

03061
2eje.

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

UNIDAD ACADEMICA DE LOS CICLOS PROFESIONAL Y DE POSGRADO DEL COLEGIO DE CIENCIAS Y

HUMANIDADES

INSTITUTO DE INVESTIGACIONES EN MATEMATICAS APLICADAS Y EN SISTEMAS

ANALISIS BAYESIANO DE LA SELECCION DEL TAMAÑO DE MUESTRA EN UN CONTRASTE DE

HIPOTESIS MULTIPLES

TESIS

PARA OBTENER EL GRADO DE MAESTRO EN ESTADISTICA E INVESTIGACION DE OPERACIONES

PRESENTA EL ACTUARIO

JOSE GABRIEL HUERTA GOMEZ

México, D.F.

1994

TESIS CON
FALLA DE ORIGEN



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

DEDICATORIA

A mis padres. Gracias a su apoyo he podido culminar esta meta.

A mis hermanas. Por su afecto y amistad.

A la memoria de mi abuelita Herlinda.

A la memoria de Mimi.

A Dios.

AGRADECIMIENTOS

*Un agradecimiento especial para mi director de tesis, el Dr. Manuel Mendoza Ramírez.
Por el tiempo y dedicación que le dió a este trabajo. Por su paciencia y amistad.
El es el gran culpable de que me interese la Estadística Bayesiana.*

Un agradecimiento especial al Jurado revisor:

Dr. Federico O'Reilly Togno.

Dr. Ignacio Méndez Ramírez.

Dr. Manuel Mendoza Ramírez.

Dra. Belem Trejo Valdívila.

M. en C. Raúl Rueda Díaz del Campo.

Si algún acierto tiene este trabajo, en buena medida se debe a sus valiosos comentarios.

A mis profes de la maestría: Silvia, Lety, Belem, Raúl, Alejandro, Federico, Nacho, Manuel y Martín.

A mis cuates y cuatas de la maestría y de la especialización. Con ellos he tenido la oportunidad de convivir en el pasillo del primer piso, en los salones del cuarto piso y en muchos otros lugares dentro y fuera de la U.N.A.M. Gracias por los momentos tan divertidos que hemos tenido juntos.

A mis cuates y cuatas de la Facultad de Ciencias. Por su nobleza y tolerancia conmigo.

A la Facultad de Ciencias de la U.N.A.M. Institución en la que estude la licenciatura y en la que trabaje mientras estudiaba la maestría.

Al Instituto de Investigaciones en Matemáticas y en Sistemas de la U.N.A.M. por todo el apoyo dado durante la realización de este trabajo.

INDICE

CAPÍTULO I: INTRODUCCIÓN	
I.1. PLANTEAMIENTO DEL PROBLEMA	1
I.2. CONTENIDO	3
CAPÍTULO II: FUNDAMENTOS DE TEORÍA DE LA DECISIÓN Y ESTADÍSTICA BAYESIANA	
II.1. TEORIA DE LA DECISION	4
II.2. INFERENCIA ESTADISTICA Y TEORIA DE LA DECISION	11
II.3. TAMAÑO DE MUESTRA	21
CAPÍTULO III: PLANTEAMIENTO BAYESIANO DEL PROBLEMA ORIGINAL	
III.1. EL CONTRASTE DE HIPOTESIS	32
III.2. DISTRIBUCIONES INICIALES Y FINALES	34
III.3. FUNCIONES DE PERDIDA	42
III.4. TAMAÑO DE MUESTRA	56
CAPÍTULO IV: VARIANTE DEL PROBLEMA ORIGINAL	
IV.1. MODIFICACION AL PROCEDIMIENTO DE TAMAÑO DE MUESTRA	77
IV.2. DECISIONES DE REFERENCIA	79
IV.3. CONTRASTE DE HIPOTESIS Y TAMAÑO DE MUESTRA	83
CAPÍTULO V: COMENTARIOS FINALES	107
BIBLIOGRAFIA	109

CAPITULO I

INTRODUCCION

I.1.-PLANTEAMIENTO DEL PROBLEMA

La determinación del tamaño de muestra óptimo es un tema de interés en cualquier proceso de inferencia estadística. Es un hecho que el tamaño de muestra más conveniente es aquel que satisfaga una determinada condición de equilibrio entre costo (del muestreo) y precisión (de la inferencia). Es claro así, que la selección de un tamaño muestral específico tiene asociadas distintas consecuencias y que el problema puede plantearse como uno de decisión.

Desde esa perspectiva, resulta natural abordar el problema con un enfoque que, por principio, establezca explícitamente la situación como una toma de decisiones y que aborde la solución partiendo de principios generales que garanticen, en algún sentido, la coherencia de la respuesta que se produzca.

Este es el caso del enfoque bayesiano de la estadística y en este trabajo, a partir de una aplicación concreta, se formula el problema de selección del tamaño de muestra para realizar un contraste de hipótesis tal como se describe a continuación.

Un elemento relevante que se toma en cuenta al decidir la realización de un cultivo es el porcentaje de germinación de la semilla utilizada en la siembra. Existe una amplia experiencia en los procesos experimentales y el análisis estadístico para la estimación de la proporción de semillas de un lote que habrá de germinar en unas condiciones ambientales específicas. De hecho, esta experiencia ha dado lugar a procedimientos rutinarios de certificación de semillas.

Típicamente se trata, como ya se ha indicado, de obtener una estimación del porcentaje de germinación o de contrastar la hipótesis de que ese porcentaje rebasa un nivel mínimo aceptable para su utilización comercial. Un tema colateral, que resulta de interés, es el problema de seleccionar el tamaño de muestra óptimo para

realizar estas pruebas. De nuevo, cuando se trata de una estimación puntual o de un contraste de hipótesis convencional este aspecto no reviste mayor problema.

Sin embargo, cuando se trata de clasificar una variedad de semilla con mayor detalle se presentan aspectos que merece la pena considerar. Especialmente ocurre que el cálculo del tamaño de muestra óptimo no es ya, en general, simple.

Concretamente, si dadas unas condiciones de almacenamiento, una variedad de semilla se clasifica como susceptible cuando su porcentaje de germinación θ satisface la condición $\theta < \theta_1$, como resistente cuando $\theta_2 < \theta$ y como intermedia cuando $\theta_1 \leq \theta \leq \theta_2$, con $0 < \theta_1 < \theta_2 < 1$, entonces el problema de clasificar la variedad se convierte en el de contrastar el juego de hipótesis

$$H_i: \theta \in \Theta_i \quad i = 1, 2, 3 \quad (1)$$

en donde $\Theta_1 = [0, \theta_1]$, $\Theta_2 = [\theta_1, \theta_2]$ y $\Theta_3 = (\theta_2, 1]$, constituyen una partición del espacio parametral $\theta = [0, 1]$.

En la situación que aquí se trata, se tiene una variable aleatoria X para cada semilla de la variedad bajo estudio. Esta variable se define como

$$X = \begin{cases} 1 & \text{si la semilla germina} \\ 0 & \text{en caso contrario} \end{cases} \quad (2)$$

Naturalmente, X sigue una distribución Bernoulli con parámetro desconocido θ . Cuando se cuenta con información experimental producto de un muestreo de la variable X , los datos correspondientes se denotan por $Z_n = (X_1, X_2, \dots, X_n)$. Además, se supone que las variables aleatorias que conforman al vector Z_n son independientes de tal manera que X_1, X_2, \dots, X_n es una muestra aleatoria para todo entero positivo n .

En este trabajo, el problema de clasificación de variedades, formulado como un contraste de hipótesis, se aborda desde una perspectiva bayesiana mediante los conceptos más elementales de este enfoque. También se proponen procedimientos para resolver el correspondiente problema de tamaño de muestra a través del enfoque bayesiano.

1.2.-CONTENIDO

En el capítulo II se presenta una introducción general a la Teoría de la Decisión y a la Estadística Bayesiana. Se muestra como se pueden resolver los problemas de estimación y prueba de hipótesis con este enfoque y se hace énfasis en la alternativa bayesiana para tratar con el problema de selección del tamaño de muestra.

En el capítulo III se formula el problema de contraste a partir de la teoría expuesta en el capítulo II y se plantea un primer método para determinar el tamaño de muestra óptimo.

En el capítulo IV se presenta una modificación al método de tamaño de muestra del capítulo III a través del concepto de *Decisiones de Referencia* y se implementa en varios casos particulares.

Finalmente, en el capítulo V aparecen las conclusiones del trabajo.

CAPITULO II

FUNDAMENTOS DE TEORIA DE LA DECISION Y ESTADISTICA BAYESIANA

INTRODUCCION

En este capítulo se presentan los aspectos principales de la Teoría de la Decisión en ambiente de incertidumbre. Se formulan los problemas estadísticos como problemas de decisión y se comenta la utilidad del Teorema de Bayes para actualizar el conocimiento inicial a través de la información muestral. Finalmente, se da solución al problema estadístico de evaluación de tamaño de muestra mediante la estructura de un problema de decisión secuencial. Se ilustra con ejemplos.

II.1.- TEORIA DE LA DECISION

Cualquier problema de decisión es una situación en donde se tiene que elegir una opción dentro de un conjunto de posibles alternativas. Dicho conjunto de decisiones, que en este trabajo se denotará por \mathcal{D} , debe ser exhaustivo, para abarcar todas las opciones de interés al decisor y exclusivo, para considerar solo opciones alternativas.

Para poder discriminar entre las posibles decisiones es necesario que se conozcan las consecuencias a las que puede llevar un determinado curso de acción y la preferencia relativa de las mismas. Por esto, debe existir un conjunto \mathcal{C} de consecuencias y un orden de preferencia $<$ en \mathcal{C} tal que si $c_1 < c_2$ entonces, c_1 es menos preferible que c_2 , con c_1 y c_2 cualesquiera elementos en \mathcal{C} . Por extensión, cuando c_1 es no más preferible que c_2 se utiliza la notación $c_1 \leq c_2$. Obviamente \mathcal{C} debe incluir todas las consecuencias que aparecen en el problema y $<$ refleja las preferencias que un decisor particular tiene sobre ese conjunto de consecuencias.

Cuando cada elemento $d_i \in \mathcal{D}$, tiene asociada una única consecuencia $c_i \in \mathcal{C}$, se tiene un problema de decisión sin incertidumbre, ya que cada decisión lleva con toda

seguridad, sin importar si ocurre o no suceso alguno, a una sola consecuencia. En este caso, el problema de decisión se resuelve determinando $c^* \in \mathcal{C}$ tal que $c \preceq c^*$ para todo $c \in \mathcal{C}$, y especificando cualquier $d^* \in \mathcal{D}$, cuya consecuencia sea c^* . Por supuesto, si el decisor tiene dificultades para ordenar sus consecuencias, lo cual, en particular, puede ocurrir si \mathcal{C} es finito pero muy grande o cuando \mathcal{C} es infinito, se puede dificultar la obtención de c^* . De todas maneras, desde un punto de vista teórico, un problema de decisión sin incertidumbre definido por la terna $(\mathcal{D}, \mathcal{C}, \preceq)$, se resuelve buscando la opción que produce la consecuencia "más preferible".

Cuando $D=(d_1, d_2, \dots, d_k)$ y $\mathcal{C}=(c_1, c_2, \dots, c_k)$, el problema de decisión sin incertidumbre se puede representar mediante un **Árbol de Decisión** donde cada "rama" esta asociada a un elemento de \mathcal{D} y cada consecuencia a una "hoja" del árbol. El cuadro representa un nodo de decisión. Ver figura II.1.1.

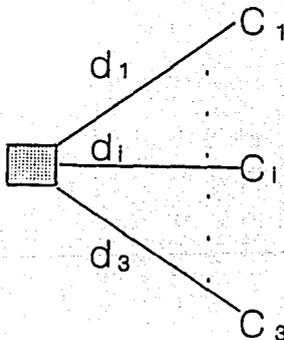


FIGURA II.1.1

ARBOL DE DECISION SIN INCERTIDUMBRE

Graficamente, la solución a un problema de decisión sin incertidumbre, consiste en "podar el árbol" quitando todas aquellas ramas cuya hoja no es c^* .

Existen problemas de decisión en donde cada opción no necesariamente lleva a una única consecuencia. En estos casos, para cada $d_i \in \mathcal{D}$, existe un conjunto de sucesos inciertos Θ_i , donde la ocurrencia eventual de algún $\theta_{ij} \in \Theta_i$ lleva a la consecuencia c_{ij} . En otras palabras, un problema de decisión en ambiente de incertidumbre está definido por:

\mathcal{D} conjunto de posibles cursos de acción.

Θ_i conjunto de sucesos o eventos inciertos para cada $d_i \in \mathcal{D}$,
con $i \in I$, un conjunto de índices.

\prec un orden de preferencia en \mathcal{E} .

f_i una función definida para cada decisión en \mathcal{D} , con dominio Θ_i y contradominio \mathcal{E} que a cada $\theta_{ij} \in \Theta_i$, con $j \in J_i$, conjunto de índices, le asocia la consecuencia c_{ij} . En términos simples, c_{ij} es la consecuencia de decidirse por d_i y de que ocurra el evento θ_{ij} .

Es necesario que el decisor construya Θ_i en forma exhaustiva y excluyente para así considerar todos los sucesos relevantes en el proceso de toma de decisiones y para que la ocurrencia de cada θ_{ij} lleve a una única consecuencia. Para lograr que los eventos sean mutuamente excluyentes basta considerar espacios producto apropiados. Ver Lindley (1984), Bernardo (1981), Bernardo y Smith (1994). Sin embargo, la exhaustividad no es posible garantizarla únicamente con espacios producto. Esta propiedad requiere un conocimiento amplio del área de aplicación y exige que $\Omega = \bigcup_j \theta_{ij}$, con Ω el evento seguro. Esto último equivale a decir que al menos un θ_{ij} ocurre con toda certeza.

Sí $\mathcal{D} = \{d_1, d_2, \dots, d_i, \dots, d_k\}$ y $\Theta_i = \{\theta_{i1}, \theta_{i2}, \dots, \theta_{ij}, \dots, \theta_{in_i}\}$ es posible representar el problema de decisión en ambiente de incertidumbre también mediante un diagrama de árbol. El diagrama comienza en un nodo de decisión representado por un cuadro. Las ramas representan decisiones que conducen a un nodo de incertidumbre (representado por un círculo). De cada uno de estos *nodos de incertidumbre* salen ramas que representan a los sucesos θ_{ij} . Al final de estas ramas, aparecen las

consecuencias. Ver figura II.1.2.

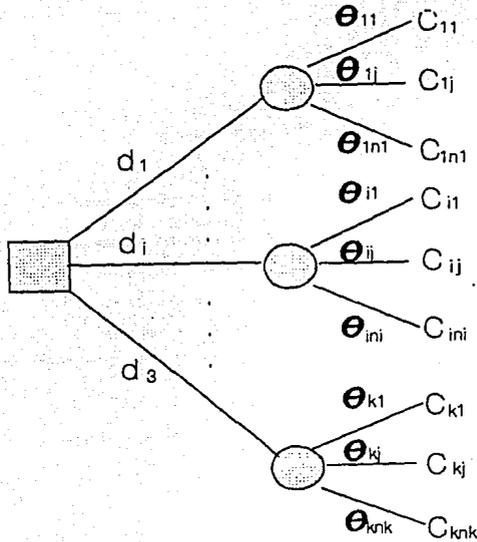


FIGURA II.1.2

ARBOL DE DECISION CON INCERTIDUMBRE

Muchos problemas de decisión aparentemente no presentan este esquema, sin embargo, siempre se puede reformular un problema de decisión con la estructura simple que aquí se ha expuesto. De hecho, es posible incorporar una simplificación más: correspondientes espacios producto es posible transformar el problema de decisión a uno donde el espacio de sucesos inciertos coincida para todas las decisiones. Para

estudiar esto con más detalle, el lector puede consultar Lindley (1984), Bernardo (1981), Bernardo y Smith (1994).

De esta forma, se tiene la siguiente definición:

DEFINICION II.1.1. Un Problema de Decisión en Ambiente de Incertidumbre, está definido por la quinteta $\mathcal{D}, \Theta, \mathcal{C}, <$ y f , con $\mathcal{D}, \Theta, \mathcal{C}, <$ como antes y f una función con dominio en el producto cartesiano $\mathcal{D} \times \Theta$, contradominio en \mathcal{C} y que a cada par $d \in \mathcal{D}$ y $\theta \in \Theta$ le asocia una consecuencia $c \in \mathcal{C}$.

La incertidumbre que cada decisor tiene sobre la ocurrencia de los eventos en Θ se puede cuantificar a través de probabilidades. Si Θ es finito o numerable bastaría con especificar números $p(\theta_j)$ para toda θ_j en Θ , tales que $0 \leq p(\theta_j) \leq 1$ y que $\sum_j p(\theta_j) = 1$. En este caso, la cantidad $p(\theta_j)$ resulta ser la probabilidad que el decisor asigna al evento θ_j . Cuando Θ es no numerable la cuantificación de la incertidumbre se puede lograr mediante una función de densidad $p(\theta)$ definida sobre una σ -álgebra que contenga a Θ . En lo que resta del trabajo se utilizará el término función de densidad de probabilidad generalizada (f.d.p.g.) para denominar una función de probabilidades o una función de densidad.

Por otra parte, en lo que toca a las preferencias, existe una función u de \mathcal{C} en \mathbb{R} tal que

$$c_1 < c_2 \text{ si y solo si } u(c_1) < u(c_2),$$

donde $<$ denota el orden de preferencias en \mathcal{C} y $<$ el orden de los números reales. u recibe el nombre de Función de Utilidad y mide las preferencias que sobre las consecuencias tiene un decisor, en una escala numérica.

En vez de u se puede tener l de \mathcal{C} a \mathbb{R} tal que

$$c_1 < c_2 \text{ si y solo si } l(c_2) < l(c_1).$$

En este caso, l define una Función de Pérdida e igual le sirve al decisor para

cuantificar numéricamente sus preferencias. Por supuesto, si l es una pérdida, $u = -l$ define una utilidad y análogamente, si u es una utilidad $l = -u$ es una pérdida.

Con l se puede definir una función L de dominio $\mathcal{D} \times \Theta$ y contradominio \mathbb{R} e igual a la composición de f seguida de l , es decir, $L = l \circ f$. Para cada $d \in \mathcal{D}$ y $\theta \in \Theta$, $L(d, \theta)$ denota la pérdida que se obtiene al tomar la decisión d , cuando ocurre el suceso θ . Si se tiene u en vez de l , el problema queda definido por la correspondiente $U = u \circ f$. Muchas veces es posible construir directamente L o U , sin hacer explícitas a las funciones l y f , de tal manera que refleje adecuadamente las preferencias del decisor.

Con Θ finito o numerable y especificadas $p(\theta_j)$ y $L(d, \theta_j)$, una manera de resolver el problema de decisión es calculando la pérdida esperada de la decisión d para toda d en \mathcal{D} , es decir

$$E_p(L(d, \theta)) = \sum_{\theta_j \in \Theta} L(d, \theta_j) p(\theta_j), \quad (1)$$

y encontrando $d^* \in \mathcal{D}$ tal que

$$E_p(L(d^*, \theta)) = \min_{\mathcal{D}} E_p(L(d, \theta)). \quad (2)$$

Si Θ es no numerable pero están determinadas $p(\theta)$ y $L(d, \theta)$, el problema se podría resolver calculando

$$E_p(L(d, \theta)) = \int_{\Theta} L(d, \theta) dp(\theta) \quad (3)$$

y minimizando $E_p(L(d, \theta))$ sobre el conjunto \mathcal{D} . Aquí, $E_p(L(d, \theta))$ también es la pérdida esperada de la decisión d y el símbolo de integral denota una integral de Lebesgue respecto a la f.d.p.g $p(\theta)$. Esta manera de tomar decisiones se conoce como *Principio de la Pérdida Esperada Mínima*. La decisión por Pérdida Esperada Mínima d^* se conoce como *Decisión de Bayes* y a $E_p(L(d^*, \theta))$ se le denomina el *Valor de Bayes*.

En vez de $L(d, \theta)$ se puede tener $U(d, \theta)$ P -medible, así, el problema de decisión

se puede resolver maximizando $Ep(U(d,\theta))$ sobre \mathcal{D} , lo que significa actuar conforme al *Principio de la Utilidad Esperada Máxima*. En este trabajo únicamente se manejarán funciones de pérdida y en consecuencia sólo se usará el Principio de la Pérdida Esperada Mínima.

Visto sobre el árbol de decisión de la figura II.1.2, el Principio de la Pérdida Esperada consiste en sustituir (podar) el nodo de incertidumbre para cada opción por el valor $Ep(L(d,\theta))$ transformando el árbol en uno de decisión sin incertidumbre, para el que ya se ha establecido la solución óptima.

Existen otras formas de resolver un problema de decisión en ambiente de incertidumbre como el método Minimax, (Ferguson (1969), Berger (1985)) o como el método de la Consecuencia Más Probable (Bernardo (1981)). Sin embargo, a partir de una serie de axiomas que definen lo que significa que un decisor sea coherente y racional ante un problema de decisión, se ha demostrado que, si se desea respetar dichos axiomas, deben existir probabilidades sobre Θ (o una función de densidad sobre Θ) que cuantifique la incertidumbre sobre Θ y una pérdida $L(d,\theta)$ o una utilidad $U(d,\theta)$, que cuantifique las preferencias. Además, la única forma coherente de resolver el problema de decisión es mediante el Principio de la Pérdida (Utilidad) Esperada Mínima (Máxima). En textos como Berger (1985), De Groot (1970), Bernardo (1981), Bernardo y Smith (1994) aparecen discusiones detalladas respecto a los Axiomas de Coherencia de la Teoría de la Decisión.

En la práctica, la especificación adecuada de las probabilidades y de la pérdida suele complicarse. De todas maneras, en este trabajo todos los problemas de decisión en ambiente de incertidumbre que aparezcan se resolverán mediante el Principio de la Pérdida Esperada Mínima. De hecho, en los problemas de decisión que se discuten en el Capítulo III se propondrá una amplia gama de posibilidades para determinar probabilidades (densidades) y pérdidas satisfactoriamente.

II.2.- INFERENCIA ESTADÍSTICA Y TEORIA DE LA DECISION

Comúnmente los problemas de inferencia estadística se pueden formular como problemas de decisión con los elementos \mathcal{D} , Θ , $p(\theta)$ y $L(d, \theta)$. En este caso, Θ está determinado por el conjunto de posibles valores que puede tomar un parámetro desconocido θ asociado a un modelo $f(X|\theta)$ que describe el comportamiento de un vector aleatorio observable X y $p(\theta)$ se interpreta como la probabilidad o función de densidad asociada al evento de que el parámetro tome precisamente el valor θ .

Si θ es una magnitud continua, el valor esperado de $L(d, \theta)$ respecto de p , denotado por $E_p(L(d, \theta))$, satisface

$$E_p(L(d, \theta)) = \int_{\Theta} L(d, \theta) p(\theta) d\theta. \quad (1)$$

Si θ es discreta,

$$E_p(L(d, \theta)) = \sum_{\theta_j \in \Theta} L(d, \theta_j) P(\theta_j). \quad (2)$$

En ambos casos, la solución d^* , por Pérdida Esperada Mínima, debe ser tal que

$$E_p(L(d^*, \theta)) = \min_{\mathcal{D}} (E_p(L(d, \theta))). \quad (3)$$

La densidad $p(\theta)$ puede cuantificar la incertidumbre sobre θ basándose únicamente en el conocimiento que el decisor tenga sobre θ , sin considerar resultados experimentales actuales. En este caso $p(\theta)$ recibe el nombre de distribución inicial o distribución a priori de θ y se denotará por $P(\theta)$.

Cuando $p(\theta)$ no sólo toma en cuenta experiencia previa sino que también considera la información dada por n experimentos X_1, X_2, \dots, X_n , a $p(\theta)$ se le llama distribución final o distribución a posteriori de θ y se denota por $P(\theta|X_1, X_2, \dots, X_n)$.

Si se ha determinado a $P(\theta)$ y se conoce la f.d.p.g. conjunta de X_1, X_2, \dots, X_n

dado θ , que se denotará por $f(X_1, X_2, \dots, X_n | \theta)$, $P(\theta | X_1, X_2, \dots, X_n)$ se puede obtener vía el Teorema de Bayes, es decir

$$P(\theta | X_1, X_2, \dots, X_n) = \frac{f(X_1, X_2, \dots, X_n | \theta)P(\theta)}{P(X_1, X_2, \dots, X_n)} \quad (4)$$

con $P(X_1, X_2, \dots, X_n)$ la f.d.p.g conjunta del vector $Z_n = (X_1, X_2, \dots, X_n)$ y que no depende de θ . $P(X_1, X_2, \dots, X_n)$ también recibe el nombre de distribución predictiva.

En el caso continuo,

$$P(X_1, X_2, \dots, X_n) = \int_{\Theta} f(X_1, X_2, \dots, X_n | \theta)P(\theta)d\theta \quad (5)$$

y en el caso discreto,

$$P(X_1, X_2, \dots, X_n) = \sum_{\theta_j \in \Theta} f(X_1, X_2, \dots, X_n | \theta_j)P(\theta_j). \quad (6)$$

Por facilidad, todas las f.d.p. generalizadas donde aparece X_1, X_2, \dots, X_n se escribirán usando la notación $Z_n = (X_1, X_2, \dots, X_n)$. Con esto, el Teorema de Bayes toma la siguiente forma

$$P(\theta | Z_n) = \frac{f(Z_n | \theta)P(\theta)}{P(Z_n)}. \quad (7)$$

Como $P(Z_n)$ no depende de θ , $P(\theta | Z_n)$ queda caracterizada por la siguiente expresión

$$P(\theta | Z_n) \propto f(Z_n | \theta)P(\theta). \quad (8)$$

Como función de θ , la expresión anterior establece que la distribución final es proporcional al producto de la verosimilitud por la inicial. De aquí que se puede obtener directamente el núcleo de la distribución final a partir de la verosimilitud

y la inicial y si el núcleo de $P(\theta|Z_n)$ corresponde al de un modelo conocido, no es necesario calcular $P(Z_n)$ mediante un proceso de integración.

Ejemplo II.2.1.

