permite demostrar el siguiente

TEOREMA 4.2.3

En el modelo de población uniforme y densidad inicial de Pareto, expresado el valor esperado del experimento $E(n)$ a través de $w_2(E(n))$, se verifica que $n_{opt}^1 = n_{opt}^2$.

Demostración.- Siguiendo idéntica técnica que la empleada para demostrar el teorema 4.2.1, sea

$$n_{opt}^1 = \left\lfloor \left(\sqrt{k} \sqrt{\frac{g_2}{c_1}} - k \right) \right\rfloor$$

Así pues, $w(E(n_{opt}^1)) \geq w(E(n_{opt}^1 + 1))$ y $w(E(n_{opt}^1)) > w(E(n_{opt}^1 - 1))$. Por

tanto, el cambio de signo se produce más adelante de n_{opt}^1 . Por el lema anterior:

$$n_{opt}^1 \leq \frac{1}{2} \left(\frac{1}{2} \left(1 + \sqrt{1 + \frac{4g_2 k}{c_1}} \right) - k \right) < n_{opt}^1 + 1$$

esto es,

$$n_{opt}^1 = \int \left(\frac{1}{2} \left(1 + \sqrt{1 + \frac{4g_2 k}{c_1}} \right) - k \right) < n_{opt}^1 + 1 .$$

En definitiva,

$$n_{opt}^1 = n_{opt}^2 .$$

Supongamos ahora que

$$n_{opt}^1 = \int \left(\sqrt{k} \sqrt{\frac{g_2}{c_1}} - k \right) + 1 .$$

Así pues, $W_2(n_{opt}^1) > W_2(E(n_{opt}^1 - 1))$ y $W_2(E(n_{opt}^1)) \geq W_2(E(n_{opt}^1 + 1))$.

De ahí que el cambio de signo se produzca más allá de n_{opt}^1 . Por el lema anterior,

$$\int \left(\sqrt{k} \sqrt{\frac{g_2}{c_1}} - k \right) + 1 \leq \frac{1}{2} \left(1 + \sqrt{1 + \frac{4g_2 k}{c_1}} \right) - k < \int \left(\sqrt{k} \sqrt{\frac{g_2}{c_1}} - k \right) + 2$$

y considerando las correspondientes partes enteras resulta, finalmente,

$$n_{opt}^1 = \int \left(\sqrt{k} \sqrt{\frac{g_2}{c_1}} - k \right) + 1 = \int \left(\frac{1}{2} \left(1 + \sqrt{1 + \frac{4g_2 k}{c_1}} \right) - k \right) = n_{opt}^2$$

toda vez que n_{opt}^1 es, por definición, un número entero.

(c.q.d.)

En estos dos primeros casos el tamaño muestral óptimo responde a una estructura general de la forma

$$f(k) \sqrt{\frac{g_i}{c_1}} - k \quad i=1,2$$

siendo f una función positiva del parámetro k . Por otra parte, se supone implícitamente que

$$f(k) \sqrt{\frac{g_i}{c_1}} > k \quad i=1,2$$

BIBLIOTECA
FACULTAD C. MATEMATICAS
VALENCIA

a fin de que el tamaño muestral óptimo sea positivo. Alcaide (1966) interpreta el parámetro k de la densidad de Pareto como un número superior a la uni

dad (para que tenga sentido el valor medio de la distribución) que corresponde a una distribución de la renta tanto más justa cuanto más se aleja de la unidad, entendiendo por distribución "justa" la que se consigue con una mayor igualdad de rentas. Pena (1965) considera otra posibilidad de justicia en el sentido de la "distribución que permite alcanzar niveles elevados de renta a porcentajes elevados de la población". Sea en un caso o en otro, de la consideración de la anterior expresión general del tamaño muestral óptimo deducido cuando el valor del experimento $E(n)$ viene expresado a través de las dos primeras funciones de la utilidad esperada, $U_i(E(n))$ ($i=1,2$), y con un costo lineal, se deduce que al realizar una investigación sobre una población que respecto a cierta característica desconocida sigue un comportamiento uniforme respecto a ella, el tamaño muestral óptimo de dicha investigación decrecerá conforme la renta esté mejor distribuida (suponemos que la característica desconocida es, precisamente, la renta). Este resultado es lógico sin más que considerar que en poblaciones con renta equidistribuida las unidades muestrales son, evidentemente, mucho más representativas, por lo que con pocas unidades muestrales se alcanzará un significativo nivel de información.

III. Si las circunstancias reales de un problema aconsejan el empleo de la función de la utilidad esperada $U_3(E(n))$, bien porque empíricamente sea aconsejable que no exista disminución en las utilidades marginales conforme crezca el tamaño muestral, bien porque la estimación de g_3 sea factible como utilidad de una especial y determinada cantidad de información tal como está establecido en su definición, la emplearemos en la expresión del valor esperado del experimento $E(n)$, de forma que se trabajará con la expresión correspondiente a $W_3(E(n))$.

En el ejemplo de aplicación, con una población uniforme en $(0, \theta)$ y una densidad inicial de Pareto definida sobre \mathcal{D} , la expresión $W_3(E(n))$ toma la forma

$$\frac{g_3 \cdot n}{\theta_0 \cdot e^{\frac{(1+k)}{k}}} - (c_0 + c_1 \cdot n)$$

Tal como puede apreciarse, dicha expresión es lineal respecto al tamaño muestral, por lo que el criterio (1) de maximización es claro: si el coeficiente de n es positivo, esto es, si

$$g_3 / \vartheta_0 \cdot \exp\left(\frac{1+k}{k}\right) > c_1$$

se gana constantemente conforme se obtienen unidades muestrales. Así pues, debe tomarse el mayor tamaño muestral que permita el presupuesto al que debe ajustarse la investigación. Por el contrario, si el coeficiente lineal es negativo, esto es, si la pendiente de la recta que expresa $W_3(E(n))$ en función de n es negativa, se perdería valor de forma constante ante la extracción de cada nueva unidad muestral. En este caso, el no muestreo sería la alternativa válida a seguir, de forma que la decisión final sobre la cantidad de interés ϑ debería realizarse sin más que considerar las opiniones iniciales.

Las dos situaciones anteriores corresponden a los casos límite que hemos excluido del tratamiento general. En efecto, la primera situación corresponde a aquella en la que el valor esperado del experimento es estrictamente creciente. En cambio, la segunda es la planteada por una utilidad que es siempre inferior al costo asociado al experimento. Ambas están implicadas por la doble linealidad - a la que ya se hacía referencia en la sección 4.1 - de la utilidad y el costo esperados. Sin embargo, cuando no se verifica la linealidad del costo el resultado es, como era de esperar, completamente distinto. En efecto, existen situaciones prácticas en las que aparecen costos marginales no necesariamente decrecientes e, incluso - aunque con mucho menos frecuencia - costos convexos como funciones de n . Considerando ahora una de estas últimas funciones veamos cuál sería el comportamiento del valor esperado $W_3(E(n))$ en el cálculo del tamaño muestral óptimo y para el modelo que seguimos aplicando.

Sea, pues,

$$U_3(E(n)) = \frac{g_3 \cdot n}{\vartheta_0 \cdot \exp\left(\frac{1+k}{k}\right)} \quad \text{y} \quad C(E(n)) = c_0 + c_1 \cdot n^2$$

siendo el valor esperado del experimento la diferencia de las anteriores dos funciones. Bajo estas hipótesis, $W_3(E(n))$ no es una función estrictamente creciente y, sin embargo, sí lo es estrictamente cóncava. Así pues, por el teorema 4.1.1 existe un máximo único para dicho valor esperado, independientemente de la definición utilizada en su obtención (teorema 4.1.2).

La maximización de $w_3(E(n))$ conduce a la ecuación:

$$\frac{g_3}{\vartheta_0 \cdot \exp\left(\frac{1+k}{k}\right)} - 2c_1 \cdot n = 0$$

de solución igual a

$$x_0 = g_3 / 2c_1 \cdot \vartheta_0 \cdot \exp\left(\frac{1+k}{k}\right)$$

de forma que el tamaño muestral óptimo resultará ser

$$n_{\text{opt}}^1 = \xi\left(g_3 / 2c_1 \cdot \vartheta_0 \cdot \exp\left(\frac{1+k}{k}\right)\right) \quad \text{o} \quad n_{\text{opt}}^1 = \xi\left(g_3 / 2c_1 \cdot \vartheta_0 \cdot \exp\left(\frac{1+k}{k}\right)\right) + 1$$

dependiendo de cuál de esos valores maximice $w_3(E(n))$. Notemos que nuevamente aparece el factor g_3/c_1 .

La aplicación del criterio expresado por la condición (3) conduce también a la misma solución. En efecto, teniendo en cuenta los incrementos

$$\Delta U_3(E(n)) = g_3 / \vartheta_0 \cdot \exp\left(\frac{1+k}{k}\right) \quad \text{y} \quad \Delta C(E(n)) = c_1 \cdot (2n - 1)$$

el incremento del valor esperado resultará ser:

$$\Delta w_3(E(n)) = g_3 / \vartheta_0 \cdot \exp\left(\frac{1+k}{k}\right) - c_1 \cdot (2n - 1)$$

de forma que al ser $\Delta w_3(E(n))$ una función lineal de pendiente negativa, el cambio de signo se producirá donde la recta anterior corte al eje de abscisas, esto es, en el punto supuestamente positivo

$$\frac{g_3}{2c_1 \cdot \vartheta_0 \cdot \exp\left(\frac{1+k}{k}\right)} + \frac{1}{2}$$

Teniendo en cuenta que a la derecha del punto de corte anterior el incremento $\Delta w_3(E(n))$ es positivo, es evidente que el tamaño muestral óptimo aplicando el criterio del cambio de signo de (3) es:

$$n_{\text{opt}}^2 = \xi\left(\frac{g_3}{2c_1 \cdot \vartheta_0 \cdot \exp\left(\frac{1+k}{k}\right)} + \frac{1}{2}\right)$$

Dado que la diferencia existente entre el punto x_0 con el punto en el que cambia el signo del incremento del valor esperado es $\frac{1}{2}$ (y, por tanto, menor

que la unidad), siguiendo idéntica técnica de demostración que la empleada en los teoremas 4.2.1 y 4.2.3 , resulta finalmente el siguiente y evidente teorema:

TEOREMA 4.2.4

En el modelo de población uniforme y densidad inicial de Pareto, expresado el valor de $W_3(E(n))$ a través de $U_3(E(n))$ y un costo dado por la expresión $c_0 + c_1 \cdot n^2$, se verifica que $n_{opt}^1 = n_{opt}^2$.

Notemos finalmente que los resultados de todas las aplicaciones anteriores y para ambos modelos son independientes del costo fijo c_0 , toda vez que, por su misma definición, dicho costo aparece inmediatamente se deci de realizar una experimentación e independientemente de cuál es el tamaño muestral a considerar.

CAPITULO 5

EXTENSION AL MUESTREO CON ESTRATOS

En este capítulo se generaliza para el muestreo estratificado los conceptos desarrollados con anterioridad para aplicarlos, nuevamente, a resolver el común problema de diseño consistente en determinar el tamaño muestral. Tanto desde un punto de vista teórico como, sobre todo, práctico, la gran mayoría de encuestas por muestreo probabilístico se realizan dividiendo la población en estratos lo más heterogéneos posible entre ellos y, en cambio, con unidades de muestreo últimas lo más parecidas dentro de cada estrato respecto la característica objeto de interés. En definitiva, y por definición de muestreo estratificado, la población se divide en subpoblaciones tan distintas como sea posible en relación con el fenómeno objeto de estudio, obteniéndose posteriormente muestras independientes en cada subpoblación. Esta selección de estratos homogéneos "dentro" y heterogéneos "entre" ellos queda debidamente justificada mediante el correspondiente análisis de la varianza, a fin de ganar precisión en los resultados y estimaciones muestrales.

De la importancia que en la actualidad tiene el muestreo estratificado en el diseño de encuestas por muestreo probabilístico baste citar, como ejemplos recogidos por Sánchez-Crespo (1973), algunas de las encuestas en las que existe una estratificación en un momento determinado de su diseño y que están realizadas por los más diversos organismos de varios países. Así, en la Encuesta Sanitaria mediante entrevista a los hogares de los Estados Unidos, las unidades primarias - que son o un condado o un grupo de ellos adyacentes, o un Area Metropolitana - están estratificadas en 372 estratos. Así mismo, en el diseño de la Encuesta Sanitaria mediante examen médico completo del mismo país, existen 42 estratos de aproximadamente igual número de habitantes. La Encuesta Sanitaria mediante entrevista realizada en Japón obtiene también una estratificación previa de los municipios según su población. En la Encuesta de Transporte de Mercancías por carretera, rea

lizada en Alemania en 1970, se estratificaron los vehículos de acuerdo con las principales clasificaciones de las tablas proyectadas. En la correspondiente encuesta española, realizada en 1969, se obtuvo una estratificación en consonancia a ciertas características, como la capacidad de carga y tipo de vehículo, la provincia de matriculación, etc. La Encuesta de Presupuestos Familiares realizada en España en 1980-81, así como la Encuesta continua de Población Activa, al estar integradas en el diseño de la Encuesta General de Población del Instituto Nacional de Estadística, disponen de los estratos empleados en el diseño de ésta última. Y así, las unidades primarias que son las secciones censales se han estratificado atendiendo a las siguientes características:

1º Tipo de municipio al que pertenecen las secciones censales. Dicho tipo está relacionado con la importancia demográfica del municipio dentro de su provincia o de la influencia que sobre el mismo ejerza la capital.

2º Proporción de población activa que pertenece a una serie de profesiones, agrupadas en tres grandes grupos que corresponden, respectivamente, a profesionales con nivel cultural superior al medio, al ambiente rural puro y al conjunto de obreros o trabajadores con más o menos especialización que, en general, no se distinguen por su nivel cultural.

Por último, la Annual Survey of Manufactures de los Estados Unidos y a fin de recabar informaciones industriales, realiza una estratificación según el número de empleados en los distintos establecimientos. Esta diversidad de encuestas con distintos objetivos y realizadas en diversos países que emplean en su diseño el muestreo estratificado nos indica claramente que, en la práctica, deben tenerse en cuenta estratos de unidades con mucha frecuencia.

Así pues, en este capítulo vamos a indicar cómo afecta la existencia de estratos al problema general desarrollado en los capítulos anteriores. Estableceremos unas hipótesis de trabajo comunes distinguiendo, después, distintos casos según sendas hipótesis concretas. Estas permitirán, precisamente, realizar un estudio de aplicación práctica para la obtención del tamaño muestral óptimo por optimización del valor esperado de un experimento.

5.1 OBJETIVOS Y NOTACIONES. HIPOTESIS DE TRABAJO COMUNES

En esta generalización a un proceso de muestreo estratificado el objetivo propuesto coincide plenamente con el expresado para los capítulos anteriores, si bien las circunstancias e hipótesis de trabajo van a modificarse ligeramente, teniendo en cuenta la especial obtención de la muestra con la consideración de la partición introducida en la población a consecuencia de los estratos. Por tanto, se pretende esbozar una teoría normativa, de fácil comprensión, sobre el problema de diseño frecuente y sencillo como es el de la determinación del tamaño muestral cuando el proceso de muestreo que permite recabar la información es el estratificado. Los argumentos expresados seguirán perteneciendo al esquema Bayesiano.

Nuevamente supondremos que el objetivo que se pretende al recabar información procedente de una muestra es el de mejorar nuestro conocimiento sobre el valor de una cierta cantidad θ que toma valores en un determinado conjunto Θ . La incógnita θ sigue recibiendo el nombre de "cantidad de interés". Tal como demuestra Bernardo (1979) y ya se ha indicado con anterioridad, el concepto de información esperada proporciona un poderoso instrumento para analizar el contenido de la información que, según apuntábamos, se obtiene a través de una muestra. Y a su vez, esta muestra no es, en realidad, sino el resultado de haber diseñado previamente un experimento realizado por el investigador sobre una población cuyas unidades dependen de θ de alguna manera especificada a través de una densidad de probabilidad.

Precisamente, el experimento a considerar supone una estratificación de la población objeto de estudio y la consiguiente obtención de una muestra estratificada. Lo importante de este proceso de estratificación reside en que su existencia afecta doblemente, por una parte a la población muestral y, por otra, al espacio paramétrico en el que toma valores la cantidad aleatoria θ . Efectivamente, y por su misma definición, el modelo matemático básico a considerar para el ulterior desarrollo consistirá del consabido espacio muestral total X de elementos x , indistintamente llamados muestra, dato o resultado experimental, dividido en L subespacios o estratos, X_j ($j=1,2,\dots,L$), y que constituyen una partición del espacio primitivo X , esto es,

$$X = \bigcup_{j=1}^L X_j \quad \text{y} \quad X_j \cap X_k = \emptyset \quad \text{para todo } j \neq k$$

A su vez, para cada uno de los anteriores subespacios el modelo a considerar va a ser idéntico - en cuanto que dispondrá de los mismos elementos - que el desarrollado en el capítulo primero. Y así, habrá que considerar la existencia de un espacio paramétrico para cada uno de ellos. En efecto, tal como acabamos de decir el proceso de estratificación afecta de forma muy íntima a la naturaleza del espacio paramétrico Θ . Efectivamente, si bien en un principio la cantidad de interés es el parámetro θ , al restringimos a una subpoblación de la población total X (y no olvidemos que, por construcción de los estratos, las subpoblaciones X_i ($i=1,2,\dots,L$) son escogidas de la forma más heterogénea posible para aumentar la precisión de la estimación, tal como queda recogido en el correspondiente análisis de la varianza) el parámetro θ pierde su condición de tal para transformarse en una nueva cantidad de interés, la θ_i , de clara interpretación: θ_i admite idéntica definición conceptual que el parámetro general θ pero referida al nuevo espacio muestral X_i . Por ejemplo, ante una estratificación regional en una encuesta de población activa en la que se estudia la proporción θ de parados, θ_i representaría dicha proporción referida a la región i -ésima. De esta definición, pues, se deduce una doble implicación. Por una parte, los parámetros o nuevas cantidades de interés θ_i son efectivamente parámetros distintos entre sí y respecto a la cantidad de interés θ referida a toda la población. Ello implica que no necesariamente los espacios paramétricos Θ_i y Θ en los que toman valores los parámetros θ_i y θ , respectivamente, son iguales sino que puede haber circunstancias en las que $\Theta_i \neq \Theta$, sin ningún criterio de posible inclusión. Por otra parte, el conocimiento exacto de todas las cantidades aleatorias θ_i debe permitir conocer, a su vez, al parámetro global θ . En efecto, por misma definición de θ_i , conocidas todas las cantidades aleatorias θ_i ($i=1,2,\dots,L$) y siendo conocido, evidentemente, por el investigador el conjunto de estratos $\{X_i\}_{i=1,2,\dots,L}$ en los que ha quedado particionado el espacio muestral global X , la determinación de θ queda establecida a partir de una relación funcional entre las distintas θ_i ($i=1,2,\dots,L$) y θ . Así pues, supondremos la existencia de una función

$$\theta = f(\theta_1, \theta_2, \dots, \theta_L)$$

que relaciona al parámetro global θ con los parámetros θ_i , conceptualmente idénticos a θ , pero de ámbito de aplicación restringido a su estrato respectivo.

Por regla general las características que son objeto de estudio en las encuestas e investigaciones estadísticas vienen representadas por un total, por un valor medio o por una proporción, que no es sino un valor medio para una característica que toma como posibles valores el 1 y el 0, esto es, en poblaciones $X = \{0,1\}$. Con gran generalidad son estas tres las características que se estudian en las encuestas por muestreo probabilístico. En cualquiera de estos tres casos la relación funcional que liga las cantidades θ_i con θ es lineal, por lo que supondremos, salvo especificación al contrario, que la relación f anterior puede venir expresada a través de un sumatorio de la forma

$$\theta = \sum_{i=1}^L h_i \cdot \theta_i$$

siendo las h_i ($i=1,2,\dots,L$) cantidades conocidas y positivas. En el caso de poblaciones finitas (que es evidentemente el caso de mayor investigación práctica) y denotando por N el número total de unidades muestrales para la población global X y por N_i el correspondiente número para el estrato X_i (se supondrá siempre que N es lo suficientemente grande como para asegurar que todos los tamaños parciales N_i también lo son), siendo

$$N = \sum_{i=1}^L N_i$$

se verifican las evidentes igualdades:

$$\theta = \sum_{i=1}^L \theta_i$$

cuando tanto θ como las θ_i representan totales, y

$$\theta = \sum_{i=1}^L w_i \cdot \theta_i$$

con $w_i = N_i/N$ ($i=1,2,\dots,L$), cuando θ y las θ_i representan medias y proporciones (Azorín, 1962; Cochran, 1974).

Resumiendo, la existencia de un proceso de estratificación condu

ce a la consideración de L nuevas cantidades de interés (tantas como estratos) que determinan completamente el parámetro ϑ primitivo. Así pues, se dispondrá de un vector-parámetro o vector de interés

$$\underline{\vartheta} = \{ \vartheta_1, \vartheta_2, \dots, \vartheta_L \}$$

como posible (ya insistiremos en ello más adelante) objeto de la atención del investigador. Con esta notación, la anterior relación funcional que liga ϑ con las ϑ_i podrá reescribirse como

$$\vartheta = f(\underline{\vartheta}) = f(\{ \vartheta_i \}_{i=1,2,\dots,L})$$

Una vez especificado el vector paramétrico $\underline{\vartheta}$ y tal como se ha indicado anteriormente, el modelo matemático básico a considerar en cada uno de los estratos X_i es idéntico al desarrollado en el capítulo primero, y donde el nuevo parámetro ϑ_i , componente i -ésima del vector $\underline{\vartheta}$, es la nueva cantidad de interés. Así pues, supondremos que el estrato i -ésimo está provisto de una apropiada σ -álgebra de conjuntos sobre los que existe una familia dada de medidas de probabilidad. Estas medidas están cada una de ellas identificada por los valores del parámetro ϑ_i , que toma valores en el espacio paramétrico \mathcal{D}_i .

La muestra total extraída a fin de recabar información sobre el vector de interés $\underline{\vartheta}$ o sobre la cantidad aleatoria de interés ϑ está constituida por el conjunto de submuestras extraídas de los respectivos estratos. Estas son aleatorias y, entre ellas, independientes. Es decir, el proceso de obtención de una unidad de una submuestra no afecta a la de la siguiente, así como el proceso de obtención de una submuestra no afecta tampoco a la obtención de la correspondiente a otro estrato. Esta doble independencia es la que permitirá, un poco más adelante, encontrar una densidad de probabilidad para la muestra total definida a partir de un doble productorio. El cociente resultante de dividir el número de unidades de la submuestra i -ésima, n_i , por el del estrato total (supuesto finito, pero suficientemente grande) N_i , recibe el nombre de fracción de muestreo para el estrato i -ésimo, esto es,

$$f_i = n_i/N_i \quad (i=1,2,\dots,L)$$

y que indica la intensidad con que se muestrea en cada estrato. Es claro que

cuando dichas fracciones sean iguales en todos los estratos, la muestra obtenida representa a la población en escala reducida, tratándose por ello de una especie de población en miniatura (Sánchez-Crespo, 1973). En este caso el muestreo estratificado recibe el nombre de proporcional.

De acuerdo con el párrafo anterior, denotemos por Z_i la muestra obtenida en el estrato i -ésimo. Dicha muestra no es sino el resultado de la realización repetida n_i veces (tantas como indica el tamaño parcial de la submuestra) del experimento $E = \{X_i, \Theta_i, p(x_i/\theta_i)\}$ consistente en la extracción de una unidad de muestreo del estrato i -ésimo, la cual se distribuye, para algún $\theta_i \in \Theta_i$ determinado, de acuerdo con la densidad de probabilidad $p(x_i/\theta_i)$ y con respecto a alguna medida dominante σ -finita definida sobre X_i . Seguimos suponiendo que si θ_{i1} y θ_{i2} son valores diferentes de Θ_i las distribuciones correspondientes de x_i , $p(x_i/\theta_{i1})$ y $p(x_i/\theta_{i2})$, difieren salvo un conjunto de medida dominante nula. Por tanto, la muestra del estrato i -ésimo consistirá de n_i observaciones, resultado de sendas realizaciones del experimento definido ahora mismo. Usaremos la notación

$$z_i = \{x_{i1}, x_{i2}, \dots, x_{in_i}\}$$

para expresar la muestra obtenida en el estrato i -ésimo. Además, por misma definición del experimento $E = \{X_i, \Theta_i, p(x_i/\theta_i)\}$ y al ser z_i una muestra aleatoria simple, resultará ser:

$$p(z_i/\theta_i) = \prod_{j=1}^{n_i} p(x_{ij}/\theta_i)$$

como la densidad de probabilidad que describe a la muestra z_i del estrato i -ésimo.

En un segundo estrato, el j -ésimo por ejemplo, consideraremos análogamente la repetición n_j veces del experimento $E = \{X_j, \Theta_j, p(x_j/\theta_j)\}$ que proporcionará como resultado la muestra z_j . Dado que se han escogido estratos lo más heterogéneos posible entre ellos como base a una ganancia en la precisión de las estimaciones puede ocurrir, evidentemente, que las densidades de probabilidad que describen los comportamientos de las cantidades aleatorias x_{ij} y x_{kl} pertenecientes a dos estratos diferentes, X_i y X_k respectivamente, sean, a su vez, diferentes, perteneciendo a dos diferentes

familias de densidades de probabilidad. Este hecho, sin embargo, no afecta para nada el desarrollo teórico posterior por lo que siempre consideraremos que las densidades $p(x_{ij}/\theta_i)$ ($i=1,2,\dots,L$) que intervienen en las definiciones de los experimentos realizados en los estratos respectivos pertenecen a una misma familia de densidades de probabilidad. Por otra parte, en muchas investigaciones prácticas en las que, además, la estratificación suele realizarse atendiendo, entre otros, a criterios políticos y de organización administrativa, la misma variación expresada a través de los parámetros distintos θ_i es suficiente, al poder tomar valores en conjuntos Θ_i no necesariamente iguales. Así pues, y salvo especificación en contra, todas las densidades $p(x_{ij}/\theta_i)$ ($i=1,2,\dots,L$) se describirán por una misma densidad de probabilidad, pero ciertamente con parámetro θ_i distinto.

De la realización de los distintos experimentos $E(n_i)$ en sendos estratos se dispondrá del conjunto de L submuestras

$$z_1 = \{x_{11}, x_{12}, \dots, x_{1n_1}\}$$

$$z_2 = \{x_{21}, x_{22}, \dots, x_{2n_2}\}$$

$$\dots$$

$$z_L = \{x_{L1}, x_{L2}, \dots, x_{Ln_L}\}$$

cuyo conjunto constituye la muestra global de la población estratificada X , muestra que denotaremos por

$$z = \{z_1, z_2, \dots, z_L\}$$

Por la misma obtención de la muestra z , y atendiendo al hecho de que las realizaciones de los experimentos $E(n_i)$ que proporcionan las submuestras z_i son independientes entre sí, la densidad de probabilidad que describe el comportamiento de la muestra total vendrá dada por el doble productorio

$$p(z/\underline{\theta}) = \prod_{k=1}^L p(z_k/\theta_k) = \prod_{k=1}^L \prod_{j=1}^{n_k} p(x_{kj}/\theta_k)$$

donde como puede apreciarse dicha densidad depende del vector cantidad de interés $\underline{\theta}$ a través de productos de densidades en las que intervienen sus componentes θ_i . Notemos que por la relación funcional que liga la primitiva cantidad de interés $\underline{\theta}$ con las θ_i se verifica que

$$p(z/\underline{\vartheta}, \theta) = p(z/\underline{\vartheta})$$

toda vez que el conocimiento de $\underline{\vartheta}$ implica que θ sea conocida también.

Así pues, la consideración de un proceso de estratificación conduce a pensar en la existencia de un vector paramétrico y de un conjunto de L experimentos, tantos como estratos, que dependen cada uno de una componente distinta de dicho vector paramétrico a través de una determinada densidad de probabilidad. Esto permite generalizar fácilmente el concepto de experimento al caso de existir una estratificación. Y así, denotando por

$$E^S(\underline{n}) = \left\{ \{X_i\}, \{\Theta_i\}, \{p(x_i/\theta_i)\} \right\}_{i=1,2,\dots,L}$$

representaremos el experimento consistente en la realización de L experimentos diferentes, $E(n_i)$, realizados en sendos estratos. De esta definición se deduce que el tamaño muestral debe considerarse transformado, a su vez, en un vector tamaño muestral, cuyas componentes son precisamente los tamaños muestrales de las respectivas submuestras. Denotaremos, pues, por $\underline{n} = (n_1, n_2, \dots, n_L)$ dicho vector tamaño muestral. Para un \underline{n} concreto, la suma de sus componentes expresa el tamaño muestral total del experimento $E^S(\underline{n})$ y lo denotaremos por n , esto es,

$$n = \sum_{i=1}^L n_i$$

La existencia de este tamaño muestral como vector condicionará la obtención del tamaño muestral óptimo, como tendremos ocasión de comprobar más adelante.

Hasta ahora todos los elementos considerados y el tratamiento dado están encuadrados dentro de la visión que la Estadística Clásica aporta a un proceso de estratificación. Para completarlo dentro de la línea de los capítulos anteriores, esto es, al considerar un argumento Bayesiano, hay que extender el modelo básico anterior y suponer que el conjunto en el que toma valores el parámetro $\underline{\vartheta}$

$$\underline{\vartheta} = \prod_{i=1}^L \vartheta_i$$

es soporte de una \mathcal{G} -álgebra, existiendo una medida de probabilidad conjunta sobre ella, que viene descrita a través de su función de densidad

$$p(\underline{\vartheta}) = p(\vartheta_1, \vartheta_2, \dots, \vartheta_L)$$

y definida con respecto a alguna medida dominante σ -finita sobre \mathcal{G} . Nuevamente la interpretación de esta medida de probabilidad es la ya conocida para una densidad inicial. Así pues, $p(\underline{\vartheta})$ describe las opiniones iniciales que el investigador, antes de realizar experimento alguno, tiene sobre el vector aleatorio $\underline{\vartheta}$. De ahí que sobre $p(\underline{\vartheta})$ puedan formularse todos los comentarios realizados sobre una densidad inicial, sin más que considerar que $\underline{\vartheta}$ se trata de una variable aleatoria definida en L dimensiones. Por otra parte, la misma densidad anterior expresa la posible dependencia existente entre las distintas componentes ϑ_i del vector $\underline{\vartheta}$. En efecto, por regla general y debido a cómo nacen las cantidades ϑ_i , es evidente que no deben ser independientes, influyendo los valores de una de ellas sobre los de las otras. La formulación de esta circunstancia conduce a la consideración de una densidad para $\underline{\vartheta}$ que no pueda expresarse como producto de densidades definidas para sus componentes ϑ_i por separado. Sin embargo, con posterioridad resolveremos un problema suponiendo dicha independencia que facilita extraordinariamente el trabajo al simplificar las expresiones a optimizar. En definitiva, el investigador deberá ser capaz de describir sus opiniones iniciales sobre $\underline{\vartheta}$ a través de una densidad a priori. Esta descripción puede realizarla bien directamente, bien describiendo sus opiniones sobre cada componente ϑ_i a través de una densidad inicial de la forma $p(\vartheta_i/\psi)$, donde ψ es un nuevo parámetro global que expresa la dependencia de las distintas cantidades ϑ_i entre sí, y que admite también una descripción inicial a través de la densidad $p(\psi)$. Entonces, la descripción del vector paramétrico $\underline{\vartheta}$ vendrá expresada a través de la densidad

$$p(\underline{\vartheta}) = \int \prod_{i=1}^L p(\vartheta_i/\psi) \cdot p(\psi) \cdot d\psi$$

Las opiniones del investigador sobre una ϑ_j determinada vendrán dadas por la marginal correspondiente

$$p(\vartheta_j) = \int_{\mathcal{G}'} p(\underline{\vartheta}) \cdot d\underline{\vartheta}' = \int_{\mathcal{G}_1} \int_{\mathcal{G}_2} \dots \int_{\mathcal{G}_{j-1}} \int_{\mathcal{G}_{j+1}} \dots \int_{\mathcal{G}_L} p(\vartheta_1, \dots, \vartheta_L) \cdot d\vartheta_1 \cdot \dots \cdot d\vartheta_{j-1} \cdot d\vartheta_{j+1} \cdot \dots \cdot d\vartheta_L$$

$$\omega' = \prod_{i \neq j}^L \omega_i$$

Una vez más y por simplicidad en la notación, la integración respecto la medida dominante definida sobre el conjunto Ω de posibles valores de una cantidad aleatoria ω se denotará simplemente por

$$\int \dots d\omega$$

Por otra parte, el rango de integración en las fórmulas utilizadas siempre estarán extendidas a todo el espacio, por lo que generalmente lo omitiremos. En caso contrario se especificarán los límites convenientemente.

La densidad "predictiva" del resultado muestral global z viene definida por

$$p(z) = \int p(z/\underline{\vartheta}) \cdot p(\underline{\vartheta}) \cdot d\underline{\vartheta}$$

Análogamente a la situación contenida en la sección 1.2, también el teorema de Bayes permite que, una vez realizado el experimento $E^S(\underline{n})$ y obtenido el resultado muestral $z = \{z_1, z_2, \dots, z_L\}$, pueda obtenerse la densidad de probabilidad que describa las opiniones posteriores que el investigador posea sobre $\underline{\vartheta}$ tras estudiar la información que sobre ella contiene la muestra z . Dicha densidad posterior viene dada por

$$p(\underline{\vartheta}/z) = p(z/\underline{\vartheta}) \cdot p(\underline{\vartheta}) / p(z)$$

esto es,

$$p(\underline{\vartheta}/z) = \frac{p(\vartheta_1, \vartheta_2, \dots, \vartheta_L) \cdot \prod_{i=1}^L p(z_i/\vartheta_i)}{\iint \dots \int p(\vartheta_1, \vartheta_2, \dots, \vartheta_L) \cdot \prod_{i=1}^L p(z_i/\vartheta_i) \cdot d\vartheta_1 \cdot d\vartheta_2 \dots d\vartheta_L}$$

Hasta ahora se han desarrollado una serie de elementos comunes a todo problema en el que, bajo una visión Bayesiana del mismo, interviene un proceso de muestreo estratificado. Como ya ha quedado indicado, la existencia de la estratificación no afecta solamente a la población X sobre la que se realiza, sino que también implica por una parte la existencia de tantos tamaños muestrales como estratos haya y cuya suma constituye el tamaño muestral total; por otra, la estratificación modifica íntimamente la estructura de la cantidad de interés que pasa a describirse por un vector parámetros

trico $\underline{\vartheta}$ cuyas componentes ϑ_i guardan con θ una relación funcional determinada y conocida de antemano. De ahí que al ser capaz el investigador de describir sus opiniones sobre $\underline{\vartheta}$ a través de la densidad $p(\underline{\vartheta})$, pueda también describir dichas opiniones sobre θ a través de la densidad inicial $p(\theta)$ deducida de la $p(\underline{\vartheta})$ a través de la relación funcional conocida $\theta = f(\underline{\vartheta})$. Considerado de esta forma, el problema introducido por la existencia de una estratificación es análogo, de momento, en cuanto que dispone de idénticos elementos que el resuelto en capítulos anteriores, sin más que considerar que tanto el tamaño muestral \underline{n} como la cantidad de interés $\underline{\vartheta}$ son vectores L -dimensionales. La diferencia más incisiva consiste en el modo de obtención de la muestra a partir de los estratos, toda vez que las expresiones (formales, claro está) de las densidades $p(z)$ y $p(\underline{\vartheta}/z)$ son idénticas a las ya utilizadas. Y así, la realización del experimento $E^s(\underline{n})$ supondrá la extracción de L submuestras independientes entre sí de cada uno de los estratos y de acuerdo con unas densidades de probabilidad que dependen de parámetros distintos. De ahí que al pasar de una unidad de muestreo a la siguiente deberá tenerse en cuenta a qué estrato pertenece la nueva unidad muestral para conocer qué densidad de probabilidad es la que modifica la expresión de la anterior densidad $p(\underline{\vartheta}/z)$. Sobre este tema volveremos inmediatamente.

5.2 HIPOTESIS ESPECIFICAS

Hasta ahora el tratamiento efectuado con el experimento $E^s(\underline{n})$ que considera un proceso de estratificación en la población X ha sido el general, sin más que considerar que tanto el tamaño muestral \underline{n} como el parámetro desconocido $\underline{\vartheta}$ son vectores en L dimensiones. Sin embargo, con todos los elementos anteriores no puede resolverse todavía el especial problema considerado como objetivo del presente trabajo: la determinación del tamaño muestral óptimo. Efectivamente, es necesario realizar previamente dos nuevas hipótesis para completar los elementos anteriores, sin olvidar que la optimización del tamaño muestral va a realizarse maximizando el valor esperado del experimento (como diferencia entre su utilidad y costo esperados) dentro del contexto de actuación realizado en el cuarto capítulo. La primera de estas hipótesis específicas se refiere al número de incógnitas que deben calcu

larse. En efecto, cuando la afijación de la estratificación (entendiendo por afijación la forma de distribuir el tamaño muestral global en los tamaños de las muestras extraídas en los estratos respectivos) es fijada de antemano por el investigador el vector tamaño muestral quedará reducido a una sola variable, tal como se comprobará ahora mismo. En caso contrario, se dispondrá evidentemente, de tantas variables como tamaños muestrales parciales haya en el problema, esto es, tantos como estratos. La segunda de las hipótesis específicas va a referirse a definir con exactitud el objetivo último que se pretende al realizar la experimentación expresada a través de $E^S(\underline{n})$. Ello es debido a que la aparición del vector cantidad de interés $\underline{\vartheta}$, formado por las L cantidades ϑ_i , exige que el investigador se defina sobre si su último deseo es recabar información sobre el parámetro primitivo ϑ exclusivamente, ó sobre el vector paramétrico $\underline{\vartheta}$ ó pretende optimizar la información esperada sobre cada una de las componentes ϑ_i por separado. Es evidente que la metodología a utilizar dependerá de la adopción de una u otra de estas hipótesis concretas. Notemos, sin embargo, que sean cuáles sean las hipótesis adicionales realizadas sobre ambas cuestiones, tanto el tamaño muestral global n , como la primitiva cantidad de interés ϑ , son funciones de sus respectivos n_i y ϑ_i a través de las relaciones

$$n = \sum_{i=1}^L n_i \quad \text{y} \quad \vartheta = f(\underline{\vartheta})$$

Pasemos, pues, a comentar las formulaciones concretas que admiten estos nuevos elementos de trabajo y que permiten enfocar y resolver de forma definitiva el problema propuesto.

a) Se ha definido como afijación de una estratificación al modo de distribuir las n unidades que constituyen la muestra total entre los diferentes estratos. A partir de ella, por tanto, pueden obtenerse las fracciones de muestreo que son las que indican la intensidad con que se muestrea en cada estrato. Supongamos que el investigador determina la afijación antes de realizar el experimento $E^S(\underline{n})$. Ello supone que para un determinado tamaño total n conoce de antemano los tamaños parciales n_i ($i=1,2,\dots,L$). En definitiva, trabajar con una afijación conocida supone que las distintas componentes del vector tamaño muestral quedan determinadas al conocer el tamaño muestral total y no al contrario. Así pues, establecida la hipótesis de ese conocimiento previo puede admitirse la existencia de L relaciones conq

cidas de la forma

$$n_i = f_i(n) \quad i=1,2,\dots,L \quad \text{siendo} \quad n = \sum_{i=1}^L f_i(n) \quad (1)$$

dependientes todas ellas del tamaño muestral total n y que permiten obtener inmediatamente el vector tamaño muestral $\underline{n} = (n_1, n_2, \dots, n_L)$ conocido n . Puede, pues, escribirse que

$$\underline{n} = (f_1(n), f_2(n), \dots, f_L(n))$$

igualdad que pone de manifiesto que el vector tamaño muestral queda reducido a la única variable tamaño muestral total n . Generalmente, las L relaciones anteriores suelen venir expresadas a través de un conjunto de L constantes o factores de proporcionalidad

$$\{f_i\}_{i=1,2,\dots,L}$$

que verifican las condiciones

$$n_i = f_i \cdot n \quad i=1,2,\dots,L \quad \text{siendo} \quad \sum_{i=1}^L f_i = 1$$

Por ejemplo, (Azorín, 1962) considerando poblaciones finitas pero lo suficientemente grandes como para poder aplicar las ideas anteriores, y para una afijación uniforme, las relaciones (1) tomarían la forma

$$n_i = n/L \quad i=1,2,\dots,L \quad ;$$

para una afijación proporcional,

$$n_i = n \cdot N_i / N \quad i=1,2,\dots,L \quad ;$$

y para la afijación óptima o de Neymann obtenida asignando un n_i a cada estrato proporcionalmente al producto $N_i \cdot S_i$, donde S_i denota la variabilidad del estrato i -ésimo, las relaciones (1) se expresarían como

$$n_i = n \cdot N_i \cdot S_i / \sum_{i=1}^L N_i \cdot S_i \quad i=1,2,\dots,L$$

Lo verdaderamente importante a destacar del hecho de que la afijación esté determinada previamente por el investigador es que el vector tamaño muestral \underline{n} queda perfectamente determinado conocido el tamaño muestral total n . Esto supone evidentemente que el problema planteado por la optimización de expre-

siones en las que aparezca el vector \underline{n} (como lo es la determinación del tamaño muestral óptimo) queda reducido, en realidad, a un problema de optimización de una expresión dependiente de una única variable. Notemos que bajo esta hipótesis la familia que define al experimento $E^S(\underline{n})$ resultaría

$$\left\{ E(f_i(n)) \right\}_{i=1,2,\dots,L}$$

de forma que bajo esta hipótesis dicho experimento podría expresarse con la notación más adecuada de $E^S(n)$, esto es, dependiente de la única variable: tamaño muestral total.

La hipótesis alternativa, el no conocimiento previo de la afijación de la estratificación, supone dejar en libertad a las componentes n_i del vector \underline{n} , de forma que el tamaño muestral total solamente quedará determinado al conocerse todos los valores n_i o tamaños de las submuestras respectivas. La introducción de un proceso de estratificación trae consigo una mayor riqueza, tanto en el tratamiento de cualquier problema que utiliza el muestreo estratificado, como en la posibilidad de estimación de cantidades y parámetros desconocidos. Esta riqueza, referida al tamaño muestral, puede traducirse en la obtención de afijaciones que respondan a unos determinados criterios de optimización y utilidad. En nuestro caso, la determinación de las variables n_i ($i=1,2,\dots,L$) se obtendrá por maximización de una expresión relacionada con el valor esperado del experimento $E^S(\underline{n})$. Así pues desde este punto de vista, el problema de determinación del vector \underline{n} consistirá en un problema de determinación de una afijación óptima. Pero, además, con la salvedad de que esta afijación es única en el sentido de que determina, a su vez, con total exactitud al tamaño muestral total n , al ser éste suma de los tamaños componentes n_i . Evidentemente, la maximización de expresiones en las que intervenga el valor esperado de $E^S(\underline{n})$ deberá realizarse teniendo en cuenta que se trata de expresiones dependientes de L variables distintas. Notemos que el objetivo propuesto de determinar un tamaño muestral óptimo o un conjunto de ellos, según los casos, es un problema que debe resolverse previamente a la realización práctica del experimento que permite la obtención de la muestra cuya investigación proporciona conocimiento sobre la cantidad de interés. De ahí que, en todo caso, se trabaje con valores esperados.

Destaquemos el distinto carácter que la hipótesis del conocimiento

to previo o no de la afijación de la estratificación confiere a un problema concreto. Al realizar el diseño de una encuesta donde la población se ha estratificado el investigador debe tomar dos decisiones que afectan, respectivamente, al tamaño muestral y a la afijación correspondiente. Si la decisión sobre la afijación es tomada previamente, una vez decidido el tamaño muestral total quedan inmediatamente fijados los tamaños parciales n_i correspondientes a las muestras extraídas de los respectivos estratos. En cambio, si la primera decisión es la efectuada sobre los valores de dichos tamaños parciales, su determinación supondrá, a la vez, la determinación del tamaño muestral total y, evidentemente, de la afijación a considerar que ya viene indicada por los particulares tamaños n_i escogidos. De ahí que en este segundo caso la afijación particular obtenida no pueda considerarse como una condición general válida para cualquier tamaño muestral total, en cuanto que éste viene también determinado como una suma de los tamaños parciales. Desde el punto de vista operativo la aceptación o no de la hipótesis considerada conduce, respectivamente, a que las expresiones con las que se trabaja dependan de una o varias variables. En aquél caso se tratará del tamaño muestral total, mientras que en éste serán L tamaños parciales n_i los considerados.

Problema aparte es el planteado al investigador cuando acepte la existencia de una afijación previa y deba decidir, por tanto, sobre la afijación concreta a considerar. De las afijaciones anteriormente mencionadas la proporcional es la más usual en casi todas las investigaciones, atendiendo fundamentalmente a que produce una moderada ganancia en la precisión de las estimaciones, a que obtiene una muestra autoponderada y a la simplificación de cálculo de los errores de muestreo (Sánchez-Crespo, 1973). En último término, tanto para la elección de la particular afijación cuando ésta se supone conocida a priori como cuando se acepta la hipótesis contraria, la elección en un sentido o en otro debe realizarse atendiendo a las condiciones que cada investigación concreta pueda exigir en la formulación de los elementos a considerar en dicho problema.

b) La segunda de las hipótesis específicas que el investigador debe responder para completar el conjunto de elementos necesarios para resolver el problema de determinación del tamaño muestral óptimo está precisamente relacionada con el proceso de optimización a seguir. Ya se ha indicado que la estratificación de una población conlleva un incremento de difi-

cultad en el tratamiento formal de cualquier proceso de experimentación asociado con esa población. Pero, por el contrario, aumenta la riqueza de sus resultados, entendiéndose por riqueza el aumento de información obtenida por la consideración de nuevas cantidades que pueden estimarse y ser investigadas. En efecto, tal como ha podido apreciarse en capítulos anteriores la obtención de una muestra de tamaño óptimo n se realizaba teniendo en cuenta la información que contenía sobre una cierta cantidad de interés ϑ que se pretendía estimar como último objetivo. Sin embargo, tal como se ha visto en la sección 5.1, la estratificación transforma a su vez la anterior cantidad de interés en el nuevo vector paramétrico o de interés $\underline{\vartheta}$, de L dimensiones (tantas como estratos) y cuyas componentes ϑ_i , en el ámbito de aplicación respectivo de su estrato, admiten idéntica interpretación que la satisfecha por ϑ para la población total. Pero aún hay más: existe una relación conocida que liga ϑ con el vector $\underline{\vartheta}$ y que permite transformar densidades de probabilidad de la cantidad $\underline{\vartheta}$ (y, en particular, la que describe las opiniones iniciales que el investigador posee sobre ella) en las correspondientes densidades de probabilidad del parámetro ϑ (en concreto, siempre que el investigador sea coherente en sus opiniones, también obtendrá de esta forma la densidad posterior). Por otra parte, el experimento $E^S(\underline{n})$ se ha definido como en conjunto de L experimentos independientes realizados en sus estratos respectivos de acuerdo con unas densidades de probabilidad que dependen para cada estrato de la respectiva cantidad de interés ϑ_i . De cada una de ellas se obtiene la submuestra respectiva, constituyendo el conjunto de todas ellas la muestra total. Por tanto, también para cada una de las componentes ϑ_i puede calcularse la información esperada proporcionada por el también experimento componente $E(n_i)$ o por todo el experimento $E^S(\underline{n})$, toda vez que la muestra total contiene información sobre la cantidad de interés ϑ_i , y ello para $i=1,2,\dots,L$. Para esto basta considerar como la densidad inicial que describe las opiniones iniciales que el investigador posee sobre ϑ_i a la densidad $p(\vartheta_i)$ marginal correspondiente de la $p(\underline{\vartheta})$, a fin de asegurar la coherencia en la actuación de dicho investigador. En resumen, nos encontramos con las posibles cantidades de interés ϑ , $\underline{\vartheta}$ y ϑ_i ($i=1,2,\dots,L$), consideradas estas últimas por separado, y cada una de ellas con su correspondiente densidad inicial: $p(\vartheta)$ deducida de la densidad $p(\underline{\vartheta})$ definida por el investigador para describir sus opiniones iniciales

sobre $\underline{\vartheta}$ y obtenida a través de la relación funcional $\vartheta = f(\underline{\vartheta})$; y $p(\vartheta_i)$ ($i=1,2,\dots,L$) como correspondiente marginal de $p(\underline{\vartheta})$. En definitiva, descritas sus opiniones iniciales por la densidad $p(\underline{\vartheta})$ y a fin de asegurar la coherencia del investigador, las densidades $p(\vartheta)$ y $p(\vartheta_i)$ ($i=1,2,\dots,L$) se deducen de la forma indicada. Así pues, este mismo investigador deberá centrar el problema fijando con exactitud qué objetivo último pretende con la realización del experimento $E^S(\underline{n})$ y qué cantidad o cantidades de interés está interesado en estimar. Esta decisión es de fundamental importancia en cuanto que condiciona el proceso de optimización y especifica qué expresiones deben utilizarse en dicho proceso. Y todo esto porque la optimización puede perseguir el conocimiento de la única cantidad ϑ , o de todo el vector paramétrico $\underline{\vartheta}$ ó, por último, puede pretender optimizar la información esperada sobre una ó varias de las cantidades ϑ_i por separado. Realicemos un breve comentario sobre cada una de estas posibilidades.

Supongamos, primeramente, que la estratificación de la población X se realiza única y exclusivamente con objeto de acceder de la forma más sencilla posible a una muestra que proporcione información sobre la cantidad aleatoria ϑ . Esto es, el investigador está interesado tan sólo en la cantidad ϑ , no importándole por tanto la posible estimación de todo el vector $\underline{\vartheta}$. Para ello deberá ser capaz de encontrar una expresión adecuada para medir la información esperada que sobre ϑ contiene el experimento $E^S(\underline{n})$ que da lugar a un proceso de muestreo estratificado. Precisamente, la relación funcional que liga ϑ con el vector $\underline{\vartheta}$ es la que proporciona la solución al problema. En efecto, recordemos que el investigador ha descrito sus opiniones iniciales a través de la densidad de probabilidad $p(\underline{\vartheta})$. Con ella y a partir también de la conocida densidad $p(z/\underline{\vartheta})$ que describe a la muestra total obtenida del experimento $E^S(\underline{n})$ en función del vector $\underline{\vartheta}$ puede deducirse la densidad posterior $p(\vartheta/z)$ por el teorema de Bayes tal como se indicó en la sección 5.1. Pues bien, a partir de ambas densidades, $p(\underline{\vartheta})$ y $p(\vartheta/z)$ y por la relación funcional $\vartheta = f(\underline{\vartheta})$ puede, a su vez, deducirse las nuevas densidades de probabilidad

$$p(\vartheta) \text{ a partir de } p(\underline{\vartheta}) \text{ por la relación } \vartheta = f(\underline{\vartheta});$$

$$p(\vartheta/z) \text{ a partir de } p(\underline{\vartheta}/z) \text{ por la relación } \vartheta = f(\underline{\vartheta});$$

de forma que, conocidas $p(\vartheta)$ y $p(\vartheta/z)$ puede aplicarse la doble integral

de Lindley (1956). Resulta, pues, la siguiente

DEFINICION. La información esperada sobre ϑ proporcionada por el experimento $E^S(\underline{n})$ cuando el resultado muestral obtenido es z y la densidad inicial $p(\vartheta)$ es la deducida de la densidad $p(\underline{z})$ a partir de la relación funcional que liga ϑ con \underline{z} , viene dada a través de la doble integral

$$I^{\vartheta}\{E^S(\underline{n}), p(\underline{z})\} = \iint p(z) \cdot p(\vartheta/z) \cdot \log \frac{p(\vartheta/z)}{p(\vartheta)} \cdot d\vartheta \cdot dz$$

La definición anterior admite como expresiones equivalentes

$$\iint p(\vartheta) \cdot p(z/\vartheta) \cdot \log \frac{p(\vartheta/z)}{p(\vartheta)} \cdot d\vartheta \cdot dz = \iint p(\vartheta, z) \cdot \log \frac{p(\vartheta/z)}{p(\vartheta)} \cdot d\vartheta \cdot dz$$

y donde z toma valores en el espacio

$$\prod_{i=1}^L X_i^{(n_i)} = \prod_{i=1}^L X_i^{x_i} \times \dots \times X_i^{(n_i)}$$

En todas las expresiones anteriores la densidad $p(\vartheta)$ actúa como densidad inicial para ϑ a fin de que el investigador sea coherente. Notemos también que cualquiera de las submuestras z_i contiene información sobre la cantidad de interés ϑ , de forma que efectivamente tiene sentido el empleo de la densidad $p(\vartheta/z)$ en las expresiones anteriores. Además, la información esperada anterior sigue siendo función del vector tamaño muestral \underline{n} dado que sus componentes n_i aparecen en las expresiones de $p(\vartheta/z)$ al ser

$$p(\underline{z}/z) = \prod_{k=1}^L \prod_{i=1}^{n_k} p(x_{ki}/\theta_k)$$

Así pues, la información esperada que sobre ϑ contiene la muestra estratificada obtenida como resultado del experimento $E^S(\underline{n})$ puede cuantificarse mediante la expresión ya conocida y desarrollada por Lindley (1956). En este sentido el problema sigue siendo análogo al resuelto en capítulos anteriores sin más que considerar con especial cuidado la procedencia de las densidades de probabilidad que aparecen en la doble integral. No obstante, el problema real nace de la obtención práctica de dichas densidades $p(\vartheta)$ y $p(\vartheta/z)$ ya que puede resultar tan extraordinariamente difícil que

desaconje, incluso, el ceñirse como último objetivo a la estimación de la única cantidad de interés ϑ . Pensemos, por ejemplo, que ante la sencilla y frecuente relación lineal como función que liga ϑ con el vector $\underline{\vartheta}$, esto es, siendo

$$\vartheta = \sum_{i=1}^L h_i \cdot \vartheta_i \quad h_i > 0 \quad i=1,2,\dots,L$$

la correspondiente densidad inicial para ϑ se calculará, según comprobaremos más adelante, por la expresión general (Sixto Ríos, 1970)

$$p(\vartheta) = \frac{1}{h_L} \int_{\vartheta_1} \int_{\vartheta_2} \dots \int_{\vartheta_{L-1}} p(\vartheta_1, \vartheta_2, \dots, \vartheta_{L-1}, \frac{\vartheta - \sum_{i=1}^{L-1} h_i \cdot \vartheta_i}{h_L}) \cdot d\vartheta_1 \cdot d\vartheta_2 \dots d\vartheta_{L-1}$$

donde la densidad de probabilidad que aparece en el integrando es la misma que la densidad de probabilidad que describe las opiniones iniciales que el investigador posee sobre $\underline{\vartheta}$. Dependerá de dicha densidad $p(\underline{\vartheta})$ la facilidad de cálculo de la expresión anterior. Idéntico razonamiento es aplicable a la densidad posterior $p(\vartheta/z)$.

