



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

**PRUEBAS DE HIPOTESIS EN ESTADISTICA  
BAYESIANA**

**T E S I S**  
QUE PARA OBTENER EL TITULO DE  
M A T E M A T I C O  
P R E S E N T A I  
JOSE ANDRES CHRISTEN GRACIA

AGOSTO DE 1989

TESIS CON  
FALTA 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 esta 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.

## Prólogo

Los contrastes o pruebas de hipótesis es uno de los temas más estudiados dentro de la estadística. Dentro de la literatura correspondiente, en Inglés, se coincide en llamar a estos *Hypotheses tests*; en Español, se han usado varios términos como *testar*, *probar* o *contrastar* hipótesis. El más usado (en traducciones y trabajos originales) es *pruebas de hipótesis* sin embargo, el que me parece más adecuado para lo que son los métodos que estudiaremos en esta tesis es *contraste de hipótesis*. De aquí en adelante utilizaré este término.

En esta tesis restringiremos nuestro estudio a las hipótesis paramétricas sin embargo, no todas las hipótesis en estadística son de este tipo: También puede haber hipótesis no paramétricas. Tampoco hicimos ningún comentario acerca del tamaño de muestra óptimo para llevar a cabo un contraste de hipótesis y tampoco estudiamos los contrastes de hipótesis secuenciales (tamaño de muestra aleatorio). Todo esto tendrá que quedar para un estudio posterior.

Dentro de la estadística Bayesiana existen dos maneras de abordar el problema de los contrastes de hipótesis, uno es considerándolo como un problema de inferencia y otro visto como un problema de decisión; de esta manera, en el capítulo 1 presentamos los conceptos básicos de la teoría (Bayesiana) de decisiones, que después serán utilizados en la tesis. En el capítulo 2 se presentan los métodos para contrastar hipótesis incluyendo ejemplos diversos. En el capítulo 3 se estudian los contrastes de hipótesis que se suelen plantear acerca de los modelos lineales con la intención de estudiar algunos de los problemas que se tienen al intentar aplicar los métodos vistos en el capítulo 2; al mismo tiempo, hacemos una exposición de lo que serían las hipótesis multiparamétricas (que son una sencilla generalización de las uniparamétricas, estudiadas en el capítulo 2). Proponemos también, al final del capítulo, algunas opciones para solucionar los problemas encontrados. En el capítulo 4 se encuentran las conclusiones de todo el trabajo.

Esta tesis es una recopilación bibliográfica acerca de los contrastes de hipótesis en la estadística Bayesiana, considero que es una exposición original y también crítica del tema. La exposición de la tesis está dirigida para alguien que haya estudiado los elementos básicos de la estadística Bayesiana sin embargo, a lo largo del trabajo se utilizan conceptos de la teoría de la medida; en el apéndice A hay una exposición de teoría de la medida que puede servir como base para comprender lo que se usa de esta teoría en la tesis.

Espero haber sido suficientemente claro en la exposición de este trabajo:

Andrés Christen, Agosto 1989.

## INDICE

Capítulo 1	
Teoría de Decisiones	1
1.1 Conjunto de Acciones	2
1.2 Consecuencias y Funciones de Pérdida	3
1.3 Espacio Parametral	4
1.4 Pérdida Esperada y Decisión Óptima	5
Capítulo 2	
Contraste de Hipótesis	7
2.1 Definiciones	8
2.1.1 Contraste de Hipótesis	9
2.1.2 Contraste de Hipótesis Múltiples	9
2.2 Métodos para Contrastar Hipótesis	11
2.2.1 Comparación de Probabilidades	11
2.2.2 Problema de Decisión	19
2.2.3 Contraste de Significancia de Lindley	28
Capítulo 3	
Hipótesis Sobre Modelos Lineales	33
3.1 El Modelo Lineal	34
3.2 Hipótesis Lineal General	38
3.2 Ejemplos Comparativos	42
Capítulo 4	
Conclusiones	48
Apéndice A	52
Apéndice B	60
Apéndice C	64
Bibliografía	74

# Capítulo 1

## Teoría de Decisiones

# Capítulo 1

## Teoría de Decisiones

La intención de este capítulo es dar una revisión rápida a la teoría estadística de las decisiones, no con la intención de realizar un estudio profundo de ésta sino, más bien, exponer los temas que se van a utilizar en la tesis para, sobre todo; unificar la notación. La teoría es mucho más amplia de lo expuesto aquí, para una exposición completa se puede consultar Berger (1980), DeGroot (1970) o Lindley (1971). En este capítulo y en el resto de la tesis, se hace uso de conceptos básicos de teoría de la medida. Esto no debe de ser un obstáculo para entender este trabajo; para aquellos que no hayan estudiado medida antes, en el apéndice A hay una explicación rápida que puede ayudar a comprender lo que aquí se utiliza de esta teoría.

### 1.1 Conjunto de Acciones

Vamos a analizar lo que es un problema de decisión. Alguna persona, grupo de personas, empresa, etc. tiene un cierto problema. Para *actuar* ante el problema (no necesariamente resolverlo), se pueden llevar a cabo varias acciones, entonces: ¿ qué acción es la más adecuada ? ¿Cómo *decidir* qué acción es la más adecuada?

A la persona, grupo de personas, empresa etc. las llamaré decisor. A las acciones las agrupamos en un conjunto  $D$ . Para que exista algún problema de decisión, la cardinalidad de  $D$  debe de ser mayor que uno por que si no, la mejor acción es la única en  $D$  y el problema está resuelto. Es importante aclarar que  $D$  queda estipulado por el decisor y sólo por él. Si el problema es comprar un coche, el decisor estipula que coches estarán en  $D$  y cuales no, o si bien desea ponerlos todos, o sólo los de cuatro bujías, etc. Lo que esperaríamos es que el decisor incluyera en  $D$  *todos* los coches en los que estuviese interesado, y esto depende de las características del problema o sea: del decisor, de las necesidades del decisor y en general, de cómo enfrente el decisor su problema. Puede también depender de lo que haya estudiado acerca de él, de si se había enfrentado a un problema similar, de lo negligente que sea, de lo imaginativo que se comporte, etc. En algunos problemas  $D$  también puede ser considerado infinito, pero para el tema de esta tesis sólo se requerirán  $D$  finitos.

Es importante aclarar que sólo se consideran las opciones de  $D$  y por lo tanto la mejor decisión tendrá que pertenecer a  $D$ , NO importa si puede haber alguna otra decisión supuestamente mejor que las contenidas en  $D$ . A lo que vamos a aspirar es a encontrar la mejor acción contemplada en  $D$ , y en ese sentido será la

la decisión óptima y cualquier otra acción, por buena que pudiese considerarse, no podrá ser seleccionada.

## 1.2 Consecuencias y Funciones de Pérdida

Tomando en cuenta el conjunto de acciones  $D$  el problema es decidir que acción es la más adecuada. Para lograr esto necesitamos poder evaluar las consecuencias de decidirnos por cada acción es decir, si queremos decidir entre comprar el coche  $A$  o el coche  $B$  y conocemos las características de ambos, sus precios, sus medidas, sus limitantes; podríamos evaluar las consecuencias de preferir el coche  $A$  o el  $B$ . Por ejemplo, podemos evaluar la consecuencia de que el consumo de gasolina de el coche  $A$  sea de, por ejemplo, 6.0 Km. por litro, si es que estamos interesados en el consumo de gasolina. Esto podría ser bueno para una camioneta de carga pero tal vez no tanto para un automóvil particular, y las consecuencias dependerían del tipo de uso que le vayamos a dar al coche. Podríamos considerar también el número de puertas del coche, si es que el automóvil va a ser usado para transportar varias personas, y entonces la acción más conveniente sería un coche con muchas puertas; o también podríamos evaluar que tan caro es cada coche o una mezcla de varias características por ejemplo, si tenemos dinero para comprar el coche  $A$  o el  $B$  pero uno es de dos puertas y el otro es de cuatro, entonces tendríamos que mezclar ambas características para decidir que acción (comprar el coche  $A$  o el  $B$ ) es la más adecuada. En esta tesis vamos a suponer que el decisor evalúa las consecuencias de cada decisión mediante alguna función de pérdida. En la bibliografía mencionada se suelen usar funciones de utilidad o de pérdida indistintamente y como puede esperarse, los resultados son los mismos. En la tesis voy sólo a usar funciones que miden la pérdida de las consecuencias de tomar cada decisión, a decir,  $L(\cdot)$ , y entonces si  $L(A) < L(B)$  preferiríamos el coche  $A$  por tener menos inconvenientes que  $B$ . Para aclarar un poco más esta idea veamos otro ejemplo:

El problema es el siguiente, digamos que Jorge quiere hacer deporte el próximo domingo y tres amigos lo invitan a jugar tres distintos deportes. El primero lo invita a jugar Tenis, el segundo Squash, y el tercero Futbol. Jorge puede sólo considerar estas tres opciones, pero también decide incluir en  $D$  (el conjunto de acciones) el no ir con ningún amigo e ir a correr a un parque. Entonces procede a evaluar las inconveniencias de cada una de las acciones y optar por una de las cuatro considerando, por ejemplo; qué deporte le interesa más, si llueve o no, qué amigo es mejor deportista, quien es el mejor amigo, si cayera un meteorito sobre la cancha de Futbol o Tenis etc. Para evaluar las consecuencias de decidirse por tal o cual deporte Jorge realiza las siguientes consideraciones:

Alguien le propuso la posibilidad de que le cayera un meteorito en la cabeza mientras juega Fútbol, pero considera esta idea ridícula por lo poco probable (y esto es muy importante) que le parece y no se hubiera tomado la molestia ni siquiera de pensarlo. Su experiencia y su conocimiento le indican que de considerar ese evento la probabilidad sería demasiado pequeña y finalmente sólo estorbaría para el proceso de decisión. Por lo que respecta a las invitaciones, no tiene ningún inconveniente en preferir a algún amigo, pues con los tres tiene igual confianza y su relación es similar; no tiene ningún compromiso anterior con alguno y puede aceptar o rechazar la invitación de cualquiera de los tres amigos sin ningún resentimiento. Tampoco puede influir en su decisión que tan bien juegan su deporte pues, como él, los tres son aficionados. Lo que más le interesa a Jorge es qué deporte va a jugar el domingo, y en esos términos evaluar cada decisión. Pero esto depende (según Jorge) de si llueve o no. Procede entonces a realizar una lista de consecuencias, donde, por ejemplo: Si va a jugar Tenis y llueve no podrá jugar y se va a enojar mucho. Si va a Jugar Fútbol y llueve hará deporte pero no le parece muy agradable, etc. Jorge procede entonces a evaluar cada consecuencia con una función de pérdida  $L(\cdot)$ , pero vemos entonces que evaluar la consecuencia o pérdida de cada acción depende de un fenómeno aleatorio o incierto (introducido por Jorge) y este es el problema estadístico de la toma de decisiones, del cual nos vamos a ocupar en esta tesis (específicamente en el punto 2.2). En los puntos siguientes introducimos una exposición formal de las ideas anteriores.

### 1.3 Espacio Parametral

Sea  $X$  una variable aleatoria que toma valores en los reales. Supongamos que dicha variable pertenece a alguna familia de distribuciones indicada por algún parámetro, a decir:

$$X \in F_{\theta}, \theta \in \Theta.$$

Donde para cada  $\theta \in \Theta$  tenemos una distribución para  $X$ . No tenemos la información de cual es la  $\theta \in \Theta$  apropiada. Esto puede darse por ejemplo, si  $X$  es la cuantificación de algún aspecto de cierto fenómeno, es decir; la distribución de  $X$  es un modelo del conocimiento que tenemos acerca del comportamiento de  $X$ , y el valor de  $\theta$  sería un suceso incierto. Entonces, para cada  $\theta \in \Theta$  tenemos un modelo, o como suele decirse, algún estado de este fenómeno. Hay que notar que el conocimiento acerca de  $X$  puede variar por muchos factores, pero lo que estamos suponiendo es que no dudamos que  $X$  pertenezca a  $F_{\theta}$ . Ciertos autores coinciden en llamar a  $\Theta$  el *conjunto de estados de la naturaleza*, otros lo llaman *conjunto de modelos* y también ha recibido los nombres de *espacio parametral* y el de *conjunto de sucesos inciertos*. En esta tesis vamos a referirnos a  $\Theta$  indistintamente como el espacio parametral o el conjunto de sucesos inciertos.



Por lo visto en el párrafo anterior, no sabemos qué  $\theta \in \Theta$  sea la más conveniente para explicar el comportamiento de  $X$ . Según la interpretación subjetiva de la probabilidad, para expresar el conocimiento que tenemos acerca de  $\theta$  le asignamos una distribución de probabilidad, considerando a  $\theta$  una variable aleatoria. Entonces, la distribución de  $X$  junto con la de  $\theta$  modelan el conocimiento que se tiene acerca del comportamiento de  $X$ . Sea entonces,  $f(\theta)$  la función de densidad generalizada<sup>1</sup> de  $\theta$ . También puede darse el caso de que queramos obtener una muestra de  $X$  para aumentar la información que tenemos acerca del fenómeno y modificar la distribución de  $\theta$ , es por esto que a la distribución de  $\theta$  definida por la densidad  $f(\theta)$  se la llama la distribución *a priori* o anterior a la obtención de una muestra. En esta tesis me voy a referir a ella como la distribución inicial.

Al obtener una muestra de  $X$ , sea esta  $z_n = (x_1, \dots, x_n)$ , nuestro conocimiento se modifica por consiguiente, la distribución de  $\theta$  se modifica con esta muestra. Según el teorema de Bayes tenemos:

$$f(\theta | z_n) \propto h(z_n | \theta) f(\theta),$$

donde  $h(z_n | \theta)$  es la distribución conjunta de la muestra y entonces,  $f(\theta | z_n)$  es la distribución de  $\theta$  después de obtener la muestra, o distribución *posterior*.

#### 1.4 Pérdida Esperada y Decisión Óptima

Supongamos entonces que el fenómeno aleatorio involucrado en el proceso de decisión está dado por una variable  $X$  con las características de 1.3. Definimos entonces a:

$$L: D \times \Theta \rightarrow R,$$

que sería la función de pérdida. Es decir, dados  $d \in D$  y  $\theta \in \Theta$ ,  $L(d, \theta)$  evalúa la pérdida obtenida al tomar la decisión  $d$  si el estado es  $\theta$ .

<sup>1</sup> Nos referimos a una función de densidad generalizada de  $\theta$  como una función real positiva  $f(\theta)$  con las siguientes características: (las integrales son en el sentido de Lebesgue)

$$\int_0^1 f(\theta) d\mu(\theta) = 1$$

$$P(\theta < a) = \int_{(\theta < a)} f(\theta) d\mu(\theta),$$

donde  $\mu(\theta)$  es una medida en  $R$  con respecto a la cual  $f(\theta)$  es densidad y entonces, si la medida es la de Lebesgue en  $R$ ,  $f(\theta)$  es una función de densidad común, si es la de conteo en  $R$  entonces  $f(\theta)$  es una función de probabilidad, etc.

Tenemos ahora la función de pérdida  $L$ ; como  $L$  es función de  $\theta$  entonces, dada  $d \in D$ ,  $L(d, \theta)$  es una variable aleatoria y su distribución depende directamente de la distribución de  $\theta$ . Entonces definimos a<sup>2</sup>:

$$L^*(d) = E_{\theta}[L(d, \theta)] = \int_{\Theta} L(d, \theta) f(\theta) d\mu(\theta),$$

que sería la pérdida esperada para la decisión  $d$ , y a:

$$L^*(d | z_n) = E_{\theta}[L(d, \theta) | z_n] = \int_{\Theta} L(d, \theta) f(\theta | z_n) d\mu(\theta)$$

que sería la pérdida esperada después de obtener la muestra de  $X$ , bajo la decisión  $d$ .

Puede consultarse la bibliografía mencionada para revisar que  $d^* \in D$  es la decisión óptima si:

$$L^*(d^* | z_n) = \min_{d \in D} L^*(d | z_n),$$

o sea,  $d^*$  minimiza la pérdida esperada después de obtener la muestra. Puede también considerarse la decisión que minimice  $L^*(\cdot)$  si es que no vamos a obtener una muestra y confiamos completamente en nuestra información inicial, pero en el problema de pruebas de hipótesis paramétricas, que es del que nos vamos a ocupar en esta tesis, se considerarán sólo las decisiones después de incluir una muestra en el análisis.

<sup>2</sup> Las integrales que aparecen son en el sentido de Lebesgue y  $\mu(\theta)$  es la medida apropiada para la densidad de  $\theta$  (la de Lebesgue en  $\mathbb{R}$ , la de conteo, etc.).

## Capítulo 2

### Contraste de Hipótesis

## Capítulo 2

### Contraste de Hipótesis

#### 2.1 Definiciones

Una hipótesis es una proposición que está sujeta a verificación. Considérese una variable aleatoria como en 1.3, una hipótesis paramétrica es una proposición acerca del comportamiento del parámetro  $\theta$  que, como ya vimos, lo vamos a considerar una variable aleatoria. En concreto, una hipótesis paramétrica queda determinada por un conjunto  $\Theta_1 \subset \Theta$  (subconjunto del espacio parametral) y se estipula de la manera siguiente:

$$H_1 : \theta \in \Theta_1.$$

En esta tesis sólo vamos a estudiar las hipótesis paramétricas (desde el punto de vista de la estadística Bayesiana). El cómo se estipulan las hipótesis resulta ser un problema muy controvertido en casi todas las ramas de la ciencia y ha generado grandes discusiones, este problema no será abordado en esta tesis por estar más allá de los alcances de la misma; siempre vamos a partir de que ya están propuestas las hipótesis. Ahora, ¿para qué se estipulan las hipótesis? Dentro de la literatura de estadística Bayesiana existen dos tendencias principales. En una de estas tendencias se resuelven los problemas al inferir, mientras que en la otra el problema es tomar una decisión. Las pruebas de hipótesis no pueden quedar al margen de estos dos puntos de vista, y es por eso que los métodos que vamos a estudiar en el punto 2.2 están motivados por éstos, viendo a las hipótesis como un problema de inferencia o como un problema de decisión. Las hipótesis son planteadas entonces, como guía para hacer inferencias, plantean lo que se ha de inferir; o definir sobre lo que se ha de decidir, definir el espacio de acciones, como se verá adelante.