Sea X_1, X_2, \dots, X_n una muestra aleatoria con f.d.p.g. (dado θ) Bernoulli (θ). Suponga que $P(\theta)$ es una densidad Beta (α, β), es decir,

$$P(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} I_{(0,1)}(\theta), \quad \alpha > 0, \beta > 0. \quad (9)$$

con $\Gamma(\cdot)$ la función Gamma e $I_A(\cdot)$ la función indicadora sobre el conjunto A . Calcular $P(\theta|Z_n)$. (En el resto del trabajo cuando alguna función de densidad $\pi(\theta)$ sea Beta (α, β), se usará la notación $\pi(\theta) = \text{Be}(\alpha, \beta)$).

Se sabe que $f(X_i|\theta) = \theta^{x_i} (1-\theta)^{1-x_i} I_{(0,1)}(X_i)$. Como las observaciones conforman una muestra aleatoria

$$f(Z_n|\theta) = \theta^Y (1-\theta)^{n-Y} \prod_{i=1}^n I_{(0,1)}(X_i) \quad (10)$$

$$\text{con } Y = \sum_{i=1}^n X_i.$$

Por otra parte,

$$P(\theta|Z_n) \propto P(Z_n|\theta)P(\theta)$$

$$\propto \theta^Y (1-\theta)^{n-Y} \theta^{\alpha-1} (1-\theta)^{\beta-1} I_{(0,1)}(\theta). \quad (11)$$

Por lo que

$$P(\theta|Z_n) \propto \theta^{\alpha+Y-1} (1-\theta)^{\beta+n-Y-1} I_{(0,1)}(\theta) \quad (12)$$

Notar que el lado derecho de esta última expresión es el núcleo de una Beta de

parámetros $\alpha+Y$, $\beta+n-Y$ y por esto $P(\theta|Z_n)$ es una Beta con estos parámetros. Observe que

$$P(Z_n) = \frac{\Gamma(\alpha+Y) \Gamma(\beta+Y)}{\Gamma(\alpha+\beta+n)} \prod_{i=1}^n \binom{1}{(0,1)} (X_i) \quad (13)$$

Este ejemplo presenta un par de aspectos interesantes. Primero, se prueba que bajo muestreo Bernoulli con una inicial Beta, la final vuelve a ser Beta. Por esta razón, a la familia Beta se le denomina *Conjugada* para el muestreo Bernoulli. En general, toda familia paramétrica que sea cerrada bajo la aplicación del Teorema de Bayes con respecto a una verosimilitud específica se le llama *Familia Conjugada* para ese tipo de muestreo.

Segundo, $P(\theta|Z_n)$ depende de la muestra únicamente a través de $Y = \sum_{i=1}^n X_i$. Esto implica, que en el sentido bayesiano, Y es suficiente para θ . En general, si T_n es una estadística tal que $P(\theta|Z_n) = P(\theta|T_n)$ con

$$P(\theta|T_n) \propto f(T_n|\theta)P(\theta) \quad (14)$$

donde $f(T_n|\theta)$ es la f.d.p.g. de T_n condicional a θ entonces, T_n es una estadística suficiente para θ . El concepto de suficiencia aquí expuesto es equivalente al concepto de suficiencia de Neyman-Pearson (De Groot (1970)).

La idea de Familias Conjugadas tiene relevancia cuando esa familia distribuciones iniciales es suficientemente amplia como para que el decisor pueda expresar su incertidumbre sobre θ con un miembro de la familia. Hay una conexión directa entre familias conjugadas y suficiencia, ya que si existe una estadística suficiente de dimensión fija, siempre es posible construir una Familia Conjugada Paramétrica. Todos estos puntos se encuentran detallados en De Groot (1970), Berger (1985), Bernardo y Smith (1994).

Con lo que se ha expuesto en esta sección, se puede notar que para tomar "Decisiones Estadísticas" a través del Principio de la Pérdida Esperada Mínima, hay

que determinar una función de pérdida $L(d, \theta)$ que cuantifique adecuadamente las preferencias entre las consecuencias y especificar una distribución inicial $P(\theta)$, que exprese el conocimiento inicial sobre θ . Con base en este principio y si no se incorpora información muestral, la decisión óptima se obtiene al minimizar $E_p(L(d, \theta))$ sobre \mathcal{D} . Si se usa la información dada por Z_n , es necesario usar Bayes para obtener $P(\theta|Z_n)$ y después encontrar la decisión óptima al minimizar $E_{p|Z_n}(L(d, \theta))$ sobre \mathcal{D} , donde $E_{p|Z_n}(L(d, \theta))$ denota al valor esperado de $L(d, \theta)$ respecto de $P(\theta|Z_n)$.

En problemas de estimación puntual \mathcal{D} es el espacio parametral y en pruebas de hipótesis \mathcal{D} está formado por las hipótesis que se piensa contrastar. Además, usualmente es fácil determinar modelos interesantes para $L(d, \theta)$ en problemas de estimación puntual y contraste de hipótesis. $P(\theta)$ puede ser más difícil de precisar, aunque el concepto de conjugadas a veces resulta útil en este sentido. Más aún, según Diaconis e Ylvisaker (1985) la incertidumbre sobre θ siempre se puede aproximar por una distribución mezcla de conjugadas, aunque en la práctica su determinación puede complicarse. Parte de este trabajo está dedicada a proponer y estudiar el comportamiento de diferentes alternativas para $L(d, \theta)$ y $P(\theta)$ en el contexto del problema de contraste de hipótesis planteado en el capítulo I.

A continuación se describen algunos ejemplos de estimación puntual y pruebas de hipótesis analizados como problemas de decisión.

Ejemplo II.2.2.

En el contexto del ejemplo II.2.1, suponga que se quiere estimar puntualmente el parámetro θ . En este caso $\mathcal{D} = [0, 1]$ y por facilidad de notación, $\Theta = [0, 1]$.

Si se usa pérdida cuadrática, es decir si

$$L(d, \theta) = k(d - \theta)^2 \quad \text{con } k > 0, \quad (15)$$

la decisión de Bayes es aquella que minimiza

$$kE_p(d - \theta)^2. \quad (16)$$

sea cual sea la densidad $p(\theta)$.

Una simple cálculo permite probar que la decisión de Bayes es

$$d^* = E_p(\theta), \quad (17)$$

con $E_p(\theta)$ la media de la función de densidad $p(\theta)$.

El valor de Bayes resulta ser

$$k E_p(E_p(\theta) - \theta)^2 = k \text{Var}_p(\theta), \quad (18)$$

con $\text{Var}_p(\theta)$ la varianza de $p(\theta)$.

Cuando, como en el ejemplo II.2.1., $P(\theta)$ es una Beta (α, β) ,

$$d^* = \frac{\alpha}{\alpha + \beta} \quad (19)$$

y

$$k \text{Var}_p(\theta) = \frac{k\alpha\beta}{(\alpha + \beta + 1)(\alpha + \beta)^2} \quad (20)$$

Si se usan datos y una inicial Beta, entonces

$$d^* = \frac{\alpha + Y}{\alpha + \beta + n} \quad (21)$$

y

$$k \text{Var}_{P|Z_n}(\theta) = \frac{k(\alpha + Y)(\beta + n - Y)}{(\alpha + \beta + n + 1)(\alpha + \beta + n)^2} \quad (22)$$

con $\text{Var}_{P|Z_n}(\theta)$ la varianza de la final $P(\theta|Z_n)$.

Si $L(d, \theta) = k |d - \theta|$ la solución de Bayes es la mediana de la distribución final o inicial, dependiendo si se incorporan o no datos (Consultar Berger (1985) o De Groot (1970)). En el caso Beta puede ser necesario usar tablas de Beta incompleta para calcular la mediana de la inicial o de la final.

Ejemplo II.2.3.

Nuevamente, en el mismo contexto del ejemplo II.2.1, suponga que se quiere probar

$$H_0: \theta = \theta_0 \quad \text{vs.} \quad H_1: \theta = \theta_1 \quad (23)$$

con $\theta_0 < \theta_1$ y θ_0, θ_1 los únicos posibles valores que puede tomar el parámetro θ .

Ahora $\mathcal{D} = \{H_0, H_1\}$, $\Theta = \{\theta_0, \theta_1\}$ y $P(\theta)$ es cualquier densidad discreta en Θ . Supongamos $P(\theta_0) = \pi$, $P(\theta_1) = 1 - \pi$ y que la pérdida vale 0 si se toma una decisión correcta y uno en caso contrario. Esto se sintetiza en la tabla que se presenta a continuación.

TABLA DE PERDIDA

		θ		
		θ_0	θ_1	
\mathcal{D}	θ_0	0	1	$P(\theta_0) = \pi$ $P(\theta_1) = 1 - \pi$
	θ_1	1	0	

Usando la distribución inicial,

$$E_p(L(H_0, \theta)) = 1 - \pi \quad (24)$$

y

$$E_p(L(H_1, \theta)) = \pi. \quad (25)$$

De esto se concluye que H_0 es Bayes cuando $\pi > \frac{1}{2}$ y el valor de Bayes es $1-\pi$. Si $\pi < \frac{1}{2}$, H_1 es Bayes y el valor de Bayes es π . Con $\pi = \frac{1}{2}$, tanto H_0 como H_1 son decisiones de Bayes ambas con valor de Bayes igual a $\frac{1}{2}$.

Si se quiere efectuar el contraste usando datos es necesario calcular $P(\theta_0|Z_n)$ y $P(\theta_1|Z_n)$, las probabilidades finales de cada una de las hipótesis. Por el Teorema de Bayes y como los datos tienen distribución Bernoulli

$$P(\theta_0|Z_n) = \frac{\pi \theta_0^Y (1-\theta_0)^{n-Y}}{\pi \theta_0^Y (1-\theta_0)^{n-Y} + (1-\pi) \theta_1^Y (1-\theta_1)^{n-Y}} \quad (26)$$

con $Y = \sum_{i=1}^n X_i$ y

$$P(\theta_1|Z_n) = 1 - P(\theta_0|Z_n). \quad (27)$$

Con la distribución final,

$$E_{p|Z_n}(L(H_0, \theta)) = 1 - P(\theta_0|Z_n) \quad (28)$$

y

$$E_{p|Z_n}(L(H_1, \theta)) = P(\theta_0|Z_n). \quad (29)$$

En consecuencia, con datos, H_0 es Bayes sí y solo si $P(\theta_0|Z_n) > \frac{1}{2}$ con valor de Bayes $1 - P(\theta_0|Z_n)$. H_1 es Bayes cuando $P(\theta_0|Z_n) < \frac{1}{2}$ y su valor de Bayes es $P(\theta_0|Z_n)$. Finalmente, si $P(\theta_0|Z_n) = \frac{1}{2}$, H_0 y H_1 son decisiones de Pérdida Esperada Mínima con

valor de Bayes igual a $\frac{1}{2}$.

Mediante un poco de álgebra es fácil verificar que H_0 es Bayes cuando $Y > K_n$, H_1 es Bayes si $Y < K_n$ y tanto H_0 como H_1 son Bayes si $Y = K_n$ con

$$K_n = \frac{\ln(\pi/(1-\pi)) + n \ln((1-\theta_1)/(1-\theta_0))}{\ln((\theta_0(1-\theta_1))/(\theta_1(1-\theta_0)))} \quad (30)$$

Es decir, H_0 se "rechaza" si $Y > K_n$ con K_n una constante. Esta regla de decisión es similar al criterio clásico de Neyman-Pearson.

Ejemplo II.2.4.

Una vez más en la situación del ejemplo II.2.1, suponga que ahora se quiere probar

$$H_0: \theta \leq \theta_0 \quad \text{vs.} \quad H_1: \theta > \theta_1. \quad (31)$$

Como en el ejemplo II.2.3, $\mathcal{D} = (H_0, H_1)$ pero ahora $\theta = [0, 1]$. En todo momento se supone que la pérdida es del tipo 0 o 1, es decir, si se decide correctamente la pérdida vale cero y si no vale uno. La tabla siguiente expresa la pérdida en forma explícita.

TABLA DE PERDIDA
 θ

		$\theta \leq \theta_0$	$\theta > \theta_0$
\mathcal{D}	$\theta \leq \theta_0$	0	1
	$\theta > \theta_0$	1	0

Si $P(\theta \leq \theta_0)$ es la probabilidad inicial y $P(\theta \leq \theta_0 | Z_n)$ la probabilidad final de H_0 respectivamente entonces,

$$E_p(L(H_0, \theta)) = 1 - P(\theta \leq \theta_0), \quad (32)$$

$$E_p(L(H_1, \theta)) = P(\theta \leq \theta_0), \quad (33)$$

$$E_{P|Z_n}(L(H_0, \theta)) = 1 - P(\theta \leq \theta_0 | Z_n) \quad (34)$$

y

$$E_{P|Z_n}(L(H_1, \theta)) = P(\theta \leq \theta_0 | Z_n). \quad (35)$$

De estas expresiones se puede observar que H_0 es Bayes, sin datos cuando $P(\theta \leq \theta_0) \geq \frac{1}{2}$ y que H_0 es Bayes con datos si $P(\theta \leq \theta_0 | Z_n) \geq \frac{1}{2}$. Análogamente, H_1 es Bayes sin datos cuando $P(\theta \leq \theta_0) \leq \frac{1}{2}$ y H_1 es Bayes con datos si $P(\theta \leq \theta_0 | Z_n) \leq \frac{1}{2}$.

El árbol deja ver que el problema de tamaño de muestra, visto como uno de decisión, presenta más de un nodo de decisiones. Cualquier problema de decisión con esta estructura recibe el nombre de *Problema de Decisión Secuencial*. Otro aspecto que se puede observar de la figura II.3.1 es que para el caso $n=0$ la pérdida sólo depende de d y θ . Esto porque la opción $n=0$ equivale a tomar decisiones en \mathcal{D} sin incorporar información muestral.

En resumen, los elementos que intervienen en el problema de tamaño de muestra visto como uno de decisión son:

N el conjunto de los enteros no negativos cuyos elementos se van a denotar por n .

Z_n el conjunto de posibles muestras de tamaño n con elementos denotados por Z_n ($Z_0 = \emptyset$) con distribución condicional $f(Z_n|\theta)$.

\mathcal{D} el conjunto de decisiones definidas por el problema de inferencia.

Θ el espacio parametral con f.d.p.g inicial $P(\theta)$.

$f(Z_n|\theta)$ la f.d.p.g. de la muestra Z_n dado el parámetro θ .

$L(n, Z_n, d, \theta)$ la pérdida que resulta de elegir el tamaño de muestra n , observar Z_n , decidirse por d y con θ el verdadero valor del parámetro. Por notación, $L(0, Z_0, d, \theta) = L(d, \theta)$.

En Bernardo y Smith (1994) se presenta y justifica la solución general de cualquier problema de decisión secuencial. Esta consiste en podar el árbol, de atrás hacia adelante, mediante el Principio de la Pérdida Esperada. En nuestro caso, esto implica lo siguiente:

Para $n=0$, determinar d_0^* tal que

$$E_p(L(d_0^*, \theta)) = \min_{\mathcal{D}} E_p(L(d, \theta)). \quad (1)$$

Asignar como consecuencia a la decisión $n=0$ el valor

$$L^*(0) = \min_D E_p(L(d, \theta)). \quad (2)$$

Si $n > 0$, para cada $Z_n \in Z_n$ determinar $d^*_{Z_n}$ tal que

$$E_{p|Z_n}(L(n, Z_n, d^*_{Z_n}, \theta)) = \min_D E_{p|Z_n}(L(n, Z_n, d, \theta)) \quad (3)$$

con $E_{p|Z_n}(L(n, Z_n, d, \theta))$ el valor esperado de $L(n, Z_n, d, \theta)$ respecto de $P(\theta|Z_n)$. Considerar como consecuencia de la decisión $n > 0$ y del suceso Z_n la pérdida

$$L^*(n, Z_n) = \min_D E_{p|Z_n}(L(n, Z_n, d, \theta)). \quad (4)$$

Efectuados estos pasos, el árbol de la figura II.3.1 se reduce a un árbol con la estructura sencilla de problemas de decisión descrita en la sección II.1 de este trabajo. Ver figura II.3.2.

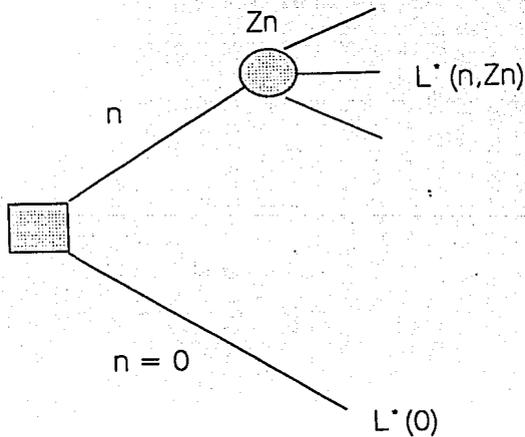


FIGURA II.3.2
ÁRBOL DE DECISION SIMPLIFICADO. PROBLEMA DE TAMAÑO DE MUESTRA

Ahora, la solución al problema consiste en buscar $n^* > 0$ tal que

$$L(n^*) = E_{P_n}(L^*(n^*, Z_n)) = \min_{n>0} E_{P_n}(L^*(n, Z_n)) \quad (5)$$

con $E_{P_n}(L^*(n, Z_n))$ el valor esperado de $L^*(n, Z_n)$ respecto de la f.d.p.g. $P(Z_n)$ y comparar $L^*(n)$ con $L^*(0)$. Si $L^*(0) \leq L^*(n)$ el tamaño óptimo es $n_0 = 0$ y si $L^*(n) < L^*(0)$ el tamaño óptimo es $n_0 = n^*$. Elegido n_0 y obtenida la muestra Z_{n_0} la decisión óptima en \mathcal{D} es $d_{z_{n_0}}^*$. Si $n_0 = 0$, la decisión óptima en \mathcal{D} es d_0^* .

Comúnmente se supone que la pérdida $L(n, Z_n, d, \theta)$ es igual a la suma de una pérdida que sólo depende de d y θ más el costo de muestrear n piezas de información y observar Z_n . Esto es,

$$L(n, Z_n, d, \theta) = L(d, \theta) + C(n, Z_n) \quad (6)$$

con $C(n, Z_n)$ la función de costo que se debe expresar en las mismas unidades de $L(d, \theta)$ y tal que $C(0, Z_0) = 0$. Con este supuesto, $d_{z_n}^*$ satisface

$$E_{P|Z_n}(L(d^* Z_n, \theta)) = \min_{\mathcal{D}} E_{P|Z_n}(L(d, \theta)), \quad (7)$$

$$L^*(n, Z_n) = \min_{\mathcal{D}} E_{P|Z_n}(L(d, \theta)) + C(n, Z_n) \quad (8)$$

y

$$E_{P_n}(L^*(n, Z_n)) = L^*(n) + E_{P_n}(C(n, Z_n)) \quad (9)$$

con

$$L^*(n) = E_{P_n}(\min_{\mathcal{D}} E_{P|Z_n}(L(d, \theta))). \quad (10)$$

En De Groot (1970), $L^*(n)$ recibe el nombre de *Riesgo de Bayes* que será el término con que se designará esta función de n a lo largo de este trabajo. $E_{P_n}(C(n, Z_n))$ es el costo esperado para una muestra de tamaño n .

En la práctica, es difícil especificar la función de costo en las mismas unidades de $L(d, \theta)$. En estos casos el uso directo del Riesgo de Bayes parece una alternativa razonable para evaluar el tamaño de muestra. En el siguiente capítulo se estudia el uso de $L^*(n)$ para evaluar el tamaño de muestra en el contexto del contraste de hipótesis planteado en el Capítulo I. A continuación un ejemplo.

Ejemplo II.3.1. (Continuación Ejemplo II.2.2.)

Suponga que además de estimar θ se desea evaluar el tamaño de muestra óptimo para efectuar la estimación. Se considera adecuado pensar que la función de costo sea proporcional al tamaño de muestra, es decir, $C(n, Z_n) = cn$ con $c > 0$.

Notar que

$$L(n, Z_n, d, \theta) = k(d - \theta)^2 + cn \quad (11)$$

expresión que también vale para $n = 0$.

Con una inicial Beta (α, β) ,

$$d^*_{Z_n} = \frac{\alpha + Y}{\alpha + \beta + n} \quad (12)$$

con $Y = \sum_{i=1}^n X_i$ y

$$L^*(n, Z_n) = \frac{k(\alpha + Y)(\alpha + n - Y)}{(\alpha + \beta + n + 1)(\alpha + \beta + n)^2} + cn. \quad (13)$$

Como Y dado θ tiene distribución Binomial (n, θ) y puesto que la esperanza de la esperanza condicional es la esperanza no condicional, resulta que

$$E_Y(Y) = E_p(n\theta) = n \frac{\alpha}{\alpha + \beta} \quad (14)$$

y

$$E_Y(Y^2) = E_P(n\theta - n\theta^2 + n^2\theta^2)$$

$$= \frac{n\alpha}{\alpha+\beta} - \frac{n\alpha(\alpha+1)}{(\alpha+\beta+1)(\alpha+\beta)} + \frac{n^2\alpha(\alpha+1)}{(\alpha+\beta+1)(\alpha+\beta)}, \quad (15)$$

con $E_Y(Y)$ y $E_Y(Y^2)$ iguales al primer y segundo momento de la f.d.p.g. $P(Y)$. De las últimas tres ecuaciones, se tiene que para toda $n \in \mathbb{N}$

$$L^*(n) = \frac{k\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)(\alpha+\beta+n)} \quad (16)$$

y

$$E_{P_n}(L^*(n, Z_n)) = L^*(n) + cn. \quad (17)$$

De esta forma, es posible encontrar explícitamente n_0 (el tamaño óptimo), extendiendo a toda n en los reales, derivando $E_{P_n}(L^*(n, Z_n))$ e igualando a cero.

Es fácil probar que

$$\frac{\partial E_{P_n}(L^*(n, Z_n))}{\partial n} = \frac{-k\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)(\alpha+\beta+n)^2} + c, \quad (18)$$

y que

$$\frac{\partial E_{P_n}(L^*(n, Z_n))}{\partial n} = 0 \quad (19)$$

cuando

$$n = \sqrt{\frac{k\alpha\beta}{c(\alpha+\beta)(\alpha+\beta+1)}} - (\alpha+\beta). \quad (20)$$

Si este punto crítico se denota por n_c , se buscan dos enteros consecutivos

entre los que se encuentre n_c y n_c es el entero, entre los dos consecutivos, con menor pérdida esperada. El lector puede verificar que n_c es un mínimo ya que la segunda derivada parcial de $E_{P_n}(L^*(n, Z_n))$ respecto de n es mayor que cero para todo número n .

Es interesante notar que en el ejemplo II.3.1, $L^*(n)$ es una función monótona decreciente en n y acotada inferiormente por cero. Normalmente $L^*(n)$ es monótona decreciente en n y tiene una cota inferior, el *Valor Esperado de la Información Perfecta* que se define a continuación.

DEFINICION II.3.1. Sea $(\mathcal{D}, \Theta, P, L)$ un problema de decisión. Sea Θ un espacio parametral con f.d.p.g. inicial $P(\theta)$. Si para toda $d \in \mathcal{D}$, $L(d, \theta)$ tiene un mínimo en θ y la función $L_M(d, \theta) = \min_{\mathcal{D}} L(d, \theta)$ es integrable respecto de la distribución asociada con $P(\theta)$, el *Valor Esperado de la Información Perfecta*, denotado por C , se define como

$$C = \int_{\Theta} L_M(d, \theta) dF_P(\theta). \tag{21}$$

La función $L_M(d, \theta)$ es la pérdida que se obtendría si el decisor supiera con certeza cuál es el verdadero valor del parámetro θ . Por esto, $L_M(d, \theta)$ es la pérdida por información perfecta y de ahí que a C se le denomine *Valor Esperado de la Información Perfecta*.

TEOREMA II.3.1. Sea $(\mathcal{D}, \Theta, P, L)$ un problema de decisión con Θ un espacio parametral. Si $C \in \mathbb{R}$, $L(d, \theta)$ no depende de Z_n y para todo $n \in \mathbb{N}$, existe $L^*(n)$, entonces

$$C \leq L^*(n). \tag{22}$$

Demostración.

Por la definición II.3.1,

$$L_M(d, \theta) \leq L(d, \theta), \quad \forall d \in \mathcal{D}. \quad (23)$$

Entonces,

$$E_{P|Z_n}(L_M(d, \theta)) \leq E_{P|Z_n}(L(d, \theta)), \quad \forall d \in \mathcal{D}. \quad (24)$$

En consecuencia,

$$E_{P|Z_n}(L_M(d, \theta)) \leq \min_{\mathcal{D}} E_{P|Z_n}(L(d, \theta)). \quad (25)$$

y por tanto,

$$E_{P_n}(E_{P|Z_n}(L_M(d, \theta))) \leq E_{P_n}(\min_{\mathcal{D}} E_{P|Z_n}(L(d, \theta))) = L^*(n). \quad (26)$$

Pero,

$$E_{P_n}(E_{P_n|Z_n}(L_M(d, \theta))) = E_P(E_{Z_n}(L_M(d, \theta))) \quad (27)$$

con $E_{Z_n}(L_M(d, \theta))$ igual al valor esperado de $L_M(d, \theta)$ respecto de $P(Z_n|\theta)$.

Como $L(d, \theta)$ no depende de Z_n ,

$$E_{Z_n}(L_M(d, \theta)) = L_M(d, \theta). \quad (28)$$

De las dos últimas ecuaciones, se desprende que

$$E_{P_n}(E_{P|Z_n}(L_M(d, \theta))) = C. \quad (29)$$

Con (26) y (29), se obtiene la conclusión del Teorema. ■

Bernardo y Smith (1994) presentan análogos a la Definición II.3.1 y al Teorema

II.3.1 en un contexto de Teoría de Decisión en general. Para los fines de este trabajo sólo es de interés la definición de información perfecta y la desigualdad del Teorema II.3.1, cuando Θ es un espacio parametral.

TEOREMA II.3.2. Bajo las condiciones del Teorema II.3.1 y bajo el supuesto de que Z_n es un vector de variables aleatorias independientes (v.a.i.) dado θ ,

$$L^*(n+1) \leq L^*(n) \quad (30)$$

para toda $n \in \mathbb{N}$.

Demostración. (Por claridad, únicamente se presenta la demostración para el caso θ continuo y Z_n un vector de variables aleatorias continuas.)