La segunda de las actitudes que el investigador puede tomar frente al proceso de optimización es la que se refiere a la consideración del vector paramétrico en su conjunto. Esta situación responde, tal vez, a la deducción normal asociada al proceso de estratificación y resuelve el problema de la forma más semejante a la realizada en el modelo unidimensional. Decíamos antes que la estratificación aumenta la riqueza en cuanto al número de características a investigar. Efectivamente, la introducción de los estratos en la población X lleva consigo la consideración de L nuevas cantidades de interés ϑ_i que constituyen las componentes del vector aleatorio $\underline{\vartheta}$ y tales que su exacta determinación implica la exacta determinación de la cantidad primitiva ϑ a través de la relación funcional $\vartheta = f(\underline{\vartheta})$. De ahí que el acopio de información sobre el vector $\underline{\vartheta}$ permita, a su vez, tener información sobre la cantidad ϑ obtenida, por tanto, como consecuencia de la investigación y estudio centrados en $\underline{\vartheta}$. Parece lógico, pues, que una buena cantidad de información sobre las cantidades aleatorias ϑ_i ($i=1,2,\dots,L$) consideradas en su conjunto permitan, a su vez, tener un buen conocimiento del paráme

tro ϑ , permitiendo, además, determinar el comportamiento que respecto a la característica objeto de estudio tienen todos y cada uno de los estratos. Este es el sentido de la riqueza en la estratificación que antes apuntábamos.

Supongamos, por tanto, que para el proceso de optimización de la información el investigador está interesado en la obtenida para el vector $\underline{\vartheta}$. La cuantificación de esta información esperada sobre $\underline{\vartheta}$ proporcionada por el experimento $E^S(\underline{n})$ se realiza por la expresión ya conocida de Lindley (1956). En este caso, de las densidades que aparecen en la formulación, $p(\underline{\vartheta})$ es la densidad inicial que describe las opiniones que el investigador tiene sobre el vector paramétrico y la $p(\underline{\vartheta}/z)$ es la correspondiente densidad obtenida por Bayes a partir de la también conocida $p(z/\underline{\vartheta})$. Puede, pues, darse la siguiente

DEFINICION. La información esperada sobre $\underline{\vartheta}$ proporcionada por el experimento $E^S(\underline{n})$ cuando el resultado muestral obtenido es z y la densidad inicial es $p(\underline{\vartheta})$, viene dada por la conocida doble integral

$$I^{\underline{\vartheta}}\{E^S(\underline{n}), p(\underline{\vartheta})\} = \iint p(z) \cdot p(\underline{\vartheta}/z) \cdot \log \frac{p(\underline{\vartheta}/z)}{p(\underline{\vartheta})} \cdot d\underline{\vartheta} \cdot dz$$

Esta información, que aparecerá en la expresión del valor esperado del experimento, es la que debe maximizarse ante el problema de determinación del óptimo tamaño muestral. Notemos, en este sentido, que $I^{\underline{\vartheta}}\{E^S(\underline{n}), p(\underline{\vartheta})\}$ es función del vector tamaño muestral \underline{n} , toda vez que sus componentes n_i aparecen en el cálculo de las densidades $p(z/\underline{\vartheta})$ y $p(\underline{\vartheta}/z)$.

Comparando los dos criterios anteriores expresados a través de sus informaciones esperadas respectivas puede apreciarse la ganancia de información al utilizar la segunda de las expresiones, resultado completamente lógico puesto que la estimación del vector $\underline{\vartheta}$ implica, según hemos dicho, la estimación de la cantidad ϑ . Notemos que todo el aparato introducido por la estratificación tiene mucho más sentido, evidentemente, para la estimación del vector $\underline{\vartheta}$ que para la de la cantidad ϑ . En efecto, Bernardo (1978) estudia el concepto de información esperada ante transformaciones $\psi = \psi(\vartheta)$ de la cantidad de interés, definiendo - tal como mencionábamos en el capítulo primero - la información esperada útil que no es sino la información esperada sobre la nueva cantidad ψ . También es conocido el resultado de que

la información esperada útil sobre ψ no es mayor que la información esperada sobre θ , siendo iguales en el caso de transformaciones biyectivas de la cantidad de interés. Aplicando este resultado a nuestro caso (posteriormente comprobaremos las condiciones de aplicabilidad del mismo), siendo la transformación no necesariamente biyectiva $\vartheta = f(\underline{\vartheta})$, deducimos el resultado

$$I^{\vartheta} \{E^S(\underline{n}), p(\underline{\vartheta})\} \leq I^{\vartheta} \{E^S(\underline{n}), p(\underline{\vartheta})\}$$

esto es, la información esperada sobre el vector $\underline{\vartheta}$ proporcionada por el experimento $E^S(\underline{n})$ cuando las opiniones iniciales se describen por la densidad L-dimensional $p(\underline{\vartheta})$ no es menor que la información esperada sobre θ proporcionada por dicho experimento. Dado que las transformaciones lineales citadas no son biyectivas, el resultado anterior se verifica con desigualdad para estas transformaciones de tanta aplicación. Más adelante volveremos sobre esta situación para justificarla debidamente.

Una tercera posibilidad que tiene el investigador para abordar el problema planteado por la determinación del tamaño muestral consiste en considerar las informaciones esperadas que la muestra total z contiene sobre una o varias de las cantidades ϑ_i ($i=1,2,\dots,L$) por separado. Pensemos que la interpretación dada a estos parámetros, interpretación nacida de su misma construcción, es idéntica a todos ellos y al parámetro general ϑ , dependiendo su comportamiento única y exclusivamente de su campo de aplicación. Por otra parte, la muestra z_i extraída en el estrato i -ésimo no sólo contiene información sobre su correspondiente parámetro ϑ_i , sino también sobre el resto de las cantidades ϑ_j , toda vez que dichas cantidades no tienen por qué ser independientes. Pensemos nuevamente en una encuesta de población activa cuyo último objetivo sea la estimación del nivel de desempleo ϑ y en la que, por ejemplo, se realiza una estratificación regional. Es ilógico pensar que, salvo excepcionales circunstancias, los niveles de paro en las distintas regiones del país difieran excesivamente y tengan, incluso, un comportamiento independiente. Lo lógico sería considerar que, con las diferencias esperadas y posibles, el comportamiento de las ϑ_i siga una tendencia determinada, puesto que la fuerte interrelación de las economías regionales en una economía nacional no permita, evidentemente, comportamientos excesivamente diferenciados. En definitiva, si bien el proceso que permite obtener las distintas muestras parciales z_i cuyo conjunto constituye la muestra total es independien

te, lo cierto es que al no serlo necesariamente las cantidades de interés θ_i , todas y cada una de las muestras z_i contienen, en general, información sobre todas y cada una de las cantidades θ_i ($i=1,2,\dots,L$). Por tanto la muestra total z proporciona información sobre todas las θ_i por separado y en su conjunto. De ahí que tengan sentido las densidades $p(\theta_i/z)$ ($i=1,2,\dots,L$) que vamos a considerar.

Conocidas, pues, las densidades $p(\underline{\theta})$ y $p(z/\underline{\theta})$ dadas como datos en cualquier problema donde intervenga el experimento $E^S(\underline{n})$, puede conocerse a su vez por Bayes la densidad $p(\underline{\theta}/z)$ que describe las opiniones posteriores del investigador una vez que ha realizado $E^S(\underline{n})$ y obtenido z como resultado muestral. Tienen sentido, por tanto, las nuevas densidades de probabilidad $p(\theta_i)$ y $p(\theta_i/z)$, que son las correspondientes densidades marginales para θ_i de las $p(\underline{\theta})$ y $p(\underline{\theta}/z)$, respectivamente, esto es,

$$p(\theta_i) = \int_{\Theta'} p(\underline{\theta}) \cdot d\underline{\theta}'$$

$$p(\theta_i/z) = \int_{\Theta'} p(\underline{\theta}/z) \cdot d\underline{\theta}'$$

donde

$$\Theta' = \prod_{j \neq i}^L \Theta_j \quad \text{y} \quad d\underline{\theta}' = d\theta_1 \cdot d\theta_2 \cdots d\theta_{i-1} \cdot d\theta_{i+1} \cdots d\theta_L$$

A partir de estas dos últimas densidades de probabilidad puede darse la siguiente

DEFINICION. La información esperada sobre θ_i proporcionada por el experimento $E^S(\underline{n})$ cuando el resultado muestral obtenido es z y la densidad inicial es $p(\theta_i)$ (esto es, la marginal i -ésima de la densidad $p(\underline{\theta})$), viene dada a través de la conocida doble integral

$$I^{\theta_i} \{ E^S(\underline{n}), p(\underline{\theta}) \} = \iint p(z) \cdot p(\theta_i/z) \cdot \log \frac{p(\theta_i/z)}{p(\theta_i)} \cdot d\theta_i \cdot dz$$

Esta definición responde a la estructura de toda información esperada tal como la define Lindley (1956) y donde la densidad

$$p(z) = \int p(z/\underline{\theta}) \cdot p(\underline{\theta}) \cdot d\underline{\theta}$$

es conocida.

Resulta imprescindible destacar dos circunstancias en la doble integral anterior. Por una parte, $I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\}$ es, nuevamente, función del vector tamaño muestral, pues sus distintas componentes aparecen todas en la obtención de la densidad posterior $p(\vartheta_i/z)$. En efecto, notemos que

$$p(\vartheta_i/z) = \frac{1}{p(z)} \int_{\vartheta_1} \dots \int_{\vartheta_{i-1}} \int_{\vartheta_{i+1}} \dots \int_{\vartheta_L} p(\vartheta_1, \dots, \vartheta_L).$$

$$= \prod_{k=1}^L \prod_{j=1}^n p(x_{jk}/\vartheta_k) \cdot d\vartheta_1 \dots d\vartheta_{i-1} \cdot d\vartheta_{i+1} \dots d\vartheta_L$$

Por otra parte, la anterior información está calculada incluyendo la influencia de todo el experimento $E^S(\underline{n})$ y no solamente de la incidencia que sobre ϑ_i tiene el experimento respectivo componente $E(n_i)$, esto es, la información $I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\}$ es claramente distinta de la $I^{\vartheta_i} \{E(n_i), p(\vartheta_i)\}$, pues basta comparar sus expresiones, siendo

$$I^{\vartheta_i} \{E(n_i), p(\vartheta_i)\} = \iint p(\vartheta_i) \cdot p(z_i/\vartheta_i) \cdot \log \frac{p(\vartheta_i/z_i)}{p(\vartheta_i)} \cdot d\vartheta_i \cdot dz_i$$

En resumen, $I^{\vartheta_i} \{E^S(\underline{n})\}$ es función del vector tamaño muestral \underline{n} y responde a la consideración de una expresión que contiene la información que sobre la cantidad de interés ϑ_i lleva consigo toda la muestra z resultado global de la realización del experimento completo $E^S(\underline{n})$.

Otra vez el resultado encontrado por Bernardo (1978) es de aplicación en estas circunstancias a fin de determinar la relación entre las informaciones esperadas que sobre ϑ_i proporcionan, respectivamente, los experimentos $E^S(\underline{n})$ y $E(n_i)$, sin más que considerar la transformación m_i , marginal i -ésima. En efecto, siendo

$$\vartheta_i = m_i(\underline{\vartheta})$$

deducimos el resultado

$$I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\} > I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\}$$

lo que, en analogía a lo expresado para $I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\}$, parece enfatizar una vez más que el experimento $E^S(\underline{n})$ asociado a un proceso de estra-

tificación cobra su verdadera dimensión actuando sobre el vector paramétrico $\underline{\vartheta}$, más bien que sobre la función ϑ o la componente i -ésima ϑ_i . Una vez más, volveremos más adelante sobre esta situación para justificarla debidamente.

La tercera de las actitudes que puede tomar el investigador se centra, pues, en el estudio de una sola de las cantidades de interés ϑ_i que el proceso de estratificación ha definido. Supongamos, pues, que el investigador está interesado únicamente en el estudio de ϑ_i y para ello idea el diseño del experimento $E^S(\underline{n})$ a fin de recabar información sobre ella. La información esperada sobre ϑ_i proporcionada por $E^S(\underline{n})$ la cuantificará a través de la ya conocida $I^{\vartheta_i}\{E^S(\underline{n}), p(\underline{\vartheta})\}$. Será esta información la que aparecerá en las expresiones del valor esperado de $E^S(\underline{n})$ y que, a su vez, permitirán resolver el problema planteado por la determinación del tamaño muestral óptimo.

Evidentemente que puede resultar curioso el hecho de que el investigador considere la realización de todo el experimento $E^S(\underline{n})$ cuando tan sólo se interese por la cantidad ϑ_i . Sin embargo esta situación es factible en la práctica cuando el investigador, deseando obtener información sobre ϑ_i se enfrenta con una muestra o un diseño conocidos para los que existe una determinada estratificación. Entonces debe ser capaz de cuantificar esta información, teniendo en cuenta que la información proporcionada por $E^S(\underline{n})$ sobre ϑ_i es mayor que la proporcionada por el experimento componente $E(n_i)$ como se comprobará más adelante. De ahí que disponiendo ya de una muestra estratificada resulte, por término medio, más interesante conocer la información que sobre ϑ_i contiene la muestra global z que la correspondiente al estrato en cuestión, la $z_i(n_i)$ resultado de $E(n_i)$. Pero aún hay más, hay ocasiones en las que, incluso, resulta interesante realizar una estratificación de antemano a fin de recabar información sobre cierto parámetro ϑ_i . Supongamos, por ejemplo, que pretende estimarse una cierta cantidad ϑ para lo cual se dispone de una población X cuyas unidades guardan cierta relación con dicho parámetro. Dentro de esa población, existe un subconjunto $X_1 \subset X$ cuyo comportamiento respecto ϑ es considerado como normal, existiendo otro subconjunto $X_2 = X - X_1$ formado por aquellas unidades cuyo comportamiento sobre ϑ es completamente atípico, pero sin que ello quiera decir

que sus unidades no contengan información (que en este caso será confirmación sobre lo que no debe valer ϑ) sobre el parámetro objeto de estudio. Es evidente que en este caso una adecuada estratificación puede proporcionar una magnífica información sobre ϑ , superior a la que se obtendría considerando la población X sin estratos.

Así pues, el tercer camino que se le abre al investigador ante una población estratificada consiste en optimizar la información que sobre una cantidad de interés concreta ϑ_i proporciona el experimento $E^S(\underline{n})$ considerado globalmente. Dado que $I^{\vartheta_i}\{E^S(\underline{n}), p(\underline{\vartheta})\}$ es función del vector tamaño muestral el proceso de optimización de la misma sigue estando en la línea del ideado para los casos anteriores. Ello conduce a considerar un proceso que maximice a la información que proporciona $E^S(\underline{n})$ sobre una única cantidad de interés ϑ_i y no sobre el vector $\underline{\vartheta}$ considerado como conjunto de los parámetros ϑ_i ($i=1,2,\dots,L$). Notemos que, por otra parte, la imposibilidad existente para optimizar al conjunto de valores esperados

$$w(I^{\vartheta_i}\{E^S(\underline{n}), p(\underline{\vartheta})\}) - C(E^S(\underline{n})) \quad i=1,2,\dots,L$$

toda vez que el vector tamaño muestral óptimo para ϑ_i no tiene por qué coincidir con el obtenido para ϑ_j . Sin embargo, existe una situación especial de independencia en la que sí puede obtenerse el vector \underline{n} óptimo que maximiza todos los valores anteriores del experimento $E^S(\underline{n})$, esto es, que optimiza las informaciones sobre ϑ_i proporcionadas por $E^S(\underline{n})$ conjuntamente. Es el caso de independencia de los parámetros ϑ_i . También sobre estas ideas volveremos en su momento.

Pasemos ya al estudio exhaustivo de los casos mencionados, discutiendo las propiedades satisfechas por las distintas informaciones proporcionadas por el único experimento $E^S(\underline{n})$ sobre las cantidades de interés $\underline{\vartheta}$, $\underline{\vartheta}$ y ϑ_i . Para estas informaciones encontraremos las relaciones de dominancia de unas sobre otras. Resolveremos también el problema teórico planteado por la determinación del tamaño muestral óptimo.

5.3 INFORMACION ESPERADA $I^{\vartheta}\{E^S(\underline{n}), p(\underline{\vartheta})\}$

Supongamos una investigación que pretende recabar información so

bre la cantidad de interés desconocida ϑ . A tal fin, el investigador idea un experimento consistente en la obtención de una muestra estratificada de u na población cuyos elementos tienen cierta característica que guarda con ϑ una relación determinada, expresada por sendas funciones de probabilidad definidas en sus estratos respectivos. Al realizar la estratificación aparecen las nuevas cantidades aleatorias ϑ_i ($i=1,2,\dots,L$) sobre las que el investigador tiene unas opiniones iniciales expresadas por una densidad conjunta que, al mismo tiempo, especifica la posible interdependencia entre ellas. Dicha densidad inicial se ha expresado como

$$p(\underline{\vartheta}) = p(\vartheta_1, \vartheta_2, \dots, \vartheta_L)$$

Entre las cantidades aleatorias ϑ y ϑ_i ($i=1,2,\dots,L$) existe una relación funcional (que se supone lineal) dada por

$$\vartheta = f(\underline{\vartheta}) = \sum_{i=1}^L h_i \cdot \vartheta_i \quad h_i > 0 \quad i=1,2,\dots,L$$

Sin embargo, y a pesar de que el conocimiento exacto de los parámetros ϑ_i (o su buena estimación) suponga conocer con exactitud a la cantidad aleatoria ϑ (o una buena estimación) a partir de la relación funcional mencionada, el investigador desea ceñirse al estudio de ϑ , de forma que debe maximizar la información que la muestra z contiene sobre ella, no importándole, pues, la información que sobre las otras ϑ_i contiene z .

Para resolver el problema, supongamos realizado el experimento $E^S(\underline{n})$ y conocido el resultado muestral obtenido z como conjunto de las L submuestras de los estratos, $z = (z_1, z_2, \dots, z_L)$, con

$$z_i = (x_{i1}, x_{i2}, \dots, x_{in_i}) \quad i=1,2,\dots,L$$

La densidad de probabilidad que describe a la muestra obtenida como dependiente del vector paramétrico $\underline{\vartheta}$ viene dada por el doble productorio

$$p(z/\underline{\vartheta}) = \prod_{k=1}^L \prod_{j=1}^{n_k} p(x_{kj}/\vartheta_k)$$

toda vez que las muestras extraídas dentro de los estratos son aleatorias.

Recordemos que $n = (n_1, n_2, \dots, n_L)$ es el vector tamaño muestral y que, cono

cidas $p(\underline{\vartheta})$ y $p(z/\underline{\vartheta})$ y aplicando Bayes, puede determinarse $p(\underline{\vartheta}/z)$ a través de la fórmula

$$p(\underline{\vartheta}/z) \propto p(z/\underline{\vartheta}) \cdot p(\underline{\vartheta}) = p(\vartheta_1, \vartheta_2, \dots, \vartheta_L) \cdot \prod_{k=1}^L \prod_{j=1}^{n_k} p(x_{kj}/\vartheta_k)$$

donde el factor de proporcionalidad es $1/p(z)$.

Teniendo en cuenta la relación lineal que liga ϑ con $\underline{\vartheta}$ y las densidades $p(\underline{\vartheta})$ y $p(\underline{\vartheta}/z)$, inicial y posterior, respectivamente, las densidades de probabilidad $p(\vartheta)$ y $p(\vartheta/z)$ referidas a la cantidad de interés ϑ (y que toma valores en el conjunto $\Theta = f(\underline{\vartheta})$) vendrán dadas a partir de la transformación siguiente:

$$\begin{aligned} \vartheta &= \sum_{i=1}^L h_i \cdot \vartheta_i \\ \alpha_1 &= \vartheta_1 \\ \alpha_2 &= \vartheta_2 \\ &\dots \\ \alpha_{L-1} &= \vartheta_{L-1} \end{aligned}$$

siendo la correspondiente transformación inversa la

$$\begin{aligned} \vartheta_1 &= \alpha_1 \\ \vartheta_2 &= \alpha_2 \\ &\dots \\ \vartheta_{L-1} &= \alpha_{L-1} \\ \vartheta_L &= \frac{\vartheta - \sum_{i=1}^{L-1} h_i \cdot \vartheta_i}{h_L} \end{aligned}$$

El correspondiente jacobiano para la transformación inversa es:

$$J_1 = \frac{\partial(\vartheta_1, \vartheta_2, \dots, \vartheta_L)}{\partial(\vartheta, \alpha_1, \dots, \alpha_{L-1})} = \frac{1}{h_L}$$

De donde la densidad de probabilidad de las nuevas variables aleatorias

$(\vartheta, \alpha_1, \dots, \alpha_{L-1})$ vendrá dada por la función

$$\frac{1}{h_L} \cdot p_{\vartheta}(\vartheta_1, \vartheta_2, \dots, \vartheta_{L-1}, \frac{\vartheta - \sum_{i=1}^{L-1} h_i \cdot \vartheta_i}{h_L})$$

donde con la notación p_{ϑ} indicamos la densidad de probabilidad inicial del vector ϑ . De ahí que las densidades $p(\vartheta)$ y $p(\vartheta/z)$ vengan dadas por las correspondientes marginales de las anteriores densidades, esto es, por

$$p(\vartheta) = \frac{1}{h_L} \int_{\Theta_1} \int_{\Theta_2} \dots \int_{\Theta_{L-1}} p_{\vartheta}(\vartheta_1, \vartheta_2, \dots, \vartheta_{L-1}, \frac{\vartheta - \sum_{i=1}^{L-1} h_i \cdot \vartheta_i}{h_L}) \cdot d\vartheta_1 \cdot d\vartheta_2 \cdot \dots \cdot d\vartheta_{L-1}$$

$$\cdot d\vartheta_1 \cdot d\vartheta_2 \cdot \dots \cdot d\vartheta_{L-1}$$

y por

$$p(\vartheta/z) = \frac{1}{p(z) \cdot h_L} \int_{\Theta_1} \int_{\Theta_2} \dots \int_{\Theta_{L-1}} p_{\vartheta}(\vartheta_1, \vartheta_2, \dots, \vartheta_{L-1}, \frac{\vartheta - \sum_{i=1}^{L-1} h_i \cdot \vartheta_i}{h_L}) \cdot d\vartheta_1 \cdot d\vartheta_2 \cdot \dots \cdot d\vartheta_{L-1}$$

$$\cdot \left(\prod_{k=1}^{L-1} \prod_{j=1}^n p(x_{jk}/\vartheta_k) \right) \cdot \prod_{j=1}^n p(x_{jL} / \frac{\vartheta - \sum_{i=1}^{L-1} h_i \cdot \vartheta_i}{h_L}) \cdot d\vartheta_1 \cdot d\vartheta_2 \cdot \dots \cdot d\vartheta_{L-1}$$

Implícitamente se ha supuesto que la existencia de todas las densidades e integrales anteriores.

Conocidas las densidades $p(\vartheta)$ y $p(\vartheta/z)$, la información esperada sobre ϑ proporcionada por el experimento $E^S(\underline{n})$ asociado a un proceso de muestreo estratificado, vendrá dada por la ya conocida doble integral (Bernardo, 1978):

$$I^{\vartheta} \{E^S(\underline{n}), p(\vartheta)\} = \int p(z) \int p(\vartheta/z) \cdot \log \frac{p(\vartheta/z)}{p(\vartheta)} \cdot d\vartheta \cdot dz$$

que admite las expresiones alternativas mencionadas con anterioridad y donde z toma valores en el conjunto

$$Z = \prod_{i=1}^L x_i \times x_i \times \dots \times x_i^{(n_i)} \times x_i$$

Notemos que $I^{\theta} \{E^S(\underline{n}), p(\underline{\theta})\}$ es la información esperada útil sobre θ ante la transformación no necesariamente biyectiva $\vartheta = f(\underline{\vartheta})$ del vector de interés $\underline{\vartheta}$, en el sentido definido por Bernardo (1978).

Es hora ya de establecer la hipótesis del conocimiento previo o no de la afijación de la estratificación, dado que dicha hipótesis condiciona el que la información esperada que acaba de definirse sea función de una o de L variables, respectivamente. Consideremos primeramente que la afijación viene determinada de antemano. Así, para un determinado tamaño muestral total n se conoce, de forma inmediata, los distintos tamaños muestrales n_i de las muestras extraídas en sus estratos respectivos y que, en conjunto, constituyen la muestra total z . Expresado de esta forma puede, por tanto, suponerse la existencia de L factores f_i verificando las relaciones

$$n_i = f_i \cdot n \quad i=1,2,\dots,L$$

y que indican qué proporción del tamaño muestral total n corresponde a la muestra de cada uno de los estratos. Evidentemente se admite que $f_i > 0$ para todos ellos, pues si existiese un estrato en el que $f_i = 0$ el problema quedaría reducido a uno idéntico sin más que considerar como población objeto de estudio la $X - X_i$. Notemos que la condición

$$n = \sum_{i=1}^L n_i$$

implica, a su vez, la exigencia

$$\sum_{i=1}^L f_i = 1$$

Para afijaciones particulares el conjunto $\{f_i\}_{i=1,2,\dots,L}$ toma, a su vez, valores determinados. Por ejemplo, en poblaciones finitas en las que se supone la aplicabilidad de nuestros resultados, la afijación uniforme en la que n_i es igual en todos los estratos, resultaría:

$$f_i = 1/L \quad i=1,2,\dots,L \quad L = \text{número de estratos}$$

En el caso de una afijación proporcional, en la que n se distribuye proporcionalmente a los valores de N_h ó tamaño del estrato h -ésimo, resulta ser:

$$f_i = n/N \quad i=1,2,\dots,L \quad \text{con} \quad N = \sum_{i=1}^L N_i$$

Para la afijación óptima o de Neyman, donde n se distribuye atendiendo a la optimización de la precisión del posible estimador, resultaría:

$$f_i = N_i \cdot S_i / \sum_{j=1}^L N_j \cdot S_j \quad i=1,2,\dots,L \quad S_i = \text{variabilidad del estrato } i\text{-ésimo}$$

Lo verdaderamente interesante a destacar es, sin embargo, que conocida de antemano la afijación correspondiente del proceso de estratificación el problema de determinación del vector tamaño muestral \underline{n} queda reducido a la determinación de la única variable n , en cuanto que conocida ésta las n_i están perfectamente determinadas a través de los productos $n \cdot f_i$ ($i=1,2,\dots,L$). En la práctica y dado que estos productos no tienen por qué ser enteros, las unidades sobrantes se van distribuyendo en aquellos estratos de mayor afijación. Esto es, asignando como tamaños n_i los respectivos valores $\xi(n \cdot f_i)$ donde ξ es la función "parte entera", las

$$n = \sum_{i=1}^L \xi(n \cdot f_i)$$

unidades sobrantes van repartiéndose atendiendo a los estratos de mayor valor para f_i , mediante un procedimiento cíclico y perfectamente determinado a priori. Suponiendo un orden creciente (lo cual no supone pérdida de generalidad) en los valores de las constantes f_i , esto es,

$$f_1 \leq f_2 \leq \dots \leq f_i \leq \dots \leq f_L$$

y tomando como período el valor máximo $1/f_1$ unidades, cada vez que se aumenta el tamaño muestral total en dicho período se verificará que todo estrato verá aumentado en, al menos, una unidad su respectivo tamaño muestral n_i . Aún más, las situaciones expresadas por los tamaños muestrales totales n , $n - (1/f_1)$ y $n + (1/f_1)$ (suponiendo que $1/f_1 \in \mathbb{Z}$) son tales que el número de unidades incrementadas en el estrato genérico i -ésimo al pasar del tamaño muestral total $n - (1/f_1)$ al n coincide con el número de unidades incrementadas en dicho estrato al pasar del tamaño muestral total n al $n + (1/f_1)$. Puede suponerse que el tamaño muestral total es, al menos, $1/f_1$ con objeto de asegurar que se muestrea en todos los estratos, de acuerdo con la definición del muestreo estartificado. Esta suposición, no obstante, no es restrictiva puesto que si para un determinado n existe un estrato tal que

el producto $n \cdot f_i$ es tan pequeño que implica la no obtención de ni siquiera una unidad muestral puede suponerse una población de partida excluyendo dicho estrato. En muestras grandes es evidente que queda garantizada la obtención de unidades en todos los estratos. Además, en muchas aplicaciones prácticas suele llegarse a situaciones de compromiso ante la exigencia de muestrear mínimamente en todos los estratos. Pensemos en el muestreo mínimo exigido en determinadas encuestas realizadas por el INE dentro del marco de la Encuesta General de Población a fin de asegurar resultados significativos en todas las provincias. Ello supone la existencia de un tamaño muestral n_0 mínimo.

Supuesta la afijación conocida el vector tamaño muestral \underline{n} queda reducido al

$$\underline{n} = (n \cdot f_1, n \cdot f_2, \dots, n \cdot f_L)$$

que, a su vez, queda determinado inmediatamente fijando un tamaño muestral total concreto n . Por otra parte, el experimento $E^S(\underline{n})$ consistirá ahora en la realización de L experimentos en sendos estratos y con tamaños n_i conocidos, esto es, será la familia de experimentos

$$E^S(\underline{n}) = \left\{ \left\{ X_i(n \cdot f_i) \right\}, \left\{ \Theta_i \right\}, \left\{ p(X_i(n \cdot f_i) / \Theta_i) \right\} \right\}_{i=1,2,\dots,L}$$

De ahí que bajo dicha hipótesis la información esperada sobre Θ proporcionada por $E^S(\underline{n})$ (donde al quedar \underline{n} determinado al conocer simplemente n empleamos la notación anterior), esto es, $I^{\Theta} \{ E^S(\underline{n}), p(\Theta) \}$ es una función de la única variable n ó tamaño muestral total. Su expresión es idéntica a la de $I^{\Theta} \{ E^S(\underline{n}), p(\Theta) \}$ sin más que considerar sustituidos en los dobles productorios que definen a $p(z)$ y $p(\Theta/z)$ los tamaños muestrales n_i por los productos $n \cdot f_i$ ($i=1,2,\dots,L$).

Recordemos que el objetivo final sigue siendo la determinación de un vector tamaño muestral óptimo que, en este caso, queda reducido a la determinación del tamaño muestral total óptimo. Para ello hay que insistir en que el investigador diseña el experimento $E^S(\underline{n})$ a fin de recabar información sobre Θ . Es conocido que siempre existe ganancia de información cuando el tamaño muestral aumenta. De ahí que, por una parte, haya que considerar se la utilidad esperada del experimento $E^S(\underline{n})$, esto es, $U(E^S(\underline{n}))$, y por otra, su coste esperado, $C(E^S(\underline{n}))$, de forma que el tamaño óptimo debe ser una solución de compromiso entre la utilidad y costo esperados. Estas ideas,

ya desarrolladas en la sección 4.1, siguen siendo válidas al aplicarlas a un experimento que suponga un proceso de estratificación. De ahí que definiendo el valor esperado del experimento $E^S(n)$ como

$$W(E^S(n)) = U(E^S(n)) - C(E^S(n))$$

esto es, como diferencia entre su utilidad y costo esperados, el investigador, si desea ser coherente al elegir el tamaño muestral óptimo, lo tomará con la condición que maximice el valor esperado de $E^S(n)$, esto es, de la ecuación

$$\frac{d W(E^S(n))}{dn} = 0 \quad (3)$$

o lo que es lo mismo, de la condición

$$\frac{d U(E^S(n))}{dn} = \frac{d C(E^S(n))}{dn}$$

Nuevamente sería de aplicación en estos momentos el estudio de qué particular función de la utilidad esperada debería emplearse para la obtención del óptimo. Análogamente a lo desarrollado en los capítulos tercero y cuarto y con una función de costo esperado genérica, $C(E^S(n))$, podrían considerarse las siguientes expresiones del valor esperado del experimento $E^S(n)$ según utilicemos sendas expresiones para la utilidad esperada del mismo:

$$W_1(E^S(n)) = g_1 \cdot I^{\vartheta} \{E^S(n), p(\vartheta)\} - C(E^S(n))$$

$$W_2(E^S(n)) = g_2 \cdot (1 - \exp(-I^{\vartheta} \{E^S(n), p(\vartheta)\})) - C(E^S(n))$$

$$W_3(E^S(n)) = g_3 \cdot \exp(-H(p(\vartheta))) \cdot (\exp(I^{\vartheta} \{E^S(n), p(\vartheta)\}) - 1) - C(E^S(n))$$

donde la interpretación de las constantes g_i es la realizada en capítulos pasados suponiendo el espacio total X y la densidad de probabilidad $p(\vartheta)$ es la deducida de la $p(\vartheta)$ a través de la transformación $\vartheta = f(\vartheta)$. Notemos además que en todas las expresiones de $W_i(E^S(n))$ ($i=1,2,3$) aparece como única variable el tamaño muestral total n . Sin embargo, la condición (3) (establecida suponiendo que la variable n es continua) puede sustituirse considerando que tanto la utilidad como el costo esperados son funciones del vector tamaño muestral \underline{n} y optimizando respecto dicho vector pero sujeto a las L condiciones

$$n_i = n \cdot f_i \quad i=1,2,\dots,L$$

Este es el sentido del siguiente

TEOREMA 5.3.1

El tamaño muestral óptimo como el entero no negativo que, maximizando el valor de $W(E^S(n))$, sea el más próximo a la solución de

$$\frac{d W(E^S(n))}{dn} = 0$$

coincide con el obtenido de la condición de optimización de $W(E^S(\underline{n}))$ sujeta a las L condiciones

$$n_i = f_i \cdot n \quad i=1,2,\dots,L$$

Demostración.- Supongamos que

$$W(E^S(n)) = U(E^S(\underline{n})) - C(E^S(\underline{n}))$$

siendo $\underline{n} = (n_1, n_2, \dots, n_L)$ y verificándose las L condiciones anteriores.

El criterio de optimización de Lagrange conduce a la expresión

$$\epsilon = U(E^S(\underline{n})) - C(E^S(\underline{n})) + \sum_{i=1}^L \lambda_i \cdot (n_i - f_i \cdot n)$$

como aquella que debe optimizarse. Este criterio conduciría al sistema de ecuaciones

$$\frac{\partial \epsilon}{\partial n_i} = \frac{\partial U(E^S(\underline{n}))}{\partial n_i} - \frac{\partial C(E^S(\underline{n}))}{\partial n_i} + \lambda_i = 0 \quad i=1,2,\dots,L$$

$$\frac{\partial \epsilon}{\partial n} = - \sum_{i=1}^L \lambda_i \cdot f_i = 0$$

$$\frac{\partial \epsilon}{\partial \lambda_i} = n_i - f_i \cdot n = 0 \quad i=1,2,\dots,L$$

Multiplicando las L primeras ecuaciones anteriores por f_i y sumándolas para todos los estratos, resultaría:

$$\sum_{i=1}^L f_i \cdot \frac{\partial U(E^S(\underline{n}))}{\partial n_i} - \sum_{i=1}^L f_i \cdot \frac{\partial C(E^S(\underline{n}))}{\partial n_i} + \sum_{i=1}^L \lambda_i \cdot f_i = 0$$

esto es,

$$\sum_{i=1}^L f_i \cdot \frac{\partial U(E^S(\underline{n}))}{\partial n_i} = \sum_{i=1}^L f_i \cdot \frac{\partial C(E^S(\underline{n}))}{\partial n_i}$$

De las L condiciones de la hipótesis se deduce que

$$\frac{d n_i}{d n} = f_i \quad i=1,2,\dots,L$$

de donde la igualdad entre sumatorios anterior puede reescribirse como

$$\sum_{i=1}^L \frac{d n_i}{d n} \cdot \frac{\partial U(E^S(\underline{n}))}{\partial n_i} = \sum_{i=1}^L \frac{d n_i}{d n} \cdot \frac{\partial C(E^S(\underline{n}))}{\partial n_i}$$

y que conduce a la condición (3) especificada en un principio

$$\frac{d U(E^S(n))}{d n} = \frac{d C(E^S(n))}{d n}$$

sin más que considerar la regla de la cadena para la diferenciación de funciones de varias variables.

(c.q.d.)

La aplicación de la condición (3) según se utilice una u otra expresión para $W_i(E^S(n))$ ($i=1,2,3$) conduciría al siguiente conjunto de ecuaciones, representativas de sendos problemas distintos, satisfechas por el tamaño muestral total óptimo:

$$g_1 \cdot \frac{d I^0\{E^S(n), p(\underline{g})\}}{d n} = \frac{d C(E^S(n))}{d n}$$

$$g_2 \cdot \frac{d I^0\{E^S(n), p(\underline{g})\}}{d n} \cdot e^{-I^0\{E^S(n), p(\underline{g})\}} = \frac{d C(E^S(n))}{d n}$$

$$g_3 \cdot e^{-H(p(\theta))} \cdot \frac{d I^{\theta} \{ E^S(n), p(\theta) \}}{dn} \cdot e^{I^{\theta} \{ E^S(n), p(\theta) \}} = \frac{d C(E^S(n))}{dn}$$

Con una función de costo lineal y siendo c_i el costo de observación de una unidad del estrato i -ésimo, el costo esperado sería:

$$C(E^S(n)) = \sum_{i=1}^L c_i \cdot n_i = n \cdot \sum_{i=1}^L c_i \cdot f_i$$

y las ecuaciones anteriores se complementarían con la derivada

$$\frac{d C(E^S(n))}{dn} = \sum_{i=1}^L c_i \cdot f_i$$

La determinación del tamaño muestral óptimo se ha realizado suponiendo que la afijación era conocida a priori y que por tanto se trataba de maximizar expresiones dependientes de la única variable n ó tamaño muestral total. En un segundo caso, la afijación es desconocida de antemano por lo que el investigador debe determinar un vector tamaño muestral óptimo. Esto supone, a su vez, especificar una afijación concreta (no olvidemos que al especificar las componentes n_i del vector \underline{n} queda fijada la manera de distribuir el tamaño muestral total) de forma que al mismo tiempo el tamaño muestral total n queda determinado a través de la suma conocida

$$n = \sum_{i=1}^L n_i$$

esto es, como suma de las componentes del vector \underline{n} . Desconocida, pues, la afijación de la estratificación el experimento $E^S(\underline{n})$, dependiente ahora del vector tamaño muestral, consistirá en la realización de la familia de experimentos

$$\{E(n_i)\}_{i=1,2,\dots,L} = \{ \{X_i(n_i)\}, \{ \theta_i \}, \{ p(x_i(n_i)/\theta_i) \} \}_{i=1,2,\dots,L}$$

realizados en sendos estratos. Notemos ya la diferencia entre este caso y el anterior consistente en que ahora se trabaja con L variables distintas frente a la variable única anterior. El vector tamaño muestral \underline{n} aparece en las expresiones que definen las densidades $p(z)$ y $p(\theta/z)$, de forma que también aparecerá en la expresión de la información esperada que sobre θ propor

ciona el experimento $E^S(\underline{n})$. Así pues, dicha información, tal como se ha definido con anterioridad, será función del vector tamaño muestral

$$\underline{n} = (n_1, n_2, \dots, n_L)$$

Análogamente, el valor esperado del experimento $E^S(\underline{n})$ se define como la diferencia entre su utilidad y costo esperados, esto es,

$$W(E^S(\underline{n})) = U(E^S(\underline{n})) - C(E^S(\underline{n}))$$

expresión que, es forzoso destacarlo nuevamente, es función de \underline{n} . Para la utilidad esperada del experimento $E^S(\underline{n})$ puede emplearse cualquiera de las funciones definidas en la sección 3.1. Lo realmente importante es que todas ellas son funciones de la información esperada $I^g\{E^S(\underline{n}), p(\underline{g})\}$ y, por tanto, del vector \underline{n} . Para la función de costo esperado puede emplearse cualquier función que se ajuste a los datos de partida de cada problema particular. Otra vez, la función de costo lineal aparece como la de mejor aplicación por su facilidad e idoneidad en muchas circunstancias. Dicha función tiene como expresión

$$C(E^S(\underline{n})) = \sum_{i=1}^L c_i \cdot n_i$$

donde c_i sigue siendo el costo de observación de una unidad del estrato i -ésimo. Por regla general se distingue el costo unitario de observación por estrato toda vez que la estratificación práctica de una población suele hacerse, entre otros criterios, atendiendo a consideraciones geográficas y de características demográficas. Es evidente, pues, que no puede ser el mismo costo el necesario para investigar una unidad informante en una capital en la que se dispone de una oficina de estadística, que el costo asociado a una unidad informante rural a la que haya que visitar. De ahí que quede justificado la existencia de costos unitarios de observación distintos según estratos.

Ya ha quedado indicado que la diferencia fundamental en el proceso de optimización del valor esperado del experimento $E^S(\underline{n})$ en los dos primeros casos estudiados radica en que, en este segundo, se determina una afijación óptima en cuanto que se obtienen los valores óptimos de los tamaños muestrales para cada uno de los estratos, quedando inmediatamente determinado el tamaño muestral total como suma de los tamaños muestrales parcia-

les anteriores. En contraposición está el proceso desarrollado en el primer caso, donde se determina el tamaño muestral total óptimo y luego, por la afijación previamente conocida, se determinan los tamaños de los estratos. En definitiva, el experimento $E^S(\underline{n})$ óptimo que presupone una estratificación de la población X y un desconocimiento previo de su afijación, realizado a fin de recabar información sobre la cantidad de interés θ , será el realizado con un vector tamaño muestral \underline{n} deducido como el vector no negativo que maximizando a $W(E^S(\underline{n}))$ sea el más cercano en valores enteros a la solución del sistema de L ecuaciones

$$\frac{\partial W(E^S(\underline{n}))}{\partial n_i} = 0 \quad i=1,2,\dots,L$$

sistema equivalente al

$$\frac{\partial U(E^S(\underline{n}))}{\partial n_i} = \frac{\partial C(E^S(\underline{n}))}{\partial n_i} \quad i=1,2,\dots,L$$

Según se utilice una u otra utilidad esperada y uno u otro costo esperado el sistema anterior tomará una forma determinada. Y así, empleando un costo lineal y una función de utilidad esperada $U_i(E^S(\underline{n}))$ ($i=1,2,3$) de las definidas en la sección 3.1 se obtendrán los siguientes tres sistemas de ecuaciones que permiten la obtención del vector óptimo \underline{n} :

$$g_1 \cdot \frac{\partial I^{\theta}\{E^S(\underline{n}), p(\underline{\theta})\}}{\partial n_i} = c_i \quad i=1,2,\dots,L$$

$$g_2 \cdot \frac{\partial I^{\theta}\{E^S(\underline{n}), p(\underline{\theta})\}}{\partial n_i} \cdot e^{-I^{\theta}\{E^S(\underline{n}), p(\underline{\theta})\}} = c_i \quad i=1,2,\dots,L$$

$$g_3 \cdot e^{-H(p(\theta))} \frac{\partial I^{\theta}\{E^S(\underline{n}), p(\underline{\theta})\}}{\partial n_i} \cdot e^{I^{\theta}\{E^S(\underline{n}), p(\underline{\theta})\}} = c_i \quad i=1,2,\dots,L$$

5.4 INFORMACION ESPERADA $I^{\vartheta} \{E^S(\underline{n}), p(\underline{\vartheta})\}$

Sabemos ya que cualquier proceso de estratificación lleva consigo la consideración de las L nuevas cantidades aleatorias ϑ_i cuya interpretación es idéntica a la de la cantidad de interés ϑ pero con actuación restringida a su estrato respectivo. Evidentemente, estas nuevas cantidades no son necesariamente independientes. Por otra parte existe una relación funcional que liga a los distintos parámetros ϑ_i ($i=1,2,\dots,L$) y ϑ entre sí, denotada por $\vartheta = f(\underline{\vartheta})$, siendo $\underline{\vartheta}$ el vector paramétrico de componentes ϑ_i . De ahí que el conocimiento exacto de los parámetros en los estratos permita un exacto conocimiento de la cantidad ϑ . Y análogamente, un buen conocimiento de las cantidades ϑ_i permite, a su vez, una buena estimación del parámetro global ϑ . En otro orden de ideas, en el diseño de una encuesta por muestreo probabilístico la estratificación no se realiza con la única finalidad de facilitar la obtención de la muestra. Un objetivo fundamental que suele presidir la decisión de qué estratos considerar - estratos de unidades análogas y lo más heterogéneos posible entre ellos - consiste en determinar de qué subpoblaciones deben estimarse también ciertas características. Así pues, y con carácter general, la consideración de un proceso de muestreo estratificado lleva implícitamente consigo una ampliación en el campo de la investigación, de forma que el investigador realiza un experimento a fin de recabar información no sólo de la cantidad ϑ sino también de todos los parámetros ϑ_i . Pero aún más, tal como se acaba de afirmar el conocimiento sobre el conjunto de las ϑ_i ($i=1,2,\dots,L$) proporciona información sobre la cantidad general ϑ . De ahí que en el próximo desarrollo y a diferencia de lo expuesto en la sección 5.3, el investigador centrará su atención sobre la información proporcionada por el experimento asociado a un proceso de muestreo estratificado sobre el valor del vector paramétrico $\underline{\vartheta} = (\vartheta_1, \vartheta_2, \dots, \vartheta_L)$, obteniendo la información sobre ϑ a partir de la anterior.

Inicialmente, el investigador especifica sus opiniones sobre $\underline{\vartheta}$ a través de la densidad $p(\underline{\vartheta})$. Por otra parte, la realización del experimento $E^S(\underline{n})$ supone la realización simultánea en sendos estratos de L experimentos de la familia

$$\{E(n_i)\}_{i=1,2,\dots,L} = \left\{ \{X_i(n_i)\}, \{\vartheta_i\}, \{p(x_i(n_i)/\vartheta_i)\} \right\}_{i=1,2,\dots,L}$$

El resultado del experimento $E(n_i)$ realizado en el estrato i -ésimo viene especificado por la muestra

$$z_i = (x_{i1}, x_{i2}, \dots, x_{in_i}) \quad x_{ij} \in X_i \quad i=1,2,\dots,L \quad j=1,2,\dots,n_i$$

de forma que el resultado global del experimento $E^S(\underline{n})$ está constituido por el conjunto de las L submuestras anteriores,

$$z(\underline{n}) = (z_1, z_2, \dots, z_L)$$

La notación anterior especifica la dependencia resultado muestral-vector tamaño muestral \underline{n} . Este vector tiene como componentes n_i los distintos tamaños de las muestras z_i , esto es,

$$\underline{n} = (n_1, n_2, \dots, n_L)$$

mientras que la suma de dichas componentes constituye el tamaño muestral total asociado al experimento $E^S(\underline{n})$. Resulta interesante destacar la diferencia entre el vector \underline{n} y el tamaño total n dado que más adelante se estudiará el comportamiento respecto a ambos de la información esperada definida más adelante.

Dado que las muestras son aleatorias y que los procesos de obtención de las mismas son independientes entre sí la densidad $p(z(\underline{n})/\underline{\theta})$ que expresa el comportamiento de la muestra global dependiente del vector paramétrico $\underline{\theta}$ viene dada por

$$p(z(\underline{n})/\underline{\theta}) = \prod_{k=1}^L \prod_{i=1}^{n_k} p(x_{ki}/\theta_k)$$

donde nuevamente puede apreciarse cómo influyen los tamaños parciales n_i en el cálculo de $p(z(\underline{n})/\underline{\theta})$.

Consideremos ahora el vector $\underline{n+1}$. Expresará a aquel vector tamaño muestral de componentes mayores en una unidad que las del vector \underline{n} , esto es,

$$\underline{n+1} = (n_1+1, n_2+1, \dots, n_L+1)$$

De ahí que

$$p(z(\underline{n+1})/\underline{\theta}) = \prod_{k=1}^L \prod_{i=1}^{n_k+1} p(x_{ik}/\theta_k)$$

existiendo la siguiente relación entre densidades

$$\begin{aligned} p(z(\underline{n+1})/\underline{\vartheta}) &= \prod_{k=1}^L \prod_{i=1}^{n_k} p(x_{ki}/\vartheta_k) \cdot \prod_{k=1}^L p(x_{kn_k+1}/\vartheta_k) = \\ &= p(z(\underline{n})/\underline{\vartheta}) \cdot \prod_{k=1}^L p(x_{kn_k+1}/\vartheta_k) \end{aligned}$$

Así pues, la consideración del vector $\underline{n+1}$ supondrá un incremento en el tamaño muestral total de L nuevas unidades.

Comparemos ahora los experimentos $E^S(\underline{n})$ y $E^S(\underline{n+1})$, consistente este último en obtener una sola unidad adicional de muestreo. Cuando la afijación de la estratificación es conocida a priori, la situación expresada por los vectores \underline{n} y $\underline{n+1}$ se traduce en considerar la existencia de un estrato i_0 -ésimo conocido en el que el experimento $E(n_{i_0})$ componente del $E^S(\underline{n})$ pasa a ser el $E(n_{i_0}+1)$ del respectivo experimento $E^S(\underline{n+1})$, permaneciendo iguales los experimentos componentes de los demás estratos. Así:

$$\begin{aligned} p(z(\underline{n+1})/\underline{\vartheta}) &= \prod_{k \neq i_0}^L \prod_{i=1}^{n_k} p(x_{ki}/\vartheta_k) \cdot \prod_{j=1}^{n_{i_0}+1} p(x_{i_0j}/\vartheta_{i_0}) = \\ &= p(x_{i_0 n_{i_0}+1}/\vartheta_{i_0}) \cdot p(z(\underline{n})/\underline{\vartheta}) \end{aligned}$$

Si la afijación no es conocida la situación es análoga a la anterior pero sin poder concretar el estrato al que pertenece la nueva unidad muestral. En este caso se dirá que existe un cierto estrato i -ésimo que verifica la anterior igualdad.