Para nuestro estudio necesitamos las siguientes definiciones: Una hipótesis es *simple* si, dada ésta, la distribución de la variable cuyo parámetro está involucrado en la hipótesis, queda totalmente determinada. Si consideramos una variable aleatoria como en 1.3, entonces una hipótesis simple es tal que el subconjunto del espacio parametral con la que queda estipulada contiene un solo elemento; en la notación anterior,  $H_1$  es simple si  $\Theta_1 = \{\theta_1\}$  o:

$$H_1 : \theta = \theta_1,$$

con  $\theta_1 \in \Theta$ . Una hipótesis es *compuesta* si no es simple.

### 2.1.1 Contraste de Hipótesis

Considérese, de nuevo, una variable aleatoria como en 1.3. Un contraste de hipótesis (paramétricas) queda estipulado por una hipótesis paramétrica sobre  $\Theta$  y su complemento. Sea  $\Theta_1 \subset \Theta$ , entonces un contraste de hipótesis es:

$$H_1 : \theta \in \Theta_1$$

$$H_2 : \theta \in \Theta_2$$

donde,  $\Theta_2 = \Theta - \Theta_1$  y  $\Theta$  es el espacio parametral.

Seguindo la idea del punto 1.3, sabemos que  $X \in F_\theta$  y que  $\theta \in \Theta$  y entonces proponemos que  $\theta \in \Theta_1$ . Desde luego, si  $\theta \notin \Theta_1$  entonces  $\theta \in \Theta_2$ . Parece ocioso afirmar lo anterior pero, si no tuviésemos bien estipulado el espacio parametral  $\Theta$ , no podríamos afirmar qué sucede si rechazamos  $H_1$ . Entonces, si después de hacer un análisis tendiente a verificar la hipótesis  $H_1$  resulta que la rechazamos, tenemos que la hipótesis  $H_2$  (no  $H_1$ ) es más conveniente. Tenemos pues, que al proponer  $H_1$  lo que en realidad vamos a hacer es comparar o *contrastar* entre  $H_1$  y  $H_2$  cuál es más conveniente.

Si por ejemplo, la familia de distribuciones es:  $F_\theta = \{\text{Distribución Normal con varianza}=1 \text{ y media } \theta\}$ , y el espacio parametral es  $\Theta = R$ , un ejemplo de contraste de hipótesis puede ser  $\Theta_1 = \{\theta > 0\}$  y:

$$H_1 : \theta > 0$$

$$H_2 : \theta \leq 0.$$

Ahora bien, la familia podría ser  $F_\theta = \{\text{Distribución Normal con media}=0 \text{ y varianza}=1 \text{ si } \theta = 1 \text{ y distribución } t \text{ con } 23 \text{ grados de libertad si } \theta = 2\}$ , el espacio parametral puede ser  $\Theta = \{1, 2\}$  y un contraste de hipótesis sería:

$$H_1 : \theta = 1$$

$$H_2 : \theta = 2.$$

Estos son sólo algunos ejemplos de contrastes de hipótesis que como se puede observar, pueden tomar muchos matices.

### 2.1.2 Contrastes de Hipótesis Múltiples

Comunmente las hipótesis en estadística se han planteado como lo expuesto en el punto anterior sin embargo, en ciertos casos podría ser muy útil tener más de dos posibilidades para describir el comportamiento de  $\theta$ . A lo que nos referimos es que también podrían considerarse hipótesis como las siguientes: Se considera una

familia disjunta de subconjuntos de  $\Theta$ , sea esta  $\Theta_i$ ;  $i = 1, \dots, h$  con  $\cup \Theta_i = \Theta$ , y se establecen las siguientes hipótesis:

$$H_i: \theta \in \Theta_i; \quad i = 1, \dots, h.$$

Lo que buscamos entonces es averiguar cual de las  $h$  hipótesis es más conveniente. En los puntos siguientes vamos a estar considerando estos tipos de contrastes de hipótesis como generalización de los contrastes con sólo dos posibilidades.

## 2.2 Métodos para Contrastar Hipótesis

A partir de este punto vamos a estudiar las técnicas para abordar el problema estadístico del contraste de hipótesis. En estas secciones se va a abordar el problema en forma general, y vamos a estudiar algunos ejemplos concretos.

### 2.2.1 Comparación de Probabilidades

El método de comparación de probabilidades es muy simple (véase Berger 1980, p.104), una vez estipulada la distribución inicial, se toma una muestra y se calcula la densidad generalizada final,  $f(\theta | z_n)$ . Después procedemos a calcular:

$$P(H_i | z_n) = P(\theta \in \Theta_i | z_n) = \int_{\Theta_i} f(\theta | z_n) d\mu(\theta).$$

con  $i = 1, \dots, h$  y entonces nos inclinamos por la hipótesis con mayor probabilidad asociada. El método de comparación de probabilidades está basado en que cada hipótesis se le puede asignar una probabilidad llamando  $P(H_i) = P(\theta \in \Theta_i)$ . La probabilidad de la hipótesis  $H_i$  debe de entenderse como la probabilidad de que el parámetro involucrado pertenezca al conjunto  $\Theta_i$  correspondiente o sea, la probabilidad de que el parámetro se comporte según la hipótesis  $H_i$ ; en este sentido,  $P(H_i)$  mide la *validez* o *verosimilitud* de la hipótesis  $H_i$ .

Cuando  $h = 2$  es muy común utilizar el cociente:

$$\frac{P(H_1 | z_n)}{P(H_2 | z_n)}$$

para comparar las probabilidades de las hipótesis, siendo la regla de decisión el preferir  $H_1$  si el cociente es mayor que 1 y al mismo tiempo observar que tan grande es una probabilidad con respecto a la otra.

Recordemos de 2.1.1 que había dos principales enfoques para realizar un contraste de hipótesis, el método expuesto en este punto está dirigido para cuando sólo queremos inferir, y no para tomar decisiones pues no tomamos en cuenta las consecuencias de preferir cualquiera de las hipótesis.

Ahora vamos a estudiar algunos ejemplos de contrastes de hipótesis utilizando el método de comparación de probabilidades. Un primer caso es cuando tenemos una hipótesis simple con otra simple; entonces,  $\Theta = \{\theta_1, \theta_2\}$  y  $\Theta_1 = \{\theta_1\}$ ,  $\Theta_2 = \{\theta_2\}$ . De esta manera, tenemos que el contraste de hipótesis está dado por:

$$H_1 : \theta = \theta_1$$

$$H_2 : \theta = \theta_2.$$

La distribución inicial  $f(\theta)$  se define como  $f(\theta_1) = p$  y  $f(\theta_2) = q$  con  $p + q = 1$ . Para preferir a  $H_1$  tenemos que:

$$P(H_1 | z_n) > P(H_2 | z_n)$$

ó

$$f(\theta_1 | z_n) > f(\theta_2 | z_n),$$

como lo vimos en el capítulo anterior  $f(\theta | z_n) = kh(z_n | \theta)f(\theta)$  con  $k$  constante positiva, entonces si:

$$f(\theta_1 | z_n) > f(\theta_2 | z_n)$$

ó

$$kh(z_n | \theta_1)f(\theta_1) > kh(z_n | \theta_2)f(\theta_2)$$

ó

$$\frac{h(z_n | \theta_1)}{h(z_n | \theta_2)} > \frac{q}{p},$$

preferimos a  $H_1$ .

**Ejemplo 1:** Si  $F_\theta = \{\text{Familia de distribuciones Bernoulli con parámetro } \theta\}$  con  $\Theta = \{\theta_1, \theta_2\} \subset (0, 1)$ . Vamos a suponer que  $\theta_1 < \theta_2$ , las hipótesis son:

$$H_1 : \theta = \theta_1$$

$$H_2 : \theta = \theta_2.$$

Si  $z_n = (x_1, \dots, x_n)$  es una muestra aleatoria, entonces:

$$h(z_n | \theta) \propto \theta^{\sum x_i} (1 - \theta)^{n - \sum x_i}$$

de donde,

$$\frac{h(z_n | \theta_1)}{h(z_n | \theta_2)} = \left(\frac{\theta_1}{\theta_2}\right)^{\sum x_i} \left(\frac{1 - \theta_1}{1 - \theta_2}\right)^{n - \sum x_i}$$

Entonces preferimos a  $H_1 : \theta = \theta_1$  si:

$$\left(\frac{\theta_1}{\theta_2}\right)^{\sum x_i} \left(\frac{1 - \theta_1}{1 - \theta_2}\right)^{n - \sum x_i} > \frac{q}{p},$$



aplicando logaritmo a ambas partes de la desigualdad, tenemos:

$$\sum x_i \ln \left( \frac{\theta_1}{\theta_2} \right) + (n - \sum x_i) \ln \left( \frac{1-\theta_1}{1-\theta_2} \right) > \ln \left( \frac{q}{p} \right)$$

ó

$$\sum x_i (\ln \left( \frac{\theta_1}{\theta_2} \right) - \ln \left( \frac{1-\theta_1}{1-\theta_2} \right)) > \ln \left( \frac{q}{p} \right) - n \ln \left( \frac{1-\theta_1}{1-\theta_2} \right).$$

En el lado izquierdo de la desigualdad, tenemos:

$$\ln \left( \frac{\theta_1(1-\theta_2)}{\theta_2(1-\theta_1)} \right) \sum x_i,$$

el logaritmo es positivo si y sólo si su argumento es mayor que 1, entonces el logaritmo del cociente es positivo si y sólo si:

$$\left( \frac{\theta_1(1-\theta_2)}{\theta_2(1-\theta_1)} \right) > 1$$

ó,

$$\frac{(1-\theta_2)}{\theta_2} > \frac{(1-\theta_1)}{\theta_1}$$

ó,

$$\frac{1}{\theta_2} - 1 > \frac{1}{\theta_1} - 1$$

ó,

$$\frac{1}{\theta_2} > \frac{1}{\theta_1}$$

ó,

$$\theta_1 > \theta_2.$$

Pero supusimos que  $\theta_1 < \theta_2$  entonces al dividir por el logaritmos se invierte la desigualdad y entonces preferimos a  $H_1$  si:

$$\sum x_i < \frac{\ln \left( \frac{q}{p} \right) - n \ln \left( \frac{1-\theta_1}{1-\theta_2} \right)}{\ln \left( \frac{\theta_1(1-\theta_2)}{\theta_2(1-\theta_1)} \right)}$$

y a  $H_2$  en otro caso. Preferir  $H_1$  significa preferir  $\theta_1$  sobre  $\theta_2$ , siendo  $\theta_1 < \theta_2$ ; entonces, preferimos al valor menor si  $\sum x_i$  es menor que una constante y al valor mayor si  $\sum x_i$  es mayor que esta constante, lo cual está de acuerdo con nuestra intuición, siendo que  $\theta$  determina la probabilidad de que  $X = 1$  y entre mayor es  $\theta$  se espera que  $\sum x_i$  sea mayor. \*

**Ejemplo 2:**  $F_\theta = \{\text{Familia de distribuciones Poisson con parámetro } \theta\}$ .  $\Theta = \{\theta_1, \theta_2\} \subset \mathbb{R}^+$ . También aquí vamos a suponer que  $\theta_1 < \theta_2$ . Las hipótesis son:

$$H_1: \theta = \theta_1$$

$$H_2: \theta = \theta_2.$$

Si  $z_n = (x_1, \dots, x_n)$  es una muestra aleatoria, entonces:

$$h(z_n | \theta) \propto \frac{\theta^{\sum x_i} \exp(-n\theta)}{x_1! x_2! \dots x_n!}$$

y

$$\frac{h(z_n | \theta_1)}{h(z_n | \theta_2)} = (\theta_1/\theta_2)^{\sum x_i} \exp(n(\theta_2 - \theta_1)).$$

Preferimos  $H_1$  si:

$$\left(\frac{\theta_1}{\theta_2}\right)^{\sum x_i} \exp(n(\theta_2 - \theta_1)) > \frac{q}{p},$$

aplicamos logaritmo en ambos lados de la desigualdad tenemos:

$$\sum x_i \ln\left(\frac{\theta_1}{\theta_2}\right) + n(\theta_2 - \theta_1) > \ln\left(\frac{q}{p}\right),$$

como  $\theta_1 < \theta_2$  entonces  $\frac{\theta_1}{\theta_2} < 1$  y  $\ln\left(\frac{\theta_1}{\theta_2}\right) + n(\theta_2 - \theta_1) < 0$ ; tenemos que preferimos a  $H_1$  si:

$$\sum x_i < \frac{\ln\left(\frac{q}{p}\right)}{\ln\left(\frac{\theta_1}{\theta_2}\right) + n(\theta_2 - \theta_1)}$$

y a  $H_2$  en otro caso. •

**Ejemplo 3:**  $F_\theta = \{\text{Familia de distribuciones Normal con varianza}=\sigma^2 \text{ y media } \theta\}$  (en este caso  $\sigma^2$  es una constante conocida).  $\Theta = \{\theta_1, \theta_2\} \subset \mathbb{R}$ . Vamos a suponer que  $\theta_1 < \theta_2$  y las hipótesis entonces son:

$$H_1: \theta = \theta_1$$

$$H_2: \theta = \theta_2.$$

Si  $z_n = (x_1, \dots, x_n)$  es una muestra aleatoria, entonces:

$$h(z_n | \theta) = k \exp\left(-\frac{1}{2\sigma^2} \sum (x_i - \theta)^2\right)$$

y

$$\frac{h(z_n | \theta_1)}{h(z_n | \theta_2)} = \exp\left\{\left(\frac{1}{2}\right)\sigma^{-2}\left(\sum((x_i - \theta_2)^2 - (x_i - \theta_1)^2)\right)\right\},$$

con un cálculo sencillo podemos ver que:

$$\frac{h(z_n | \theta_1)}{h(z_n | \theta_2)} = \exp\left\{\left(\frac{1}{2}\right)\sigma^{-2}(n(\theta_2^2 - \theta_1^2) + \sum 2x_i(\theta_1 - \theta_2))\right\}.$$

Preferimos a  $H_1$  si:

$$\exp\left\{\left(\frac{1}{2}\right)\sigma^{-2}(n(\theta_2^2 - \theta_1^2) + (\theta_1 - \theta_2)2 \sum x_i)\right\} > \frac{q}{p}$$

ó

$$(n/2)(\theta_2^2 - \theta_1^2) + (\theta_1 - \theta_2) \sum x_i > \sigma^2 \ln(q/p)$$

ó

$$(\theta_1 - \theta_2) \sum x_i > \sigma^2 \ln(q/p) - (n/2)(\theta_2^2 - \theta_1^2).$$

Como  $\theta_1 - \theta_2 < 0$  entonces preferimos a  $H_1$  si:

$$\sum x_i < \frac{\sigma^2 \ln(q/p) - (n/2)(\theta_2^2 - \theta_1^2)}{(\theta_1 - \theta_2)}$$

y preferimos a  $H_2$  en otro caso. Si  $p = q$  o sea,  $H_1$  y  $H_2$  tienen la misma probabilidad inicial entonces  $\ln(q/p) = 0$  y preferimos a  $H_1$  si:

$$\sum x_i < \frac{n(\theta_1^2 - \theta_2^2)}{2(\theta_1 - \theta_2)}$$

ó

$$\bar{X} < \frac{(\theta_1^2 - \theta_2^2)}{2(\theta_1 - \theta_2)},$$

utilizando diferencia de cuadrados para  $(\theta_1^2 - \theta_2^2)$  tenemos que preferimos la hipótesis  $H_1$  en este caso si:

$$\bar{X} < \frac{\theta_1 + \theta_2}{2}.$$

Nótese que en este caso, el preferir una u otra hipótesis no depende del valor de  $\sigma^2$ . •

Otro caso sería cuando  $\Theta$ , el espacio parametral, consta de un número a lo más numerable de elementos; en este caso la distribución de  $\theta$ ,  $f(\theta)$ , es una distribución con  $\mu(\theta)$  la medida de conteo o sea,  $f(\theta)$ , es una distribución discreta. Todas las hipótesis que se pueden proponer en este caso son a lo más numerables<sup>1</sup> y entonces procedemos a comparar probabilidades como ha sido estipulado antes y las integrales que se manejan se calculan como sumas o series. Este caso es poco común y no hacemos ningún ejemplo al respecto.

El siguiente caso es cuando  $\Theta$  tiene la cardinalidad de  $R$ , y tenemos un contraste de hipótesis compuestas. Lo usual es que  $\lambda(\Theta) > 0$ <sup>2</sup> y vamos a suponer que  $\lambda(\Theta_i) > 0$  con  $i = 1, \dots, h$ , esto quiere decir que  $\Theta$  y  $\Theta_i$  son conjuntos no sólo con más de un elemento (pues hablamos de hipótesis compuestas) si no que son conjuntos *gordos*, como intervalos o vecindades de  $R$ . Podemos ahora utilizar una distribución continua para  $\theta$  o, en otras palabras,  $f(\theta)$  sería una distribución con respecto a la medida de Lebesgue en  $R$  (como por ejemplo, la Beta, la Normal, la F, etc.) que pueden servirnos para modelar el conocimiento que del parámetro tenga el interesado en llevar a cabo el contraste de hipótesis. Se continúa entonces con el procedimiento usual de comparación de probabilidades calculando  $P(H_i | z_n)$  con  $i = 1 \dots h$  donde tal vez haya problemas de integración.