Por definición,

$$\begin{aligned} L^*(n) &= \int_{Z_n} P(Z_n) \left[\min_{\mathcal{D}} \int_{\Theta} L(d, \theta) P(\theta | Z_n) d\theta \right] dZ_n \\ &= \int_{Z_n} \left[\min_{\mathcal{D}} \int_{\Theta} L(d, \theta) P(\theta | Z_n) P(Z_n) d\theta \right] dZ_n. \end{aligned} \quad (31)$$

Como para todo $\theta \in \Theta$,

$$\int_{X_{n+1}} f(X_{n+1} | \theta) dX_{n+1} = 1, \quad (32)$$

$$\begin{aligned} L^*(n) &= \int_{Z_n} \left[\min_{\mathcal{D}} \int_{\Theta} L(d, \theta) P(\theta | Z_n) P(Z_n) \left(\int_{X_{n+1}} f(X_{n+1} | \theta) dX_{n+1} \right) d\theta \right] dZ_n \\ &= \int_{Z_n} \left[\min_{\mathcal{D}} \int_{\Theta} \int_{X_{n+1}} L(d, \theta) P(\theta | Z_n) P(Z_n) f(X_{n+1} | \theta) dX_{n+1} d\theta \right] dZ_n. \end{aligned} \quad (33)$$

Se sabe que

$$P(\theta | Z_n)P(Z_n) = f(Z_n | \theta)P(\theta) \quad (34)$$

y como $X_1, X_2, \dots, X_n, X_{n+1}$ son independientes dado θ , entonces

$$P(\theta | Z_n)P(Z_n)f(X_{n+1} | \theta) = f(Z_{n+1} | \theta)P(\theta) = P(\theta | Z_{n+1})P(Z_{n+1}). \quad (35)$$

De esta forma,

$$\begin{aligned} L^*(n) &= \int_{Z_n} \left[\min_{\mathcal{D}} \int_{\theta} \int_{X_{n+1}} L(d, \theta) P(\theta | Z_{n+1}) P(Z_{n+1}) dX_{n+1} d\theta \right] dZ_n. \\ &= \int_{Z_n} \left[\min_{\mathcal{D}} \int_{X_{n+1}} P(Z_{n+1}) \left[\int_{\theta} L(d, \theta) P(\theta | Z_{n+1}) d\theta \right] dX_{n+1} \right] dZ_n. \end{aligned} \quad (36)$$

Por otro lado, para toda $d \in \mathcal{D}$

$$\int_{\theta} L(d, \theta) P(\theta | Z_{n+1}) d\theta \geq \min_{\mathcal{D}} \int_{\theta} L(d, \theta) P(\theta | Z_{n+1}) d\theta. \quad (37)$$

Si esta desigualdad se multiplica por $P(Z_{n+1})$ y después se integra respecto de X_{n+1} , se obtiene que

$$\begin{aligned} \min_{\mathcal{D}} \int_{X_{n+1}} P(Z_{n+1}) \left[\int_{\theta} L(d, \theta) P(\theta | Z_{n+1}) d\theta \right] dX_{n+1} &\geq \\ \int_{X_{n+1}} P(Z_{n+1}) \left[\min_{\mathcal{D}} \int_{\theta} L(d, \theta) P(\theta | Z_{n+1}) d\theta \right] dX_{n+1}. \end{aligned} \quad (38)$$

Al integrar ambos lados de la última desigualdad obtenida respecto de Z_n , se llega a que

$$\int_{Z_n} \left[\min_{\mathcal{D}} \int_{X_{n+1}} P(Z_{n+1}) \left[\int_{\Theta} L(d, \theta) P(\theta | Z_{n+1}) d\theta \right] dX_{n+1} \right] dZ_n =$$

$$\int_{Z_n} \int_{X_{n+1}} P(Z_{n+1}) \left[\min_{\mathcal{D}} \int_{\Theta} L(d, \theta) P(\theta | Z_{n+1}) d\theta \right] dX_{n+1} dZ_n. \quad (39)$$

Notar que el lado derecho de (39) es igual a $L^*(n+1)$. Con (36) y (39) se obtiene la conclusión del teorema. ■

La importancia de los teoremas II.3.1 y II.3.2 será notoria en el siguiente capítulo, donde se propone un procedimiento para evaluar el tamaño de muestra usando el Riesgo de Bayes $L^*(n)$.

CAPITULO III

PLANTEAMIENTO BAYESIANO DEL PROBLEMA ORIGINAL

INTRODUCCION

El problema de contraste de tres hipótesis, expuesto en el capítulo I, se formula como problema de decisión. Se proponen distribuciones iniciales y funciones de pérdida útiles para efectuar el contraste. Para las iniciales, se calculan las correspondientes distribuciones finales y para las pérdidas se obtienen expresiones para la pérdida esperada con algunas de las iniciales y finales propuestas. Se define un método de evaluación de tamaño de muestra a través del Riesgo de Bayes y se implementa al problema de contraste de tres hipótesis.

III.1.-EL CONTRASTE DE HIPOTESIS

Recordemos que el problema de clasificación de semillas planteado en el Capítulo I se convierte en el de contrastar el juego de hipótesis

$$H_i: \theta \in \Theta_i \quad i=1,2,3 \quad (1)$$

en donde $\Theta_1 = [0, \theta_1]$, $\Theta_2 = [\theta_1, \theta_2]$ y $\Theta_3 = [\theta_2, 1]$, constituyen una partición del espacio parametral $\theta = [0, 1]$ y θ es el parámetro de una distribución Bernoulli. Las observaciones del vector Z_n conforman una muestra de variables aleatorias independientes e idénticamente distribuidas según el modelo Bernoulli, a menos que se indique otra cosa. Con esto,

$$f(Z_n | \theta) = \theta^Y (1-\theta)^{n-Y} \prod_{i=1}^n \binom{1}{(0,1)} (X_i) \quad (2)$$

$$y \text{ y } Y = \sum_{i=1}^n X_i.$$

Por lo discutido en el capítulo II, éste problema de contraste de hipótesis se puede formular como uno de decisión en donde $\mathcal{D} = \{d_1, d_2, d_3\}$, con d_1 la decisión de que H_1 es cierta, d_2 la decisión de que H_2 es cierta y d_3 la decisión de que H_3 es cierta. θ está representado por el intervalo $[0,1]$ y la pérdida, en general, es de la forma $L(d_i, \theta)$ con $i = 1, 2, 3$ y $\theta \in [0, 1]$.

Con la inicial $P(\theta)$, la solución al problema consiste en evaluar

$$L_i = E_P(L(d_i, \theta)) \quad i = 1, 2, 3 \quad (3)$$

y elegir d_j si ocurre que

$$L_j \leq L_i \quad i = 1, 2, 3. \quad (4)$$

Disponible la muestra Z_n , la solución al problema consiste en calcular $P(\theta|Z_n)$, evaluar

$$L_i^* = E_{P|Z_n}(L(d_i, \theta)) \quad i = 1, 2, 3 \quad (5)$$

y elegir d_j si ocurre que

$$L_j^* \leq L_i^* \quad i = 1, 2, 3. \quad (6)$$

III.2.-DISTRIBUCIONES INICIALES Y FINALES

Cuando en un problema de decisión estadístico el espacio de sucesos inciertos está representado por el intervalo $[0,1]$ y las observaciones son una muestra aleatoria de variables Bernoulli es común cuantificar la incertidumbre a través de un miembro de la familia \mathcal{F} con

$$\mathcal{F} = \{\Pi(\theta); \Pi(\theta) = \text{Be}(\alpha, \beta)\}. \quad (1)$$

La razón de esto es que \mathcal{F} ofrece varias posibilidades para expresar el conocimiento inicial. Como se demostró en el ejemplo II.2.1, \mathcal{F} es una familia paramétrica conjugada.

La familia de las Betas genera tres modelos que se pueden considerar dentro de la clase de distribuciones iniciales *no informativas* o de *referencia* (Bernardo (1979)). La distribución uniforme en el $[0,1]$ definida por

$$P(\theta) = \begin{cases} 1 & (\theta) \\ 0,1 \end{cases} \quad (2)$$

la distribución de Jeffreys definida por la ecuación

$$P(\theta) = \text{Be}\left(-\frac{1}{2}, -\frac{1}{2}\right) \quad (3)$$

y la distribución de referencia *Límite de Conjugadas* definida por la expresión

$$P(\theta) \propto \theta^{-1}(1-\theta)^{-1}. \quad (4)$$

La uniforme resulta ser la densidad que maximiza la entropía sobre el conjunto de todas las funciones de densidad en el $[0,1]$ con distribución continua (Bernardo (1979)) y aunque satisface el Principio de Indiferencia no es invariante ante reparametrizaciones (Box y Tiao (1973)). Por el ejemplo II.2.1, la final, cuando se

usa una inicial uniforme, resulta ser una Beta de parámetros $1 + Y$ y $n - Y + 1$.

La distribución de Jeffreys es la que maximiza la información esperada faltante y es invariante ante reparametrizaciones (Bernardo (1979)). Nuevamente, por el ejemplo II.2.1, la final, cuando se usa la inicial de Jeffreys es una Beta de parámetros $\frac{1}{2} + Y$ y $n - Y + \frac{1}{2}$.

En general, la distribución de referencia límite de conjugadas se determina buscando aquella inicial tal que los parámetros de su correspondiente final sólo dependan del tamaño de muestra y de los datos (De Groot (1970)). Con una inicial Beta (α, β) y con el supuesto de que las observaciones son Bernoulli, esto equivale a hacer que los valores de α y β sean iguales a cero. De aquí que la inicial límite de conjugadas en el problema Bernoulli-Beta sea $\theta^{-1}(1-\theta)^{-1}$. Es conveniente notar que dicha inicial no define una f.d.p.g. ya que

$$\int_0^1 \theta^{-1}(1-\theta)^{-1} d\theta = \infty. \quad (5)$$

Es decir, la distribución inicial no informativa límite de conjugadas, resulta ser una distribución impropia. Por simple sustitución en el Teorema de Bayes, la final que se obtiene al usar la inicial límite de conjugadas es una Beta de parámetros Y y $n - Y$ que también puede ser impropia. Cuando se utiliza una inicial impropia, ésta se considera como una aproximación a una situación real en la que se tiene poca información inicial.

Alternativamente, para el problema de prueba de hipótesis planteado se puede, definir la inicial de la siguiente manera.

Sea $\phi(\theta) = i$ si y solo si $\theta \in \Theta_i$, con $i = 1, 2, 3$. Para expresar el conocimiento inicial sobre ϕ es necesario determinar Π_1, Π_2, Π_3 , con $\Pi_1 + \Pi_2 + \Pi_3 = 1$ tales que

$$P(\phi = i) = \Pi_i, \quad i = 1, 2, 3 \quad (6)$$

Por Bayes, la probabilidad final de que $\phi = i$ se obtiene por la relación

$$P(\phi = i | Z_n) \propto f(Z_n | \phi = i)P(\phi = i), \quad i = 1, 2, 3 \quad (7)$$

donde

$$f(Z_n | \phi = i) = \int_{\Theta_i} f(Z_n | \theta)P(\theta | \phi = i)d\theta. \quad (8)$$

Esta última integral requiere de la especificación de $P(\theta | \phi = i)$, la f.d.p.g. inicial de θ dado $\phi = i$, $i = 1, 2, 3$. Dadas $P(\phi = i)$ y $P(\theta | \phi = i)$ resulta que

$$P(\theta) = \sum_{i=1}^3 P(\theta | \phi = i)P(\phi = i). \quad (9)$$

Mendoza (1993) propone definir

$$P(\theta | \phi = i) \propto \theta^{-r}(1-\theta)^{-r} I_{\Theta_i}(\theta), \quad r \geq 0, \quad i = 1, 2, 3 \quad (10)$$

que da lugar a la inicial

$$P(\theta) = \frac{\pi_1}{C_1} \theta^{-r}(1-\theta)^{-r} I_{\Theta_1}(\theta) + \frac{\pi_2}{C_2} \theta^{-r}(1-\theta)^{-r} I_{\Theta_2}(\theta) + \frac{\pi_3}{C_3} \theta^{-r}(1-\theta)^{-r} I_{\Theta_3}(\theta), \quad (11)$$

donde

$$C_i = \int_{\Theta_i} \theta^{-r}(1-\theta)^{-r} d\theta. \quad (12)$$

Con $r = 0$

$$P(\theta) = \frac{\pi_1}{\theta_1} I_{\Theta_1}(\theta) + \frac{\pi_2}{(\theta_2 - \theta_1)} I_{\Theta_2}(\theta) + \frac{\pi_3}{(1 - \theta_2)} I_{\Theta_3}(\theta). \quad (13)$$

Esta función de densidad es propia, se muestra graficamente en la figura III.2.1.

($\pi_1 = \pi_2 = \pi_3 = \frac{1}{3}$, $\theta_1 = .4$, $\theta_2 = .75$) y se denotará por $U(\pi_1, \pi_2, \pi_3)$.

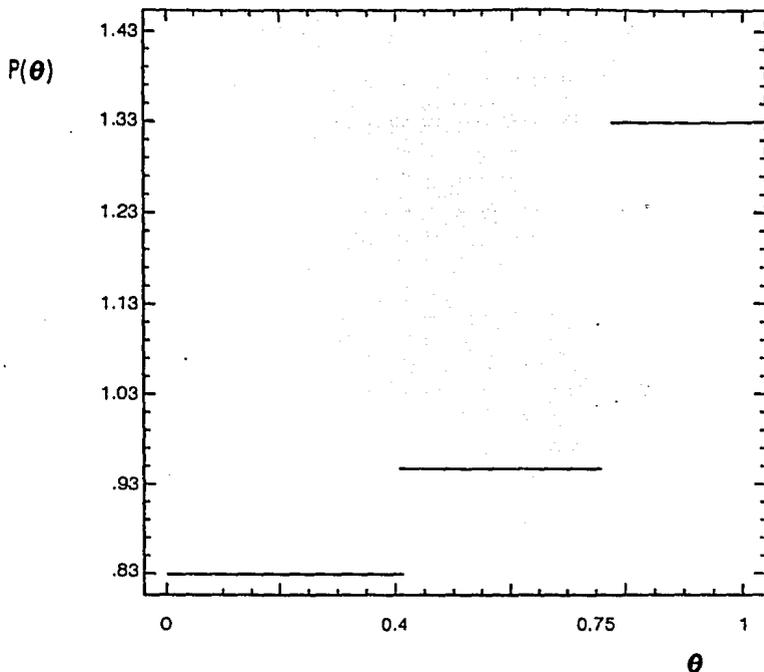


FIGURA III.2.1

INICIAL $U(\pi_1, \pi_2, \pi_3)$, $\pi_1 = \pi_2 = \pi_3 = \frac{1}{3}$, $\theta_1 = .4$ y $\theta_2 = .75$

Si $r = \frac{1}{2}$ la densidad aparece en la figura III.2.2 ($\pi_1 = \pi_2 = \pi_3 = \frac{1}{3}$, $\theta_1 = .4$, $\theta_2 = .75$).

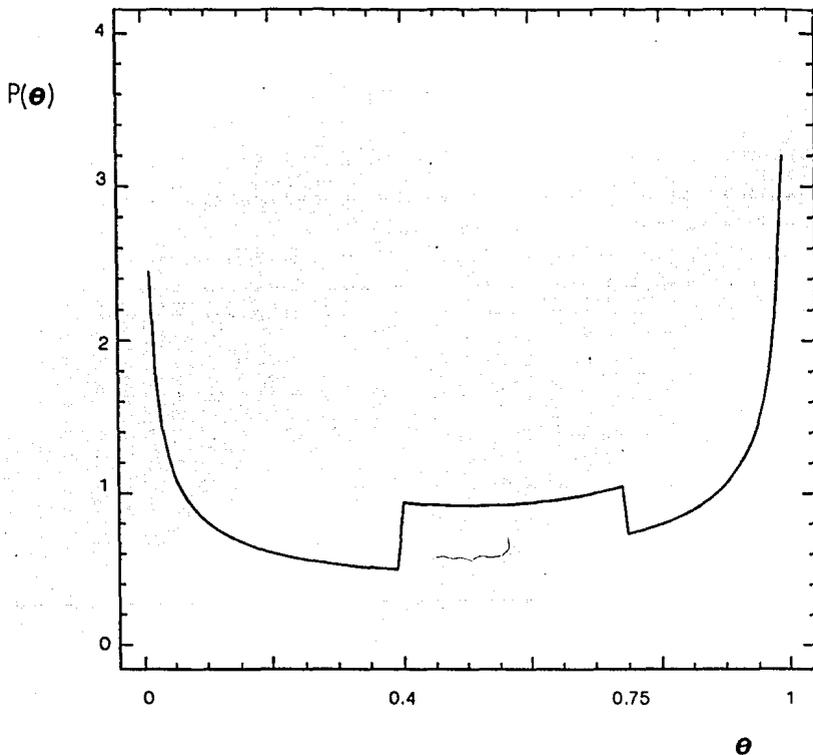


FIGURA III.2.1

INICIAL (11), $r = \frac{1}{2}$, $\pi_1 = \pi_2 = \pi_3 = \frac{1}{3}$, $\theta_1 = .4$, $\theta_2 = .75$

Si $P(\theta)$ es la inicial de (11) y Z_n es una muestra aleatoria de observaciones Bernoulli (θ),

$$P(\theta | Z_n) \propto \frac{\Pi_1}{C_1} \theta^{-r+Y} (1-\theta)^{-r+n-Y} I_{\theta_1}(\theta) + \frac{\Pi_2}{C_2} \theta^{-r+Y} (1-\theta)^{-r+n-Y} I_{\theta_2}(\theta) + \frac{\Pi_3}{C_3} \theta^{-r+Y} (1-\theta)^{-r+n-Y} I_{\theta_3}(\theta), \quad r \geq 0, Y = \sum_{i=1}^n X_i \quad (14)$$

Cuando $P(\theta) = U(\Pi_1, \Pi_2, \Pi_3)$ ($r = 0$),

$$P(\theta | Z_n) \propto \frac{\Pi_1}{\theta_1} \text{Be}(Y+1, n-Y+1) I_{\theta_1}(\theta) + \frac{\Pi_2}{(\theta_2 - \theta_1)} \text{Be}(Y+1, n-Y+1) I_{\theta_2}(\theta) + \frac{\Pi_3}{(1-\theta_2)} \text{Be}(Y+1, n-Y+1) I_{\theta_3}(\theta). \quad (15)$$

Si

$$\mu = \frac{\Pi_1}{\theta_1} (B_1(Y+1, n-Y+1)) + \frac{\Pi_2}{(\theta_2 - \theta_1)} (B_2(Y+1, n-Y+1) - B_1(Y+1, n-Y+1)) + \frac{\Pi_3}{(1-\theta_2)} (1 - B_2(Y+1, n-Y+1)), \quad (16)$$

con $B_i(\alpha, \beta) = \int_0^{\theta_i} \text{Be}(\alpha, \beta) d\theta$, $i = 1, 2$ y se define, a $M(\theta)$ como

$$M(\theta) = \frac{1}{\mu} \left[\frac{\Pi_1}{\theta_1} I_{\theta_1}(\theta) + \frac{\Pi_2}{(\theta_2 - \theta_1)} I_{\theta_2}(\theta) + \frac{\Pi_3}{(1-\theta_2)} I_{\theta_3}(\theta) \right], \quad (17)$$

la final que corresponde a la inicial $U(\Pi_1, \Pi_2, \Pi_3)$ es

$$P(\theta | Z_n) = M(\theta) \text{Be}(Y+1, n-Y+1). \quad (18)$$

Gráficamente, esta función de densidad se presenta en la figura III.2.3 para $n = 20$, $\theta_1 = .4$, $\theta_2 = .75$, $Y = 4$ y $\pi_1 = \pi_2 = \pi_3 = \frac{1}{3}$.

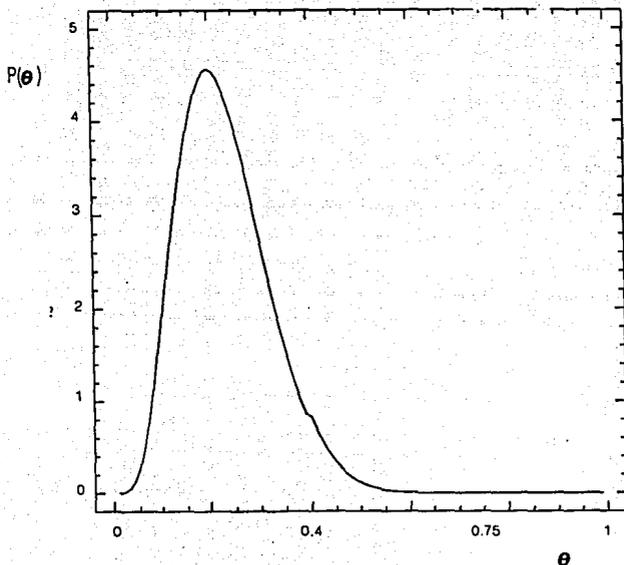


FIGURA III.2.3
FINAL PARA LA INICIAL $U(\pi_1, \pi_2, \pi_3)$

$$\theta_1 = .4, \theta_2 = .75, \pi_1 = \pi_2 = \pi_3 = \frac{1}{3}, n = 20, Y = 4$$

La ecuación (II) define una familia de iniciales útil para el problema de contraste de hipótesis discutido. Si $\pi_1 = \pi_2 = \pi_3 = \frac{1}{3}$, la inicial mencionada resulta ser una inicial no informativa sobre el parámetro ϕ .

Esta sección presentó alternativas para modelar el conocimiento inicial sobre θ , el parámetro de una variable aleatoria Bernoulli cuando se desea resolver el problema de contraste definido en la sección anterior. Las iniciales propuestas pueden modelar la incertidumbre de muchos decisores aunque estas distribuciones no son exhaustivas, es decir, no se afirma que todo decisor sea capaz de expresar su conocimiento inicial con alguna $P(\theta)$ aquí expuesta.

III.3.- FUNCIONES DE PERDIDA

Algunas posibles funciones de pérdida para el problema de contraste son:

Pérdida 0-K. Esta pérdida se define como

$$L(d_i, \theta) = k_i(1 - I(\theta)), \quad k_i > 0, \quad i = 1, 2, 3 \quad (1)$$

y es del tipo "todo o nada" ya que un acierto al decidirse por d_i produce pérdida cero y un error, pérdida k_i .

Gráficamente, para $k_i = 1$, $i = 1, 2, 3$, $\theta_1 = .4$ y $\theta_2 = .75$, la pérdida aparece en la figura III.3.1.

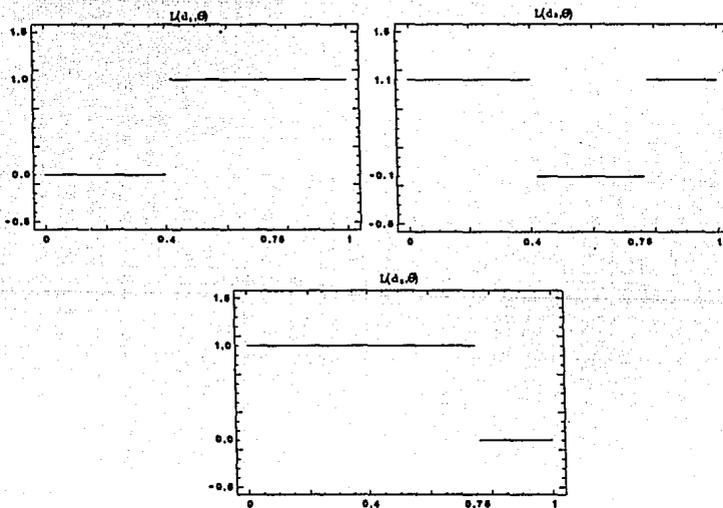


FIGURA III.3.1

PERDIDA 0-K, $K_1 = K_2 = K_3 = 1$, $\theta_1 = .4$, $\theta_2 = .75$

Si $p(\theta)$ es cualquier f.d.p.g sobre $[0,1]$ y se utiliza la pérdida O-K resulta que

$$E_p(L(d_i, \theta)) = k_i (1 - \Pr(\theta \in \Theta_i)), \quad i = 1, 2, 3 \quad (2)$$

donde $\Pr(\theta \in \Theta_i)$ es la probabilidad de que θ pertenezca a Θ_i , calculada con $p(\theta)$. Se concluye que la solución por pérdida esperada mínima se obtiene al determinar d_i que maximice $\Pr(\theta \in \Theta_i)$.

Cuando $P(\theta) = Be(\alpha, \beta)$,

$$E_p(L(d_1, \theta)) = k_1(1 - B_1(\alpha, \beta)), \quad (3)$$

$$E_p(L(d_2, \theta)) = k_2(1 - (B_2(\alpha, \beta) - B_1(\alpha, \beta))) \quad (4)$$

y

$$E_p(L(d_3, \theta)) = k_3(B_2(\alpha, \beta)), \quad (5)$$

con $B_1(\alpha, \beta)$ y $B_2(\alpha, \beta)$ definidas como antes.

Si $P(\theta|Z_n)$ es la final que corresponde a una inicial $Be(\alpha, \beta)$, $E_{p|Z_n}(L(d_i, \theta))$, $i = 1, 2, 3$, se obtiene con las tres ecuaciones anteriores sustituyendo α y β por $\alpha + Y$

y $\beta + n - Y$ respectivamente y $Y = \sum_{i=1}^n X_i$.

Cuando se usa como inicial la distribución límite de conjugadas no tiene sentido evaluar la pérdida esperada sin datos. Si los parámetros que corresponden a la final de la distribución límite son positivos, la pérdida esperada se obtiene sustituyendo en las ecuaciones (3)-(5) α y β por Y y $n - Y$.

Si $P(\theta) = U(\Pi_1, \Pi_2, \Pi_3)$, que resulta ser la inicial que distribuye uniformemente la probabilidad Π_i sobre Θ_i , entonces

$$E_p(L(d_i, \theta)) = k_i(1 - \Pi_i), \quad i = 1, 2, 3. \quad (6)$$

Cuando $P(\theta|Z_n)$ es la final que corresponde a la inicial $U(\pi_1, \pi_2, \pi_3)$ se puede verificar que

$$E_{P|Z_n}(L(d_1, \theta)) = k_1 \left(1 - \frac{\pi_1}{\mu \theta_1} B_1(Y+1, n-Y+1) \right), \quad (7)$$

$$E_{P|Z_n}(L(d_2, \theta)) = k_2 \left(1 - \left(\frac{\pi_2}{\mu(\theta_2 - \theta_1)} \right) (B_2(Y+1, n-Y+1) - B_1(Y+1, n-Y+1)) \right), \quad (8)$$

$$E_{P|Z_n}(L(d_3, \theta)) = k_3 \left(1 - \left(\frac{\pi_3}{\mu(1 - \theta_2)} \right) (1 - B_2(Y+1, n-Y+1)) \right), \quad (9)$$

con μ la probabilidad ponderada definida en la sección anterior.