Conocidas las densidades $p(\underline{\vartheta})$ y $p(z(\underline{n})/\underline{\vartheta})$, la densidad de probabilidad que describe las opiniones posteriores sobre $\underline{\vartheta}$ que el investigador posee una vez que ha estudiado la información contenida en el resultado muestral $z(\underline{n})$ puede calcularse a partir de la condición de Bayes. Denotándola por $p(\underline{\vartheta}/z(\underline{n}))$, la cuantificación de la información esperada sobre $\underline{\vartheta}$ proporcionada por el experimento $E^S(\underline{n})$ se realiza a través de la formulación ya conocida de Lindley (1956) que admite la siguiente expresión a partir de la doble integral siguiente:

$$I_{\theta}^{\mathcal{S}}\{E^{\mathcal{S}}(\underline{n}), p(\underline{\theta})\} = \iint p(\underline{\theta}) \cdot p(z(n)/\underline{\theta}) \cdot \log \frac{p(\underline{\theta}/z(n))}{p(\underline{\theta})} \cdot d\underline{\theta} \cdot dz(n)$$

A pesar de la coincidencia formal de la doble integral anterior con la definida por Lindley (1956) para la información esperada proporcionada por un experimento, el distinto comportamiento para la obtención del resultado muestral cuando el proceso es estratificado respecto al seguido en la obtención de una muestra aleatoria simple obliga a la demostración de las análogas propiedades demostradas por Lindley (1956) para esta última información. En efecto, cuando en la población X no se realiza ninguna división en estratos, la función que describe el comportamiento de la muestra $x(n) = (x_1, x_2, \dots, x_n)$ obtenida tras la realización del experimento $E(n)$ y cuando es de interés la cantidad θ , viene dada por el producto simple:

$$p(x(n)/\theta) = \prod_{i=1}^n p(x_i/\theta)$$

Sin embargo, al dividir la población total X en estratos, el experimento $E^{\mathcal{S}}(\underline{n})$ a considerar consiste, según ya es conocido, en la simultánea realización en sendos estratos de L experimentos $E(n_i)$ independientes, realizados cada uno en el ámbito de aplicación de su estrato respectivo y referidos a una sola cantidad de interés θ_i . Por ello, en el proceso de estratificación no sólo se ve afectada la población X sino también el parámetro θ que pasa a describirse por un vector paramétrico $\underline{\theta}$, de forma que el comportamiento de cada una de las muestras parciales z_i se describirá por una densidad de probabilidad dependiente solamente de la i -ésima componente del vector $\underline{\theta}$, esto es,

$$p(z_i(n_i)/\underline{\theta}) = p(z_i(n_i)/\theta_i) = \prod_{j=1}^{n_i} p(x_{ij}/\theta_i)$$

Y teniendo en cuenta que, según se ha indicado, los procesos que permiten la obtención de las muestras z_i en sus estratos son independientes, la probabilidad que describe el comportamiento de la muestra total vendrá expresada a través del producto de las anteriores densidades, esto es, a través de un doble productorio donde cada una de las densidades que en él intervienen siguen dependiendo exclusivamente de una i -ésima componente del vector $\underline{\theta}$, esto es,

$$p(z(\underline{n})/\underline{\vartheta}) = \prod_{i=1}^L p(z_i(n_i)/\vartheta_i) = \prod_{i=1}^L \prod_{j=1}^{n_i} p(x_{ij}/\vartheta_i)$$

Considerando ahora dos vectores tamaño muestral determinados, \underline{n}^1 y \underline{n}^2 , tales que $\underline{n}^1 \leq \underline{n}^2$, esto es, $n_i^1 \leq n_i^2$ ($i=1,2,\dots,L$), se verifica la igualdad

$$p(z(\underline{n}^2)/\underline{\vartheta}) = \prod_{i=1}^L \prod_{j=1}^{n_i^2} p(x_{ij}/\vartheta_i) = \prod_{i=1}^L \prod_{j=1}^{n_i^1} p(x_{ij}/\vartheta_i) \cdot$$

$$\prod_{i=1}^L \prod_{j=1}^{n_i^2 - n_i^1} p(x_{ij}/\vartheta_i) = p(z(\underline{n}^1)/\underline{\vartheta}) \cdot \prod_{i=1}^L \prod_{j=1}^{n_i^2 - n_i^1} p(x_{ij}/\vartheta_i)$$

Por todas estas circunstancias que suponen un comportamiento para $z(\underline{n})$ distinto que el seguido por una muestra obtenida de una población sin estratos - y con la consideración adicional de una familia de experimentos definidos cada uno en su estrato respectivo, en contraposición a un experimento único - resulta necesario comprobar para $I^{\vartheta}\{E^S(\underline{n}), p(\underline{\vartheta})\}$ las propiedades demostradas por Lindley (1956) para una información esperada obtenida a través de un experimento no asociado a ninguna estratificación. Así podrá comprobarse que para el nuevo proceso definido el comportamiento de la información esperada anterior sigue verificando idénticas actitudes que en el modelo no estratificado.

Lindley (1956) demuestra para una información esperada sobre ϑ , $I^{\vartheta}\{E, p(\vartheta)\}$, proporcionada por el experimento E y con densidad inicial $p(\vartheta)$, que toma valores no negativos y que solamente se anula cuando no se modifican las opiniones iniciales tras investigar el resultado muestral obtenido. Este resultado es también generalizable a la información esperada $I^{\vartheta}\{E^S(\underline{n}), p(\underline{\vartheta})\}$ tal como queda recogido en el siguiente

TEOREMA 5.4.1

Para un experimento $E^S(\underline{n})$ y densidad inicial $p(\underline{\vartheta})$,

$$I^{\vartheta}\{E^S(\underline{n}), p(\underline{\vartheta})\} \geq 0$$

con igualdad si, y sólo si (sii), $p(\underline{\vartheta}/z(\underline{n}))$ es independiente del resultado muestral $z(\underline{n})$ excepto, a lo sumo, en un conjunto de valores de $\underline{\vartheta}$ de medida nula.

Demostración.- Los argumentos empleados por Lindley (1956) siguen siendo válidos para este caso. En efecto, por misma definición

$$I^{\vartheta} \{E^S(\underline{z}(n), p(\underline{\vartheta}))\} = \iint g(z(\underline{n}), \underline{\vartheta}) \cdot \log g(z(\underline{n}), \underline{\vartheta}) \cdot p(z(\underline{n})) \cdot p(\underline{\vartheta}) \cdot dz(\underline{n}) \cdot d\underline{\vartheta}$$

donde

$$g(z(\underline{n}), \underline{\vartheta}) = p(\underline{\vartheta}/z(\underline{n})) / p(\underline{\vartheta})$$

Dado que la función $x \cdot \log x$ es convexa, aplicando la desigualdad de Jensen

$$\iint g(z(\underline{n}), \underline{\vartheta}) \cdot \log g(z(\underline{n}), \underline{\vartheta}) \cdot p(z(\underline{n})) \cdot p(\underline{\vartheta}) \cdot dz(\underline{n}) \cdot d\underline{\vartheta} \geq$$

$$\iint g(z(\underline{n}), \underline{\vartheta}) \cdot p(z(\underline{n})) \cdot p(\underline{\vartheta}) \cdot dz(\underline{n}) \cdot d\underline{\vartheta} \cdot \log \iint g(z(\underline{n}), \underline{\vartheta}) \cdot p(z(\underline{n})) \cdot p(\underline{\vartheta}) \cdot dz(\underline{n}) \cdot d\underline{\vartheta}$$

con igualdad si $g(z(\underline{n}), \underline{\vartheta})$ es una constante salvo un conjunto de medida nula. Esto es, si $p(\underline{\vartheta}/z(\underline{n})) = p(\underline{\vartheta})$ salvo un conjunto de valores de $\underline{\vartheta}$ de medida nula. Además:

$$\begin{aligned} \iint g(z(\underline{n}), \underline{\vartheta}) \cdot p(z(\underline{n})) \cdot p(\underline{\vartheta}) \cdot dz(\underline{n}) \cdot d\underline{\vartheta} &= \iint p(\underline{\vartheta}/z(\underline{n})) \cdot p(z(\underline{n})) \cdot dz(\underline{n}) \cdot d\underline{\vartheta} = \\ &= \iint p(z(\underline{n}), \underline{\vartheta}) \cdot dz(\underline{n}) \cdot d\underline{\vartheta} = 1 \end{aligned}$$

de forma que el segundo miembro de la anterior desigualdad se anula al tomar logaritmos en la doble integral. Esto completa la demostración.

(c.q.d.)

El teorema indica que, bajo condición de que la densidad posterior de $\underline{\vartheta}$ varíe con $z(\underline{n})$, el experimento $E^S(\underline{z}(n))$ resulta, por término medio, informativo. Es evidente que si realizado $E^S(\underline{z}(n))$ la información contenida en su resultado $z(\underline{n})$ no hace modificar las opiniones iniciales representadas a través de $p(\underline{\vartheta})$, no existe ganancia de información por el hecho de realizar la experimentación.

Teniendo en cuenta la especial definición de $p(\underline{\vartheta}/z(\underline{n}))$ el resultado del anterior Teorema puede expresarse de una forma alternativa. En efecto,

$$p(\underline{\vartheta}/z(\underline{n})) = \frac{p(\vartheta_1, \vartheta_2, \dots, \vartheta_L) \cdot \prod_{i=1}^L p(z_i(n_i)/\vartheta_i)}{\int_{\Theta_1} \int_{\Theta_2} \dots \int_{\Theta_L} p(\vartheta_1, \vartheta_2, \dots, \vartheta_L) \prod_{i=1}^L p(z_i(n_i)/\vartheta_i) \cdot d\vartheta_1 \cdot d\vartheta_2 \cdot \dots \cdot d\vartheta_L}$$

donde $p(z_i(n_i)/\theta_i)$ es la densidad que en el estrato i -ésimo describe el comportamiento de la muestra $z_i(n_i)$ obtenida tras la realización del experimento $E(n_i)$ componente i -ésimo del $E^S(\underline{n})$. Entonces, verificándose en todos los estratos la condición

$$p(z_i(n_i)/\theta_i) = p(z_i(n_i)) \quad i=1,2,\dots,L$$

que es equivalente a

$$p(\theta_i/z_i(n_i)) = p(\theta_i) \quad i=1,2,\dots,L \quad \text{con } p(\theta_i) \text{ marginal } i\text{-ésima de } p(\underline{\theta})$$

se deduce que

$$\int \dots \int p(\theta_1, \dots, \theta_L) \cdot \prod_{i=1}^L p(z_i(n_i)/\theta_i) \cdot d\theta_1 \dots d\theta_L = \prod_{i=1}^L p(z_i(n_i))$$

con lo que finalmente $p(\underline{\theta}/z(\underline{n})) = p(\underline{\theta})$. E inversamente, siendo cierta esta última igualdad entre densidades de probabilidad se obtiene:

$$\begin{aligned} \prod_{i=1}^L p(z_i(n_i)/\theta_i) &= \int \dots \int p(\theta_1, \dots, \theta_L) \cdot \prod_{i=1}^L p(z_i(n_i)/\theta_i) \cdot d\theta_1 \dots d\theta_L = \\ &= p(z(\underline{n})) \end{aligned}$$

Por ello, el productorio anterior no depende de ninguna de las cantidades de interés θ_i . Pero por su misma definición como producto de densidades dependientes de una única cantidad de interés se deduce finalmente que todas las densidades $p(z_i(n_i)/\theta_i)$ no varían en los conjuntos \mathcal{C}_i salvo, tal vez, en subconjuntos de medida dominante nula. O lo que es lo mismo, que las densidades $p(\theta_i/z_i(n_i))$ no varían con el resultado muestral $z_i(n_i)$ obtenido como componente i -ésima del resultado muestral total $z(n)$. Esta discusión supone la demostración del siguiente

COROLARIO 5.4.1

Para un experimento $E^S(\underline{n})$ y una densidad inicial $p(\underline{\theta})$,

$$I^{\theta} \{E^S(\underline{n}), p(\underline{\theta})\} \geq 0$$

con igualdad sii en todos los estratos las densidades $p(\theta_i/z_i(n_i))$ son independientes de los resultados muestrales respectivos $z_i(n_i)$ salvo, a lo sumo, en subconjuntos de valores de θ_i de medida dominante nula, o lo que es lo mismo, cuando las densidades $p(z_i(n_i)/\theta_i)$ no

varían en Θ_i salvo en respectivos subconjuntos de medida nula.

Claramente el corolario indica que solamente cuando en todos los estratos no exista variación en las opiniones iniciales sobre su parámetro respectivo a pesar de realizar los experimentos $E(n_i)$ la información esperada sobre ϑ será nula. Bastará que en un solo estrato se modifiquen las opiniones iniciales para que el experimento $E^S(\underline{n})$ resulte, por término medio, informativo.

Lindley (1956) demuestra (como ya se indicó en las secciones 1.3 y 3.2) que la información esperada $I^{\vartheta}\{E, p(\vartheta)\}$ verificaba cierta aditividad muy interesante. Teniendo en cuenta que la información esperada definida ahora, $I^{\vartheta}\{E^S(\underline{n}), p(\vartheta)\}$, es función también de un experimento - si bien más complejo que en el caso simple - vamos a extenderla a nuestro caso. Para ello supongamos las situaciones expresadas por los experimentos $E^S(\underline{n})$ y $E^S(\underline{n+1})$ definidos a partir de sendas familias de experimentos componentes:

$$\{E(n_i)\}_{i=1,2,\dots,L} \quad \text{y} \quad \{E(n_i+1)\}_{i=1,2,\dots,L}$$

Así pues la diferencia entre ellos radica en que $E^S(\underline{n+1})$ obtiene una nueva unidad de muestreo en todos los estratos de la población X , aumentando su tamaño muestral total en L unidades, una por estrato. Definamos ahora el experimento $E^S(\underline{(n+1)}^a/z(\underline{n}))$ consistente en la familia de L experimentos realizados en sendos estratos

$$\begin{aligned} E^S(\underline{(n+1)}^a/z(\underline{n})) &= \{E(n_i+1)^a/z_i(n_i)\}_{i=1,2,\dots,L} = \\ &= \left\{ \{X_i\}, \{\Theta_i\}, \{p(x_{i n_i+1}/\vartheta_i, z_i(n_i))\} \right\}_{i=1,2,\dots,L} \end{aligned}$$

y que consiste en obtener en cada estrato la unidad (n_i+1) -ésima supuestas obtenidas y conocidas las muestras $z_i(n_i)$ ($i=1,2,\dots,L$) resultados de los experimentos $E(n_i)$ que definen $E^S(\underline{n})$. La información esperada proporcionada por el experimento $E^S(\underline{(n+1)}^a/z(\underline{n}))$ sobre ϑ una vez que se ha realizado $E^S(\underline{n})$ y obtenido $z(\underline{n})$ viene dada por la doble integral

$$\begin{aligned} I^{\vartheta}\{E^S(\underline{(n+1)}^a/z(\underline{n})), p(\vartheta/z(\underline{n}))\} &= \iint p(\vartheta/z(\underline{n})) \cdot p(z_{n+1}/z(\underline{n}), \vartheta) \cdot \\ &\cdot \log \frac{p(\vartheta/z_{n+1}, z(\underline{n}))}{p(\vartheta/z(\underline{n}))} \cdot d\vartheta \cdot dz_{n+1} \end{aligned}$$

donde z_{n+1} denota el resultado de dicho experimento, esto es,

$$z_{n+1} = (x_{1n_1+1}, x_{2n_2+1}, \dots, x_{Ln_L+1})$$

y $p(z_{n+1}/\underline{\vartheta}, z(\underline{n}))$ es la densidad que describe el comportamiento del resultado z_{n+1} dependiendo de $z(\underline{n})$ (resultado del experimento $E^S(\underline{n})$) y del vector paramétrico $\underline{\vartheta} \in \underline{\Theta}$. De ahí que siendo independientes las realizaciones de los experimentos $E((n_i+1)^a/z_i(n_i))_{i=1,2,\dots,L}$ en sus estratos respectivos

$$p(z_{n+1}/\underline{\vartheta}, z(\underline{n})) = \prod_{i=1}^L p(x_{in_i+1}/\vartheta_i, z_i(n_i))$$

Notemos, por otra parte, que el resultado $z(\underline{n+1})$ del experimento $E^S(\underline{n+1})$ puede expresarse a través del par de resultados $(z_{n+1}, z(\underline{n}))$ de los experimentos $E^S((n+1)^a/z(\underline{n}))$ y $E^S(\underline{n})$ respectivamente. De ahí que sean idénticas las igualdades entre las densidades siguientes:

$$p(\underline{\vartheta}/z_{n+1}, z(\underline{n})) = p(\underline{\vartheta}/z(\underline{n+1})) \quad \text{y} \quad p(z_{n+1}, z(\underline{n})/\underline{\vartheta}) = p(z(\underline{n+1})/\underline{\vartheta})$$

Con todos los elementos anteriores definiremos como información esperada sobre $\underline{\vartheta}$ proporcionada por el experimento $E((n+1)^a/z(\underline{n}))$ después de que se ha realizado $E^S(\underline{n})$ como el valor esperado sobre $z(\underline{n})$ de la información

$$I^{\underline{\vartheta}}\{E^S((n+1)^a/z(\underline{n})), p(\underline{\vartheta}/z(\underline{n}))\}$$

esto es, a través de la integral

$$I^{\underline{\vartheta}}\{E^S((n+1)^a/z(\underline{n})), p(\underline{\vartheta})\} = \int p(z(\underline{n})) \int p(\underline{\vartheta}/z(\underline{n})) \cdot p(z_{n+1}/z(\underline{n}), \underline{\vartheta}) \cdot \log \frac{p(\underline{\vartheta}/z_{n+1}, z(\underline{n}))}{p(\underline{\vartheta}/z(\underline{n}))} \cdot d\underline{\vartheta} \cdot dz_{n+1} \cdot dz(\underline{n})$$

Una demostración en la línea de la del teorema 5.4.1 establece que la información anterior es también no negativa, tomando como expresión para la función $g(z(\underline{n+1}), \underline{\vartheta})$ la

$$g(\underline{\vartheta}, z(\underline{n}), z_{n+1}) = p(\underline{\vartheta}/z_{n+1}, z(\underline{n})) / p(\underline{\vartheta}/z(\underline{n}))$$

Con la notación anteriormente introducida demostraremos el siguiente resultado que corresponde a la aditividad satisfecha por la información esperada sobre $\underline{\vartheta}$ proporcionada por el experimento $E^S(\underline{n})$:

TEOREMA 5.4.2

Dados los experimentos $E^S(\underline{n})$ y $E^S(\underline{n+1})$ y una densidad inicial denotada por $p(\underline{\vartheta})$, se verifica el siguiente criterio de aditividad para la información esperada sobre $\underline{\vartheta}$:

$$I^{\underline{\vartheta}}\{E^S(\underline{n+1}), p(\underline{\vartheta})\} = I^{\underline{\vartheta}}\{E^S(\underline{n}), p(\underline{\vartheta})\} + I^{\underline{\vartheta}}\{E^S(\underline{n+1})^a / E^S(\underline{n}), p(\underline{\vartheta})\}$$

Demostración.- Nuevamente la demostración desarrollada por Lindley (1956) es válida en este caso. De sus respectivas definiciones:

$$\begin{aligned} I^{\underline{\vartheta}}\{E^S(\underline{n}), p(\underline{\vartheta})\} &= \iint p(\underline{\vartheta}) \cdot p(z(\underline{n})/\underline{\vartheta}) \cdot \log \frac{p(\underline{\vartheta}/z(\underline{n}))}{p(\underline{\vartheta})} \cdot d\underline{\vartheta} \cdot dz(\underline{n}) = \\ &= \iint p(\underline{\vartheta}) \cdot p(z(\underline{n+1})/\underline{\vartheta}) \cdot \log \frac{p(\underline{\vartheta}/z(\underline{n}))}{p(\underline{\vartheta})} \cdot d\underline{\vartheta} \cdot dz(\underline{n+1}) \end{aligned}$$

$$\begin{aligned} I^{\underline{\vartheta}}\{E^S(\underline{n+1})^a / E^S(\underline{n}), p(\underline{\vartheta})\} &= \int p(z(\underline{n})) \iint p(\underline{\vartheta}/z(\underline{n})) \cdot p(z_{\underline{n+1}}/\underline{\vartheta}, z(\underline{n})) \cdot \\ &\cdot \log \frac{p(\underline{\vartheta}/z_{\underline{n+1}}, z(\underline{n}))}{p(\underline{\vartheta}/z(\underline{n}))} \cdot d\underline{\vartheta} \cdot dz_{\underline{n+1}} \cdot dz(\underline{n}) = \\ &= \iint p(\underline{\vartheta}) \cdot p(z(\underline{n+1})/\underline{\vartheta}) \cdot \log \frac{p(\underline{\vartheta}/z(\underline{n+1}))}{p(\underline{\vartheta}/z(\underline{n}))} \cdot d\underline{\vartheta} \cdot dz(\underline{n+1}) \end{aligned}$$

Sumando ambas informaciones resulta:

$$\begin{aligned} I^{\underline{\vartheta}}\{E^S(\underline{n}), p(\underline{\vartheta})\} + I^{\underline{\vartheta}}\{E^S(\underline{n+1})^a / E^S(\underline{n}), p(\underline{\vartheta})\} &= \\ = \iint p(\underline{\vartheta}) \cdot p(z(\underline{n+1})/\underline{\vartheta}) \cdot \log \frac{p(\underline{\vartheta}/z(\underline{n+1}))}{p(\underline{\vartheta})} \cdot d\underline{\vartheta} \cdot dz(\underline{n+1}) &= I^{\underline{\vartheta}}\{E^S(\underline{n+1}), p(\underline{\vartheta})\} \\ &\quad \text{(c.q.d.)} \end{aligned}$$

Obviamente que este resultado puede generalizarse a un número finito de experimentos con un espacio paramétrico común $\underline{\vartheta}$. Esta generalización conduciría a una expresión de la forma:

$$I^{\underline{\vartheta}}\{E^S(\underline{n+k}), p(\underline{\vartheta})\} = I^{\underline{\vartheta}}\{E^S(\underline{n}), p(\underline{\vartheta})\} + \sum_{i=1}^k I^{\underline{\vartheta}}\{E^S(\underline{n+i})^a / E^S(\underline{n+i-1}), p(\underline{\vartheta})\}$$

donde

$$E^S(\underline{n} + \underline{p}) = \left\{ E(n_i + p) \right\}_{i=1,2,\dots,L}$$

proporcionando así una regla necesaria para la adición de grupos de información provenientes de diferentes caminos. Como inmediata e importancia consecuencia del teorema 5.4.2 es el siguiente

COROLARIO 5.4.2

Siendo $z(\underline{n})$ suficiente para $z(\underline{n+1}) = (z_{n+1}, z(\underline{n}))$ con respecto a $\underline{\vartheta}$, en el sentido de que $p(\underline{\vartheta}/z(\underline{n+1})) = p(\underline{\vartheta}/z(\underline{n}))$, entonces

$$I^{\underline{\vartheta}}\{E^S(\underline{n+1}), p(\underline{\vartheta})\} = I^{\underline{\vartheta}}\{E^S(\underline{n}), p(\underline{\vartheta})\}$$

Demostración.- Tal como se ha definido $I^{\underline{\vartheta}}\{E^S(\underline{n+1})/E^S(\underline{n}), p(\underline{\vartheta})\}$, si $p(\underline{\vartheta}/z(\underline{n+1})) = p(\underline{\vartheta}/z(\underline{n}))$, dicha información es nula por el teorema 5.4.1. El resultado se sigue, pues, del teorema 5.4.2 .

(c.q.d.)

El corolario 5.4.2 establece que no existe pérdida de información si ésta se ciñe a la observación de un estadístico suficiente con respecto a la cantidad de interés. E inversamente, si un estadístico es insuficiente en el sentido descrito existe una pérdida de información al ser $I^{\underline{\vartheta}}\{E^S(\underline{n+1})/E^S(\underline{n}), p(\underline{\vartheta})\}$ estrictamente positiva. Por otra parte, la interpretación del corolario anterior es clara: no vale la pena realizar el experimento $E^S(\underline{n+1})$ (recordemos que el objetivo de la realización del mismo es la de recabar información sobre la cantidad de interés) cuando las opiniones sobre $\underline{\vartheta}$ no se modifican respecto las iniciales proporcionadas por $E^S(\underline{n})$. Notemos, finalmente, que la igualdad entre las densidades $p(\underline{\vartheta}/z(\underline{n}))$ y $p(\underline{\vartheta}/z(\underline{n+1}))$ se traduce en la condición

$$p(z(\underline{n+1})) = p(z(\underline{n})) \cdot \prod_{i=1}^L p(x_{in_i+1}/\theta_i)$$

lo que implica que las densidades $p(x_{in_i+1}/\theta_i)$ ($i=1,2,\dots,L$) no varíen en los conjuntos \mathcal{G}_i , salvo subconjuntos i de medida nula.

Prosigamos estudiando el comportamiento de $I^{\underline{\vartheta}}\{E^S(\underline{n}), p(\underline{\vartheta})\}$ como función del vector tamaño muestral \underline{n} . Existe una extendida creencia sobre el decrecimiento de la información marginal de las sucesivas observaciones independientes y equidistribuidas, opinión expresada matemáticamente

por Lindley (1956) para una información esperada de la forma $I^{\vartheta}\{E(n), p(\vartheta)\}$ al demostrar que se trata de una función creciente y cóncava de n . El próximo teorema no es sino la confirmación del resultado anterior para la información esperada sobre ϑ proporcionada por el experimento $E^S(\underline{n})$:

TEOREMA 5.4.3

Para un experimento $E^S(\underline{n}) = \{E(n_i)\}_{i=1,2,\dots,L}$ y densidad inicial $p(\vartheta)$ que describe las opiniones iniciales del investigador sobre el vector paramétrico ϑ , $I^{\vartheta}\{E^S(\underline{n}), p(\vartheta)\}$ es una función creciente y cóncava del vector \underline{n} .

Demostración.- Por el teorema 5.4.2 es conocido que

$$I^{\vartheta}\{E^S(\underline{n+1}), p(\vartheta)\} - I^{\vartheta}\{E^S(\underline{n}), p(\vartheta)\} = I^{\vartheta}\{E^S(\underline{n+1})^A / E^S(\underline{n}), p(\vartheta)\} > 0$$

y, por tanto, $I^{\vartheta}\{E^S(\underline{n}), p(\vartheta)\}$ es una función creciente del vector \underline{n} .

Para probar su concavidad es suficiente demostrar que la expresión

$$\Delta = I^{\vartheta}\{E^S(\underline{n}), p(\vartheta)\} - I^{\vartheta}\{E^S(\underline{n-1}), p(\vartheta)\} - \left\{ I^{\vartheta}\{E^S(\underline{n+1}), p(\vartheta)\} - I^{\vartheta}\{E^S(\underline{n}), p(\vartheta)\} \right\}$$

es positiva. Utilizando la igualdad expresada en el teorema 5.4.2

$$\begin{aligned} \Delta &= I^{\vartheta}\{E^S(\underline{n})^A / E^S(\underline{n-1}), p(\vartheta)\} - I^{\vartheta}\{E^S(\underline{n+1})^A / E^S(\underline{n}), p(\vartheta)\} = \\ &= \iint p(\vartheta) \cdot p(z(\underline{n})/\vartheta) \cdot \log \frac{p(\vartheta/z(\underline{n}))}{p(\vartheta/z(\underline{n-1}))} \cdot d\vartheta \cdot dz(\underline{n}) - \\ &- \iint p(\vartheta) \cdot p(z(\underline{n+1})/\vartheta) \cdot \log \frac{p(\vartheta/z(\underline{n+1}))}{p(\vartheta/z(\underline{n}))} \cdot d\vartheta \cdot dz(\underline{n+1}) = \\ &= \iint p(\vartheta) \cdot p(z(\underline{n+1})/\vartheta) \cdot \log \frac{p(\vartheta/z(\underline{n}))^2}{p(\vartheta/z(\underline{n-1})) \cdot p(\vartheta/z(\underline{n+1}))} \cdot d\vartheta \cdot dz(\underline{n+1}) \end{aligned}$$

Estudiamos la expresión bajo el logaritmo.

$$\frac{p(\vartheta/z(\underline{n}))^2}{p(\vartheta/z(\underline{n-1})) \cdot p(\vartheta/z(\underline{n+1}))} = \frac{p(z(\underline{n})/\vartheta)^2}{p(z(\underline{n-1})/\vartheta) \cdot p(z(\underline{n+1})/\vartheta)} \cdot \frac{p(z(\underline{n-1})) \cdot p(z(\underline{n+1}))}{p(z(\underline{n}))^2}$$

Así pues,

$$\Delta = \iint p(\underline{\vartheta}) \cdot p(z(n+1)/\underline{\vartheta}) \cdot \log \frac{p(z(n)/\underline{\vartheta})^2}{p(z(n+1)/\underline{\vartheta}) \cdot p(z(n-1)/\underline{\vartheta})} \cdot d\underline{\vartheta} \cdot dz(n+1) +$$

$$+ \iint p(\underline{\vartheta}) \cdot p(z(n+1)/\underline{\vartheta}) \cdot \log \frac{p(z(n-1)) \cdot p(z(n+1))}{p(z(n))^2} \cdot d\underline{\vartheta} \cdot dz(n+1)$$

Estudieemos la primera de las dos integrales anteriores:

$$\frac{p(z(n)/\underline{\vartheta})^2}{p(z(n-1)/\underline{\vartheta}) \cdot p(z(n+1)/\underline{\vartheta})} = \frac{\left\{ \prod_{i=1}^L \prod_{j=1}^{n_i} p(x_{ij}/\theta_i) \right\}^2}{\prod_{i=1}^L \prod_{j=1}^{n_i-1} p(x_{ij}/\theta_i) \prod_{i=1}^L \prod_{j=1}^{n_i+1} p(x_{ij}/\theta_i)} =$$

$$= \prod_{i=1}^L \frac{p(x_{in_i}/\theta_i)}{p(x_{in_i+1}/\theta_i)}$$

Dado que las cantidades aleatorias x_{in_i} y x_{in_i+1} están equidistribuidas la primera integral se reduce a:

$$\sum_{i=1}^L \iint p(\underline{\vartheta}) \cdot p(z(n+1)/\underline{\vartheta}) \cdot \log \frac{p(x_{in_i}/\theta_i)}{p(x_{in_i+1}/\theta_i)} \cdot d\underline{\vartheta} \cdot dz(n+1) = 0$$

Para resolver la segunda integral hay que tener en cuenta que:

$$p(z(n)/\underline{\vartheta}) = p(z_n, z(n-1)/\underline{\vartheta}) = p(z_n/\underline{\vartheta}) \cdot p(z(n-1)/\underline{\vartheta}) = p(z(n-1)/\underline{\vartheta}) \cdot \prod_{i=1}^L p(x_{in_i}/\theta_i)$$

siendo por tanto

$$p(z(n)) = p(z_n, z(n-1))$$

Así pues,

$$\iint p(\underline{\vartheta}) \cdot p(z(n+1)/\underline{\vartheta}) \cdot \log \frac{p(z(n-1)) \cdot p(z(n+1))}{p(z(n))^2} \cdot d\underline{\vartheta} \cdot dz(n+1) =$$

$$= \int p(z(n+1)) \cdot \log \left(\frac{p(z(n-1))}{p(z_n, z(n-1))} \cdot \frac{p(z_{n+1}, z(n))}{p(z(n))} \right) \cdot dz(n+1) =$$

$$= \int p(z(n+1)) \cdot \log \frac{p(z_{n+1}/z(n))}{p(z_n/z(n-1))} \cdot dz(n+1) =$$

$$= \iint p(z(n)) \cdot p(z_{n+1}/z(n)) \cdot \log \frac{p(z_{n+1}/z(n))}{p(z_{n+1}/z(n-1))} \cdot dz_{n+1} \cdot dz(n)$$

toda vez que $p(z/z(n-1)) = p(z_{n+1}/z(n-1))$ al estar z_n y z_{n+1} equidistribuidas. La tesis del teorema queda demostrada al observar que la última integral es la información esperada que sobre z_{n+1} proporciona z_n , en el sentido descrito por el teorema 5.4.2. Y dicha información esperada, por el teorema 5.4.1, es positiva.

(c.q.d.)

Lindley (1956) demuestra para una información esperada de la forma $I^{\theta} \{E, p(\theta)\}$ que se trata de una funcional cóncava de la densidad inicial $p(\theta)$. También este resultado es extensible a la información esperada sobre $\underline{\theta}$ proporcionada por $E^S(\underline{n})$ como funcional de $p(\underline{\theta})$.

TEOREMA 5.4.4

$I^{\theta} \{E^S(\underline{n}), p(\underline{\theta})\}$ es una funcional cóncava de $p(\underline{\theta})$

Demostración.- Sigamos el razonamiento de Lindley (1956). Habrá que demostrar que siendo $p_1(\underline{\theta})$ y $p_2(\underline{\theta})$ dos densidades de probabilidad iniciales para $\underline{\theta}$ y $\lambda \in [0, 1]$, entonces

$$I^{\theta} \{E^S(\underline{n}), p(\underline{\theta})\} - \lambda I^{\theta} \{E^S(\underline{n}), p_1(\underline{\theta})\} - (1-\lambda) \cdot I^{\theta} \{E^S(\underline{n}), p_2(\underline{\theta})\} > 0$$

donde $p(\underline{\theta}) = \lambda p_1(\underline{\theta}) + (1-\lambda) \cdot p_2(\underline{\theta})$ es efectivamente aquella densidad de probabilidad que para un cierto $\underline{\theta}_0$ concreto verifica:

$$p(\underline{\theta}_0) = \lambda p_1(\underline{\theta}_0) + (1-\lambda) \cdot p_2(\underline{\theta}_0)$$

El primer miembro de la desigualdad anterior puede escribirse como

$$\begin{aligned} & \iint p(z(n)/\underline{\theta}) \cdot (\lambda p_1(\underline{\theta}) + (1-\lambda) \cdot p_2(\underline{\theta})) \cdot \log \frac{p(z(n)/\underline{\theta})}{p(z(n))} \cdot d\underline{\theta} \cdot dz(n) - \\ & - \lambda \iint p(z(n)/\underline{\theta}) \cdot p_1(\underline{\theta}) \cdot \log \frac{p(z(n)/\underline{\theta})}{p_1(z(n))} \cdot d\underline{\theta} \cdot dz(n) - \\ & - (1-\lambda) \cdot \iint p(z(n)/\underline{\theta}) \cdot p_2(\underline{\theta}) \cdot \log \frac{p(z(n)/\underline{\theta})}{p_2(z(n))} \cdot d\underline{\theta} \cdot dz(n) \end{aligned}$$

donde

$$p_i(z(\underline{n})) = \int p(z(\underline{n})/\underline{\vartheta}) \cdot p_i(\underline{\vartheta}) \cdot d\underline{\vartheta} \quad (i=1,2) \quad y$$

$$p(z(\underline{n})) = \lambda p_1(z(\underline{n})) + (1-\lambda) \cdot p_2(z(\underline{n}))$$

De ahí que dicho primer miembro pueda expresarse como

$$\lambda \iint p(z(\underline{n})/\underline{\vartheta}) \cdot p_1(\underline{\vartheta}) \cdot \log \frac{p_1(z(\underline{n}))}{p(z(\underline{n}))} \cdot dz(\underline{n}) \cdot d\underline{\vartheta} +$$

$$+ (1-\lambda) \cdot \iint p(z(\underline{n})/\underline{\vartheta}) \cdot p_2(\underline{\vartheta}) \cdot \log \frac{p_2(z(\underline{n}))}{p(z(\underline{n}))} \cdot dz(\underline{n}) \cdot d\underline{\vartheta}$$

y ambas integrales son positivas siguiendo idéntico razonamiento al empleado en el teorema 5.4.1 con

$$g_i(z(\underline{n}), \underline{\vartheta}) = p_i(z(\underline{n})/\underline{\vartheta}) / p(z(\underline{n}))$$

(c.q.d.)

Del anterior conjunto de teoremas queda establecida la coincidencia entre las propiedades demostradas por Lindley (1956) para la información $I^\vartheta\{E, p(\vartheta)\}$ y las satisfechas por la información esperada asociada a un proceso de muestreo estratificado, $I^\vartheta\{E^S(\underline{n}), p(\underline{\vartheta})\}$, considerada como función del vector tamaño muestral y como funcional, bien de la densidad inicial $p(\underline{\vartheta})$, bien del experimento $E^S(\underline{n})$. Por ello, recordando la definición dada por Bernardo (1978) para la información esperada útil sobre una nueva cantidad de interés ψ relacionada con la primitiva ϑ a través de la transformación $\psi = \psi(\vartheta)$, puede considerarse a la información $I^\vartheta\{E^S(\underline{n}), p(\underline{\vartheta})\}$ definida en la sección 5.3, como la correspondiente información esperada útil sobre $\vartheta = f(\underline{\vartheta})$ proporcionada por el experimento $E^S(\underline{n})$. De ahí que los resultados obtenidos por Bernardo (1978) sean de aplicación a este caso, como puede comprobarse empleando idénticas técnicas de demostración utilizadas por dicho autor. En particular, puede afirmarse que $I^\vartheta\{E^S(\underline{n}), p(\underline{\vartheta})\}$ es una función no negativa que solamente se anula cuando $p(\vartheta/z(\underline{n}))$ es independiente del resultado $z(\underline{n})$ del experimento $E^S(\underline{n})$ salvo, a lo sumo, en un conjunto de valores de ϑ de medida dominante nula, resultado que se corresponde con el expresado por el teorema 5.4.1 para $I^\vartheta\{E^S(\underline{n}), p(\underline{\vartheta})\}$. También $I^\vartheta\{E^S(\underline{n}), p(\underline{\vartheta})\}$ verifica un criterio de aditividad en la línea del demostrado para la información esperada sobre $\underline{\vartheta}$ en el teorema 5.4.2,

al verificar que $I^{\vartheta} \{E^S(\underline{n}+1), p(\underline{\vartheta})\} = I^{\vartheta} \{E^S(\underline{n}), p(\underline{\vartheta})\} + I^{\vartheta} \{E^S(\underline{n}+1)^a / E^S(\underline{n}), p(\underline{\vartheta})\}$. Y considerada como función del vector tamaño muestral \underline{n} , la información esperada $I^{\vartheta} \{E^S(\underline{n}), p(\underline{\vartheta})\}$ es creciente y cóncava, en correspondencia al resultado expresado por el teorema 5.4.3 para $I^{\vartheta} \{E^S(\underline{n}), p(\underline{\vartheta})\}$. Sin embargo, lo realmente interesante es el resultado encontrado por Bernardo (1978) en el que se comparan las informaciones útil y total al demostrar que la información esperada útil no es mayor que la información esperada total. Dicho resultado es generalizable a la situación introducida por el experimento $E^S(\underline{n})$, toda vez que se trata de una propiedad en la que no interviene la forma específica de obtención de la muestra $z(\underline{n})$ de una población estratificada. Por otra parte, teniendo en cuenta que la información esperada sobre ϑ proporcionada por $E^S(\underline{n})$, con $\vartheta = f(\underline{\vartheta})$, puede definirse también a través de la expresión

$$I^{\vartheta} \{E^S(\underline{n}), p(\underline{\vartheta})\} = \iint p(z(\underline{n})/\underline{\vartheta}) \cdot p(\underline{\vartheta}) \cdot \log \frac{p_{\vartheta}(f(\underline{\vartheta})/z(\underline{n}))}{p_{\vartheta}(f(\underline{\vartheta}))} \cdot d\underline{\vartheta} \cdot dz(\underline{n})$$

es evidente que la información $I^{\vartheta} \{E^S(\underline{n}), p(\underline{\vartheta})\}$ no varía ante transformaciones biyectivas del parámetro de interés $\underline{\vartheta}$. Demostremos, pues, el siguiente

TEOREMA 5.4.5

Para un experimento $E^S(\underline{n}) = \{E(n_i)\}_{i=1,2,\dots,L}$, densidad inicial $p(\underline{\vartheta})$ y función $\vartheta = f(\underline{\vartheta})$,

$$I^{\vartheta} \{E^S(\underline{n}), p(\underline{\vartheta})\} \geq I^{\vartheta} \{E^S(\underline{n}), p(\underline{\vartheta})\}$$

Demostración.— El argumento desarrollado por Bernardo (1978) es válido en nuestro caso. En efecto, siendo la transformación $\vartheta = f(\underline{\vartheta})$ no necesariamente biyectiva (y, en particular, no lo es la combinación lineal de la que hablábamos en la sección 5.1) puede considerarse una transformación uno-uno $\eta = \{\vartheta, \omega\}$ del parámetro $\underline{\vartheta}$. Debido a la invarianza de la información esperada bajo transformaciones biyectivas del vector paramétrico, resulta:

$$I^{\vartheta} \{E^S(\underline{n}), p(\underline{\vartheta})\} = I^{\eta} \{E^S(\underline{n}), p(\eta)\}$$

de donde

$$I^{\vartheta} \{E^S(\underline{n}), p(\underline{\vartheta})\} - I^{\vartheta} \{E^S(\underline{n}), p(\underline{\vartheta})\} = I^{\eta} \{E^S(\underline{n}), p(\eta)\} -$$

$$\begin{aligned}
- I^{\theta} \{ E^{\theta}(\underline{z}), p(\gamma) \} &= \iint p(\gamma) \cdot p(z(\underline{n})/\gamma) \cdot \log \frac{p(\gamma/z(\underline{n}))}{p(\gamma)} \cdot d\gamma \cdot dz(\underline{n}) - \\
- \iint p(\vartheta) \cdot p(z(\underline{n})/\vartheta) \cdot \log \frac{p(\vartheta/z(\underline{n}))}{p(\vartheta)} \cdot d\vartheta \cdot dz(\underline{n}) &= - \int p(\gamma) \cdot \log p(\gamma) \cdot d\gamma + \\
+ \iint p(\gamma) \cdot p(z(\underline{n})/\gamma) \cdot \log p(\gamma/z(\underline{n})) \cdot d\gamma \cdot dz(\underline{n}) &+ \int p(\vartheta) \cdot \log p(\vartheta) \cdot d\vartheta - \\
- \iint p(\vartheta) \cdot p(z(\underline{n})/\vartheta) \cdot \log p(\vartheta/z(\underline{n})) \cdot d\vartheta \cdot dz(\underline{n}) &
\end{aligned}$$

Por otra parte es evidente que

$$\begin{aligned}
\int p(\gamma) \cdot \log p(\gamma) \cdot d\gamma &= \int p(\vartheta) \cdot \log p(\vartheta) \cdot d\vartheta + \\
+ \iint p(\vartheta) \cdot p(\omega/\vartheta) \cdot \log p(\omega/\vartheta) \cdot d\omega \cdot d\vartheta &
\end{aligned}$$

De ahí que sustituyendo

$$\begin{aligned}
I^{\theta} \{ E^{\theta}(\underline{z}), p(\vartheta) \} - I^{\vartheta} \{ E^{\vartheta}(\underline{z}), p(\vartheta) \} &= - \iint p(\vartheta) \cdot p(\omega/\vartheta) \cdot \log p(\omega/\vartheta) \cdot \\
\cdot d\omega \cdot d\vartheta + \iint p(\gamma) \cdot p(z(\underline{n})/\gamma) \cdot \log p(\gamma/z(\underline{n})) \cdot d\gamma \cdot dz(\underline{n}) &- \\
- \iint p(\vartheta) \cdot p(z(\underline{n})/\vartheta) \cdot \log p(\vartheta/z(\underline{n})) \cdot d\vartheta \cdot dz(\underline{n}) &= \\
= - \iiint p(\vartheta, \omega, z(\underline{n})) \cdot \log p(\omega/\vartheta) \cdot d\omega \cdot d\vartheta \cdot dz(\underline{n}) &+ \\
+ \iiint p(\vartheta, \omega, z(\underline{n})) \cdot \log p(\vartheta, \omega/z(\underline{n})) \cdot d\vartheta \cdot d\omega \cdot dz(\underline{n}) &- \\
- \iiint p(\vartheta, \omega, z(\underline{n})) \cdot \log p(\vartheta/z(\underline{n})) \cdot d\vartheta \cdot d\omega \cdot dz(\underline{n}) &= \\
= \iiint p(\vartheta, \omega, z(\underline{n})) \cdot \log \frac{p(\vartheta, \omega/z(\underline{n}))}{p(\omega/\vartheta) \cdot p(\vartheta/z(\underline{n}))} \cdot d\vartheta \cdot d\omega \cdot dz(\underline{n}) &= \\
= \int p(\vartheta) \iint p(\omega/\vartheta) \cdot p(z(\underline{n})/\omega, \vartheta) \cdot \log \frac{p(\omega/\vartheta, z(\underline{n}))}{p(\omega/\vartheta)} \cdot d\omega \cdot dz(\underline{n}) \cdot d\vartheta &
\end{aligned}$$

que no es sino la información esperada sobre ω proporcionada por $E^S(\underline{n})$ una vez que θ es conocido. Y dicha información es positiva por un argumento similar al usado para demostrar el Teorema 5.4.1, con

$$g(\omega, z(\underline{n})) = p(\omega/\theta, z(\underline{n})) / p(\omega/\theta)$$

(c.q.d.)

La desigualdad del anterior teorema pone de manifiesto un importante hecho: diseñado un proceso de muestreo estratificado la información esperada sobre el vector paramétrico $\underline{\theta}$ que proporciona es mayor que la información esperada sobre la cantidad de interés $\vartheta = f(\underline{\vartheta})$. Dado que las cantidades ϑ_i nacen de la misma estratificación de la población X como un conjunto de parámetros que admiten idéntica interpretación que ϑ pero restringidos al ámbito de aplicación de su respectivo estrato X_i , de forma que al mismo tiempo el conjunto de las ϑ_i ($i=1,2,\dots,L$) permite obtener ϑ a través de la relación funcional que liga a todos entre sí, resulta que no es interesante realizar una estratificación de la población X y realizar el subsiguiente experimento $E^S(\underline{n})$ con el único objetivo de acceder a la información sobre la única cantidad ϑ . Puede, por tanto, apreciarse la riqueza que una adecuada estratificación introduce en el diseño de un experimento óptimo.

Hasta ahora se ha estudiado el comportamiento de $I^{\vartheta}\{E^S(\underline{n}), p(\underline{\vartheta})\}$ como función del vector tamaño muestral \underline{n} . Sin embargo, también resulta interesante estudiar propiedades más locales y conocer el comportamiento seguido por dicha información como función de la componente i -ésima de \underline{n} , esto es, del tamaño muestral de la submuestra extraída en X_i , así como respecto al tamaño muestral total n definido como suma de las componentes n_i del vector \underline{n} . Tanto las características de concavidad y crecimiento, así como la especial aditividad satisfechas por la información $I^{\vartheta}\{E^S(\underline{n}), p(\underline{\vartheta})\}$ como función de \underline{n} siguen cumpliéndose al estudiar su comportamiento como función de n_i . Y para n sigue verificándose dichas propiedades de una forma "escalonada". Estudiemos con detalle cada una de estas características.

Para estudiar $I^{\vartheta}\{E^S(\underline{n}), p(\underline{\vartheta})\}$ como función de una componente cualquiera n_i habrá que considerar el experimento $E^S(\underline{n}+1)$ ya definido con anterioridad. Comparándolo con el $E^S(\underline{n})$ resulta que los expe-

rimentos que definen $E^S(\underline{n} + 1)$ coinciden con los respectivos experimentos que definen a $E^S(\underline{n})$ en todos los estratos salvo en uno, el i_0 -ésimo, por ejemplo (estrato conocido cuando la afijación de la estratificación está determinada a priori y desconocido en caso contrario), en el que se obtiene una unidad de muestreo adicional. Así pues, al escribir $E^S(\underline{n} + 1)$ indicaremos que existe un estrato i_0 -ésimo en el que está definido el experimento $E(n_{i_0} + 1)$ frente al $E(n_{i_0})$ que es el respectivo experimento componente de $E^S(\underline{n})$. También se indicó la igualdad entre densidades

$$p(z(\underline{n} + 1)/\vartheta) = p(x_{i_0 n_{i_0} + 1} / \theta_{i_0}) \cdot p(z(\underline{n})/\vartheta)$$

siendo $z(\underline{n} + 1)$ el resultado de $E^S(\underline{n} + 1)$. Definamos ahora el experimento $E(n_{i_0} + 1)^*/z(\underline{n})$ consistente en obtener la unidad $(n_{i_0} + 1)$ -ésima en el estrato X_{i_0} supuestas obtenidas y conocidas las demás i_0 muestras $z_i(n_i)$ ($i=1, 2, \dots, L$) resultados de los experimentos $E(n_i)$ que definen a $E^S(\underline{n})$. Tendremos, pues,

$$E(n_{i_0} + 1)^*/z(\underline{n}) = \left\{ x_{i_0}, \theta_{i_0}, p(x_{i_0 n_{i_0} + 1} / \theta_{i_0}, z_i(n_i)) \right\}$$

La información esperada proporcionada por $E(n_{i_0} + 1)^*/z(\underline{n})$ sobre ϑ (recordemos que la unidad de muestreo $x_{i_0 n_{i_0} + 1}$ contiene información sobre to-

das las θ_{i_0} al no ser éstas independientes) una vez que se ha realizado $E^S(\underline{n})$ y obtenido $z(\underline{n})$ viene dada por la doble integral

$$\begin{aligned} I^{\vartheta} \{ E(n_{i_0} + 1)^*/z(\underline{n}), p(\vartheta/z(\underline{n})) \} &= \\ &= \iint p(x_{i_0 n_{i_0} + 1}, \vartheta/z(\underline{n})) \cdot \log \frac{p(\vartheta/z(\underline{n}), x_{i_0 n_{i_0} + 1})}{p(\vartheta/z(\underline{n}))} \cdot d\vartheta \cdot dx_{i_0 n_{i_0} + 1} = \\ &= \iint p(\vartheta/z(\underline{n})) \cdot p(x_{i_0 n_{i_0} + 1}/\vartheta, z(\underline{n})) \cdot \log \frac{p(\vartheta/z(\underline{n} + 1))}{p(\vartheta/z(\underline{n}))} \cdot d\vartheta \cdot dx_{i_0 n_{i_0} + 1} \end{aligned}$$

Notemos que al ser $z(\underline{n} + 1) = (x_{i_0 n_{i_0} + 1}, z(\underline{n}))$ las siguientes igualdades entre densidades

$$p(\vartheta/z(\underline{n} + 1)) = p(\vartheta/x_{i_0 n_{i_0} + 1}, z(\underline{n}))$$

$$p(x_{i_0 n_0 + 1}, z(\underline{n})/\underline{\vartheta}) = p(z(\underline{n}+1)/\underline{\vartheta})$$

son identidades.

Con todos los elementos anteriores se define como la información esperada sobre $\underline{\vartheta}$ proporcionada por el experimento $E(n_{i_0} + 1)/z(\underline{n})$ después de que se haya realizado $E^S(\underline{n})$, al valor esperado sobre $z(\underline{n})$ de la información

$$I^{\underline{\vartheta}} \left\{ E(n_{i_0} + 1)/z(\underline{n}), p(\underline{\vartheta}/z(\underline{n})) \right\}$$

esto es, a través de la integral

$$I^{\underline{\vartheta}} \left\{ E(n_{i_0} + 1)/E^S(\underline{n}), p(\underline{\vartheta}) \right\} =$$

$$= \int p(z(\underline{n})) \iint p(\underline{\vartheta}/z(\underline{n})) \cdot p(x_{i_0 n_0 + 1}/\underline{\vartheta}, z(\underline{n})) \cdot \log \frac{p(\underline{\vartheta}/x_{i_0 n_0 + 1}, z(\underline{n}))}{p(\underline{\vartheta}/z(\underline{n}))} \cdot d\underline{\vartheta} \cdot dz(\underline{n}) \cdot dx_{i_0 n_0 + 1}$$

Una demostración en la línea de la del teorema 5.4.1 establece que la información esperada anterior es no negativa.

TEOREMA 5.4.6

Dados los experimentos $E^S(\underline{n})$ y $E^S(\underline{n} + 1)$, una densidad inicial $p(\underline{\vartheta})$ y suponiendo que la nueva unidad muestral que diferencia a los dos experimentos anteriores pertenece al estrato i_0 -ésimo, se verifica el criterio de aditividad siguiente:

$$I^{\underline{\vartheta}} \left\{ E^S(\underline{n} + 1), p(\underline{\vartheta}) \right\} = I^{\underline{\vartheta}} \left\{ E^S(\underline{n}), p(\underline{\vartheta}) \right\} +$$

$$+ I^{\underline{\vartheta}} \left\{ E(n_{i_0} + 1)/E^S(\underline{n}), p(\underline{\vartheta}) \right\}$$

Demostración.- Siguiendo idéntica técnica de demostración que se utilizó en el teorema 5.4.2 y de las respectivas definiciones resulta:

$$I^{\underline{\vartheta}} \left\{ E^S(\underline{n}), p(\underline{\vartheta}) \right\} = \iint p(\underline{\vartheta}) \cdot p(z(\underline{n})/\underline{\vartheta}) \cdot \log \frac{p(\underline{\vartheta}/z(\underline{n}))}{p(\underline{\vartheta})} \cdot d\underline{\vartheta} \cdot dz(\underline{n}) =$$

$$= \iint p(\underline{\vartheta}) \cdot p(z(\underline{n}+1)/\underline{\vartheta}) \cdot \log \frac{p(\underline{\vartheta}/z(\underline{n}))}{p(\underline{\vartheta})} \cdot d\underline{\vartheta} \cdot dz(\underline{n}+1)$$

$$\begin{aligned} I^{\underline{\vartheta}} \{ E(n_{i_0} + 1) / E^S(\underline{n}), p(\underline{\vartheta}) \} &= \int p(z(\underline{n})) \iint p(\underline{\vartheta}/z(\underline{n})) \cdot \\ &\cdot p(x_{i_0 i_0} + 1 / \underline{\vartheta}, z(\underline{n})) \cdot \log \frac{p(\underline{\vartheta}/x_{i_0 i_0} + 1, z(\underline{n}))}{p(\underline{\vartheta}/z(\underline{n}))} \cdot d\underline{\vartheta} \cdot dx_{i_0 i_0} + 1 \cdot dz(\underline{n}) = \\ &= \iint p(\underline{\vartheta}) \cdot p(z(\underline{n}+1)/\underline{\vartheta}) \cdot \log \frac{p(\underline{\vartheta}/z(\underline{n}+1))}{p(\underline{\vartheta}/z(\underline{n}))} \cdot d\underline{\vartheta} \cdot dz(\underline{n}+1) \end{aligned}$$

Sumando ambas informaciones resulta finalmente

$$\begin{aligned} I^{\underline{\vartheta}} \{ E^S(\underline{n}), p(\underline{\vartheta}) \} + I^{\underline{\vartheta}} \{ E(n_{i_0} + 1) / E^S(\underline{n}), p(\underline{\vartheta}) \} &= \\ = \iint p(\underline{\vartheta}) \cdot p(z(\underline{n}+1)/\underline{\vartheta}) \cdot \log \frac{p(\underline{\vartheta}/z(\underline{n}+1))}{p(\underline{\vartheta})} \cdot d\underline{\vartheta} \cdot dz(\underline{n}+1) &= \end{aligned}$$

$$= I^{\underline{\vartheta}} \{ E^S(\underline{n} + 1), p(\underline{\vartheta}) \}$$

(c.q.d.)