**Ejemplo 4:**  $F_\theta = \{\text{Familia de distribuciones Bernoulli con parámetro } \theta\}$ , con  $h = 2$ ,  $\Theta = (0, 1)$  y  $\Theta_1 = (0, \theta_0)$  y entonces  $\Theta_2 = [\theta_0, 1]$ ; en otras palabras, las hipótesis que manejamos son:

$$H_1 : \theta < \theta_0$$

$$H_2 : \theta \geq \theta_0.$$

Tenemos que:

$$h(z_n | \theta) \propto \theta^{\sum x_i} (1 - \theta)^{n - \sum x_i}.$$

La distribución conjugada<sup>3</sup> para  $\theta$  es:

$$f(\theta) = \theta^{a-1} (1 - \theta)^{b-1} \chi_{(0,1)}(\theta)$$

<sup>1</sup> Al decir que una hipótesis es numerable o finita o con dos elementos etc. estamos hablando de la cardinalidad del conjunto  $\Theta$ ; correspondiente. Vamos a utilizar esta convención a partir de este punto.

<sup>2</sup>  $\lambda(\Theta)$ , la medida de Lebesgue en  $R$ .

<sup>3</sup> En este ejemplo y en el resto de la tesis vamos a utilizar las distribuciones conjugadas para la distribución del parámetro. Las distribuciones conjugadas facilitan mucho la obtención de la distribución posterior del parámetro. Si el conocimiento del que plantea el contraste de hipótesis (debe, en su caso) no queda bien modelado con la distribución conjugada se recomienda utilizar combinaciones lineales de distribuciones conjugadas o inclusive mezclas de éstas, que siguen siendo muy fáciles de manejar, para obtener la distribución inicial que sea conveniente; véase el artículo de Dalal y Hall (1983) para la definición de mezclas de distribuciones conjugadas y una demostración de que cualquier distribución inicial puede ser aproximada por una mezcla de conjugadas de la familia exponencial.

que es una Beta con parámetros  $a$  y  $b$  ( $\beta(a, b)$ ). Suponiendo que se utiliza ésta como distribución inicial, se tiene que: calcular  $f(\theta | z_n)$ :

$$f(\theta | z_n) \propto \theta^{(\sum x_i + a) - 1} (1 - \theta)^{(n - \sum x_i + b) - 1}$$

que es una distribución Beta con parámetros:

$$a' = (\sum x_i + a)$$

$$b' = (n - \sum x_i + b)$$

y entonces,

$$P(H_1 | z_n) = P[\theta < \theta_0 | \theta \sim \beta(a', b')]$$

y

$$P(H_2 | z_n) = 1 - P(H_1 | z_n).$$

Puede utilizarse entonces algún programa de cómputo que mediante integración numérica calcule  $P(H_i | z_n)$ . Por ejemplo el paquete STATGRAPHICS, ya que la integración de la densidad Beta en general, tiene que ser en forma numérica. •

**Ejemplo 5:** Sea  $F_\theta = \{ \text{Familia de distribuciones Normal con precisión } t \text{ y media } \theta, (N(\theta, t)) \}$  (en este caso  $t$  es una constante conocida). Sea  $h = 3$ ,  $\Theta = R$ ,  $\Theta_1 = (-\infty, \theta_a]$ ,  $\Theta_2 = (\theta_a, \theta_b]$  y  $\Theta_3 = (\theta_b, \infty)$ . Tenemos entonces que las hipótesis que manejamos son las siguientes:

$$H_1 : \theta \leq \theta_a$$

$$H_2 : \theta_a < \theta \leq \theta_b$$

$$H_3 : \theta > \theta_b.$$

La distribución conjugada para  $\theta$  es una Normal. Vamos a suponer que para modelar el conocimiento inicial del interesado vamos a utilizar una distribución Normal  $N(\mu_0, t_0)$  para  $\theta$ . Puede demostrarse que (véase Drunk 1979, p.323):

$$\theta | z_n \sim N(\mu', t')$$

con

$$t' = t_0 + nt$$

y

$$\mu' = \frac{nt \sum x_i + t_0 \mu_0}{t_0 + nt}.$$

Para calcular  $P(H_i | z_n)$  ( $i = 1, 2, 3$ ), pueden utilizarse tablas de la distribución  $N(0,1)$ , puesto que:

$$\sqrt{i}(\theta - \mu^i) \sim N(0,1). \bullet$$

Analicemos ahora los contrastes de hipótesis simples contra compuestas. Si el espacio parametral es numerable, se utiliza una distribución discreta para  $\theta$  y se procede como antes, si por el contrario tenemos que  $\Theta_1 = \{\theta_1\}$  y  $\Theta_2 = R - \Theta_1$  podemos utilizar una distribución continua para  $\theta$  y tendríamos que siempre nos inclinariamos por  $H_2$ , pues la medida de Lebesgue de  $\Theta_1$  es cero y entonces  $P(H_1 | z_n) = 0$ . Sin embargo, es lo más común en la literatura en estadística Bayesiana encontrarse con un procedimiento alternativo (véase por ejemplo, Jeffreys 1961, Box y Tiao 1973, DeGroot 1970) para probar estas hipótesis. Se argumenta en estos casos que en ciertos experimentos, según el conocimiento del experimentador, hay un valor del parámetro que es especialmente distinguido entre los demás y el conocimiento del interesado no queda bien modelado utilizando una distribución continua para  $\theta$ . El procedimiento es utilizar una medida en  $R$  de la siguiente manera: Para todo  $A \subset R$  Lebesgue medible,

$$\mu(\theta)(A) = \lambda(A) + \chi_A(\theta_1)$$

donde  $\lambda$  es la medida de Lebesgue en  $R$  y  $\chi$  la función indicadora (entiéndase  $\mu(\theta)(A)$  como la medida de  $A$  según  $\mu(\theta)$ , en el apéndice A en el punto A.1, hay una explicación más amplia de  $\mu(\theta)$ ). Entonces, se da una probabilidad  $p$  inicial para  $H_1$  y  $f(\theta_1) = p$ . La función  $f(\theta)$  integra (según Riemann)  $1 - p$  en  $\Theta$ . Entonces, con un cálculo simple podemos ver que:

$$\int_{\Theta} f(\theta) d\mu(\theta) = 1$$

y se genera una distribución de probabilidad que se le suele llamar distribución *discontinua*. Después se procede a tomar una muestra y a comparar probabilidades. Se evita entonces el problema de que  $P(H_1 | z_n)$  es siempre cero (Nótese que  $P(H_1) = p$ , antes de tomar la muestra).

Las opiniones sobre estos contrastes de hipótesis son diversas. Por ejemplo, en Phillips (1973) en el capítulo 14 dedicado al contraste de hipótesis, menciona que conceptualmente estas hipótesis no tienen mucho sentido (dentro de la estadística Bayesiana) sin embargo dedica una buena exposición al respecto. Hay que darnos cuenta que el procedimiento matemáticamente es correcto y que su utilización depende de interpretaciones y concepciones de muchos tipos, en el punto 2.2.3 y en el capítulo 3 discutimos más detenidamente este tipo de contrastes. No presentaremos ejemplos con distribuciones discontinuas.

### 2.2.2 Problema de Decisión

Veamos ahora lo que dice DeGroot (1970, p.237) de este problema

"Problemas que involucran sólo dos decisiones son conocidos como *problemas de contraste de hipótesis*"

En síntesis, el problema de decisión se plantea como menciona DeGroot. Para observar a un problema de contraste de hipótesis como un problema de decisión, lo que se hace es considerar a  $D$  (el espacio de acciones) constituido por las acciones de decidirnos por cada una de las hipótesis planteadas. En un problema de contraste de hipótesis sólo tenemos dos hipótesis y es por eso que en la referencia se dice que un problema con sólo dos acciones es un problema de contraste de hipótesis. Tenemos entonces las hipótesis  $H_i$  ( $i = 1 \dots h$ ) y consideramos el preferir la hipótesis  $H_i$  como la acción  $d_i$ . Tenemos entonces:

$$D = \{d_1, d_2, \dots, d_h\},$$

con  $d_i$  equivalente a decidirse por  $H_i$ . Lo más importante de este enfoque es que, como lo vimos en el capítulo 1, en teoría de decisiones tenemos la posibilidad de incluir funciones de pérdida en el análisis. Esto significa que vamos a considerar los inconvenientes de tomar una decisión o sea, de preferir alguna hipótesis, dado que el valor del parámetro fuera  $\theta_0$ . Comúnmente se suele hablar, al contrastar hipótesis, de dos tipos de errores en los que se puede incurrir; un error es preferir  $H_1$  siendo que  $\theta \in \Theta_2$  y el otro error es preferir  $H_2$  siendo que  $\theta \in \Theta_1$ . Al utilizar las funciones de pérdida, implícitamente estamos considerando en nuestro análisis el incurrir en estos errores, pero también tenemos la posibilidad de medir cuanto perdemos al cometer cada uno de los errores. Esa es la ventaja de tratar al problema de contraste de hipótesis como un problema de decisión. Ahora introducimos la siguiente notación:

$$L_i(\theta) = L(d_i, \theta),$$

y también:

$$L_i^* = L^*(d_i | z_n) = \int_{\Theta} L_i(\theta) f(\theta | z_n) d\mu(\theta).$$

Entonces, por lo que vimos en 1.4, la decisión óptima es la  $d_i$  correspondiente a la  $L_i^*$ ; mínima. Si  $h = 2$ , nos decidimos por  $H_1$  si:

$$L_1^* < L_2^*,$$

que sería la regla de decisión óptima para los contrastes de hipótesis.

El método que hemos estudiado en este punto, como me parece evidente, nos sirve para cuando el problema es tomar una decisión. Me parece importante aclarar lo anterior pues es muy valioso siempre tener presente a que tipo

de problema de contraste de hipótesis nos enfrentamos, y así utilizar las técnicas adecuadas.

Ahora veamos distintos casos de contrastes de hipótesis, en el más sencillo tenemos una hipótesis simple contra otra simple. Tenemos que  $\Theta_1 = \{\theta_1\}$  y  $\Theta_2 = \{\theta_2\}$ . Para la distribución inicial tenemos que  $f(\theta_1) = p$  y  $f(\theta_2) = q$  con  $p + q = 1$ . Para las funciones de pérdida podemos observar lo siguiente:

$$(1) \quad \begin{cases} L_1(\theta_1) = 0, \\ L_1(\theta_2) = a, \\ L_2(\theta_1) = b, \\ L_2(\theta_2) = 0. \end{cases}$$

La razón por la cual  $L_i(\theta_i) = 0$ , es que normalmente no hay pérdida si no hay error (la decisión  $d_i$  es preferir el valor del parámetro  $\theta_i$ , y si nos decidimos por  $d_i$  siendo el valor del parámetro  $\theta_i$  pues no hay error). Ahora,  $a$  y  $b$  tienen que ser mayores que cero porque si, por ejemplo,  $a$  es negativa o cero, de inmediato (sin molestarnos en tomar una muestra) nos decidiríamos por  $d_1$  pues decidirse por  $d_2$  sólo implica pérdidas y equivalentemente observaríamos respecto a  $b$ . Podemos observar que:

$$L_1^* = L_1(\theta_2)f(\theta_2 | z_n) = af(\theta_2 | z_n)$$

y

$$L_2^* = L_2(\theta_1)f(\theta_1 | z_n) = bf(\theta_1 | z_n).$$

Como vimos en el capítulo 1,  $f(\theta | z_n) = kh(z_n | \theta)f(\theta)$  con  $k$  una constante positiva. Para decidirnos por  $H_1$  tenemos que:

$$L_1^* < L_2^*$$

ó

$$akh(z_n | \theta_2)f(\theta_2) < bkh(z_n | \theta_1)f(\theta_1)$$

ó

$$\frac{h(z_n | \theta_2)}{h(z_n | \theta_1)} < \frac{bp}{aq}$$

y entonces, la decisión óptima es decidirse por  $H_1$  si el cociente de verosimilitudes de la izquierda es menor que la constante de la derecha, y por  $H_2$  en otro caso<sup>4</sup>.

<sup>4</sup>Esta regla de decisión tiene una gran similitud con el teorema de la estadística clásica conocido como el lema fundamental de Neyman-Pearson (1933) véase por ejemplo, Drunk (1979 p.415).



**Ejemplo 1:** Sea  $F_\theta = \{\text{Familia de distribuciones Bernoulli con parámetro } \theta\}$ . El espacio parametral es  $\Theta = \{\theta_1, \theta_2\} \subset (0, 1)$ . Sin pérdida de generalidad suponemos que  $\theta_1 < \theta_2$  y que las hipótesis son:

$$H_1 : \theta = \theta_1$$

$$H_2 : \theta = \theta_2.$$

Si  $z_n = (x_1, \dots, x_n)$  es una muestra aleatoria, entonces:

$$h(z_n | \theta) = \theta^{\sum x_i} (1 - \theta)^{n - \sum x_i}$$

y,

$$\frac{h(z_n | \theta_2)}{h(z_n | \theta_1)} = \left(\frac{\theta_2}{\theta_1}\right)^{\sum x_i} \left(\frac{1 - \theta_2}{1 - \theta_1}\right)^{n - \sum x_i}$$

La regla de decisión es decidirse por  $H_1$  si:

$$\left(\frac{\theta_2}{\theta_1}\right)^{\sum x_i} \left(\frac{1 - \theta_2}{1 - \theta_1}\right)^{n - \sum x_i} < \frac{bp}{aq},$$

aplicando logaritmos tenemos:

$$\sum x_i \ln\left(\frac{\theta_2}{\theta_1}\right) + (n - \sum x_i) \ln\left(\frac{1 - \theta_2}{1 - \theta_1}\right) < \ln\left(\frac{bp}{aq}\right)$$

ó

$$\sum x_i \left( \ln\left(\frac{\theta_2}{\theta_1}\right) - \ln\left(\frac{1 - \theta_2}{1 - \theta_1}\right) \right) < \ln\left(\frac{q}{p}\right) - n \ln\left(\frac{1 - \theta_2}{1 - \theta_1}\right). \quad (2)$$

Expresamos el lado izquierdo de (2) como:

$$\sum x_i \ln\left(\frac{\theta_2(1 - \theta_1)}{\theta_1(1 - \theta_2)}\right),$$

Vimos en el ejemplo 1 de 2.2.1 que si  $\theta_1 < \theta_2$  entonces:

$$\left(\frac{\theta_1(1 - \theta_2)}{\theta_2(1 - \theta_1)}\right) < 1$$

eso quiere decir que:

$$\left(\frac{\theta_2(1 - \theta_1)}{\theta_1(1 - \theta_2)}\right) > 1.$$

Si entonces dividimos por el logaritmo de este cociente a (2) no se invierte la desigualdad y entonces la decisión óptima es  $H_1$  si:

$$\sum x_i < \frac{\ln\left(\frac{bp}{aq}\right) - n \ln\left(\frac{1 - \theta_2}{1 - \theta_1}\right)}{\left(\frac{\theta_2(1 - \theta_1)}{\theta_1(1 - \theta_2)}\right)}$$

y es  $H_2$  en otro caso. •

**Ejemplo 2:** Tenemos  $F_\theta = \{\text{Familia de distribuciones exponencial con parámetro } \theta\}$ . El espacio parametral es  $\Theta = \{\theta_1, \theta_2\} \subset R^+$ . Como antes, vamos a suponer que  $\theta_1 < \theta_2$ , las hipótesis son:

$$H_1: \theta = \theta_1$$

$$H_2: \theta = \theta_2.$$

Si  $z_n = (x_1, \dots, x_n)$  es una muestra aleatoria, entonces:

$$h(z_n | \theta) = k\theta^n \exp(-\theta \sum x_i),$$

de donde:

$$\frac{h(z_n | \theta_2)}{h(z_n | \theta_1)} = \left(\frac{\theta_2}{\theta_1}\right)^n \exp\{-(\theta_2 - \theta_1) \sum x_i\}.$$

Por lo tanto la decisión óptima es  $H_1$  si:

$$\frac{h(z_n | \theta_2)}{h(z_n | \theta_1)} = \left(\frac{\theta_2}{\theta_1}\right)^n \exp\{-(\theta_2 - \theta_1) \sum x_i\} < \frac{b\beta}{a\alpha}.$$

Al aplicar logaritmos tenemos:

$$n \log \left(\frac{\theta_2}{\theta_1}\right) + (\theta_1 - \theta_2) \sum x_i < \ln \left(\frac{b\beta}{a\alpha}\right),$$

como  $\theta_1 - \theta_2 < 0$  se invierte la desigualdad al dividir por  $(\theta_1 - \theta_2)$  y entonces la decisión óptima es  $H_1$  si:

$$\sum x_i > \frac{\ln \left(\frac{b\beta}{a\alpha}\right) - n \ln \left(\frac{\theta_2}{\theta_1}\right)}{(\theta_1 - \theta_2)}$$

y es  $H_2$  en otro caso. En este caso tenemos que decidirnos por el valor  $\theta_i$  mayor si  $\sum x_i$  es menor que una constante, esto se puede explicar intuitivamente ya que valores mayores para el parámetro implican que los valores más probables para  $X$  son muy cercanos a cero y por lo tanto,  $\sum x_i$  tendería a ser pequeña. •

En un caso más general, si  $\Theta$  es a lo más numerable se utiliza una distribución discreta para  $\theta$  y las integrales que se manejan se calculan con sumas o series.

En el caso de que  $\Theta$  tenga medida de Lebesgue mayor que cero, al utilizar distribuciones continuas nos podemos enfrentar a problemas serios de integración

y tener que utilizar aproximaciones por métodos numéricos. Si, por ejemplo, tratamos con hipótesis compuestas, calcular:

$$\int_{\Theta} L_i(\theta) f(\theta | z_n) d\lambda$$

puede ser muy complicado puesto que  $f(\theta | z_n)$  normalmente no se puede integrar en forma analítica, complicándose aun más al estar multiplicado por  $L_i(\theta)$ .