Pérdida Lineal. Esta pérdida se define como

$$L(d_1, \theta) = k_{11}(\theta - \theta_1) \left(1 - I_{\theta_1}(\theta) \right), \quad (10)$$

$$L(d_2, \theta) = k_{21}(\theta_1 - \theta) I_{\theta_1}(\theta) + k_{22}(\theta - \theta_2) I_{\theta_3}(\theta), \quad (11)$$

$$L(d_3, \theta) = k_{31}(\theta_2 - \theta) \left(1 - I_{\theta_3}(\theta) \right). \quad (12)$$

Gráficamente para $k_{11} = 1$, $\theta_1 = .4$ y $\theta_2 = .75$, la pérdida aparece en la figura III.3.2.

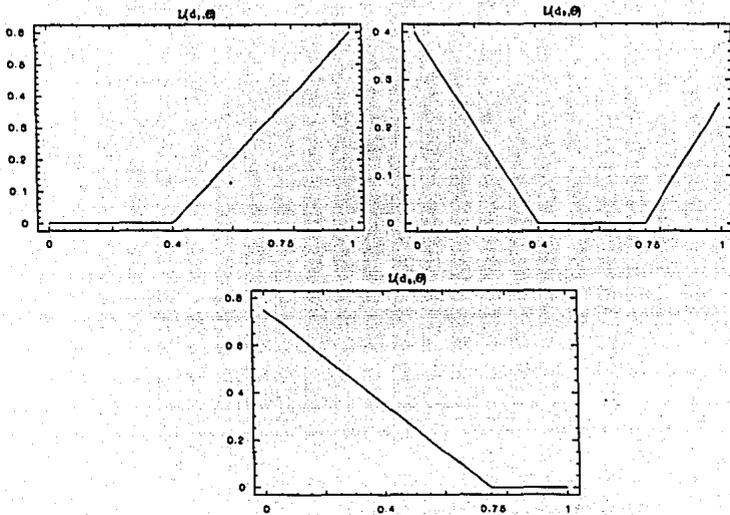


FIGURA III.3.2
 PERDIDA LINEAL $k_{ij} = 1$, $\theta_1 = .4$, $\theta_2 = .75$

Si $p(\theta)$ es cualquier f.d.p.g de probabilidad sobre el $[0,1]$ y se utiliza la pérdida lineal, entonces

$$E_p(L(d_1, \theta)) = k_{11} \left[E_p(\theta) - \int_0^{\theta_1} p(\theta) d\theta - \theta_1 (1 - \Pr(\theta \in \Theta_1)) \right] \quad (13)$$

$$E_p(L(d_2, \theta)) = k_{21} \left[\theta_1 \Pr(\theta \in \Theta_1) - \int_0^{\theta_1} p(\theta) d\theta \right] + k_{22} \left[\int_{\theta_2}^1 p(\theta) d\theta - \theta_2 \Pr(\theta \in \Theta_3) \right] \quad (14)$$

$$E_p(L(d_3, \theta)) = k_{31} \left[\theta_2 (1 - \Pr(\theta \in \Theta_3)) - E_p(\theta) + \int_{\theta_2}^1 \theta p(\theta) d\theta \right]. \quad (15)$$

En particular, cuando $P(\theta) = Be(\alpha, \beta)$,

$$E_p(L(d_1, \theta)) = k_{11} \left(\frac{\alpha}{\alpha + \beta} - \frac{\alpha(B_1(\alpha+1, \beta))}{\alpha + \beta} - \theta_1 (1 - B_1(\alpha, \beta)) \right), \quad (16)$$

$$E_p(L(d_2, \theta)) = k_{21} \left(\theta_1 B_1(\alpha, \beta) - \frac{\alpha(B_1(\alpha+1, \beta))}{\alpha + \beta} \right) + k_{22} \left(\frac{\alpha}{\alpha + \beta} (1 - B_2(\alpha+1, \beta)) - \theta_2 (1 - B_2(\alpha, \beta)) \right) \quad (17)$$

$$E_p(L(d_3, \theta)) = k_{31} \left(\theta_2 (B_2(\alpha, \beta)) - \frac{\alpha}{\alpha + \beta} + \frac{\alpha}{\alpha + \beta} (1 - B_2(\alpha+1, \beta)) \right) \quad (18)$$

en donde basta observar que $\int_a^b \theta Be(\alpha, \beta) d\theta = \int_a^b \frac{\alpha}{\alpha + \beta} Be(\alpha+1, \beta) d\theta$.

Nuevamente, si $P(\theta|Z_n)$ es la final que corresponde a una inicial Beta (α, β) , $E_{p|Z_n}(L(d_i, \theta))$, $i=1,2,3$, se obtiene con las tres últimas ecuaciones al sustituir α y β por $\alpha + Y$ y $\beta + n - Y$. Con la final correspondiente a la distribución inicial límite de conjugadas, la pérdida esperada se obtiene al cambiar α y β en las mismas tres ecuaciones (16-18) por Y y $n - Y$, siempre que estos parámetros sean mayores a cero.

Si $P(\theta) = U(\Pi_1, \Pi_2, \Pi_3)$ entonces

$$E_p(L(d_1, \theta)) = k_{11} \left[\frac{1}{2} ((2\Pi_1 + \Pi_2 - 2)\theta_1 + (\Pi_2 + \Pi_3)\theta_2 + \Pi_3) \right] \quad (19)$$

$$E_p(L(d_2, \theta)) = k_{21} \left[\frac{\theta_1 \Pi_1}{2} \right] + k_{22} \left[\frac{\Pi_3 (1 - \theta_2)}{2} \right] \quad (20)$$

$$E_p(L(d_3, \theta)) = k_{31} \left[\left(\frac{2\Pi_1 + \Pi_2}{2} \right) \theta_2 - \left(\frac{\Pi_1 + \Pi_2}{2} \right) \theta_1 \right]. \quad (21)$$

Cuando $P(\theta|Z_n)$ es la final que corresponde a la inicial $U(\Pi_1, \Pi_2, \Pi_3)$ y μ definida anteriormente, se tiene que

$$E_{P|Z_n}(L(d_1, \theta)) = k_{11} \left[\frac{Y+1}{\mu(n+2)} \left[\frac{\Pi_2}{(\theta_2 - \theta_1)} (B_2(Y+2, n-Y+1) - B_1(Y+2, n-Y+1)) \right. \right. \\ \left. \left. + \frac{\Pi_3}{(1-\theta_2)} (1 - B_2(Y+2, n-Y+1)) \right] - \theta_1 \left(1 - \frac{\Pi_1}{\mu\theta_1} (B_1(Y+1, n-Y+1)) \right) \right]. \quad (22)$$

$$E_{P|Z_n}(L(d_2, \theta)) = k_{21} \left[\frac{\Pi_1}{\mu} (B_1(Y+1, n-Y+1)) - \frac{\Pi_1(Y+1)}{\mu(\theta_1)(n+2)} (B_1(Y+2, n-Y+1)) \right] \\ + k_{22} \left[\frac{\Pi_3(Y+1)}{\mu(1-\theta_2)(n+2)} (1 - B_2(Y+2, n-Y+1)) \right. \\ \left. - \theta_2 \left(1 - \frac{1}{\mu} \right) \left[\left(\frac{\Pi_1\theta_2 - (\Pi_1 + \Pi_2)\theta_1}{\theta_1(\theta_2 - \theta_1)} \right) (B_1(Y+1, n-Y+1)) + \frac{\Pi_2}{(\theta_2 - \theta_1)} (B_2(Y+1, n-Y+1)) \right] \right]. \quad (23)$$

$$E_{P|Z_n}(L(d_3, \theta)) = k_{31} \left[\frac{\theta_2}{\mu} \left[\frac{\Pi_1\theta_2 - (\Pi_1 + \Pi_2)\theta_1}{\theta_1(\theta_2 - \theta_1)} (B_1(Y+1, n-Y+1)) + \frac{\Pi_2}{(\theta_2 - \theta_1)} (B_2(Y+1, n-Y+1)) \right] \right. \\ \left. - \frac{1}{\mu} \left[\frac{\Pi_2(Y+1)}{(\theta_2 - \theta_1)(n+2)} (B_2(Y+2, n-Y+1)) + \frac{(\Pi_1\theta_2 - (\Pi_1 + \Pi_2)\theta_1)(Y+1)}{\theta_1(\theta_2 - \theta_1)(n+2)} (B_1(Y+2, n-Y+1)) \right] \right]. \quad (24)$$

Pérdida Cuadrática. Esta pérdida se define como

$$L(d_1, \theta) = C_1(\theta - \theta_1)(\theta + \theta_1) I(\theta), \quad (25)$$

$$L(d_2, \theta) = C_2(\theta - \theta_1)(\theta - \theta_2) I(\theta), \quad (26)$$

$$L(d_3, \theta) = C_3(\theta - \theta_2)(\theta + \theta_2 - 2)I(\theta). \quad (27)$$

$$C_i > 0 \quad i = 1, 2, 3.$$

Para $\theta_1 = .4$, $\theta_2 = .75$, $C_i = 1$ con $i = 1, 2, 3$, la pérdida se muestra en la figura III.3.3.

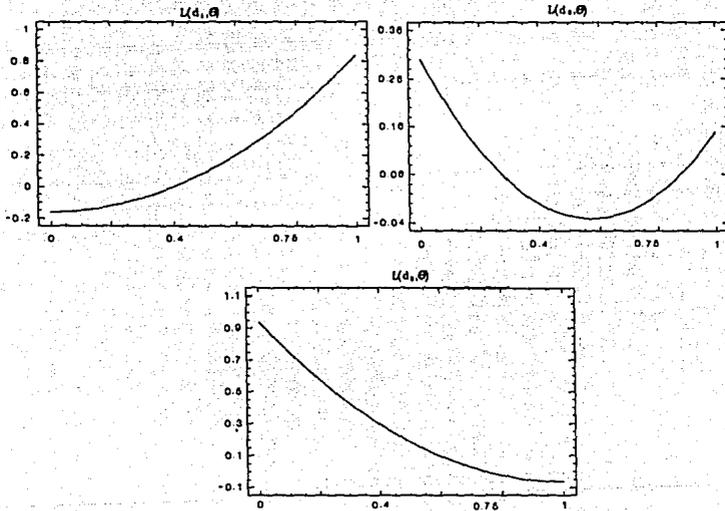


FIGURA III.3.3
 PERDIDA CUADRÁTICA, $C_1=C_2=C_3=1$, $\theta_1 = .4$, $\theta_2 = .75$

Si $p(\theta)$ es cualquier f.d.p.g. en el intervalo $[0,1]$ y se utiliza una pérdida cuadrática, entonces

$$E_p(L(d_1, \theta)) = C_1(E_p(\theta^2) - \theta_1^2), \quad (28)$$

$$E_p(L(d_2, \theta)) = C_2(E_p(\theta^2) - (\theta_1 + \theta_2)E_p(\theta) + \theta_1\theta_2), \quad (29)$$

$$E_p(L(d_3, \theta)) = C_3(E_p(\theta^2) - 2E_p(\theta) + 2\theta_2 - \theta_2^2). \quad (30)$$

De estas ecuaciones, se tiene que para determinar la pérdida esperada con Pérdida Cuadrática sólo es necesario calcular $E_p(\theta^2)$ y $E_p(\theta)$, primer y segundo momentos de $p(\theta)$, respectivamente.

Para evaluar la pérdida esperada cuando la inicial es $P(\theta) = Be(\alpha, \beta)$, basta recordar que

$$E_p(\theta) = \frac{\alpha}{\alpha + \beta} \quad (31)$$

y

$$E_p(\theta^2) = \frac{\alpha(\alpha+1)}{(\alpha+\beta+1)(\alpha+\beta)}. \quad (32)$$

Los primer y segundo momentos de la final cuando la inicial es Beta (α, β) se calculan con las dos últimas ecuaciones sustituyendo α por $\alpha + Y$ y β por $\beta + n - Y$.

Por otra parte, si la inicial $P(\theta) = U(\Pi_1, \Pi_2, \Pi_3)$,

$$E_p(\theta) = \frac{1}{2} \left[(\Pi_1 + \Pi_2)\theta_1 + (\Pi_1 + \Pi_3)\theta_2 + \Pi_3 \right] \quad (33)$$

y

$$E_p(\theta^2) = \frac{1}{3} \left[\frac{\Pi_2(\theta_2^3 - \theta_1^3)}{(\theta_2 - \theta_1)} + \frac{\Pi_3(1 - \theta_2^3)}{(1 - \theta_2)} + \Pi_1\theta_1^2 \right]. \quad (34)$$

Cuando $P(\theta|Z_n)$ es la final de una inicial $U(\Pi_1, \Pi_2, \Pi_3)$,

$$E_{P|Z_n}(\theta) = \frac{1}{\mu} \left[\frac{(\Pi_1\theta_2 - (\Pi_1 + \Pi_2)\theta_1)(Y+1)}{\theta_1(\theta_2 - \theta_1)(n+2)} B_1(Y+2, n-Y+1) + \right.$$

$$\left. \frac{(\pi_2 - (\pi_2 + \pi_3)\theta_2 + \pi_3\theta_1)(Y+1)}{(\theta_2 - \theta_1)(1 - \theta_2)(n+2)} (B_2(Y+2, n-Y+1)) + \frac{\pi_3(Y+1)}{(1 - \theta_2)(n+2)} \right] \quad (35)$$

y

$$E_{P|Z_n}(\theta^2) = \frac{1}{\mu} \left[\frac{(\pi_1\theta_2 - (\pi_1 + \pi_2)\theta_1)(Y+1)(Y+2)}{\theta_1(\theta_2 - \theta_1)(n+2)(n+3)} (B_1(Y+3, n-Y+1)) + \right. \\ \left. \frac{(\pi_2 - (\pi_2 + \pi_3)\theta_2 + \pi_3\theta_1)(Y+1)(Y+2)}{(\theta_2 - \theta_1)(1 - \theta_2)(n+2)(n+3)} (B_2(Y+3, n-Y+1)) + \frac{\pi_3(Y+1)(Y+2)}{(1 - \theta_2)(n+2)(n+3)} \right] \quad (36)$$

con μ , de nuevo, la probabilidad ponderada definida en la sección anterior

Divergencia Logarítmica Existe una cantidad importante de literatura (Shannon (1948), Lindley (1956), Lee (1964), Good (1966), Bernardo (1979), Gutiérrez-Peña (1992) y Rueda (1992) entre otros) que apoya el uso de la divergencia logarítmica como una medida de la discrepancia entre modelos de probabilidad. Si $p(x)$ y $q(x)$ son f.d.p.g con soporte común, entonces la divergencia logarítmica o discrepancia del modelo q respecto del modelo p está dada por

$$\delta(p, q) = \int p(x) \log \left(\frac{p(x)}{q(x)} \right) dx. \quad (37)$$

Si $f(X|\theta)$ es una densidad Bernoulli de parámetro desconocido θ y $\hat{\theta}$ es un valor particular de θ que satisface alguna restricción, entonces la divergencia logarítmica de $\hat{\theta}$ con respecto de θ es

$$\delta(\theta, \hat{\theta}) = \int_{x=0}^1 f(X|\theta) \log \left(\frac{f(X|\theta)}{f(X|\hat{\theta})} \right) \\ = \theta \log \left(\frac{\theta}{\hat{\theta}} \right) + (1-\theta) \log \left(\frac{1-\theta}{1-\hat{\theta}} \right). \quad (38)$$

Más aún, la discrepancia de un conjunto de valores de θ , θ_i , respecto del verdadero valor θ se puede evaluar con

$$\delta(\theta, \hat{\theta}_i) = \theta \log \left\{ \frac{\theta}{\hat{\theta}_i} \right\} + (1-\theta) \log \left\{ \frac{(1-\theta)}{(1-\hat{\theta}_i)} \right\} \quad (39)$$

donde $\hat{\theta}_i$ es un valor que se selecciona entre los valores que pertenecen a Θ_i .

Con esto, la pérdida basada en la divergencia logarítmica para el problema de contraste de tres hipótesis que aquí se trata, se define como

$$L(d_i, \theta) = A \delta(\theta, \hat{\theta}_i) + B, \quad i = 1, 2, 3. \quad A > 0 \quad (40)$$

Para un θ fijo, $\hat{\theta}_i$ puede ser seleccionado como el valor que minimiza $\delta(\theta, \hat{\theta}_i)$ en Θ_i . Sin embargo, θ es desconocido y entonces, la selección de $\hat{\theta}_i$ se debe formular como un problema de decisión bajo incertidumbre donde la pérdida puede ser la misma discrepancia entre θ y $\hat{\theta}_i$. Si $p(\theta)$ es una f.d.p.g. que describe el conocimiento sobre θ y de acuerdo al Principio de la Pérdida Esperada Mínima, $\hat{\theta}_i$ es tal que

$$D^*(\hat{\theta}_i) = \min_{\Theta_i} D^*(\hat{\theta}), \quad i = 1, 2, 3 \quad (41)$$

donde

$$\begin{aligned} D^*(\hat{\theta}) &= E_p(\delta(\theta, \hat{\theta})) \\ &= E_p(\theta \log \theta) - \log(\hat{\theta})E_p(\theta) + E_p((1-\theta) \log(1-\theta)) - \log(1-\hat{\theta})E_p(1-\theta). \end{aligned} \quad (42)$$

es la discrepancia esperada entre θ y $\hat{\theta}$ respecto de $p(\theta)$.

Es fácil verificar que el mínimo sin restricciones de $D^*(\hat{\theta})$ se alcanza en $\hat{\theta} = E_p(\theta)$, la media de $p(\theta)$. En consecuencia, y recordando la estructura de Θ_1 , Θ_2 y Θ_3 , tenemos que

$$\hat{\theta}_1 = \begin{cases} E_p(\theta) & \text{si } E_p(\theta) \leq \theta_1 \\ \theta_1 & \text{si } E_p(\theta) > \theta_1 \end{cases} \quad (43)$$

$$\hat{\theta}_2 = \begin{cases} \theta_1 & \text{si } E_p(\theta) < \theta_1 \\ E_p(\theta) & \text{si } \theta_1 \leq E_p(\theta) \leq \theta_2 \\ \theta_2 & \text{si } E_p(\theta) > \theta_2 \end{cases} \quad (44)$$

$$\hat{\theta}_3 = \begin{cases} \theta_2 & \text{si } E_p(\theta) < \theta_2 \\ E_p(\theta) & \text{si } E_p(\theta) \geq \theta_2 \end{cases} \quad (45)$$

Gráficamente con $A = 1$, $B = 0$, $\theta_1 = .4$, $\theta_2 = .75$ y $E_p(\theta) = .5$, la pérdida de la divergencia logarítmica aparece en la figura III.3.4.

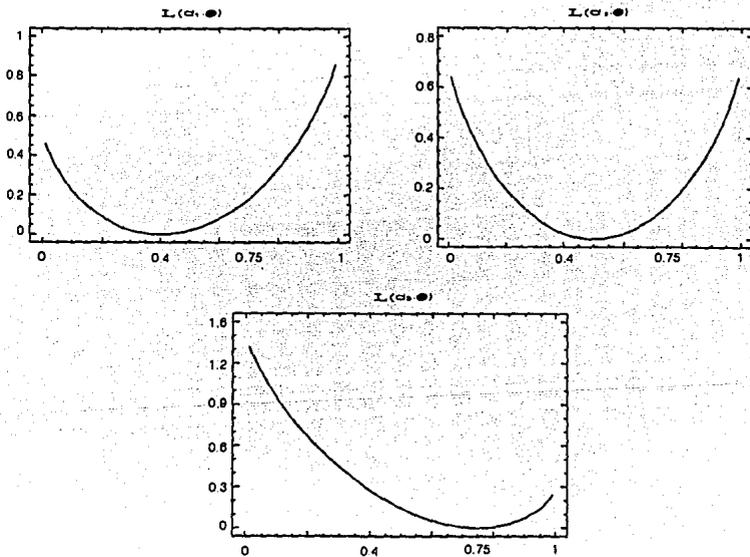


FIGURA III.3.4

PERDIDA DE LA DIVERGENCIA LOGARITMICA. $A = 1$, $B = 0$, $\theta_1 = .4$, $\theta_2 = .75$

Con todo esto, se puede concluir que si $p(\theta)$ es cualquier f.d.p.g. en el $[0,1]$,

$$E_p(L(d_i, \theta)) = A D^*(\hat{\theta}_i) + B, \quad i = 1, 2, 3, \quad (46)$$

$\hat{\theta}_i$ los representantes de clase definidos en las ecuaciones (43)-(45).

Cuando se utiliza la inicial $P(\theta) = Be(\alpha, \beta)$ y la pérdida de discrepancia logarítmica, entonces

$$E_p(L(d_i, \theta)) = A \left(\frac{\alpha}{\alpha + \beta} (\psi(\alpha + 1) - \log(\hat{\theta}_i)) + \frac{\beta}{\alpha + \beta} (\psi(\beta + 1) - \log(1 - \hat{\theta}_i)) - \psi(\alpha + \beta + 1) \right) + B, \quad i = 1, 2, 3 \quad (47)$$

donde $\psi(\cdot)$ es la función Dígamma.

La ecuación anterior se obtiene al recordar que con una Beta (α, β)

$$E_p(\theta \log(\theta)) = \frac{\alpha}{\alpha + \beta} (\psi(\alpha + 1) - \psi(\alpha + \beta + 1)), \quad (48)$$

$$E_p((1 - \theta) \log(1 - \theta)) = \frac{\beta}{\alpha + \beta} (\psi(\beta + 1) - \psi(\alpha + \beta + 1)) \quad (49)$$

y de sustituir estas expresiones en la pérdida esperada de (46).

Es posible dar una expresión equivalente a (47) a través del Teorema 2 de Gutierrez-Pefia (1992).

Para encontrar la decisión de Bayes en este caso, no es necesario evaluar la función Dígamma ya que

$$E_p(L(d_i, \theta) - L(d_{i+1}, \theta)) = A \left(\frac{\alpha}{\alpha + \beta} (\log(\hat{\theta}_{i+1}) - \log(\hat{\theta}_i)) + \frac{\beta}{\alpha + \beta} (\log(1 - \hat{\theta}_{i+1}) - \log(1 - \hat{\theta}_i)) \right), \quad i = 1, 2. \quad (50)$$

Si $P(\theta|Z_n)$ es la final que corresponde a una inicial Beta (α, β) , $E_{P|Z_n}(L(d_i, \theta))$ se calcula con la pérdida esperada (47) de la página anterior, sustituyendo α y β por $\alpha + Y$, $\beta + n - Y$, respectivamente. Notar que $\hat{\theta}$ también esta en función de los parámetros de la final. Cuando $P(\theta|Z_n)$ es la final de la inicial límite de conjugadas, $E_{P|Z_n}(L(d_i, \theta))$ se calcula con la misma ecuación (47) sustituyendo α y β por Y y $n - Y$.

De la expresión general de la pérdida esperada para la Divergencia Logarítmica (ecuación 47 de la página anterior), se puede ver que con cualquier $p(\theta)$ f.d.p.g que exprese el conocimiento del decisor, determinadas A , B , θ_1 y θ_2 , para evaluar $E_p(L(d_i, \theta))$, $i = 1, 2, 3$, sólo se necesita determinar $E_p(\theta)$, $E_p(\theta \log(\theta))$ y $E_p((1-\theta) \log(1-\theta))$. Para la inicial $U(\pi_1, \pi_2, \pi_3)$ que distribuye uniformemente la probabilidad π_i sobre θ_i , $E_p(\theta)$ es igual a $\frac{1}{2} \left[(\pi_1 + \pi_2)\theta_1 + (\pi_1 + \pi_3)\theta_2 + \pi_3 \right]$ y

$$E_p(\theta \log(\theta)) = \frac{\pi_1 \theta_1}{2} \left(\log(\theta_1) - \frac{1}{2} \right) + \frac{\pi_2}{2(\theta_2 - \theta_1)} \left(\theta_2^2 \left(\log(\theta_2) - \frac{1}{2} \right) \right) - \theta_1^2 \left(\log(\theta_1) - \frac{1}{2} \right) - \frac{\pi_3}{2(1 - \theta_2)} \left(\frac{1}{2} + \theta_2^2 \left(\log(\theta_2) - \frac{1}{2} \right) \right), \quad (51)$$

$$E_p((1-\theta) \log(1-\theta)) = - \frac{\pi_1}{2(\theta_1)} \left[(1-\theta_1)^2 \left(\log(1-\theta_1) - \frac{1}{2} \right) + \frac{1}{2} \right] + \frac{\pi_2}{2(\theta_2 - \theta_1)} \left[(1-\theta_1)^2 \left(\log(1-\theta_1) - \frac{1}{2} \right) - (1-\theta_2)^2 \left(\log(1-\theta_2) - \frac{1}{2} \right) \right] + \frac{\pi_3}{2} \left[(1 - \theta_2) \left(\log(1-\theta_2) - \frac{1}{2} \right) \right] \quad (52)$$

Para la final de una inicial $U(\pi_1, \pi_2, \pi_3)$, $E_{P|Z_n}(\theta)$ aparece en (35) pero $E_{P|Z_n}(\theta \log(\theta))$ y $E_{P|Z_n}((1-\theta) \log(1-\theta))$ se tienen que evaluar vía integración

numérica.

De esta manera, para implementar la solución al problema propuesto, utilizando cualquiera de las combinaciones de inicial (o final) y pérdida que se han presentado basta, en el peor de los casos, con recurrir a procedimientos numéricos para evaluar expresiones del tipo $E_p(\log\theta)$ lo mismo que Betas incompletas y Digamas. En particular, en este trabajo se utilizaron los algoritmos de Press et.al. (1989) y de Bernardo (1976) para evaluar las Betas y las Digamas respectivamente.

III.4.- TAMAÑO DE MUESTRA

Como se discutió en la tercera sección del capítulo II, el problema de seleccionar el tamaño de muestra (n) es a su vez un problema de decisión secuencial que requiere de la especificación de la pérdida $L(n, Z_n, d, \theta)$. Para el problema de clasificación de semillas que se ha abordado a lo largo de éste trabajo resulta difícil determinar $L(n, Z_n, d, \theta)$, aún cuando se suponga aditividad en la pérdida. La razón es que no resulta obvio como proponer una función $C(n, Z_n)$ en unidades compatibles con $L(d, \theta)$.