Así pues, puede apreciarse que también $I^{\underline{\vartheta}} \{ E^S(\underline{n}), p(\underline{\vartheta}) \}$ considerada como función de la componente i_0 -ésima del vector \underline{n} verifica una aditividad análoga a la demostrada en el teorema 5.4.2. Al considerar el paso del experimento $E^S(\underline{n})$ al $E^S(\underline{n} + 1)$ y cuando la afijación está determinada de antemano queda, a su vez, completamente especificado el estrato i_0 -ésimo al que pertenece la nueva unidad muestreada. Si la afijación es desconocida no es posible a priori adscribirla a un estrato concreto, aunque sigue siendo cierto que el aumento de información producida al extraerla coincide con la diferencia entre las informaciones esperadas proporcionadas por los experimentos $E^S(\underline{n})$ y $E^S(\underline{n} + 1)$, de acuerdo con el teorema 5.4.6. Como corolario análogo al 5.4.2 tenemos el siguiente:

COROLARIO 5.4.3

Siendo $z(\underline{n})$ suficiente para $z(\underline{n}+1)$ con respecto a $\underline{\vartheta}$ en todos los estratos, esto es, verificándose que

$$p(\underline{\vartheta}/z(\underline{n}), x_{i, n_i+1}) = p(\underline{\vartheta}/z(\underline{n})) \quad i=1, 2, \dots, L$$

entonces

$$I^{\underline{\vartheta}}\{E^S(\underline{n}), p(\underline{\vartheta})\} = I^{\underline{\vartheta}}\{E^S(\underline{n}+1), p(\underline{\vartheta})\}$$

Demostración.— Verificándose la hipótesis de igualdad entre dichas densidades y por su misma definición, $I^{\underline{\vartheta}}\{E(n_i+1)^S/E^S(\underline{n}), p(\underline{\vartheta})\}$ es nula en todos los estratos. Así pues, y con independencia del estrato en el que se extraería la última unidad muestral del experimento $E^S(\underline{n}+1)$, se sigue el resultado propuesto del teorema 5.4.6.

(c.q.d.)

El corolario anterior proporciona una interesante consecuencia que aplicaremos en el próximo capítulo para la obtención práctica de la información $I^{\underline{\vartheta}}\{E^S(\underline{n}), p(\underline{\vartheta})\}$. En efecto, de dicho corolario se deduce que no hay pérdida en la información esperada cuando la investigación se ciñe en cada estrato al estudio de un estadístico suficiente para todo el vector aleatorio $\underline{\vartheta}$.

Evidentemente que el resultado del teorema 5.4.6 puede generalizarse a un número finito de experimentos con espacio paramétrico $\underline{\vartheta}$ común, siempre y cuando las nuevas unidades muestrales obtenidas pertenezcan a diferentes estratos. Así por ejemplo, para dos nuevas unidades de muestreo pertenecientes a los estratos i_0 e i_1 -ésimos respectivamente y correspondientes a las unidades adicionales de los experimentos $E^S(\underline{n}+1)$ y $E^S(\underline{n}+2)$, se verifica:

$$\begin{aligned} I^{\underline{\vartheta}}\{E^S(\underline{n}+1), p(\underline{\vartheta})\} &= I^{\underline{\vartheta}}\{E^S(\underline{n}), p(\underline{\vartheta})\} + I^{\underline{\vartheta}}\{E(n_{i_0}+1)^S/E^S(\underline{n}), p(\underline{\vartheta})\} \\ I^{\underline{\vartheta}}\{E^S(\underline{n}+2), p(\underline{\vartheta})\} &= I^{\underline{\vartheta}}\{E^S(\underline{n}+1), p(\underline{\vartheta})\} + I^{\underline{\vartheta}}\{E(n_{i_1}+1)^S/E^S(\underline{n}+1), p(\underline{\vartheta})\} = \\ &= I^{\underline{\vartheta}}\{E^S(\underline{n}), p(\underline{\vartheta})\} + I^{\underline{\vartheta}}\{E(n_{i_0}+1)^S/E^S(\underline{n}), p(\underline{\vartheta})\} + I^{\underline{\vartheta}}\{E(n_{i_1}+1)^S/E^S(\underline{n}+1), p(\underline{\vartheta})\} \end{aligned}$$

resultado que admite una evidente generalización al caso de $k \leq L$ nuevas unidades muestrales extraídas en diferentes estratos, generalización expresada a través de la ecuación

$$I^{\underline{\vartheta}}\{E^S(\underline{n}+k), p(\underline{\vartheta})\} = I^{\underline{\vartheta}}\{E^S(\underline{n}), p(\underline{\vartheta})\} + \sum_{i=1}^k I^{\underline{\vartheta}}\{E(n_{i_1}+1)^S/E^S(\underline{n}+i-1), p(\underline{\vartheta})\}$$

Combinando este último resultado con el obtenido del teorema 5.4.2 se deduce la existencia de una aditividad perfecta para la información esperada $I^{\vartheta}\{E^S(\underline{n}), p(\underline{\vartheta})\}$ en el sentido de que para un vector tamaño muestral \underline{n} dicha información puede obtenerse como suma de las informaciones obtenidas una a una y proporcionadas por cada nueva unidad muestral extraída. Evidentemente suponiendo conocido en cada etapa el resultado de la anterior. Aplicando este hecho a los experimentos $E^S(\underline{n})$ y $E^S(\underline{n+1})$ se cumple que

$$I^{\vartheta}\{E^S(\underline{n+1}), p(\underline{\vartheta})\} = I^{\vartheta}\{E^S(\underline{n}), p(\underline{\vartheta})\} + \sum_{i=1}^L I^{\vartheta}\{E(n_i+1)^{\#}/E^S(\underline{n+1-1}), p(\underline{\vartheta})\}$$

de donde deducimos el siguiente

COROLARIO 5.4.4

Bajo las hipótesis y notación de los teoremas 5.4.2 y 5.4.6

$$I^{\vartheta}\{E^S(\underline{n+1})^{\#}/E^S(\underline{n}), p(\underline{\vartheta})\} = \sum_{i=1}^L I^{\vartheta}\{E(n_i+1)^{\#}/E^S(\underline{n+1-1}), p(\underline{\vartheta})\}$$

Estudiada la aditividad anterior pasemos a comprobar el comportamiento de $I^{\vartheta}\{E^S(\underline{n}), p(\underline{\vartheta})\}$ como función de una componente particular n_i del vector \underline{n} , para así deducir su carácter cóncavo y creciente. A fin de fijar qué componente concreta es la considerada se utilizará la notación

$$I^{\vartheta}\{E^S(\underline{n}, n_i), p(\underline{\vartheta})\}$$

para indicar que se estudia la información esperada $I^{\vartheta}\{E^S(\underline{n}), p(\underline{\vartheta})\}$ como función de la componente i -ésima del vector tamaño muestral.

TEOREMA 5.4.7

$I^{\vartheta}\{E^S(\underline{n}, n_i), p(\underline{\vartheta})\}$ es una función creciente y cóncava respecto la componente i -ésima del vector \underline{n} .

Demostración.- Es creciente, puesto que

$$I^{\vartheta}\{E^S(\underline{n+1}, n_i+1), p(\underline{\vartheta})\} - I^{\vartheta}\{E^S(\underline{n}, n_i), p(\underline{\vartheta})\} = I^{\vartheta}\{E(n_i+1)^{\#}/E^S(\underline{n}, n_i), p(\underline{\vartheta})\} \geq 0$$

tal como se indicó en la demostración del teorema 5.4.6.

Para demostrar su concavidad será suficiente ver que la expresión

$$\Delta = I^{\vartheta} \{ E^S(\underline{n}, n_1), \rho(\underline{\vartheta}) \} - I^{\vartheta} \{ E^S(\underline{n-1}, n_1-1), \rho(\underline{\vartheta}) \} - \\ - \left\{ I^{\vartheta} \{ E^S(\underline{n+1}, n_1+1), \rho(\underline{\vartheta}) \} - I^{\vartheta} \{ E^S(\underline{n}, n_1), \rho(\underline{\vartheta}) \} \right\}$$

es mayor que cero. En efecto, Δ puede reescribirse como

$$\Delta = I^{\vartheta} \{ E(n_1)^{\vartheta} / E^S(\underline{n-1}, n_1-1), \rho(\underline{\vartheta}) \} - I^{\vartheta} \{ E(n_1+1)^{\vartheta} / E^S(\underline{n}, n_1), \rho(\underline{\vartheta}) \} - \\ - \iint \rho(\underline{\vartheta}) \cdot \rho(z(\underline{n}) / \underline{\vartheta}) \cdot \log \frac{\rho(\underline{\vartheta} / x_{1n_1}, z(\underline{n-1}))}{\rho(\underline{\vartheta} / z(\underline{n-1}))} \cdot d\underline{\vartheta} \cdot dz(\underline{n}) - \\ - \iint \rho(\underline{\vartheta}) \cdot \rho(z(\underline{n+1}) / \underline{\vartheta}) \cdot \log \frac{\rho(\underline{\vartheta} / x_{1n_1+1}, z(\underline{n}))}{\rho(\underline{\vartheta} / z(\underline{n}))} \cdot d\underline{\vartheta} \cdot dz(\underline{n+1}) - \\ - \iint \rho(\underline{\vartheta}) \cdot \rho(z(\underline{n+1}) / \underline{\vartheta}) \cdot \log \frac{\rho(\underline{\vartheta} / z(\underline{n}))^2}{\rho(\underline{\vartheta} / z(\underline{n-1})) \cdot \rho(\underline{\vartheta} / z(\underline{n+1}))} \cdot d\underline{\vartheta} \cdot dz(\underline{n+1})$$

La expresión bajo logaritmos resulta ser

$$\frac{\rho(z(\underline{n}) / \underline{\vartheta})^2}{\rho(z(\underline{n-1}) / \underline{\vartheta}) \cdot \rho(z(\underline{n+1}) / \underline{\vartheta})} \cdot \frac{\rho(z(\underline{n-1})) \cdot \rho(z(\underline{n+1}))}{\rho(z(\underline{n}))^2} - \\ - \frac{\rho(x_{1n_1} / \vartheta_1)}{\rho(x_{1n_1+1} / \vartheta_1)} \cdot \frac{\rho(z(\underline{n-1})) \cdot \rho(z(\underline{n+1}))}{\rho(z(\underline{n}))^2}$$

Y dado que x_{1n_1} y x_{1n_1+1} están equidistribuidos, se deduce que

$$\iint \rho(\underline{\vartheta}) \cdot \rho(z(\underline{n+1}) / \underline{\vartheta}) \cdot \log \frac{\rho(x_{1n_1} / \vartheta_1)}{\rho(x_{1n_1+1} / \vartheta_1)} \cdot d\underline{\vartheta} \cdot dz(\underline{n+1}) = 0$$

mientras que

$$\frac{\rho(z(\underline{n-1})) \cdot \rho(z(\underline{n+1}))}{\rho(z(\underline{n}))^2} = \frac{\rho(x_{1n_1+1} / z(\underline{n}))}{\rho(x_{1n_1} / z(\underline{n-1}))}$$

De ahí que

$$\Delta = \iint \rho(\underline{\vartheta}) \cdot \rho(z(\underline{n+1}) / \underline{\vartheta}) \cdot \log \frac{\rho(x_{1n_1+1} / z(\underline{n}))}{\rho(x_{1n_1} / z(\underline{n-1}))} \cdot d\underline{\vartheta} \cdot dz(\underline{n+1}) -$$

$$\begin{aligned}
&= \int p(z(\underline{n}+1)) \cdot \log \frac{p(x_{i_{n_i+1}}/x_{i_{n_i}}, z(\underline{n}-1))}{p(x_{i_{n_i+1}}/z(\underline{n}-1))} \cdot dz(\underline{n}+1) = \\
&= \iint p(z(\underline{n})) \cdot p(x_{i_{n_i+1}}/z(\underline{n})) \cdot \log \frac{p(x_{i_{n_i+1}}/x_{i_{n_i}}, z(\underline{n}-1))}{p(x_{i_{n_i+1}}/z(\underline{n}-1))} \cdot dz(\underline{n}) \cdot dx_{i_{n_i+1}}
\end{aligned}$$

que no es sino la información esperada que sobre $x_{i_{n_i+1}}$ proporciona $x_{i_{n_i}}$ en el sentido descrito por el teorema 5.4.6. Por un razonamiento análogo al utilizado en el teorema 5.4.1 dicha información es positiva.

(c.q.d.)

El comportamiento de $I^{\vartheta} \{ E^{\vartheta}(\underline{n}), p(\vartheta) \}$ como función del tamaño muestral total n es ligeramente distinto del estudiado hasta ahora para \underline{n} y n_i ($i=1,2,\dots,L$) en el sentido de que sus propiedades van a tener un marcado carácter escalonado como se tendrá ocasión de comprobar al conjugar, precisamente, las demostradas para el vector tamaño muestral \underline{n} y para su i -ésima componente n_i . A tal fin, consideraremos dos tamaños muestrales totales n' y n'' tales que verifican las L condiciones

$$n'_i < n''_i \quad i=1,2,\dots,L$$

donde

$$\underline{n}' = (n'_1, n'_2, \dots, n'_L) \quad \text{y} \quad \underline{n}'' = (n''_1, n''_2, \dots, n''_L)$$

Sigamos idéntico razonamiento al empleado ya en dos ocasiones anteriores. Sean, pues, los experimentos $E^{\vartheta}(\underline{n}')$ y $E^{\vartheta}(\underline{n}'')$. Para pasar de uno a otro es necesario obtener un total de $n''-n'$ nuevas unidades de muestreo, repartiéndose $n''-n'_i$ en cada uno de los estratos. Siendo $z(\underline{n}')$ y $z(\underline{n}'')$ los resultados de los experimentos anteriores, respectivamente, existe la siguiente relación entre las densidades $p(z(\underline{n}')/\vartheta)$ y $p(z(\underline{n}'')/\vartheta)$:

$$p(z(\underline{n}'')/\vartheta) = \prod_{i=1}^L \prod_{j=1}^{n''_i} p(x_{ij}/\vartheta_i) = \prod_{i=1}^L \prod_{j=1}^{n'_i} p(x_{ij}/\vartheta_i) \cdot$$

$$\prod_{i=1}^L \prod_{j=1}^{n''_i - n'_i} p(x_{ij}/\vartheta_i) = p(z(\underline{n}')/\vartheta) \cdot \prod_{i=1}^L \prod_{j=1}^{n''_i - n'_i} p(x_{ij}/\vartheta_i)$$

Considerando el experimento $E^{\vartheta}(\underline{n}'' - \underline{n}')$ consistente en obtener esas $n''-n'$

nuevas unidades muestrales y repartidas entre los estratos siguiendo las indicaciones anteriores, esto es, con $\underline{n''-n'} = (n''_1-n'_1, n''_2-n'_2, \dots, n''_L-n'_L)$, la igualdad anterior puede reescribirse como

$$p(z(\underline{n''})/\underline{\vartheta}) = p(z(\underline{n'})/\underline{\vartheta}) \cdot p(z(\underline{n''-n'})/\underline{\vartheta}) = p(z(\underline{n'}), z(\underline{n''-n'})/\underline{\vartheta})$$

Definamos ahora el experimento $E^S((\underline{n''-n'})^{\#}/z(\underline{n'}))$ consistente en la familia de L experimentos realizados en sendos estratos

$$E^S((\underline{n''-n'})^{\#}/z(\underline{n'})) = \left\{ E(n''_i-n'_i)^{\#}/z_i(n'_i) \right\}_{i=1,2,\dots,L}$$

$$= \left\{ \left\{ X_i(n''_i-n'_i) \right\}, \left\{ \vartheta_i \right\}, \left\{ p(z_i(n''_i-n'_i)/\vartheta_i, z_i(n'_i)) \right\} \right\}_{i=1,2,\dots,L}$$

y que consista en obtener en el estrato i -ésimo las $n''_i-n'_i$ unidades de muestreo que faltan para completar el experimento $E(n''_i)$, supuestas obtenidas y conocidas las muestras $z_i(n'_i)$ ($i=1,2,\dots,L$) resultados de los experimentos $E(n'_i)$ que definen $E^S(\underline{n'})$. La información esperada proporcionada por el experimento $E^S((\underline{n''-n'})^{\#}/z(\underline{n'}))$ sobre $\underline{\vartheta}$ una vez que se ha realizado $E^S(\underline{n'})$ y obtenido $z(\underline{n'})$ viene dada por la doble integral

$$I^{\underline{\vartheta}} \left\{ E^S((\underline{n''-n'})^{\#}/z(\underline{n'})), p(\underline{\vartheta}/z(\underline{n'})) \right\} =$$

$$= \iint p(\underline{\vartheta}/z(\underline{n'})) \cdot p(z(\underline{n''-n'})/\underline{\vartheta}, z(\underline{n'})) \cdot \log \frac{p(\underline{\vartheta}/z(\underline{n''-n'}), z(\underline{n'}))}{p(\underline{\vartheta}/z(\underline{n'}))} \cdot d\underline{\vartheta} \cdot dz(\underline{n''-n'})$$

donde $p(z(\underline{n''-n'})/\underline{\vartheta}, z(\underline{n'}))$ es la densidad que describe el comportamiento del resultado $z(\underline{n''-n'})$ que depende, a su vez, del resultado $z(\underline{n'})$ del experimento $E^S(\underline{n'})$ y del vector paramétrico $\underline{\vartheta} \in \underline{\Omega}$. De ahí que siendo independientes las realizaciones de los experimentos $E((n''_i-n'_i)^{\#}/z_i(n'_i))$ ($i=1,2,\dots,L$) en sus estratos respectivos

$$p(z(\underline{n''-n'})/\underline{\vartheta}, z(\underline{n'})) = \prod_{i=1}^L \prod_{j=1}^{n''_i-n'_i} p(x_{ij}/\vartheta_i, z_i(n'_i))$$

Notemos que, por otra parte, el resultado $z(\underline{n''})$ del experimento $E^S(\underline{n''})$ puede expresarse a través del par de resultados $(z(\underline{n''-n'}), z(\underline{n'}))$ de los experimentos $E^S((\underline{n''-n'})^{\#}/z(\underline{n'}))$ y $E^S(\underline{n'})$ respectivamente. De ahí que sean identidades las igualdades entre densidades

$$p(\underline{\vartheta}/z(\underline{n''-n'}), z(\underline{n'})) = p(\underline{\vartheta}/z(\underline{n''}))$$

$$p(z(\underline{n}''-\underline{n}'), z(\underline{n}')/\underline{\vartheta}) = p(z(\underline{n}'')/\underline{\vartheta})$$

Con todos los elementos anteriores definiremos como información esperada sobre $\underline{\vartheta}$ proporcionada por el experimento $E((\underline{n}''-\underline{n}')^{\mathbb{S}}/z(\underline{n}'))$ después de que se haya realizado $E^{\mathbb{S}}(\underline{n}')$ como el valor esperado sobre $z(\underline{n}')$ de la información

$$I^{\underline{\vartheta}}\{E^{\mathbb{S}}((\underline{n}''-\underline{n}')^{\mathbb{S}}/z(\underline{n}')), p(\underline{\vartheta}/z(\underline{n}'))\}$$

esto es, a través de la integral

$$I^{\underline{\vartheta}}\{E^{\mathbb{S}}(\underline{n}''-\underline{n}')^{\mathbb{S}}/E^{\mathbb{S}}(\underline{n}'), p(\underline{\vartheta})\} =$$

$$= \int p(z(\underline{n}')) \iint p(\underline{\vartheta}/z(\underline{n}')) \cdot p(z(\underline{n}''-\underline{n}')/\underline{\vartheta}, z(\underline{n}')) \cdot \log \frac{p(\underline{\vartheta}/z(\underline{n}''-\underline{n}'), z(\underline{n}'))}{p(\underline{\vartheta}/z(\underline{n}'))} \cdot$$

$$\cdot d\underline{\vartheta} \cdot dz(\underline{n}''-\underline{n}') \cdot dz(\underline{n}')$$

Una demostración según la línea del teorema 5.4.1 establece que la información anterior es también no negativa, tomando como expresión para $g(\underline{\vartheta}, z(\underline{n}''))$ la

$$g(\underline{\vartheta}, z(\underline{n}'')) = p(\underline{\vartheta}/z(\underline{n}'')) / p(\underline{\vartheta}/z(\underline{n}'))$$

Con la notación introducida demosetremos el siguiente resultado que corresponde a la aditividad satisfecha por la información esperada sobre $\underline{\vartheta}$ proporcionada por el experimento $E^{\mathbb{S}}(\underline{n})$ y considerada como función del tamaño muestral total n , resultado que no es sino una generalización de los contenidos en los teoremas 5.4.2 y 5.4.6.

TEOREMA 5.4.8

Dados los tamaños muestrales totales n' y n'' tales que su adscripción en los diferentes estratos en tamaños muestrales parciales da lugar a los vectores tamaños muestrales \underline{n}' y \underline{n}'' que verifican que $n'_i < n''_i$ $i=1,2,\dots,L$, y siendo $p(\underline{\vartheta})$ la densidad inicial, se cumple el siguiente criterio de aditividad:

$$I^{\underline{\vartheta}}\{E^{\mathbb{S}}(\underline{n}''), p(\underline{\vartheta})\} = I^{\underline{\vartheta}}\{E^{\mathbb{S}}(\underline{n}'), p(\underline{\vartheta})\} + I^{\underline{\vartheta}}\{E^{\mathbb{S}}(\underline{n}''-\underline{n}')^{\mathbb{S}}/E^{\mathbb{S}}(\underline{n}'), p(\underline{\vartheta})\}$$

Demostración.- Utilizando idéntica técnica de demostración que la empleada en los teoremas 5.4.2 y 5.4.6, resulta:



$$I^{\vartheta} \{ E^S(\underline{n}'), p(\underline{\vartheta}) \} = \iint p(\underline{\vartheta}) \cdot p(z(\underline{n}')/\underline{\vartheta}) \cdot \log \frac{p(\underline{\vartheta}/z(\underline{n}'))}{p(\underline{\vartheta})} \cdot d\underline{\vartheta} \cdot dz(\underline{n}') =$$

$$= \iint p(\underline{\vartheta}) \cdot p(z(\underline{n}'')/\underline{\vartheta}) \cdot \log \frac{p(\underline{\vartheta}/z(\underline{n}''))}{p(\underline{\vartheta})} \cdot d\underline{\vartheta} \cdot dz(\underline{n}'')$$

$$I^{\vartheta} \{ E^S(\underline{n}''-\underline{n}')^S/E^S(\underline{n}'), p(\underline{\vartheta}) \} = \int p(z(\underline{n}')) \iint p(\underline{\vartheta}/z(\underline{n}')) \cdot p(z(\underline{n}''-\underline{n}')/\underline{\vartheta}, z(\underline{n}')) \cdot$$

$$\cdot \log \frac{p(\underline{\vartheta}/z(\underline{n}''-\underline{n}'), z(\underline{n}'))}{p(\underline{\vartheta}/z(\underline{n}'))} \cdot d\underline{\vartheta} \cdot dz(\underline{n}''-\underline{n}') \cdot dz(\underline{n}') =$$

$$= \iint p(\underline{\vartheta}) \cdot p(z(\underline{n}'')/\underline{\vartheta}) \cdot \log \frac{p(\underline{\vartheta}/z(\underline{n}''))}{p(\underline{\vartheta}/z(\underline{n}'))} \cdot d\underline{\vartheta} \cdot dz(\underline{n}'')$$

Sumando ambas informaciones resulta finalmente

$$I^{\vartheta} \{ E^S(\underline{n}'), p(\underline{\vartheta}) \} + I^{\vartheta} \{ E^S(\underline{n}''-\underline{n}')^S/E^S(\underline{n}'), p(\underline{\vartheta}) \} =$$

$$= \iint p(\underline{\vartheta}) \cdot p(z(\underline{n}'')/\underline{\vartheta}) \cdot \log \frac{p(\underline{\vartheta}/z(\underline{n}''))}{p(\underline{\vartheta}/z(\underline{n}'))} \cdot d\underline{\vartheta} \cdot dz(\underline{n}'') = I^{\vartheta} \{ E^S(\underline{n}''), p(\underline{\vartheta}) \}$$

(c.q.d.)

Notemos primeramente que cuando la afijación de la estratificación es conocida a priori, esto es, cuando se conoce una familia de L constantes $\{f_i\}_{i=1,2,\dots,L}$ tales que $n_i = f_i \cdot n$ ($i=1,2,\dots,L$), dados los tamaños totales n' y n'' se conocen, de forma automática, los vectores \underline{n}' y \underline{n}'' . En segundo lugar, es de destacar que el teorema 5.4.8 generaliza los ya demostrados con anterioridad al considerar incrementos no iguales de unidades muestreadas en los diferentes estratos, sirviendo para reforzar la idea de la aditividad perfecta satisfecha por la información esperada sobre ϑ proporcionada por el experimento $E^S(\underline{n})$ asociado a un proceso de estratificación.

Para caracterizar el comportamiento cóncavo de $I^{\vartheta} \{ E^S(\underline{n}), p(\underline{\vartheta}) \}$ como función del tamaño total n es necesario introducir la siguiente notación: sean n', n'' y n''' tres tamaños muestrales totales que una vez adscritos en los correspondientes tamaños muestrales parciales n'_i, n''_i y n'''_i ($i=1,2,\dots,L$) verifican las condiciones

$$n_1'' - n_1' = n_1''' - n_1'' = \alpha_1 \quad i=1,2,\dots,L$$

Así pues, considerando los vectores muestrales definidos con dichas componentes $\underline{n}^\wedge = (n_1^\wedge, n_2^\wedge, \dots, n_L^\wedge)$ y los experimentos consiguientes $E^S(\underline{n}^\wedge)$, el paso de $E^S(\underline{n}')$ a $E^S(\underline{n}'')$ se realiza estudiando el mismo número α_1 de unidades muestrales adicionales en cada estrato respecto $E^S(\underline{n}')$ que las unidades de muestreo adicionales estudiadas en el respectivo estrato al pasar de $E^S(\underline{n}'')$ a $E^S(\underline{n}''')$. Por otra parte, destaquemos nuevamente que con la afijación fijada de antemano la adscripción de los tamaños muestrales t_{ij} tales en los distintos tamaños parciales de las muestras de los estratos es un proceso cíclico y perfectamente determinado. Bajo estas condiciones y teniendo en cuenta la definición de $I^{\vartheta}\{E(\underline{n}''-\underline{n}')/E^S(\underline{n}'), p(\vartheta)\}$ pueda escribirse que

$$I^{\vartheta}\{E^S(\underline{n}''), p(\vartheta)\} - I^{\vartheta}\{E^S(\underline{n}'), p(\vartheta)\} = I^{\vartheta}\{E^S(\underline{\alpha})^{\#}/E^S(\underline{n}'), p(\vartheta)\} \geq 0$$

$$I^{\vartheta}\{E^S(\underline{n}'''), p(\vartheta)\} - I^{\vartheta}\{E^S(\underline{n}''), p(\vartheta)\} = I^{\vartheta}\{E^S(\underline{\alpha})^{\#}/E^S(\underline{n}''), p(\vartheta)\} \geq 0$$

y donde $\underline{\alpha}$ denota el vector diferencia $\underline{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_L)$. Pues bien, para los tamaños n' , n'' y n''' que verifican las anteriores hipótesis se cumple una especie de concavidad en cuanto que satisfacen la condición exigida por ésta. Este comportamiento, satisfecho por tamaños muestrales totales que difieren entre sí el mismo número de unidades y que respetan este hecho en todos los tamaños parciales de los estratos, produce el especial efecto de concavidad en escalera al que antes nos referíamos. Demostremos ya el siguiente

TEOREMA 5.4.9

Para unos tamaños muestrales totales n' , n'' y n''' tales que su adscripción en los diferentes estratos en tamaños muestrales parciales dan lugar a tres vectores muestrales \underline{n}' , \underline{n}'' y \underline{n}''' cuyas componentes satisfacen las condiciones

$$n_1'' - n_1' = n_1''' - n_1'' = \alpha_1 \quad i=1,2,\dots,L$$

y siendo $p(\vartheta)$ la densidad inicial, se verifica:

$$1^{\circ} \quad I^{\vartheta}\{E^S(\underline{n}'''), p(\vartheta)\} \geq I^{\vartheta}\{E^S(\underline{n}''), p(\vartheta)\} \geq I^{\vartheta}\{E^S(\underline{n}'), p(\vartheta)\}$$

$$2^{\circ} \quad 2 \cdot I^{\vartheta}\{E^S(\underline{n}''), p(\vartheta)\} - I^{\vartheta}\{E^S(\underline{n}'''), p(\vartheta)\} - I^{\vartheta}\{E^S(\underline{n}'), p(\vartheta)\} \geq 0$$

Demostración.— La primera condición de crecimiento queda recogida en las dos últimas desigualdades escritas inmediatamente antes del teorema. Por otra parte, demostrar la condición segunda es equivalente a ver que

$$\Delta = I^{\vartheta} \{ E^S(\underline{\alpha})^2 / E^S(n'), p(\underline{\vartheta}) \} - I^{\vartheta} \{ E^S(\underline{\alpha})^2 / E^S(n''), p(\underline{\vartheta}) \} \geq 0$$

Como Δ puede reescribirse a través de la doble integral

$$\Delta = \iint p(\underline{\vartheta}) \cdot p(z(n''')/\underline{\vartheta}) \cdot \log \frac{p(\underline{\vartheta}/z(n'''))^2}{p(\underline{\vartheta}/z(n')) \cdot p(\underline{\vartheta}/z(n'''))} \cdot d\underline{\vartheta} \cdot dz(n''')$$

la expresión bajo logaritmos resulta ser

$$\frac{p(z(n'')/\underline{\vartheta})^2 \cdot p(z(n''')) \cdot p(z(n'))}{p(z(n')/\underline{\vartheta}) \cdot p(z(n''')/\underline{\vartheta}) \cdot p(z(n''))^2} =$$

$$= \frac{p(z(n''')) \cdot p(z(n')) \prod_{i=1}^L \prod_{j=n'_i+1}^{n''_i} p(x_{ij}/\vartheta_i)}{p(z(n''))^2 \prod_{i=1}^L \prod_{j=n''_i+1}^{n'_i} p(x_{ij}/\vartheta_i)}$$

$$= \prod_{i=1}^L \prod_{j=1}^{n''_i} \frac{p(x_{in''_i+j}/\vartheta_i) \cdot p(z(n''')) \cdot p(z(n'))}{p(x_{in''_i+j}/\vartheta_i) \cdot p(z(n''))^2}$$

Dado que las cantidades aleatorias $x_{in''_i+j}$ y $x_{in''_i+j}$ están equidistribuidas se deduce que

$$\iint p(\underline{\vartheta}) \cdot p(z(n''')/\underline{\vartheta}) \cdot \log \frac{\prod_{i=1}^L \prod_{j=1}^{n''_i} p(x_{in''_i+j}/\vartheta_i)}{\prod_{i=1}^L \prod_{j=1}^{n''_i} p(x_{in''_i+j}/\vartheta_i)} \cdot d\underline{\vartheta} \cdot dz(n''') =$$

$$= \sum_{i=1}^L \sum_{j=1}^{\alpha_i} \iint p(\underline{\vartheta}) \cdot p(z(n''')/\underline{\vartheta}) \cdot \log \frac{p(x_{in''_i+j}/\vartheta_i)}{p(x_{in''_i+j}/\vartheta_i)} \cdot d\underline{\vartheta} \cdot dz(n''') = 0$$

mientras que

$$\frac{\rho(z(\underline{n}')) \cdot \rho(z(\underline{n}'''))}{\rho(z(\underline{n}''))^2} = \frac{\rho(z(\underline{n}''''-\underline{n}'')/z(\underline{n}''))}{\rho(z(\underline{n}''-\underline{n}')/z(\underline{n}'))} =$$

$$= \frac{\rho(z(\underline{n}''''-\underline{n}'')/z(\underline{n}''-\underline{n}'), z(\underline{n}'))}{\rho(z(\underline{n}''-\underline{n}')/z(\underline{n}'))}$$

De ahí que

$$\Delta = \iiint \rho(\underline{z}) \cdot \rho(z(\underline{n}''')) / \underline{z} \cdot \log \frac{\rho(z(\underline{n}''''-\underline{n}'')/z(\underline{n}'), z(\underline{n}''-\underline{n}'))}{\rho(z(\underline{n}''-\underline{n}')/z(\underline{n}'))} \cdot d\underline{z} \cdot dz(\underline{n}''') =$$

$$= \iiint \rho(z(\underline{n}'')) \cdot \rho(z(\underline{n}''''-\underline{n}'')/z(\underline{n}'')) \cdot \log \frac{\rho(z(\underline{n}''''-\underline{n}'')/z(\underline{n}''-\underline{n}'), z(\underline{n}'))}{\rho(z(\underline{n}''-\underline{n}')/z(\underline{n}'))} \cdot d\underline{z} \cdot dz(\underline{n}''')$$

que no es sino la información esperada que sobre $z(\underline{n}''''-\underline{n}'')$ proporciona $z(\underline{n}''-\underline{n}')$ en el sentido descrito en el teorema anterior. Por un razonamiento análogo al utilizado en el teorema 5.4.1 dicha información es positiva.

(c.q.d.)

La desigualdad demostrada ahora mismo responde a una concavidad escalonada en cuanto que asegura dicho comportamiento en situaciones en las que los tamaños muestrales totales n' , n'' y n''' verifican las condiciones exigidas en el teorema anterior. Estas condiciones, cuando la afijación viene determinada de antemano, admiten una formulación mucho más elegante. En efecto, conocidas las L constantes f_i tales que $n_i = f_i \cdot n$ ($i=1,2, \dots, L$) para cualquier tamaño muestral total y denotando por f_1 la mayor de ellas (se supondrá que $1/f_1 \in \mathbb{Z}$), la adscripción de $1/f_1$ nuevas unidades se realiza de forma cíclica y perfectamente determinada. En efecto, supuesto un vector tamaño muestral \underline{n} , la adscripción de una nueva unidad muestral adicional se realizará en aquel estrato que por su afijación le corresponde. Al considerar $1/f_1$ nuevas unidades se verificará, ante ese carácter cíclico y determinado de la adscripción, que todo estrato verá incrementado en al menos una unidad su tamaño muestral respectivo. Comparando, pues, dos situaciones que supongan sendos incrementos de $1/f_1$ nuevas unidades muestreadas los incrementos de nuevas unidades en cada uno de los estratos se producirán añadiendo el mismo número de ellas. Esta situación es precisamente la descrita en las hipótesis del teorema 5.4.9. Resulta,

finalmente, el siguiente y evidente

COROLARIO 5.4.5

Con una afijación de la estratificación conocida previamente, esto es, dadas las constantes f_i tales que $n_i = f_i \cdot n$ ($i=1,2,\dots,L$) para cualquier tamaño muestral total n , y respetando un criterio cíclico y determinado para la adscripción de nuevas unidades de muestreo, siendo f_1 la menor de dichas constantes ($1/f_1 \in \mathbb{Z}$) y $p(\underline{\vartheta})$ la densidad inicial, se verifica que

$$1^\circ I^{\underline{\vartheta}} \{ E^S(\underline{n} + (1/f_1)), p(\underline{\vartheta}) \} - I^{\underline{\vartheta}} \{ E^S(\underline{n}), p(\underline{\vartheta}) \} \geq 0$$

$$2^\circ 2 \cdot I^{\underline{\vartheta}} \{ E^S(\underline{n}), p(\underline{\vartheta}) \} - I^{\underline{\vartheta}} \{ E^S(\underline{n} + (1/f_1)), p(\underline{\vartheta}) \} - I^{\underline{\vartheta}} \{ E^S(\underline{n} - (1/f_1)), p(\underline{\vartheta}) \} \geq 0$$

Centremos ahora el estudio a un caso particular de gran importancia. Para ello seguiremos con los mismos elementos de trabajo hasta ahora considerados sobre los que se realiza una hipótesis adicional: la de la independencia entre las nuevas cantidades de interés ϑ_i . Esta hipótesis adicional simplifica considerablemente la fuerte estructuración introducida por el experimento $E^S(\underline{n})$ definido como una familia de experimentos realizados en sendos estratos. Al mismo tiempo, también facilitará sensiblemente el cálculo y obtención de la información esperada $I^{\underline{\vartheta}} \{ E^S(\underline{n}), p(\underline{\vartheta}) \}$ y de otras magnitudes relacionadas con ella, como son la utilidad y el costo esperados del experimento $E^S(\underline{n})$. Precisamente estas simplificaciones pueden justificar de por sí el empleo de esta independencia aunque, como acaba de afirmarse, la hipótesis sea fuertemente restrictiva. En el mismo mecanismo de construcción de la estratificación de la población X se ha indicado que la actuación sobre el espacio paramétrico Θ a que da lugar implica la consideración de L nuevas cantidades aleatorias ϑ_i que, generalmente, son interdependientes. De ahí que la densidad $p(\underline{\vartheta})$ que describe las opiniones iniciales que el investigador tiene sobre el vector paramétrico $\underline{\vartheta}$ formado por las cantidades ϑ_i tenga la consideración de una densidad conjunta que indirectamente da idea de la interdependencia de sus componentes. Suponer, por tanto, que los parámetros ϑ_i son independientes implica inmediatamente la existencia de una densidad de la forma

$$p(\underline{\vartheta}) = \prod_{i=1}^L p(\vartheta_i / \gamma_i) = \prod_{i=1}^L p(\vartheta_i)$$

con

$$\int p(\vartheta_i / \gamma_i) \cdot d\vartheta_i = 1 \quad i=1,2,\dots,L$$

esto es, obtenida como un producto de densidades definidas para cada componente ϑ_i por separado y donde los γ_i son parámetros propios de la densidad respectiva. Precisamente la descripción de las opiniones iniciales sobre $\underline{\vartheta}$ a partir del conjunto de densidades $p(\varphi)$ y $p(\vartheta_i / \varphi)$ ($i=1,2,\dots,L$) tiene la ventaja práctica de que, lógicamente, le resultará más fácil al investigador describir sus opiniones iniciales sobre cada ϑ_i por separado que considerando su conjunto $\underline{\vartheta}$. Bajo la hipótesis de independencia, sin embargo, es evidente que las densidades a considerar serán de la forma $p(\vartheta_i / \gamma_i)$, con γ_i un parámetro propio de la distribución de ϑ_i , no teniendo entonces la consideración del parámetro φ con su propia densidad inicial. Si bien la hipótesis de independencia es fuertemente restrictiva, el investigador podrá actuar sobre las densidades $p(\vartheta_i / \gamma_i)$ para intentar suavizar dicha hipótesis. En este sentido, no sólo deberá escoger de entre una de las posibles familias de densidades de probabilidad para describir sus opiniones, sino que también deberá asignar valores concretos a los parámetros γ_i de esas densidades para que, en probabilidad, las cantidades aleatorias ϑ_i tomen valores bastante alejados o próximos, más o menos centrados, etc., a fin de ajustar de esta forma las dependencias de las cantidades ϑ_i . Lo cierto es que bajo la independencia de las cantidades ϑ_i se produce una gran simplificación en el tratamiento de la información esperada sobre $\underline{\vartheta}$ proporcionada por $E^S(\underline{n})$, como queda recogido en el siguiente

TEOREMA 5.4.10

Para un experimento $E^S(\underline{n}) = \{E(n_i)\}_{i=1,2,\dots,L}$ y una densidad inicial

$$p(\underline{\vartheta}) = \prod_{i=1}^L p(\vartheta_i)$$

la información esperada sobre $\underline{\vartheta}$ proporcionada por $E^S(\underline{n})$ es suma de las informaciones esperadas sobre cada ϑ_i proporcionadas por

los experimentos $E(n_i)$, tal como las define Lindley (1956).

Demostración.- Siendo $z = (z_1(n_1), z_2(n_2), \dots, z_L(n_L))$ el resultado de $E^S(\underline{n})$, la densidad posterior que describe el comportamiento del vector paramétrico $\underline{\vartheta}$ conocido dicho resultado viene dada por

$$\begin{aligned}
 p(\underline{\vartheta}/z) &= \frac{\prod_{i=1}^L \prod_{j=1}^{n_i} p(x_{ij}/\vartheta_i) \cdot p(\vartheta_i)}{\iint \dots \int \prod_{i=1}^L \prod_{j=1}^{n_i} p(\vartheta_i) \cdot p(x_{ij}/\vartheta_i) \cdot d\vartheta_1 \cdot d\vartheta_2 \cdot \dots \cdot d\vartheta_L} \\
 &= \frac{\prod_{i=1}^L p(\vartheta_i) \cdot p(z_i(n_i)/\vartheta_i)}{\iint \dots \int \prod_{i=1}^L p(\vartheta_i) \cdot p(z_i(n_i)/\vartheta_i) \cdot d\vartheta_1 \cdot d\vartheta_2 \cdot \dots \cdot d\vartheta_L} \\
 &= \frac{\prod_{i=1}^L p(\vartheta_i) \cdot p(z_i(n_i)/\vartheta_i)}{\prod_{i=1}^L \int p(\vartheta_i) \cdot p(z_i(n_i)/\vartheta_i) \cdot d\vartheta_i} = \prod_{i=1}^L \frac{p(\vartheta_i) \cdot p(z_i(n_i)/\vartheta_i)}{\int p(\vartheta_i) \cdot p(z_i(n_i)/\vartheta_i) \cdot d\vartheta_i} \\
 &= \prod_{i=1}^L p(\vartheta_i/z_i(n_i))
 \end{aligned}$$

esto es, el producto extendido a todos los estratos de las densidades posteriores de las cantidades ϑ_i conocidos los resultados $z_i(n_i)$ de los experimentos $E(n_i)$ componentes del $E^S(\underline{n})$. Tomando logaritmos en el cociente

$$p(\underline{\vartheta}/z) / p(\underline{\vartheta})$$

resulta:

$$\begin{aligned}
 \log \frac{p(\underline{\vartheta}/z)}{p(\underline{\vartheta})} &= \log \frac{\prod_{i=1}^L p(\vartheta_i/z_i(n_i))}{\prod_{i=1}^L p(\vartheta_i)} = \log \prod_{i=1}^L \frac{p(\vartheta_i/z_i(n_i))}{p(\vartheta_i)} \\
 &= \sum_{i=1}^L \log \frac{p(\vartheta_i/z_i(n_i))}{p(\vartheta_i)}
 \end{aligned}$$

Finalmente, con $Z_i = X_{i1} X_{i2} \dots X_{i(n_i)}$, resulta

$$\begin{aligned}
 I^{\vartheta} \{ E^S(\underline{n}), p(\underline{\vartheta}) \} &= \iint p(\underline{\vartheta}) \cdot p(z/\underline{\vartheta}) \cdot \log \frac{p(\underline{\vartheta}/z)}{p(\underline{\vartheta})} \cdot d\underline{\vartheta} \cdot dz = \\
 &= \iint \dots \int \prod_{i=1}^L p(\vartheta_i) \cdot p(z_i(n_i)/\vartheta_i) \left\{ \sum_{i=1}^L \log \frac{p(\vartheta_i/z_i(n_i))}{p(\vartheta_i)} \right\} d\vartheta_1 \dots d\vartheta_L \cdot \\
 &\cdot dz_1(n_1) \dots dz_L(n_L) = \sum_{i=1}^L \int_{Z_i} \int_{\Theta_i} p(\vartheta_i) \cdot p(z_i(n_i)/\vartheta_i) \cdot \log \frac{p(\vartheta_i/z_i(n_i))}{p(\vartheta_i)} \cdot \\
 &\cdot d\vartheta_i \cdot dz_i(n_i) = \sum_{i=1}^L I^{\vartheta_i} \{ E(n_i), p(\vartheta_i) \}
 \end{aligned}$$

donde, como puede apreciarse, $I^{\vartheta_i} \{ E(n_i), p(\vartheta_i) \}$ es la información esperada sobre ϑ_i proporcionada por el resultado del experimento $E(n_i)$ definida por Lindley (1956).

(c.q.d.)

Este teorema transforma una información que está concebida como la proporcionada por un proceso global en una suma de informaciones esperadas locales, entendiendo por ello informaciones cuyos elementos definitorios actúan exclusivamente en cada estrato: parámetro ϑ_i y muestra $z_i(n_i)$. Este resultado es completamente lógico ante la hipótesis de partida. En efecto, se ha supuesto que los parámetros ϑ_i componentes del vector $\underline{\vartheta}$ son independientes. De ahí que las observaciones obtenidas en cada estrato - obtención que, por otra parte, es independiente de la realizada en los demás - carezcan de información sobre las cantidades de interés que no sean la suya respectiva, de forma que la información total sobre el vector paramétrico $\underline{\vartheta}$ queda reducida a la suma de las informaciones parciales sobre cada una de sus componentes. Así pues el sumatorio anterior queda justificado ante la doble independencia - parámetros de interés y métodos de obtención de las muestras en los distintos estratos - del proceso de muestreo estratificado asociado a la realización del experimento $E^S(\underline{n})$.

Por otra parte, y teniendo en cuenta las propiedades ya demostradas, el sumatorio de informaciones obtenido en el teorema 5.4.10 como

caso particular de la información esperada $I^{\vartheta} \{E^S(\underline{n}), p(\underline{\vartheta})\}$, toma valores no negativos y solamente se anula cuando en todos los estratos no se modifican las opiniones iniciales tras la realización de los experimentos $E(n_i)$, esto es, cuando $p(\vartheta_i/z_i(n_i)) = p(\vartheta_i)$ ($i=1,2,\dots,L$). Verifica idénticos criterios de aditividad que los recogidos en los teoremas 5.4.2, 5.4.6 y 5.4.8. En particular son ciertas las ecuaciones

$$\begin{aligned} \sum_{i=1}^L I^{\vartheta_i} \{E(n_i+1), p(\vartheta_i)\} &= \sum_{i=1}^L I^{\vartheta_i} \{E(n_i), p(\vartheta_i)\} + \\ &+ \sum_{i=1}^L I^{\vartheta_i} \{E(n_i+1)^S/E(n_i), p(\vartheta_i)\} \\ \sum_{i \neq i_0}^L I^{\vartheta_i} \{E(n_i), p(\vartheta_i)\} + I^{\vartheta_{i_0}} \{E(n_{i_0}+1), p(\vartheta_{i_0})\} &= I^{\vartheta} \{E^S(\underline{n}+1), p(\underline{\vartheta})\} - \\ &= \sum_{i=1}^L I^{\vartheta_i} \{E(n_i), p(\vartheta_i)\} + I^{\vartheta_{i_0}} \{E(n_{i_0}+1)^S/E(n_{i_0}), p(\vartheta_{i_0})\} \end{aligned}$$

toda vez que se satisfacen componente a componente (Lindley, 1956). Se trata también de una expresión creciente y cóncava como función del vector tamaño muestral \underline{n} y de cualquiera de sus componentes n_i , verificando así mismo idéntica concavidad escalonada como función del tamaño muestral total n que la demostrada para $I^{\vartheta} \{E^S(\underline{n}), p(\underline{\vartheta})\}$ en el teorema 5.4.9. Por último, se trata de una funcional cóncava de la densidad inicial $p(\underline{\vartheta})$, considerando densidades relacionadas a través de igualdades de la forma

$$\prod_{i=1}^L p(\vartheta_i) = \lambda \prod_{i=1}^L p_1(\vartheta_i) + (1-\lambda) \prod_{i=1}^L p_2(\vartheta_i)$$

Aparte de estas propiedades que son las generales de cualquier información esperada el sumatorio anterior es una función del vector \underline{n} pero dependiendo de cada una de sus componentes a través de un único sumatorio. Esta circunstancia lo hace de especial aplicación al trabajar con expresiones de la forma

$$f\left(\sum_{i=1}^L I^{\vartheta_i} \{E(n_i), p(\vartheta_i)\}\right) = \varphi(n_1, n_2, \dots, n_L)$$

siendo f una función cualquiera, que responden a una expresión generalizada del valor esperado del experimento $E^S(\underline{n})$. En efecto, y como se

comprobará inmediatamente, la optimización respecto al vector n de la expresión anterior conduciría al sistema de ecuaciones

$$\frac{d I^{\vartheta_i} \{E(n_i), p(\vartheta_i)\}}{dn_i} \cdot f'(x) = \frac{\partial \psi(n_1, n_2, \dots, n_L)}{\partial n_i} \quad i=1, 2, \dots, L$$

donde $x = \sum_{i=1}^L I^{\vartheta_i} \{E(n_i), p(\vartheta_i)\}$.

Teniendo en cuenta la igualdad mencionada en el capítulo primero y válida en todo estrato

$$I^{\vartheta_i} \{E(n_i), p(\vartheta_i)\} = H(p(\vartheta_i)) - E_{z_i(n_i)} H(p(\vartheta_i/z_i(n_i)))$$

a partir de la que se definía precisamente a la información esperada, y en el supuesto de que $p(\underline{\vartheta})$ sea un producto de densidades de probabilidad definidas para cada componente ϑ_i del vector $\underline{\vartheta}$, puede escribirse que

$$I^{\underline{\vartheta}} \{E^s(\underline{n}), p(\underline{\vartheta})\} = \sum_{i=1}^L H(p(\vartheta_i)) - \sum_{i=1}^L E_{z_i(n_i)} H(p(\vartheta_i/z_i(n_i)))$$

donde, evidentemente, se verifica que la entropía de $p(\underline{\vartheta})$ es

$$H(p(\underline{\vartheta})) = \sum_{i=1}^L H(p(\vartheta_i))$$

Suponiendo aplicable el mismo modelo en todos los estratos y que las cantidades de interés ϑ_i son independientes, las expresiones que tomaría $I^{\underline{\vartheta}} \{E^s(\underline{n}), p(\underline{\vartheta})\}$ utilizando las informaciones esperadas desarrolladas en el capítulo segundo darían lugar a los ejemplos siguientes:

EJEMPLO 1º. Supongamos que la cantidad aleatoria x se distribuye en todos los estratos según una densidad uniforme en el intervalo $(0, \vartheta_i)$, $(i=1, 2, \dots, L)$. Supongamos también que las opiniones iniciales sobre ϑ_i vengan descritas, en el estrato i -ésimo, por una densidad de probabilidad de Pareto, con parámetros ϑ_{i0} y γ_i . Es conocido que

$$I^{\vartheta_i} \{E(n_i), p(\vartheta_i/\vartheta_{i0}, \gamma_i)\} = \log(1 + (n_i/\gamma_i))$$

$$H(p(\vartheta_i/\vartheta_{i0}, \gamma_i)) = \log(\vartheta_{i0}/\gamma_i) + (1 + (1/\gamma_i))$$

De ahí que la información esperada sobre el vector paramétrico $\underline{\vartheta}$ de componentes ϑ_i independientes obtenida a partir del experimento $E^S(\underline{n})$ y cuando las opiniones iniciales se describen por la densidad de probabilidad

$$p(\vartheta_1, \vartheta_2, \dots, \vartheta_L) = \prod_{i=1}^L \frac{\vartheta_i^{\nu_i} \cdot \vartheta_i}{\vartheta_i^{\nu_i+1}}$$

tome como expresión

$$I^{\underline{\vartheta}} \left\{ E^S(\underline{n}), \prod_{i=1}^L \frac{\vartheta_i^{\nu_i} \cdot \vartheta_i}{\vartheta_i^{\nu_i+1}} \right\} = \log \prod_{i=1}^L \left(1 + \frac{n_i}{\nu_i} \right) = \sum_{i=1}^L \log \left(1 + \frac{n_i}{\nu_i} \right), \text{ y}$$

$$H \left(\prod_{i=1}^L \frac{\vartheta_i^{\nu_i} \cdot \vartheta_i}{\vartheta_i^{\nu_i+1}} \right) = \sum_{i=1}^L \log \frac{\vartheta_{oi}}{\nu_i} + \sum_{i=1}^L \left(1 + \frac{1}{\nu_i} \right)$$

EJEMPLO 2º. Supongamos ahora que la cantidad aleatoria x se describe, en el estrato h -ésimo ($h=1,2,\dots,L$), por una $N(\vartheta_h, \sigma_h)$, con σ_h conocida en todos los estratos. Supongamos también que para representar las opiniones iniciales sobre ϑ_h se acepta una densidad de probabilidad $N(\vartheta_{oh}, \sigma_{oh})$. Es conocido (Lindley, 1956) que tras la realización en el estrato h -ésimo del experimento $E(n_h)$ la información esperada obtenida

$$\text{es } I^{\vartheta_h} \left\{ E(n_h), N(\vartheta_{oh}, \sigma_{oh}) \right\} = \frac{1}{2} \log \left(1 + n_h \cdot \frac{\sigma_{oh}^2}{\sigma_h^2} \right)$$

$$H(N(\vartheta_{oh}, \sigma_{oh})) = \frac{1}{2} \log(2\pi e \sigma_{oh}^2)$$

Así pues, la información esperada obtenida sobre $\underline{\vartheta}$ a partir del experimento $E^S(\underline{n})$ cuando las opiniones iniciales sobre el vector de interés se describen por la densidad

$$p(\vartheta_1, \vartheta_2, \dots, \vartheta_L) = \prod_{h=1}^L N(\vartheta_{oh}, \sigma_{oh})$$

admite como expresión

$$I^{\underline{\vartheta}} \left\{ E^S(\underline{n}), \prod_{h=1}^L N(\vartheta_{oh}, \sigma_{oh}) \right\} = \sum_{h=1}^L \frac{1}{2} \log \left(1 + n_h \cdot \frac{\sigma_{oh}^2}{\sigma_h^2} \right)$$

$$H\left(\prod_{h=1}^L N(\vartheta_{oh}, \sigma_{oh}^2)\right) = \frac{1}{2} \sum_{h=1}^L \log(2\pi e \sigma_{oh}^2)$$

EJEMPLO 3º. Sea ahora una combinación de los dos ejemplos anteriores. Sean, por tanto, $i=1,2,\dots,L_1$ los estratos en los que es aplicable el primer modelo; y sean $h=1,2,\dots,L_2$ los estratos en los que se aplica el modelo normal, con $L_1 + L_2 = L$. Dado que la información esperada sobre ϑ proporcionada por $E^S(\underline{n})$ y la correspondiente entropía son, bajo la hipótesis de independencia de las componentes ϑ_i del vector ϑ , a su vez suma de las informaciones esperadas y entropías, respectivamente, deducidas de cada experimento $E(n_i)$, es evidente el resultado

$$I^{\vartheta}\left\{E^S(\underline{n}), \prod_{i=1}^{L_1} \frac{\vartheta_i \cdot \vartheta_{i0}^{\vartheta_i}}{\vartheta_i^{\vartheta_i+1}} \cdot \prod_{h=1}^{L_2} \frac{1}{\sigma_{oh} \sqrt{2\pi}} e^{-\frac{1}{2\sigma_{oh}^2}(\vartheta_h - \vartheta_{oh}^2)^2}\right\} =$$

$$= \sum_{i=1}^{L_1} \log\left(1 + \frac{n_i}{\vartheta_i}\right) + \sum_{h=1}^{L_2} \frac{1}{2} \log\left(1 + n_h \cdot \frac{\sigma_{oh}^2}{\sigma_h^2}\right)$$

$$H\left(\prod_{i=1}^{L_1} \frac{\vartheta_i \cdot \vartheta_{i0}^{\vartheta_i}}{\vartheta_i^{\vartheta_i+1}} \cdot \prod_{h=1}^{L_2} \frac{1}{\sigma_{oh} \sqrt{2\pi}} e^{-\frac{1}{2\sigma_{oh}^2}(\vartheta_h - \vartheta_{oh}^2)^2}\right) =$$

$$= \sum_{i=1}^{L_1} \log(\vartheta_{oi}/\vartheta_i) + \sum_{i=1}^{L_1} \left(1 + \frac{1}{\vartheta_i}\right) + \frac{1}{2} \sum_{h=1}^{L_2} \log(2\pi e \sigma_{oh}^2)$$

Notamos en los tres ejemplos anteriores cómo el investigador puede suavizar la hipótesis de independencia de las cantidades de interés ϑ_i actuando sobre los parámetros ϑ_i y σ_{oh} de las densidades iniciales que describen el comportamiento a priori de las ϑ_i .