Hay muchos tipos de funciones de pérdida, lo más común es que se siga la idea expuesta en (1). Generalizando a (1) tenemos que:

$$L_i(\theta) = c_i \chi_{\Theta - \Theta_i}(\theta),$$

con  $c_i$  ( $i = 1 \dots h$ ) una constante (de pérdida) positiva y  $\chi$  la función indicadora. Lo que quiere decir la función  $L_i$  es que si  $\theta$  no está en  $\Theta_i$  entonces tenemos una pérdida de magnitud  $c_i$ . En otros casos, y por las necesidades del decisor, puede requerirse de funciones de pérdida más sofisticadas con la intención de medir mejor las consecuencias de nuestros errores. Como se suele proceder es no sólo midiendo si el parámetro está o no en el conjunto  $\Theta_i$  correspondiente si no, en su caso, que tan lejos está el parámetro del conjunto en cuestión. Un ejemplo de lo anterior es la siguiente función de pérdida: Supongase que se tiene  $\Theta = R$  y que  $h = 2$ , con  $\Theta_1 = V_r(\theta_1)$  una vecindad de  $\theta_1$  con radio  $r \in R$ , y  $\Theta_2 = \Theta - \Theta_1$ . Ahora:

$$L_1(\theta) = a | \theta - \theta_1 | \chi_{\Theta_2}(\theta),$$

con  $a$  una constante positiva. Este es sólo un ejemplo de una función de pérdida para hipótesis compuestas. Hay que tener siempre en cuenta que las funciones de pérdida deben de ser estipuladas por el decisor, como habíamos visto en el capítulo anterior, y por consiguiente pueden variar mucho de un problema a otro.

**Ejemplo 3:** En una universidad existe una red de computadora con muchas terminales. A la comisión de cómputo de la universidad han llegado muchas quejas acerca del servicio que presta el CPU (Unidad Central de Proceso) de la computadora; se dice en las quejas que el servicio se ha vuelto muy lento.

La comisión de cómputo de la universidad, el decisor, encarga un análisis del sistema. Deciden que cambiarán el CPU si el tiempo de respuesta esperado para que el CPU atienda a la terminal es mayor que  $t_a$  segundos, que es el tiempo de respuesta esperado especificado por el fabricante de la red. Se toma a  $X$  como la variable aleatoria que mide el tiempo de respuesta del CPU, el decisor coincide en que  $F_\theta = \{ \text{Familia de distribuciones exponencial con parámetro } \theta \}$ . Entonces  $X \in F_\theta$ ; podemos ver que  $E(X) = 1/\theta$ . Entonces el tiempo de respuesta esperado

es  $1/\theta$ . El problema puede plantearse como un contraste de hipótesis utilizando el método de teoría de decisiones. Las hipótesis que se plantean son:

$$H_1 : (1/\theta > t_a)$$

$$H_2 : (1/\theta \leq t_a),$$

y entonces decidirse por  $H_1$  es cambiar el CPU y por  $H_2$  es NO cambiar el CPU. El espacio parametral es  $\Theta = R^+$  y  $\Theta_1 = (0, 1/t_a)$  y  $\Theta_2 = [1/t_a, \infty)$ . Las hipótesis, en términos de  $\theta$ , son:

$$H_1 : \theta < 1/t_a$$

$$H_2 : \theta \geq 1/t_a.$$

Suponga que las quejas que ha recibido la comisión de cómputo sobre el CPU indican que por cada segundo de más que espera un usuario en respuesta del CPU, la universidad pierde  $S$  pesos y existen más o menos  $m$  usuarios<sup>5</sup>. La comisión de cómputo utilizando esta información estipula la siguiente función de pérdida para la decisión  $H_2$ :

$$L_2(t_e) = Sm(t_e - t_a)\chi_{(t_a, \infty)}(t_e),$$

donde  $t_e$  es el tiempo de respuesta esperado o sea,  $1/\theta$ . Esta es la pérdida esperada si no se cambia el CPU siendo que debería cambiarse y  $(t_e - t_a)$  modela el tiempo que espera de más un usuario. Lo anterior en términos del parámetro  $\theta$  nos da:

$$L_2(\theta) = Sm((1/\theta) - t_a)\chi_{(0, 1/t_a)}(\theta).$$

El costo de cambiar el CPU, incluida la instalación y el costo del tiempo en que no habrá servicio en la red durante la instalación, es de  $r$  pesos. La comisión elabora esta otra función de pérdida para la decisión  $H_1$ :

$$L_1(t_e) = r\chi_{(0, t_a)}(t_e).$$

Esto, expresado en el parámetro  $\theta$  es:

$$L_1(\theta) = r\chi_{[1/t_a, \infty)}(\theta).$$

La distribución conjugada de  $\theta$  es una Gama,  $G(a, b)$ . Después de una amplia discusión al interior de la comisión se coincide en utilizar una distribución

<sup>5</sup> En este ejemplo, por sencillez, no estamos considerando cuantas llamadas hace cada usuario al procesador ni tampoco un estudio a lo largo de un periodo de tiempo dado (semanas, meses, etc.)

Gama para modelar el conocimiento inicial que tiene la comisión acerca de  $\theta$ , entonces:

$$f(\theta) \propto \theta^{a-1} \exp(-\theta/b).$$

En las especificaciones del fabricante se estipula que  $t_a$  (el tiempo de respuesta esperado) es de 2.3 segundos. Sabemos que si  $\theta \sim G(a, b)$ , entonces  $E(\theta) = ab$ . La comisión de cómputo escogió  $b = 1$  y  $a = 1/2.3 = 0.4348$ . Se toma entonces una muestra aleatoria<sup>6</sup> de tamaño 100 (en el apéndice B aparece una tabla con los datos de la muestra). La densidad conjunta de la muestra  $z_n$  es:

$$h(z_n | \theta) = k\theta^n \exp(-\theta \sum x_i),$$

entonces la distribución posterior de  $\theta$  es:

$$f(\theta | z_n) \propto \theta^n \exp(-\theta \sum x_i) \theta^{a-1} \exp(-\theta/b),$$

ó

$$f(\theta | z_n) \propto \theta^{(n+a)-1} \exp\{-\theta(\sum x_i + 1/b)\}.$$

Entonces,

$$\theta | z_n \sim G(A, B)$$

con

$$A = n + a$$

$$B = \frac{1}{\sum x_i + 1/b}.$$

La distribución inicial de  $\theta$  es  $G(0.4348, 1)$ . Puede verse en el apéndice B que  $\sum x_i = 327.65$ ; después de hacer los cálculos necesarios tenemos que:

$$\theta | z_n \sim G(100.4348, 0.003).$$

Ahora sólo falta calcular las pérdidas esperadas para cada una de las hipótesis, tenemos que:

$$L_1^* = \int_{0.4348}^{\infty} r f(\theta | z_n) d\theta = r P\{Y > 0.4348 | Y \sim G(A, B)\},$$

<sup>6</sup> Para este ejemplo se realizó una simulación por computadora para obtener los valores de la muestra. El tiempo de respuesta promedio en la simulación es de 3.3 segundos.

llamemos  $\alpha_1 = P\{Y > 0.4348 \mid Y \sim G(A, B)\}$ , que después calcularemos. Para calcular la pérdida esperada  $L_2^*$  tenemos:

$$L_2^* = \int_0^{0.4348} Sm(1/\theta - 2.3)f(\theta \mid z_n)d\theta$$

entonces,

$$L_2^* = Sm \int_0^{0.4348} (1/\theta)f(\theta \mid z_n)d\theta - Sm2.3P\{Y < 0.4348 \mid Y \sim G(A, B)\}.$$

La probabilidad de la derecha es  $1 - \alpha_1$ , para calcular la integral de la izquierda tenemos:

$$\begin{aligned} Sm \int_0^{0.4348} (1/\theta)f(\theta \mid z_n)d\theta &= Sm \int_0^{0.4348} (1/\theta) \frac{\theta^{A-1} \exp(-\theta/B)}{\Gamma(A)B^A} d\theta \\ &= Sm \frac{1}{\Gamma(A)B^A} \int_0^{0.4348} \theta^{(A-1)-1} \exp(-\theta/B) d\theta, \end{aligned}$$

notando que esta integral es el núcleo de una densidad Gama, podemos entonces calcularla como una probabilidad multiplicando por la constante apropiada:

$$= \frac{Sm}{\Gamma(A)B^A} \Gamma(A-1)B^{A-1} P\{Y < 0.4348 \mid Y \sim G(A-1, B)\}.$$

Llamemos  $\alpha_2 = P\{Y < 0.4348 \mid Y \sim G(A-1, B)\}$ , sabemos que  $\Gamma(c) = (c-1)\Gamma(c-1)$ . Tenemos entonces que:

$$Sm \int_0^{0.4348} (1/\theta)f(\theta \mid z_n)d\theta = \alpha_2 Sm \frac{\Gamma(A-1)B^{A-1}}{(A-1)\Gamma(A-1)B^A} = \frac{\alpha_2 Sm}{(A-1)B}.$$

Finalmente tenemos que:

$$L_2^* = Sm \left\{ \frac{\alpha_2}{(A-1)B} - 2.3(1 - \alpha_1) \right\}.$$

Recordemos que la decisión óptima es decidirse por la hipótesis  $H_2$  (NO cambiar el CPU) si:

$$L_2^* < L_1^*$$

ó

$$Sm \left\{ \frac{\alpha_2}{(A-1)B} - 2.3(1 - \alpha_1) \right\} < \alpha_1 r,$$

para calcular  $\alpha_1$  y  $\alpha_2$  utilizamos el paquete STATGRAPHICS, resulta que  $1 - \alpha_1 = 0.999$ , entonces  $\alpha_1 = 10^{-3}$ ; por otro lado,  $\alpha_2 = 0.9999$ . Con un cálculo aproximado podemos ver que la comisión de cómputo NO cambia el CPU si:

$$r > (2,280)Sm.$$

Recordemos que  $r$  es costo del CPU nuevo, entonces NO se cambia el CPU si el valor del nuevo es mayor que la constante de la derecha. Supóngase que los que se quejan dicen que  $S$ , el costo de espera de un segundo de más, es de diez pesos, lo cual a la comisión le parece muy exagerado. Suponiendo que hay 2000 usuarios, haciendo los cálculos pertinentes tenemos que, NO cambiamos el CPU si este es más caro que \$45,600,000. El CPU nuevo es mucho más caro que eso. •

Veamos ahora que sucede con las hipótesis de simple contra compuesta. Como lo vimos al final de 2.2.1, si tenemos  $H_1 : \theta = \theta_1$  con  $H_2 : \theta \neq \theta_1$ , y  $\Theta$  con medida mayor que cero, podemos utilizar una distribución continua y resultaría que preferiríamos  $H_2$  independientemente de la muestra. Sin embargo, el conocimiento que el decisor tiene acerca de  $\theta$  puede requerir que  $P(H_1) > 0$ . Se utiliza entonces una distribución discontinua y se procede con las técnicas de teoría de decisiones. No comentaremos más al respecto pero podemos recomendar el artículo de Ferrandiz (1985) donde se expone otra forma de contrastar que  $\mu = 0$  (la media de una distribución normal) con  $\mu \neq 0$  utilizando distribuciones discontinuas. Lo importante del artículo es que se hace depender a la función de utilidad (pérdida) de una función de distancia, midiendo la magnitud de  $\mu$  al origen. Se establece al final del artículo una relación entre el nivel de significancia que se utiliza en los contrastes de hipótesis en estadística clásica y la función:  $l(\theta) = L_2(\theta) - L_1(\theta)$ . Se observa que es inconsistente tener un nivel de significancia fijo mientras el tamaño de muestra crece y esto tiene el mismo efecto que la paradoja de Lindley<sup>7</sup>. Establece, también, una forma de cambiar consistentemente el nivel de significancia según crece el tamaño de muestra. Esto puede dar pautas para la utilización del método de distribuciones discontinuas.

<sup>7</sup> Esta es una famosa paradoja que aparece en la estadística clásica dentro de el problema de contrastes de hipótesis, véase Lindley (1957).

### 2.2.3 Contraste de Significancia de Lindley

Este método fue propuesto por Lindley (1970 p.58) y fué desarrollado para contrastar hipótesis simples contra compuestas para parámetros con distribución continua. Es difícil pensar en generalizaciones de este método para algún otro tipo de contrastes de hipótesis. Para estudiar este método necesitamos la definición de región de máxima densidad<sup>1</sup>:

Se tiene  $\theta \in \Theta$  con distribución generada por la densidad generalizada  $f(\theta)$ , la región de *Máxima Densidad* (MD) con probabilidad  $\alpha$  es la región  $I_\alpha$  ( $I_\alpha \subset \Theta$ ) tal que:

$$P(I_\alpha) = \alpha$$

y si  $x \in I_\alpha$  y  $y \notin I_\alpha$ , entonces:

$$f(x) \geq f(y).$$

Se puede probar que si  $f(\theta)$  no es localmente uniforme (que  $f(\theta)$  nunca sea constante) entonces  $I_\alpha$  es único para cada  $\alpha$ . Se puede probar también que si  $f(\theta)$  es continua entonces existe una región para cualquier probabilidad; también, si  $I_\alpha$  es una región de MD con probabilidad  $\alpha$ , entonces si  $f(\theta)$  nunca es constante (o es constante sólo en conjuntos de probabilidad cero) entonces  $I_\alpha = \{\theta \in \Theta \mid f(\theta) \geq C_\alpha\}$  con  $C_\alpha = \inf_{\theta \in I_\alpha} f(\theta)$ . Como gran parte de las densidades continuas que se utilizan nunca son constantes (o son constantes sólo en conjuntos de probabilidad cero), cuando hablemos de una región de MD con probabilidad  $\alpha$  esta se puede pensar como el conjunto  $\{\theta \in \Theta \mid f(\theta) \geq C_\alpha\}$  para una cierta  $C_\alpha$ . Nosotros vamos a estar trabajando con lo que se conoce como regiones de Máxima Densidad Posterior (MDP) que son regiones de MD pero calculadas con la densidad posterior del parámetro.

El contraste de hipótesis que se quiere realizar es el siguiente:

$$H_1 : \theta = \theta_0$$

$$H_2 : \theta \neq \theta_0,$$

con  $f(\theta)$  una densidad continua y  $\Theta$  el espacio parametral. Se obtiene una muestra  $z_n$  y se calcula la distribución posterior  $f(\theta \mid z_n)$ ; se dice entonces que se rechaza la hipótesis  $H_1$  con un nivel de significancia  $1 - \alpha$  si  $\theta_0$  no pertenece a la región de MDP con probabilidad  $\alpha$ . Si la distribución posterior de  $\theta$  nunca es constante entonces  $I_\alpha = \{\theta \in \Theta \mid f(\theta) \geq C_\alpha\}$  para cierta  $C_\alpha$  y entonces rechazamos  $H_1$  a un nivel de significancia  $1 - \alpha$  si  $\theta_0 \notin \{\theta \in \Theta \mid f(\theta) \geq C_\alpha\}$  ó  $f(\theta_0) < C_\alpha$ .

<sup>1</sup>La definición y demas resultados al respecto fueron tomados de Box y Tiao (1973, p.122 en adelante).



El contraste de hipótesis por el procedimiento de significancia de Lindley está enfocado para problemas de inferencia y está basado en el hecho en que las regiones de MD son los conjuntos que menor volumen tienen a una cierta probabilidad; esto quiere decir que si  $I_\alpha$  es una región de MD con probabilidad  $\alpha$ , entonces cualquier otro conjunto con la misma probabilidad tiene mayor volumen que  $I_\alpha$ <sup>2</sup>. Lo anterior se interpreta diciendo que  $I_\alpha$  contiene a los valores más probables de  $\theta$  y por eso rechazamos la hipótesis  $H_1: \theta = \theta_0$  si  $\theta_0 \notin I_\alpha$ . El principal problema para aplicar el contraste de significancia de Lindley es encontrar la región de MD, una vez resuelto el problema de encontrar la distribución posterior de  $\theta$ . En los ejemplos siguientes vamos a ilustrar como pueden encontrarse las regiones de MD para algunas distribuciones continuas.

**Ejemplo 1:** Si se tiene como densidad posterior de  $\theta$  una densidad unimodal y simétrica con respecto a la moda entonces la región de MD a cualquier probabilidad es una vecindad centrada en la moda y con radio  $r$  que depende de la probabilidad de la región. Por ejemplo, si  $\theta | z_n \sim N(\mu, \sigma^2)$  sabemos que  $\mu$  es la moda de la distribución Normal y es simétrica con respecto a ésta; para calcular la región de MDP con probabilidad  $\alpha$  necesitamos encontrar  $r$  tal que  $P(|\theta - \mu| < r) = \alpha$ , esto lo podemos hacer encontrando un punto  $\mu - r$  tal que:

$$P(\theta < \mu - r | z_n) = \frac{1 - \alpha}{2},$$

de esta manera  $P(\theta > \mu + r | z_n) = \frac{1 - \alpha}{2}$ , por simetría, y entonces  $P(|\theta - \mu| < r) = \alpha$ ; por lo tanto,  $r$  es el radio que buscábamos. Esto se puede hacer con tablas de la Normal (ya que  $\frac{\theta - \mu}{\sigma} \sim N(0, 1)$ ) o con alguna rutina de integración numérica.

Otro ejemplo de una distribución continua, simétrica y unimodal es la  $t$  de Student, en el apéndice C aparece un listado de un programa en pascal que tiene implementada una rutina para calcular regiones de MD para la distribución  $t$  a cualquier probabilidad. Este calcula numéricamente las probabilidades y utiliza la idea del párrafo anterior para encontrar la región de MD deseada. •

**Ejemplo 2:** Si la distribución del parámetro  $\theta$  no es simétrica el problema de encontrar la región de MD es mucho más complicado que lo visto en el ejemplo 1. Un método analítico implicaría lo siguiente: (consideramos que la distribución en cuestión es unimodal, como suele suceder)

1) Elegir  $K \in R$

<sup>2</sup> En términos formales queremos decir que si  $P(A) = \alpha$  entonces  $\mu(\theta)(A) \geq \mu(\theta)(I_\alpha)$ .