Por otro lado, con $L(d, \theta)$ es posible determinar el Riesgo de Bayes $L^*(n)$. Como se puede verificar en los teoremas II.3.1 y II.3.2 esta función de n es no creciente, está acotada inferiormente por el valor esperado de la información perfecta C (ver definición II.3.1) y

$$\lim_{n \rightarrow \infty} L^*(n) = C. \quad (1)$$

En otras palabras, C es el Riesgo de Bayes que corresponde a $n = \infty$.

Aquí, es conveniente observar que en la práctica C es finita, entonces es posible que para alguna n finita se obtenga un Riesgo de Bayes que si bien es mayor que C , en términos prácticos excede a esa constante por una fracción insignificante. De esta manera, una solución aproximada, pero razonable, consiste en elegir el valor de n más pequeño tal que

$$L^*(n) - C < \epsilon \quad (2)$$

con ϵ una constante positiva apropiada. Es decir, se pretende elegir la n que proporcione un Riesgo de Bayes prácticamente indistinguible con el riesgo que se obtendría con información perfecta.

Un camino para determinar ϵ consiste en seleccionar $t \in (0,1)$ y hacer $\epsilon = (L^*(0) - C)t$, de esta manera ϵ es una proporción del rango del Riesgo de Bayes. En principio y con fines ilustrativos se usaran valores de t como .01, .05 y .1.

$L^*(0) - C$ se puede interpretar como la disminución en pérdida que a un decisor le reporta la información perfecta cuando parte de un tamaño de muestra igual a cero (es decir cuando realiza el contraste sólo con la información inicial).

Para el problema de prueba de hipótesis planteado en este capítulo, las dos secciones anteriores presentan una variedad de distribuciones iniciales y funciones de pérdida para las cuales C es finita. De hecho, basta con que $p(\theta)$ sea propia y $L(d_i, \theta)$ sea acotada inferiormente en $[0, 1]$ ($i = 1, 2, 3$) para garantizar que $-\infty < C < \infty$. Hay que tener presente que en el contexto del mismo problema y dado que

$Y = \sum_{i=1}^n X_i$ es una estadística suficiente, cuando las observaciones son Bernoulli, se tiene que

$$E_{P|Z_n}(L(d_i, \theta)) = L_i^*(Y), \quad i = 1, 2, 3 \quad (3)$$

y

$$\min_{\mathcal{D}} E_{P|Z_n}(L(d, \theta)) = h(Y). \quad (4)$$

De esta forma,

$$L^*(n) = E_P(h(Y)) = \sum_{Y=0}^n h(Y)P(Y) \quad (5)$$

con

$$P(Y) = \int_0^1 f(Y|\theta)P(\theta) d\theta, \quad (6)$$

en donde $P(\theta)$ es la distribución inicial y $f(Y|\theta)$ la función de densidad Binomial de parámetros (n, θ) .

Se puede observar que con el fin de evaluar $L^*(n)$ es necesario determinar la f.d.p.g. $P(Y)$. A continuación se presentan expresiones para $P(Y)$ correspondientes a algunas de las iniciales que aparecen en la sección 2 de este capítulo.

Si $P(\theta) = Be(\alpha, \beta)$, entonces

$$\begin{aligned}
 P(Y) &= \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} \binom{n}{Y} \theta^Y(1-\theta)^{n-Y} d\theta \\
 &= \binom{n}{Y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+Y)}{\Gamma(\alpha)} \frac{\Gamma(\beta-Y)}{\Gamma(\beta)} \frac{1}{\Gamma(\alpha+\beta+n)}, \quad Y = 0, 1, \dots, n. \quad (7)
 \end{aligned}$$

Esta expresión define a la distribución Beta-Binomial de parámetros n , α y β de donde se desprende que si $P(\theta)$ es una uniforme $[0,1]$,

$$P(Y) = \frac{1}{n+1}, \quad Y = 0, 1, \dots, n. \quad (8)$$

También con (7) se puede verificar que si se usa la inicial de Jeffreys,

$$P(Y) = \frac{\Gamma(1/2 + Y)\Gamma(n + 1/2 - Y)}{\pi(n-Y)! Y!}, \quad Y = 0, 1, \dots, n. \quad (9)$$

Si $P(\theta) = U(\pi_1, \pi_2, \pi_3)$ entonces

$$\begin{aligned}
 P(Y) &= \int_0^{\frac{\pi_1}{\theta_1}} \frac{\pi_1}{\theta_1} \binom{n}{Y} \theta^Y(1-\theta)^{n-Y} d\theta + \int_{\frac{\pi_1}{\theta_2-\theta_1}}^{\frac{\pi_2}{\theta_2-\theta_1}} \frac{\pi_2}{(\theta_2-\theta_1)} \binom{n}{Y} \theta^Y(1-\theta)^{n-Y} d\theta + \int_{\frac{\pi_3}{1-\theta_2}}^1 \frac{\pi_3}{(1-\theta_2)} \binom{n}{Y} \theta^Y(1-\theta)^{n-Y} d\theta \\
 &= \frac{1}{n+1} \left[\left(\frac{\pi_1}{\theta_1} - \frac{\pi_2}{(\theta_2-\theta_1)} \right) (B_1(Y+1, n-Y+1)) + \right. \\
 &\quad \left. \left(\frac{\pi_2}{(\theta_2-\theta_1)} - \frac{\pi_3}{(1-\theta_2)} \right) (B_2(Y+1, n-Y+1)) + \frac{\pi_3}{(1-\theta_2)} \right]. \quad (10)
 \end{aligned}$$

Aunque se tienen expresiones cerradas para $P(Y)$ con las iniciales mencionadas no es posible, mediante el uso de las pérdidas de la sección anterior, determinar una ecuación para $L^*(n)$ que sea útil para encontrar un entero n tal $L^*(n) - C$ sea menor que ϵ . Sin embargo, se puede diseñar un algoritmo numérico que para cada valor de n , calcule los valores de Bayes y $P(Y)$ desde $Y = 0, 1, 2, \dots, n$ y que posteriormente evalúe

$L^*(n)$ a través de una suma. De esta forma, conocidos c y C , se puede obtener una solución numérica para la desigualdad $L^*(n) - C < \epsilon$.

Por otra parte, para evaluar el tamaño de muestra según lo discutido en esta sección es necesario calcular C , que como se puede ver de la Definición 11.3.1, depende de $L(d, \theta)$ y de la inicial $P(\theta)$. Enseguida se presentan expresiones para C con las pérdidas de la sección previa y algunas de las iniciales de la sección 2 de este capítulo. De hecho, si se usa Pérdida 0-k o Pérdida Lineal,

$$\min_{\mathcal{D}} L(d, \theta) = 0, \quad \forall \theta \in [0, 1], \quad (11)$$

en consecuencia $C = 0$ para toda inicial $P(\theta)$.

Con Pérdida Cuadrática,

$$\min_{\mathcal{D}} L(d, \theta) = C_1(\theta - \theta_1)(\theta + \theta_1)I_{\theta_1}(\theta) + C_2(\theta - \theta_1)(\theta - \theta_2)I_{\theta_2}(\theta) + C_3(\theta - \theta_2)(\theta + \theta_2 - 2)I_{\theta_3}(\theta). \quad (12)$$

Mediante esta ecuación es fácil verificar que si $P(\theta) = Be(\alpha, \beta)$ y se utiliza la Pérdida Cuadrática, entonces

$$\begin{aligned} C = & \frac{(C_1 - C_2)\alpha(\alpha + 1)}{(\alpha + \beta + 1)(\alpha + \beta)} (B_1(\alpha + 2, \beta)) - (C_1\theta_1^2 + C_2\theta_1\theta_2)(B_1(\alpha, \beta)) + \frac{(C_2 - C_3)\alpha(\alpha + 1)}{(\alpha + \beta + 1)(\alpha + \beta)} (B_2(\alpha + 2, \beta)) \\ & + \frac{(2C_3 - (\theta_1 + \theta_2)C_2)\alpha}{(\alpha + \beta)} (B_2(\alpha + 1, \beta)) + \frac{C_2(\theta_1 + \theta_2)\alpha}{(\alpha + \beta)} (B_1(\alpha + 1, \beta)) \\ & + (C_2(\theta_1\theta_2) - C_3(2\theta_2 - \theta_2^2))(B_2(\alpha, \beta)) + \frac{C_3\alpha(\alpha + 1)}{(\alpha + \beta + 1)(\alpha + \beta)} - \frac{2C_3\alpha}{(\alpha + \beta)} + C_3(2\theta_2 - \theta_2^2). \end{aligned} \quad (13)$$

Cuando $P(\theta) = U(\Pi_1, \Pi_2, \Pi_3)$, con la misma pérdida cuadrática se tiene que

$$C = \int_0^{\theta_1} C_1(\theta - \theta_1)(\theta + \theta_1) \frac{\Pi_1}{\theta_1} d\theta + \int_{\theta_1}^{\theta_2} \frac{\Pi_2}{(\theta_2 - \theta_1)} C_2(\theta - \theta_1)(\theta - \theta_2) d\theta + \int_{\theta_2}^1 \frac{\Pi_3}{(1 - \theta_2)} C_3(\theta - \theta_2)(\theta + \theta_2 - 2) d\theta$$

$$= -\frac{2C_1\pi_1\theta_1^2}{3} + \frac{C_2\pi_2}{(\theta_2-\theta_1)} \left(\frac{1}{6}(\theta_1^3-\theta_2^3) + \frac{1}{2}(\theta_1\theta_2(\theta_2-\theta_1)) \right) + \frac{C_3\pi_3}{(1-\theta_2)} \left(2\theta_2(1-\theta_2) - \frac{2}{3}(1-\theta_2^3) \right). \quad (14)$$

La teoría que apoya el uso del Riesgo de Bayes para evaluar el tamaño de muestra requiere que $L(d_i, \theta)$ no dependa de Z_n , $i=1,2,3$. Ahora bien, como se recordará para el caso de la divergencia logarítmica, $L(d_i, \theta) = L(\hat{\theta}_i, \theta)$, donde $\hat{\theta}_i$ se obtiene con $p(\theta)$ que describe el conocimiento sobre θ . En la propuesta original para seleccionar los representantes $\hat{\theta}_i$ ($i=1,2,3$) (Rueda (1992)) se emplea la distribución final para θ . En consecuencia, $\hat{\theta}_i$ depende de Z_n lo mismo que $L(d_i, \theta)$. Para resolver éste problema, en éste trabajo, cuando se determinan los representantes, la distribución que se utiliza será siempre la inicial. Para determinar expresiones del valor esperado de la información perfecta, con la pérdida de la divergencia logarítmica, es necesario considerar lo siguiente. Si $P(\theta)$ es la f.d.p.g inicial y se utiliza la pérdida de la divergencia, entonces

i) $\hat{\theta}_1 < \hat{\theta}_2 < \hat{\theta}_3$ implica que

$$\min_{\mathcal{D}} L(d, \theta) = A \left(\delta(\theta, \hat{\theta}_1) I_{\Omega_1}(\theta) + \delta(\theta, \hat{\theta}_2) I_{\Omega_2}(\theta) + \delta(\theta, \hat{\theta}_3) I_{\Omega_3}(\theta) \right) + B. \quad (15)$$

ii) $\hat{\theta}_1 = \hat{\theta}_2 < \hat{\theta}_3$ implica que

$$\min_{\mathcal{D}} L(d, \theta) = A \left(\delta(\theta, \hat{\theta}_2) I_{[0, \theta_2^*]}(\theta) + \delta(\theta, \hat{\theta}_3) I_{\{\theta_2^*, 1\}}(\theta) \right) + B. \quad (16)$$

iii) $\hat{\theta}_1 < \hat{\theta}_2 = \hat{\theta}_3$ implica que

$$\min_{\mathcal{D}} L(d, \theta) = A \left(\delta(\theta, \hat{\theta}_1) I_{[0, \theta_1^*]}(\theta) + \delta(\theta, \hat{\theta}_2) I_{\{\theta_1^*, 1\}}(\theta) \right) + B. \quad (17)$$

en donde $\Omega_1 = [0, \theta_1^*]$, $\Omega_2 = (\theta_1^*, \theta_2^*)$ y $\Omega_3 = [\theta_2^*, 1]$,

$$\theta_1^* = \frac{\log((1-\hat{\theta}_1)/(1-\hat{\theta}_2))}{\log((\hat{\theta}_2(1-\hat{\theta}_1))/(\hat{\theta}_1(1-\hat{\theta}_2)))} \quad (18)$$

y

$$\theta_2^* = \frac{\log((1-\hat{\theta}_2)/(1-\hat{\theta}_3))}{\log((\hat{\theta}_3(1-\hat{\theta}_2))/(\hat{\theta}_2(1-\hat{\theta}_3)))} \quad (19)$$

A continuación se presenta la demostración del caso 1). Los otros dos casos se prueban de manera similar.

Como

$$0 < \hat{\theta}_2 \log \left(\frac{\hat{\theta}_2}{\hat{\theta}_1} \right) + (1-\hat{\theta}_2) \log \left(\frac{(1-\hat{\theta}_2)}{(1-\hat{\theta}_1)} \right), \quad (20)$$

entonces

$$\log \left(\frac{(1-\hat{\theta}_1)}{(1-\hat{\theta}_2)} \right) < \hat{\theta}_2 \log \left(\frac{\hat{\theta}_2(1-\hat{\theta}_1)}{\hat{\theta}_1(1-\hat{\theta}_2)} \right). \quad (21)$$

Por tanto,

$$\theta_1^* < \hat{\theta}_2. \quad (22)$$

Análogamente y dado que

$$0 < \hat{\theta}_2 \log \left(\frac{\hat{\theta}_2}{\hat{\theta}_1} \right) + (1-\hat{\theta}_2) \log \left(\frac{(1-\hat{\theta}_2)}{(1-\hat{\theta}_1)} \right), \quad (23)$$

entonces

$$\hat{\theta}_2 < \theta_2^*. \quad (24)$$

En consecuencia,

$$\theta_1^* < \theta_2^* \quad (25)$$

Con argumentos similares es fácil ver también que $\hat{\theta}_1 < \theta_1^*$ y $\theta_2^* < \hat{\theta}_3$.

Por otro lado,

$$L(d_2, \theta) - L(d_1, \theta) = A \left(\log\left(\frac{1-\hat{\theta}_1}{1-\hat{\theta}_2}\right) + \left(\log\left(\frac{\hat{\theta}_1(1-\hat{\theta}_2)}{\hat{\theta}_2(1-\hat{\theta}_1)}\right) \right) \theta \right) \quad (26)$$

es una función decreciente en θ que se hace cero si $\theta = \theta_1^*$.

Entonces

$$L(d_1, \theta) \leq L(d_2, \theta), \quad \text{si } \theta \in [0, \theta_1^*] \quad (27)$$

y

$$L(d_2, \theta) < L(d_1, \theta), \quad \text{si } \theta \in (\theta_1^*, 1]. \quad (28)$$

De la misma manera, se puede probar que

$$L(d_2, \theta) \leq L(d_3, \theta), \quad \text{si } \theta \in [0, \theta_2^*] \quad (29)$$

y

$$L(d_3, \theta) < L(d_2, \theta), \quad \text{si } \theta \in (\theta_2^*, 1]. \quad (30)$$

Con (27)-(30) se obtiene la expresión de $\min_{\theta} L(d, \theta)$ para el caso i). El caso ii) considerado, aparece cuando $E(\theta) = \theta_1$. El caso iii) se presenta si $E(\theta) = \theta_2$. Si $E(\theta) = \theta_1$ las decisiones d_i y d_{i+1} son igualmente preferibles ya que $\delta(\theta, \hat{\theta}_i) = \delta(\theta, \hat{\theta}_{i+1})$, $i = 1, 2$. No puede ocurrir que $\hat{\theta}_1 = \hat{\theta}_2 = \hat{\theta}_3$ porque $\theta_1 < \theta_2$.

Suponiendo que $P(\theta) = Be(\alpha, \beta)$ y que se utiliza la pérdida de la divergencia logarítmica

i) Si $\hat{\theta}_1 < \hat{\theta}_2 < \hat{\theta}_3$ entonces

$$C = A \left[\log\left(\frac{1-\hat{\theta}_2}{1-\hat{\theta}_1}\right) B(\alpha, \beta, \theta_1^*) + \log\left(\frac{1-\hat{\theta}_3}{1-\hat{\theta}_2}\right) B(\alpha, \beta, \theta_2^*) \right]$$

$$\begin{aligned}
& + \left[\frac{\alpha}{\alpha+\beta} \right] \log \left(\hat{\theta}_2(1-\hat{\theta}_1) / (\hat{\theta}_1(1-\hat{\theta}_2)) \right) B(\alpha+1, \beta, \theta_1^*) \\
& + \left[\frac{\alpha}{\alpha+\beta} \right] \log \left(\hat{\theta}_3(1-\hat{\theta}_2) / (\hat{\theta}_2(1-\hat{\theta}_3)) \right) B(\alpha+1, \beta, \theta_2^*) \\
& - \log(1-\hat{\theta}_3) - \left[\frac{\alpha}{\alpha+\beta} \right] \log(\hat{\theta}_3 / (1-\hat{\theta}_3)) + \left[\frac{\beta}{\alpha+\beta} \right] \Psi(\beta) + \left[\frac{\alpha}{\alpha+\beta} \right] \Psi(\alpha+1) - \Psi(\alpha+\beta) \Big] + B
\end{aligned} \tag{31}$$

donde

$$B(\alpha, \beta, k) = \int_0^k \text{Beta}(\alpha, \beta) d\theta. \tag{32}$$

ii) Si $\hat{\theta}_1 = \hat{\theta}_2 < \hat{\theta}_3$ entonces

$$\begin{aligned}
C = A \Bigg[& \log \left((1-\hat{\theta}_3) / (1-\hat{\theta}_2) \right) B(\alpha, \beta, \theta_2^*) + \left[\frac{\alpha}{\alpha+\beta} \right] \log \left(\hat{\theta}_3(1-\hat{\theta}_2) / (\hat{\theta}_2(1-\hat{\theta}_3)) \right) B(\alpha+1, \beta, \theta_2^*) \\
& - \log(1-\hat{\theta}_3) - \left[\frac{\alpha}{\alpha+\beta} \right] \log(\hat{\theta}_3 / (1-\hat{\theta}_3)) + \left[\frac{\beta}{\alpha+\beta} \right] \Psi(\beta) + \left[\frac{\alpha}{\alpha+\beta} \right] \Psi(\alpha+1) \\
& - \Psi(\alpha+\beta) \Big] + B.
\end{aligned} \tag{33}$$

iii) Si $\hat{\theta}_1 < \hat{\theta}_2 = \hat{\theta}_3$ entonces

$$\begin{aligned}
C = A \Bigg[& \log \left((1-\hat{\theta}_2) / (1-\hat{\theta}_1) \right) B(\alpha, \beta, \theta_1^*) + \left[\frac{\alpha}{\alpha+\beta} \right] \log \left(\hat{\theta}_2(1-\hat{\theta}_1) / (\hat{\theta}_1(1-\hat{\theta}_2)) \right) B(\alpha+1, \beta, \theta_1^*) \\
& - \log(1-\hat{\theta}_2) - \left[\frac{\alpha}{\alpha+\beta} \right] \log(\hat{\theta}_2 / (1-\hat{\theta}_2)) + \left[\frac{\beta}{\alpha+\beta} \right] \Psi(\beta) + \left[\frac{\alpha}{\alpha+\beta} \right] \Psi(\alpha+1) \\
& - \Psi(\alpha+\beta) \Big] + B.
\end{aligned} \tag{34}$$

A continuación se verifica el caso 1). Los otros dos casos son análogos. La expresión de $\min_{\theta} L(d, \theta)$ que aparece en (15) lleva a que

$$C = A \left(\int_0^{\theta_1^*} \delta(\theta, \hat{\theta}_1) \text{Be}(\alpha, \beta) d\theta + \int_{\theta_1^*}^{\theta_2^*} \delta(\theta, \hat{\theta}_2) \text{Be}(\alpha, \beta) d\theta + \int_{\theta_2^*}^1 \delta(\theta, \hat{\theta}_3) \text{Be}(\alpha, \beta) d\theta \right) + B \quad (35)$$

Si en esta ecuación se sustituye $\delta(\theta, \hat{\theta}_i)$ y la densidad $\text{Beta}(\alpha, \beta)$, como

$$\int_a^b \delta(\theta, \hat{\theta}_i) \text{Be}(\alpha, \beta) d\theta = \int_0^b \delta(\theta, \hat{\theta}_i) \text{Be}(\alpha, \beta) d\theta - \int_0^a \delta(\theta, \hat{\theta}_i) \text{Be}(\alpha, \beta) d\theta \quad (36)$$

se obtiene

$$\begin{aligned} C = A & \left(-\log(1-\hat{\theta}_1) B(\alpha, \beta, \theta_1^*) - \log(1-\hat{\theta}_2) B(\alpha, \beta, \theta_2^*) + \log(1-\hat{\theta}_2) B(\alpha, \beta, \theta_1^*) \right. \\ & - \left(\frac{\alpha}{\alpha+\beta} \right) \log(\hat{\theta}_1/(1-\hat{\theta}_1)) B(\alpha+1, \beta, \theta_1^*) \\ & - \left(\frac{\alpha}{\alpha+\beta} \right) (\log(\hat{\theta}_2/(1-\hat{\theta}_2))) (B(\alpha+1, \beta, \theta_2^*) - B(\alpha+1, \beta, \theta_1^*)) \\ & - \log(1-\hat{\theta}_3) (1 - B(\alpha, \beta, \theta_2^*)) + E((1-\theta)\log(1-\theta)) + E(\theta \log(\theta)) \\ & \left. - \left(\frac{\alpha}{\alpha+\beta} \right) \log(\hat{\theta}_3/(1-\hat{\theta}_3)) (1 - B(\alpha+1, \beta, \theta_2^*)) \right) + B \quad (37) \end{aligned}$$

Cuando $P(\theta) = \text{Be}(\alpha, \beta)$, $E((1-\theta)\log(1-\theta))$ y $E(\theta \log(\theta))$ se encuentran en (48)-(49) de la sección anterior. Al sustituir estas expresiones en la última ecuación obtenida para C y usar el hecho de que

$$\Psi(x+1) = \Psi(x) + \frac{1}{x} \quad x > 0 \quad (38)$$

se obtiene C como se presentó en (31).

Con la inicial $U(\Pi_1, \Pi_2, \Pi_3)$, que distribuye uniformemente la probabilidad Π_1 sobre θ_1 , el cálculo de C , con la pérdida de la divergencia logarítmica, requiere considerar cinco posibilidades, $E(\theta) < \theta_1$, $E(\theta) = \theta_1$, $\theta_1 < E(\theta) < \theta_2$, $E(\theta) = \theta_2$ y $\theta_2 < E(\theta)$, con $E(\theta)$ la media de la inicial. Esto porque $U(\Pi_1, \Pi_2, \Pi_3)$ define una f.d.p.g que no es continua como función de θ . En este trabajo, sólo se presenta la expresión para C en el caso en que $E(\theta) < \theta_1$. De hecho, con esta condición resulta que $\hat{\theta}_1 = E(\theta)$, $\hat{\theta}_2 = \theta_1$ y $\hat{\theta}_3 = \theta_2$. Como $P(\theta) = U(\Pi_1, \Pi_2, \Pi_3)$ y con la expresión (15) para el $\min_{\mathcal{D}} L(d, \theta)$ se llega a que

$$\begin{aligned}
 C = A & \left[\left(\frac{\Pi_1}{\theta_1} \right) \left(\int_0^{\theta_1^*} \delta(\theta, \hat{\theta}_1) d\theta + \int_{\theta_1^*}^{\theta_1} \delta(\theta, \hat{\theta}_2) d\theta \right) + \right. \\
 & + \left(\frac{\Pi_2}{(\theta_2 - \theta_1)} \right) \left(\int_{\theta_1}^{\theta_2^*} \delta(\theta, \hat{\theta}_2) d\theta + \int_{\theta_2^*}^{\theta_2} \delta(\theta, \hat{\theta}_3) d\theta \right) + \left(\frac{\Pi_3}{(1 - \theta_2)} \right) \left(\int_{\theta_2}^1 \delta(\theta, \hat{\theta}_3) d\theta \right) \Big] \\
 & + B.
 \end{aligned} \tag{39}$$

Si se sustituye $\delta(\theta, \hat{\theta}_1)$ en esta expresión y se recuerda que

$$\int \theta \log \theta d\theta = \left(\frac{\theta^2}{2} \right) \left(\log \theta - \frac{1}{2} \right) \tag{40}$$

y

$$\int (1 - \theta) \log(1 - \theta) d\theta = - \frac{(1 - \theta)^2}{2} \left(\log(1 - \theta) - \frac{1}{2} \right), \tag{41}$$

entonces

$$C = A \left[\left(\frac{\Pi_1 \theta_2 - (\Pi_2 + \Pi_1) \theta_1}{(\theta_1 - \theta_1) \theta_1} \right) \left(\frac{\theta_1^2 (\log(\theta_1) - 1/2)}{2} - \frac{(1 - \theta_1)^2 (\log(1 - \theta_1) - 1/2)}{2} \right) \right]$$

$$\begin{aligned}
& + \left(\frac{\pi_2 + \pi_3 \theta_1 - (\pi_2 + \pi_3) \theta_2}{(\theta_2 - \theta_1)(1 - \theta_2)} \right) \left(\frac{\theta_2^2 (\log(\theta_2))}{2} - \frac{\theta_2^2}{2} - \frac{(1 - \theta_2)^2 (\log(1 - \theta_2) - 1/2)}{2} \right) \\
& - \frac{1}{4} \left(\frac{\pi_1}{\theta_1} + \frac{\pi_3}{(1 - \theta_2)} \right) - \frac{\pi_1}{\theta_1} (\log(1 - \hat{\theta}_1)(\theta_1^*)) - \frac{\pi_1}{\theta_1} (\log(\hat{\theta}_1/(1 - \hat{\theta}_1))) \frac{(\theta_1^*)^2}{2} \\
& - \left(\frac{\pi_1(\theta_1 - \theta_1^*)}{\theta_1} + \frac{\pi_2(\theta_2^* - \theta_1)}{(\theta_2 - \theta_1)} \right) (\log(1 - \hat{\theta}_2)) \\
& - (\log(\hat{\theta}_2/(1 - \hat{\theta}_2))) \left(\frac{\pi_1(\theta_1^2 - (\theta_1^*)^2)}{2\theta_1} + \frac{\pi_2((\theta_2^*)^2 - \theta_1^2)}{2(\theta_2 - \theta_1)} \right) \\
& - (\log(1 - \hat{\theta}_3)) \left(\frac{\pi_2(\theta_2 - \theta_2^*)}{(\theta_2 - \theta_1)} + \pi_3 \right) \\
& - (\log(\hat{\theta}_3/(1 - \hat{\theta}_3))) \left(\frac{\pi_2(\theta_2^2 - (\theta_2^*)^2)}{2(\theta_2 - \theta_1)} + \frac{\pi_3(1 + \theta_2)}{2} \right) + B
\end{aligned} \tag{42}$$

Otros casos se pueden abordar en forma análoga, por ejemplo, si $E(\theta) = \theta_1$,

$$\begin{aligned}
C = A \left(\int_0^{\theta_1} \frac{\pi_1}{\theta_1} \delta(\theta, \hat{\theta}_2) d\theta + \int_{\theta_1}^{\theta_2^*} \frac{\pi_2}{(\theta_2 - \theta_1)} \delta(\theta, \hat{\theta}_2) d\theta + \int_{\theta_2^*}^{\theta_2} \frac{\pi_2}{(\theta_2 - \theta_1)} \delta(\theta, \hat{\theta}_3) d\theta \right. \\
\left. + \int_{\theta_2}^1 \frac{\pi_3}{(1 - \theta_2)} \delta(\theta, \hat{\theta}_3) d\theta \right) + B.
\end{aligned} \tag{43}$$

Las integrales de esta última ecuación son fáciles de obtener.