No hay que olvidar que el último objetivo a considerar sigue siendo el de encontrar una normativa de fácil comprensión para determinar un tamaño muestral óptimo. Dicha normativa es idéntica a la expresada en el capítulo cuarto y aplicada ya en la sección 5.3, actuando al maximizar el valor esperado del experimento $E^S(\underline{n})$ como diferencia entre su utilidad y costo esperados. Por ahora se dispone de la información esperada so-

bre ϑ proporcionada por $E^S(\underline{n})$, esto es, la información $I^{\vartheta}\{E^S(\underline{n}), p(\vartheta)\}$ definida considerando las distintas cantidades de interés ϑ_i en su conjunto. A partir de ella se obtiene la utilidad esperada de $E^S(\underline{n})$ con las distintas expresiones definidas en la sección 3.1. En este sentido y como vía abierta a una ulterior investigación está la confirmación de la unicidad de la expresión logarítmica para medir esta utilidad esperada en la línea del resultado encontrado por Bernardo (1979). Por otra parte, teniendo en cuenta las propiedades satisfechas por $I^{\vartheta}\{E^S(\underline{n}), p(\vartheta)\}$ y siguiendo idénticas técnicas de demostración que las utilizadas en el capítulo tercero, pueden demostrarse para las distintas $U_i(E^S(\underline{n}))$ las mismas propiedades que las contenidas en la sección 3.2. Lo importante, en nuestro caso, es confrontar esta utilidad esperada $U(E^S(\underline{n}))$ con su costo esperado $C(E^S(\underline{n}))$. Su diferencia es, según ya se ha definido, el valor esperado $W(E^S(\underline{n}))$ del experimento $E^S(\underline{n})$ asociado a un proceso de estratificación. Además se trata de una expresión que es función del vector tamaño muestral \underline{n} . Las distintas expresiones que tomaría $W(E^S(\underline{n}))$ en los ejemplos ahora mismo mencionados, con un costo genérico $C(E^S(\underline{n}))$ y bajo la hipótesis de independencia de las cantidades de interés ϑ_i , serían las siguientes para un modelo de densidad inicial de Pareto y poblaciones uniformes:

$$W_1(E^S(\underline{n})) = g_1 \cdot \sum_{i=1}^L \log\left(1 + \frac{n_i}{\sigma_i}\right) - C(E^S(\underline{n}))$$

$$W_2(E^S(\underline{n})) = g_2 \cdot \left(1 - \frac{1}{\prod_{i=1}^L \left(1 + \frac{n_i}{\sigma_i}\right)}\right) - C(E^S(\underline{n}))$$

$$W_3(E^S(\underline{n})) = g_3 \cdot \left(\prod_{i=1}^L \left(1 + \frac{n_i}{\sigma_i}\right) - 1\right) / \prod_{i=1}^L \frac{\theta_{oi}}{\gamma_i} \cdot \exp\left(\sum_{i=1}^L \left(1 + \frac{1}{\sigma_i}\right)\right) - C(E^S(\underline{n}))$$

mientras que para el modelo normal resultarían:

$$W_1(E^S(\underline{n})) = \frac{1}{2}g_1 \cdot \sum_{i=1}^L \log\left(1 + n_i \cdot \frac{\sigma_{oi}^2}{\sigma_i^2}\right) - C(E^S(\underline{n}))$$

$$W_2(E^S(\underline{n})) = g_2 \cdot \left(1 - \frac{1}{\prod_{i=1}^L \sqrt{1 + n_i \cdot \frac{\sigma_{oi}^2}{\sigma_i^2}}}\right) - C(E^S(\underline{n}))$$

$$W_3(E^S(\underline{n})) = g_3 \cdot \left(\frac{L}{\prod_{i=1}^L \sqrt{1 + n_i \cdot \frac{\sigma_{oi}^2}{\sigma_i^2}}} - 1\right) / \prod_{i=1}^L \sqrt{2\pi e \sigma_{oi}^2} - C(E^S(\underline{n}))$$

donde las constantes g_i ($i=1,2,3$) admiten idéntica interpretación que la indicada en la sección 3.1, aunque destacando que, a semejanza de lo ocurrido en dicha sección, responden a propiedades o comportamientos satisfechos por unidades de la población considerada en su conjunto, independientemente de la estratificación efectuada. Además, es de destacar que cualquier criterio de optimización de las expresiones $W_i(E^S(\underline{n}))$ conduce, por una parte, a la maximización de la utilidad esperada del experimento y, por ello, a la maximización de la información esperada que proporciona, considerada ésta sobre el vector paramétrico $\underline{\theta}$ en su conjunto; y, por otra, minimiza el costo esperado de $E^S(\underline{n})$. En definitiva, el tamaño muestral óptimo se reduce a la solución de compromiso derivada de alcanzar la mayor información al costo más pequeño posible, lo que es equivalente a afirmar que es la solución que proporciona la mayor utilidad esperada al menor costo posible.

Esta es el momento de diferenciar sobre si la afijación de la estratificación es o no conocida previamente. Como ya se indicó en los dos primeros casos de la sección 5.3 el conocimiento previo o no de la afijación se traduce en que la expresión de $W_i(E^S(\underline{n}))$ ($i=1,2,3$) dependerá de la única variable n tamaño muestral total ó de las L variables n_i tamaños muestrales parciales, respectivamente. Supongamos primeramente que existe la familia de constantes $\{f_i\}_{i=1,2,\dots,L}$ verificando las L condiciones

$$n_i = f_i \cdot n \quad i=1,2,\dots,L$$

para cualquier tamaño muestral total. Esto implica que dado un tamaño total n la adscripción de esas unidades en los distintos estratos queda perfectamente determinada. De ahí que las expresiones $W_i(E^S(\underline{n}))$ ($i=1,2,3$) dependan exclusivamente de n , sin más que sustituir en ellas los valores de n_i por los de $n \cdot f_i$. Así pues la condición de optimización se traduce

en este caso en la ecuación

$$\frac{d W(E^S(n))}{dn} = 0$$

que es equivalente a la

$$\frac{d U(E^S(n))}{dn} = \frac{d C(E^S(n))}{dn}$$

y en donde puede apreciarse que el tamaño muestral total óptimo supone una solución de compromiso entre la utilidad y el costo esperados del experimento $E^S(n)$ como ya apuntábamos con anterioridad. En definitiva, el tamaño muestral total óptimo será el entero no negativo que maximizando a $W(E^S(n))$ sea el más cercano a la solución de la anterior ecuación. Conocido dicho tamaño, la afijación óptima se obtendrá sin más que emplear las L condiciones $n_i = f_i \cdot n$ ($i=1,2,\dots,L$). De ahí que en este caso de afijación previamente conocida el experimento asociado al proceso de estratificación pase a denotarse por $E^S(n)$ para indicar que depende exclusivamente del tamaño muestral total n , toda vez que el vector tamaño muestral queda perfectamente determinado al conocer a aquél.

Con total analogía con el resultado demostrado en el teorema 5.3.1 (ya se demostró con completa generalidad entonces) la condición $d W(E^S(n))/dn = 0$ puede sustituirse considerando que $W(E^S(n))$ es función de L variables n_i pero sujetas a las L exigencias derivadas de la existencia de una afijación preestablecida. Así pues, $d W(E^S(n))/dn = 0$ es equivalente, en este caso, a la solución del sistema

$$\frac{\partial \phi}{\partial n_i} = 0 \quad i=1,2,\dots,L$$

donde

$$\phi = U(E^S(n)) - C(E^S(n)) + \sum_{i=1}^L \lambda_i \cdot (n_i - n \cdot f_i)$$

tal como quedó demostrado en el teorema citado.

En el caso de que la afijación fuese desconocida el investigador deberá determinar todos los tamaños parciales n_i cuya suma constituirá, a posteriori, el tamaño muestral total óptimo. Así pues, en este caso el investigador debe encontrar una afijación óptima correspondiente a un

tamaño muestral óptimo único. En este caso, la condición de maximización del valor esperado $W(E^S(\underline{n}))$ conduce al sistema de L ecuaciones

$$\frac{\partial W(E^S(\underline{n}))}{\partial n_i} = 0 \quad i=1,2,\dots,L \quad \underline{n} = (n_1, n_2, \dots, n_L)$$

que es equivalente al

$$\frac{\partial U(E^S(\underline{n}))}{\partial n_i} = \frac{\partial C(E^S(\underline{n}))}{\partial n_i} \quad i=1,2,\dots,L$$

y en donde puede apreciarse nuevamente que cada tamaño muestral parcial óptimo n_i se obtiene como solución de compromiso entre la utilidad y costo esperados que proporciona el experimento $E^S(\underline{n})$ del que dicho tamaño parcial n_i forma parte como componente del vector tamaño muestral \underline{n} .

Suponiendo cierta la hipótesis de independencia para las componentes θ_i del vector paramétrico $\underline{\theta}$, las distintas expresiones que tomaría el valor esperado $W(E^S(\underline{n}))$ según se utilice una u otra función de la información esperada serían:

$$W_1(E^S(\underline{n})) = g_1 \cdot \sum_{i=1}^L I^{\theta_i} \{E(n_i), p(\theta_i)\} - C(E^S(\underline{n}))$$

$$W_2(E^S(\underline{n})) = g_2 \cdot \left(1 - \frac{1}{\exp\left(\sum_{i=1}^L I^{\theta_i} \{E(n_i), p(\theta_i)\}\right)} \right) - C(E^S(\underline{n}))$$

$$W_3(E^S(\underline{n})) = g_3 \cdot e^{-\sum_{i=1}^L H(p(\theta_i))} \cdot \left(e^{\sum_{i=1}^L I^{\theta_i} \{E(n_i), p(\theta_i)\}} - 1 \right) - C(E^S(\underline{n}))$$

toda vez que la información esperada sobre $\underline{\theta}$ proporcionada por $E^S(\underline{n})$ se reduce a la suma de informaciones esperadas sobre cada θ_i proporcionadas por los experimentos componentes $E(n_i)$. Y análogamente, la entropía de la densidad inicial $p(\underline{\theta})$ es también suma de las entropías de las densidades componentes $p(\theta_i)$. De ahí que la última condición de optimización expresada por un sistema de L ecuaciones conduzca, a su vez, a los tres sistemas de ecuaciones resultados de la aplicación de sendas expresiones para $W(E^S(\underline{n}))$:

$$g_1 \cdot \frac{d I^{\vartheta_i} \{E(n_i), p(\vartheta_i)\}}{dn_i} = \frac{\partial C(E^S(\underline{n}))}{\partial n_i} \quad i=1,2,\dots,L$$

$$g_2 \cdot \frac{d I^{\vartheta_i} \{E(n_i), p(\vartheta_i)\}}{dn_i} \cdot \exp\left(-\sum_{i=1}^L I^{\vartheta_i} \{E(n_i), p(\vartheta_i)\}\right) = \frac{\partial C(E^S(\underline{n}))}{\partial n_i} \quad i=1,\dots,L$$

$$g_3 \cdot \frac{d I^{\vartheta_i} \{E(n_i), p(\vartheta_i)\}}{dn_i} \cdot \exp\left(-\sum_{i=1}^L \left[H(p(\vartheta_i)) - I^{\vartheta_i} \{E(n_i), p(\vartheta_i)\} \right] \right) =$$

$$= \frac{\partial C(E^S(\underline{n}))}{\partial n_i} \quad i=1,2,\dots,L$$

Como ejemplo de aplicación de las ideas anteriores vamos a calcular para los dos modelos desarrollados en el capítulo segundo y generalizados al caso de una población estratificada, las expresiones satisfechas por el tamaño muestral total óptimo (en el caso de que la afijación se conozca previamente) o por el vector tamaño muestral óptimo (si la afijación no es conocida de antemano), cuando el valor esperado del experimento viene dado a través de su primera expresión $W_1(E^S(\underline{n}))$, y suponiendo un costo lineal dado a través del sumatorio

$$C(E^S(\underline{n})) = \sum_{i=1}^L c_i \cdot n_i$$

donde c_i es el costo unitario de observación de una unidad del estrato i -ésimo. Supondremos finalmente la independencia de las cantidades de interés ϑ_i .

TEOREMA 5.4.11

Cuando el valor esperado del experimento $E^S(\underline{n})$ viene dado por el sumatorio

$$W(E^S(\underline{n})) = g_1 \cdot \sum_{i=1}^L I^{\vartheta_i} \{E(n_i), p(\vartheta_i)\} - \sum_{i=1}^L c_i \cdot n_i$$

satisfaciéndose las L condiciones para cualquier tamaño muestral total n

$$n_i = n \cdot f_i \quad i=1,2,\dots,L$$

resulta que los tamaños muestrales totales óptimos son los enteros no negativos que maximizando $W(E^S(\underline{n}))$ son los más cercanos a las soluciones de las ecuaciones

$$a) \quad g_1 \cdot \sum_{i=1}^L \frac{f_i}{\sigma_i + n \cdot f_i} = \sum_{i=1}^L c_i \cdot f_i$$

cuando se trata del modelo de densidades iniciales de Pareto y estratos uniformes;

$$b) \quad \frac{1}{2} g_1 \cdot \sum_{i=1}^L \frac{f_i \cdot \sigma_{oi}^2}{\sigma_i^2 + n \cdot f_i \cdot \sigma_{oi}^2} = \sum_{i=1}^L c_i \cdot f_i$$

en el modelo normal.

Demostración.-

a) La función ϕ toma la forma

$$\phi = g_1 \cdot \sum_{i=1}^L \log\left(1 + \frac{n \cdot f_i}{\sigma_i}\right) - \sum_{i=1}^L c_i \cdot n_i + \sum_{i=1}^L \lambda_i \cdot (n_i - n \cdot f_i)$$

De ahí que las condiciones

$$\frac{\partial \phi}{\partial n_i} = 0 \quad (i=1,2,\dots,L); \quad \frac{\partial \phi}{\partial n} = 0; \quad \frac{\partial \phi}{\partial \lambda_i} = 0 \quad (i=1,2,\dots,L)$$

conduzcan al sistema

$$\frac{g_1}{\sigma_i + n_i} - c_i + \lambda_i = 0 \quad (i=1,2,\dots,L); \quad \sum_{i=1}^L \lambda_i \cdot f_i = 0; \quad n_i - f_i \cdot n = 0 \quad (i=1,2,\dots,L)$$

De donde multiplicando la primera ecuación por f_i y sumando para todos los estratos resulta la condición a) de la tesis del teorema, como condición satisfecha por el tamaño muestral total óptimo.

b) En el modelo normal, la expresión a optimizar por máximos condicionados resulta ser

$$\phi = \frac{1}{2} g_1 \cdot \sum_{i=1}^L \log\left(1 + n_i \cdot \frac{\sigma_{oi}^2}{\sigma_i^2}\right) - \sum_{i=1}^L c_i \cdot n_i + \sum_{i=1}^L \lambda_i \cdot (n_i - n \cdot f_i)$$

De ahí que las mismas condiciones indicadas en el anterior apartado conduzan al sistema

$$\frac{g_1 \cdot \sigma_{oi}^2}{2(\sigma_i^2 + n_i \cdot \sigma_{oi}^2)} - c_i + \lambda_i = 0 \quad (i=1,2,\dots,L) ; \quad \sum_{i=1}^L \lambda_i \cdot f_i = 0 ;$$

$$n_i - n \cdot f_i = 0 \quad (i=1,2,\dots,L)$$

Y efectuando idéntica operación que en el caso anterior resulta finalmente la ecuación que constituye la condición b) de la tesis del teorema, condición satisfecha por el tamaño muestral total óptimo en el modelo normal.

(c.q.d.)

TEOREMA 5.4.12

Suponiendo que el valor esperado del experimento $E^S(\underline{n})$ viene dado a través del sumatorio

$$W(E^S(\underline{n})) = g_1 \cdot \sum_{i=1}^L I^{\vartheta_i} \{E(n_i), p(\vartheta_i)\} - \sum_{i=1}^L c_i \cdot n_i$$

los vectores tamaños muestrales óptimos tienen como componentes los enteros no negativos que maximizando a $W(E^S(\underline{n}))$ sean los más cercanos a las soluciones

$$a) \quad n_i = \frac{g_1}{c_i} - \vartheta_i \quad i=1,2,\dots,L$$

$$b) \quad n_i = (g_1/2c_i) - (\sigma_i^2 / \sigma_{oi}^2) \quad i=1,2,\dots,L$$

donde la solución a) es válida para el modelo de densidades iniciales de Pareto y estratos uniformes; y la solución b) es válida en el modelo normal.

Demostración.- Las condiciones

$$g_1 \cdot \frac{d I^{\vartheta_i} \{E(n_i), p(\vartheta_i)\}}{dn_i} = \frac{\partial C(E^S(\underline{n}))}{\partial n_i} \quad i=1,2,\dots,L$$

conducen a:

$$a) \quad g_1 \cdot \frac{1}{\vartheta_i + n_i} = c_i \quad i=1,2,\dots,L$$

sistema que proporciona los tamaños muestrales parciales óptimos como los enteros no negativos que maximizando a $W(E^S(\underline{n}))$ son los más cercanos a

$$(g_1/c_i) = \gamma_i \quad i=1,2,\dots,L$$

lo que supone en este caso en el que el valor esperado del experimento viene expresado a través de $W_1(E^S(\underline{n}))$ una total coincidencia con el resultado obtenido en la sección 4.2 .

b) En el modelo normal el sistema de ecuaciones a considerar es

$$\frac{1}{2}g_1 \cdot \frac{\sigma_{oi}^2}{\sigma_i^2 + n \cdot \sigma_{oi}^2} = c_i \quad i=1,2,\dots,L$$

que proporciona como tamaños muestrales parciales óptimos los enteros no negativos que maximizando a $W(E^S(\underline{n}))$ sean los más cercanos a

$$(g_1/2c_i) = (\sigma_i / \sigma_{oi})^2 \quad i=1,2,\dots,L$$

resultado que también coincide plenamente con el obtenido en el capítulo cuarto al venir expresado el valor del experimento $E^S(\underline{n})$ simplemente como suma de los valores esperados de los experimentos componentes $E(n_i)$.

(c.q.d.)

5.5 INFORMACION ESPERADA $I^{\theta_1}\{E^S(\underline{n}), p(\underline{\theta})\}$

La última de las actitudes que puede tomar el investigador al enfrentarse con el experimento $E^S(\underline{n})$ asociado a un proceso de estratificación consiste en considerar la información esperada que dicho experimento proporciona sobre cada una de las cantidades de interés θ_i . Efectivamente, el resultado experimental z de $E^S(\underline{n})$ está constituido por un conjunto de submuestras z_i ($i=1,2,\dots,L$) obtenidas en sendos estratos y que son los resultados de los experimentos componentes $E(n_i)$. Pero cada una de estas submuestras no sólo contienen información sobre su parámetro respectivo sino que también contienen información sobre las cantidades de interés con ámbito de aplicación en los demás estratos, y ello debido a que dichas cantidades θ_i no son necesariamente independientes. Así pues, tienen sentido las densidades $p(\theta_i/z)$ que permiten al investigador centrar la investigación sobre la información sobre cada θ_i por separado contenida en el experimento $E^S(\underline{n})$, disponiendo por tanto de la familia de

L informaciones esperadas que ahora mismo definiremos

$$\left\{ I^{\theta_i} \left\{ E^S(\underline{n}), p(\underline{\vartheta}) \right\} \right\}_{i=1,2,\dots,L}$$

Cada una de las informaciones anteriores cobra su verdadero sentido cuando el investigador está interesado únicamente en adquirir conocimiento sobre una cierta θ_i exclusivamente, no interesándole las demás cantidades θ_j ($j \neq i$), e, incluso, el parámetro global ϑ deducible de todos los parámetros θ_i a través de la relación $\theta = f(\underline{\vartheta})$. En tales casos y realizado el experimento $E^S(\underline{n})$ el investigador deberá trabajar con la información que dicho experimento proporciona sobre la única cantidad de interés θ_i , información contenida en la muestra total z resultado de su realización. Las circunstancias que pueden motivar al investigador a utilizar la información que todo un experimento tan complejo como es el $E^S(\underline{n})$ proporciona sobre una única cantidad θ_i pueden proceder, en muchos casos, de la existencia de una muestra estratificada existente antes de haber surgido la necesidad de información sobre θ_i , por lo que puede aprovecharse sin recurrir al diseño de un nuevo experimento. Esta actitud se ve avalada, por otra parte, por el resultado que posteriormente demostraremos y que afirma que cualquiera de las informaciones esperadas de la familia anterior no es menor que la información proporcionada sobre la respectiva cantidad θ_i por el respectivo experimento componente $E(n_i)$ de $E^S(\underline{n})$. De ahí que la utilización de una muestra estratificada global previamente obtenida quede plenamente justificada frente al empleo de la respectiva muestra parcial $z_i(n_i)$ de la muestra estratificada global. Y todo ello, evidentemente, en ausencia de costo. Estudiemos, por tanto, la información que el experimento $E^S(\underline{n})$ proporciona sobre la cantidad de interés θ_i por separado, cuando ésta es efectivamente la única cantidad de interés para el investigador.

A fin de obtener la información esperada sobre θ_i proporcionada por $E^S(\underline{n})$ el investigador parte de las densidades conocidas $p(\underline{\vartheta})$ y $p(z/\underline{\vartheta})$. La segunda de ellas - definida con exhaustividad en la sección 5.1 - describe el comportamiento de la muestra z en función del vector paramétrico $\underline{\vartheta}$ y se expresa como producto de densidades conocidas $p(x_{ij}/\vartheta_i)$ que intervienen en la definición de los distintos $E(n_i)$. Re-

sultará:

$$p(z/\underline{\vartheta}) = \prod_{i=1}^L \prod_{j=1}^{n_i} p(x_{ij}/\vartheta_i) = \prod_{i=1}^L p(z_i(n_i)/\vartheta_i)$$

con

$$z_i(n_i) = (x_{i1}, x_{i2}, \dots, x_{in_i}) \quad (i=1, 2, \dots, L)$$

Así pues, $p(z/\underline{\vartheta})$ dependerá de los distintos tamaños muestrales parciales n_i que constituyen el vector tamaño muestral \underline{n} . Por otra parte, la densidad $p(\underline{\vartheta})$ completa la especificación Bayesiana del problema al ser la densidad de probabilidad que describe las opiniones iniciales que el investigador tiene sobre el vector $\underline{\vartheta}$. A partir de la información contenida en la muestra $z = (z_1(n_1), z_2(n_2), \dots, z_L(n_L))$ las opiniones iniciales sufren una modificación debido a la información que sobre $\underline{\vartheta}$ contiene z . De ahí que las opiniones posteriores se describan por la nueva densidad posterior deducida de la condición de Bayes

$$p(\underline{\vartheta}/z) \propto p(\underline{\vartheta}) \cdot p(z/\underline{\vartheta})$$

siendo el factor de proporcionalidad $1/p(z)$, con

$$p(z) = \int p(\underline{\vartheta}) \cdot p(z/\underline{\vartheta}) \cdot d\underline{\vartheta}$$

Nuevamente a partir de las densidades anteriores y cuando el interés del investigador está centrado en la cantidad ϑ_i , la cuantificación de la información esperada que sobre ella proporciona el experimento $E^S(\underline{n})$ la realizará a través de las densidades marginales

$$p(\vartheta_i) = \int_{\underline{\vartheta}'} p(\underline{\vartheta}) \cdot d\underline{\vartheta}' \quad , \quad p(\vartheta_i/z) = \int_{\underline{\vartheta}'} p(\underline{\vartheta}/z) \cdot d\underline{\vartheta}'$$

donde

$$\underline{\vartheta}' = \prod_{j \neq i}^L \vartheta_j \quad \text{y} \quad d\underline{\vartheta}' = \prod_{j \neq i}^L d\vartheta_j$$

Evidentemente que la utilización de estas densidades asegura la coherencia del investigador a describir tanto sus opiniones iniciales como finales sobre ϑ_i , teniendo en cuenta las densidades que describen dichas opiniones sobre todo el vector $\underline{\vartheta}$ y del que ϑ_i no es sino una de sus componentes. Conocidas ambas densidades y empleando la formulación tan conocida de Lindley (1956), puede definirse la información esperada que sobre ϑ_i

proporciona el experimento $E^S(\underline{n})$ a través de la conocida doble integral

$$I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\} = \iint p(z) \cdot p(\vartheta_i/z) \cdot \log \frac{p(\vartheta_i/z)}{p(\vartheta_i)} \cdot d\vartheta_i \cdot dz$$

que admite como expresiones alternativas las

$$\begin{aligned} \iint p(\vartheta_i) \cdot p(z/\vartheta_i) \cdot \log \frac{p(\vartheta_i/z)}{p(\vartheta_i)} \cdot d\vartheta_i \cdot dz &= \\ = \iint p(\vartheta_i, z) \cdot \log \frac{p(z/\vartheta_i)}{p(z)} \cdot d\vartheta_i \cdot dz \end{aligned}$$

donde, según se ha indicado, $p(\vartheta_i)$ y $p(\vartheta_i/z)$ son las correspondientes marginales de las densidades $p(\underline{\vartheta})$ y $p(\underline{\vartheta}/z)$ y la densidad $p(z)$ es conocida. Como ya se hizo en su momento, es interesante destacar dos circunstancias en la doble integral anterior. Por una parte, nuevamente $I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\}$ es función del vector tamaño muestral \underline{n} puesto que sus distintas componentes n_i aparecen en la obtención de la densidad $p(\underline{\vartheta}/z)$ y, por ello, en $p(\vartheta_i/z)$. Este hecho permitirá utilizarla en criterios de optimización como los aplicados en casos anteriores para determinar tamaños muestrales óptimos. Por otra parte, en la información anterior influye todo el resultado muestral z del experimento $E^S(\underline{n})$ en cuanto que $I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\}$ es una medida de la información que sobre ϑ_i contiene toda la muestra z resultado del experimento $E^S(\underline{n})$. Teniendo en cuenta la información esperada sobre ϑ_i proporcionada por el experimento componente $E(n_i)$ definida por la conocida doble integral (Lindley, 1956) :

$$I^{\vartheta_i} \{E(n_i), p(\vartheta_i)\} = \iint p(\vartheta_i) \cdot p(z_i/\vartheta_i) \cdot \log \frac{p(\vartheta_i/z_i)}{p(\vartheta_i)} \cdot d\vartheta_i \cdot dz_i$$

donde $p(\vartheta_i)$ es también la marginal i -ésima de $p(\underline{\vartheta})$, parece lógico afirmar que $I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\}$ debe ser mayor que $I^{\vartheta_i} \{E(n_i), p(\vartheta_i)\}$ puesto que para la primera información se investigan un mayor número de unidades de observación y todas las unidades investigadas por el experi-

mento componente $E(n_i)$ lo son, evidentemente, por $E^S(\underline{n})$. Esta afirmación será demostrada más adelante. Para ello, denotemos por $z = (z_1(n_1), \dots, z_L(n_L))$ la muestra resultado de $E^S(\underline{n})$, siendo

$$z' = (z_1(n_1), z_2(n_2), \dots, z_{i-1}(n_{i-1}), z_{i+1}(n_{i+1}), \dots, z_L(n_L))$$

Es evidente que las siguientes igualdades entre densidades son identidades:

$$p(\vartheta_i/z) = p(\vartheta_i/z', z_i(n_i)) \quad ; \quad p(z/\vartheta_i) = p(z', z_i(n_i)/\vartheta_i) \quad ;$$

$$p(z, \vartheta_i) = p(z', z_i(n_i), \vartheta_i) \quad .$$

Demostremos ya el siguiente

TEOREMA 5.5.1

Con la notación expresada, la información esperada sobre ϑ_i proporcionada por el experimento $E^S(\underline{n})$ no es menor que la información esperada sobre ϑ_i proporcionada por el experimento componente $E(n_i)$

Demostración.- De sus mismas definiciones

$$\begin{aligned} \Delta &= I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\} - I^{\vartheta_i} \{E(n_i), p(\vartheta_i)\} = \iint p(z) \cdot p(\vartheta_i/z) \cdot \log \frac{p(\vartheta_i/z)}{p(\vartheta_i)} \\ &\cdot d\vartheta_i \cdot dz - \iint p(\vartheta_i, z_i(n_i)) \cdot \log \frac{p(\vartheta_i/z_i(n_i))}{p(\vartheta_i)} \cdot d\vartheta_i \cdot dz_i(n_i) = \\ &= \iiint p(\vartheta_i, z', z_i(n_i)) \cdot \log \frac{p(\vartheta_i/z', z_i(n_i))}{p(\vartheta_i/z_i(n_i))} \cdot d\vartheta_i \cdot dz_i(n_i) \cdot dz' \end{aligned}$$

Llamando

$$g(\vartheta_i, z_i(n_i), z') = \frac{p(\vartheta_i/z', z_i(n_i))}{p(\vartheta_i/z_i(n_i))}$$

resulta

$$\Delta = \iint g(\vartheta_i, z', z_i(n_i)) \cdot \log g(\vartheta_i, z', z_i(n_i)) \cdot p(z) \cdot p(\vartheta_i/z_i(n_i)) \cdot d\vartheta_i \cdot dz$$

expresión que es no negativa por idéntica razón a la utilizada para demostrar el teorema 5.4.1 .

(c.q.d.)

Notemos que Δ no es sino la información esperada sobre ϑ_i proporcionada por la muestra adicional z' supuesto realizado $E(n_i)$ y

obtenido $z_i(n_i)$. Aplicando, pues, el teorema 5.4.1 puede afirmarse que esta información se anula si y sólo si la densidad $p(\vartheta_i/z)$ no varía respecto a la $p(\vartheta_i/z_i(n_i))$ salvo en un subconjunto de ϑ_i de medida dominante nula. En definitiva, deducimos que las informaciones esperadas $I^{\vartheta_i}\{E^S(\underline{n}), p(\underline{\vartheta})\}$ e $I^{\vartheta_i}\{E(n_i), p(\vartheta_i)\}$ coinciden si y solamente si se verifica la igualdad entre las densidades de probabilidad anteriores. Esto equivale, a su vez, a afirmar que ambas informaciones coincidirán cuando la densidad $p(\vartheta_i/z)$ sea independiente del resultado muestral z' salvo, a lo sumo, en un subconjunto de ϑ_i de medida nula. Sobre este resultado, de gran interés, volveremos más adelante.

El teorema 5.5.1 pone de manifiesto un importante hecho: el diseño de un proceso de estratificación es, por término medio, útilmente informativo a la hora de recabar información sobre una determinada cantidad de interés ϑ_i . Así pues, el resultado anterior justifica el diseño de una estratificación a fin de adquirir conocimiento sobre una única cantidad de interés ϑ_i . Sin embargo es evidente que, en este caso, la decisión última deberá tomarse en términos de los respectivos valores esperados de los experimentos $E(n_i)$ y $E^S(\underline{n})$ y comparar el incremento de información con un posible incremento del costo asociado a la existencia de los estratos. No obstante y según ya se afirmó, el diseño de $E^S(\underline{n})$ tiene especial sentido cuando el investigador está interesado en todas las cantidades ϑ_i y, consiguientemente, en el parámetro $\underline{\vartheta}$. De ahí que resulte interesante comparar cualquiera de las informaciones $I^{\vartheta_i}\{E^S(\underline{n}), p(\underline{\vartheta})\}$ con la $I^{\underline{\vartheta}}\{E^S(\underline{n}), p(\underline{\vartheta})\}$ que expresa la información esperada que sobre el conjunto de valores ϑ expresado a través del vector paramétrico $\underline{\vartheta}$ contiene el experimento $E^S(\underline{n})$. Demostremos el siguiente teorema de comparación:

TEOREMA 5.5.2

La información esperada $I^{\vartheta_i}\{E^S(\underline{n}), p(\underline{\vartheta})\}$ es la información esperada útil (en el sentido definido por Bernardo, (1978) y expresado en la sección 1.4) de la información $I^{\underline{\vartheta}}\{E^S(\underline{n}), p(\underline{\vartheta})\}$ por la transformación no biyectiva $\vartheta_i = \vartheta_i(\underline{\vartheta})$.

Demostración.- La transformación no biyectiva $\vartheta_i = \vartheta_i(\underline{\vartheta})$ puede completarse con las identidades

$$\begin{aligned}
 \alpha_1 &= \vartheta_1 \\
 \alpha_2 &= \vartheta_2 \\
 &\dots \\
 \alpha_{i-1} &= \vartheta_{i-1} \\
 \vartheta_i &= \vartheta_i \\
 \alpha_{i+1} &= \vartheta_{i+1} \\
 &\dots \\
 \alpha_L &= \vartheta_L
 \end{aligned}$$

siendo, trivialmente, la correspondiente transformación inversa

$$\begin{aligned}
 \vartheta_1 &= \alpha_1 \\
 \vartheta_2 &= \alpha_2 \\
 &\dots \\
 \vartheta_{i-1} &= \alpha_{i-1} \\
 \vartheta_i &= \vartheta_i \\
 \vartheta_{i+1} &= \alpha_{i+1} \\
 &\dots \\
 \vartheta_L &= \alpha_L
 \end{aligned}$$

cuyo jacobiano es

$$J_1 = \frac{\partial(\vartheta_1, \vartheta_2, \dots, \vartheta_L)}{\partial(\alpha_1, \alpha_2, \dots, \alpha_{i-1}, \vartheta_i, \alpha_{i+1}, \dots, \alpha_L)} = \begin{vmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{vmatrix}_{L \times L} = 1$$

De ahí que las densidades de probabilidad inicial y posterior de las nuevas variables aleatorias $(\alpha_1, \alpha_2, \dots, \alpha_{i-1}, \vartheta_i, \alpha_{i+1}, \dots, \alpha_L)$ vendrán dadas por las funciones

$$\begin{aligned}
 |J_1| \cdot p(\alpha_1, \dots, \alpha_{i-1}, \vartheta_i, \alpha_{i+1}, \dots, \alpha_L) &= p_{\vartheta}(\vartheta_1, \vartheta_2, \dots, \vartheta_L) = p_{\vartheta}(\vartheta) \\
 |J_1| \cdot p(\alpha_1, \dots, \alpha_{i-1}, \vartheta_i, \alpha_{i+1}, \dots, \alpha_L / z) &= p_{\vartheta}(\vartheta_1, \dots, \vartheta_L / z) = p_{\vartheta}(\vartheta / z)
 \end{aligned}$$

donde con la notación p_{ϑ} indicamos las densidades de probabilidad conocidas del vector ϑ . De ahí que las densidades $p(\vartheta_i)$ y $p(\vartheta_i / z)$ vengan

dadas por las correspondientes marginales de las densidades anteriores, es to es, a través de

$$p(\vartheta_i) = \int_{\underline{\vartheta}} p(\underline{\vartheta}) \cdot d\underline{\vartheta}' \quad ; \quad p(\vartheta_i/z) = \int_{\underline{\vartheta}} p(\underline{\vartheta}/z) \cdot d\underline{\vartheta}'$$

donde

$$\underline{\vartheta}' = \prod_{j \neq i}^L \vartheta_j \quad ; \quad d\underline{\vartheta}' = \prod_{j \neq i}^L d\vartheta_j$$

En definitiva, por su misma definición la información esperada útil de la $I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\}$ tras la transformación $\vartheta_i = \vartheta_i(\underline{\vartheta})$ del vector $\underline{\vartheta}$ no es sino la información $I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\}$ definida con anterioridad.
(c.q.d.)

Desde un punto de vista teórico nos encontramos, pues, en una situación completamente análoga a la planteada en la sección 5.4 al comparar las informaciones $I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\}$ con $I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\}$. Teniendo en cuenta las propiedades satisfechas por $I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\}$ puede afirmarse que, siguiendo idénticas técnicas de demostración para las respectivas propiedades que las empleadas por Bernardo (1978), la información esperada $I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\}$ sobre ϑ_i proporcionada por $E^S(\underline{n})$ es una función no negativa que solamente se anula cuando $p(\vartheta_i/z)$ es independiente del resultado z del experimento $E^S(\underline{n})$ salvo, a lo sumo, en un subconjunto de \mathcal{G}_i de medida nula, resultado que se corresponde con el expresado por el teorema 5.4.1 para $I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\}$. También $I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\}$ verifica un criterio de aditividad en la línea del demostrado para la información esperada sobre $\underline{\vartheta}$ en el teorema 5.4.2, al cumplir que $I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\} + I^{\vartheta_i} \{E^S(\underline{n}+1)/E^S(\underline{n}), p(\underline{\vartheta})\} = I^{\vartheta_i} \{E^S(\underline{n}+1), p(\underline{\vartheta})\}$. Considerada como función del vector tamaño muestral \underline{n} la información esperada $I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\}$ es creciente y cóncava, en correspondencia al resultado expresado por el teorema 5.4.3 para $I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\}$. Y, finalmente, el teorema 5.4.5 puede también aplicarse a nuestro caso pues su demostración está efectuada con una función $\vartheta = f(\underline{\vartheta})$ cualquiera, ob teniendo la desigualdad entre informaciones

$$I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\} \leq I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\}$$

que expresa que la información esperada sobre el vector $\underline{\vartheta}$ proporcionada por el experimento $E^S(\underline{n})$ no es menor que la que proporciona dicho expe-

rimento sobre cualquiera de las componentes ϑ_i del vector paramétrico. Notemos que todas las propiedades anteriores son válidas independientemente de la cantidad de interés ϑ_i escogida. Por ello y reuniendo los resultados anteriores aparece la siguiente cadena de desigualdades

$$I^{\vartheta_i} \{E(n_i), p(\vartheta_i)\} \leq I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\} \leq I^{\vartheta} \{E^S(\underline{n}), p(\underline{\vartheta})\} \quad i=1,2,\dots,L$$

de las que se deduce la desigualdad entre sumatorios

$$\sum_{i=1}^L I^{\vartheta_i} \{E(n_i), p(\vartheta_i)\} \leq \sum_{i=1}^L I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\} \leq L \cdot I^{\vartheta} \{E^S(\underline{n}), p(\underline{\vartheta})\}$$

sumatorios que aparecerán ahora mismo al hablar del proceso de optimización de las informaciones $I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\}$ a fin de determinar los tamaños muestrales óptimos.

Finalmente, dentro de las propiedades satisfechas por $I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\}$, existe una que verificada por la información esperada total no lo es por la información esperada útil: la que describe el comportamiento de la información esperada como funcional de la densidad inicial. Tal como se afirmó en la sección 1.4, si bien la información esperada total es cóncava como funcional de la densidad inicial, no lo es la información esperada útil, a no ser que se considere para un $p(\vartheta/\psi)$ fijo (Bernardo, 1978). Para $I^{\vartheta} \{E^S(\underline{n}), p(\underline{\vartheta})\}$ se demostró este carácter cóncavo en el teorema 5.4.4. Pues bien, debido precisamente a la forma de definir $I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\}$ en términos de densidades marginales de las funciones de probabilidad que intervienen en la definición de $I^{\vartheta} \{E^S(\underline{n}), p(\underline{\vartheta})\}$, sigue verificándose para la información esperada sobre ϑ_i proporcionada por $E^S(\underline{n})$ idéntico comportamiento cóncavo como funcional de la densidad $p(\underline{\vartheta})$, toda vez que lo es evidentemente respecto a la densidad particular $p(\vartheta_i)$. Demostremos, pues, el siguiente

TEOREMA 5.5.3

$I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\}$ es una funcional cóncava de $p(\underline{\vartheta})$.

Demostración.- El razonamiento seguido en el teorema 5.4.4 es válido en este caso sin más que considerar que la igualdad

$$p(\underline{\vartheta}) = \lambda p_1(\underline{\vartheta}) + (1-\lambda) \cdot p_2(\underline{\vartheta})$$

con

$$\int p_1(\underline{\vartheta}) \cdot d\underline{\vartheta} = \int p_2(\underline{\vartheta}) \cdot d\underline{\vartheta} = 1$$

conduce a

$$\int_{\underline{\vartheta}'} p(\underline{\vartheta}) \cdot d\underline{\vartheta}' = \lambda \int_{\underline{\vartheta}'} p_1(\underline{\vartheta}) \cdot d\underline{\vartheta}' + (1-\lambda) \int_{\underline{\vartheta}'} p_2(\underline{\vartheta}) \cdot d\underline{\vartheta}'$$

esto es, a

$$p(\vartheta_i) = \lambda p_1(\vartheta_i) + (1-\lambda) \cdot p_2(\vartheta_i) \quad i=1,2,\dots,L$$

(c.q.d.)

Situémonos nuevamente ante el problema específico planteado por la determinación del tamaño muestral óptimo. Para ello habrá que seguir considerando el valor esperado del experimento $E^S(\underline{n})$ como diferencia entre su utilidad y costo esperados, expresiones ambas definidas en función del vector tamaño muestral \underline{n} y que debidamente optimizadas proporcionan una expresión satisfecha por el tamaño óptimo. Una vez más se considerarán las funciones de utilidad definidas en la sección 3.1 como medidas de la utilidad esperada para $E^S(\underline{n})$. De ahí que empleando la conocida notación de $U(E^S(\underline{n}))$ y $C(E^S(\underline{n}))$ como expresiones de la utilidad y costo esperados, respectivamente, del experimento $E^S(\underline{n})$ pueda, por tanto, definirse su valor esperado como la diferencia ya conocida

$$W^i(E^S(\underline{n})) = U^i(E^S(\underline{n})) - C(E^S(\underline{n}))$$

donde $U^i(E^S(\underline{n}))$ es la utilidad esperada del experimento $E^S(\underline{n})$ realizado tan sólo para recabar información sobre la única cantidad de interés ϑ_i . De ahí que según se utilice una u otra de las funciones de la utilidad esperada, el valor esperado $W^i(E^S(\underline{n}))$ admita como expresiones concretas

$$W_1^i(E^S(\underline{n})) = g_{1i} \cdot I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\} - C(E^S(\underline{n}))$$

$$W_2^i(E^S(\underline{n})) = g_{2i} \cdot (1 - \exp(-I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\})) - C(E^S(\underline{n}))$$

$$W_3^i(E^S(\underline{n})) = g_{3i} \cdot \exp(-H(p(\vartheta_i))) (\exp(I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\}) - 1) - C(E^S(\underline{n}))$$

donde puede apreciarse que $W^i(E^S(\underline{n}))$ en sus distintas expresiones es, efectivamente, función de \underline{n} al serlo $I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\}$ y $C(E^S(\underline{n}))$. Notemos también que mientras las expresiones de $U(E^S(\underline{n}))$ están definidas sobre la información $I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\}$ y responden, por tanto, al

concepto de una utilidad esperada de la información proporcionada por $E^S(\underline{n})$ sobre una cantidad de interés ϑ_i de ámbito de aplicación en un único estrato, el costo esperado $C(E^S(\underline{n}))$ expresa, en cambio, el costo asociado a la realización de $E^S(\underline{n})$, tratándose por ello de un costo válido independientemente de la información $I^{\vartheta_i}\{E^S(\underline{n}), p(\vartheta)\}$ considerada. En cuanto a las constantes g_i que aparecen en las expresiones de las utilidades $U^i(E^S(\underline{n}))$ admiten idéntica interpretación que la indicada en sus respectivas definiciones de la sección 3.1 sin más que considerar que deben referirse a su parámetro respectivo ϑ_i . De ahí que deban distinguirse en función de dicho parámetro.

La aplicación del criterio de optimización expresado en la sección 4.1 conduce a idénticas exigencias sobre $W^i(E^S(\underline{n}))$ que las indicadas en las secciones 5.3 y 5.4. Y así, supuestamente conocida previamente la afijación de la estratificación, esto es, existiendo las ya conocidas L relaciones $n_i = n \cdot f_i$ ($i=1,2,\dots,L$) que adscriben un determinado tamaño global n en los distintos tamaños parciales n_i , la condición de optimización del valor esperado $E^S(\underline{n})$ se traduce en encontrar las soluciones de la ecuación

$$\frac{d W^i(E^S(\underline{n}))}{dn} = \frac{d U^i(E^S(\underline{n}))}{dn} - \frac{d C(E^S(\underline{n}))}{dn} = 0$$

donde con $E^S(\underline{n})$ queremos indicar que dicho experimento depende finalmente de la única variable n ó tamaño muestral total, al ser el vector tamaño muestral $\underline{n} = (n f_1, n f_2, \dots, n f_L)$. Así pues, el tamaño muestral total óptimo será el entero no negativo que maximizando a $W^i(E^S(\underline{n}))$ sea el más cercano a la solución de la ecuación anterior y del que, a partir de las relaciones previamente establecidas, se determinarán los tamaños muestrales parciales óptimos. Evidentemente, la condición anterior es equivalente a la resolución mediante mínimos condicionados del sistema expresado en el teorema 5.3.1, con la expresión ϕ a optimizar dada por

$$\phi = U^i(E^S(\underline{n})) - C(E^S(\underline{n})) + \sum_{j=1}^L \lambda_j (n_j - n \cdot f_j)$$

El tamaño muestral total óptimo así conseguido maximiza la utilidad esperada proporcionada por la información $I^{\vartheta_i}\{E^S(\underline{n}), p(\vartheta)\}$. Es evidente

que si consideramos la familia de valores esperados

$$\left\{ W^i(E^S(\underline{n})) \right\}_{i=1,2,\dots,L}$$

cada uno de ellos proporcionará un tamaño muestral total óptimo que, lógicamente, serán distintos entre sí. De ahí la imposibilidad de aplicar este criterio conjuntamente para obtener el tamaño muestral que optimice todas y cada una de las informaciones esperadas (o valores esperados) que todo el experimento $E^S(\underline{n})$ proporciona sobre todas y cada una de las cantidades de interés θ_i . Sin embargo, este resultado resulta también lógico en cuanto que por su misma definición cada información $I^i\{E^S(\underline{n}), p(\underline{\theta})\}$ se restringe a su parámetro θ_i respectivo que, por otra parte, es en este caso la única cantidad de interés para el investigador.

También con total analogía a los procesos de optimización en los que intervenía $I^{\theta}\{E^S(\underline{n}), p(\underline{\theta})\}$ e $I^{\underline{\theta}}\{E^S(\underline{n}), p(\underline{\theta})\}$, cuando la afijación no está determinada a priori la condición de optimización del valor esperado $W^i(E^S(\underline{n}))$ conduce al sistema de ecuaciones

$$\frac{\partial W^i(E^S(\underline{n}))}{\partial n_j} = \frac{\partial U^i(E^S(\underline{n}))}{\partial n_j} - \frac{\partial C(E^S(\underline{n}))}{\partial n_j} = 0 \quad j=1,2,\dots,L$$

satisfecho por el vector tamaño muestral óptimo, de forma que éste tendrá como componentes los enteros no negativos que maximizando $W^i(E^S(\underline{n}))$ sean los más cercanos a las respectivas componentes del vector solución del sistema anterior. Notemos nuevamente que la aplicación del criterio anterior sobre todos los valores esperados $W^i(E^S(\underline{n}))$ ($i=1,2,\dots,L$) proporciona un conjunto de L sistemas de L ecuaciones cada uno cuyas sendas soluciones no son necesariamente iguales y, por ello, compatibles. Sin embargo, la hipótesis de independencia de las componentes θ_i entre sí permite maximizar conjuntamente todos los valores esperados anteriores.

Así pues y en total semejanza con la sección 5.4, supongamos la existencia de independencia para las cantidades de interés θ_i . Todos los comentarios efectuados en su momento sobre esta restrictiva hipótesis siguen siendo válidos en estas circunstancias. En particular, existe una gran simplificación en las expresiones con las que se trabaja como pueda apreciarse ahora mismo mediante el siguiente

TEOREMA 5.5.4

Suponiendo que la densidad que describe las opiniones iniciales del investigador sobre el vector paramétrico $\underline{\vartheta}$ venga dada por un producto de densidades de sus componentes ϑ_i , se verifica que la información que sobre ϑ_i proporciona $E^S(\underline{n})$ es igual que la proporcionada por el respectivo experimento componente $E(n_i)$, siendo la densidad inicial $p(\vartheta_i)$ la marginal correspondiente de $p(\underline{\vartheta})$. Y este resultado es válido independientemente de la cantidad ϑ_i considerada.