2) Encontrar  $a_1 < a_2$  tales que  $f(a_1) = K = f(a_2)$  (existen sólo dos).

3) Calcular  $\beta = P(\theta \in [a_1, a_2] | z_n)$ .

4) Si  $\beta = \alpha$ ,  $[a_1, a_2]$  es la región de MD buscada; si  $\beta < \alpha$  elegir  $K' < K$  y elegir  $K' > K$  en otro caso. Repetir los pasos del 2) al 4) con  $K = K'$ .

Como puede verse, llevar a cabo la implementación de lo anterior en algún lenguaje de programación tiene muchas complicaciones técnicas, aun suponiendo que podemos calcular  $a_1$  y  $a_2$  ya que resulta difícil y tardado calcular  $P(\theta \in [a_1, a_2] | z_n)$ .

Una forma alternativa de encontrar la región de MD con probabilidad  $\alpha$  es utilizar una gráfica de la función de densidad posterior para encontrar  $a_1$  y  $a_2$ . En la figura 1 aparece una gráfica de una densidad Gamma con parámetros 9.4 y 1.5, para calcular la región de MD con  $\alpha = 0.9$  se tiró una línea a altura  $C_1 = 0.024$  estimándose  $a_1 = 2.2$  y  $a_2 = 10.7$  (los puntos en las abscisas donde la recta  $y = 0.024$  se interseca con la gráfica), se calculó entonces la probabilidad  $P(2.2 \leq \theta \leq 10.7 | z_n) = 0.9671$ . Se tomó  $C_2 > C_1$  ( $C_2 = 0.04$ ) y se continuó con el proceso hasta que con  $C_4 = 0.056$  resultó  $a_1 = 3.0$  y  $a_2 = 9.3$  y  $P(3.0 \leq \theta \leq 9.3 | z_n) = 0.8905$  entonces  $[3.0, 9.3]$  es una aproximación de la región de MD deseada.

En la figura 2 encontramos la gráfica de una densidad Beta con parámetros 8.4 y 3.0, se siguió el mismo proceso del párrafo anterior para encontrar el intervalo de MD con probabilidad 0.9; con  $C_1 = 0.9$  obtuvimos  $a_1 = 0.53$  y  $a_2 = 0.94$  y la  $P(0.53 \leq \theta \leq 0.94 | z_n) = 0.9079$ , con lo cual  $[0.53, 0.94]$  es una buena aproximación a la región de MD que buscábamos. •

Densidad Gama con parámetros 9.4 y 1.5

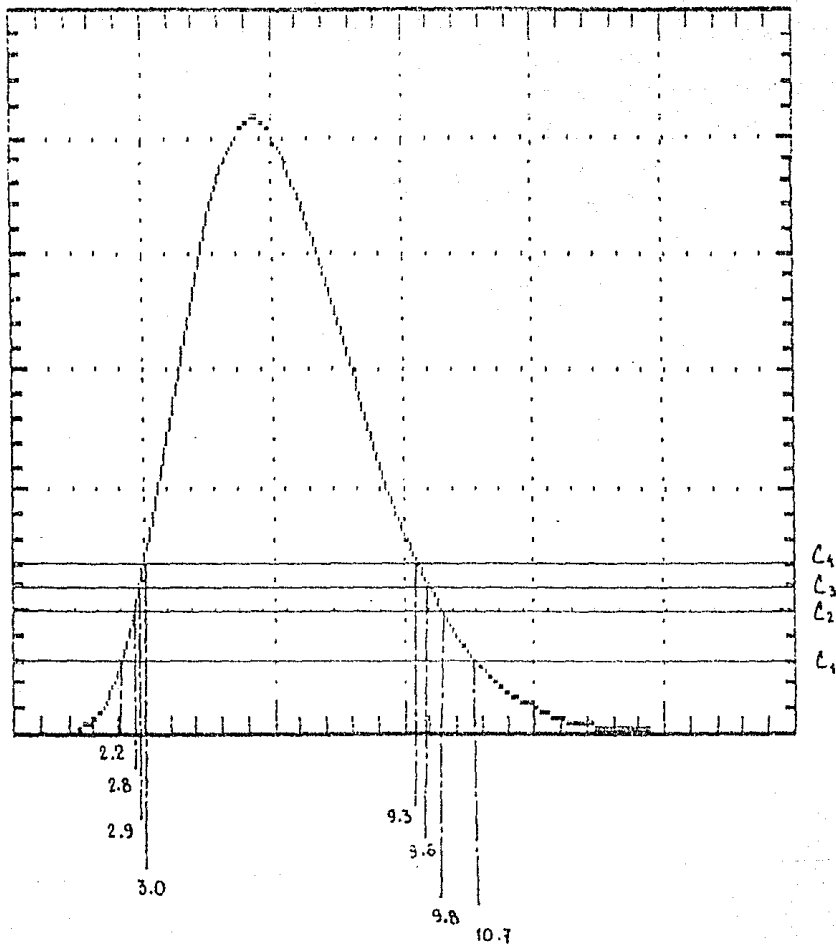


Figura 1

Densidad Beta con parámetros 8.4 y 3.0

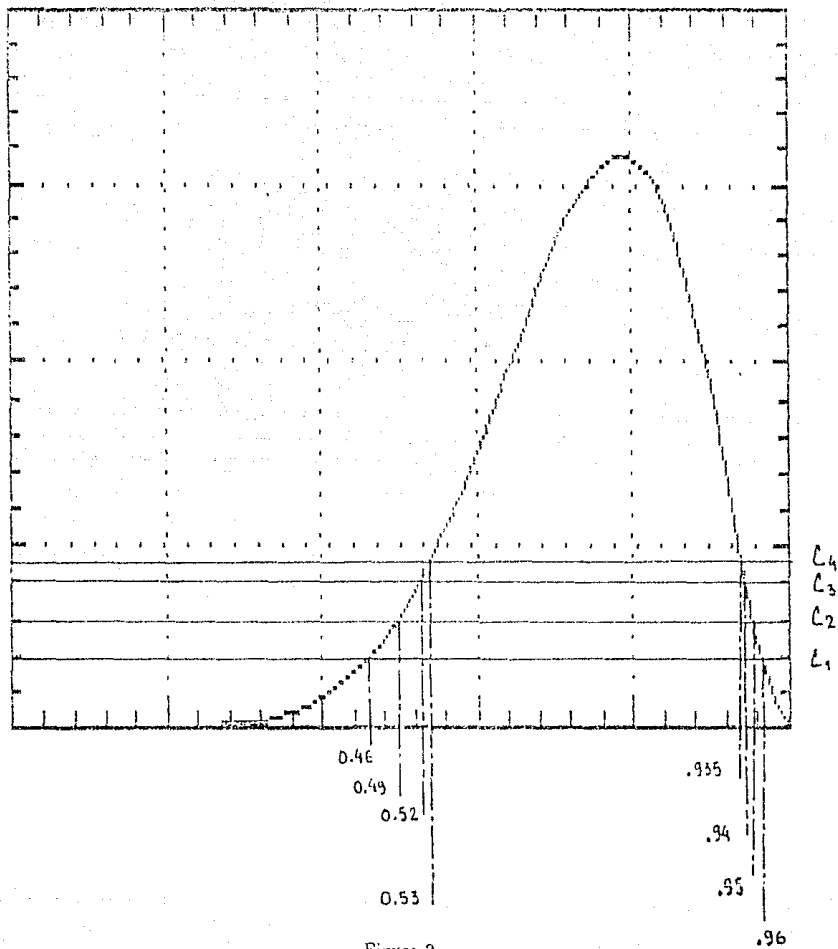


Figura 2

## Capítulo 3

### Hipótesis Sobre Modelos Lineales

### Capítulo 3

## Hipótesis Sobre Modelos Lineales

La intención de este capítulo es ejemplificar los métodos de contrastes de hipótesis expuestos en el capítulo anterior. A diferencia de los ejemplos expuestos en el capítulo 2, ahora nos vamos a enfrentar a un contraste de hipótesis intentando aplicar cada uno de los métodos de contrastes de hipótesis vistos antes. Decidimos introducir este capítulo debido a la problemática que tienen los contrastes de hipótesis en el contexto de los modelos lineales. En la literatura sobre estadística Bayesiana se notan ciertas inconsistencias al tratar las hipótesis sobre modelos lineales ya que se les suele dar un tratamiento aparte de los demás contrastes de hipótesis. Tradicionalmente, este tipo de contrastes ( en estadística clásica) se han manejado como contrastes de hipótesis simple con compuesta, tratándose de parámetros con distribuciones continuas. Como vimos en el capítulo anterior, cuando tenemos este problema podemos optar por cambiar el modelo y hacer de la distribución del parámetro una distribución discontinua, podemos también utilizar el contraste de significancia de Lindley u optar por contrastar las hipótesis por el método de comparación de probabilidades o (si es el caso) mediante teoría de decisiones utilizando la distribución continua y preferir siempre la hipótesis compuesta, independientemente de la muestra. Sin embargo, antes de preocuparnos por el contraste deberíamos preguntarnos:

¿ El contraste de estas hipótesis nos va a proporcionar la información sobre la que estamos interesados ?

Tal vez las hipótesis sobre modelos lineales planteadas en la forma tradicional, no son las hipótesis en que estamos interesados. No espero aquí dar respuesta a todos estos cuestionamientos, pero al menos espero hacer notar que aún no está completamente claro el problema de contrastes de hipótesis dentro de la estadística Bayesiana.

### 3.1 El Modelo Lineal

Tenemos una variable aleatoria  $Y$  y decimos que hay una relación lineal entre  $Y$  y las variables  $x_1, x_2, \dots, x_p$  si:

$$Y = \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_p x_p.$$

A las variables  $x_i$  las llamamos variables explicativas y a  $Y$  variable dependiente. Se puede interpretar lo anterior como que las  $x_i$  ( $i = 1, \dots, p$ ) representan ciertas condiciones, y dadas éstas observamos la resultante  $Y$ . Vamos a utilizar la siguiente

notación. Tenemos un vector de observaciones de  $Y = (y_1, y_2, \dots, y_n)'$  y una matriz  $X$  de  $n \times p$  formada por las columnas  $X_i$  donde las entradas de cada una de éstas son los valores que observamos de la variable explicativa  $x_i$ . El modelo lineal es:

$$Y = X\theta$$

con  $\theta = (\theta_1, \theta_2, \dots, \theta_p)'$ . Debido al conocimiento que tenemos sobre la variable  $Y$ , como de la relación lineal de ésta con  $x_i$  ( $i = 1, \dots, p$ ), se introduce al modelo un factor de error como un vector  $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)'$  y el modelo queda:

$$Y = X\theta + \epsilon.$$

Por las características del modelo,  $\epsilon$  es una variable aleatoria y usualmente se supone que  $\epsilon \sim N_n(0, \tau^{-1}I_n)$ , donde  $I_n$  es la matriz identidad de  $n \times n$  y  $\tau^{-1}I_n$  la matriz de precisión de la distribución Normal  $n$ -variada. No es mi intención dar aquí una justificación amplia del modelo lineal, para una exposición completa puede consultarse Broemeling (1985, p.1 en adelante) o Zellner (1971, p.65 en adelante). A continuación se presentan dos ejemplos de modelos lineales:

**Ejemplo 1:** Se está estudiando una reacción química en la que intervienen dos compuestos A y B, y al darse la reacción se incrementa la temperatura del recipiente donde se mezclan A y B; los compuestos A y B se mezclan con una cantidad fija de agua. Digamos que la cantidad del compuesto A es  $x_1$  y la cantidad del compuesto B es  $x_2$ , al llevar a cabo la reacción química, la temperatura máxima que medimos es  $Y$ . Los interesados en la reacción química quieren encontrar la relación:

$$Y = \theta_1 x_1 + \theta_2 x_2.$$

Por la experiencia acumulada anteriormente, se sabe que hay cierto error en la medición de la temperatura, así como en el cálculo del peso de los compuestos A y B que serán utilizados en la reacción. Es por esto que el modelo lineal se le incluye un factor de error  $\epsilon$ . Usualmente se supone que  $\epsilon$  es una variable aleatoria con distribución Normal  $n$ -variada centrada en cero, y en este caso también se hace así; el modelo lineal es:

$$Y = X\theta + \epsilon,$$

para distintas cantidades del compuesto A (entradas de la columna 1 de  $X$ ) y distintas cantidades del compuesto B (entradas de la columna 2 de  $X$ ).

Al manejar el modelo se hace la siguiente observación: Si no hay compuestos A y B ( $x_i = 0$ ) no hay ninguna reacción y la temperatura máxima no

necesariamente es cero ( $Y = \theta_1 \cdot 0 + \theta_2 \cdot 0 + \epsilon$ ); para corregir lo anterior se introduce el parámetro  $\theta_3$  de la siguiente manera:

$$Y = \theta_1 x_1 + \theta_2 x_2 + \theta_3 + \epsilon,$$

añadiendo una tercera columna a  $X$  formada por  $X_3 = (1, 1, \dots, 1)'$  y el modelo sigue siendo  $Y = X\theta + \epsilon$ . Entonces si  $x_1 = x_2 = 0$  se tiene que  $Y = \theta_3 + \epsilon$ .

Una posible hipótesis que se puede plantear acerca de este modelo sería:  $H_1$ : El compuesto B no influye en que se eleve la temperatura; o también,  $H_1$ : El compuesto A no influye en que se eleve la temperatura. •

**Ejemplo 2:** Se quiere medir la cantidad de luz que emite un foco dado el voltaje a que trabaja. El voltaje a que trabaja es la variable  $x_1$  y la luz emitida es medida por la variable  $Y$ . Si el voltaje es cero ( $x_1 = 0$ ) entonces  $Y = 0$ . Se plantea igualmente el modelo lineal:

$$Y = X\theta + \epsilon.$$

La intención de los investigadores es saber si existe una relación lineal entre el voltaje y la cantidad de luz emitida, para eso plantean la hipótesis: No existe relación lineal entre la cantidad de luz emitida y el voltaje de trabajo ( $H_1$ ). •

Analícemos ahora un poco más el modelo lineal. Como  $\epsilon$  es una variable aleatoria con distribución Normal  $n$ -variada y el vector  $X$  es fijo entonces  $Y$  también es Normal. En realidad  $(y_1, y_2, \dots, y_n)'$  es una muestra de  $Y$ , tenemos entonces que  $z_n = (y_1, y_2, \dots, y_n)'$ . Si  $z_n$  es una muestra aleatoria entonces:

$$h(z_n | \theta, \tau) \propto \tau^{n/2} \exp\left\{-\frac{\tau}{2}(Y - X\theta)'(Y - X\theta)\right\},$$

y por el teorema de Bayes:

$$f(\theta, \tau | z_n) \propto h(z_n | \theta, \tau)f(\theta, \tau).$$

La familia conjugada para la distribución inicial conjunta de  $\theta$  y  $\tau$  adquiere la siguiente fórmula:

$$f(\theta, \tau) = f_1(\theta | \tau)f_2(\tau)$$

donde,

$$f_1(\theta | \tau) \propto \tau^{p/2} \exp\left\{-\frac{\tau}{2}(\theta - \mu)'P(\theta - \mu)\right\}$$



con  $\mu$  un vector de  $p \times 1$  y  $\tau P$  matriz de precisión positiva definida de  $p \times p$ .  
Tenemos entonces que:

$$\theta | \tau \sim N_p(\mu, \tau P).$$

Ahora,  $\tau$  se distribuye Gama con parámetros  $a$  y  $b$ , entonces:

$$f_2(\tau) = \tau^{a-1} \exp\left(-\frac{\tau}{b}\right); \quad \tau > 0.$$

Para calcular la distribución marginal de  $\theta$  tenemos que:

$$f_1(\theta) \propto \int_0^{\infty} f(\theta, \tau) d\tau$$

entonces, después de los cálculos necesarios (Broemeling 1985, p.3):

$$f_1(\theta) \propto \{2/b + (\theta - \mu)' P (\theta - \mu)\}^{-\frac{(b+2a)}{2}}$$

que es una distribución  $t$   $p$ -dimensional con  $2a$  grados de libertad, vector de medias  $\mu$  y matriz de precisión  $(2a)(b/2)P$ . Tenemos pues que:

$$E(\theta) = \mu,$$

y la matriz de dispersión de  $\theta$  es:

$$D(\theta) = \frac{P^{-1}(2/b)}{p + 2(a - 1)}.$$

Ahora, la distribución posterior  $f_1(\theta | z_n)$  es también una distribución  $t$  con  $n + 2a$  grados de libertad, vector de medias:

$$\mu^* = (X'X + P)^{-1}(X'Y + P\mu)$$

y matriz de precisión:

$$D^* = (X'X + P)(n + 2a)\{(2/b) + Y'Y - (X'Y + P\mu)'(X'X + P)^{-1}(X'Y + P\mu)\}^{-1}.$$

**Ejemplo 3:** Consideremos el caso más sencillo del modelo lineal, que traté en el ejemplo 2. Recordemos que  $p = 1$ ,  $X$  matriz de  $n \times 1$ , y vector de observaciones  $Y$  de  $n \times 1$ ; el modelo lineal es  $Y = X\theta + \epsilon$  que puede reescribirse como:

$$y_i = x_i\theta + \epsilon_i; \quad i = 1, \dots, n.$$

Si utilizamos  $f_1(\theta) \propto \{(2/b) + (\theta - \mu)^2\}^{-\frac{(1+2a)}{2}}$  tenemos que  $\theta \sim t(2a, \mu, ab)$ , que es la distribución conjugada para  $\theta$ . En los términos vistos antes,  $P = 1$ ,  $a, b > 0$  y  $\mu \in R$ . Ahora, considerando que  $X'X = \sum x_i^2$  y  $X'Y = \sum x_i y_i$ , entonces:

$$\mu^* = \frac{\mu + \sum x_i y_i}{1 + \sum x_i^2}$$

y

$$r^* = \frac{(n + 2a)}{\frac{\sum y_i^2 + (2/b)}{\sum x_i^2 + 1} - (\mu^*)^2}$$

entonces,

$$\theta | z_n \sim t(n + 2a, \mu^*, r^*).$$

### 3.2 Hipótesis Lineal General

La hipótesis lineal general está dada por una matriz  $A$  de  $k \times p$ ,  $k \leq p$  y un vector  $b \in R^k$  y el contraste de hipótesis es:

$$H_1 : A\theta = b$$

$$H_2 : A\theta \neq b.$$

El contraste anterior nos puede servir para estipular muy diversas hipótesis utilizando la matriz  $A$  que sea conveniente, uno de los contrastes más importantes es el que se lleva a cabo para revisar la validez del modelo lineal, usando  $b = 0$  y  $A$  igual a la matriz identidad de  $p \times p$  ( $I_p$ ). Si  $A\theta = 0$  tenemos que  $\theta = 0$  y entonces no existe una relación lineal entre las variables  $Y$  y  $X$ , resultando que el modelo lineal no es adecuado, no es correcto el ajuste lineal, como se suele decir. Observemos ahora la hipótesis lineal general con  $A = I_p$  y  $b = \theta_0$ , ésta se quedaría como:

$$H_1 : \theta = \theta_0$$

$$H_2 : \theta \neq \theta_0.$$

Si llevamos a cabo este contraste con el método de comparación de probabilidades o vista como un problema de decisión (si es el caso) aceptaríamos  $H_2$  independientemente del resultado de la muestra, pues  $P(\Theta_1) = 0$  y, por lo tanto (por el teorema de Bayes),  $P(\Theta_1 | z_n) = 0$ ; para que tuviese sentido utilizar alguno de estos métodos, tendríamos que cambiar la distribución de  $\theta$  a una distribución discontinua. Supongamos que confiamos en el planteamiento del modelo, que el

modelo es correcto y no hay porque cambiarlo; el paso siguiente es decidir entre estas dos opciones:

1) Utilizar el contraste de significancia de Lindley para contrastar la hipótesis lineal general.