Antes de presentar unos ejemplos numéricos de evaluación de tamaño de muestra conforme a los principios de esta sección, se presenta el Teorema III.4.1 que establece condiciones para la convergencia del Riesgo de Bayes al valor esperado de la información perfecta, en el contexto del problema de contraste definido en este trabajo.

TEOREMA III.4.1. Sea $Z_n = (X_1, X_2, \dots, X_n)$ una colección de variables aleatorias con f.d.p.g. $f(X|\theta)$, $\theta \in \Theta = [a, b]$, donde $-\infty \leq a < b \leq \infty$ y $P(\theta)$ es una distribución inicial continua para θ . Si se considera el contraste de las hipótesis

$$d_1 : H_1: [a, \theta_1], \quad d_2 : H_2: [\theta_1, \theta_2], \quad d_3 : H_3: (\theta_2, b] \quad (44)$$

con $\theta_1 < \theta_2$ y la función de pérdida $L(d_i, \theta)$; $i = 1, 2, 3$ satisface las condiciones

i) acotada en $[a, b]$,

ii) no depende de los datos Z_n ,

$$\text{iii) } \min_{\mathcal{D}} L(d, \theta) = L(d_1, \theta) I_{\Omega_1}(\theta) + L(d_2, \theta) I_{\Omega_2}(\theta) + L(d_3, \theta) I_{\Omega_3}(\theta) \quad (45)$$

donde $\mathcal{D} = (d_1, d_2, d_3)$, $\Omega_1 = [a, \tilde{\theta}_1]$, $\Omega_2 = [\tilde{\theta}_1, \tilde{\theta}_2]$, $\Omega_3 = (\tilde{\theta}_2, b]$, $\tilde{\theta}_1 \leq \tilde{\theta}_2$. Entonces, si existe un estimador consistente para θ ,

$$\lim_{n \rightarrow \infty} L^*(n) = C \quad (46)$$

con

$$L^*(n) = E_{P_n} \left(\min_{\mathcal{D}} E_P \{ L(d, \theta) \} \right). \quad (47)$$

Demostración

Por el Teorema II.3.1 se tiene que

$$L^*(n) \geq C, \quad \forall n \in \mathbb{N}. \quad (48)$$

Por otra parte,

$$L^*(n) = \int_{Z_n} P(Z_n) \int_{\Theta} L(\delta_B(Z_n), \theta) P(\theta | Z_n) d\theta \, dZ_n. \quad (49)$$

con $\delta_B(Z_n)$ la decisión de Bayes dado Z_n . Ahora bien, $\delta_B(\cdot)$ define una regla de decisión (de Bayes) con dominio en Z_n y contradominio en \mathcal{D} . Sea ahora $\delta(Z_n)$ cualquier otra regla de decisión. Por ser $\delta_B(\cdot)$ de Bayes se tiene que

$$E_{P|Z_n}(L(\delta_B(Z_n), \theta)) \leq E_{P|Z_n}(L(\delta(Z_n), \theta)) \quad (50)$$

para toda $Z_n \in Z_n$. Es inmediato entonces que

$$L^*(n) \leq L_\delta(n) \quad (51)$$

con

$$L_\delta(n) = \int_{Z_n} P(Z_n) \int_{\Theta} L(\delta(Z_n), \theta) P(\theta|Z_n) d\theta dZ_n. \quad (52)$$

En estas condiciones, basta con exhibir una regla de decisión $\delta(Z_n)$ tal que

$$\lim_{n \rightarrow \infty} L_\delta(n) = C \quad (53)$$

para obtener la conclusión del teorema.

Si $\delta(Z_n)$ es una regla de decisión cualquiera entonces, intercambiando límites de integración, se tiene

$$\begin{aligned} L_\delta(n) &= \int_{\Theta} P(\theta) \int_{Z_n} L(\delta(Z_n), \theta) P(Z_n|\theta) dZ_n d\theta \\ &= \int_{\Theta} R(\delta(\cdot), \theta) P(\theta) d\theta. \end{aligned} \quad (54)$$

Sea, finalmente, $\hat{\theta}_M$ un estimador de consistente para θ y considere la siguiente regla de decisión

$$\delta(Z_n) = d_1 \quad \text{si } \hat{\theta}_M \leq \bar{\theta}_1, \quad (55)$$

$$\delta(Z_n) = d_2 \quad \text{si } \bar{\theta}_1 < \hat{\theta}_M < \bar{\theta}_2 \quad (56)$$

y

$$\delta(Z_n) = d_3 \quad \text{si } \hat{\theta}_M \geq \bar{\theta}_2. \quad (57)$$

Es fácil verificar que para esta regla de decisión

$$\begin{aligned} R(\delta(\cdot), \theta) &= L(d_1, \theta) \Pr\{\hat{\theta}_M \leq \bar{\theta}_1 | \theta\} + L(d_2, \theta) \Pr\{\bar{\theta}_1 < \hat{\theta}_M < \bar{\theta}_2 | \theta\} \\ &+ L(d_3, \theta) \Pr\{\hat{\theta}_M \geq \bar{\theta}_2 | \theta\}. \end{aligned} \quad (58)$$

Al sustituir esta ecuación en la expresión de $L_{\delta}(n)$ de (54) se obtiene

$$\begin{aligned} L_{\delta}(n) &= \int_a^b L(d_1, \theta) \Pr\{\hat{\theta}_M \leq \bar{\theta}_1 | \theta\} P(\theta) d\theta + \int_a^b L(d_2, \theta) \Pr\{\bar{\theta}_1 < \hat{\theta}_M < \bar{\theta}_2 | \theta\} P(\theta) d\theta \\ &+ \int_a^b L(d_3, \theta) \Pr\{\hat{\theta}_M \geq \bar{\theta}_2 | \theta\} P(\theta) d\theta \\ &= \int_a^{\bar{\theta}_1} \left[L(d_1, \theta) \Pr\{\hat{\theta}_M \leq \bar{\theta}_1 | \theta\} + L(d_2, \theta) \Pr\{\bar{\theta}_1 < \hat{\theta}_M < \bar{\theta}_2 | \theta\} + L(d_3, \theta) \Pr\{\hat{\theta}_M \geq \bar{\theta}_2 | \theta\} \right] P(\theta) d\theta \\ &+ \int_{\bar{\theta}_1}^{\bar{\theta}_2} \left[L(d_1, \theta) \Pr\{\hat{\theta}_M \leq \bar{\theta}_1 | \theta\} + L(d_2, \theta) \Pr\{\bar{\theta}_1 < \hat{\theta}_M < \bar{\theta}_2 | \theta\} + L(d_3, \theta) \Pr\{\hat{\theta}_M \geq \bar{\theta}_2 | \theta\} \right] P(\theta) d\theta \\ &+ \int_{\bar{\theta}_2}^b \left[L(d_1, \theta) \Pr\{\hat{\theta}_M \leq \bar{\theta}_1 | \theta\} + L(d_2, \theta) \Pr\{\bar{\theta}_1 < \hat{\theta}_M < \bar{\theta}_2 | \theta\} + L(d_3, \theta) \Pr\{\hat{\theta}_M \geq \bar{\theta}_2 | \theta\} \right] P(\theta) d\theta. \end{aligned} \quad (59)$$

Por otro lado,

$$C = \int_a^{\bar{\theta}_1} L(d_1, \theta) P(\theta) d\theta + \int_{\bar{\theta}_1}^{\bar{\theta}_2} L(d_2, \theta) P(\theta) d\theta + \int_{\bar{\theta}_2}^b L(d_3, \theta) P(\theta) d\theta, \quad (60)$$

entonces,

$$\begin{aligned} L_\delta(n) - C &= \int_a^{\bar{\theta}_1} \left[-L(d_1, \theta) \Pr\{\hat{\theta}_M > \bar{\theta}_1 | \theta\} + L(d_2, \theta) \Pr\{\bar{\theta}_1 < \hat{\theta}_M < \bar{\theta}_2 | \theta\} \right. \\ &\quad \left. + L(d_3, \theta) \Pr\{\hat{\theta}_M \geq \bar{\theta}_2 | \theta\} \right] P(\theta) d\theta + \int_{\bar{\theta}_1}^{\bar{\theta}_2} \left[L(d_1, \theta) \Pr\{\hat{\theta}_M \leq \bar{\theta}_1 | \theta\} \right. \\ &\quad \left. - L(d_2, \theta) (\Pr\{\hat{\theta}_M \leq \bar{\theta}_1 | \theta\} + \Pr\{\hat{\theta}_M \geq \bar{\theta}_2 | \theta\}) + L(d_3, \theta) \Pr\{\hat{\theta}_M \geq \bar{\theta}_2 | \theta\} \right] P(\theta) d\theta \\ &\quad + \int_{\bar{\theta}_2}^b \left[L(d_1, \theta) \Pr\{\hat{\theta}_M \leq \bar{\theta}_1 | \theta\} + L(d_2, \theta) \Pr\{\bar{\theta}_1 < \hat{\theta}_M < \bar{\theta}_2 | \theta\} \right. \\ &\quad \left. - L(d_3, \theta) \Pr\{\hat{\theta}_M < \bar{\theta}_2 | \theta\} \right] P(\theta) d\theta. \end{aligned} \quad (61)$$

Puesto que $\hat{\theta}_M$ es consistente, el límite, cuando n tiende a infinito, de cada una de las probabilidades que aparecen en la ecuación anterior es cero, excepto para algunos valores de θ con P -medida cero. Dado que $L(d_i, \theta)$ es acotada es posible aplicar el Teorema de Convergencia Dominada de Lebesgue en la expresión (61) y como la pérdida no depende de los datos,

$$\lim_{n \rightarrow \infty} L_\delta(n) - C = 0 \quad (62)$$

El problema de clasificación de semillas fué abordado, con el enfoque bayesiano, y sin utilizar función de pérdida alguna, en Mendoza y Valencia (1986) para clasificar distintas variedades de maíz. En ese caso, se utilizó, de acuerdo a la rutina del laboratorio correspondiente $n = 800$. Marginalmente, es interesante notar que este tamaño de muestra garantiza, desde la perspectiva frecuentista, un error máximo de 0.05 en la estimación puntual de θ con una confianza del 99%. Por otra parte, los límites para la clasificación de las variedades se fijaron en $\theta_1 = 0.40$ y $\theta_2 = 0.75$. Estas mismas condiciones se consideraron en este trabajo para utilizar el procedimiento descrito esta sección.

n	$L^*(n)$
0	0.6000
1	0.4612
50	0.1030
100	0.0731
150	0.0599
200	0.0519
250	0.0464
300	0.0424
800	0.0260
∞	0.0000

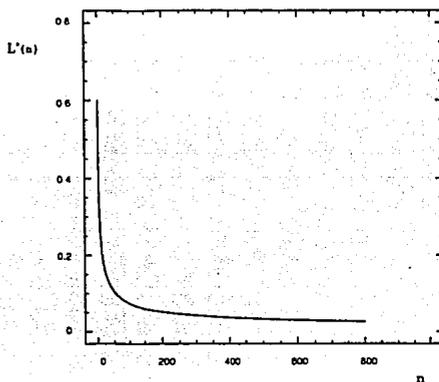


FIGURA III.4.1

VALORES DEL RIESGO DE BAYES. PERDIDA 0-K. INICIAL UNIFORME

$$k_1=k_2=k_3=1. \theta_1= .40, \theta_2= .75.$$

Con la pérdida 0-k ($k_1 = k_2 = k_3 = 1$) y una distribución inicial uniforme, la gráfica de $L^*(n)$, junto con algunos de sus valores se reportan en la figura III.4.1. De dicha tabla se puede notar que $L^*(0) - C = 0.6$ y con $t = 0.01$, se obtiene $c = .006$. El valor mínimo valor de n que satisface $L^*(n) - C < c$ es mayor a 800. Si $t = 0.1$ entonces $c = .06$ y el menor valor de n que satisface $L^*(n) - C < c$ es 150.

Con la pérdida lineal ($k_{ij} = 1$) y de nuevo con una inicial uniforme, la gráfica de $L^*(n)$, junto con algunos de sus valores, se presentan en la figura III.4.2.

n	$L^*(n)$
0	0.1112
1	0.0753
50	0.0041
100	0.0020
150	0.0014
200	0.0010
250	0.0008
300	0.0007
800	0.0004
∞	0.0000

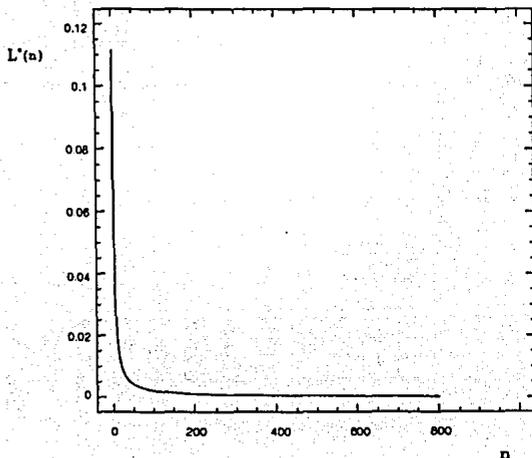


FIGURA III.4.2
VALORES DEL RIESGO DE BAYES. PERDIDA LINEAL. INICIAL UNIFORME

$$k_{ij} = 1, \theta_1 = .40, \theta_2 = .75$$

Ahora, $L^*(0) - C = .1112$ y si se utiliza $t = 0.01$, $\epsilon = .0011$. Con $n = 193$ se cumple la condición $L^*(0) - C < \epsilon$. Si $t = 0.1$, $\epsilon = .0111$ y el tamaño de muestra resulta ser igual a 19.

Con la pérdida cuadrática ($C_1 = C_2 = C_3 = 1$) y una inicial uniforme, la gráfica de $L^*(n)$, junto con algunos de sus valores, aparecen en la figura III.4.3.

n	$L^*(n)$
0	0.0583
1	0.0200
50	-0.0560
100	-0.0581
150	-0.0588
200	-0.0592
250	-0.0594
300	-0.0595
800	-0.0600
∞	-0.0602

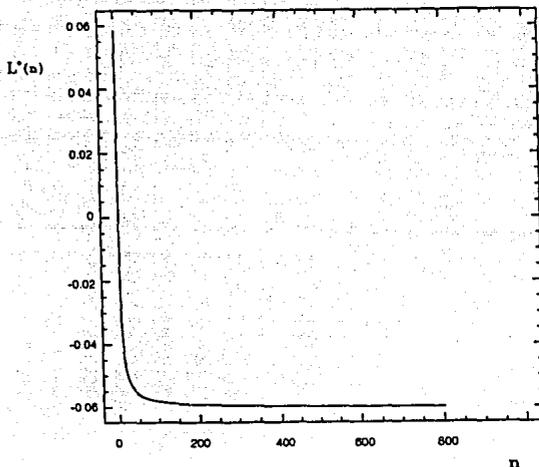


FIGURA III.4.3

VALORES DEL RIESGO DE BAYES. PERDIDA CUADRÁTICA. INICIAL UNIFORME

$$C_1 = C_2 = C_3 = 1. \quad \theta_1 = .40, \theta_2 = .75$$

De la información contenida en la figura se tiene que, $L^*(0) - C = 0.1185$. Con $t = 0.01$, $\epsilon = .0011$, por tanto $n = 196$ es un tamaño de muestra adecuado. Si $t = 0.1$, $\epsilon = .0118$ el valor de n que satisface $L^*(n) - C < \epsilon$ resulta ser 17.

Con la divergencia logarítmica ($A = 1$, $B = 0$) y una inicial uniforme, la gráfica de $L^*(n)$ aparece, junto con algunos de sus valores, en la figura III.4.4.

n	$L^*(n)$
0	0.1931
1	0.1499
50	0.0807
100	0.0791
150	0.0785
200	0.0782
250	0.0780
300	0.0779
800	0.0776
∞	0.0773

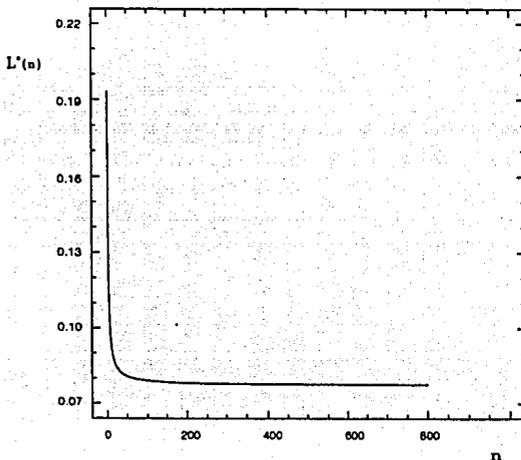


FIGURA III.4.4.

VALORES DEL RIESGO DE BAYES. DIVERGENCIA LOGARÍTMICA. INICIAL UNIFORME

$$A = 1, B = 0. \theta_1 = .40, \theta_2 = .75.$$

De la figura es fácil obtener que $L^*(0) - C = 0.1158$. Si $t = 0.01$, $\epsilon = (L^*(0) - C)t = .0011$. En consecuencia, $n \approx 161$ es más que suficiente. Si se decide usar $t = 0.1$, $\epsilon = .0115$ y por tanto n vale 15.

En la tabla III.4.1 se presenta un resumen de los cálculos efectuados para obtener el tamaño de muestra, así como los correspondientes valores de n para cada una de las cuatro pérdidas que se han manejado en este trabajo. En todo momento se supone que $P(\theta) = Be(1,1)$, $\theta_1 = .4$ y $\theta_2 = .75$.

	0-K (k=1)	LINEAL (k _{1j} =1)	CUADRATICA (C _i =1)	LOGARITMICA (A=1,B=0)
L*(0)-C	0.6000	0.1112	0.1185	0.1158

$$t = .01$$

e	0.0060	0.0011	0.0011	0.0011
n	> 800	193	196	161
L*(n)-C		0.0010	0.0010	0.0010

$$t = .1$$

e	0.0600	0.0111	0.0118	0.0115
n	150	19	17	15
L*(n)-C	0.5990	0.0101	0.0114	0.0105

TABLA III.5.5
TAMAÑO DE MUESTRA. INICIAL UNIFORME
 $\theta_1 = .40$, $\theta_2 = .75$

De los ejemplos se puede notar que los valores de n que da la desigualdad $L^*(n) - C < \epsilon$ para las pérdidas lineal y cuadrática son esencialmente los mismos, siempre que la inicial sea uniforme. Para la divergencia logarítmica el valor de n es un tanto menor y en el caso de la pérdida $0-k$ los valores de n resultan ser mayores respecto a los obtenidos con las otras pérdidas. Resultados similares se pueden obtener con otras Betas.

En Huerta y Mendoza (1993b) se calculan los tamaños de muestra según lo discutido en esta sección para la pérdida de la divergencia logarítmica ($A = 1$, $B = 0$) con la inicial de Jeffreys y con una inicial Beta (2,2) usando $t = 0.01$. Los tamaños de muestra obtenidos fueron $n = 72$ y $n = 359$ respectivamente. Estos valores de n , junto con el de 161 que se obtuvo para la uniforme puede llevar a la conclusión de que a mayor información, el procedimiento exige más muestra. Los resultados no son del todo sorprendentes si se toma en cuenta que al usar una inicial muy informativa la diferencia $L^*(0) - C$ más cercana a cero. En consecuencia, la reducción en pérdida que produce una muestra de tamaño n , $L^*(0) - L^*(n)$, aumenta más lentamente cuando se trabaja en una situación que incorpore una gran cantidad de información inicial. En otras palabras, una inicial informativa conduce a una situación en donde queda poco por aprender ($L^*(0)$ cercano a C) y para reducir, en términos relativos, esa pérdida es necesario un esfuerzo de muestreo mayor. A juicio del autor, el método propuesto en la sección III.4 no presenta contradicciones lógicas. Lo que puede llevar a resultados con poco interés práctico, es el uso constante de un mismo valor de t . En el siguiente capítulo, se propone una modificación al método de tamaño de muestra basada en la idea de *Decisiones de Referencia* que aparece en Bernardo y Smith (1994).

CAPITULO IV

VARIANTE DEL PROBLEMA ORIGINAL

INTRODUCCION

Se define un procedimiento de evaluación de tamaño de muestra a través de la distribución de referencia asociada a un problema de decisión. Se implementa esta modificación al problema de contraste original y con las cuatro pérdidas presentadas en el capítulo anterior. Con la familia Beta como la clase de iniciales admisibles, se demuestra, que para tres de las pérdidas es necesario incorporar una restricción adicional para obtener un tamaño de muestra razonable. Se ilustra con varios ejemplos.

IV.1.- MODIFICACION AL PROCEDIMIENTO DE TAMAÑO DE MUESTRA

En la sección III.4 del capítulo anterior, se propuso el procedimiento que determina n tal que $L^*(n) - C < t(L^*(0) - C)$, para un valor de t fijo y cercano a cero. Se ilustró el comportamiento de esta estrategia en una variedad de situaciones y en general se obtuvieron resultados interesantes. Sin embargo, también se indicó que esta idea puede producir tamaños de muestra mayores conforme la inicial es *más informativa*. Cómo se explicó en la sección III.4, ésto se debe a que tanto $L^*(0) - C$ como $L^*(0) - L^*(n)$ son casi cero cuando la inicial $P(\theta)$ es *muy informativa*.

La diferencia $L^*(0) - C$, que a partir de este momento se denotará por $L_P^*(0) - C_P$, dada su dependencia de la inicial $P(\theta)$, se puede interpretar como la disminución en pérdida esperada que corresponde al paso de una situación en la que se tiene información inicial $P(\theta)$ a otra en la que se cuenta con información perfecta. Cuando $L_P^*(0) - C_P$ es casi cero, el decisor considera que su información inicial es valiosa en el sentido de que lo coloca en una situación que ya es cercana a la de información perfecta y por tanto podría no estar dispuesto a tomar información adicional. En especial, podría rehusarse a emplear una n , posiblemente grande, en donde $L^*(n) - C_P$ fuese menor que $t(L_P^*(0) - C_P)$. Por otra parte, si la inicial es *no informativa*,

es decir, si $L^*(0) - C_p$ es grande respecto de cero, lo más razonable es que el decisor esté dispuesto a usar información adicional y en particular a utilizar un tamaño de muestra n tal que

$$L^*(n) - C_p < t(L^*(0) - C_p). \quad (1)$$

La idea en general es que para una misma función de pérdida un decisor debería de requerir menos información adicional (un menor tamaño de muestra) si su inicial es más informativa. Una forma de instrumentar esta idea, modificando el procedimiento del capítulo anterior, es la siguiente.

Si ante un problema de decisión es posible definir una inicial $\Pi(\theta)$ no informativa y para $\Pi(\theta)$ se tiene que el tamaño de muestra óptimo es el menor entero n_r tal que

$$L^*(n_r) - C_\pi < t(L^*(0) - C_\pi), \quad (2)$$

con t un número positivo, cercano a cero, entonces, el tamaño de muestra n_o con cualquier otra inicial $P(\theta)$ se puede definir como

$$n_o = n_r - n_e, \quad (3)$$

donde n_e es el número de piezas experimentales a las que equivale la información contenida en $P(\theta)$.

Para poder implementar esta estrategia es necesario formalizar conceptos como *ignorancia*, *no informativo* o *número de piezas experimentales equivalentes*. Esto se discute en la sección IV.2.

IV.2.- DECISIONES DE REFERENCIA

En un problema de decisión definido por \mathcal{D} , Θ , $P(\theta)$, $L(d, \theta)$ con Θ un espacio parametral, la solución por pérdida esperada mínima, sin datos, se obtiene al determinar $d_0 \in \mathcal{D}$ tal que

$$E_p(L(d_0, \theta)) = \min_{\mathcal{D}} E_p(L(d, \theta)). \quad (1)$$

Si se dispone de un vector de observaciones $Z_n = (X_1, X_2, \dots, X_n)$ con f.d.p.g. $f(Z_n | \theta)$, la solución de Bayes con ésta información muestral se obtiene al buscar $d^*(z_n) \in \mathcal{D}$ tal que

$$E_{p|Z_n}(L(d^*(z_n), \theta)) = \min_{\mathcal{D}} E_{p|Z_n}(L(d, \theta)). \quad (2)$$

Para algunos decisores, la dependencia de la decisión óptima $d^*(z_n)$ en la inicial $P(\theta)$ puede ser incómoda. Muchas veces se prefiere un procedimiento en el cual la decisión óptima sólo se seleccione en base al vector Z_n y al modelo $f(Z_n | \theta)$. Necesariamente, un criterio de decisión de éste tipo es inconsistente con los Axiomas de Coherencia de la Teoría de la Decisión, sin embargo, y como parte de un análisis de sensibilidad, un decisor coherente puede estar interesado en la decisión óptima sugerida sólo por los datos.

Por esto, es de interés identificar una distribución inicial sobre θ , $\Pi(\theta)$, para la cual la información muestral resulte lo más determinante posible. La decisión óptima $\delta(z_n)$, obtenida con $\Pi(\theta | Z_n)$, que es la distribución final que corresponde a la inicial $\Pi(\theta)$, es una *decisión de referencia* que se puede interpretar como la decisión óptima en una situación donde se espera que los datos aporten la mayor contribución posible al proceso de decisión.