Demostración.- Para una cantidad ϑ_i cualquiera y siendo

$$p(\underline{\vartheta}) = \prod_{i=1}^L p(\vartheta_i)$$

se verifican las siguientes igualdades entre densidades:

$$\begin{aligned} p(\vartheta_i/z) &= \int_{\underline{\vartheta}'} p(\underline{\vartheta}/z) \cdot d\underline{\vartheta}' = \frac{1}{p(z)} \int_{\underline{\vartheta}'} p(z/\underline{\vartheta}) \cdot p(\underline{\vartheta}) \cdot d\underline{\vartheta}' = \\ &= \frac{1}{p(z)} \int_{\underline{\vartheta}'} \prod_{j=1}^L p(z/\vartheta_j) \cdot \prod_{j=1}^L p(\vartheta_j) \cdot d\underline{\vartheta}' = \\ &= \frac{1}{p(z)} \int_{\underline{\vartheta}'} \prod_{j=1}^L p(z/\vartheta_j) \cdot p(\vartheta_j) \cdot d\underline{\vartheta}' = \\ &= \frac{1}{p(z)} \left\{ \prod_{j \neq i}^L \int_{\vartheta_j} p(z/\vartheta_j) \cdot p(\vartheta_j) \cdot d\vartheta_j \right\} p(z/\vartheta_i) \cdot p(\vartheta_i) = \\ &= \frac{1}{p(z)} \cdot p(z/\vartheta_i) \cdot p(\vartheta_i) \cdot \prod_{j \neq i}^L p(z_j) \quad ; \quad y \end{aligned}$$

$$p(z) = \int p(z/\underline{\vartheta}) \cdot p(\underline{\vartheta}) \cdot d\underline{\vartheta} = \prod_{i=1}^L \int_{\vartheta_i} p(z/\vartheta_i) \cdot p(\vartheta_i) \cdot d\vartheta_i = \prod_{i=1}^L p(z_i)$$

De donde

$$p(\vartheta_i/z) = p(z/\vartheta_i) \cdot p(\vartheta_i) / p(z_i) = p(\vartheta_i/z_i)$$

De ahí que

$$\begin{aligned}
I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\} &= \iint p(z) \cdot p(\vartheta_i/z) \cdot \log \frac{p(\vartheta_i/z)}{p(\vartheta_i)} \cdot d\vartheta_i \cdot dz = \\
&= \iint p(z) \cdot p(\vartheta_i/z_i) \cdot \log \frac{p(\vartheta_i/z_i)}{p(\vartheta_i)} \cdot d\vartheta_i \cdot dz = \\
&= \iint p(z_i) \cdot p(\vartheta_i/z_i) \cdot \log \frac{p(\vartheta_i/z_i)}{p(\vartheta_i)} \cdot d\vartheta_i \cdot dz_i = I^{\vartheta_i} \{E(n_i), p(\vartheta_i)\} \\
&\quad \text{(c.q.d.)}
\end{aligned}$$

Notemos que este mismo resultado está implícitamente contenido en el teorema 5.5.1 en el que se comprobaba la relación entre ambas informaciones. En efecto, al pie de dicho teorema afirmábamos que las informaciones esperadas $I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\}$ e $I^{\vartheta_i} \{E(n_i), p(\vartheta_i)\}$ coincidían si y sólo si las densidades $p(\vartheta_i/z)$ y $p(\vartheta_i/z_i)$ eran iguales. Dado que la muestra total está formada por las submuestras z_i ($i=1,2,\dots,L$) la afirmación anterior es equivalente a decir que ϑ_i es independiente del resultado muestral $z' = (z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_L)$ que es el que contiene información sobre el resto de parámetros ϑ_j , $j \neq i$. Si esta circunstancia es extendida a todos los parámetros deduciremos que cada muestra z_i sólo contiene información sobre su propio parámetro, lo que supone una implícita confirmación de la independencia de las cantidades ϑ_i . E inversamente, el resultado del teorema 5.5.4 es completamente lógico: aceptar que las ϑ_i son independientes presupone afirmar que todas las submuestras z_j , con $j \neq i$, no contienen información alguna sobre la cantidad ϑ_i , por lo que la información que sobre ella proporciona $E^S(\underline{n})$ debe coincidir con la proporcionada por su experimento componente $E(n_i)$.

Supuesta la independencia de las cantidades de interés ϑ_i es inmediato el siguiente

COROLARIO 5.5.1

Siendo $p(\underline{\vartheta}) = \prod_{i=1}^L p(\vartheta_i)$, se verifican las igualdades

$$\sum_{i=1}^L I^{\vartheta_i} \{E(n_i), p(\vartheta_i)\} = \sum_{i=1}^L I^{\vartheta_i} \{E^S(\underline{n}), p(\underline{\vartheta})\} = I^{\underline{\vartheta}} \{E^S(\underline{n}), p(\underline{\vartheta})\}$$

Demostración.- El resultado se sigue de forma inmediata de los teoremas 5.4.10 y 5.5.4 .

(c.q.d.)

De este corolario se deduce que un experimento $E^S(\underline{n})$ óptimo realizado a fin de obtener información sobre el vector paramétrico $\underline{\vartheta}$ (esto es, sobre el conjunto de parámetros ϑ_i y, en definitiva, sobre el valor de ϑ) es aquel que optimiza su valor esperado en el que interviene, a través de la utilidad esperada, la información $I^{\vartheta}\{E^S(\underline{n}), p(\underline{\vartheta})\}$. De ahí que suponiendo la independencia de las cantidades aleatorias ϑ_i ($i=1, 2, \dots, L$) dicho experimento óptimo se obtiene, a su vez, optimizando expresiones en las que aparece la suma de las informaciones que proporciona sobre cada cantidad de interés por separado, suma que coincide también con la información que cada experimento componente $E(n_i)$ proporciona sobre su parámetro respectivo.

Destaquemos que cuando los parámetros ϑ son independientes no tiene sentido la realización del experimento $E^S(\underline{n})$ para recabar exclusivamente información sobre una única componente ϑ_{i_0} . En efecto, $E^S(\underline{n})$ obliga por su misma definición a la obtención de muestras en todos los estratos cuando, en el caso de independencia, toda la información que sobre ϑ_{i_0} proporciona es la canalizada a través del experimento componente $E(n_{i_0})$, de forma que, obviamente, la estrategia óptima pasa por la obtención de una única muestra total del estrato en cuestión dejando a un lado la realización de todo el experimento $E^S(\underline{n})$. Esto puede verse de forma muy gráfica al considerar el conjunto de L ecuaciones que conducen a la obtención del tamaño muestral óptimo. En efecto, dado un costo lineal en todos los estratos y siendo

$$C(E^S(\underline{n})) = \sum_{i=1}^L c_i \cdot n_i$$

resulta que bajo la hipótesis de independencia de las ϑ_i , $I^{\vartheta_i}\{E(n_i), p(\vartheta_i)\} = I^{\vartheta_i}\{E^S(\underline{n}), p(\underline{\vartheta})\}$, de forma que las L condiciones

$$\frac{\partial U^{i_0}(E^S(\underline{n}))}{\partial n_j} = \frac{\partial C(E^S(\underline{n}))}{\partial n_j} \quad j=1, 2, \dots, L$$

se traducen, considerando por ejemplo la primera de las funciones de la

utilidad,

$$U^{i_0}(E^S(\underline{n})) = g_{1i_0} \cdot I^{\theta_{i_0}} \{E^S(\underline{n}), p(\underline{\theta})\},$$

en

$$g_{1i_0} \cdot \frac{\partial I^{\theta_{i_0}} \{E^S(\underline{n}), p(\underline{\theta})\}}{\partial n_{i_0}} = g_{1i_0} \cdot \frac{d I^{\theta_{i_0}} \{E(n_{i_0}), p(\theta_{i_0})\}}{dn_{i_0}} = c_{i_0}$$

mientras que las restantes condiciones conducen al absurdo de considerar unos costos unitario nulos, $c_k = 0$ para todo $k \neq i_0$. Este absurdo desaparece, obviamente, al considerar esos costes unitarios nulos como solución y tratar a toda la muestra como concentrada en el único estrato i_0 -ésimo.

Por otra parte, independientemente de la expresión concreta para cada uno de los costos asociados a los experimentos componentes $E(n_i)$, resulta claro que la estructura del costo asociado al experimento $E^S(\underline{n})$ puede venir expresada como una suma de los costos parciales, toda vez que los experimentos $E(n_i)$ son independientes. Puede, por tanto, considerarse como expresión para $C(E^S(\underline{n}))$ la siguiente:

$$C(E^S(\underline{n})) = \sum_{i=1}^L C(E(n_i)) + c_0$$

donde c_0 es el costo fijo de realización de $E^S(\underline{n})$ y que depende de la especial estratificación realizada en la población X .

Notemos finalmente que con un costo respetando la estructura aditiva anterior y cuando exista independencia entre los parámetros θ_i la condición

$$\frac{\partial W^i(E^S(\underline{n}))}{\partial n_i} = 0$$

se traduce en la igualdad entre derivadas

$$\frac{d I^{\theta_i} \{E(n_i), p(\theta_i)\}}{dn_i} = \frac{d C(E(n_i))}{dn_i} \quad (a)$$

Dado que bajo la hipótesis de independencia, $I^{\underline{\theta}} \{E^S(\underline{n}), p(\underline{\theta})\}$ es una suma de informaciones $I^{\theta_i} \{E(n_i), p(\theta_i)\}$, deducimos finalmente que si el costo respeta la estructura lineal, el óptimo obtenido de las condicio-

nes

$$\frac{\partial W(E^s(\underline{n}))}{\partial n_i} = 0 \quad i=1,2,\dots,L$$

coincide con el conjunto de óptimos deducidos de las L igualdades (a) anteriores. Así pues, en este caso la optimización de la información $I^s\{E^s(\underline{n}), p(\underline{\theta})\}$ coincide con la optimización en cada estrato de las informaciones $I^i\{E(n_i), p(\theta_i)\}$.

CAPITULO 6

CÁLCULO DE LA INFORMACION ESPERADA PROPORCIONADA POR EL EXPERIMENTO $E^S(\underline{n})$.
 RESULTADOS PRACTICOS

En este último capítulo van a aplicarse las ideas teóricas contenidas en el capítulo quinto. Para ello se verá primeramente dos métodos que permiten obtener de forma práctica y exacta la información esperada que sobre \underline{g} proporciona el experimento $E^S(\underline{n})$, para más tarde aplicarlos al modelo normal. En segundo lugar, se obtendrá una expresión para aproximar la información anterior ante transformaciones aproximadamente normales del vector \underline{g} . Finalmente, esta aproximación será utilizada para describir un problema que tiene como objetivo la determinación de una distribución de proporciones.

6.1 DOS METODOS PARA LA OBTENCION EXACTA DE $I^g\{E^S(\underline{n}), p(\underline{g})\}$

Es conocido (como ya se ha comentado en varias ocasiones a lo largo de la presente tesis) que el cálculo de la información esperada puede resultar difícil en la práctica, ante la complejidad del integrando que aparece en su definición. Precisamente por ello existen métodos de aproximación que estudian el comportamiento asintótico de dicha información esperada en muestras grandes (ver, al respecto, los comentarios efectuados en la sección 2.3 y los trabajos de Bernardo (1979)^{*} y Ibragimov & Hans' Minsky (1973)). Si esta dificultad es ya notoria cuando la cantidad de interés es una variable aleatoria unidimensional y la muestra resultado del experimento $E(n)$ se extrae de un único espacio muestral, es evidente entonces la dificultad adicional que supone manejar un vector paramétrico como cantidad de interés, \underline{g} , y una muestra extraída mediante un procedimiento complejo, como lo es el resultado $z(\underline{n})$ del experimento $E^S(\underline{n})$ asociado a

un proceso de estratificación. Precisamente una de las mayores dificultades radica en que el experimento $E^S(\underline{n})$ considera un vector tamaño muestral \underline{n} cuyas componentes - que son los tamaños muestrales parciales de las muestras extraídas en los estratos respectivos - no son necesariamente iguales, existiendo distintos espacios muestrales (tantos como componentes tiene el vector $\underline{\theta}$). En efecto, en el caso contrario con tamaños parciales iguales (esto es, con una afijación uniforme) la obtención de la información $I^{\underline{\theta}}\{E^S(\underline{n}), p(\underline{\theta})\}$ podría simplificarse considerablemente suponiendo que la muestra z es extraída de la población L -dimensional

$$\prod_{i=1}^L x_i$$

De esta forma, la determinación de la información esperada quedaría encuadrada dentro de la teoría general, con una cantidad de interés L -dimensional $\underline{\theta}$ y un espacio muestral también L -dimensional. Sin embargo, esta situación no es la general, por lo que resulta necesario encontrar procedimientos (aparte de la obtención directa por aplicación de su definición) que permitan obtener de manera exacta o aproximada dicha información. En la presente sección van a estudiarse dos métodos que permiten determinar exactamente a $I^{\underline{\theta}}\{E^S(\underline{n}), p(\underline{\theta})\}$, dejando para la sección 6.3 el estudio del comportamiento asintótico de dicha información ante transformaciones biyectivas de la cantidad de interés.

I. OBTENCION DE $I^{\underline{\theta}}\{E^S(\underline{n}), p(\underline{\theta})\}$ POR EXTENSION DE LA CANTIDAD DE INTERES. Consideremos el vector tamaño muestral $\underline{n} = (n_1, n_2, \dots, n_L)$ con

$$n = \sum_{i=1}^L n_i$$

Para este vector genérico se define la matriz de dimensión $n \times L$

$$A(\underline{n}) = (a_{ij}) \begin{cases} 1 & \text{si } j=1, 2, \dots, L; \\ & i = \sum_k^{j-1} n_k + 1, \sum_k^{j-1} n_k + 2, \dots, \sum_k^{j-1} n_k + n_j \\ 0 & \text{en los demás casos} \end{cases}$$

esto es,

$$A(\underline{n}) = \left(\begin{array}{cccc} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & 1 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{array} \right) \begin{array}{l} \left. \begin{array}{l} \dots \\ \dots \\ \dots \end{array} \right\} n_1 \\ \left. \begin{array}{l} \dots \\ \dots \\ \dots \end{array} \right\} n_2 \\ \left. \begin{array}{l} \dots \\ \dots \\ \dots \end{array} \right\} n_L \end{array} \quad n \times L$$

Expresemos el vector paramétrico como un vector columna $L \times 1$

$$\underline{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_L \end{pmatrix} \quad L \times 1$$

y consideremos el nuevo vector paramétrico $\underline{\mu}$ obtenido por la transformación definida a través de la matriz $A(\underline{n})$:

$$\underline{\mu} = A \cdot \underline{\theta}$$

de forma que el vector columna $\underline{\mu}$ de dimensión $n \times 1$ tendrá como componentes

$$\underline{\mu} = (\mu_{i1}) \begin{cases} \theta_1 & 1 \leq i \leq n_1 \\ \theta_2 & n_1 + 1 \leq i \leq n_2 \\ \dots & \dots \\ \theta_L & n_{L-1} + 1 \leq i \leq n_L \end{cases}$$

y tomará valores en el conjunto

$$\underline{\mu} = \prod_{i=1}^L \prod_{j=1}^{n_i} \Theta_{ij} = \prod_{i=1}^L \Theta_{i \times \dots \times i \times \dots \times i}^{(n_i)} \Theta_{i1}$$

Consideremos una transformación inversa cualquiera de la anterior, la definida por la matriz, por ejemplo, $B(\underline{n})$ de expresión

$$B(\underline{n}) = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ & & & & & & & & & & & & \\ & & & & & & & & & & & & \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & 1 & 0 & \dots & 0 \end{pmatrix} L \times n$$

siendo

$$\underline{\vartheta} = B \cdot \underline{\mu}$$

Evidentemente, siendo $p(\underline{\vartheta})$ la densidad de probabilidad que describe las opiniones iniciales que el investigador posea sobre $\underline{\vartheta}$, las opiniones iniciales sobre el nuevo parámetro $\underline{\mu}$ vendrán descritas por la correspondiente densidad transformada (supuestamente existente) $p(\underline{\mu}) = A(p(\underline{\vartheta}))$, que corresponde a una distribución degenerada. En efecto, toda la masa de probabilidad de la variable $\underline{\mu}$ está contenida en el hiperplano de L -dimensiones definido por las igualdades

$$\mu_{i1} = \vartheta_j \quad \text{si} \quad \sum_k^{j-1} n_k + 1 \leq i \leq \sum_k^{j-1} n_k + n_j \quad ; \quad j=1,2,\dots,L$$

A partir de los elementos anteriores puede considerarse que $\underline{\mu} = (\underline{\vartheta}, \underline{\mu}^*)$, donde $\underline{\mu}^*$ es el vector formado por las $n-1$ componentes ϑ_j restantes ($j=1,2,\dots,L$). Denotemos por $m(\underline{\vartheta})$ la densidad marginal obtenida de $p(\underline{\mu})$ proyectando las $n-L$ componentes restantes que definen a $\underline{\mu}^*$, esto es,

$$m(\underline{\vartheta}) = \int p(\underline{\mu}) \cdot d\underline{\mu}^*$$

donde $\underline{\mu}^*$ toma valores en

$$\underline{\mu}^* = \prod_{i=1}^L \prod_{j=1}^{n_i-1} \vartheta_i = \prod_{i=1}^L \vartheta_i \times \vartheta_i \times \dots \times \vartheta_i$$

Con la notación expresada demostraremos el siguiente

LEMA 6.1.1

La densidad marginal $m(\underline{\vartheta})$ deducida de $p(\underline{\mu})$ coincide con $p(\underline{\vartheta})$

Demostración.- Es conocida (por ejemplo, Lindgren (1968)) la relación existente entre las funciones características de las variables $\underline{\vartheta}$ y $\underline{\mu}$, dada por

$$\varphi_{\underline{\vartheta}, p(\underline{\vartheta})}(t) = \varphi_{\underline{\mu}, p(\underline{\mu})}(B't)$$

donde t es un vector columna $L \times 1$. Resulta

$$\begin{aligned} \varphi_{\underline{\vartheta}, p(\underline{\vartheta})}(t) &= E_{\underline{\vartheta}, p(\underline{\vartheta})}(e^{it' \underline{\vartheta}}) = E_{\underline{\mu}, p(\underline{\mu})}(e^{it'B \underline{\mu}}) = E_{\underline{\mu}, p(\underline{\mu})}(e^{it' \underline{\vartheta}}) = \\ &= E_{(\underline{\vartheta}, \underline{\mu}^*), p(\underline{\vartheta}, \underline{\mu}^*)}(e^{it' \underline{\vartheta}}) = E_{\underline{\vartheta}, m(\underline{\vartheta})}(e^{it' \underline{\vartheta}}) = \varphi_{\underline{\vartheta}, m(\underline{\vartheta})}(t) \end{aligned}$$

y donde $E_{\omega, p(\omega)}$ denota una esperanza definida sobre ω con la densidad de probabilidad $p(\omega)$.

(c.q.d.)

Aplicamos el resultado del lema 6.1.1 para encontrar un procedimiento que permita la obtención exacta de la información $I^{\vartheta} \{E^{\vartheta}(z), p(\underline{\vartheta})\}$. Para ello demostramos que las siguientes situaciones son equivalentes: por una parte, realizar el experimento $E^{\vartheta}(z)$ a fin de recabar información sobre el vector cantidad de interés $\underline{\vartheta}$; por otra, realizar el experimento $E(E^{\vartheta}(z)) = \{z, \underline{\mu}, p(z/\underline{\mu})\}$ consistente en la extracción de una unidad muestral de la población n -dimensional

$$z = \prod_{i=1}^L \prod_{j=1}^{n_i} x_{ij} = \prod_{i=1}^L x_{i1} \times x_{i2} \times \dots \times x_{in_i}$$

con objeto de obtener información sobre la cantidad de interés transformada $\underline{\mu} = A \cdot \underline{\vartheta}$. Notemos la existencia de la igualdad entre las densidades de probabilidad siguiente:

$$p(z/\underline{\mu}) = \prod_{i=1}^L \prod_{j=1}^{n_i} p(x_{ij}/\vartheta_i) = p(z/\underline{\vartheta})$$

y donde z , en $p(z/\underline{\vartheta})$, representa un conjunto de L submuestras extraídas de sendos estratos, mientras que en $p(z/\underline{\mu})$ no es sino una unidad muestral de una población n -dimensional. Notemos además que

$$\begin{aligned} p_{p(\underline{\mu})}(z) &= \int p(z/\underline{\mu}) \cdot p(\underline{\mu}) \cdot d\underline{\mu} = \iint p(z/\underline{\vartheta}) \cdot p(\underline{\vartheta}, \underline{\mu}^*) \cdot d\underline{\vartheta} \cdot d\underline{\mu}^* = \\ &= \int p(z/\underline{\vartheta}) \cdot p(\underline{\vartheta}) \cdot d\underline{\vartheta} = p_{p(\underline{\vartheta})}(z) \end{aligned}$$

en virtud del lema 6.1.1. Con la notación anterior demostramos el siguiente

TEOREMA 6.1.1

$$I^{\underline{\theta}} \{E^S(\underline{n}), p(\underline{\theta})\} = I^{\underline{\mu}} \{E(E^S(\underline{n})), p(\underline{\mu})\}$$

Demostración.- Siendo z el resultado muestral común a los experimentos $E^S(\underline{n})$ y $E(E^S(\underline{n}))$ y de sus mismas definiciones resulta

$$\begin{aligned} I^{\underline{\mu}} \{E(E^S(\underline{n})), p(\underline{\mu})\} &= \iint p(\underline{\mu}) \cdot p(z/\underline{\mu}) \cdot \log \frac{p(z/\underline{\mu})}{p(z)} \cdot d\underline{\mu} \cdot dz = \\ &= \iint p(\underline{\mu}) \cdot p(z/\underline{\theta}) \cdot \log \frac{p(z/\underline{\theta})}{p(z)} \cdot d\underline{\mu} \cdot dz = \\ &= \iint p(\underline{\theta}) \cdot p(z/\underline{\theta}) \cdot \log \frac{p(z/\underline{\theta})}{p(z)} \cdot d\underline{\theta} \cdot dz = I^{\underline{\theta}} \{E^S(\underline{n}), p(\underline{\theta})\} \end{aligned}$$

habiendo aplicado el resultado del lema 6.1.1.

(c.q.d.)

Así pues el teorema anterior permite transformar el problema de obtención de la información esperada asociada a un proceso de estratificación a aquel que pretende determinar dicha información asociada a un experimento consistente en la extracción de una unidad muestral única de una población n -dimensional. Como la dimensión del parámetro transformado $\underline{\mu}$ es también n , el problema queda, pues, reducido al caso n -dimensional del estudiado por Lindley (1956).

Transformando el anterior resultado al caso límite de una población con un estrato único, esto es, a la teoría general unidimensional, el teorema 6.1.1 indica que, bajo el punto de vista de la información esperada, las situaciones expresadas por los experimentos $E(n)$ y $E(E(n))$ consistente éste en la obtención de una unidad muestral cuando ésta se distribuye según

$$\prod_{i=1}^n p(x_i/\theta)$$

a fin de recabar información sobre $\underline{\mu} = A \cdot \theta$ - siendo A un vector columna $n \times 1$ de valores unidad -, son equivalentes. Comprobémoslo para el modelo normal. Sea, pues, θ una variable $N(\theta_0, \sigma_0^2)$. La nueva variable

$\underline{\mu} = A \cdot \theta$, con

$$A = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}_{n \times 1}$$

es (DeGroot, 1970) una normal degenerada n -dimensional, con vector media $A \theta$ y matriz de momentos $A \cdot \sigma_o^2 \cdot A' = \sigma_o^2 \cdot AA'$. Considerando la población

$$Z = \prod_{i=1}^n X = X_1 X_2 \dots X_n$$

distribuida como una normal n -dimensional $N^n(A\theta, H)$ de componentes independientes, con

$$H = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix}_{n \times n}$$

y aplicando formalmente el resultado obtenido por Lindley (1956) para el modelo normal n -dimensional

$$I^{\mu} \{E(E(n)), N^n(A\theta_o, \sigma_o^2 \cdot AA')\} = \frac{1}{2} \log \frac{|H + A \sigma_o^2 A'|}{|H|}$$

Y al ser

$$|A \sigma_o^2 A' + H| = \begin{vmatrix} \sigma_o^2 + \sigma^2 & \sigma_o^2 & \dots & \sigma_o^2 \\ \sigma_o^2 & \sigma_o^2 + \sigma^2 & \dots & \sigma_o^2 \\ \dots & \dots & \dots & \dots \\ \sigma_o^2 & \sigma_o^2 & \dots & \sigma_o^2 + \sigma^2 \end{vmatrix} =$$

$$= \sigma^{2n} + n \sigma_o^2 \sigma^{2(n-1)}, \quad \text{y} \quad |H| = \sigma^{2n}$$

resulta finalmente que

$$I^{\mu} \{E(E(n)), N^n(A\theta_o, \sigma_o^2 \cdot AA')\} = \frac{1}{2} \log \frac{\sigma^{2n} + n \sigma_o^2 \sigma^{2(n-1)}}{\sigma^{2n}} =$$

$$= \frac{1}{2} \log \left(1 + n \cdot \frac{\sigma_0^2}{\sigma^2} \right) = I^\theta \left\{ E(n), N(\vartheta_0, \sigma_0^2) \right\}$$

resultado que corrobora de forma práctica la equivalencia de las situaciones recogidas en el teorema 5.1.1 .

II. OBTENCION DE $I^\theta \{ E^S(\underline{n}), p(\underline{\vartheta}) \}$ POR CONSIDERACION DE ESTOS DISTICOS SUFICIENTES EN TODOS LOS ESTRATOS. Un segundo procedimiento para la obtención exacta de $I^\theta \{ E^S(\underline{n}), p(\underline{\vartheta}) \}$ es una consecuencia inmediata del corolario 5.4.3 . Notemos que la situación expresada por la existencia de un experimento $E^S(\underline{n})$ es también equivalente a la indicada por otro experimento realizado para obtener información también sobre $\underline{\vartheta}$ y consistente en obtener una muestra unitaria de la población L-dimensional

$$Z = Z_1 \times Z_2 \times \dots \times Z_L$$

cuando Z_j es, a su vez, la población n_j -dimensional

$$Z_j = \prod_{i=1}^{n_j} X_{j,i} = X_{j,1} \times X_{j,2} \times \dots \times X_{j,n_j}$$

Es evidente que el problema anterior quedará reducido a la situación L-dimensional estudiada ya por Lindley (1956) cuando, en cada estrato, la observación se cifa al estudio de un estadístico suficiente para todo el vector $\underline{\vartheta}$, de acuerdo con el corolario mencionado. Así pues y desde el punto de vista de la información esperada, las situaciones expresadas por los experimentos $E^S(\underline{n})$ y $E(\underline{1}, S)$ consistente éste en la obtención de una unidad muestral de un espacio L-dimensional S en el que toma valores un conjunto de L estadísticos suficientes para $\underline{\vartheta}$, son equivalentes. De ahí que sustituyendo los espacios muestrales Z_i por sendos S_i en donde toman valores un conjunto de estadísticos φ_i tales que

$$p(\underline{\vartheta} / z_1, \dots, z_{i-1}, \varphi_i, z_{i+1}, \dots, z_L) = p(\underline{\vartheta} / z) \quad \text{para todo } i=1, 2, \dots, L$$

y aplicando el corolario 5.4.3, resulte finalmente que

$$I^\theta \{ E^S(\underline{n}), p(\underline{\vartheta}) \} = I^\theta \{ E(\underline{1}, S), p(\underline{\vartheta}) \}$$

Apliquemos otra vez el criterio anterior al caso unidimensional en un espacio sin estratificar. La realización del experimento $E(n)$, a fin de obtener información sobre θ , efectuada sobre el espacio muestral X resultaría equivalente - desde el punto de vista de la información es-

perada proporcionada - a realizar un experimento $E(1, S)$ consistente en obtener una unidad muestral de un espacio S en el que toma valores un estadístico suficiente para θ . La igualdad de ambos criterios queda garantizada por los resultados encontrados por Lindley (1956) y Bernardo (1978).

La aplicación de este criterio al modelo normal conduciría a la consideración de, por ejemplo, la media muestral \bar{x} como estadístico suficiente. Supongamos, pues, que $p(x/\theta)$ es $N(\theta, \sigma)$ y que $p(\theta) = N(\theta_0, \sigma_0)$. Sabemos que \bar{x} será $N(\theta, \sigma/\sqrt{n})$ con valores en un cierto conjunto S . Así pues,

$$I^{\theta} \{E(1, S), N(\theta_0, \sigma_0)\} = \frac{1}{2} \log \left(1 + \frac{\sigma_0^2}{\sigma^2 / n} \right) = I^{\theta} \{E(n), N(\theta_0, \sigma_0)\}$$

Ya se indicó en su momento que la información esperada proporcionada por un experimento $E^S(\underline{n})$ no variaba cuando toda la muestra en su conjunto (y no consideradas las submuestras por separado) se sustituye por un estadístico suficiente respecto al vector $\underline{\theta}$ (teorema 5.4.2). También comentábamos que este resultado seguía siendo totalmente válido ante cualquier transformación de la cantidad de interés, pues bastaba seguir idéntica técnica de demostración que la utilizada por Bernardo (1978). Pues bien, a su vez el resultado del teorema 5.4.6 puede aplicarse cuando la investigación se centra sobre una nueva cantidad de interés $\underline{\mu} = T(\underline{\theta})$ resultado de una transformación cualquiera del vector paramétrico $\underline{\theta}$. Demostremos, pues, el siguiente

TEOREMA 6.1.2

Considerando la transformación $\underline{\mu} = T(\underline{\theta})$ de la cantidad de interés y bajo las hipótesis y notación del teorema 5.4.6, resulta:

$$I^{\underline{\mu}} \{E^S(\underline{n}+1), p(\underline{\theta})\} = I^{\underline{\mu}} \{E^S(\underline{n}), p(\underline{\theta})\} + I^{\underline{\mu}} \{E(n_1 + 1)^S / E^S(\underline{n}), p(\underline{\theta})\}$$

Demostración.- Siguiendo el razonamiento de Bernardo (1978) y de sus definiciones respectivas resulta

$$I^{\underline{\mu}} \{E^S(\underline{n}), p(\underline{\theta})\} = \iint p(\underline{z}(\underline{n})) \cdot p(\underline{\mu}/\underline{z}(\underline{n})) \cdot \log \frac{p(\underline{\mu}/\underline{z}(\underline{n}))}{p(\underline{\mu})} \cdot d\underline{\mu} \cdot d\underline{z}(\underline{n}) =$$

$$= \iint p(z(n+1)) \cdot p(\underline{\mu}/z(n+1)) \cdot \log \frac{p(\underline{\mu}/z(n))}{p(\underline{\mu})} \cdot d\underline{\mu} \cdot dz(n+1)$$

$$I^{\underline{\mu}} \{ E(n_1+1)^{\underline{\theta}} / E^S(\underline{n}), p(\underline{\theta}) \} = E_{z(n)} \{ I^{\underline{\mu}} \{ E(n_1+1)^{\underline{\theta}} / z(n), p(\underline{\theta}/z(n)) \} \} =$$

$$= \iint p(z(n+1)) \cdot p(\underline{\mu}/z(n+1)) \cdot \log \frac{p(\underline{\mu}/z(n+1))}{p(\underline{\mu}/z(n))} \cdot d\underline{\mu} \cdot dz(n+1)$$

Sumando ambas integrales tenemos

$$\iint p(z(n+1)) \cdot p(\underline{\mu}/z(n+1)) \cdot \log \frac{p(\underline{\mu}/z(n+1))}{p(\underline{\mu})} \cdot d\underline{\mu} \cdot dz(n+1) =$$

$$= I^{\underline{\mu}} \{ E^S(n+1), p(\underline{\theta}) \}$$

(c.q.d.)

La generalización a estadísticos suficientes en todos los estratos es inmediata. Resulta entonces el siguiente

COROLARIO 6.1.1

Siendo $z(\underline{n})$ suficiente para $z(n+1)$ con respecto a $\underline{\mu}$ en todos los estratos, esto es, verificándose que $p(\underline{\mu}/z(\underline{n}), x_{i n_1+1}) = p(\underline{\mu}/z(\underline{n}))$ para todo $i=1,2,\dots,L$, entonces

$$I^{\underline{\mu}} \{ E^S(\underline{n}), p(\underline{\theta}) \} = I^{\underline{\mu}} \{ E^S(n+1), p(\underline{\theta}) \}$$

Demostración.- Es inmediata sin más que considerar que en todo estrato $I^{\underline{\mu}} \{ E(n_1+1)^{\underline{\theta}} / E^S(\underline{n}), p(\underline{\theta}) \} = 0$ y aplicar el teorema anterior.

(c.q.d.)

Así como el teorema 5.4.6 y su respectivo corolario 5.4.3 permitían definir el segundo de los procedimientos de obtención exacta para $I^{\underline{\theta}} \{ E^S(\underline{n}), p(\underline{\theta}) \}$, el corolario 6.1.1 permite aplicarlo para la obtención exacta de las informaciones esperadas $I^{\underline{\theta}} \{ E^S(\underline{n}), p(\underline{\theta}) \}$ y $I^{\underline{\theta}^i} \{ E^S(\underline{n}), p(\underline{\theta}) \}$ definidas en la sección 5.2 a partir de sendas transformaciones de la cantidad de interés $\underline{\theta}$. Todas estas ideas las aplicaremos ahora mismo en el modelo normal. Puede, pues, escribirse bajo la notación ya expresada que

$$I^{\vartheta} \{E^S(\underline{n}), p(\underline{\vartheta})\} = I^{\vartheta} \{E(\underline{1}, S), p(\underline{\vartheta})\}, \text{ y}$$

$$I^{\vartheta_1} \{E^S(\underline{n}), p(\underline{\vartheta})\} = I^{\vartheta_1} \{E^S(\underline{1}, S), p(\underline{\vartheta})\}.$$

6.2 EL MODELO NORMAL

La extensión a un experimento de la forma $E^S(\underline{n})$ del modelo normal desarrollado por Lindley (1956) y Bernardo (1975) supone considerar un conjunto de L estratos o subpoblaciones sobre las que hay definidas sendas densidades normales y dependientes de cada componente ϑ_i de $\underline{\vartheta}$, dadas por

$$p(x_i / \vartheta_i) = \frac{1}{\sqrt{2\pi} \sigma_i} e^{-\frac{1}{2} \left(\frac{x_i - \vartheta_i}{\sigma_i} \right)^2} \quad x_i \in \mathbb{R}$$

con σ_i conocida en todos los estratos. Debido a la independencia de las muestras parciales, la densidad conjunta que describe toda la muestra z vendrá dada por

$$p(z / \underline{\vartheta}) = \prod_{i=1}^L \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi} \sigma_i} e^{-\frac{1}{2} \left(\frac{x_{ij} - \vartheta_i}{\sigma_i} \right)^2}$$

con $z = (z_1, z_2, \dots, z_L)$, y $z_i = (x_{i1}, x_{i2}, \dots, x_{in_i})$ $i = 1, 2, \dots, L$. Por tanto z puede considerarse como una muestra cuya distribución, dependiente del vector aleatorio columna $L \times 1$ $\underline{\vartheta}$, tiene como densidad de probabilidad un producto de densidades - recordemos que las submuestras son aleatorias e independientes entre sí - multinormales de n_i -dimensiones, respectivamente.

La densidad inicial $p(\underline{\vartheta})$ especifica completamente el modelo. En general supondremos que $p(\underline{\vartheta})$ es también una multinormal que, describiendo las opiniones iniciales que el investigador posee sobre $\underline{\vartheta}$, tiene como vector media $\underline{\vartheta}_0$ y como matriz de momentos H_0 , de componentes σ_{ij}^0 . Así pues,

$$p(\underline{\vartheta}) = \frac{1}{(\sqrt{2\pi})^L |H_0|^{\frac{1}{2}}} e^{-\frac{1}{2} (\underline{\vartheta} - \underline{\vartheta}_0)' H_0^{-1} (\underline{\vartheta} - \underline{\vartheta}_0)}$$

con $\underline{\vartheta} \in R^L$, y siendo

$$\underline{\vartheta}_0 = \begin{pmatrix} \vartheta_{10} \\ \vartheta_{20} \\ \vdots \\ \vartheta_{L0} \end{pmatrix} \quad \text{y} \quad H_0 = \begin{pmatrix} \sigma_{11}^0 & \sigma_{12}^0 & \dots & \sigma_{1L}^0 \\ \sigma_{12}^0 & \sigma_{22}^0 & \dots & \sigma_{2L}^0 \\ \dots & \dots & \dots & \dots \\ \sigma_{1L}^0 & \sigma_{2L}^0 & \dots & \sigma_{LL}^0 \end{pmatrix}$$

La anterior hipótesis acerca de la distribución de $\underline{\vartheta}$ abarca también una serie de casos en los que el investigador prefiere describir sus opiniones iniciales sobre cada componente ϑ_i por separado a través de un conjunto de L densidades $p(\vartheta_i/\varphi)$, donde φ es un nuevo parámetro que indica la dependencia entre las ϑ_i . Especificando una densidad inicial $p(\varphi)$ para el nuevo parámetro, la densidad conjunta se obtendrá de la integral ya mencionada en la sección 5.1

$$\int \prod_{i=1}^L p(\vartheta_i/\varphi) \cdot p(\varphi) \cdot d\varphi = p(\underline{\vartheta})$$

Pues bien, el modelo normal descrito ahora mismo abarca también la situación anterior cuando las $p(\vartheta_i/\varphi)$ son normales con el mismo valor medio φ . Este es el sentido del siguiente

TEOREMA 6.2.1

Siendo $p(\vartheta_i/\varphi)$ normales con media φ y varianza σ_{ii}^0 conocida para todo $i = 1, 2, \dots, L$, siendo $p(\varphi)$ normal de media φ_0 y varianza σ_0^2 , resulta que la distribución conjunta $p(\underline{\vartheta})$ es multinormal L -dimensional.

Demostración.- Por misma definición

$$p(\underline{\vartheta}) = \int \prod_{i=1}^L p(\vartheta_i/\varphi) \cdot p(\varphi) \cdot d\varphi = \int \frac{1}{(\sqrt{2\pi})^{L+1} \sigma_0 \sqrt{\prod_{i=1}^L \sigma_{ii}^0}} \cdot$$

$$\cdot \exp \left\{ -\frac{1}{2} \left(\sum_{i=1}^L \frac{(\vartheta_i - \varphi)^2}{\sigma_{ii}^0} + \frac{(\varphi - \varphi_0)^2}{\sigma_0^2} \right) \right\} d\varphi$$

Consideremos el exponente

$$\sum_{i=1}^L \frac{(\vartheta_i - \varphi)^2}{\sigma_{ii}^0} + \frac{(\varphi - \varphi_0)^2}{\sigma_0^2} = \varphi^2 \left\{ \sum_{i=1}^L \frac{1}{\sigma_{ii}^0} + \frac{1}{\sigma_0^2} \right\} -$$

$$- 2\varphi \left\{ \sum_{i=1}^L \frac{\vartheta_i}{\sigma_{ii}^0} + \frac{\varphi_0}{\sigma_0^2} \right\} + \sum_{i=1}^L \frac{\vartheta_i^2}{\sigma_{ii}^0} + \frac{\varphi_0^2}{\sigma_0^2} =$$

$$\frac{\left(\varphi - \frac{\sum_{i=1}^L \frac{\vartheta_i}{\sigma_{ii}^0} + \frac{\varphi_0}{\sigma_0^2}}{\sum_{i=1}^L \frac{1}{\sigma_{ii}^0} + \frac{1}{\sigma_0^2}} \right)^2}{\frac{1}{\sum_{i=1}^L \frac{1}{\sigma_{ii}^0} + \frac{1}{\sigma_0^2}}} + \sum_{i=1}^L \frac{\vartheta_i^2}{\sigma_{ii}^0} + \frac{\varphi_0^2}{\sigma_0^2} +$$

Por tanto

$$p(\underline{\vartheta}) = \frac{1 / \sqrt{\sum_{i=1}^L \frac{1}{\sigma_{ii}^0} + \frac{1}{\sigma_0^2}}}{(\sqrt{2\pi})^L \sigma_0 \sqrt{\prod_{i=1}^L \sigma_{ii}^0}} \cdot \exp$$

$$\exp \left\{ -\frac{1}{2} \left(\sum_{i=1}^L \frac{\vartheta_i^2}{\sigma_{ii}^0} + \frac{\varphi_0^2}{\sigma_0^2} - \frac{\left(\sum_{i=1}^L \frac{\vartheta_i}{\sigma_{ii}^0} + \frac{\varphi_0}{\sigma_0^2} \right)^2}{\sum_{i=1}^L \frac{1}{\sigma_{ii}^0} + \frac{1}{\sigma_0^2}} \right) \right\} =$$

$$= \frac{1}{\sqrt{s} (\sqrt{2\pi})^L} \cdot \exp \left\{ -\frac{1}{2} (\underline{\vartheta} - \underline{\varphi}_0)' \tau (\underline{\vartheta} - \underline{\varphi}_0) \right\}$$

con

$$\underline{\vartheta} - \underline{\varphi}_0 = \begin{pmatrix} \theta_1 - \varphi_0 \\ \theta_2 - \varphi_0 \\ \vdots \\ \theta_L - \varphi_0 \end{pmatrix}, \quad S = \frac{1}{|T|} = \sum_{i=0}^L \prod_{j \neq i}^L \sigma_{jj}^0, \quad y$$

$$T = \frac{1}{S} \begin{pmatrix} \sum_{\substack{i \neq 1 \\ i \geq 0}}^L \prod_{j \neq 1, i}^L \sigma_{jj}^0 & - \prod_{\substack{i \neq 1, 2 \\ i \geq 0}}^L \sigma_{ii}^0 & \dots & - \prod_{\substack{i \neq 1, L \\ i \geq 0}}^L \sigma_{ii}^0 \\ - \prod_{\substack{i \neq 1, 2 \\ i \geq 0}}^L \sigma_{ii}^0 & \sum_{\substack{i \neq 2 \\ i \geq 0}}^L \prod_{j \neq 2, i}^L \sigma_{jj}^0 & \dots & - \prod_{\substack{i \neq 2, L \\ i \geq 0}}^L \sigma_{ii}^0 \\ \dots & \dots & \dots & \dots \\ - \prod_{\substack{i \neq 1, L \\ i \geq 0}}^L \sigma_{ii}^0 & - \prod_{\substack{i \neq 2, L \\ i \geq 0}}^L \sigma_{ii}^0 & \dots & \sum_{\substack{i \neq L \\ i \geq 0}}^L \prod_{j \neq L, i}^L \sigma_{jj}^0 \end{pmatrix}$$

donde por exigencias de notación admitimos que $\sigma_{00}^0 = \sigma_0^2$.
(c.q.d.)

Así pues, y con generalidad, consideraremos como hipótesis básicas del modelo normal asociado al experimento $E^S(\underline{n})$ las dos siguientes:

- I. $p(x_i / \theta_i)$ es $N(\theta_i, \sigma_i^2)$, $i=1, 2, \dots, L$
- II. $p(\underline{\vartheta})$ es multinormal $N^L(\underline{\vartheta}_0, H_0)$

De esta forma el experimento $E^S(\underline{n})$ se definirá como la familia de L experimentos $\{E_i(n_i)\}_{i=1, 2, \dots, L}$:

$$E_i = \{x_i, \theta_i, N(\theta_i, \sigma_i^2)\}$$

Pasemos ya a obtener la información esperada $I^{\underline{\vartheta}}\{E^S(\underline{n}), N^L(\underline{\vartheta}_0, H_0)\}$ aplicando los resultados de la sección anterior.

I. Obtención de la información esperada en el modelo normal por extensión de la cantidad de interés. Apliquemos los resultados encontrados por el primero de los métodos propuestos en la sección 6.1. La hipótesis I permite considerar a la muestral global z como una muestra unitaria de

una población n -dimensional normal multivariante, de media $A \cdot \underline{\vartheta}$ y matriz de momentos M , siendo las matrices

$$A = (a_{ij}) \begin{cases} 1 & \text{si } j=1,2,\dots,L \\ i = \sum_k^{j-1} n_k + 1, \sum_k^{j-1} n_k + 2, \dots, \sum_k^{j-1} n_k + n_j \\ 0 & \text{en los demás casos} \end{cases}$$

ya conocida, y

$$M = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \sigma_1^2 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ \dots & \dots \\ 0 & 0 & \dots & \sigma_1^2 & \dots & 0 & 0 & \dots & 0 \\ \dots & \dots \\ 0 & 0 & \dots & 0 & \dots & \sigma_L^2 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & \dots & 0 & \sigma_L^2 & \dots & 0 \\ \dots & \dots \\ 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & \sigma_L^2 \end{pmatrix} \begin{matrix} \left. \begin{matrix} \dots \\ \dots \\ \dots \end{matrix} \right\} n_1 \\ \left. \begin{matrix} \dots \\ \dots \\ \dots \end{matrix} \right\} n_L \end{matrix} \quad n \times n$$

De la hipótesis II del modelo normal, considerando extendido el parámetro original $\underline{\vartheta}$ al nuevo vector paramétrico $\underline{\mu} = A \cdot \underline{\vartheta}$, resulta (DeGroot, 1970) que la distribución de $\underline{\mu}$ es una normal multivariante n -dimensional degenerada, con toda la masa contenida en un hiperplano de L dimensiones y con un vector media $A \cdot \underline{\vartheta}_0$ y matriz de momentos singular $AH A'$.

Con la transformación anterior y aplicando el resultado de Lindley (1956) para la información esperada del modelo normal multivariante, resulta por el teorema 6.1.1 que

$$I^{\underline{\vartheta}} \{ E^S(\underline{n}), N^L(\underline{\vartheta}_0, H_0) \} = I^{\underline{\mu}} \{ E^S(\underline{n}), N^n(A \underline{\vartheta}_0, AH A') \} = \frac{1}{2} \log \frac{|AH A' + M|}{|M|} \quad (1)$$

como la información esperada sobre $\underline{\vartheta}$ proporcionada por $E^S(\underline{n})$ cuando se satisfacen las hipótesis I y II del modelo normal antes mencionadas.

Ya se comentó en el capítulo quinto que bajo la hipótesis de independencia de las componentes θ_i la expresión de la información esperada sobre $\underline{\vartheta}$ proporcionada por $E^S(\underline{n})$ se simplificaba notablemente y venía expresada como suma de las informaciones esperadas sobre cada θ_i proporcionadas por los experimentos componentes. La traducción de este resultado a nuestro caso se realiza a partir del siguiente

TEOREMA 6.2.2

En el modelo normal, siendo independientes las componentes θ_i del vector paramétrico $\underline{\vartheta}$, resulta ser

$$I^{\underline{\vartheta}}\{E^S(\underline{n}), N^L(\underline{\vartheta}_0, H_0)\} = \sum_{i=1}^L \frac{1}{2} \log\left(1 + n_i \cdot \frac{\sigma_{11}^0}{\sigma_i^2}\right)$$

Demostración.- Bajo la hipótesis del teorema, la matriz H_0 es diagonal

$$H_0 = \begin{pmatrix} \sigma_{11}^0 & 0 & \dots & 0 \\ 0 & \sigma_{22}^0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_{LL}^0 \end{pmatrix}$$

de forma que

$$AH_0A' + M = \begin{pmatrix} \sigma_{11}^0 + \sigma_1^2 & \sigma_{11}^0 & \dots & \sigma_{11}^0 & \dots & 0 & 0 & \dots & 0 \\ \sigma_{11}^0 & \sigma_{11}^0 + \sigma_1^2 & \dots & \sigma_{11}^0 & \dots & 0 & 0 & \dots & 0 \\ \dots & \dots \\ \sigma_{11}^0 & \sigma_{11}^0 & \dots & \sigma_{11}^0 + \sigma_1^2 & \dots & 0 & 0 & \dots & 0 \\ \dots & \dots \\ 0 & 0 & \dots & 0 & \dots & \sigma_{LL}^0 + \sigma_L^2 & \sigma_{LL}^0 & \dots & \sigma_{LL}^0 \\ 0 & 0 & \dots & 0 & \dots & \sigma_{LL}^0 & \sigma_{LL}^0 + \sigma_L^2 & \dots & \sigma_{LL}^0 \\ \dots & \dots \\ 0 & 0 & \dots & 0 & \dots & \sigma_{LL}^0 & \sigma_{LL}^0 & \dots & \sigma_{LL}^0 + \sigma_L^2 \end{pmatrix}$$

verificándose que

$$|A H_0 A' + M| = \prod_{i=1}^L (\sigma_i^{2n_i} + n_i \sigma_{ii}^0 \sigma_i^{2(n_i-1)}) \quad \text{y} \quad |M| = \prod_{i=1}^L \sigma_i^{2n_i}$$

resultando, por (1), que

$$I^{\underline{\vartheta}} \{E^S(\underline{n}), N^L(\underline{\vartheta}_0, H_0)\} = \frac{1}{2} \log \prod_{i=1}^L \left(1 + n_i \cdot \frac{\sigma_{ii}^0}{\sigma_i^2}\right) =$$

$$= \sum_{i=1}^L \frac{1}{2} \log \left(1 + n_i \cdot \frac{\sigma_{ii}^0}{\sigma_i^2}\right) = \sum_{i=1}^L I^{\vartheta_i} \{E(n_i), N(\vartheta_{i0}, \sigma_{ii}^0)\} \quad (\text{c.q.d.})$$

El teorema anterior no es sino confirmación práctica del corolario 5.5.1.

II. Obtención de la información esperada en el modelo normal por consideración de estadísticos suficientes en todos los estratos. Consideremos previamente la obtención de la densidad posterior $p(\underline{\vartheta}/z)$. Para ello, siendo $x_i \in X_i$ ($i=1,2,\dots,L$), la distribución conjunta del vector $\underline{x} = (x_1, x_2, \dots, x_L)$ formado por componentes independientes vendrá expresada a través de la densidad de probabilidad

$$p(\underline{x}/\underline{\vartheta}) = \frac{1}{(\sqrt{2\pi})^L \prod_{i=1}^L \sigma_i} e^{-\frac{1}{2} (\underline{x} - \underline{\vartheta})' R (\underline{x} - \underline{\vartheta})}$$

donde R es la matriz de precisión, inversa de la correspondiente matriz H de covarianzas

$$H = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_L^2 \end{pmatrix}$$

esto es,

$$R = \begin{pmatrix} r_1 & 0 & \dots & 0 \\ 0 & r_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & r_L \end{pmatrix} \quad \text{con} \quad r_i = \frac{1}{\sigma_i^2} \quad \text{para todo } i = 1, 2, \dots, L$$

Denotando por T a la matriz de precisión de la densidad inicial $p(\underline{\theta})$, esto es, $T = H_0^{-1}$, demosremos el siguiente

TEOREMA 6.2.3

Bajo las hipótesis del modelo normal y con la notación anterior, $p(\underline{\theta}/z)$ es multinormal L -dimensional, con vector media

$$\underline{\theta}^* = (T + NR)^{-1}(T\underline{\theta}_0 + NR\bar{x})$$

y matriz de precisión igual a $T + NR$, siendo

$$N = \begin{pmatrix} n_1 & 0 & \dots & 0 \\ 0 & n_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & n_L \end{pmatrix}, \text{ y}$$

\bar{x} el vector columna cuyas componentes son las medias muestrales de las submuestras extraídas en los respectivos estratos.

Demostración.- La función de verosimilitud satisface la relación siguiente:

$$\begin{aligned} p(z/\underline{\theta}) &\propto \exp\left(-\frac{1}{2} \sum_{i=1}^L \sum_{j=1}^{n_i} \left(\frac{x_{ij} - \theta_i}{\sigma_i}\right)^2\right) = \\ &= \prod_{i=1}^L \exp\left(-\frac{1}{2} \sum_{j=1}^{n_i} r_i \cdot (x_{ij} - \theta_i)^2\right) \end{aligned}$$

Pero al ser

$$\sum_{j=1}^{n_i} (x_{ij} - \theta_i)^2 = n_i \cdot (\theta_i - \bar{x}_i)^2 + \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$$

$$\text{con } \bar{x}_i = \sum_{j=1}^{n_i} x_{ij} / n_i$$

resulta finalmente

$$\begin{aligned} p(z/\underline{\theta}) &\propto \prod_{i=1}^L \exp\left(-\frac{1}{2} n_i r_i \cdot (\theta_i - \bar{x}_i)^2\right) = \exp\left(-\frac{1}{2} \sum_{i=1}^L n_i r_i (\theta_i - \bar{x}_i)^2\right) = \\ &= \exp\left(-\frac{1}{2} (\underline{\theta} - \bar{x})' NR (\underline{\theta} - \bar{x})\right) \end{aligned}$$

Dado que

$$p(\underline{\vartheta}) \propto \exp\left(-\frac{1}{2}(\underline{\vartheta} - \underline{\vartheta}_0)' T(\underline{\vartheta} - \underline{\vartheta}_0)\right)$$

la tesis del teorema se alcanza sin más que seguir idéntica técnica de demostración que la utilizada por DeGroot (1970), pág 176 .

(c.q.d.)

De este resultado puede observarse que la densidad $p(\underline{\vartheta}/z)$ sería la misma si en vez de considerar la muestra total z el investigador se hubiese ceñido a la observación del vector $\bar{\underline{x}}$. En efecto, dicho vector - cuyas componentes son independientes entre sí al serlo las submuestras sobre las que están definidas - se distribuye según una multinormal de media $\underline{\vartheta}$ y matriz de precisión NR , puesto que cada componente es normal de media ϑ_i y varianza σ_i^2/n_i . Así pues, resulta ser $p(\underline{\vartheta}/z) = p(\underline{\vartheta}/\bar{\underline{x}})$ y $p(\underline{\vartheta}/z) = p(\underline{\vartheta}/z_1, \dots, z_{i-1}, \bar{x}_i, z_{i+1}, \dots, z_L)$, $i = 1, 2, \dots, L$. De ahí que aplicando el segundo de los métodos de la sección 6.1, la información esperada $I^{\underline{\vartheta}}\{E^S(\underline{n}), N^L(\underline{\vartheta}_0, H_0)\}$ puede obtenerse estudiando la información contenida por una muestra unitaria del espacio

$$\bar{\underline{x}} = \prod_{i=1}^L \bar{x}_i$$

en el que toman valores el conjunto de estadísticos suficientes $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_L$. Dado que, según acaba de afirmarse, $\bar{\underline{x}}$ es $N^L(\underline{\vartheta}, (NR)^{-1})$, aplicando el resultado de Lindley (1956) para la información esperada sobre $\underline{\vartheta}$ proporcionada por una unidad muestral en el modelo normal n -dimensional, resulta finalmente que

$$\begin{aligned} I^{\underline{\vartheta}}\{E^S(\underline{n}), N^L(\underline{\vartheta}_0, H_0)\} &= I^{\underline{\vartheta}}\{E^S(\underline{1}, \bar{\underline{x}}), N^L(\underline{\vartheta}_0, H_0)\} = \\ &= \frac{1}{2} \log \frac{|H_0 + (NR)^{-1}|}{|(NR)^{-1}|} = \frac{1}{2} \log \frac{|H_0 N + H|}{|H|} \end{aligned} \quad (2)$$

como expresión de la información esperada que, sobre $\underline{\vartheta}$, proporciona el experimento $E^S(\underline{n})$ y bajo las hipótesis del modelo normal.