2) Replantear la hipótesis lineal general.

Analicemos la opción 1): La forma de llevar a cabo el contraste de significancia de Lindley es primero encontrando la región de Máxima Densidad Posterior (región de MDP)  $I_\alpha$ , de  $\theta$  para la probabilidad  $\alpha$  deseada. Sabemos que la distribución posterior de  $\theta$  es una  $t$   $p$ -dimensional, utilizando las propiedades de las regiones de MDP tenemos que  $I_\alpha = \{\theta \in \Theta \mid f_1(\theta \mid z_n) > C_\alpha\}$  para alguna  $C_\alpha \in R^+$ . Entonces se rechaza  $H_1$  si  $\theta_0 \notin I_\alpha$  ó si  $f_1(\theta_0 \mid z_n) < C_\alpha$  con una significancia de  $(1 - \alpha)100\%$ . Para el caso  $p = 1$ , se tiene una distribución  $t$  univariada, como ésta es simétrica con respecto a su media  $\mu^*$ , y la media coincide con la moda, entonces  $I_\alpha$  va a ser una vecindad de  $\mu^*$  de radio  $r_\alpha$  tal que:

$$\int_{V_{r_\alpha}(\mu^*)} f_1(\theta \mid z_n) d\theta = \alpha.$$

En el apéndice C se encuentra el listado de un programa en pascal que calcula  $I_\alpha$  para cualquier  $\alpha$ , siempre que  $2a$  sea entero (recuérdese que la distribución posterior de  $\theta$  tiene  $n + 2a$  grados de libertad).

Una forma alternativa de llevar a cabo este contraste de hipótesis para ciertos valores de  $\alpha$ , si no se cuenta con una computadora, es considerar lo siguiente:

$$r^*(\theta - \mu^*)^2 \sim F_{(1, n+2a)}.$$

Podemos buscar en tablas de la distribución  $F$  el cuantil de  $(1 - \alpha) \cdot 100\%$ , sea este  $F^*$ , entonces:

$$P(r^*(\theta - \mu^*)^2 \leq F^* \mid z_n) = \alpha$$

entonces,

$$P(|\theta - \mu^*| \leq \sqrt{\frac{F^*}{r^*}} \mid z_n) = \alpha$$

y esto es una vecindad centrada en  $\mu^*$  con radio  $r = \sqrt{\frac{F^*}{r^*}}$ , y:

$$P(\theta \in V_r(\mu^*) \mid z_n) = \alpha.$$

Tenemos que  $V_r(\mu^*)$  es la región de MDP con probabilidad  $\alpha$  y entonces rechazamos la hipótesis  $H_1$  con un nivel de significancia de  $(1 - \alpha)100\%$  si  $\theta_0 \notin V_r(\mu^*)$  y en el caso que  $\theta_0 = 0$  rechazamos  $H_1$  si  $|\mu^*| > \sqrt{\frac{F^*}{r^*}}$ . El problema aquí es que sólo existen tablas para ciertos valores de  $\alpha$  como 0.95, 0.9, etc. y bien puede no

quererse utilizar un nivel de significancia como ese y también que;  $2a$  tiene que ser entero.

En el caso general ( $H_1 : A\theta = b$ ), el método anterior funciona también para; como  $A\theta$  es una transformación lineal de  $\theta$  entonces  $A\theta$  se distribuye como una  $t$ ,  $k$ -dimensional, con media  $A\mu^*$  y matriz de precisión  $(AD^*A')^{-1}$  si  $AD^*A'$  es no singular. Se considera entonces la transformación:

$$G(A\theta) = k^{-1} \{A(\theta - \mu^*)\}' (AD^*A')^{-1} \{A(\theta - \mu^*)\},$$

y entonces  $G$  tiene una distribución  $F$  con  $k$  y  $n + 2a$  grados de libertad y si  $F^*$  es cuantil de  $F_{(k, n+2a)}$  a  $(1 - \alpha) \cdot 100\%$  entonces:

$$I_\alpha = \{A\theta \in R^k \mid G(A\theta) \leq F^*\},$$

que es la región de MPD con probabilidad  $\alpha$  para la variable aleatoria  $A\theta$ , como la hipótesis lineal general es  $H_1 : A\theta = b$ , entonces se rechaza  $H_1$  si  $b \notin I_\alpha$  o si:

$$G(b) > F^*.$$

Hasta aquí todo parece funcionar bien con el contraste de significancia de Lindley.

Analicemos ahora la opción 2). Volvamos a la hipótesis lineal general, tomando  $A = I_p$  y  $b = 0$ :

$$H_1 : \theta = 0$$

$$H_2 : \theta \neq 0.$$

Si  $H_1$  es correcta, no existe una relación lineal entre las variables  $X$  y  $Y$  y el modelo que se planteó es inoperante, inadecuado (nótese que en este caso el que  $H_1$  sea correcta significa que  $\theta = 0$ , y por eso es que no hay relación lineal entre las variables  $X$  y  $Y$ ); el problema es darnos cuenta que la mayoría de las veces se está interesado en la hipótesis de si es o no adecuado el modelo que hemos planteado y el contraste de arriba es sólo una interpretación, en términos del parámetro  $\theta$ , de esta hipótesis original. Las hipótesis originales de que hablo podrían quedar expresadas en estos términos:

$H_1$  : El modelo lineal no es adecuado.

$H_2$  : El modelo lineal es adecuado.

Obviamente, lo que es o no es adecuado debe de ser estipulado por el que plantea las hipótesis. En una primera aproximación para intentar resolver el problema, intuitivamente pensamos que podríamos proponer la hipótesis  $H_1 : \|\theta\| \leq \epsilon$  para interpretar la hipótesis "El modelo lineal no es adecuado", y de esta manera transformando la hipótesis  $H_1 : \theta = 0$  en una hipótesis compuesta; ahora proponemos que  $\theta$  esté "cerca" de 0 ( $\|\theta\| \leq \epsilon$ ) en lugar de que "sea" 0 y desde

luego el valor de  $\epsilon$  debe de ser estipulado por el que plantea la hipótesis. Parece ser entonces que el contraste de hipótesis anterior lo podríamos plantear, en términos del parámetro  $\theta$ , como:

$$H_1 : \|\bar{\theta}\| \leq \epsilon$$

$$H_2 : \|\bar{\theta}\| > \epsilon.$$

Generalizando esta idea, la hipótesis lineal general podría plantearse de la siguiente forma:

$$H_1 : \|A\theta - b\| \leq \epsilon$$

$$H_2 : \|A\theta - b\| > \epsilon.$$

Obsérvese que  $\|A\theta - b\|$  es alguna norma de  $(A\theta - b)$  en  $R^k$  y se puede entender como la norma común de  $R^k$ , pero también esta podría cambiarse según los requerimientos del interesado. Para el caso  $p = 1$  el contraste queda:

$$H_1 : |\theta - \theta_0| \leq \epsilon$$

$$H_2 : |\theta - \theta_0| > \epsilon$$

o viéndolo como vecindades:

$$H_1 : \theta \in V_\epsilon(\theta_0)$$

$$H_2 : \theta \notin V_\epsilon(\theta_0).$$

Ahora es posible utilizar el método de comparación de probabilidades o (si es el caso) teoría de decisiones para llevar a cabo el contraste anterior; el problema al utilizar el método de comparación de probabilidades es calcular:

$$P(H_1 | z_n) = \int_{V_\epsilon(\theta_0)} f(\theta | z_n) d\theta.$$

Para el caso de  $p = 1$ , el programa que aparece en el apéndice C puede calcular esta probabilidad para cualquier  $\theta_0$  y para cualquier  $\epsilon$ ; para valores de  $p$  mayores que 1 los cálculos son mucho más complicados pues se trata de calcular numéricamente integrales múltiples y los algoritmos de integración numérica son mucho más complejos<sup>1</sup>. Si se desea utilizar el método de teoría de decisiones podemos utilizar una función de pérdida como  $L_1 = \alpha \chi_{\theta_2}(\theta)$  y  $L_1^* = \alpha P(H_2 | z_n)$  y los problemas de cálculo son iguales que en el caso de comparación de probabilidades. Si se utilizan funciones de pérdida más complicadas los cálculos pueden ser prácticamente imposibles.

<sup>1</sup> De hecho, los algoritmos de integración numérica múltiple que aparecen en la literatura correspondiente, sólo se realizan para regiones simples (como cuadros o cubos) y aún así éstos son bastante complicados y aparte; es muy difícil estimar los errores de aproximación. Véase Burden y Faires (1985, p.211) o Bakhvalov (1977, p.298).

### 3.3 Ejemplos Comparativos

Por lo visto en el punto anterior, parece ser que no existe, dentro de la estadística Bayesiana, un método concreto para atacar el contraste de la hipótesis lineal general. Sin intentar cambiar el modelo, las dos opciones que planteamos antes parecen ser igualmente adecuadas sin embargo, si vemos el problema desde el punto de vista de la teoría de decisiones utilizando las funciones de pérdida  $L_i(\theta) = c_i \chi_{\Theta - \Theta_i}(\theta)$   $i = 1, 2$  con  $c_1 = c_2$ , entonces la decisión óptima es elegir  $H_1$  si:

$$L_1^* < L_2^*$$

ó

$$c_1 P(H_2 | z_n) < c_2 P(H_1 | z_n)$$

ó

$$P(H_1 | z_n) > P(H_2 | z_n),$$

y en este sentido el procedimiento de comparación de probabilidades es óptimo, como ya fué explicado en el capítulo anterior. Por otro lado, calcular  $P(H_1 | z_n)$  puede ser bastante complicado puesto que  $\Theta_1 = \{\theta \in R^k \mid \|A\theta - b\| \leq \epsilon\}$  y se tendría que hacer una integración numérica múltiple en  $R^k$  sobre  $\Theta_1$  y esto puede acarrear muchos problemas; puede ser entonces que en algunos casos, calcular  $P(H_1 | z_n)$  sea muy complicado y por lo tanto, llevar a cabo el método de comparación de probabilidades sea prácticamente imposible. Comparando ésto con el método de significancia de Lindley observamos que el único problema es encontrar el cuantil de la distribución  $F$  correspondiente para el nivel de significancia  $\alpha$  deseado; si éste es como de los que usualmente aparecen en tablas (0.9, 0.95, etc.) el problema está resuelto. Si no existiera una tabla para un nivel de significancia dado el cuantil puede ser calculado numéricamente e inclusive, el paquete STATGRAPHICS tiene implementada una rutina para calcular esos cuantiles. Fuera de lo anterior, para realizar el contraste de la hipótesis lineal general por el método de significancia de Lindley, tenemos que calcular algunos productos de matrices e invertir una matriz (que por cierto es simétrica), los cálculos pueden ser tediosos pero el problema está resuelto.

Analicemos las dos opciones planteadas en 3.2:

1) Utilizar el contraste de significancia de Lindley para contrastar la hipótesis lineal general.

2) Replantear la hipótesis lineal general como en 3.2, fijar un radio  $\epsilon$  y hacer un esfuerzo de cálculo para obtener  $P(H_1 | z_n)$ .

A continuación vamos a estudiar algunos ejemplos comparativos del uso de las dos opciones, y lo vamos a hacer precisamente cuando es fácil calcular  $P(H_1 | z_n)$ , para  $p = 1$ . Antes de comenzar con el análisis observemos que para preferir  $H_1$  es necesario (con  $p = 1$ ) que  $\mu^*$ , la media de la distribución posterior de  $\theta$  (que es una  $t$ ), esté dentro de  $\Theta_1 = V_\epsilon(\theta_0)$ . Esto es fácil de ver, ya que si  $\mu^* \notin V_\epsilon(\theta_0)$ , y suponemos que  $\theta_0 < \mu^*$ , tenemos que  $V_\epsilon(\theta_0) \subset (-\infty, \mu^*)$  y como  $P((-\infty, \mu^*) | z_n) = 0.5$  entonces  $P(V_\epsilon(\theta_0) | z_n) < 0.5$  lo cual implica que preferimos  $H_2$ . Análogamente sucede para cuando  $\theta_0 > \mu^*$ ; entonces si  $\mu^* \notin V_\epsilon(\theta_0)$  resulta que preferimos a  $H_2^2$ .

Ahora, en el punto anterior hablamos de una  $\epsilon$  (pequeña) y que debería ser estipulada por el interesado en el contraste de hipótesis. Tal vez una  $\epsilon = 0.1$  sea apropiada o tal vez en otro caso una  $\epsilon = 0.5$  sea la conveniente etc. para llevara cabo este ejemplo vamos a considerar que  $\epsilon = 1.1$  ya es una  $\epsilon$  grande, y lo que queremos decir es que normalmente se utilizaran valores de  $\epsilon$  menores que, digamos, 1.1. Veamos el siguiente ejemplo:

Tenemos  $\theta_0 = 0$  y la hipótesis lineal general es:

$$H_1 : \theta = 0$$

$$H_2 : \theta \neq 0,$$

y siguiendo la opción 2) tenemos:

$$H_1 : \theta \in V_\epsilon(0)$$

$$H_2 : \theta \notin V_\epsilon(0)$$

para  $p = 1$  y  $1.1 > \epsilon > 0$  dada. Tenemos que la distribución posterior de  $\theta$  es  $t(15, 1.1, 0.2)$ , que es la distribución posterior del parámetro de un modelo lineal con  $p = 1$  y  $n = 13$ . Como  $\mu^* = 1.1$  y  $\epsilon < 1.1$  tenemos que  $\mu^* \notin V_\epsilon(0)$  y entonces vamos preferir  $H_2$  si seguimos el método de comparación de probabilidades. Ahora al calcular el intervalo de MDP con nivel de significancia  $\alpha = 0.1$  tenemos que éste es  $(-2.821, 5.021)$  ¡ y entonces resulta que según el contraste de significancia de Lindley no rechazamos  $H_1$ , con  $\alpha = 0.1$  !

Para que  $P(H_1 | z_n) > 0.5$ , y preferir  $H_1$ , es necesario que  $\epsilon > 1.8$  lo cual parece ser muy grande. Podemos ver en la figura 1 (al final del capítulo) una gráfica del nivel de significancia contra el extremo izquierdo y derecho del intervalo de MDP correspondiente y tenemos que para que sea rechazada la hipótesis  $H_1$  es necesario que  $\alpha < 0.37$ . Tenemos entonces que según el contraste de significancia de Lindley se debería NO rechazar  $H_1$ , aún con niveles de significancia tan bajos

<sup>2</sup> Nótese que si  $\mu^* \in V_\epsilon(\theta_0)$  esto no implica que  $P(H_1 | z_n) > 0.5$ .

como 0.5 y sin embargo, según el método de comparación de probabilidades, se debería preferir  $H_2$  puesto que  $P(H_2 | z_n)$  es 15 veces mayor que  $P(H_1 | z_n)$  para  $\epsilon = 0.2$  y 5.5 veces mayor que  $P(H_1 | z_n)$  si  $\epsilon = 0.5$ .

En el ejemplo anterior el intervalo de MPD resultó tan amplio porque la precisión es muy baja, si aumentamos la precisión a 1 tenemos que el intervalo de MPD, para  $\alpha = 0.9$ , es  $(-0.654, 2.854)$  y para rechazar  $H_1$  es necesario que  $\alpha < 0.71$ . En este caso seguimos teniendo el mismo problema que antes puesto que  $P(H_2 | z_n)$  es 22 veces mayor que  $P(H_1 | z_n)$ , para  $\epsilon = 0.2$ . Si aumentamos la precisión ahora a 5 el intervalo de MPD para  $\alpha = 0.9$  es  $(0.316, 1.884)$  y en este caso coincide el método de comparación de probabilidades y el de significancia de Lindley, igual si tomamos precisión de 10 o cualquiera más grande. En las figuras 2 y 3 aparecen las gráficas del nivel de significancia contra los extremos derecho e izquierdo del intervalo de MDP correspondiente a precisiones 1 y 5 respectivamente.