Como siempre, \mathcal{D} denota el espacio de decisiones, Θ el espacio parametral y $L(d, \theta)$ la pérdida por seleccionar $d \in \mathcal{D}$ cuando θ es el verdadero valor del parámetro.

Se supondrá que $P(\theta)$, la distribución inicial, pertenece a una clase \mathcal{F} de distribuciones permisibles para modelar el conocimiento inicial. Finalmente, $Z_n = (X_1, X_2, \dots, X_n)$ es un vector de observaciones donde cada X_i es el resultado de un experimento $\mathcal{E} = \{ X, \theta, f(X|\theta) \}$ que permite mejorar el conocimiento sobre θ .

Antes de obtener el vector Z_n , la pérdida esperada de la decisión óptima d_0 es

$$L_p^*(0) = E_p(L(d_0, \theta)). \quad (3)$$

Con la muestra Z_n , la pérdida esperada de la decisión óptima $d^*(z_n)$ es

$$L_p^*(n, Z_n) = E_{p|Z_n}(L(d^*(z_n), \theta)). \quad (4)$$

De aquí, es posible definir la disminución en pérdida dada por una muestra de tamaño n como

$$\begin{aligned} V(n, P(\theta)) &= L_p^*(0) - \int_{Z_n} L_p^*(n, Z_n) P(Z_n) dZ_n \\ &= L_p^*(0) - L_p^*(n), \end{aligned} \quad (5)$$

donde $L_p^*(n)$ es el Riesgo de Bayes para una muestra de tamaño n , calculado con la inicial $P(\theta)$. Cuando n tiende a infinito, $V(n, P(\theta))$ se puede interpretar como la utilidad faltante debido al desconocimiento de θ . Dicho límite se denotará por $V(\infty, P(\theta))$.

Si se trata de identificar la distribución inicial sobre θ para la que la información de los datos resulte lo más útil posible, parece natural buscar $\Pi(\theta)$ tal que maximice $V(\infty, P(\theta))$ sobre la clase \mathcal{F} . Una vez determinada $\Pi(\theta)$, la decisión $\delta(z_n)$ que minimice

$$\int_{\mathcal{D}} L(d, \theta) \Pi(\theta | Z_n) d\theta \quad (6)$$

sobre \mathcal{D} con

$$\Pi(\theta|Z_n) \propto f(Z_n|\theta)\Pi(\theta), \quad (7)$$

será entonces, la decisión de referencia buscada.

Bernardo (1979) menciona que el límite de $V(n, P(\theta))$ cuando n tiende a infinito no siempre existe, ni la maximización de $V(\omega, P(\theta))$ sobre \mathcal{F} es siempre posible, a menos que \mathcal{F} se restrinja en alguna forma. Sin embargo, se puede definir $\Pi(\theta)$ mediante un proceso límite, al determinar $\Pi_n(\theta)$ tal que maximice $V(n, P(\theta))$ y hacer

$$\Pi(\theta) = \lim_{n \rightarrow \infty} \Pi_n(\theta). \quad (8)$$

En las condiciones del Teorema III.5.1 siempre existe $V(\omega, P(\theta))$, más aún,

$$V(\omega, P(\theta)) = L P^*(0) - C_P \quad (9)$$

con C_P el valor esperado de la información perfecta con la inicial $P(\theta)$.

DEFINICION IV.2.1- (Bernardo (1979)) Considere un problema de decisión $(\mathcal{D}, \theta, L(d, \theta), P(\theta))$. Sea $Z_n = (X_1, X_2, \dots, X_n)$ un vector de variables aleatorias con f.d.p.g. conjunta $f(Z_n|\theta)$ y \mathcal{F} una clase de iniciales admisibles para θ . Una decisión de referencia $\delta(z_n)$, después de haber observado Z_n , se define como una decisión en \mathcal{D} tal que minimiza (6), con $\Pi(\theta|Z_n)$ la final que corresponde a la inicial $\Pi(\theta)$ que maximiza $V(\omega, P(\theta))$ sobre \mathcal{F} . $\Pi(\theta)$ recibe el nombre de distribución de referencia o distribución no informativa para el problema de decisión definido por $(\mathcal{D}, \theta, L(d, \theta), P(\theta))$.

El concepto de distribución de referencia permite determinar el número de piezas experimentales al que equivale la información contenida en cualquier inicial $P(\theta)$ en \mathcal{F} para el caso en que $\lim_{n \rightarrow \infty} L P^*(n)$ existe y es finito. Si $\Pi(\theta)$ es la inicial de referencia, entonces

$$L \Pi^*(0) - C_\Pi \geq L P^*(0) - C_P \quad (10)$$

para toda inicial $P(\theta) \in \mathcal{F}$. Si el Riesgo de Bayes para $\Pi(\theta)$ se denota por $L \Pi^*(n)$,

por la desigualdad anterior, existe un entero no negativo n_0 tal que aproximadamente

$$L\pi^*(n_0) - C\pi = Lp^*(0) - C\pi. \quad (11)$$

Notemos que n_0 es un tamaño de muestra que deja a un "decisor" que utiliza como inicial $\Pi(\theta)$ en las mismas condiciones, en cuanto a pérdida se refiere, que a un decisor que usa la inicial $P(\theta)$ y que no tiene información muestral. Por esta razón, a n_0 se le denominará el tamaño de muestra al que equivale la información en $P(\theta)$.

En base a lo discutido en ésta sección y en la sección anterior y suponiendo que $\lim_{n \rightarrow \infty} Lp^*(n)$ existe, se puede estructurar una variante al procedimiento de tamaño de muestra de la sección III.4, ésta consiste en

i) Determinar, para el problema de decisión definido por $(\mathcal{D}, \Theta, L(d, \theta), P(\theta))$, una distribución de referencia $\Pi(\theta)$ para la clase \mathcal{F} .

ii) Para un $t \in (0,1)$ y con el Riesgo de Bayes $L^*\pi(n)$, calcular n_r tal que

$$L\pi^*(n_r) - C\pi < t(L\pi^*(0) - C\pi). \quad (12)$$

iii) Si la inicial del decisor es $P(\theta)$, determinar n_0 que satisfaga (11).

iv) Definir el tamaño de muestra como

$$n_0 = \max(0, n_r - n_0). \quad (13)$$

En la siguiente sección se discute como implementar esta nueva forma de seleccionar el tamaño de muestra en el contexto del problema de contraste de hipótesis que ha sido abordado a lo largo de este trabajo.

IV.3.- CONTRASTE DE HIPOTESIS Y TAMAÑO DE MUESTRA

Si es de interés probar

$$H_1: \theta \in [0, \theta_1], \quad H_2: \theta \in (\theta_1, \theta_2), \quad H_3: \theta \in [\theta_2, 1] \quad (1)$$

donde las observaciones son Bernoulli de parámetro θ , se decide usar la Pérdida 0-k ($k_1 = k_2 = k_3 = k$) y además \mathcal{F} es una familia de distribuciones en el $[0,1]$ cuyos elementos satisfacen que

$$\Pr\{\theta \in [0, \theta_1]\} = P_1, \quad \Pr\{\theta \in [\theta_1, \theta_2]\} = P_2, \quad \Pr\{\theta \in [\theta_2, 1]\} = P_3 \quad (2)$$

con $0 \leq P_i \leq 1$, $i = 1, 2, 3$ y $P_1 + P_2 + P_3 = 1$, la distribución de referencia sobre \mathcal{F} , según la definición IV.2.1, es cualquier distribución tal que

$$\Pr\{\theta \in [0, \theta_1]\} = \Pr\{\theta \in [\theta_1, \theta_2]\} = \Pr\{\theta \in [\theta_2, 1]\} = \frac{1}{3} \quad (3)$$

La prueba de este resultado es como sigue.

Sea $P(\theta)$ cualquier inicial en \mathcal{F} . Por definición de la pérdida 0-k

$$E_p(L(d_1, \theta)) = k(1-P_1)$$

$$E_p(L(d_2, \theta)) = k(1-P_2)$$

$$E_p(L(d_3, \theta)) = k(P_1+P_2). \quad (4)$$

Es fácil notar que H_1 es Bayes si y solo si

$$P_2 \leq P_1$$

y

$$1-2P_1 \leq P_2. \quad (5)$$

H₂ es Bayes siempre que

$$P_1 \leq P_2$$

y

$$1 - P_1 \leq 2P_2, \quad (6)$$

y por último, H₃ es Bayes si

$$P_2 \leq 1 - 2P_1$$

y

$$2P_2 \leq 1 - P_1. \quad (7)$$

Se sabe que con la pérdida O-K, C_p = 0 para toda inicial P(θ) ∈ ℱ. En consecuencia,

$$L_p^*(0) - C_p = \begin{cases} k(1-P_1) & \text{si } P_2 \leq P_1 \text{ y } 1-2P_1 \leq P_2 \\ k(1-P_2) & \text{si } P_1 \leq P_2 \text{ y } 1-P_1 \leq 2P_2 \\ k(P_1+P_2) & \text{si } P_2 \leq 1-2P_1 \text{ y } 2P_2 \leq 1-P_1, \end{cases} \quad (8)$$

0 ≤ P_i ≤ 1, i = 1, 2, P₁+P₂ ≤ 1. Al maximizar esta última expresión como función de P₁ y P₂ se obtiene que la distribución de referencia en ℱ es cualquier que le da un tercio de probabilidad a cada hipótesis.

Es interesante observar que en este caso la distribución de referencia coincide con la que se considera no informativa para un parámetro de interés discreto con tres posibles valores. Cuando las constantes que definen a la pérdida O-k no son todas iguales, es posible encontrar la distribución de referencia mediante la misma técnica que se utilizó para probar el caso en que k₁=k₂=k₃=k.

Ejemplo IV.3.1.

Suponga que se desea evaluar el tamaño de muestra para el problema de contraste definido en (1) con θ₁ = .4, θ₂ = .75 y que se ha decidido utilizar la Pérdida O-k, con k_i = 1, i = 1, 2, 3. También se considera que el conocimiento inicial sobre θ se puede modelar adecuadamente a través de una Beta (α, β), es decir, la familia Beta define la clase de iniciales admisibles.

La distribución de referencia $\Pi(\theta)$ es una Beta que le da $\frac{1}{3}$ de probabilidad inicial a cada una de las hipótesis. La Beta de parámetros $\alpha = 1, \beta = .79$, aproximadamente cumple con esta condición, es decir, $\Pi(\theta) = \text{Be}(1, .79)$. La diferencia $L\pi^*(0) - C\pi$ es igual a .6655. Con $t = .05$, $t(L\pi^*(0) - C\pi)$ es igual a .033275, de donde se puede mostrar que $n_r = 454$ ($L\pi^*(454) - C\pi = .033222$). En la tabla IV.3.1 aparecen valores de n_0 , n_e , $L\pi^*(0) - C\pi$ y de $L\pi^*(n) - C\pi$ para las iniciales Beta (.5,.5), Beta (1,1), Beta (10,10), Beta (20,20). Las iniciales Beta (.5,.5) y Beta (1,1) se seleccionaron para implementar el método modificado de tamaño de muestra en situaciones que se pueden considerar de no información. Por el contrario, las densidades Beta (10,10) y Beta (20,20) representan situaciones con bastante información inicial y en estas circunstancias se obtuvieron los tamaños de muestra correspondientes.

	n_0	$L\pi^*(0) - C\pi$	n_e	$L\pi^*(n_e) - C\pi$
Be(.5,.5)	454	0.5640	0	0.6655
Be(1,1)	454	0.6000	0	0.6655
Be(10,10)	442	0.1949	12	0.1949
Be(20,20)	407	0.1024	47	0.1018

$t = .05$ $\epsilon = .0332$ $n_r = 454$

TABLA IV.3.1
TAMAÑOS DE MUESTRA. METODO MODIFICADO
PERDIDA 0-K. $k_1 = 1, \theta_1 = .40, \theta_2 = .75$

Se puede notar que las iniciales Beta (1,1) y Beta (.5,.5) se comportan como distribuciones de referencia dado que en ambos casos $n_0 = 0$.

Para la Pérdida Lineal y con la familia Beta como la clase de distribuciones admisibles, no es fácil determinar una expresión para $Lp^*(0) - Cp$ en función de α y β , los parámetros de la inicial $P(\theta)$ que permita determinar analíticamente la distribución de referencia, $\Pi(\theta)$. Con las constantes que definen a la pérdida iguales a uno y con $\theta_1 = .4$ y $\theta_2 = .75$, se efectuó una exploración numérica para determinar, aproximadamente, los valores de α y β que maximizan $Lp^*(0) - Cp$. Los resultados de la exploración aparecen en la Tabla IV.3.2.

α	β	$Lp^*(0) - Cp$
.5	.5	.16684
.6	.4	.15656
.4	.6	.15444
.6	.6	.15189
.4	.4	.18493
.5	.3	.17347
.3	.5	.15118
.3	.3	.20731
.4	.2	.19414
.2	.4	.14133
.2	.2	.23575
.3	.1	.15559
.1	.3	.11256
.1	.1	.27167
.15	.05	.16937
.05	.15	.12853
.05	.05	.28462
.0001	.0001	.29996
.1xE-6	.1xE-6	.29999

TABLA IV.3.2

EXPLORACION NUMÉRICA PARA DETERMINAR LA DISTRIBUCION DE REFERENCIA

PERDIDA LINEAL. $k_{ij} = 1$. $\theta_1 = .40$, $\theta_2 = .75$

Como se puede notar de la tabla, se tomó $\alpha = .5$, $\beta = .5$ como punto inicial para hacer la evaluación. A partir de ahí se consideraron cuatro puntos, todos equidistantes del punto inicial, dos de ellos sobre la recta $\alpha = \beta$ y los otros dos

sobre la recta perpendicular a $\alpha = \beta$ que pasa por (.5,.5). De estos puntos el de mayor valor de $Lp^*(0) - Cp$ es (.4,.4). ($Lp^*(0) - Cp = .18493$). Este se uso como nuevo punto inicial para proceder de la misma manera en que se trabajó con el punto (.5,.5). Después de algunas iteraciones se observó que $Lp^*(0) - Cp$ crece a medida que α y β tienden a cero con $\alpha = \beta$.

La exploración sugiere que con Pérdida Lineal no existe la distribución de referencia, cuando la familia de distribuciones admisibles es la familia Beta. Una posibilidad consiste en tomar el kernel de una Beta, hacer $\alpha = 0$, $\beta = 0$ y obtener entonces $\Pi(\theta) \propto \theta^{-1}(1-\theta)^{-1}$. Sin embargo, como en este caso $\Pi(\theta)$ define una distribución impropia, no existe $Lp^*(n)$ y por tanto no se puede evaluar el tamaño de muestra según lo discutido en las dos primeras secciones de este capítulo. Por otro lado, se puede tomar $\alpha = \beta$ e identificar el límite, cuando α tiende a cero, de la distribución Beta (α, α). Si este límite existe y define una f.d.p.g. propia, es posible evaluar el Riesgo de Bayes para la distribución límite y con ella calcular el tamaño de muestra. A continuación se enuncian dos propiedades que permiten garantizar que el límite existe y que además, define una distribución propia.

i) Si $F_{\alpha, m}(\theta)$ es la función de distribución de una $Be(\alpha, \alpha+m)$, $m > 0$ entonces,

$$\lim_{\alpha \rightarrow 0} F_{\alpha, m}(\theta) = \begin{cases} 1 & \text{si } \theta \geq 0 \\ 0 & \text{en otro caso.} \end{cases} \quad (9)$$

ii) Si $F_{\alpha, m}(\theta)$ es la función de distribución de una $Beta(\alpha+m, \alpha)$ $m > 0$, entonces

$$\lim_{\alpha \rightarrow 0} F_{\alpha, m}(\theta) = \begin{cases} 1 & \text{si } \theta \geq 1 \\ 0 & \text{en otro caso.} \end{cases} \quad (10)$$

Para demostrar i) basta considerar que la media de una Beta ($\alpha, \alpha+m$) es

$$\frac{\alpha}{2\alpha+m} \quad (11)$$

y su varianza es

$$\frac{\alpha(\alpha+m)}{(2\alpha+m+1)(2\alpha+m)^2} \quad (12)$$

Cuando α tiende a cero, se puede ver que la media y la varianza se hacen cero y por tanto $Be(\alpha, \alpha+m)$ converge en media cuadrática a la distribución de (9). Como convergencia en media cuadrática implica convergencia en distribución, se obtiene la propiedad i). La propiedad ii) se prueba de manera similar.

Por otra parte, es fácil verificar que

$$(\alpha+\beta)Be(\alpha, \beta) = \alpha Be(\alpha+1, \beta) + \beta Be(\alpha, \beta+1) \quad (13)$$

de donde se obtiene que

$$F\alpha(\theta) = \frac{1}{2} F_1(\theta) + \frac{1}{2} F_2(\theta), \quad (14)$$

con $F_1(\theta)$ la función de distribución de una Beta $(\alpha, \alpha+1)$, $F_2(\theta)$ la función de distribución de una Beta $(\alpha+1, \alpha)$ y $F\alpha(\theta)$ la función de distribución de una Beta (α, α) .

Por las propiedades i) y ii) dadas en (9) y (10) se concluye que

$$\lim_{\alpha \rightarrow 0} F\alpha(\theta) = \frac{1}{2} I_{(0,1)}(\theta) + \frac{1}{2} I_{(1,\infty)}(\theta). \quad (15)$$

Esta última igualdad y la tabla IV.3.1 sugieren definir, para la Pérdida Lineal, a la distribución de referencia como una distribución discreta que le da probabilidad $\frac{1}{2}$ al valor $\theta = 0$ y probabilidad $\frac{1}{2}$ al valor $\theta = 1$. A continuación se calcula el Riesgo de Bayes que corresponde a esta f.d.p.g.

Si

$$\Pi(\theta) = \frac{1}{2} I_{(0)}(\theta) + \frac{1}{2} I_{(1)}(\theta), \quad (16)$$

$L\pi^*(n)$ es su Riesgo de Bayes y se utiliza la Pérdida Lineal para el contraste de hipótesis en (1) entonces,

$$E_{\pi}(L(d_1, \theta)) = \frac{k_{11}(1-\theta_1)}{2}, \quad (17)$$

$$E_{\pi}(L(d_2, \theta)) = \frac{k_{21}\theta_1}{2} + \frac{k_{22}(1-\theta_2)}{2}, \quad (18)$$

$$E_{\pi}(L(d_3, \theta)) = \frac{k_{31}\theta_2}{2}, \quad (19)$$

con $E_{\pi}(L(d_i, \theta))$, $i = 1, 2, 3$, el valor esperado de la pérdida $L(d_i, \theta)$ respecto de $\Pi(\theta)$. Como $L\pi^*(0)$ es el valor de Bayes sin datos,

$$L\pi^*(0) = \min \left\{ \frac{k_{11}(1-\theta_1)}{2}, \frac{k_{21}\theta_1}{2} + \frac{k_{22}(1-\theta_2)}{2}, \frac{k_{31}\theta_2}{2} \right\}. \quad (20)$$

Si $Y = \sum_{i=1}^n X_i$, por el Teorema de Bayes,

$$\Pi(\theta | Y) \propto \theta^Y (1-\theta)^{n-Y} \left(\frac{1}{2} I_{(0)}(\theta) + \frac{1}{2} I_{(1)}(\theta) \right) \quad (21)$$

$Y = 0, 1, \dots, n$

de donde

$$\Pi(\theta | Y) = \begin{cases} 1 & \text{si } \theta = 1, Y = n \\ 1 & \text{si } \theta = 0, Y = 0 \\ 0 & \text{en otro caso} \end{cases} \quad (22)$$

Por otra parte, la distribución marginal de Y para la inicial $\Pi(\theta)$ es

$$Pr(Y) = \frac{1}{2} (0)^Y (1-0)^{n-Y} + \frac{1}{2} (1)^Y (0)^{n-Y}$$

$$= \begin{cases} 1/2 & \text{si } Y = 0, Y = n \\ 0 & \text{en otro caso} \end{cases}, \quad (23)$$

por lo que el Riesgo de Bayes para $\Pi(\theta)$, independientemente de la pérdida y con $n > 0$, satisface

$$L\pi^*(n) = \frac{1}{2} L^*(n, Y=0) + \frac{1}{2} L^*(n, Y=n), \quad (24)$$

donde $L^*(n, Y=i)$, es el valor de Bayes con una muestra de tamaño n y dado que se observó $Y=i$, $i = 0$ o $i = n$.

Si $Y = 0$ y se está usando la Pérdida Lineal, resulta que

$$E_{\pi|Y}(L(d_1, \theta)) = 0, \quad (25)$$

$$E_{\pi|Y}(L(d_2, \theta)) = k_{21}\theta_1, \quad (26)$$

$$E_{\pi|Y}(L(d_3, \theta)) = k_{31}\theta_2, \quad (27)$$

con $E_{\pi|Y}(L(d_i, \theta))$, $i = 1, 2, 3$ el valor esperado de la pérdida $L(d_i, \theta)$ respecto de $\Pi(\theta|Y)$. De estas tres ecuaciones se puede notar que con la Pérdida Lineal y la inicial $\Pi(\theta)$

$$L^*(n, Y=0) = 0. \quad (28)$$

Con $Y = n$,

$$E_{\pi|Y}(L(d_1, \theta)) = k_{11}(1-\theta_1), \quad (29)$$

$$E_{\pi|Y}(L(d_2, \theta)) = k_{22}(1-\theta_2), \quad (30)$$

y

$$E_{\pi|Y}(L(d_3, \theta)) = 0. \quad (31)$$

En consecuencia,

$$L^*(n, Y=n) = 0, \quad (32)$$

y por tanto para la Pérdida Lineal y la inicial $\Pi(\theta)$ de (16) se tiene que

$$L\pi^*(n) = 0 \quad (33)$$

para todo n mayor que cero.

De esta manera, si para la Pérdida Lineal se decide utilizar como distribución de referencia la f.d.p.g. $\Pi(\theta)$ que da $\frac{1}{2}$ de probabilidad a $\theta = 0$ y a $\theta = 1$, su Riesgo de Bayes, $L\pi^*(n)$, garantiza que $n_r = 0$ y por tanto $n_o = 0$, para toda inicial $P(\theta)$. Por supuesto, ésto no proporciona una solución satisfactoria al problema de tamaño de muestra. Es interesante notar que el Riesgo de Bayes de la inicial $\Pi(\theta)$ mencionada establece que con una pieza de información es más que suficiente para que el riesgo sea igual al valor esperado de la información perfecta. Este resultado es de esperarse dada la forma de la distribución final $\Pi(\theta|Y)$ que aparece en (22). Por todas estas dificultades, se decidió buscar la distribución que maximice $Lp^*(0) - Cp$, con la restricción de maximizar sólo sobre aquellas Betas tales $\alpha + \beta = 1$, para usarse como de Referencia en la evaluación del tamaño de muestra n_o . Siguiendo a De Groot (1970), una Beta tal que $\alpha + \beta = 1$, contiene información equivalente a una pieza experimental.

Con la Pérdida Lineal es necesario usar algún método numérico para encontrar una Beta $(\alpha, 1-\alpha)$ que maximice $Lp^*(0) - Cp$. Con las constantes de la pérdida iguales a uno, $\theta_1 = .4$ y $\theta_2 = .75$, la tabla IV.3.3 presenta valores de $Lp^*(0) - Cp$ para diferentes iniciales Beta $(\alpha, 1-\alpha)$.

α	$LP^*(0)-C_p$
.05	.01259
.10	.02704
.15	.04340
.20	.06170
.25	.08195
.30	.10417
.35	.12833
.40	.15444
.45	.17686
.50	.16684
.55	.16006
.60	.15656
.65	.15639
.70	.15959
.75	.12826
.80	.09841
.85	.07058
.90	.04485
.95	.02130

TABLA IV.3.3

EXPLORACION NUMERICA PARA DETERMINAR LA DISTRIBUCION DE REFERENCIA

$\alpha + \beta = 1$. PERDIDA LINEAL. $k_{ij} = 1$. $\theta_1 = .40$, $\theta_2 = .75$

Es fácil notar, de la información de la tabla, que un buen candidato para la distribución de referencia con la restricción $\alpha + \beta = 1$ es la Beta (.45,55) que como distribución difiere poco de la Beta (.5,.5).

Ejemplo IV.3.2.

Suponga que se desea obtener el tamaño de muestra para el problema de contraste definido en IV.3.1 con $\theta_1 = .4$, $\theta_2 = .75$ y que se ha decidido utilizar la

Pérdida Lineal con $k_{11} = k_{21} = k_{22} = k_{31} = 1$. Con $\Pi(\theta) = \text{Be}(.45, .55)$, de la tabla IV.3.2 se obtiene que $L\pi^*(0) - C\pi = .17686$. Si $t = .01$, $t(L\pi^*(0) - C\pi) = .00176$, además $L\pi^*(80) - C\pi = .00177$, entonces $n_r = 80$. En la tabla IV.3.4 aparecen valores de n_o , n_e , $L\pi^*(0) - C\pi$, $L\pi^*(n_e) - C\pi$ usando como iniciales: Be (.5, .5), Be (1, 1), Be (10, 10) y Be (20, 20).

	n_o	$L\pi^*(0) - C\pi$	n_e	$L\pi^*(n_e) - C\pi$
Be(.5, .5)	80	0.1668	0	0.1768
Be(1, 1)	79	0.1112	1	0.0781
Be(10, 10)	68	0.0109	12	0.0109
Be(20, 20)	42	0.0036	38	0.0036

$t = .01$ $c = .0017$ $n_r = 80$

TABLA IV.3.4

TAMAÑOS DE MUESTRA. METODO MODIFICADO. $\alpha + \beta = 1$

PERDIDA LINEAL. $k_{ij} = 1$. $\theta_1 = .40$, $\theta_2 = .75$

Ejemplo IV.3.3.

En el mismo contexto del ejemplo anterior, se quiere obtener el tamaño de muestra para las iniciales: Beta (.5, .5), Beta (1, 1), Beta (10, 10) y Beta (20, 20), pero ahora con $\theta_1 = .2$ y $\theta_2 = .8$. La tabla IV.3.5 presenta valores de $L\pi^*(0) - C\pi$, para diferentes distribuciones Beta (α, β), con la restricción $\alpha + \beta = 1$.