Los resultados establecidos por (1) y (2) coinciden, como queda recogido en el próximo

TEOREMA 6.2.4

Bajo la notación e hipótesis que conducen a los resultados (1) y (2), se verifica que

$$\frac{|AH_o A^o + M|}{|M|} = \frac{|H_o N + H|}{|H|}$$

Demostración.-

$$AH_o A^o + M = \begin{pmatrix} \sigma_{11}^o + \sigma_1^2 & \sigma_{11}^o & \dots & \sigma_{11}^o & \dots & \sigma_{1L}^o & \sigma_{1L}^o & \dots & \sigma_{1L}^o \\ \sigma_{11}^o & \sigma_{11}^o + \sigma_1^2 & \dots & \sigma_{11}^o & \dots & \sigma_{1L}^o & \sigma_{1L}^o & \dots & \sigma_{1L}^o \\ \dots & \dots \\ \sigma_{11}^o & \sigma_{11}^o & \dots & \sigma_{11}^o + \sigma_1^2 & \dots & \sigma_{1L}^o & \sigma_{1L}^o & \dots & \sigma_{1L}^o \\ \dots & \dots \\ \sigma_{1L}^o & \sigma_{1L}^o & \dots & \sigma_{1L}^o & \dots & \sigma_{LL}^o + \sigma_L^2 & \sigma_{LL}^o & \dots & \sigma_{LL}^o \\ \sigma_{1L}^o & \sigma_{1L}^o & \dots & \sigma_{1L}^o & \dots & \sigma_{LL}^o & \sigma_{LL}^o + \sigma_L^2 & \dots & \sigma_{LL}^o \\ \dots & \dots \\ \sigma_{1L}^o & \sigma_{1L}^o & \dots & \sigma_{1L}^o & \dots & \sigma_{LL}^o & \sigma_{LL}^o & \dots & \sigma_{LL}^o + \sigma_L^2 \end{pmatrix}$$

Restando la 2ª línea de la 1ª, la 3ª de la 2ª, ..., la n₁ª de la anterior, la (n₁ + 2)ª de la (n₁ + 1)ª, ..., la (n₂ + n₁)ª de la anterior, la (n₁ + n₂ + 2)ª de la (n₁ + n₂ + 1)ª, y así sucesivamente hasta restar la nª línea de la (n-1)ª, resulta: al tomar determinantes que

$$|AH_o A^o + M| = \det \begin{pmatrix} \sigma_1^2 & -\sigma_1^2 & \dots & \sigma_{11}^o & 0 & \dots & 0 & 0 & \dots & \sigma_{1L}^o & 0 \\ 0 & \sigma_1^2 & \dots & \sigma_{11}^o & 0 & \dots & 0 & 0 & \dots & \sigma_{1L}^o & 0 \\ \dots & \dots \\ 0 & 0 & \dots & -\sigma_1^2 & \dots & 0 & 0 & \dots & 0 & \dots & 0 \\ \sigma_{11}^o & \sigma_{11}^o & \dots & \sigma_{11}^o + \sigma_1^2 & \dots & \sigma_{1L}^o & \sigma_{1L}^o & \dots & \sigma_{1L}^o & \dots & \sigma_{1L}^o \end{pmatrix}$$

$$\left(\begin{array}{cccccccc}
 \dots & \dots \\
 0 & 0 & \dots & 0 & \dots & \sigma_L^2 & -\sigma_L^2 & \dots & 0 \\
 0 & 0 & \dots & 0 & \dots & 0 & \sigma_L^2 & \dots & 0 \\
 \dots & \dots \\
 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & -\sigma_L^2 \\
 \sigma_{1L}^0 & \sigma_{1L}^0 & \dots & \sigma_{1L}^0 & \dots & \sigma_{LL}^0 & \sigma_{LL}^0 & \dots & \sigma_{LL}^0 + \sigma_L^2
 \end{array} \right) \Bigg\}^{n_L}$$

Sumando la 1ª columna a la 2ª y desarrollando por el elemento (1,1) resulta

$$= \sigma_1^2 \left(\begin{array}{cccccccc}
 \sigma_1^2 & -\sigma_1^2 & \dots & \sigma_1^{n-1} & 0 & \dots & 0 & 0 & \dots & 0 \\
 \dots & \dots \\
 0 & 0 & \dots & -\sigma_1^2 & \dots & 0 & 0 & \dots & 0 & 0 \\
 2\sigma_{11}^0 & \sigma_{11}^0 & \dots & \sigma_{11}^0 + \sigma_1^2 & \dots & \sigma_{1L}^0 & \sigma_{1L}^0 & \dots & \sigma_{1L}^0 & \dots \\
 \dots & \dots \\
 0 & 0 & \dots & 0 & \dots & \sigma_L^2 & -\sigma_L^2 & \dots & 0 & 0 \\
 0 & 0 & \dots & 0 & \dots & 0 & \sigma_L^2 & \dots & 0 & 0 \\
 \dots & \dots \\
 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & -\sigma_L^2 & 0 \\
 2\sigma_{1L}^0 & \sigma_{1L}^0 & \dots & \sigma_{1L}^0 & \dots & \sigma_{LL}^0 & \sigma_{LL}^0 & \dots & \sigma_{LL}^0 + \sigma_L^2 & \dots
 \end{array} \right) \Bigg\}^{n_L}$$

Repetiendo idéntica operación anterior otras $n_1 - 2$ veces y desarrollando siempre por el elemento (1,1) resulta:

$$= \sigma_1^{2(n_1-1)} \left(\begin{array}{cccccccc}
 n_1 \sigma_{11}^0 + \sigma_1^2 & \dots & \sigma_{1L}^0 & \sigma_{1L}^0 & \dots & \sigma_{1L}^0 & \dots & \dots \\
 \dots & \dots \\
 n_1 \sigma_{12}^0 & \dots & \sigma_{2L}^0 & \sigma_{2L}^0 & \dots & \sigma_{2L}^0 & \dots & \dots \\
 \dots & \dots \\
 0 & \dots & \sigma_L^2 & -\sigma_L^2 & \dots & 0 & \dots & \dots \\
 \dots & \dots \\
 0 & 0 & 0 & \dots & -\sigma_L^2 & \dots & \dots & \dots
 \end{array} \right) \Bigg\}^{n_L}$$

(sigue)

$$\left(\begin{array}{cccccc} n_1 \sigma_{1L}^0 & \dots & \sigma_{LL}^0 & \sigma_{LL}^0 & \dots & \sigma_{LL}^0 + \sigma_L^2 \end{array} \right)$$

Sumando la segunda columna a la 3ª y desarrollando por el elemento (2,2)

$$= \sigma_1^{2(n_1-1)} \sigma_2^2 \left(\begin{array}{cccccc} n_1 \sigma_{11}^0 + \sigma_1^2 & \sigma_{12}^0 & \dots & \sigma_{12}^0 & \dots & \sigma_{1L}^0 & \sigma_{1L}^0 \\ 0 & \sigma_2^2 & \dots & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & -\sigma_2^2 & \dots & 0 & 0 \\ n_1 \sigma_{12}^0 & 2 \sigma_{22}^0 & \dots & \sigma_{22}^0 + \sigma_2^2 & \dots & \sigma_{2L}^0 & \sigma_{2L}^0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & \dots & \sigma_L^2 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ n_1 \sigma_{1L}^0 & 2 \sigma_{2L}^0 & \dots & \sigma_{2L}^0 & \dots & \sigma_{LL}^0 & \sigma_{LL}^0 + \sigma_L^2 \end{array} \right)$$

Repetiendo idéntica operación otras n_2-2 veces y desarrollando siempre por el elemento (2,2) resulta:

$$= \sigma_1^{2(n_1-1)} \sigma_2^{2(n_2-1)} \left(\begin{array}{cccccc} n_1 \sigma_{11}^0 + \sigma_1^2 & n_2 \sigma_{12}^0 & \dots & \sigma_{1L}^0 & \dots & \sigma_{1L}^0 \\ n_1 \sigma_{12}^0 & n_2 \sigma_{22}^0 + \sigma_2^2 & \dots & \sigma_{2L}^0 & \dots & \sigma_{2L}^0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_L^2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & \dots & -\sigma_L^2 \\ n_1 \sigma_{1L}^0 & n_2 \sigma_{2L}^0 & \dots & \sigma_{LL}^0 & \dots & \sigma_{LL}^0 + \sigma_L^2 \end{array} \right)$$

Iterando estas operaciones para $j = 3, 4, \dots, L-1$, resulta:

$$\left(\begin{array}{cccccc} n_1 \sigma_{11}^0 + \sigma_1^2 & n_2 \sigma_{12}^0 & \dots & n_{L-1} \sigma_{1L-1}^0 & \sigma_{1L}^0 & \dots & \sigma_{1L}^0 \\ n_1 \sigma_{12}^0 & n_2 \sigma_{22}^0 + \sigma_2^2 & \dots & n_{L-1} \sigma_{2L-1}^0 & \sigma_{2L}^0 & \dots & \sigma_{2L}^0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{array} \right)$$

(sigue)

$$= \prod_{i=1}^{L-1} \sigma_i^{2(n_i-1)} \left| \begin{array}{cccccc} n_1 \sigma_{1L-1}^0 & n_2 \sigma_{2L-1}^0 & \dots & n_{L-1} \sigma_{L-1L-1}^0 + \sigma_{L-1}^2 & \sigma_{L-1}^0 & \dots & \sigma_{L-1}^0 \\ 0 & 0 & \dots & 0 & \sigma_L^2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 0 & \dots & -\sigma_L^2 \\ n_1 \sigma_{1L}^0 & n_2 \sigma_{2L}^0 & \dots & n_{L-1} \sigma_{L-1L}^0 & \sigma_{1L}^0 & \dots & \sigma_{LL}^0 + \sigma_L^2 \end{array} \right|^{n_L}$$

Repetiendo idénticos procesos anteriores para $j=L$, resulta finalmente:

$$= \prod_{i=1}^L \sigma_i^{2(n_i-1)} \left| \begin{array}{cccc} n_1 \sigma_{11}^0 + \sigma_1^2 & n_2 \sigma_{12}^0 & \dots & n_L \sigma_{1L}^0 \\ n_1 \sigma_{12}^0 & n_2 \sigma_{22}^0 + \sigma_2^2 & \dots & n_L \sigma_{2L}^0 \\ \dots & \dots & \dots & \dots \\ n_1 \sigma_{1L}^0 & n_2 \sigma_{2L}^0 & \dots & n_L \sigma_{LL}^0 + \sigma_L^2 \end{array} \right| =$$

$$= \prod_{i=1}^L \sigma_i^{2(n_i-1)} \cdot |H_0 N + H|$$

Y al ser $|H| = \prod_{i=1}^L \sigma_i^2$ y $|M| = \prod_{i=1}^L \sigma_i^{2n_i}$

se obtiene que

$$\frac{|AH_0 A^0 + M|}{|M|} = |H_0 N + H| \cdot \prod_{i=1}^L \frac{\sigma_i^{2(n_i-1)}}{\sigma_i^{2n_i}} = \frac{|H_0 N + H|}{|H|}$$

(c.q.d.)

En el caso de independencia de las componentes ϑ_i del vector paramétrico $\underline{\vartheta}$ se obtiene el mismo resultado que el contenido en el teorema 6.2.2, como era de esperar. En efecto, siendo H_0 diagonal:

$$H_0 N + H = \begin{pmatrix} n_1 \sigma_{11}^0 + \sigma_1^2 & 0 & \dots & 0 \\ 0 & n_2 \sigma_{22}^0 + \sigma_2^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & n_L \sigma_{LL}^0 + \sigma_L^2 \end{pmatrix}, |H_0 N + H| = \prod_{i=1}^L (\sigma_i^{2+n_i} \sigma_{ii}^0)$$

de donde

$$I^g \{E^S(\underline{n}), N^L(\underline{\vartheta}_0, H_0)\} = \frac{1}{2} \log \prod_{i=1}^L \frac{\sigma_i^2 + n_i \sigma_{ii}^0}{\sigma_i^2} = \sum_{i=1}^L \frac{1}{2} \log \left(1 + n_i \frac{\sigma_{ii}^0}{\sigma_i^2} \right) =$$

$$= \sum_{i=1}^L I^g \{E(n_i), N(\vartheta_{i0}, \sigma_{ii}^0)\}$$

Veamos dos casos particulares. En un proceso de muestreo biestratificado las matrices anteriores resultan

$$H_0 = \begin{pmatrix} \sigma_{11}^0 & \sigma_{12}^0 \\ \sigma_{12}^0 & \sigma_{22}^0 \end{pmatrix}, \quad N = \begin{pmatrix} n_1 & 0 \\ 0 & n_2 \end{pmatrix}, \quad H = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$$

de forma que

$$|H_0 N + H| = (n_1 \sigma_{11}^0 + \sigma_1^2)(n_2 \sigma_{22}^0 + \sigma_2^2) - n_1 n_2 \sigma_{12}^0 \sigma_{12}^0$$

de donde

$$I^g \{E^S(n_1, n_2), N^2(\underline{\vartheta}_0, H_0)\} = \left(1 + n_1 \frac{\sigma_{11}^0}{\sigma_1^2} \right) \left(1 + n_2 \frac{\sigma_{22}^0}{\sigma_2^2} \right) - n_1 n_2 \frac{(\sigma_{12}^0)^2}{\sigma_1^2 \sigma_2^2}$$

Análogamente, la existencia de tres estratos da lugar a la siguiente expresión de la información esperada en el modelo normal:

$$I^g \{E^S(n_1, n_2, n_3), N^3(\underline{\vartheta}_0, H_0)\} = 1 + \sum_{i=1}^3 n_i \frac{\sigma_{ii}^0}{\sigma_i^2} +$$

$$+ \sum_{i=1}^3 \frac{|\text{Adj } \sigma_{ii}^0|}{|\text{Adj } \sigma_i^2|} \prod_{j \neq i} n_j + \frac{|H_0|}{|H|} \prod_{i=1}^3 n_i$$

donde la notación Adj representa el respectivo adjunto en la matriz correspondiente.

III. Obtención de las informaciones esperadas sobre θ y ϑ_i .

Cuando el investigador, al realizar el experimento $E^S(\underline{n})$ asociado a un proceso de estratificación, está más interesado en conocer la cantidad aleatoria unidimensional

$$\theta = J' \underline{\vartheta} \quad \text{con} \quad J' = (h_1, h_2, \dots, h_L)$$

deberá cuantificar la información esperada que sobre ϑ proporciona $E^S(\underline{n})$, esto es, deberá calcular $I^{\vartheta} \{E^S(\underline{n}), N^L(\underline{\vartheta}_0, H_0)\}$. Precisamente el corolario 6.1.1 permite emplear para su obtención idéntica técnica a través de estadísticos suficientes en todos los estratos. Así pues, consideraremos nuevamente el vector $\underline{\bar{x}} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_L)$ que se distribuye según una multi-normal L-dimensional

$$p(\underline{\bar{x}} / \underline{\vartheta}) = N^L(\underline{\vartheta}, HN^{-1}) \quad \underline{\bar{x}} \in \bar{X}$$

Siendo $p(\underline{\vartheta}) = N^L(\underline{\vartheta}_0, H_0)$, el teorema 6.2.3 indica que

$$p(\underline{\vartheta} / \underline{\bar{x}}) = N^L((H_0^{-1} + NH^{-1})^{-1}(H_0^{-1}\underline{\vartheta}_0 + NH^{-1}\underline{\bar{x}}), (H_0^{-1} + NH^{-1})^{-1})$$

De acuerdo, pues, con el corolario 6.1.1 se verifica que

$$I^{\vartheta} \{E^S(\underline{n}), N^L(\underline{\vartheta}_0, H_0)\} = I^{\vartheta} \{E(\underline{1}, \underline{\bar{X}}), N^L(\underline{\vartheta}_0, H_0)\}$$

Por otra parte, de las anteriores densidades de probabilidad se deduce que $p(\vartheta)$ es una normal unidimensional, de media $J'\underline{\vartheta}_0$ y varianza $J'H_0J$. Así pues

$$p(\vartheta) = N(J'\underline{\vartheta}_0, J'H_0J)$$

mientras que, análogamente,

$$p(\vartheta / \underline{\bar{x}}) = N(J'(H_0^{-1} + NH^{-1})^{-1}(H_0^{-1}\underline{\vartheta}_0 + NH^{-1}\underline{\bar{x}}), J'(H_0^{-1} + NH^{-1})^{-1}J)$$

Demostremos primeramente el siguiente

LEMA 6.2.1

Bajo la notación desarrollada

$$p(\underline{\bar{x}}) = N^L(\underline{\vartheta}_0, H_0 + HN^{-1})$$

Demostración.- En efecto,

$$p(\underline{\bar{x}}) = \int p(\underline{\bar{x}} / \underline{\vartheta}) \cdot p(\underline{\vartheta}) \cdot d\underline{\vartheta} = \int N^L(\underline{\vartheta}, HN^{-1}) \cdot N^L(\underline{\vartheta}_0, H_0) \cdot d\underline{\vartheta}$$

Denotemos por $P = (HN^{-1})^{-1}$ y $P_0 = H_0^{-1}$ a las respectivas matrices de precisión de las densidades $p(\underline{\bar{x}} / \underline{\vartheta})$ y $p(\underline{\vartheta})$. Resultará

$$p(\underline{\bar{x}}) = \frac{\sqrt{|PP_0|}}{(2\pi)^L} \exp\left(-\frac{1}{2}(\underline{\bar{x}} - \underline{\vartheta})' P (\underline{\bar{x}} - \underline{\vartheta}) - \frac{1}{2}(\underline{\vartheta} - \underline{\vartheta}_0)' P_0 (\underline{\vartheta} - \underline{\vartheta}_0)\right) d\underline{\vartheta}$$

Estudie el exponente.

$$(\bar{x} - \underline{\theta})' P (\bar{x} - \underline{\theta}) + (\underline{\theta} - \underline{\theta}_0)' P_0 (\underline{\theta} - \underline{\theta}_0) = A + B$$

donde

$$A = \{ \underline{\theta} - (P + P_0)^{-1} (P_0 \underline{\theta}_0 + P \bar{x}) \}' (P + P_0) \{ \underline{\theta} - (P + P_0)^{-1} (P_0 \underline{\theta}_0 + P \bar{x}) \} =$$

$$= \{ (P + P_0)^{-1} ((P + P_0) \underline{\theta} - P_0 \underline{\theta}_0 - P \bar{x}) \}' (P + P_0)$$

$$\{ (P + P_0)^{-1} ((P + P_0) \underline{\theta} - P_0 \underline{\theta}_0 - P \bar{x}) \}' =$$

$$= \{ P_0 (\underline{\theta} - \underline{\theta}_0) + P (\underline{\theta} - \bar{x}) \}' (P + P_0)^{-1} \{ P_0 (\underline{\theta} - \underline{\theta}_0) + P (\underline{\theta} - \bar{x}) \}$$

$$B = (\bar{x} - \underline{\theta}_0)' P (P + P_0)^{-1} P_0 (\bar{x} - \underline{\theta}_0) =$$

$$= \{ (\bar{x} - \underline{\theta}) + (\underline{\theta} - \underline{\theta}_0) \}' P (P + P_0)^{-1} P_0 \{ (\bar{x} - \underline{\theta}) + (\underline{\theta} - \underline{\theta}_0) \}$$

resulta

$$p(\bar{x}) = \frac{\sqrt{|PP_0|}}{(2\pi)^L} \exp(-\frac{1}{2} B) \int \exp(-\frac{1}{2} A) \cdot d\underline{\theta} =$$

$$= \frac{\sqrt{|PP_0|}}{(2\pi)^L} \exp(-\frac{1}{2} B) \cdot \frac{(\sqrt{2\pi})^L}{\sqrt{|P + P_0|}} =$$

$$= \frac{1}{(\sqrt{2\pi})^L} \sqrt{\frac{|PP_0|}{|P + P_0|}} \exp(-\frac{1}{2} (\bar{x} - \underline{\theta}_0)' P (P + P_0)^{-1} P_0 (\bar{x} - \underline{\theta}_0))$$

esto es,

$$p(\bar{x}) = N^L(\underline{\theta}_0, (P(P + P_0)^{-1} P_0)^{-1}) = N^L(\underline{\theta}_0, H_0 + H N^{-1})$$

(c.q.d.)

Este resultado permite, a su vez, demostrar que la densidad sobre θ calculada a través de $p(\bar{x})$ y $p(\theta/\bar{x})$ es igual a la obtenida de $p(\theta)$ a través de la transformación lineal que define a θ . Este es el sentido del próximo

LEMA 6.2.2

$$\int p(\bar{x}) \cdot p(\theta/\bar{x}) \cdot d\bar{x} = N^L(J' \underline{\theta}_0, J' H_0 J)$$

Demostración.- En efecto, la función característica de la distribución de θ definida a partir de $p(\bar{x})$ y $p(\theta/\bar{x})$ viene dada por

$$\begin{aligned}
\varphi(t) &= \int e^{it\theta} \left\{ \int p(\underline{\bar{x}}) \cdot p(\theta / \underline{\bar{x}}) \cdot d\underline{\bar{x}} \right\} d\theta = \\
&= \int p(\underline{\bar{x}}) \left\{ \int e^{it\theta} p(\theta / \underline{\bar{x}}) \cdot d\theta \right\} d\underline{\bar{x}} = \int p(\underline{\bar{x}}) \left\{ \right. \\
&\left. \int e^{it\theta} N(J'(H_0^{-1} + NH^{-1})^{-1}(H_0^{-1}\underline{\theta}_0 + NH^{-1}\underline{\bar{x}}), J'(H_0^{-1} + NH^{-1})^{-1}J) \cdot d\theta \right\} d\underline{\bar{x}} = \\
&= \int p(\underline{\bar{x}}) \cdot e^{itJ'(H_0^{-1} + NH^{-1})^{-1}(H_0^{-1}\underline{\theta}_0 + NH^{-1}\underline{\bar{x}}) - \frac{1}{2}t^2 J'(H_0^{-1} + NH^{-1})^{-1}J} \cdot d\underline{\bar{x}} = \\
&= e^{itJ'(H_0^{-1} + NH^{-1})^{-1}H_0^{-1}\underline{\theta}_0 - \frac{1}{2}t^2 J'(H_0^{-1} + NH^{-1})^{-1}J} \\
&\cdot \int e^{itJ'(H_0^{-1} + NH^{-1})^{-1}NH^{-1}\underline{\bar{x}} - \frac{1}{2}t^2 J'(H_0^{-1} + NH^{-1})^{-1}J} \cdot p(\underline{\bar{x}}) \cdot d\underline{\bar{x}} = (\text{aplicando el lema 6.2.1}) = \\
&= \exp(itJ'(H_0^{-1} + NH^{-1})^{-1}H_0^{-1}\underline{\theta}_0 - \frac{1}{2}t^2 J'(H_0^{-1} + NH^{-1})^{-1}J + \\
&+ itJ'(H_0^{-1} + NH^{-1})^{-1}NH^{-1}\underline{\theta}_0 - \frac{1}{2}t^2 J'(H_0^{-1} + NH^{-1})^{-1}NH^{-1}(H_0 + NH^{-1})NH^{-1}(\\
&(H_0^{-1} + NH^{-1})^{-1}J) = \exp(itJ'(H_0^{-1} + NH^{-1})^{-1}(H_0^{-1} + NH^{-1})\underline{\theta}_0 - \\
&- \frac{1}{2}t^2 J'(H_0^{-1} + NH^{-1})^{-1}J - \frac{1}{2}t^2 J'H_0NH^{-1}(H_0^{-1} + NH^{-1})^{-1}J) = \\
&= \exp(itJ'\underline{\theta}_0 - \frac{1}{2}t^2 J'(I + H_0NH^{-1})(H_0^{-1} + NH^{-1})^{-1}J) = \\
&= \exp(itJ'\underline{\theta}_0 - \frac{1}{2}t^2 J'H_0J)
\end{aligned}$$

(c.q.d.)

Los lemas anteriores permiten obtener la información esperada sobre θ proporcionada por el experimento $E^S(\underline{n})$ en el modelo normal.

TEOREMA 6.2.5

En el modelo normal y bajo la notación anterior, siendo $\theta = J'\underline{\theta}$,

$$I^{\theta} \left\{ E^S(\underline{n}), N^L(\underline{\theta}_0, H_0) \right\} = \frac{1}{2} \log \frac{J'H_0J}{J'(H_0^{-1} + NH^{-1})^{-1}J}$$

Demostración.- Por el corolario 6.1.1

$$\begin{aligned}
I^\theta \{E^S(\underline{n}), N^L(\underline{\theta}_0, H_0)\} &= I^\theta \{E(\underline{1}, \bar{x}), N^L(\underline{\theta}_0, H_0)\} = \\
&= \iint p(\bar{x}) \cdot p(\theta / \bar{x}) \cdot \log \frac{p(\theta / \bar{x})}{p(\theta)} \cdot d\theta \cdot d\bar{x} = \\
&= \iint p(\bar{x}) \cdot p(\theta / \bar{x}) \cdot \log \left\{ \sqrt{\frac{J^* H J}{J^*(H_0^{-1} + NH^{-1})^{-1} J}} \cdot \exp \left(\right. \right. \\
&\quad \left. \left. \left(-\frac{1}{2} \frac{(\theta - J^*(H_0^{-1} + NH^{-1})^{-1}(H_0^{-1} \underline{\theta}_0 + NH^{-1} \bar{x}))^2}{J^*(H_0^{-1} + NH^{-1})^{-1} J} + \frac{1}{2} \frac{(\theta - J^* \underline{\theta}_0)^2}{J^* H J} \right) \right\} \cdot \\
&= \log \sqrt{\frac{J^* H J}{J^*(H_0^{-1} + NH^{-1})^{-1} J}} - \frac{1}{2} \iint p(\bar{x}) \cdot p(\theta / \bar{x}) \cdot \frac{(\theta - E_{p(\theta / \bar{x})}(\theta))^2}{V_{p(\theta / \bar{x})}(\theta)} \cdot d\theta \cdot d\bar{x} + \\
&+ \frac{1}{2} \iint p(\bar{x}) \cdot p(\theta / \bar{x}) \cdot \frac{(\theta - E_{p(\theta)}(\theta))^2}{V_{p(\theta)}(\theta)} \cdot d\theta \cdot d\bar{x} = (\text{por lema 6.2.2.}) = \\
&= \frac{1}{2} \log \frac{J^* H J}{J^*(H_0^{-1} + NH^{-1})^{-1} J} - \frac{1}{2} + \frac{1}{2} \int p(\theta) \cdot \frac{(\theta - E_{p(\theta)}(\theta))^2}{V_{p(\theta)}(\theta)} \cdot d\theta = \\
&= \frac{1}{2} \log \frac{J^* H J}{J^*(H_0^{-1} + NH^{-1})^{-1} J}
\end{aligned}$$

y donde $E_{p(\omega)}$ y $V_{p(\omega)}$ denotan, respectivamente, la media y la varian-
za calculadas sobre la distribución de la variable aleatoria ω .

(c.q.d.)

El resultado anterior permite obtener la información esperada $I^\theta \{E^S(\underline{n}), N^L(\underline{\theta}_0, H_0)\}$ como caso particular. Cuando el investigador es
tá interesado en la estimación de la única cantidad θ_i y dispone de la
muestra z resultado del experimento $E^S(\underline{n})$, deberá ser capaz de calcu-
lar o cuantificar la anterior información esperada.

Sea el vector J definido como

$$J' = (0 \dots 0 \quad 1 \quad 0 \dots 0) \quad \text{siendo} \quad \theta_i = J' \vartheta$$

y cuyos elementos son todos nulos salvo el que ocupa el lugar i -ésimo. Notemos que

$$J' H_0 J = \sigma_{ii}^0 \quad \text{y} \quad J' (H_0^{-1} + N H^{-1})^{-1} J = a_{ii}$$

donde a_{ii} es el elemento (i,i) de la matriz $(H_0^{-1} + N H^{-1})^{-1}$. Así pues, siendo m_{ii} el elemento (i,i) de la matriz $H_0^{-1} + N H^{-1}$, que es la matriz de precisión de la densidad $p(\vartheta / \bar{x})$, resultará ser:

$$J' (H_0^{-1} + N H^{-1})^{-1} J = \frac{|\text{Adj } m_{ii}|}{|H_0^{-1} + N H^{-1}|} = \frac{|\text{Adj} \left(\frac{\text{Adj } \sigma_{ii}^0}{|H_0|} + \frac{n_i}{\sigma_i^2} \right)|}{|H_0^{-1} + N H^{-1}|}$$

De ahí que por el teorema 6.2.5 y con la notación anterior sea evidente el siguiente

COROLARIO 6.2.1

Siendo m_{ii} el elemento (i,i) de la matriz $H_0^{-1} + N H^{-1}$,

$$I^{\theta_i} \{ E^S(\underline{n}), N^L(\vartheta_0, H_0) \} = \frac{1}{2} \log \frac{\sigma_{ii}^0 |H_0^{-1} + N H^{-1}|}{|\text{Adj } m_{ii}|}$$

Notemos finalmente que las expresiones de todas las anteriores informaciones esperadas son función del vector tamaño muestral \underline{n} . Este hecho - obvio, por otra parte - las hace de especial aplicación para determinar el tamaño muestral óptimo, conforme ya se indicó en la exposición realizada en el capítulo quinto.

6.3 COMPORTAMIENTO ASINTOTICO DE $I^{\vartheta} \{ E^S(\underline{n}), p(\vartheta) \}$

Como ya se estableció en la sección 6.1 el cálculo exacto de la información esperada sobre ϑ proporcionada por el experimento $E^S(\underline{n})$

resulta difícil en una gran cantidad de casos. De ahí que sea esencial el conocer al menos su comportamiento asintótico a fin de poder diseñar adecuadamente el experimento $E^S(\underline{n})$ y determinar, por ejemplo, su tamaño óptimo. Precisamente en esta sección va a generalizarse para la situación expresada por $E^S(\underline{n})$ el resultado de aproximación obtenido por Bernardo (1979)* y al que ya se hizo referencia en la sección 2.3 al comparar las informaciones esperadas de los dos modelos allí desarrollados.

Bernardo (1978) demuestra que $I^S\{E(n), p(\vartheta)\}$ es invariante ante transformaciones biyectivas de la cantidad de interés ϑ . Este resultado es satisfecho por la información esperada $I^S\{E^S(\underline{n}), p(\underline{\vartheta})\}$ como queda recogido en el siguiente

TEOREMA 6.3.1

Si $\underline{\psi} = \psi(\underline{\vartheta})$ es una transformación biyectiva de la cantidad de interés $\underline{\vartheta}$, las informaciones esperadas que el experimento $E^S(\underline{n})$ proporciona sobre ambos vectores $\underline{\psi}$ y $\underline{\vartheta}$, coinciden.

Demostración.- Siendo $\underline{\psi} = \psi(\underline{\vartheta})$ una transformación biyectiva se verifica que

$$p(\underline{\psi}) = \{p(\underline{\vartheta})\}_{\underline{\vartheta} = \underline{\psi}} \cdot |J|, \quad y$$

$$p(\underline{\psi}/z) = \{p(\underline{\vartheta}/z)\}_{\underline{\vartheta} = \underline{\psi}} \cdot |J|.$$

donde

$$|J| = \left| \frac{\partial \underline{\vartheta}}{\partial \underline{\psi}} \right| = \begin{vmatrix} \frac{\partial \vartheta_1}{\partial \psi_1} & \dots & \frac{\partial \vartheta_L}{\partial \psi_1} \\ \dots & \dots & \dots \\ \frac{\partial \vartheta_1}{\partial \psi_L} & \dots & \frac{\partial \vartheta_L}{\partial \psi_L} \end{vmatrix}$$

De ahí que siendo z el resultado obtenido tras la realización de $E^S(\underline{n})$ y definiendo

$$I^S\{z, p(\underline{\vartheta})\} = \int p(\underline{\vartheta}/z) \cdot \log \frac{p(\underline{\vartheta}/z)}{p(\underline{\vartheta})} \cdot d\underline{\vartheta}, \quad e$$

$$I^{\underline{\vartheta}}\{E^S(\underline{n}), p(\underline{\vartheta})\} = \int p(z) \cdot I^{\underline{\vartheta}}\{z, p(\underline{\vartheta})\} \cdot dz$$

ante el cambio múltiple de variables representado por la transformación biyectiva resulta que

$$\begin{aligned} I^{\underline{\vartheta}}\{z, p(\underline{\vartheta})\} &= \int p(\underline{\vartheta}/z) \cdot \log \frac{p(\underline{\vartheta}/z) \cdot |J|}{p(\underline{\vartheta}) \cdot |J|} \cdot d\underline{\vartheta} = \\ &= \int p(\underline{\psi}/z) \cdot \log \frac{p(\underline{\psi}/z)}{p(\underline{\psi})} \cdot d\underline{\psi} = I^{\underline{\psi}}\{z, p(\underline{\psi})\} \end{aligned}$$

De ahí que trivialmente

$$\begin{aligned} I^{\underline{\vartheta}}\{E^S(\underline{n}), p(\underline{\vartheta})\} &= \int p(z) \cdot I^{\underline{\vartheta}}\{z, p(\underline{\vartheta})\} \cdot dz = \int p(z) \cdot I^{\underline{\psi}}\{z, p(\underline{\psi})\} \cdot dz = \\ &= I^{\underline{\psi}}\{E^S(\underline{n}), p(\underline{\psi})\} \end{aligned}$$

(c.q.d.)

En consecuencia, puede simplificarse el cálculo de $I^{\underline{\vartheta}}$ considerando una transformación $\underline{\psi} = \underline{\psi}(\underline{\vartheta})$ cuya densidad inicial $p(\underline{\psi}) = \{p(\underline{\vartheta})\}_{\underline{\vartheta}=\underline{\psi}} \cdot |J|$ sea aproximadamente multinormal y calculando $I^{\underline{\psi}}$ en su lugar. Evidentemente que existe siempre una transformación de $\underline{\vartheta}$ cuya distribución es exactamente multinormal pero, en general, tal transformación exacta es complicada (ver, por ejemplo, al respecto la sección 2.3). Puesto que lo que se pretende es proporcionar un resultado manejable resultará generalmente más práctico utilizar una transformación sencilla aunque su densidad $p(\underline{\psi})$ sea tan sólo aproximadamente multinormal.

Veamos ahora que si $p(\underline{\psi})$ es aproximadamente multinormal, la densidad posterior $p(\underline{\psi}/z)$ es asintóticamente normal multivariante. Para ello recordemos que la muestra z resultado de $E^S(\underline{n})$ está formada por un conjunto de L submuestras aleatorias independientes z_i , de forma que

$$p(z/\underline{\vartheta}) = \prod_{i=1}^L p(z_i/\vartheta_i) \quad \text{y} \quad p(z_i/\vartheta_i) = \prod_{j=1}^{n_i} p(x_{ij}/\vartheta_i)$$

Por otra parte, siendo $\underline{\psi} = \underline{\psi}(\underline{\vartheta})$ biyectiva, puede escribirse que $\underline{\vartheta} = F(\underline{\psi})$ donde $F^{-1} = \underline{\psi}$ es la correspondiente transformación inversa. Particularizando para cada componente escribiremos que $\vartheta_i = F_i(\underline{\psi})$, de forma que resultará ser

$$\underline{\vartheta} = (F_1(\underline{\psi}), F_2(\underline{\psi}), \dots, F_L(\underline{\psi}))$$

De ahí que pueda escribirse que

$$p(z/\underline{\psi}) = p(z/\underline{\psi}(\underline{\theta})) = p(z/F_1(\underline{\psi}), F_2(\underline{\psi}), \dots, F_L(\underline{\psi})) = \prod_{i=1}^L p(z_i/F_i(\underline{\psi}))$$

Bajo esta notación puede ya demostrarse el resultado citado.

TEOREMA 6.3.2

Siendo $p(\underline{\psi})$ aproximadamente multinormal de vector media $\underline{\psi}_0$ y matriz de precisión T_0 y para valores grandes en todas sus componentes del vector tamaño muestral \underline{n} , se verifica que $p(\underline{\psi}/z)$ es también aproximadamente normal multivariante con vector media $\underline{\psi}^*$ y matriz de precisión $T_0 + H(z)$, donde

$$\underline{\psi}^* = (T_0 + H(z))^{-1} \cdot (T_0 \underline{\psi}_0 + H(z) \cdot \underline{\hat{\psi}})$$

$$H(z) = \sum_{i=1}^L H^i(z_i)$$

$H^i(z_i)$ es la matriz de información de elementos iguales a

$$-\frac{\partial^2}{\partial \psi_p \partial \psi_q} \log p(z_i/F_i(\underline{\psi})) \Big|_{\underline{\psi}^i}, \quad p, q = 1, 2, \dots, L$$

$$\underline{\hat{\psi}} = H(z)^{-1} \sum_{i=1}^L H^i(z_i) \cdot \underline{\hat{\psi}}^i$$

y $\underline{\hat{\psi}}^i$ es el estimador de máxima verosimilitud de $\underline{\psi}$ calculado a partir de la densidad $p(z_i/F_i(\underline{\psi}))$.

Demostración.- Puesto que $p(\underline{\psi})$ es aproximadamente normal se tiene, en virtud del teorema de Bayes, que

$$p(\underline{\psi}/z) \propto p(\underline{\psi}) \cdot p(z/\underline{\psi}) \simeq \frac{|T_0|^{1/2}}{(\sqrt{2\pi})^L} \cdot \exp\left(-\frac{1}{2} (\underline{\psi} - \underline{\psi}_0)' T_0 (\underline{\psi} - \underline{\psi}_0)\right) \cdot p(z/\underline{\psi}) \propto$$

$$\propto p(z/\underline{\psi}) \cdot \exp\left(-\frac{1}{2} (\underline{\psi} - \underline{\psi}_0)' T_0 (\underline{\psi} - \underline{\psi}_0)\right) =$$

$$= \exp\left(\log p(z/\underline{\psi}) - \frac{1}{2} (\underline{\psi} - \underline{\psi}_0)' T_0 (\underline{\psi} - \underline{\psi}_0)\right)$$

Consideremos el factor $\exp(\log p(z/\underline{\psi}))$. Resulta ser

$$\log p(z/\underline{\psi}) = \sum_{i=1}^L \log p(z_i/F_i(\underline{\psi})) ,$$

esto es, una suma de logaritmos de densidades de probabilidad con ámbito de aplicación en un único estratos. Para cada uno de esos logaritmos es conocido (Lindley, 1965, y Bernardo, 1979) que, siendo $\hat{\psi}_i$ el estimador de máxima verosimilitud de $\underline{\psi}$ calculado a partir de su respectiva densidad

$p(z_i/F_i(\underline{\psi}))$, se cumple en un entorno de dicha estimación que

$$\log p(z_i/F_i(\underline{\psi})) = \log p(z_i/\hat{\psi}_i) - \frac{1}{2} (\underline{\psi} - \hat{\psi}_i)' H^i(z_i) (\underline{\psi} - \hat{\psi}_i) + R_i$$

donde R_i es, en muestras grandes, despreciable frente al segundo término del segundo miembro de la igualdad anterior (es del orden de $O(n_i^{-2})$) y

$$\hat{\psi}_i = \begin{pmatrix} \hat{\psi}_1^i \\ \vdots \\ \hat{\psi}_L^i \end{pmatrix} , \quad \lim_{n_i \rightarrow \infty} R_i = 0 ,$$

y $H^i(z_i)$ es la matriz simétrica de información, de elementos $h_{jk}^i(z_i)$ definidos como

$$h_{jk}^i(z_i) = - \frac{\partial^2}{\partial \psi_j \partial \psi_k} \log p(z_i/F_i(\underline{\psi})) \Big|_{\hat{\psi}_i} ,$$

verificándose además por la ley fuerte de los grandes números que

$$\lim_{n_i \rightarrow \infty} \frac{1}{n_i} \log p(z_i/F_i(\underline{\psi})) = \int p(x_{ij}/F_i(\underline{\psi})) \cdot \log p(x_{ij}/F_i(\underline{\psi})) \cdot dx_{ij} , \text{ y}$$

$$\lim_{n_i \rightarrow \infty} \frac{1}{n_i} \frac{\partial^r}{\partial \psi_1^m \dots \partial \psi_L^m} \log p(z_i/F_i(\underline{\psi})) =$$

$$= \int p(x_{ij}/F_i(\underline{\psi})) \cdot \frac{\partial^r}{\partial \psi_1^m \dots \partial \psi_L^m} \log p(x_{ij}/F_i(\underline{\psi})) \cdot dx_{ij}$$

De ahí que

$$\sum_{i=1}^L \log p(z_i/F_i(\underline{\psi})) = \sum_{i=1}^L \log p(z_i/\hat{\psi}_i) - \frac{1}{2} \sum_{i=1}^L (\underline{\psi} - \hat{\psi}_i)' H^i(z_i) (\underline{\psi} - \hat{\psi}_i) +$$

$$+ \sum_{i=1}^L R_i = \sum_{i=1}^L \log p(z_i / \hat{\psi}^i) - \frac{1}{2} (\underline{\psi} - \hat{\underline{\psi}})' H(z) (\underline{\psi} - \hat{\underline{\psi}}) + Q +$$

$$+ \sum_{i=1}^L R_i$$

donde Q es una suma de términos en los que no aparece el parámetro $\underline{\psi}$ y

$$H(z) = \sum_{i=1}^L H^i(z_i), \quad \hat{\underline{\psi}} = H(z)^{-1} \cdot \sum_{i=1}^L H^i(z_i) \cdot \hat{\underline{\psi}}^i$$

De donde

$$p(z / \underline{\psi}) = \exp(\log p(z / \underline{\psi})) \propto \prod_{i=1}^L p(z_i / \hat{\underline{\psi}}^i) \cdot \exp\left(-\frac{1}{2} (\underline{\psi} - \hat{\underline{\psi}})' H(z) (\underline{\psi} - \hat{\underline{\psi}}) + \sum_{i=1}^L R_i\right) \cdot \exp(Q) \propto$$

$$\propto \exp\left(-\frac{1}{2} (\underline{\psi} - \hat{\underline{\psi}})' H(z) (\underline{\psi} - \hat{\underline{\psi}})\right)$$

puesto que ni Q ni el productorio $\prod_{i=1}^L p(z_i / \hat{\underline{\psi}}^i)$ dependen de $\underline{\psi}$.

Finalmente, sustituyendo el resultado anterior en el producto de densidades $p(\underline{\psi}) \cdot p(z / \underline{\psi})$ resulta:

$$p(\underline{\psi} / z) \propto \exp\left(-\frac{1}{2} (\underline{\psi} - \hat{\underline{\psi}})' H(z) (\underline{\psi} - \hat{\underline{\psi}})\right) \cdot$$

$$\cdot \exp\left(-\frac{1}{2} (\underline{\psi} - \underline{\psi}_0)' T_0 (\underline{\psi} - \underline{\psi}_0)\right) \cdot \exp\left(\sum_{i=1}^L R_i\right) \propto$$

$$\propto \exp\left(-\frac{1}{2} (\underline{\psi} - (H(z) + T_0)^{-1} \cdot (H(z) \hat{\underline{\psi}} + T_0 \underline{\psi}_0))' \cdot \left(\right.\right.$$

$$\left. (T_0 + H(z)) (\underline{\psi} - (H(z) + T_0)^{-1} (H(z) \hat{\underline{\psi}} + T_0 \underline{\psi}_0)) + \sum_{i=1}^L R_i\right) =$$

$$= \exp\left(-\frac{1}{2} (\underline{\psi} - \underline{\psi}^*)' (T_0 + H(z)) (\underline{\psi} - \underline{\psi}^*) + \sum_{i=1}^L R_i\right)$$

y donde

$$\sum_{i=1}^L R_i = \sum_{i=1}^L O(n_i^{-\frac{1}{2}}) = O(\underline{n}^{-\frac{1}{2}})$$

Comparando la última expresión con la definición de una distribución multi normal se sigue el resultado apetecido.

(c.q.d.)

Utilizando el resultado anterior vamos a estimar la integral $I^{\underline{\vartheta}}\{z, p(\underline{\vartheta})\}$ y obtener, de esta forma, una aproximación válida para muestras grandes en todos los estratos.

TEOREMA 6.3.3

Si $\underline{\psi} = \psi(\underline{\vartheta})$ es una transformación biyectiva de la cantidad aleatoria $\underline{\vartheta}$ con densidad aproximadamente normal y con matriz de precisión T_0 , entonces

$$I^{\underline{\vartheta}}\{z, p(\underline{\vartheta})\} = \frac{1}{2} \log \frac{|T_0 + H(z)|}{|T_0|} - \frac{1}{2} \sum_{p=1}^L \sum_{q=1}^L h_{pq}(z).$$

$$\cdot \frac{|\text{Adj}(t_{qp} + h_{qp}(z))|}{|T_0 + H(z)|} + \frac{1}{2} (\underline{\hat{\psi}} - \underline{\psi}_0)' H(z) (T_0 + H(z))^{-1} T_0 ($$

$$(T_0 + H(z))^{-1} H(z) (\underline{\hat{\psi}} - \underline{\psi}_0) + \varepsilon$$

donde la interpretación de las anteriores matrices y vectores es la expresada en el teorema 6.3.2 y ε es una cantidad despreciable, en muestras grandes, frente a $I^{\underline{\vartheta}}\{z, p(\underline{\vartheta})\}$.

Demostración.- Por el teorema anterior y la invarianza de $I^{\underline{\vartheta}}\{z, p(\underline{\vartheta})\}$ ante transformaciones biyectivas de $\underline{\vartheta}$ recogida en el teorema 6.3.1, resulta:

$$I^{\underline{\vartheta}}\{z, p(\underline{\vartheta})\} = I^{\underline{\psi}}\{z, p(\underline{\psi})\} = \int N^L(\underline{\psi}^*, T_0 + H(z)) \cdot \log \frac{N(\underline{\psi}^*, T_0 + H(z))}{N(\underline{\psi}_0, T_0)} \cdot d\underline{\psi} + \varepsilon$$

donde las correspondientes expresiones de las multinormales vienen dadas a través de las matrices de precisión respectivas. La integral anterior es igual a la suma

$$\frac{1}{2} \log \frac{|T_0 + H(z)|}{|T_0|} - \frac{1}{2} \int N^L(\underline{\psi}^*, T_0 + H(z)) \cdot (\underline{\psi} - \underline{\psi}^*)' (T_0 + H(z)) (\underline{\psi} - \underline{\psi}^*) \cdot d\underline{\psi} +$$

$$+ \frac{1}{2} \int N^L(\underline{\psi}^*, T_0 + H(z)) \cdot (\underline{\psi} - \underline{\psi}_0)' T_0 (\underline{\psi} - \underline{\psi}_0) \cdot d\underline{\psi} =$$

$$= \frac{1}{2} \log \frac{|T_0 + H(z)|}{|T_0|} - \frac{1}{2} \sum_{p=1}^L \sum_{q=1}^L (t_{pq} + h_{pq}(z)) \cdot \frac{|\text{Adj}(t_{qp} + h_{qp}(z))|}{|T_0 + H(z)|} +$$

$$\begin{aligned}
& + \frac{1}{2} \int N^L(\underline{\psi}^*, T_0 + H(z)) \cdot (\underline{\psi} - \underline{\psi}^*)' T_0 (\underline{\psi} - \underline{\psi}^*) \cdot d\underline{\psi} + \\
& + \frac{1}{2} \int N^L(\underline{\psi}^*, T_0 + H(z)) \cdot (\underline{\psi} - \underline{\psi}^*)' T_0 (\underline{\psi}^* - \underline{\psi}_0) \cdot d\underline{\psi} + \\
& + \frac{1}{2} \int N^L(\underline{\psi}^*, T_0 + H(z)) \cdot (\underline{\psi}^* - \underline{\psi}_0)' T_0 (\underline{\psi} - \underline{\psi}^*) \cdot d\underline{\psi} + \\
& + \frac{1}{2} \int N^L(\underline{\psi}^*, T_0 + H(z)) \cdot (\underline{\psi}^* - \underline{\psi}_0)' T_0 (\underline{\psi}^* - \underline{\psi}_0) \cdot d\underline{\psi} = \\
& = \frac{1}{2} \log \frac{|T_0 + H(z)|}{|T_0|} - \frac{1}{2} \sum_{p=1}^L \sum_{q=1}^L (t_{pq} + h_{pq}(z)) \cdot \frac{|\text{Adj}(t_{qp} + h_{qp}(z))|}{|T_0 + H(z)|} + \\
& + \frac{1}{2} \sum_{p=1}^L \sum_{q=1}^L t_{pq} \cdot \frac{|\text{Adj}(t_{qp} + h_{qp}(z))|}{|T_0 + H(z)|} + \frac{1}{2} (\underline{\psi}^* - \underline{\psi}_0)' T_0 (\underline{\psi}^* - \underline{\psi}_0) = \\
& = \frac{1}{2} \log \frac{|T_0 + H(z)|}{|T_0|} - \frac{1}{2} \sum_{p=1}^L \sum_{q=1}^L h_{pq}(z) \cdot \frac{|\text{Adj}(t_{qp} + h_{qp}(z))|}{|T_0 + H(z)|} + \\
& + \frac{1}{2} (\underline{\psi}^* - \underline{\psi}_0)' T_0 (\underline{\psi}^* - \underline{\psi}_0)
\end{aligned}$$

El resultado del teorema se sigue sin más que considerar que

$$\underline{\psi}^* - \underline{\psi}_0 = (T_0 + H(z))^{-1} H(z) (\hat{\underline{\psi}} - \underline{\psi}_0)$$

de forma que el último término en la cadena de igualdades anterior resulta ser

$$\frac{1}{2} (\hat{\underline{\psi}} - \underline{\psi}_0)' H(z) (T_0 + H(z))^{-1} T_0 (T_0 + H(z))^{-1} H(z) (\hat{\underline{\psi}} - \underline{\psi}_0) \quad (\text{c.q.d.})$$

A partir del resultado expresado por el teorema anterior puede obtenerse el comportamiento asintótico de la información esperada sobre $\underline{\theta}$ proporcionada por $E^S(\underline{n})$, toda vez que dicha información es el valor esperado sobre el resultado muestral de la información ya aproximada $I^{\underline{\theta}}\{z, \rho(\underline{\theta})\}$. Para ello, demos­tramos primeramente el siguiente

LEMA 6.3.1

Si el vector tamaño muestral \underline{n} es suficientemente grande en todas sus componentes, dado $\underline{\psi}$, la distribución de $\hat{\underline{\psi}}$ es aproximadamente normal multivariante con vector media igual a $\underline{\psi}$ y matriz de precisión

$$I(\underline{\psi}) = \sum_{i=1}^L n_i \cdot I^i(\underline{\psi})$$

donde $I^i(\underline{\psi})$ es la matriz de elementos $c_{pq}^i(\underline{\psi})$ definidos como

$$c_{pq}^i(\underline{\psi}) = - \int p(x_{ij}/F_i(\underline{\psi})) \frac{\partial^2}{\partial \psi_p \partial \psi_q} \log p(x_{ij}/F_i(\underline{\psi})) \cdot dx_{ij}$$

Demostración.- Dado $\underline{\psi}$, las distintas estimaciones máximo-verosímiles $\hat{\underline{\psi}}^i$ son independientes entre sí, toda vez que así lo son las submuestras parciales sobre las que están definidas. Por otra parte, es conocido (Bernardo, 1979*) que, según la ley fuerte de los grandes números y tal como se comentaba en el teorema 6.3.2, para valores grandes de n_i la constante $h_{pq}^i(z_i)$ se comporta, para todo $p, q = 1, 2, \dots, L$, y dado $\underline{\psi}$, como n_i veces la constante

$$c_{pq}^i(\underline{\psi}) = - \int p(x_{ij}/F_i(\underline{\psi})) \frac{\partial^2}{\partial \psi_p \partial \psi_q} \log p(x_{ij}/F_i(\underline{\psi})) \cdot dx_{ij}$$

Sea la matriz $I^i(\underline{\psi})$ definida con los elementos $c_{pq}^i(\underline{\psi})$ ($p, q = 1, 2, \dots, L$). Del comentario anterior deducimos que si n_i es grande, la matriz $H^i(z_i)$ se comportará, dado $\underline{\psi}$, como n_i veces la matriz $I^i(\underline{\psi})$. Así pues, siendo el vector \underline{n} lo suficientemente grande en todas sus componentes resultará que

$$H(z) = \sum_{i=1}^L H^i(z_i) \simeq \sum_{i=1}^L n_i \cdot I^i(\underline{\psi}) = I(\underline{\psi})$$

Podrá, pues, escribirse que

$$\hat{\underline{\psi}} = H(z)^{-1} \sum_{i=1}^L H^i(z_i) \cdot \hat{\underline{\psi}}^i = \sum_{i=1}^L H(z)^{-1} \cdot H^i(z_i) \cdot \hat{\underline{\psi}}^i$$

Además y bajo ciertas débiles condiciones de regularidad es conocido (Cox & Hinkley, 1974) que dado $\underline{\psi}$ la distribución del estimador máximo verosímil $\hat{\underline{\psi}}^i$ es, en muestras grandes, aproximadamente multinormal, con vector media $\underline{\psi}$ y matriz de precisión $n_i \cdot I^i(\underline{\psi})$. Así pues, si n_i es grande

$$p(\hat{\underline{\psi}}^i / \underline{\psi}) = N^L(\underline{\psi}, n_i \cdot I^i(\underline{\psi})) \text{ como matriz de precisión}$$

De ahí que siendo \underline{n} lo suficientemente grande en todas sus componentes

$$H(z)^{-1} \cdot H^i(z_i) \cdot \hat{\psi}^i \simeq I(\underline{\psi})^{-1} \cdot n_i \cdot I^i(\underline{\psi}) \cdot \hat{\psi}^i$$

Y así resulta que dado $\underline{\psi}$

$$\begin{aligned} \rho(I(\underline{\psi})^{-1} n_i I^i(\underline{\psi}) \hat{\psi}^i / \underline{\psi}) &\simeq N^L(I(\underline{\psi})^{-1} n_i I^i(\underline{\psi}) \cdot \underline{\psi}, \text{ matriz de momentos} = \\ &= I(\underline{\psi})^{-1} n_i I^i(\underline{\psi}) I(\underline{\psi})^{-1}) \end{aligned}$$

Dado que $\hat{\psi}$ es, dado $\underline{\psi}$, una suma de multinormales independientes resulta (Graybill, 1961) finalmente que siendo \underline{n} lo suficientemente grande como para asegurar que lo sean todas sus componentes

$$\begin{aligned} \rho(\hat{\psi} / \underline{\psi}) &\simeq N^L\left(\sum_{i=1}^L I(\underline{\psi})^{-1} n_i I^i(\underline{\psi}) \cdot \underline{\psi}, \sum_{i=1}^L I(\underline{\psi})^{-1} n_i I^i(\underline{\psi}) I(\underline{\psi})^{-1}\right) = \\ &= N^L(\underline{\psi}, I(\underline{\psi})^{-1}) = \text{matriz de momentos} \\ &\hspace{15em} (\text{c.q.d.}) \end{aligned}$$

Estamos ya en condiciones de poder encontrar una expresión que se aproxime asintóticamente a $I^{\underline{\vartheta}}\{E^S(\underline{n}), \rho(\underline{\vartheta})\}$.