Notemos que el único problema serio para aplicar el método de comparación de probabilidades en el caso general es calcular  $P(H_1 | z_n)$ ; por lo visto en los párrafos anteriores parece ser que si:

$$\|A\mu^* - b\| > \epsilon$$

ó

$$A\mu^* \notin V_\epsilon(b),$$

entonces:

$$P(A\theta \in V_\epsilon(b) | z_n) < 0.5.$$

Si la conjetura anterior es verdadera significa que si  $A\mu^*$ , la media de la distribución de  $A\theta$ , NO pertenece a la vecindad  $V_\epsilon(b)$  de  $R^k$  ( $6 \|A\mu^* - b\| > \epsilon$ ) entonces debemos preferir  $H_2$ , puesto que  $P(A\theta \in V_\epsilon(b) | z_n) = P(H_1 | z_n) < 0.5$ ; lo anterior es una generalización de lo visto antes para la  $t$  unidimensional y parece ser cierto en general aunque no he podido demostrarlo. Esto nos daría una gran ayuda puesto que calcular  $\|A\mu^* - b\|$  es bastante fácil y si  $\|A\mu^* - b\| > \epsilon$  preferiríamos  $H_2$ . Observemos que si  $\|A\mu^* - b\| < \epsilon$  tendríamos que calcular  $P(H_1 | z_n)$ , pero sólo en ese caso.

Analícemos un poco más la hipótesis  $H_1 : \|\theta\| \leq \epsilon$  y como interpreta ésta a la hipótesis planteada antes "El modelo lineal no es adecuado". En principio nos parece mejor que la hipótesis  $H_1 : \theta = 0$  pues  $P(H_1) = 0$  ahora, intuitivamente supusimos que la hipótesis  $H_1 : \|\theta\| \leq \epsilon$  interpretaba la proposición "El modelo lineal no es adecuado" pero ¿que tan cierto nos parece que un valor pequeño de  $\epsilon$  refleja un mal ajuste en el modelo?



Gráficas del nivel de significancia contra los extremos izquierdo y derecho del intervalo de MDP correspondiente:

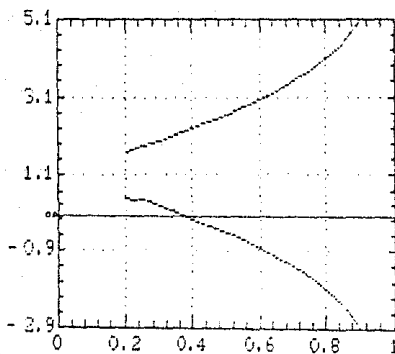


Figura 1,  $t(15, 1.1, 0.2)$ .

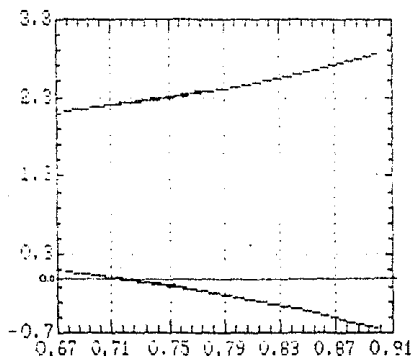


Figura 2,  $t(15, 1.1, 1)$ .

Gráfica del nivel de significancia contra los extremos izquierdo y derecho del intervalo de MDP correspondiente:

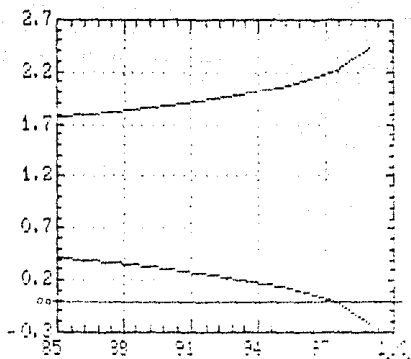


Figura 3,  $t(15, 1.1, 5)$ .

Supongamos que tenemos un modelo que se ajusta muy bien (los datos se encuentran sobre una recta) y que el modelo es  $Y = 10X + \epsilon$ ; ahora podemos crear otro modelo con la variable  $X' = aX$  y el modelo sería  $Y = (10/a)X' + \epsilon$ , si ahora ponemos  $a = 1000$  el modelo nos queda  $Y = 0.01X' + \epsilon$  y  $\theta = 0.01$  que es pequeño, sin embargo el modelo sigue siendo igual de bueno. Notamos que la elección de  $\epsilon$  debe depender de la escala que estamos utilizando pero aun más: Supongamos que tenemos un modelo muy malo (una nube de puntos con poca relación entre si) pero hay unos cuantos puntos por debajo de la nube y resulta que la pendiente de la recta ajustada ( $\mu^*$ ) es muy grande y en realidad el ajuste del modelo es muy malo aun cuando  $\theta$  sea grande, entonces si  $\|\mu^*\| > \epsilon$  tenemos que  $P(H_2 | z_n) > 0.5$  y preferimos a  $H_2$  ("El modelo es adecuado") aun cuando esto no nos parezca muy correcto. De lo anterior podemos hacer dos observaciones: 1) Aun cuando sabemos que  $P(H_2 | z_n) > 0.5$  no la hemos calculado y si la variabilidad en  $\theta$  es grande (que es el caso al trabajar con puntos muy dispersos) muy posiblemente tendremos que  $P(H_2 | z_n)$  sea muy parecida a 0.5, lo cual nos haría dudar de la validez de  $H_2$  y por otro lado, lo más importante, 2) tal vez nos parezca que  $\|\theta\| < \epsilon$  no sea una buena medida de la calidad del ajuste lineal, pero finalmente quien debe decir que es un modelo adecuado o no es el interesado en llevar a cabo el contraste de hipótesis.

Nuestro intento de interpretar la hipótesis "El modelo lineal no es adecuado" como una hipótesis paramétrica compuesta estuvo restringido a  $\|\theta\| < \epsilon$  sin embargo, el camino que hemos esquematizado aquí en este sentido me parece viable. En una investigación futura una opción sería encontrar una medida conveniente de la calidad del ajuste lineal (como por ejemplo la  $R^2$  o el coeficiente de correlación múltiple) y entonces estipular una hipótesis compuesta usando esta medida; seguramente que resultará complicado, pero como hemos visto aquí, podemos encontrarnos con resultados interesantes que nos conduzcan a técnicas prácticas. Lo que siempre tenemos que tener en cuenta (y no sólo para los modelos lineales) es que si la hipótesis suele ser  $H_1 : \theta = \theta_0$ , en estadística Bayesiana no tiene porque estipularse como  $H_1 : \theta \in V_\epsilon(\theta_0)$ .

## Capítulo 4

### Conclusiones

## Capítulo 4

### Conclusiones

En esta tesis intenté ubicar, al problema de contraste de hipótesis, dentro de los métodos generales usados en estadística Bayesiana. Como ya expliqué antes, dentro de la estadística Bayesiana se distinguen dos tendencias principales: En una de estas se resuelven los problemas al inferir (como en Box y Tiao 1973). En la otra tendencia el problema es tomar una decisión (como Berger 1980). Debido a lo anterior, presenté en el capítulo 1 el teorema de Bayes (inferencia) y los conceptos más importantes de la teoría de decisiones; de esta manera, en el punto 2.2.1 traté el problema de contraste de hipótesis como un problema de inferencia y en el punto 2.2.2 como un problema de decisión. En el punto 2.2.3 expongo otro método (contraste de significancia de Lindley) que maneja a los contraste de hipótesis como un problema de inferencia. Cada uno de estos métodos se exponen junto con algunos ejemplos, en donde todo resulta bien. Sin embargo, en el capítulo 3 se expone un ejemplo (modelos lineales) donde las cosas pueden no funcionar correctamente, al intentar contrastar ciertas hipótesis. La importancia de este capítulo, aparte de la comparación de los métodos expuestos en 2.2, es ejemplificar las hipótesis multiparamétricas que son generalizaciones de las uniparamétricas, tratadas en el capítulo 2.

Los problemas importantes que noté a lo largo de la exposición de este trabajo, aparecen con el contraste de las hipótesis:

$$(1) \quad \begin{cases} H_1 : \theta = \theta_0 \\ H_2 : \theta \neq \theta_0, \end{cases}$$

siendo  $\theta$  una variable aleatoria continua. Como ya lo he discutido en los capítulos anteriores: Tenemos que  $P(H_1) = 0$  y entonces  $P(H_1 | z_n) = 0$ , lo cual nos llevaría a preferir  $H_2$  independientemente del resultado de la muestra  $z_n$ . Como dice Winkler (1972, p.424): "Esto no refleja una falla en el enfoque Bayesiano; al contrario, refleja lo absurdo de la hipótesis  $H_1$ ". Si estos tipos de contrastes parecen ser tan absurdos, desde el punto de vista de probabilidad tan claramente a favor de la hipótesis  $H_2$ , ¿Por qué se insiste en plantearlos ?

La respuesta a la pregunta anterior me parece bastante clara: Se insiste el contraste de hipótesis (1) debido a una muy defectuosa importación de un concepto de la estadística Clásica.

Dentro de la estadística clásica se supone que el parámetro  $\theta$ , que determina la distribución de la variable en estudio, es un valor fijo pero desconocido. En este contexto, plantear la hipótesis:

$$H_1 : \theta = \theta_0$$

tiene mucho sentido:  $H_1$  propone que el valor de  $\theta$  sea  $\theta_0$ . Desde luego, las técnicas para realizar el contraste de hipótesis(1), dentro de la estadística Clásica, son completamente distintas a las expuestas en esta tesis; partiendo del hecho de que  $\theta$  no se considera una variable aleatoria.

El contraste de significancia de Lindley, estudiado en el punto 2.2.3, propone un método, que técnicamente no tiene grandes complicaciones, para llevar a cabo el contraste (1). Pero: ¿ Cual es la justificación de este método ?

Vimos ya a lo largo de la tesis, y en particular al principio del punto 3.3, que la manera óptima de elegir entre  $H_1$  y  $H_2$  es utilizando el método de comparación de probabilidades, observando a este como un caso particular de la teoría de decisiones. Lo anterior nos conduce a una contradicción: Puede darse el caso de que utilizando el contraste de significancia de Lindley rechacemos la hipótesis  $H_1$  con un nivel de significancia muy alto mientras que, siguiendo el método de comparación de probabilidades, se prefiera a  $H_2$  sin lugar a dudas ya que  $P(H_2 | z_n) = 1$ . Aún Lindley menciona (Lindley, 1970 p.59), acerca de los contrastes de significancia, que igual que con el concepto de región de máxima densidad, un contraste de significancia es utilizado por ventajas prácticas más que por que matemáticamente se conveniente.

No existe razón teórica para aplicar un contraste de significancia, sólo que técnicamente es fácil aplicarlo.

Sin considerar ahora el contraste de significancia de Lindley, otra opción para llevar a cabo el contraste (1), es cambiar la distribución de  $\theta$  a una distribución discontinua. Una distribución discontinua en  $\theta_0$  implicaría que el único punto en todo el espacio parametral  $\Theta$ , que acumula una probabilidad mayor que cero, es precisamente  $\theta_0$ . Resulta que, siendo  $\Theta$  un conjunto no numerable, de entre todos los puntos en  $\Theta$  escogemos uno,  $\theta_0$ , y sólo a este le otorgamos una probabilidad mayor que cero. Después planteamos el contraste de hipótesis:

$$H_1 : \theta = \theta_0$$

$$H_2 : \theta \neq \theta_0.$$

Realmente no existe un argumento en contra de el uso de un modelo discontinuo, dentro de la teoría de probabilidades y la concepción de la probabilidad subjetiva, pues:

1) Matemáticamente el modelo es correcto.

2) Alguien, por alguna razón, puede requerir que  $P(\theta = \theta_0) > 0$  y sin embargo, tener con espacio parametral un conjunto con medida de Lebesgue mayor que cero (en particular, no numerable). En este caso, no quedaría otra opción que utilizar una distribución discontinua para  $\theta$ .

Aún cuando no podemos decir que el uso de una distribución discontinua, para llevar a cabo el contraste (1), sea un procedimiento incorrecto; hay que hacer notar que: El caso planteado en 2) se refiere a un conocimiento muy extraño; extraordinario.

En el capítulo 3 discutimos que, tal vez, no estemos interesados en que  $\theta = \theta_0$  sino, más bien, en que  $\theta \in V_\epsilon(\theta_0)$ . El convertir la hipótesis simple  $H_1 : \theta = \theta_0$  en la hipótesis compuesta  $H_1 : \theta \in V_\epsilon(\theta_0)$ , y así tener un contraste de hipótesis de compuesta contra compuesta, tiene su precio: aun cuando sepamos que valor se le va a dar a  $\epsilon$  es necesario preguntarnos si en verdad la hipótesis  $H_1 : \theta \in V_\epsilon(\theta_0)$  es en la que estamos interesados. En resumen, de estas conclusiones proponemos que:

a) Deberíamos dejar de utilizar el contraste de significancia de Lindley, y utilizar el método de comparación de probabilidades o (en su caso) teoría de decisiones, como el método general para contrastar hipótesis en estadística Bayesiana.

b) Hacer un cuestionamiento minucioso de la hipótesis  $H_1 : \theta = \theta_0$ , si nos enfrentamos a un contraste como (1), intentando replantearla en una hipótesis compuesta, antes de utilizar un modelo discontinuo para  $\theta$ .

## Apéndice A



## Apéndice A

### Medida e Integración de Lebesgue

La teoría de la medida e integración de Lebesgue es una de las ramas más amplias e importantes del análisis matemático, la intención de este apéndice no es, ni remotamente, desarrollar este tema en general; por el contrario, la idea de este apéndice es entender someramente lo que es una medida en  $R$  y saber como calcular integrales de Lebesgue sólo para algunos casos particulares importantes.

Los textos sobre teoría de la medida e integración son bastante avanzados y se requiere de conocimientos de análisis como continuidad en espacios métricos, integral de Riemann-Stieljes etc. y nociones básicas de topología. Para una exposición completa del tema puede consultarse Pfeffer (1977) o Fernández (1976).

En los puntos A.1, A.2 y A.3 se hace una exposición muy somera de los preliminares más esenciales para el estudio de esta teoría y decidí incluirlos sólo para hacer más redondo el trabajo, mi intención no es explicar detalladamente estos tres primeros puntos, pero los puntos A.4, A.5 y sobre todo A.6 deberán leerse con cuidado para entender lo que es una medida en  $R$  y como podrían calcularse las integrales de Lebesgue que aparecen en la tesis.

En el apéndice, como en toda la tesis, vamos a denotar el conjunto de los números reales con  $R$ , al conjunto de los números reales extendidos con  $R^*$  ( $R^* = R \cup \{-\infty, \infty\}$ ), con  $Q$  al conjunto de los números racionales, con  $P(\Omega)$  a la familia de subconjuntos del conjunto  $\Omega$  y con  $\chi_A(x)$  a la función indicadora de  $A$  ( $\chi_A(x) = 1$  si  $x \in A$  y  $\chi_A(x) = 0$  si  $x \notin A$ ).

#### A.1 Espacio Medible

Para estudiar los espacios medibles necesitamos la siguiente definición:

Def: Sea  $\Omega$  un conjunto, decimos que  $\mathcal{A} \subset P(\Omega)$  es una  $\sigma$ -álgebra si:

- 1)  $\Omega \in \mathcal{A}$ .
- 2) Si  $A \in \mathcal{A} \Rightarrow \Omega - A \in \mathcal{A}$ .
- 3) Si  $\{A_i\} \ i = 1, \dots, \infty$  con  $A_i \in \mathcal{A}$  entonces:

$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}.$$

Como  $\Omega \in \mathcal{A}$ , entonces  $\Omega - \Omega = \emptyset \in \mathcal{A}$  y si  $\{A_i\}$  con  $A_i \in \mathcal{A}$  como  $\Omega - A_i \in \mathcal{A}$  entonces  $\Omega - \bigcup (\Omega - A_i) \in \mathcal{A}$ . El ejemplo más sencillo de  $\sigma$ -álgebra de un conjunto  $\Omega$  es  $\{\Omega, \emptyset\}$ . Para estudiar otros ejemplos veamos la siguiente definición:

<sup>1</sup> Véase Pfeffer (1977, p.12).

**Def:** Sea  $\Omega$  un conjunto y  $\mathcal{F} \subset P(\Omega)$ , llamamos a  $\sigma(\mathcal{F})$  la  $\sigma$ -álgebra generada por  $\mathcal{F}$ , a la intersección de todas las  $\sigma$ -álgebras que contienen a  $\mathcal{F}$ .

Es fácil demostrar que  $\sigma(\mathcal{F})$  en realidad es una  $\sigma$ -álgebra y que  $\mathcal{F} \subset \sigma(\mathcal{F})$ ; se suele decir que  $\sigma(\mathcal{F})$  es la mínima  $\sigma$ -álgebra que contiene a  $\mathcal{F}$ .

Si tenemos  $\Omega = \mathbb{R}$  la  $\sigma$ -álgebra de Borel, que llamaremos  $\mathcal{B}$ , es la  $\sigma$ -álgebra generada por la familia  $\mathcal{F} = \{(-\infty, a) \mid a \in \mathbb{R}\}$ ; como  $\mathcal{F} \in \mathcal{B}$  entonces  $(-\infty, a) \in \mathcal{B}$  para toda  $a \in \mathbb{R}$ , tenemos entonces que  $[a, \infty) \in \mathcal{B}$  y  $(-\infty, a] = \bigcap_{i=1}^{\infty} (-\infty, a + 1/i) \in \mathcal{B}$ ; de esta manera observamos que todos los intervalos cerrados o abiertos de  $\mathbb{R}$  están en  $\mathcal{B}$ , y también como  $\{a\} = (-\infty, a] \cap [a, \infty)$  resulta que todo conjunto unitario  $\{a\}$  está en  $\mathcal{B}$ . Observemos entonces que si  $A$  es numerable  $A = \cup\{a_i\}$  con  $a_i \in \mathbb{R}$   $i = 1, \dots, \infty$  y resulta que  $A \in \mathcal{B}$ , y tenemos que todo conjunto numerable está en  $\mathcal{B}$ .