α	$Lp^*(0) - C_p$
.05	.02507
.10	.05240
.15	.08189
.20	.11343
.25	.10761
.30	.09674
.35	.08832
.40	.08232
.45	.07874
.50	.07754
.55	.07874
.60	.08232
.65	.08832
.70	.09674
.75	.10761
.80	.11343
.85	.08189
.90	.05240
.95	.02507

TABLA IV.3.5
 EXPLORACION NUMERICA PARA DETERMINAR LA DISTRIBUCION DE REFERENCIA
 $\alpha + \beta = 1$. PERDIDA LINEAL. $k_{ij} = 1$. $\theta_1 = .20$, $\theta_2 = .80$.

La tabla muestra dos candidatos para la distribución de referencia con la restricción $\alpha + \beta = 1$, la Beta (.2,.8) y la Beta (.8,.2). La tabla IV.3.6 muestra una evaluación más cuidadosa de $Lp^*(0) - C_p$ para valores de α cercanos a .2 y a .8.

α	$L_{P^*}(0)C_P$
.16	.08804
.17	.09427
.18	.10058
.19	.10696
.21	.11809
.22	.11532
.23	.11265
.24	.11008
.76	.11008
.77	.11265
.78	.11532
.79	.11809
.81	.10696
.82	.10058
.83	.09427
.84	.08804

TABLA IV.3.6

EXPLORACION NUMERICA PARA DETERMINAR LA DISTRIBUCION DE REFERENCIA

$$\alpha + \beta = 1. \text{ PERDIDA LINEAL. } k_{ij} = 1. \theta_1 = .20, \theta_2 = .80$$

De ésta segunda exploración se puede proponer a $\Pi(\theta) = \text{Be}(.21,.79)$ o bien a $\Pi(\theta) = \text{Be}(.79,.21)$. La existencia de dos distribuciones Beta que maximicen $L_{P^*}(0) - C_P$, con la restricción $\alpha + \beta = 1$, obliga al decisor a evaluar no para cada una de las distribuciones $\Pi(\theta)$ y tomar como tamaño de muestra el mayor de los valores no obtenidos. Con $\Pi(\theta) = \text{Be}(.21,.79)$, $L_{P^*}(0) - C_P = .11809$. Si $t = .01$, $t(L_{P^*}(0) - C_P) = .00118$. Como $L_{P^*}(.71) - C_P = .00117$ entonces $n_r = 71$. Resultados para las iniciales que se han manejado en otros ejemplos aparecen en la siguiente tabla.

	n_0	$Lp^*(0)-Cp$	n_e	$L\pi^*(n_e)-C\pi$
Be(.5, .5)	71	0.0775	0	0.1180
Be(1, 1)	70	0.0400	1	0.0425
Be(10,10)	0	0.6E-4	> 400	< 0.0002
Be(20,20)	0	0.3E-6	> 400	< 0.0002

$$t = .01 \quad \varepsilon = .0011 \quad n_r = 71$$

TABLA IV.3.7

TAMAÑOS DE MUESTRA. METODO MODIFICADO. $\alpha + \beta = 1$
 PERDIDA LINEAL. $k_{ij} = 1$. $\theta_1 = .20$, $\theta_2 = .80$

Si se toma $\Pi(\theta) = \text{Be}(.79,.21)$ los resultados son idénticos que con $\Pi(\theta) = \text{Be}(.21,.79)$. Note que las distribuciones Beta (10,10) y Beta (20,20) son relativamente (para $t = .01$) muy informativas ya que en ambos casos $n_0 = 0$ dado que el tamaño de muestra equivalente n_e es mayor a 400. Esto se presenta ya que $L\pi^*(400)-C\pi = 0.0002$ y para ambas distribuciones ocurre que $Lp^*(0)-Cp < L\pi^*(400)-C\pi$.

Para la Pérdida Cuadrática tampoco es fácil determinar, analíticamente, la distribución de referencia con la familia Beta como la familia de iniciales admisibles. La tabla IV.3.8 muestra una exploración numérica similar a la que se efectuó con Pérdida Lineal, para determinar la distribución de referencia en el caso de la Pérdida Cuadrática con $C_1 = C_2 = C_3 = 1$, $\theta_1 = .4$ y $\theta_2 = .75$.

α	β	$LP^*(0)-Cp$
.5	.5	.17552
.6	.4	.15634
.4	.6	.20789
.5	.7	.18698
.3	.5	.20616
.3	.7	.13810
.6	.6	.16025
.4	.4	.19397
.5	.3	.17114
.3	.5	.20616
.3	.3	.21676
.4	.2	.18812
.2	.4	.19558
.2	.2	.24569
.3	.1	.20599
.1	.3	.15837
.1	.1	.28373
.15	.05	.23482
.05	.15	.18643
.05	.05	.30769
.0001	.0001	.33618

TABLA IV.3.8

EXPLORACION NUMERICA PARA DETERMINAR LA DISTRIBUCION DE REFERENCIA

PERDIDA CUADRATICA. $C_1 = 1$. $\theta_1 = .40$, $\theta_2 = .75$.

La tabla parece indicar que una posible distribución de referencia con Pérdida Cuadrática, está dada, de nuevo, por la f.d.p. discreta $\Pi(\theta)$ que la da $\frac{1}{2}$ de probabilidad a $\theta = 0$ y a $\theta = 1$. Se puede verificar, con el mismo procedimiento que se uso para la Pérdida Lineal, que el Riesgo de Bayes para esta f.d.p.g con la Pérdida Cuadrática es,

$$L\pi^*(0) = \min \left\{ C_1 \left(\frac{1}{2} - \theta_1^2 \right), C_2 \left(\frac{1 - (\theta_1 + \theta_2)}{2} + \theta_1 \theta_2 \right), C_3 \left(2\theta_2 - \theta_2^2 - \frac{1}{2} \right) \right\} \quad (34)$$

y

$$L\pi^*(n) = \frac{1}{2} \left[-C_1 \theta_1^2 + C_3 (2\theta_2 - \theta_2^2 - 1) \right], \quad (35)$$

para toda n mayor que cero.

Si se decide utilizar como distribución de referencia la f.d.p.g $\Pi(\theta)$, las dos ecuaciones anteriores permiten asegurar que para cualquier inicial, $n_0 = 0$ y $n_0 = 0$. Al igual que con la Pérdida Lineal se decidió, para la Pérdida Cuadrática, incorporar la restricción $\alpha + \beta = 1$ para determinar la distribución de referencia. En la tabla IV.3.9 aparecen valores de $L\pi^*(0) - C_p$ con Pérdida Cuadrática, $C_1 = C_2 = C_3 = 1$, $\theta_1 = .4$, $\theta_2 = .75$ para diferentes iniciales Beta (α , $1-\alpha$). En la tabla IV.3.10, los valores de $L\pi^*(0) - C_p$ se calcularon con $C_1 = C_2 = C_3 = 1$, $\theta_1 = .2$ y $\theta_2 = .8$.

α	$L\pi^*(0) - C_p$
.05	.01614
.10	.03489
.15	.05635
.20	.08065
.25	.10787
.30	.13810
.35	.17141
.40	.20789
.45	.19007
.50	.17552
.55	.16427
.60	.15634
.65	.15175
.70	.15052
.75	.15263
.80	.11557
.85	.08182
.90	.05133
.95	.02408

TABLA IV.3.9

EXPLORACION NUMERICA PARA DETERMINAR LA DISTRIBUCION DE REFERENCIA

$\alpha + \beta = 1$. PERDIDA CUADRATICA. $C_1 = 1$. $\theta_1 = .40$, $\theta_2 = .75$.

α	$Lp^*(0) - Cp$
.05	.02631
.10	.05527
.15	.08683
.20	.12096
.25	.10761
.30	.09674
.35	.08832
.40	.08232
.45	.07874
.50	.07754
.55	.07874
.60	.08232
.65	.08832
.70	.09674
.75	.10761
.80	.12096
.85	.08683
.90	.05527
.95	.02631

TABLA IV.3.10

EXPLORACION NUMERICA PARA DETERMINAR LA DISTRIBUCION DE REFERENCIA

$\alpha + \beta = 1$. PERDIDA CUADRATICA. $C_1 = 1$. $\theta_1 = .20$, $\theta_2 = .80$.

Para $\theta_1 = .4$ y $\theta_2 = .75$, la tabla IV.3.9 muestra que un buen candidato para la distribución de referencia es la Beta (.4, .6). De hecho se exploró con más detalle la función $Lp^*(0) - Cp$ alrededor de .4 y se confirmó que el punto máximo es, numéricamente indistinguible de $\alpha = .4$. Con $\theta_1 = .2$ y $\theta_2 = .8$, la tabla IV.3.10 muestra dos posibles distribuciones de referencia, la Beta (.2, .8) y la Beta (.8, .2).

Ejemplo IV.3.4.

Para la Pérdida Cuadrática, $C_1 = C_2 = C_3 = 1$, $\theta_1 = .4$ y $\theta_2 = .75$, se quiere evaluar el tamaño de muestra con las iniciales: Beta (.5, .5), Beta (1,1), Beta (10,10) y Beta (20,20). Si se usa como distribución de referencia

$\Pi(\theta) = \text{Be}(4.6)$, $L\pi^*(0) - C\pi$ es igual a .20789. Con $t = .01$, $t(L\pi^*(0) - C\pi) = .00207$ y $n_r = 66$ ya que $L\pi^*(66) - C\pi = .002068$ ($L\pi^*(n_r) - C\pi < .00207$). Resultados para iniciales que se han manejado en otros ejemplos aparecen en la tabla IV.3.11.

	n_o	$L\pi^*(0) - C\pi$	n_e	$L\pi^*(n_e) - C\pi$
Be(.5,.5)	66	0.1775	0	0.2078
Be(1,1)	65	0.1185	1	0.0698
Be(10,10)	56	0.0124	11	0.0112
Be(20,20)	34	0.0041	32	0.0041

$t = .01$ $\epsilon = .0020$ $n_r = 66$

TABLA IV.3.11

TAMAÑOS DE MUESTRA. METODO MODIFICADO. $\alpha + \beta = 1$.

PERDIDA CUADRATICA. $C_1 = 1$. $\theta_1 = .40$, $\theta_2 = .75$.

Si $\theta_1 = .2$, $\theta_2 = .8$ y se decide usar como distribución de referencia $\Pi(\theta) = \text{Be}(.2,.8)$ entonces $L\pi^*(0) - C\pi = .12096$. Con $t = .01$, $t(L\pi^*(0) - C\pi) = .00120$ y $n_r = 66$ ya que $L\pi^*(66) - C\pi = .00121$. Resultados de n_o , n_e , $L\pi^*(0) - C\pi$ y $L\pi^*(n_e) - C\pi$ con las iniciales que se han utilizado en otros ejemplos, aparecen en la tabla IV.3.12.

	n_o	$L\pi^*(0) - C\pi$	n_e	$L\pi^*(n_e) - C\pi$
Be(.5,.5)	66	0.0775	0	0.1209
Be(1,1)	65	0.0400	1	0.0409
Be(10,10)	0	$0.6E-4$	> 400	$< 0.210E-3$
Be(20,20)	0	$0.3E-6$	> 400	$< 0.210E-3$

$t = .01$ $\epsilon = .0012$ $n_r = 66$

TABLA IV.3.12

TAMAÑOS DE MUESTRA. METODO MODIFICADO. $\alpha + \beta = 1$.

PERDIDA CUADRATICA. $C_1 = 1$. $\theta_1 = .20$, $\theta_2 = .80$.

Es de interés notar que las distribuciones Beta (.5,.5) y Beta (1,1) esencialmente se comportan como de referencia para los dos casos discutidos. Las distribuciones Beta (10,10) y la Beta (20,20) por otra parte son informativas para el caso $\theta_1 = .2$ y $\theta_2 = .8$, con $t = 0.01$.

Para la Pérdida de la Divergencia Logarítmica de con $A = 1$, $B = 0$, $\theta_1 = .4$, $\theta_2 = .75$ y con la familia Beta como la familia de iniciales admisibles, se trató de determinar numéricamente la distribución de referencia con una exploración similar a la efectuada con la Pérdida Lineal y con la Pérdida Cuadrática. Los resultados aparecen en la tabla IV.3.13.

α	β	$Lp^*(0)-Cp$
.5	.5	.16361
.6	.4	.17008
.4	.6	.12652
.6	.6	.15094
.4	.4	.17881
.5	.3	.18130
.3	.5	.14376
.3	.3	.19747
.4	.2	.18745
.2	.4	.16777
.2	.2	.22099
.3	.1	.16217
.1	.3	.19539
.1	.1	.25172
.15	.05	.18631
.05	.15	.22871
.05	.05	.27100
.0001	.0001	.29384

TABLA IV.3.13

EXPLORACION NUMERICA PARA DETERMINAR LA DISTRIBUCION DE REFERENCIA

$\alpha + \beta = 1$. PERDIDA LOGARITMICA. $A = 1$, $B = 0$. $\theta_1 = .40$, $\theta_2 = .75$.

La evaluación parece indicar que no existe la distribución de referencia sobre la familia de las Betas en caso de usarse la Pérdida de Divergencia Logarítmica. Se

Se puede intentar utilizar como distribución de referencia la f.d.p.g. que le da $\frac{1}{2}$ de probabilidad al valor $\theta = 1$ y $\frac{1}{2}$ al valor $\theta = 0$. Su Riesgo de Bayes, con la discrepancia logarítmica, es

$$L\pi^*(0) = A \left(\min_1 - \left(\frac{\log(1-\hat{\theta}_1) + \log(\hat{\theta}_1)}{2} \right) \right) + B \quad (36)$$

y

$$L\pi^*(n) = A \left(- \left(\frac{\log(1-\hat{\theta}_1) + \log(\hat{\theta}_1)}{2} \right) \right) + B \quad (37)$$

para toda n mayor que cero. Estas expresiones se pueden obtener en la misma forma en que se obtuvieron las correspondientes para la Pérdida Lineal.

Por otro lado, este último resultado permite asegurar que si se usa la distribución discreta $\Pi(\theta) = \frac{1}{2} I_{(0)}(\theta) + \frac{1}{2} I_{(1)}(\theta)$, como distribución de referencia en el caso de la divergencia logarítmica, el tamaño de muestra óptimo no es igual a ceropara toda inicial $P(\theta)$. Por esta razón, se decidió incorporar la restricción $\alpha + \beta = 1$.

En la tabla IV.3.14 aparecen valores de $L\pi^*(0) - C_p$ para iniciales Beta $(\alpha, 1-\alpha)$ para la Pérdida de la Divergencia Logarítmica, $A = 1$, $B = 0$, $\theta_1 = .4$ y $\theta_2 = .75$.

α	$Lp^*(0)-Cp$
.05	.07623
.10	.10548
.15	.12159
.20	.13040
.25	.13427
.30	.13437
.35	.13135
.40	.12552
.45	.14854
.50	.16361
.55	.17081
.60	.17008
.65	.16112
.70	.14343
.75	.11618
.80	.11774
.85	.11382
.90	.10195
.95	.07592

TABLA IV.3.14

EXPLORACION NUMERICA PARA DETERMINAR LA DISTRIBUCION DE REFERENCIA

$\alpha + \beta = 1$. DIVERGENCIA LOGARITMICA. $\theta_1 = .40$, $\theta_2 = .75$

De la tabla se puede inferir que un candidato para la distribución de referencia con la restricción $\alpha + \beta = 1$ es la Beta (.55,.45). Se realizó una exploración más cuidadosa de $Lp^*(0) - Cp$ alrededor de $\alpha = .55$. Los resultados de ésta evaluación se reportan en la tabla IV.3.15.

α	$L\pi^*(0) - C\pi$
.51	.16568
.52	.16743
.53	.16887
.54	.17000
.55	.17081
.56	.17131
.57	.17148
.58	.17134
.59	.17087

TABLA IV.3.15
 EXPLORACION NUMERICA PARA DETERMINAR LA DISTRIBUCION DE REFERENCIA
 $\alpha + \beta = 1$. DIVERGENCIA LOGARITMICA. $\theta_1 = .40$, $\theta_2 = .75$.

De ésta última tabla se puede observar que un buen candidato para la distribución de referencia es $\Pi(\theta) = \text{Be}(.57, .43)$. A continuación se presenta un ejemplo.

Ejemplo IV.3.5.

Para la Pérdida de Divergencia Logarítmica , con $A = 1$, $B = 0$, $\theta_1 = .4$ y $\theta_2 = .75$, se quiere evaluar el tamaño de muestra para las iniciales Beta (.5,.5), Beta (1,1), Beta (10,10) y Beta (20,20). Si $\Pi(\theta) = \text{Be}(.57, .43)$, $L\pi^*(0) - C\pi = .17148$. Con $t = .01$, $t(L\pi^*(0) - C\pi) = .001714$ y $n_r = 67$ dado que $L\pi^*(67) - C\pi = .00169$. Los resultados para las iniciales mencionadas aparecen en la tabla IV.3.16.

	n_0	$Lp^*(0)-Cp$	n_e	$L\pi^*(n_e)-C\pi$
Be(.5,.5)	67	0.1631	0	0.1714
Be(1,1)	67	0.1158	0	0.1714
Be(10,10)	60	0.0160	7	0.0143
Be(20,20)	60	0.0065	17	0.0064

$$t = .01 \quad \epsilon = .0017 \quad n_r = 67$$

TABLA IV.3.16

TAMAÑOS DE MUESTRA. METODO MODIFICADO. $\alpha + \beta = 1$.
 DIVERGENCIA LOGARITMICA. $A = 1$, $B = 0$. $\theta_1 = .40$, $\theta_2 = .75$.

Nuevamente, la inicial uniforme y la de Jeffreys vuelven a comportarse como distribuciones no informativas.

En la tabla IV.3.17 se presentan los tamaños de muestra obtenidos para las pérdidas Lineal, Cuadrática y de Divergencia Logarítmica, con $\theta_1 = .4$, $\theta_2 = .75$ y las distribuciones: Beta de Referencia con la restricción $\alpha + \beta = 1$, Beta(.5,.5), Beta(1,1), Beta(10,10) y Beta(20,20).

	BE REF.	BE(.5.5)	BE(1,1)	BE(10,10)	BE(20,20)
LINEAL	80	80	80	68	42
CUADRATICA	66	66	66	55	34
LOGARITMICA	67	67	67	60	50

TABLA IV.3.17

TAMAÑOS DE MUESTRA. METODO MODIFICADO. $\alpha + \beta = 1$.
 $\theta_1 = .40$, $\theta_2 = .75$.

Como para construir esta tabla se usaron los valores $\theta_1 = .40$ y $\theta_2 = .75$, se puede utilizar directamente al problema de clasificación abordado por Mendoza y Valencia (1986). Notemos que en ningún caso de los considerados en la tabla, los tamaños de muestra son tan grandes como el de 800 que rutinariamente se utilizaba en laboratorio. También se puede ver de la tabla que la Pérdida Lineal es la que produce mayores tamaños de muestra y las pérdidas Logarítmica y Cuadrática prácticamente producen los mismos tamaños muestrales.

CAPITULO V

COMENTARIOS FINALES

En este trabajo se presentan dos alternativas para evaluar el tamaño de muestra para el problema de contraste definido por

$$H_1: \theta \in \Theta_1, \quad i = 1, 2, 3, \quad (1)$$

con $\Theta_1 = \{0, \theta_1\}$, $\Theta_2 = \{\theta_1, \theta_2\}$, $\Theta_3 = \{\theta_2, 1\}$ y en donde las observaciones son Bernoulli (θ). Ambos métodos necesitan de la especificación de una distribución inicial $P(\theta)$ y de una función acotada de pérdida $L(d_1, \theta)$, $i = 1, 2, 3$.

El primer procedimiento (sección III.4), que consiste en buscar un tamaño de muestra n tal que

$$L^*(n) - C < t(L^*(0) - C) \quad (2)$$

produce, para un valor de t fijo y constante, un tamaño de muestra más grande a medida que se incorpora mayor conocimiento inicial. El segundo procedimiento, basado en el concepto de Decisiones de Referencia (secciones IV.2-IV.3), exige menos muestra a mayor información inicial, pero puede requerir una restricción adicional en la clase de iniciales admisibles. Para el problema de contraste de hipótesis discutido en este trabajo y con algunas de las pérdidas de III.3, parece natural incorporar la restricción $\alpha + \beta = 1$ cuando la familia Beta define la clase de distribuciones iniciales admisibles. Esta condición se utiliza con el objetivo de evaluar el tamaño de muestra con distribuciones de referencia para varios casos particulares (sección IV.3) y los resultados obtenidos son bastante razonables. Es inmediato notar que ambos métodos pueden necesitar de técnicas numéricas para su implementación pero que conceptualmente son de carácter general y en particular no dependen del número de hipótesis que se estén considerando en el contraste.

Cuando las observaciones tienen f.d.p.g. arbitraria $f(X|\theta)$ $\theta \in \mathbb{R}$ y se desea probar

$$H_1: \theta \in [a, \theta_1], H_2: \theta \in (\theta_1, \theta_2], H_3: \theta \in (\theta_2, b)$$

(3)

con $-\infty \leq a < \theta_1 < \theta_2 < b \leq \infty$, se puede utilizar el método de III.4 para el evaluar el tamaño de muestra siempre que $L(d_i, \theta)$, $i = 1, 2, 3$ sea acotada (ver Teorema III.4.1). Si en este mismo contexto se quiere usar el método de la sección IV.2 podría ser necesario restringir la clase de iniciales admisibles a una familia paramétrica con miembros $P(\theta|\gamma)$ y γ un vector de hiperparámetros. Si existe $H(\gamma)$ una función que se pueda interpretar como el número de piezas experimentales equivalentes a la información contenida en $P(\theta|\gamma)$, es posible incorporar la restricción $H(\gamma) = 1$ en la búsqueda de la distribución de referencia. Cuando $f(X|\theta)$ es un miembro de la familia Exponencial regular, θ el parámetro natural, la familia conjugada básica es paramétrica y tiene definida a la función $H(\gamma)$ (Diaconis e Ylvisaker (1979)).

Las pérdidas de III.3 se pueden reformular para el caso en que θ sea el parámetro de cualquier modelo $f(X|\theta)$. De hecho, Gutiérrez-Peña (1991) plantea la pérdida de la divergencia logarítmica en un problema de contraste de dos hipótesis y para cualquier modelo $f(X|\theta)$. Cuando el espacio parametral θ no es acotado, las generalizaciones de las Pérdidas Lineal, Cuadrática y Logarítmicas resultan no acotadas sobre θ . Esto impide garantizar que el Riesgo de Bayes se aproxima, cuando n tiende a infinito, a la información perfecta C y por tanto puede no ser posible el cálculo del tamaño de muestra según lo que se ha desarrollado en este trabajo. Sería interesante plantear las pérdidas de III.3 con θ el parámetro natural de un modelo en la familia exponencial, θ no acotado y usar como iniciales elementos en la familia conjugada para estudiar el comportamiento del Riesgo de Bayes.

En el caso de no tener especificada una función de Pérdida y sí una distribución inicial $P(\theta)$, se puede intentar determinar n tal que alguna de las probabilidades finales de las hipótesis sea mayor que .95, por ejemplo. Sin embargo, si el verdadero valor de θ es θ_1 ó θ_2 (los puntos de corte), no siempre es posible concentrar la probabilidad final en alguna de las hipótesis, aunque n sea grande. Alternativamente, se puede utilizar la Pérdida O-K con $k_1 = k_2 = k_3 = 1$ y evaluar el tamaño de muestra conforme a lo que se ha presentado en el trabajo. Se recomienda hacer esto ya que usar la Pérdida O-k, $k_i = 1$, $i = 1, 2, 3$, equivale a decidirse por la hipótesis que presenta la mayor probabilidad final.

BIBLIOGRAFIA

- Adcock, C.J. (1987). *A bayesian approach to calculating sample sizes for multinomial sampling*. The Statistician, Vol 36, 155-159.
- Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*. 2nd.ed. Springer-Verlag: New York.
- Bernardo, J.M. (1976). *Psi (Digamma) Function, Algorithm ASI03*. Appl. Statist. 25, 1,315-316.
- Bernardo, J.M. (1979). *Reference posterior distributions for bayesian inference (with discussion)*. J.R. Statist. Soc. B, 41,113-147.
- Bernardo, J.M. (1979). *Reference Decisions*. Read at the International Meeting on Statistical Decision Theory, Rome, 4-7 June, 1979.
- Bernardo, J.M. (1981). *Bloestadística: Una perspectiva bayesiana*. Vicens-Vives: Barcelona.
- Bernardo, J.M. and Smith, A.F.M. (1994). *Bayesian Theory*. John Wiley and Sons: New York.
- Box, G.E.P. and Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley: Reading.
- De Groot, M.H. (1970). *Optimal Statistical Decisions*. Mc Graw-Hill: New York.
- Diaconis, P. and Ylvisaker, D. (1979) *Conjugate Priors for Exponential Families*. The Annals of Statistics, Vol 7, No.2, 269-281.

- Ferguson T.S. (1967) *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press: New York.
- Good, I.J. (1966). *A derivation of the probabilistic explanation of information*. J. Roy. Statist. Soc. B, Vol. 28, 578-581.
- Gutiérrez-Peña, E. (1991). *Contraste bayesiano de hipótesis paramétricas*. Tesis de Maestría. UACP y P del C.C.H.-UNAM.
- Gutiérrez-Peña, E. (1992). *Expected logarithmic divergence for exponential families*. Bayesian Statistics 4 (Bernardo, J.M. et. al. eds.). Oxford University Press. 669-674.
- Huerta, G. y Mendoza, M. (1993a). *El tamaño de muestra en un problema de clasificación de semillas*. Aceptado en Agrociencia.
- Huerta, G. and Mendoza, M. (1993b). *A bayesian approach to sample size selection in a Multiple Hypothesis Test*. Sometido a publicación.
- Johnson, B.R. and Truax, D.R. (1987). *Asymptotic properties of Bayes Risk for one-sided tests*. The Canadian Journal of Statistics. Vol. 15., No.1, 53-61.
- Lee, P.M. (1964). *On the axioms of information theory*. Ann. Math. Statist. 35, 415-418.
- Lindley, D.V. (1956). *On the measure of the information provided by an experiment*. Ann. Math. Statist. 27, 986-1005.
- Lindley, D.V. (1984). *Making Decisions*. 2nd.ed. John Wiley and Sons: New York.
- Mendoza, M. y Valencia, G. (1986). *Estudio comparativo de 105 colecciones de razas mexicanas de maíz almacenadas bajo condiciones adversas de temperatura y humedad*. Comunicación Interna 10. Departamento de Matemáticas, Facultad de Ciencias, UNAM.