TEOREMA 6.3.4

Siendo $\underline{\psi} = \psi(\underline{\vartheta})$ una transformación biyectiva de la cantidad de interés $\underline{\vartheta}$ con densidad aproximadamente multinormal de matriz de precisión T_0 , entonces

$$I^{\underline{\vartheta}}\{E^S(\underline{n}), \rho(\underline{\vartheta})\} = \frac{1}{2} \log \frac{|T_0 + R|}{|T_0|} + \varepsilon$$

donde R es la matriz esperanza de la $I(\underline{\psi})$ definida en el teorema 6.3.1 y Lema 6.3.1

$$R = \int \rho(\underline{\psi}) \cdot I(\underline{\psi}) \cdot d\underline{\psi}$$

y ε es una cantidad que, en muestras grandes, es despreciable frente al logaritmo del cociente $|T_0 + R| / |T_0|$.

Demostración.— Sabemos por el lema 6.3.1 que si el vector tamaño muestral \underline{n} es lo suficientemente grande para que todas sus componentes lo sean la matriz $H(z)$ se comporta, dado $\underline{\psi}$, como la matriz

$$I(\underline{\psi}) = \sum_{i=1}^L n_i I^i(\underline{\psi})$$

De ahí que siendo \underline{n} grande, el valor esperado de cualquier función de $H(z)$ sea, aproximadamente, igual al valor que toma la función en el valor esperado de $H(z)$, esto es, en

$$\int p(z) \cdot H(z) \cdot dz = \int p(\underline{\psi}) \cdot I(\underline{\psi}) \cdot d\underline{\psi} = R$$

verificándose las igualdades

$$\begin{aligned} \int p(z) \cdot H(z) \cdot dz &= \int \sum_{i=1}^L H^i(z_i) \cdot p(z) \cdot dz = \sum_{i=1}^L \int H^i(z_i) \cdot p(z_i) \cdot dz_i = \\ &= \sum_{i=1}^L n_i \int p(\underline{\psi}) \cdot I^i(\underline{\psi}) \cdot d\underline{\psi} = \int \left\{ \sum_{i=1}^L n_i I^i(\underline{\psi}) \right\} p(\underline{\psi}) \cdot d\underline{\psi} = \\ &= \int I(\underline{\psi}) \cdot p(\underline{\psi}) \cdot d\underline{\psi} \end{aligned}$$

De ahí que

$$\begin{aligned} I^{\theta} \{ E^s(\underline{n}), p(\underline{\theta}) \} &= \int p(z) \cdot I^{\theta} \{ z, p(\underline{\theta}) \} \cdot dz = \int p(z) \cdot I^{\psi} \{ z, p(\underline{\psi}) \} \cdot dz = \\ &= I^{\psi} \{ E^s(\underline{n}), p(\underline{\psi}) \} \end{aligned}$$

según el resultado del teorema 6.3.1. Por el teorema 6.3.3 puede escribirse que

$$\begin{aligned} I^{\theta} \{ E^s(\underline{n}), p(\underline{\theta}) \} &\simeq \frac{1}{2} \log \frac{|T_0 + R|}{|T_0|} - \frac{1}{2} \sum_{p=1}^L \sum_{q=1}^L r_{pq} \cdot \frac{|\text{Adj}(t_{qp} + r_{qp})|}{|T_0 + R|} + \\ &+ \frac{1}{2} \int p(z) \cdot (\hat{\underline{\psi}} - \underline{\psi}_0)' R (T_0 + R)^{-1} T_0 (T_0 + R)^{-1} R (\hat{\underline{\psi}} - \underline{\psi}_0) \cdot dz \end{aligned}$$

donde r_{pq} es el elemento (p,q) de la matriz R , esto es,

$$r_{pq} = \sum_{i=1}^L n_i \int c_{pq}^i(\underline{\psi}) \cdot p(\underline{\psi}) \cdot d\underline{\psi}$$

Consideremos la penúltima integral:

$$\begin{aligned} \int p(z) \cdot (\hat{\underline{\psi}} - \underline{\psi}_0)' R (T_0 + R)^{-1} T_0 (T_0 + R)^{-1} R (\hat{\underline{\psi}} - \underline{\psi}_0) \cdot dz = \\ = \int p(\hat{\underline{\psi}}) \cdot (\hat{\underline{\psi}} - \underline{\psi}_0)' R (T_0 + R)^{-1} T_0 (T_0 + R)^{-1} R (\hat{\underline{\psi}} - \underline{\psi}_0) \cdot d\hat{\underline{\psi}} = \end{aligned}$$

$$\begin{aligned}
&= \int p(\underline{\psi}) \int p(\underline{\hat{\psi}} / \underline{\psi}) \cdot (\underline{\hat{\psi}} - \underline{\psi}_0)' R(T_0 + R)^{-1} T_0 (T_0 + R)^{-1} R(\underline{\hat{\psi}} - \underline{\psi}_0) \cdot d\underline{\hat{\psi}} \cdot d\underline{\psi} = \\
&= \int p(\underline{\psi}) \int p(\underline{\hat{\psi}} / \underline{\psi}) \cdot (\underline{\hat{\psi}} - \underline{\psi})' R(T_0 + R)^{-1} T_0 (T_0 + R)^{-1} R(\underline{\hat{\psi}} - \underline{\psi}) \cdot d\underline{\hat{\psi}} \cdot d\underline{\psi} + \quad (1) \\
&+ \int p(\underline{\psi}) \int p(\underline{\hat{\psi}} / \underline{\psi}) \cdot (\underline{\hat{\psi}} - \underline{\psi})' R(T_0 + R)^{-1} T_0 (T_0 + R)^{-1} R(\underline{\psi} - \underline{\psi}_0) \cdot d\underline{\hat{\psi}} \cdot d\underline{\psi} + \\
&+ \int p(\underline{\psi}) \int p(\underline{\hat{\psi}} / \underline{\psi}) \cdot (\underline{\psi} - \underline{\psi}_0)' R(T_0 + R)^{-1} T_0 (T_0 + R)^{-1} R(\underline{\hat{\psi}} - \underline{\psi}) \cdot d\underline{\hat{\psi}} \cdot d\underline{\psi} + \\
&+ \int p(\underline{\psi}) \int p(\underline{\hat{\psi}} / \underline{\psi}) \cdot (\underline{\psi} - \underline{\psi}_0)' R(T_0 + R)^{-1} T_0 (T_0 + R)^{-1} R(\underline{\psi} - \underline{\psi}_0) \cdot d\underline{\hat{\psi}} \cdot d\underline{\psi} \quad (2) = \\
&= (1) + (2) .
\end{aligned}$$

Por el lema 6.3.1 es conocido que

$$p(\underline{\hat{\psi}} / \underline{\psi}) \simeq N^L(\underline{\psi}, \text{matriz de precisión } I(\underline{\psi}))$$

Así pues, la nueva variable $\underline{Y} = (T_0 + R)^{-1} R(\underline{\hat{\psi}} - \underline{\psi})$ se distribuirá, dado $\underline{\psi}$, según la multinormal

$$p(\underline{Y} / \underline{\psi}) \simeq N^L(\underline{0}, \text{matriz de momentos } (T_0 + R)^{-1} R I(\underline{\psi})^{-1} R(T_0 + R)^{-1})$$

De ahí que

$$\begin{aligned}
(1) &= \int p(\underline{\psi}) \int p(\underline{Y} / \underline{\psi}) \cdot \underline{Y}' T_0 \underline{Y} \cdot d\underline{Y} \cdot d\underline{\psi} = \\
&= \int p(\underline{\psi}) \cdot \sum_{p=1}^L \sum_{q=1}^L t_{pq} a_{pq}(\underline{\psi}) \cdot d\underline{\psi} = \sum_{p=1}^L \sum_{q=1}^L t_{pq} b_{pq}
\end{aligned}$$

donde a_{pq} es el elemento (p,q) de la matriz $(T_0 + R)^{-1} R I(\underline{\psi})^{-1} R(T_0 + R)^{-1}$, t_{pq} el correspondiente de la matriz T_0 y b_{pq} , el elemento (p,q) de la matriz $(T_0 + R)^{-1} R(T_0 + R)^{-1}$, toda vez que

$$\int p(\underline{\psi}) \cdot I(\underline{\psi})^{-1} \cdot d\underline{\psi} = R^{-1}$$

Por otra parte, sea la nueva variable $\underline{X} = (T_0 + R)^{-1} R(\underline{\psi} - \underline{\psi}_0)$.

Dado que la distribución que describe las opiniones iniciales que el investigador posee sobre $\underline{\psi}$ es, por hipótesis, aproximadamente multinormal de vector media $\underline{\psi}_0$ y matriz de precisión T_0 , resultará que

$$p(\underline{X}) \simeq N^L(\underline{0}, \text{matriz de momentos } (T_0 + R)^{-1} R T_0^{-1} R(T_0 + R)^{-1})$$

De ahí que

$$(2) = \int p(\underline{\psi}) \cdot (\underline{\psi} - \underline{\psi}_0)' R(T_0 + R)^{-1} T_0 (T_0 + R)^{-1} R(\underline{\psi} - \underline{\psi}_0) \cdot d\underline{\psi} =$$

$$= \int p(\underline{x}) \cdot \underline{x}^T T_0 \underline{x} \cdot d\underline{x} = \sum_{p=1}^L \sum_{q=1}^L t_{pq} d_{pq}$$

donde d_{pq} es el elemento (p,q) de la matriz $(T_0 + R)^{-1} R (T_0 + R)^{-1}$.

Dado que las matrices T_0 y R son simétricas y reuniendo todos los resultados parciales se obtiene para muestras grandes que

$$I^{\vartheta} \{ E^S(\underline{n}), p(\underline{\vartheta}) \} \simeq \frac{1}{2} \log \frac{|T_0 + R|}{|T_0|} - \frac{1}{2} \sum_{p=1}^L \sum_{q=1}^L r_{pq} \cdot \frac{|\text{Adj}(t_{pq} + r_{pq})|}{|T_0 + R|} +$$

$$+ \sum_{p=1}^L \sum_{q=1}^L t_{pq} (b_{pq} + d_{pq})$$

Y al ser

$$\sum_{p=1}^L \sum_{q=1}^L r_{pq} \cdot \frac{|\text{Adj}(t_{pq} + r_{pq})|}{|T_0 + R|} = \frac{L}{2} - \sum_{p=1}^L \sum_{q=1}^L t_{pq} \cdot \frac{|\text{Adj}(t_{pq} + r_{pq})|}{|T_0 + R|}$$

Denotando por g_{pq} el elemento (p,q) de la matriz $(T_0 + R)^{-1}$ resulta:

$$I^{\vartheta} \{ E^S(\underline{n}), p(\underline{\vartheta}) \} \simeq \frac{1}{2} \log \frac{|T_0 + R|}{|T_0|} - \frac{L}{2} + \frac{1}{2} \sum_{p=1}^L \sum_{q=1}^L t_{pq} (b_{pq} + d_{pq} + g_{pq})$$

donde $b_{pq} + d_{pq} + g_{pq}$ es el elemento (p,q) de la matriz

$$\begin{aligned} & (T_0 + R)^{-1} R (T_0 + R)^{-1} + (T_0 + R)^{-1} R T_0^{-1} R (T_0 + R)^{-1} + (T_0 + R)^{-1} \\ & = (T_0 + R)^{-1} \{ R + R T_0^{-1} R + T_0 + R \} (T_0 + R)^{-1} = \\ & = (T_0 + R)^{-1} \{ (T_0 + R) T_0^{-1} (T_0 + R) \} (T_0 + R)^{-1} = T_0^{-1} \end{aligned}$$

Por tanto

$$b_{pq} + d_{pq} + g_{pq} = \frac{|\text{Adj}(t_{qp})|}{|T_0|} = \frac{|\text{Adj}(t_{pq})|}{|T_0|}$$

Finalmente, pues,

$$I^{\vartheta} \{ E^S(\underline{n}), p(\underline{\vartheta}) \} \simeq \frac{1}{2} \log \frac{|T_0 + R|}{|T_0|} - \frac{L}{2} + \frac{1}{2} \sum_{p=1}^L \sum_{q=1}^L t_{pq} \cdot \frac{|\text{Adj}(t_{pq})|}{|T_0|} =$$

$$= \frac{1}{2} \log \frac{|T_0 + R|}{|T_0|}$$

(c.q.d.)

Tal como era de esperar la coherencia del resultado anterior queda comprobada al aplicar la anterior aproximación al modelo normal y ver la coincidencia de los resultados obtenidos.

COROLARIO 6.3.1 .

En el modelo normal, la información sobre $\underline{\theta}$ proporcionada según el resultado del teorema anterior coincide con la información exacta.

Demostración.- Siendo $p(\underline{\theta})$ ya multinormal consideraremos la transformación identidad

$$\psi_i = \theta_i \quad i=1,2,\dots,L$$

Para ella,

$$c_{pq}^i(\underline{\psi}) = c_{pq}^i(\underline{\theta}_i) = 0 \quad \text{para todo } p,q \text{ tales que } p \neq i \text{ ó } q \neq i$$

$$c_{ii}^i(\underline{\theta}_i) = - \int \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x - \theta_i}{\sigma_i} \right)^2} \frac{\partial^2}{\partial \theta_i^2} \left(-\frac{1}{2} \left(\frac{x - \theta_i}{\sigma_i} \right)^2 \right) \cdot dx = \frac{1}{\sigma_i^2}$$

De ahí que

$$I(\underline{\psi}) = I(\underline{\theta}) = \begin{pmatrix} n_1/\sigma_1^2 & 0 & \dots & 0 \\ 0 & n_2/\sigma_2^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & n_L/\sigma_L^2 \end{pmatrix} =$$

$$= \begin{pmatrix} n_1 & 0 & \dots & 0 \\ 0 & n_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & n_L \end{pmatrix} \begin{pmatrix} 1/\sigma_1^2 & 0 & \dots & 0 \\ 0 & 1/\sigma_2^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1/\sigma_L^2 \end{pmatrix} = NH^{-1}$$

y que

$$R = \int p(\underline{\psi}) \cdot I(\underline{\psi}) \cdot d\underline{\psi} = NH^{-1}$$

Finalmente,

$$I_{\text{aproximada}}^{\underline{\theta}} \left\{ E^S(\underline{n}), N^L(\underline{\theta}_0, T_0^{-1} = H_0) \right\} = \frac{1}{2} \log \frac{|H_0^{-1} + R|}{|H_0^{-1}|} =$$

$$= \frac{1}{2} \log \frac{|H_0^{-1} + NH^{-1}|}{|H_0^{-1}|} = \frac{1}{2} \log \frac{|H + H_0 N|}{|H|} = I^{\tilde{\theta}} \left\{ E^S(\tilde{n}), N^L(\tilde{\theta}_0, H_0) \right\}$$

(c.q.d.)

6.4 EJEMPLO DE CALCULO APROXIMADO: MODELO MULTINOMIAL

Vamos a aplicar la aproximación desarrollada en la anterior sección al diseño de una investigación que pretende determinar la distribución en porcentaje y entre los diferentes estratos de una cierta característica cualitativa.

Supongamos, pues, que el espacio muestral total es X y en el que existe un determinado conjunto de unidades que satisfacen cierta característica cualitativa. La población total está dividida en $L+1$ subpoblaciones o estratos, de forma que el investigador está interesado en determinar qué proporción θ_i de esas unidades que satisfacen la característica cualitativa corresponden al estrato i -ésimo, X_i , y esto para todos los estratos. En el ejemplo ya comentado de una investigación tendente a determinar el nivel regional de paro, la situación ahora planteada correspondería a aquella en la que se pretende conocer su distribución porcentual en las distintas regiones, si bien no se está interesado en el conocimiento de los valores absolutos de los mismos, tal vez porque ya sea conocido a nivel de toda la población en su conjunto.

Evidentemente que la existencia de los $L+1$ estratos X_i que constituyen una partición del espacio total X y de las correspondientes proporciones θ_i asociadas a cada uno de ellos implica las dos igualdades

$$X = \bigcup_{i=1}^{L+1} X_i \quad \text{y} \quad \sum_{i=1}^{L+1} \theta_i = 1$$

Se deduce, pues, que conocidas L cantidades de interés θ_i correspondientes a sendos estratos, la $(L+1)$ -ésima viene inmediatamente determinada por el sumatorio anterior. De ahí que el investigador pueda ceñir su trabajo a tan sólo L estratos y sobre ellos diseñar un experimento de la forma $E^S(\tilde{n})$, toda vez que la estimación de las L primeras cantidades

ϑ_i implica la estimación de la restante. Supondremos, pues, una investigación centrada sobre el espacio muestral $X = X_{L+1}$, espacio estratificado en L subpoblaciones

$$X = X_{L+1} = \bigcup_{i=1}^L X_i$$

así como que en todos los estratos X_i existen unidades satisfaciendo la característica cualitativa objeto de estudio.

La consideración de tan sólo L estratos introduce la siguiente modificación sobre los parámetros $\vartheta_i \in (0,1)$, en cuanto que permite considerar los L primeros parámetros ϑ_i tales que

$$\sum_{i=1}^L \vartheta_i = \vartheta$$

siendo $\vartheta = 1 - \vartheta_{L+1}$, y estando definido el vector paramétrico $\vartheta' = (\vartheta_1, \vartheta_2, \dots, \vartheta_L)$ en el cuadrado L -dimensional abierto

$$\vartheta \in (0,1)^L$$

y con la condición adicional de que la suma de sus componentes sea estrictamente menor que la unidad, esto es,

$$\sum_{i=1}^L \vartheta_i = \vartheta < 1$$

Notemos, por otra parte, que el parámetro ϑ suma de los L primeros parámetros ϑ_i admite idéntica interpretación que la realizada para cada ϑ_i en su estrato respectivo, esto es, como el tanto por ciento de individuos de la población $X = X_{L+1}$ que verifican la característica cualitativa cuya distribución en los estratos se pretende determinar.

Así pues, tal como ha quedado planteado, el problema resulta encuadrado dentro de la teoría general desarrollada en el capítulo quinto, con la existencia de un experimento $E^S(\underline{n})$ diseñado sobre una población con L estratos y realizado a fin de recabar información sobre el vector de L componentes $\vartheta' = (\vartheta_1, \vartheta_2, \dots, \vartheta_L)$. Por tanto será necesario especificar dos familias de densidades de probabilidad a fin de describir, primeramente, las densidades $p(x_{ij} / \vartheta_i)$ en cada uno de los estratos y, por otra, una densidad inicial que interprete las opiniones iniciales que el investigador posee sobre el vector ϑ .

En todo proceso de muestreo de proporciones (o totales) de in-

dividuos que pueden verificar cierta característica cualitativa el modelo a aplicar se realiza a través de la distribución de Bernouilli

$$p(x_{ij}/\vartheta_i) = \vartheta_i^{x_{ij}} (1 - \vartheta_i)^{1-x_{ij}} \quad x_{ij} \in \{0,1\}$$

donde x_{ij} toma los valores 1 ó 0 (según que la cualidad objeto de estudio sea satisfecha o no, respectivamente, por el individuo particular x_{ij} obtenido en el proceso muestral) y ϑ_i es, precisamente, el porcentaje de individuos en X_i que satisfacen la característica cualitativa.

La densidad inicial $p(\underline{\vartheta})$ que describe las opiniones iniciales que el investigador posee sobre $\underline{\vartheta}$ puede venir expresada por la correspondiente a una distribución de Dirichlet, con vector paramétrico $\underline{\alpha}' = (\alpha_1, \alpha_2, \dots, \alpha_L, \alpha_{L+1})$ ($\alpha_i > 0$; $i = 1, 2, \dots, L+1$). Así pues,

$$p(\underline{\vartheta}) = p(\vartheta_1, \vartheta_2, \dots, \vartheta_L) = \frac{\Gamma(\alpha_0)}{\prod_{i=1}^{L+1} \Gamma(\alpha_i)} \vartheta_1^{\alpha_1-1} \cdot \vartheta_2^{\alpha_2-1} \cdot \dots \cdot \vartheta_L^{\alpha_L-1} \cdot (1 - \sum_{i=1}^L \vartheta_i)^{\alpha_{L+1}-1}$$

siendo

$$\alpha_0 = \sum_{i=1}^{L+1} \alpha_i, \quad \underline{\vartheta} \in (0,1)^L, \quad \sum_{i=1}^L \vartheta_i < 1$$

y donde la densidad $p(\underline{\vartheta})$ ya se ha expresado en función exclusivamente de las L primeras cantidades aleatorias ϑ_i y no a través de la última ϑ_{L+1} , lo que traería consigo la consideración de una distribución degenerada con toda la masa de probabilidad en el hiperplano

$$\sum_{i=1}^{L+1} \vartheta_i = 1$$

Es evidente que la distribución de Dirichlet se ajusta plenamente a la descripción del vector $\underline{\vartheta}$ por sus especiales condiciones y recinto de definición. Además verifica dos propiedades que la hacen de especial aplicabilidad para nuestro problema. En primer lugar, la distribución de Dirichlet es, en cierto sentido, una distribución beta multivariante. De hecho, si una cierta variable aleatoria x tiene una distribución beta, el vector aleatorio $y = (x, 1-x)'$ tiene una distribución de Dirichlet.

Inversamente, si el vector aleatorio $\underline{\vartheta} = (\vartheta_1, \vartheta_2, \dots, \vartheta_L)'$ tiene una distribución de Dirichlet con vector paramétrico $\underline{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_L)'$, entonces la distribución marginal de cualquier componente de $\underline{\vartheta}$, por ejemplo ϑ_j , es una distribución beta con parámetros α_j y $\alpha_0 - \alpha_j$. Este último resultado implica que los modelos para cada uno de los estratos X_i ($i=1,2,\dots,L$) por separado responden al modelo binomial estudiado con exhaustividad por Bernardo (1975) y por Basulto & Bernardo (1978). Así pues, siendo $E(n_i)$ el experimento componente i -ésimo del $E^S(\underline{n})$ definido sobre el espacio estratificado $X = X_{L+1}$, y siendo $p(\underline{\vartheta})$ una distribución de Dirichlet, puede escribirse que

$$I^{\vartheta_i} \{E(n_i), m_i(p(\underline{\vartheta}))\} = H \{Be(\alpha_i, \alpha_0 - \alpha_i)\} - \sum_{r=0}^n H \{Be(\alpha_i + r, \alpha_0 - \alpha_i + n - r), Bb_r(r/\alpha_i, \alpha_0 - \alpha_i, n)\}$$

donde $m_i(p(\underline{\vartheta}))$ es la marginal i -ésima de $p(\underline{\vartheta})$ y Bb_r una densidad predictiva Beta-binomial, y H la entropía de las densidades beta (Lindley, 1957). En segundo lugar, la distribución de Dirichlet verifica una segunda propiedad que la hace de interés en nuestro caso y que está recogida en DeGroot (1970), pág. 64. En efecto, el vector $(\vartheta_1/\vartheta, \vartheta_2/\vartheta, \dots, \vartheta_L/\vartheta)'$ es independiente de la suma ϑ y, por tanto, del parámetro ϑ_{L+1} estimado inmediatamente por la relación que lo liga con las componentes de $\underline{\vartheta}$.

Si bien son conocidas las informaciones esperadas

$$I^{\vartheta_i} \{E(n_i), Be(\alpha_i, \alpha_0 - \alpha_i)\} \quad i=1,2,\dots,L$$

donde Be representa a una densidad de probabilidad beta, es conocido por la sección 5.5 que a partir de ellas no puede deducirse la información esperada

$$I^{\vartheta} \{E^S(\underline{n}), D^L(\underline{\alpha})\}$$

(donde $D^L(\underline{\alpha})$ representa a una densidad de probabilidad de Dirichlet, de vector paramétrico $\underline{\alpha}$). De su misma definición

$$I^{\vartheta} \{E^S(\underline{n}), D^L(\underline{\alpha})\} = \iint p(z/\underline{\vartheta}) \cdot p(\underline{\vartheta}) \cdot \log \frac{p(\underline{\vartheta}/z)}{p(\underline{\vartheta})} \cdot d\underline{\vartheta} \cdot dz$$

donde la densidad posterior es

$$p(\underline{\theta}/z) \propto (1 - \sum_{i=1}^L \theta_i)^{\alpha_{L+1}-1} \cdot \frac{\Gamma(\alpha_0)}{\prod_{i=1}^{L+1} \Gamma(\alpha_i)} \cdot \prod_{i=1}^L \theta_i^{\sum_{j=1}^{n_i} x_{ij} + \alpha_i - 1} \cdot (1 - \theta_i)^{n_i - \sum_{j=1}^{n_i} x_{ij}}$$

Observando $p(\underline{\theta}/z)$ se comprende que la obtención exacta de $I^{\underline{\theta}}\{E^S(\underline{n}), D^L(\underline{\alpha})\}$ resulta extraordinariamente compleja. De ahí que a fin de poder diseñar adecuadamente el experimento $E^S(\underline{n})$ y así determinar, por ejemplo, el vector tamaño muestral óptimo, sea conveniente acudir a la información aproximada deducida del conjunto de teoremas de la sección 6.3. En este sentido notemos que bajo nuestras hipótesis no existe familia conjugada de densidades a posteriori.

Para poder aplicar los teoremas de la sección 6.3 consideremos la siguiente transformación

$$\psi_i = \log \frac{\theta_i}{1 - \sum_{j=1}^L \theta_j} = \log \frac{\theta_i}{\theta_{L+1}} \quad i=1,2,\dots,L$$

que efectivamente es biyectiva puesto que

$$\frac{\partial \psi_i}{\partial \theta_i} = \frac{1 - \sum_{j \neq i}^L \theta_j}{\theta_i (1 - \sum_{j=1}^L \theta_j)}, \quad \frac{\partial \psi_i}{\partial \theta_j} = \frac{1}{1 - \sum_{j=1}^L \theta_j}$$

verificándose que el jacobiano

$$J = \left| \frac{\partial \underline{\psi}}{\partial \underline{\theta}} \right| = \frac{1}{(1 - \sum_{j=1}^L \theta_j)^L} \begin{vmatrix} \frac{1 - \sum_{j \neq 1}^L \theta_j}{\theta_1} & 1 & \dots & 1 \\ 1 & \frac{1 - \sum_{j \neq 2}^L \theta_j}{\theta_2} & \dots & 1 \\ \dots & \dots & \dots & \dots \end{vmatrix} =$$

(sigue)

$$p(\underline{\psi}) = \frac{\Gamma(\alpha_0)}{\prod_{i=1}^{L+1} \Gamma(\alpha_i)} \cdot \prod_{i=1}^L \frac{e^{\psi_i(\alpha_i-1)}}{(1 + \sum_{j=1}^L e^{\psi_j})^{\alpha_i-1}} \cdot \frac{1}{(1 + \sum_{j=1}^L e^{\psi_j})^{\alpha_L-1}}$$

$$\cdot \frac{\prod_{i=1}^L e^{\psi_i}}{(1 + \sum_{j=1}^L e^{\psi_j})^L} = \frac{\Gamma(\alpha_0)}{\prod_{i=1}^{L+1} \Gamma(\alpha_i)} \cdot \frac{1}{(1 + \sum_{j=1}^L e^{\psi_j})^{\alpha_0}} \cdot \prod_{i=1}^L e^{\psi_i \alpha_i}$$

distribución que tiene como función característica en su dominio de definición

$$\psi(t_1, \dots, t_L) = \frac{\Gamma(\alpha_{L+1} - i \sum_{j=1}^L t_j)}{\Gamma(\alpha_{L+1})} \cdot \prod_{j=1}^L \frac{\Gamma(\alpha_j + it_j)}{\Gamma(\alpha_j)}$$

La densidad $p(\underline{\psi})$ anterior puede aproximarse por una multinomial L -dimensional de vector media $\underline{\psi}_0$ y matriz de precisión T_0 , donde

$$\underline{\psi}_0 = \begin{pmatrix} \log(\alpha_1/\alpha_{L+1}) \\ \log(\alpha_2/\alpha_{L+1}) \\ \vdots \\ \log(\alpha_L/\alpha_{L+1}) \end{pmatrix}, \text{ y}$$

$$T_0 = \begin{pmatrix} \alpha_1 - \frac{\alpha_1^2}{\alpha_0} & -\frac{\alpha_1 \alpha_2}{\alpha_0} & \dots & -\frac{\alpha_1 \alpha_L}{\alpha_0} \\ -\frac{\alpha_1 \alpha_2}{\alpha_0} & \alpha_2 - \frac{\alpha_2^2}{\alpha_0} & \dots & -\frac{\alpha_2 \alpha_L}{\alpha_0} \\ \dots & \dots & \dots & \dots \\ -\frac{\alpha_1 \alpha_L}{\alpha_0} & -\frac{\alpha_2 \alpha_L}{\alpha_0} & \dots & \alpha_L - \frac{\alpha_L^2}{\alpha_0} \end{pmatrix}$$

En efecto,

$$p(\underline{\psi}) = \frac{\Gamma(\alpha_0)}{\prod_{i=1}^{L+1} \Gamma(\alpha_i)} \cdot \exp\left(\sum_{i=1}^L \psi_i \alpha_i - \alpha_0 \cdot \log\left(1 + \sum_{j=1}^L e^{\psi_j}\right)\right) \propto \exp(Q)$$

Desarrollando el exponente en serie de Taylor alrededor de su máximo, deducido de las condiciones

$$\frac{\partial Q}{\partial \psi_i} = \alpha_i - \alpha_0 \cdot \frac{e^{\psi_i}}{1 + \sum_{j=1}^L e^{\psi_j}} = 0 \quad i = 1, 2, \dots, L$$

y que resulta ser el vector $\underline{\psi}_0$ que acabamos arriba de indicar, se obtiene que

$$p(\underline{\psi}) \approx \frac{\Gamma(\alpha_0)}{\prod_{i=1}^{L+1} \Gamma(\alpha_i)} \cdot \exp\left(\sum_{i=1}^L \alpha_i \log \frac{\alpha_i}{\alpha_{L+1}} - \alpha_0 \log \frac{\alpha_0}{\alpha_{L+1}} - \frac{1}{2} (\underline{\psi} - \underline{\psi}_0)' T_0 (\underline{\psi} - \underline{\psi}_0)\right)$$

$$\propto \exp\left(-\frac{1}{2} (\underline{\psi} - \underline{\psi}_0)' T_0 (\underline{\psi} - \underline{\psi}_0)\right) = N^L(\underline{\psi}_0, \text{matriz de precisión } T_0)$$

y donde los elementos de T_0 se obtienen de las igualdades

$$\frac{\partial^2 Q}{\partial \psi_i^2} = -\alpha_0 \cdot \frac{e^{\psi_i} (1 + \sum_{j \neq i}^L e^{\psi_j})}{(1 + \sum_{j=1}^L e^{\psi_j})^2} \quad \left| \quad \underline{\psi}'_0 = \left(\log \frac{\alpha_1}{\alpha_{L+1}}, \dots, \log \frac{\alpha_L}{\alpha_{L+1}}\right) \right. , y$$

$$\frac{\partial^2 Q}{\partial \psi_i \partial \psi_j} = \alpha_0 \cdot \frac{e^{\psi_i} e^{\psi_j}}{(1 + \sum_{j=1}^L e^{\psi_j})^2} \quad \left| \quad \underline{\psi}'_0 = \left(\log \frac{\alpha_1}{\alpha_{L+1}}, \dots, \log \frac{\alpha_L}{\alpha_{L+1}}\right) \right.$$

Notemos que

$$|T_0| = \frac{1}{\alpha_0} \cdot \prod_{i=1}^{L+1} \alpha_i$$

Con los elementos anteriores pasemos a aproximar la información exacta $I^S\{E^S(\underline{n}), D^L(\underline{\alpha})\}$ mediante el resultado del teorema 6.3.4.

Para ello, y para un cierto i tal que $1 \leq i \leq L$, resulta que

$$p(x_{ij}/F_i(\psi)) = \frac{e^{\psi_i x_{ij}} (1 + \sum_{j \neq i}^L e^{\psi_j})^{1-x_{ij}}}{1 + \sum_{j=1}^L e^{\psi_j}}$$

De donde las siguientes derivadas del logaritmo de la densidad anterior resultan

$$\frac{\partial^2 \log p(x_{ij}/F_i(\psi))}{\partial \psi_p \partial \psi_q} = -(1 - x_{ij}) \cdot \frac{e^{\psi_p} e^{\psi_q}}{(1 + \sum_{j \neq i}^L e^{\psi_j})^2} +$$

$$+ \frac{e^{\psi_p} e^{\psi_q}}{(1 + \sum_{j=1}^L e^{\psi_j})^2} \quad \text{si } p, q \neq i \text{ y } p \neq q$$

$$\frac{\partial^2 \log p(x_{ij}/F_i(\psi))}{\partial \psi_p^2} = (1 - x_{ij}) \cdot \frac{e^{\psi_p} (1 + \sum_{j \neq p, i}^L e^{\psi_j})}{(1 + \sum_{j \neq i}^L e^{\psi_j})^2} -$$

$$- \frac{e^{\psi_p} (1 + \sum_{j=1}^L e^{\psi_j})}{(1 + \sum_{j=1}^L e^{\psi_j})^2} \quad \text{si } p \neq i$$

$$\frac{\partial^2 \log p(x_{ij}/F_i(\psi))}{\partial \psi_p \partial \psi_i} = \frac{e^{\psi_i} e^{\psi_p}}{(1 + \sum_{j=1}^L e^{\psi_j})^2} \quad \text{si } p \neq i$$

$$\frac{\partial^2 \log p(x_{ij}/F_i(\psi))}{\partial \psi_i^2} = - \frac{e^{\psi_i} (1 + \sum_{j \neq i}^L e^{\psi_j})}{(1 + \sum_{j=1}^L e^{\psi_j})^2}$$

De la definición de $c_{pq}^i(\psi)$, con $p, q = 1, 2, \dots, L$, resulta:

$$c_{pq}^i(\psi) = - \sum_{x_{ij}=0,1} p(x_{ij}/F_i(\psi)) \cdot \frac{\partial^2}{\partial \psi_p \partial \psi_q} \log p(x_{ij}/F_i(\psi))$$

de forma que la matriz $I^i(\underline{\psi})$ será la de los elementos $c_{pq}^i(\underline{\psi})$ siguientes:

$$c_{ii}^i(\underline{\psi}) = \frac{e^{\psi_i} (1 + \sum_{j \neq i}^L e^{\psi_j})}{(1 + \sum_{j=1}^L e^{\psi_j})^2}$$

$$c_{pi}^i(\underline{\psi}) = c_{ip}^i(\underline{\psi}) = - \frac{e^{\psi_i} e^{\psi_p}}{(1 + \sum_{j=1}^L e^{\psi_j})^2} \quad \text{con } p \neq i$$

$$c_{pp}^i(\underline{\psi}) = \frac{e^{\psi_i} 2 e^{\psi_p}}{(1 + \sum_{j=1}^L e^{\psi_j}) \cdot (1 + \sum_{j \neq i}^L e^{\psi_j})} \quad \text{con } p \neq i$$

$$c_{pq}^i(\underline{\psi}) = c_{qp}^i(\underline{\psi}) = \frac{e^{\psi_i} e^{\psi_p} e^{\psi_q}}{(1 + \sum_{j \neq i}^L e^{\psi_j}) \cdot (1 + \sum_{j=1}^L e^{\psi_j})^2} \quad \text{con } p \neq q, p, q \neq i$$

Denotando por R^i a la matriz esperanza de la $I^i(\underline{\psi})$ es evidente que

$$R = \sum_{i=1}^L n_i \cdot R^i$$

Así pues, R^i será la matriz de elementos r_{pq}^i donde

$$r_{pi}^i = r_{ip}^i = \int c_{pi}^i(\underline{\psi}) \cdot p(\underline{\psi}) \cdot d\underline{\psi} = - \frac{\alpha_i \alpha_p}{\alpha_0 (\alpha_0 + 1)} \quad \text{si } p \neq i$$

$$r_{pp}^i = \int c_{pp}^i(\underline{\psi}) \cdot p(\underline{\psi}) \cdot d\underline{\psi} = \frac{\alpha_i \alpha_p (\alpha_p + 1)}{\alpha_0 (\alpha_0 + 1) (\alpha_0 - \alpha_i + 1)} \quad \text{si } p \neq i$$

$$r_{pq}^i = r_{qp}^i = \int c_{pq}^i(\underline{\psi}) \cdot p(\underline{\psi}) \cdot d\underline{\psi} = \frac{\alpha_i \alpha_p \alpha_q}{\alpha_0 (\alpha_0 + 1) (\alpha_0 - \alpha_i + 1)} \quad \text{si } p \neq q \text{ y } p, q \neq i$$

$$r_{ii}^i = \int c_{ii}^i(\underline{\psi}) \cdot p(\underline{\psi}) \cdot d\underline{\psi} = \frac{\alpha_i (\alpha_0 - \alpha_i)}{\alpha_0 (\alpha_0 + 1)}$$

donde las integrales anteriores están calculadas suponiendo que

$$\frac{e^{\Psi_m}}{1 + \sum_{j=1}^L e^{\Psi_j}} = x$$

Finalmente, la matriz R será la matriz de elementos r_{pq} ($p, q=1, 2, \dots, L$), donde

$$r_{pq} = \sum_{i=1}^L n_i \cdot r_{pq}^i$$

y la matriz $T_0 + R$ tendrá como elementos q_{pq} donde

$$q_{pq} = r_{pq} + t_{pq} = \begin{cases} r_{pp} + \alpha_p - \frac{\alpha^2}{\alpha_0} & p=q \\ r_{pq} - \frac{\alpha_p \alpha_q}{\alpha_0} & p \neq q \end{cases}$$

esto es,

$$T_0 + R = (q_{ij}) = \begin{cases} q_{ii} = \frac{n_i \alpha_i (\alpha_0 - \alpha_i)}{\alpha_0 (\alpha_0 + 1)} + \sum_{j \neq i}^L \frac{n_j \alpha_i \alpha_j (\alpha_i + 1)}{\alpha_0 (\alpha_0 + 1) (\alpha_0 - \alpha_j + 1)} + \\ + \frac{\alpha_i (\alpha_0 - 1)}{\alpha_0} \\ q_{ij} = - \frac{(n_i + n_j) \alpha_i \alpha_j}{\alpha_0 (\alpha_0 + 1)} - \frac{\alpha_i \alpha_j}{\alpha_0} + \\ + \sum_{h \neq i, j}^L \frac{n_h \alpha_h \alpha_i \alpha_j}{\alpha_0 (\alpha_0 + 1) (\alpha_0 - \alpha_h + 1)} \end{cases}$$

Aplicando, pues, el teorema 6.3.4 resulta en último término que

$$I^g \{E^S(\underline{n}), D^L(\underline{\alpha})\} \approx \frac{1}{2} \log \frac{|T_0 + R|}{|T_0|} = \frac{1}{2} \log \frac{\alpha_0 \cdot |T_0 + R|}{L+1 \prod_{i=1}^L \alpha_i}$$

donde las matrices R y T_0 han quedado definidas en todo el desarrollo anterior.

Tal como era de esperar la coherencia de este resultado queda comprobada al aplicarlo al modelo binomial y ver la coincidencia con la aproximación obtenida para dicho modelo por Bernardo (1979)*. En efecto, en el caso particular de que sean tan sólo dos los estratos en los que se divide la población total, esto es, si $L+1=2$, la determinación de una única componente θ_i ($i=1,2$) determina exactamente a la otra, al ser

$$\theta_1 + \theta_2 = 1$$

La correspondiente distribución sobre θ_1 , por ejemplo, es una beta de parámetros α_1 y α_2 (pues $\alpha_1 + \alpha_2 = \alpha_0$). En este caso resulta ser:

$$T_0 = \alpha_1 - \frac{\alpha_1^2}{\alpha_0} = \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2} = \frac{\alpha_1 \alpha_2}{\alpha_0}$$

$$R = n_1 \cdot \frac{\alpha_1 \alpha_2}{\alpha_0 (\alpha_0 + 1)}$$

de forma que

$$I_{\theta_1}^{\text{aproximada}} \left\{ E(n_1), \text{Ba}(\alpha_1, \alpha_2) \right\} \approx \frac{1}{2} \log \frac{\alpha_1 + \alpha_2}{\alpha_1 \alpha_2} \left(n_1 \cdot \frac{\alpha_1 \alpha_2}{\alpha_0 (\alpha_0 + 1)} + \frac{\alpha_1 \alpha_2}{\alpha_0} \right) = \frac{1}{2} \log \left(1 + \frac{n_1}{\alpha_1 + \alpha_2 + 1} \right)$$

resultado que coincide plenamente con el encontrado por Bernardo (1979)*.

CONCLUSIONES Y AREAS PARA INVESTIGACIONES ADICIONALES

En la tesis presentada se ha abordado el problema de la determinación del tamaño muestral que debe darse a un experimento realizado con el fin de obtener información sobre cierta cantidad de interés. Con este objeto, la inferencia estadística ha quedado descrita como un caso particular de la Teoría de la Decisión general en el que la clase de densidades de probabilidad posteriores constituye el espacio de decisiones. A fin de poder resolver el problema así planteado se definieron tres funciones para representar la utilidad esperada del experimento y que eran, a su vez, funciones de la información esperada. Con estas utilidades esperadas se obtuvieron los respectivos tamaños muestrales óptimos para los dos modelos concretos desarrollados en el capítulo segundo.

Precisamente son necesarias nuevas investigaciones que intenten encontrar las relaciones entre las distintas funciones anteriores para así poder comparar los resultados obtenidos cuando se emplean cada una de ellas por separado. La comparación entre las distintas constantes g_i resulta, así, una necesidad evidente. También resulta necesario comprobar en qué medida se ven afectadas ante distintas transformaciones de la cantidad de interés. Las transformaciones biyectivas son las primeras que debieran estudiarse.

La comparación entre el modelo desarrollado para una población uniforme con densidad inicial de Pareto con una población normal es imprescindible a fin de determinar hasta qué punto el primero de los modelos puede ser representativo de aquellos otros para los que los espacios muestrales están definidos en función del parámetro desconocido. En caso afirmativo, podría pensarse en la existencia de aproximaciones asintóticas para el comportamiento seguido por las informaciones esperadas de estos modelos con la información deducida para el modelo de población uniforme con densidad inicial de Pareto.

La generalización del concepto de información esperada a procesos de muestreo estratificados es otro de los resultados teóricos obtenidos en la presente tesis y que pone de manifiesto cómo dicho con-

cepto puede abarcar y ser introducido para otros diseños experimentales. En concreto, podría pensarse en diseñar adecuadamente toda una encuesta por muestreo, en el que el proceso de obtención de la muestra fuese una combinación de procesos estratificados y multietápicos.

También dentro de la generalización introducida por la presencia de muestras estratificadas podría pensarse en la necesidad de demostrar los resultados de unicidad de la expresión logarítmica para expresar la utilidad esperada de un experimento, en analogía con los resultados ya existentes para muestras aleatorias simples. En este sentido todas las discusiones al respecto, incluyendo la situación planteada por transformaciones biyectivas de la cantidad de interés, pueden plantearse en las nuevas circunstancias introducidas por la existencia de la muestra estratificada.

Finalmente, y también para la información esperada proporcionada por un proceso de muestreo estratificado, podrían estudiarse los comportamientos asintóticos de expresiones definidas como funciones de ella y sin necesidad de recurrir a hipótesis sobre la normalidad de las densidades empleadas. Un esfuerzo en este sentido resultaría de gran interés en cuanto que permitiría dar un conocimiento - aunque fuese aproximado - sobre informaciones esperadas de difícil obtención.

BIBLIOGRAFIA

- AL-BAYYATI, H.A. (1971) . A rule of thumb for determining a sample size in comparing two proportions. Technometrics 13 , 675 .
- ALCAIDE, A. (1966) . Lecciones de Econometría y Métodos Estadísticos. Madrid.
- APOSTOL, T.M. (1960) . Análisis Matemático. Barcelona: Reverté.
- ASH, R. (1965) . Information Theory. New York: Wiley-Intersciences.
- AZORIN, F. (1962) . Curso de Muestreo y Aplicaciones. 2ª Edic. Madrid: INE.
- BAYES, T.R. (1763) . Essay towards solving a problem in the doctrine of chances. Trabajo aparecido en Biometrika 45 (1958) , 293-315 .
- BASULTO, J. & BERNARDO, J.M. (1978) . Análisis Bayesiano de un proceso binomial. Trabajos de Estadística e Investigación Operativa 29 , 3-27 .
- BERNARDO, J.M. (1975) . The use of information in the design and analysis of scientific experimentation. Ph. D. Tesis. Universidad de Londres.
- BERNARDO, J.M. (1975)* . Information Theory and Decision Making. En Theories of Decision in Practice (White & Bowen eds.) , 247-251 . Londres: Hodder & Stoughton.
- BERNARDO, J.M. (1978) . Una medida de la información útil proporcionada por un experimento. Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales de Madrid, 72 . 419-440 .
- BERNARDO, J.M. (1979) . Expected information as expected utility. The Annals of Statistics 7 , 686-690 ,
- BERNARDO, J.M. (1979)* . Comportamiento asintótico de la información proporcionada por un experimento. Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales de Madrid 73 , 491-502 .
- COCHRAN, W.G. (1971) . Técnicas de Muestreo. México: CECOSA.
- COX, D.R. (1958) . Some Problems connected with Statistical Inference. Ann. Math. Statist. 29 , 357-372 .

- COX, D.R. & HINKLEY, D.V. (1974) . Theoretical Statistics. Londres: Chapman & Hall .
- DeFINETTI, B. (1937) . La prévision, ses lois logiques, ses sources subjectives. Trabajo aparecido en Studies in Subjective Probability (Kyburg & Smokler eds., 1964) . New York: Mac Graw-Hill.
- DeGROOT, M.H. (1970) . Optimal Statistical Decisions. New York: Mc Graw.
- FEDOROV, V.V. (1972) . Theory of Optimal Experiments. New York: Academic Press.
- FLEMING, W.H. (1969) . Funciones de diversas variables. Méjico: CECSA .
- FISHER, R.A. (1959) . Statistical Methods and Scientific Inference. Edinburgh: Oliver & Boyd.
- GARCIA ESPAÑA, E. (1974) . Diseño de la Encuesta General de Población. Madrid: INE .
- GRAYBILL, F.A. (1961) . An Introduction to Linnear Statistical Models. New York: Mc Graw-Hill.
- HAGSTROEM, K.G. (1925) . La loi de Pareto et la réassurance. S.A., pág 65.
- HAGSTROEM, K.G. (1944) . Inkomstutjämnigen i Sverige. Skand. Bankens Kwart. skr. (1944)
- HARDY, G.H. & LITTLEWOOD, J.E. & POLYA, G. (1952) . Inequalities. Cambridge University Press.
- HUANG, J.S. (1975) . A Note On Order Statistics From Pareto Distribution. Scandinavian Journal of Statistics, 187-190 .
- IBRAGIMOV, I.A. & HAS'MINSKY, R.Z. (1973) . On the information in a sample about a parameter. 2nd. Internat. symp. Information Theory. (Petrov & Csaki eds.). Budapest: Akadémiaikiadó.
- JOHNSON, N.L. (1962) . Estimation of sample size. Technometrics 4 , 59.
- LEE, P.M. (1964) . On the axioms of information theory. Ann. Math. Satist. 35 , 415-8 .
- LINDGREN B.W. (1968) . Statistical Theory. 2ª Edc. New York: Macmillan.

- LINDLEY, D.V. (1956). On a Measure of the Information provided by an Experiment. Ann. Math. Statist. 27 , 986-1005 .
- LINDLEY, D.V. (1957) . Binomial Sampling and the Concept of Information. Biometrika 44 , 179-46 .
- LINDLEY, D.V. (1965) . Introduction to Probability and Statistics. Cambridge University Press.
- LINDLEY, D.V. (1971) . Bayesian Statistics, a review. Reg. Conf. Ser. in Appl. Math. 2 . Philadelphia: SIAM .
- OSTEYEE, D.B. & GOOD, I.J. (1974) . Information, Weight of Evidence, the Singularity between Probability Measure and Signal Detection. Lecture Notes in Mathematics 376 . Berlin: Springer-Verlag .
- PENA, J.B. (1965) . Modelo Económico para el estudio de la distribución personal de la renta en España. Estadística Española, abril-junio 1965, 38-49 .
- PRATT, J.W. & RAIFFA, H. & SCHLAIFER, R. (1964) . The foundations of decision under uncertainty: an elementary exposition. J. Amer. Statist. Assoc. 59 , 353-75 .
- RAIFFA, H. & SCHLAIFER, R. (1961) . Applied Statistical Decision Theory. Cambridge Mass.: the MIT Press.
- RAMSEY, F.P. (1926) . Truth and Probability. Trabajo aparecido en Studies in Subjective Probability. (Kyburg & Smokler eds., 1964). 61-82 .
New York: Wiley.
- RENYI, A. (1970) . Probability Theory. Amsterdam: North Holland.
- RIOS, S. (1970) . Métodos Estadísticos. 6ª Edc. Madrid: Ediciones del Castillo.
- ROBINSON, G.K. (1975) . Conservative Statistical Inference. Tesis no publicada. Universidad de Londres.
- SANCHEZ-CRESPO, J.L. (1973) . Muestreo de poblaciones finitas aplicado al diseño de encuestas. Madrid: INE .
- SAVAGE, L.J. (1954) . The foundations of Statistics. New York: Dover.

- SHANNON, C.E. (1948) . A Mathematical Theory of Communication. Bell System Tech. J. 27 , 379-423, 623-56 .
- STONE, M. (1958) . Studies with a measure of information. Tesis no publicada. Universidad de Cambridge.
- VERES, E. (1978) . Utilidad e Información. Estadística Española, 80-81 , 39-56 .
- VERES, E. (1979) . La concavidad en la información esperada. Estadística Española, 82-83 , 79-93 .
- VILLEGAS, C. (1967) . On qualitative probability. Ann. Math. Mon. 74 , 661-9 .

Reunido el Tribunal que suscribe, en el día de la fecha, acordó otorgar, por unanimidad, a esta Tesis doctoral de

D. ERNESTO VERES FERRER

la calificación de Sobresaliente cum laude

Valencia, a 22 de Mayo de 1981

El Secretario,

El Presidente