Veamos este otro ejemplo de  $\sigma$ -álgebra; Sea  $M \subset \mathbb{R}$  finito o numerable y tomemos  $\Omega = M$ , a  $M$  lo vamos a llamar el conjunto de puntos de masa y vamos a tomar como  $\sigma$ -álgebra de este a  $P(M)$ . Nótese que  $P(\Omega)$  es siempre una  $\sigma$ -álgebra pues  $\Omega, \Omega - A$  y  $\cup A_i$  pertenecen a  $P(\Omega)$  porque son subconjuntos de  $\Omega$ .

Veamos ahora lo que es un espacio medible:

**Def:** Sea  $\Omega$  un conjunto y  $\mathcal{A} \subset P(\Omega)$ , llamamos a la dupla  $(\Omega, \mathcal{A})$  espacio medible si  $\mathcal{A}$  es una  $\sigma$ -álgebra.

Ejemplos de espacios medibles son entonces  $(\mathbb{R}, \mathcal{B})$  ó  $(M, P(M))$ . Vamos a llamar conjunto medible (de un espacio medible  $(\Omega, \mathcal{A})$  dado) si este conjunto pertenece a  $\mathcal{A}$ .

## A.2 Función medible

**Def:** Sea  $(\Omega, \mathcal{A})$  un espacio medible, decimos que  $f : \Omega \rightarrow \mathbb{R}$  es una función medible si para toda  $\alpha \in \mathbb{R}$ ,  $\{x \in \Omega \mid f(x) < \alpha\} \in \mathcal{A}$  ó  $f^{-1}((-\infty, \alpha)) \in \mathcal{A}$ .

Si tenemos el espacio medible  $(\mathbb{R}, \mathcal{B})$  cualquier función continua  $f : \mathbb{R} \rightarrow \mathbb{R}$  es función medible puesto que la imagen inversa de abiertos es un abierto y como todo abierto está en  $\mathcal{B}$  tenemos que  $f^{-1}((-\infty, \alpha)) \in \mathcal{B}$ . Puede demostrarse que si  $f$  tiene un conjunto finito de discontinuidades también es medible; hay muchas más funciones medibles para  $(\mathbb{R}, \mathcal{B})$  pero para nuestro estudio con conocer estas será suficiente.

Para el caso de  $(M, P(M))$  tenemos que toda función  $f : M \rightarrow \mathbb{R}$  es medible puesto que  $f^{-1}((-\infty, \alpha)) \subset M$  y por lo tanto  $f^{-1}((-\infty, \alpha)) \in P(M)$ .

Si  $(\Omega, \mathcal{A})$  es un espacio medible y  $f : \Omega \rightarrow \mathbb{R}$  es una función medible, llamamos  $f^+(x) = f(x) \chi_{f(x) > 0}(x)$  (la parte positiva de  $f$ ) y a  $f^-(x) =$

$-f(x)\chi_{f(x)<0}(x)$  (la parte negativa de  $f$ ). Es fácil demostrar que  $f^+$  y  $f^-$  también son medibles y que  $f = f^+ - f^-$ . También se tiene que si  $f$  y  $g$  son funciones medibles entonces  $f + g$ ,  $f - g$ ,  $fg$  y  $\alpha f$  son medibles para toda  $\alpha \in \mathbb{R}$ , y que si  $A$  es un conjunto medible  $g(x) = f(x)\chi_A(x)$  es una función medible.

### A.3 Funciones Simples

**Def:** Sea  $(\Omega, \mathcal{A})$  un espacio medible y  $f : \Omega \rightarrow \mathbb{R}$  una función medible, decimos que  $f$  es una función simple si  $f(\Omega)$  (el conjunto de valores que toma  $f$ ) es un conjunto finito (subconjunto de  $\mathbb{R}$ ).

En el espacio  $(\mathbb{R}, \mathcal{B})$  las funciones simples se ven como funciones escalonadas. En el espacio  $(M, P(M))$  si  $M$  es finito todas las funciones medibles son simples. Los siguientes son dos teoremas básicos en la teoría de la medida, para las demostraciones correspondientes puede consultarse la bibliografía mencionada:

**Teorema 1:** Si  $(\Omega, \mathcal{A})$  es un espacio medible y  $f : \Omega \rightarrow \mathbb{R}$  es una función simple, entonces existen  $A_1, A_2, \dots, A_n$  conjuntos medibles y  $a_1, a_2, \dots, a_n$  números reales distintos tales que:

$$\Omega = \bigcup_{i=1}^n A_i \quad \text{y} \quad f(x) = \sum_{i=1}^n a_i \chi_{A_i},$$

y  $A_1, A_2, \dots, A_n$  y  $a_1, a_2, \dots, a_n$  son únicos salvo por el orden de estos.

**Teorema 2:** Si  $(\Omega, \mathcal{A})$  es un espacio medible y  $f : \Omega \rightarrow \mathbb{R}$  una función medible no negativa, entonces existe una sucesión  $\{s_n\}$  de funciones simples con  $0 \leq s_n \leq s_{n+1} \leq f$  tal que:

$$s_n(x) \rightarrow f(x).$$

Vemos del teorema 1 que toda función simple es suma finita de funciones de la forma  $a\chi_A$  con  $A$  medible y el teorema 2 indica que toda función medible se puede aproximar puntualmente por una sucesión creciente de funciones simples.

### A.4 Medida

**Def:** Sea  $(\Omega, \mathcal{A})$  un espacio medible, una función  $\mu : \mathcal{A} \rightarrow \mathbb{R}^+$  es una medida en  $(\Omega, \mathcal{A})$  si:

$$1) \mu(\emptyset) = 0.$$

2) Si  $\{A_i\}$   $i = 1, \dots, \infty$  es una colección de conjuntos medibles disjuntos dos a dos ( $A_i \cap A_j = \emptyset, i \neq j$ ) entonces:

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

Obsérvese que si  $\mu(\Omega) = 1$  entonces  $\mu$  es una función de probabilidad. Intuitivamente, una función de medida mide a cualquier conjunto de la  $\sigma$ -álgebra correspondiente siguiendo las reglas 1) y 2) que intuitivamente me parecen claras: La medida del conjunto vacío es cero, y la medida de una unión de conjuntos disjuntos es la suma de las medidas de cada conjunto.

Para el espacio  $(M, P(M))$  tenemos la medida de *conteo* que es la función  $c: P(M) \rightarrow R^+$  que se define para  $A \in P(M)$  finito como  $c(A) =$  número de elementos en  $A$  y  $c(A) = \infty$  si  $A$  es infinito (numerable). En este caso medimos a un conjunto según el número de elementos que tiene.

Si ahora tenemos  $\Omega = R$  vamos a trabajar normalmente con la medida de Lebesgue, esta medida es muy difícil de construir pero tiene muchas propiedades muy útiles. La medida de Lebesgue está definida en una  $\sigma$ -álgebra  $\mathcal{A}^*$  que contiene a  $\mathcal{B}$  (la  $\sigma$ -álgebra de Borel) y resulta que toda función medible usando la  $\sigma$ -álgebra  $\mathcal{B}$  lo es según la  $\sigma$ -álgebra de Lebesgue  $\mathcal{A}^*$  (en particular las funciones continuas o con un número finito de discontinuidades). La medida de Lebesgue la llamamos  $\lambda$  y como característica básica de esta tenemos que  $\lambda((a, b)) = b - a$  (la medida de un intervalo cualquiera es la longitud de este). Intuitivamente, la medida de Lebesgue mide la longitud de los conjuntos, si tenemos  $b < c$  entonces  $\lambda((a, b) \cup (c, d)) = b - a + d - c$  y si  $a < c < b$  resulta que  $\lambda((a, b) \cup (c, d)) = d - a$ . Sin embargo,  $\lambda((a, b) \cap Q) = 0$  por que  $Q$  (el conjunto de los números racionales) es numerable y  $\lambda(Q) = 0$  por lo tanto,  $\lambda((a, b) \cap Q) = 0$ . Esto nos dice que  $\lambda$  sólo asigna medidas positivas a conjuntos *llenos o gordos* pues aún cuando  $Q$  es denso en  $R$  (entre cada dos reales hay un elemento de  $Q$ )  $Q$  es muy *flaco* en comparación con todo  $R$  por que  $Q$  es numerable y  $R$  no.

En el capítulo 2 se utiliza una medida para  $(R, \mathcal{A}^*)$  que en cierta forma es una combinación de  $c$  y  $\lambda$ . La llamamos medida discontinua (en  $x_0$ ) y se define como:

$$\mu(A) = \lambda(A) + \chi_A(x_0).$$

Esta medida es igual a la medida de Lebesgue salvo que se le suma 1 si  $x_0$  está en el conjunto a medir; esta medida podría generalizarse para un conjunto de puntos  $N = \{x_1, x_2, \dots, x_n\}$  poniendo:

$$\mu(A) = \lambda(A) + \sum_{i=1}^n \chi_A(x_i).$$

$\mu(A)$  es igual a  $\lambda(A)$  más el número de elementos de  $N$  que estén en  $A$ .

### A.5 Espacios de Medida

**Def:** Sea  $(\Omega, \mathcal{A})$  un espacio medible, llamamos a la tripleta  $(\Omega, \mathcal{A}, \mu)$  un *espacio medible* si  $\mu: \mathcal{A} \rightarrow \mathbb{R}^*$  es una medida.

Cualquier espacio de probabilidad es un espacio de medida, otros ejemplos de espacios de medida son  $(M, P(M), c)$  ó  $(\mathbb{R}, \mathcal{A}^*, \lambda)$ . Cuando hablemos de funciones medibles en un espacio de medida  $(\Omega, \mathcal{A}, \mu)$  nos referimos a funciones medibles en  $(\Omega, \mathcal{A})$ .

### A.6 Integral de Lebesgue

Vamos a definir primero la integral de Lebesgue para una función simple no negativa:

**Def:** Sea  $(\Omega, \mathcal{A}, \mu)$  un espacio de medida y  $s$  una función simple en éste con  $s \geq 0$  y  $s(x) = \sum a_i \chi_{A_i}(x); i = 1, \dots, n$ . La *integral de Lebesgue* de  $s$  es  $\int s d\mu$  y se define como:

$$\int s d\mu = \sum_{i=1}^n a_i \mu(A_i).$$

En el espacio  $(M, P(M), c)$  una función simple la podemos expresar como  $s(x) = \sum a_i \chi_{A_i}(x)$  con  $a_i = s(x_i)$  y  $A_i = \{x_i\}$  para todo  $x_i \in M$  con  $s(x_i) \neq 0$ , entonces como  $c(A_i) = 1$  tenemos que  $\int s d\mu = \sum s(x_i)$ , la suma de los valores de  $s$ . En el caso de  $(\mathbb{R}, \mathcal{A}^*, \lambda)$  la integral de Lebesgue de una función simple positiva, que se puede ver como una función escalonada, es la suma de las áreas bajo los escalones.

**Def:** Sea  $(\Omega, \mathcal{A}, \mu)$  un espacio de medida y  $f \geq 0$  una función medible, definimos  $\int f d\mu$  la integral de Lebesgue de  $f$  como:

$$\int f d\mu = \sup\{\int s d\mu \mid 0 \leq s \leq f, \text{ con } s \text{ función simple}\}.$$

Como vimos en el teorema 2 existe una sucesión de funciones simples  $0 \leq s_n \leq f$  tal que  $s_n \rightarrow f$ , de esta manera  $\int s_n d\mu$  es una aproximación a lo que debe ser  $\int f d\mu$  y por eso definimos  $\int f d\mu = \sup\{\int s d\mu \mid 0 \leq s \leq f\}$ .

Finalmente definimos  $\int f d\mu$  para cualquier función medible como  $\int f d\mu = \int f^+ d\mu - \int f^- d\mu$ .

En el caso de  $(M, P(M), c)$  es fácil demostrar que para cualquier función medible  $\int f d\mu = \sum_{x \in M} f(x)$  (la suma sobre los valores que toma  $f$ ). En el

caso de  $(R, \mathcal{A}^*, \lambda)$  si tenemos una función medible  $f: R \rightarrow R$  continua o con un número finito de discontinuidades,  $\int f d\lambda = \int_{-\infty}^{\infty} f(x) dx$  (la integral de Riemann) y si tenemos una medida discontinua en  $x_0$  entonces  $\int f d\mu = \int_{-\infty}^{\infty} f(x) dx + f(x_0)$ .

Definimos para todo  $A$  medible a  $\int_A f d\mu = \int f \chi_A d\mu$  (está bien definida puesto que  $f \chi_A$  es medible) y tenemos que:

$$\int_{[a,b]} f d\lambda = \int_a^b f(x) dx$$

si  $f$  es continua o con un número finito de discontinuidades, y si  $f: M \rightarrow R$  tenemos que:

$$\int_A f dc = \sum_{z \in M \cap A} f(z).$$

### A.7 Densidades Generalizadas

Veamos ahora como se utiliza el material anterior dentro de la teoría de las probabilidades. Sea  $X$  una variable aleatoria (v.a.) discreta, podemos tomar  $M = \{x \in R \mid P(X = x) > 0\}$  (los puntos de masa de  $X$ ) y entonces tomando  $f(x) = P(X = x)$  resulta que:

$$\int_A f dc = \sum_{z \in M \cap A} f(z) = \sum_{z \in M \cap A} P(X = z) = P(A)$$

y podemos generar a la función de probabilidad de  $X$  mediante  $f$  con la relación  $P(A) = \int_A f dc$ . Decimos entonces que  $f$  es una función de densidad (generalizada) con respecto a la medida de conteo  $(c)$ . Un ejemplo de esto podría ser  $X \sim \text{Poisson}(\theta)$   $M = \{1, 2, 3, \dots\}$  y

$$f(x) = \exp(-\theta) \frac{\theta^x}{x!},$$

que es la función de probabilidad de  $X$ .

En el caso de tener  $X$  v.a. continua, su densidad es una función  $f: R \rightarrow R$  continua o con un número finito de discontinuidades (como la densidad Normal, la Uniforme continua, la densidad F, etc.); en este caso  $f$  es una densidad (generalizada) con respecto a la medida de Lebesgue  $\lambda$  puesto que:

$$P(X < \alpha) = \int_{-\infty}^{\alpha} f(x) dx = \int_{[-\infty, \alpha]} f d\lambda.$$

Se puede generar entonces la función de probabilidad de  $X$  a partir de  $f$  y en general,  $P(A) = \int_A f d\lambda$ . Dentro esta tesis tenemos que  $X \in F_H$  y el parámetro  $\theta$  se toma como una variable aleatoria, al escribir:

$$\int_A f(\theta) d\mu(\theta)$$

nos referimos a la integral de Lebesgue con respecto a la medida  $\mu(\theta)$ , que es precisamente la medida con respecto a la cual  $f$  es una densidad. Si  $\theta$  es una v.a. discreta entonces  $\mu(\theta)$  es la medida de conteo y si  $\theta$  es una v.a. continua, entonces  $\mu(\theta)$  es la medida de Lebesgue. Un ejemplo de una integral de Lebesgue que aparece en la tesis es:

$$L_1^* = \int L_1(\theta) f(\theta | z_n) d\mu(\theta)$$

entonces, si  $\theta$  es una v.a. discreta esto se calcula:

$$L_1^* = \sum L_1(\theta_i) f(\theta_i | z_n),$$

sobre los puntos  $\theta_i$  con  $P(\theta = \theta_i) > 0$ , y si  $\theta$  es una v.a. continua entonces:

$$L_1^* = \int_{-\infty}^{\infty} L_1(\theta) f(\theta | z_n) d\theta.$$

Apéndice B

ESTA TESIS NO DEBE  
SALIR DE LA BIBLIOTECA



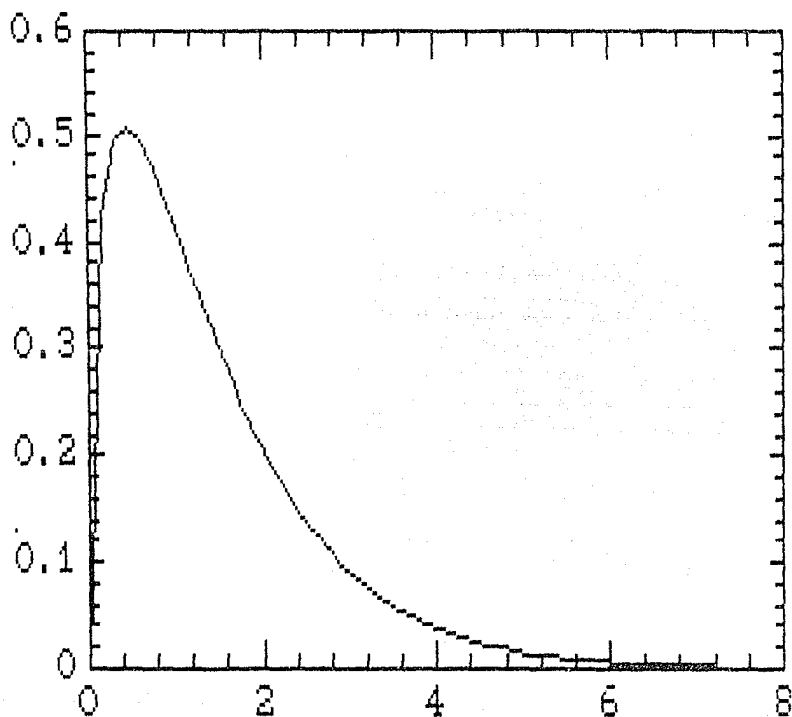
Simulación de muestreo exponencial,  $\exp(0.3)$ . 100 Observaciones:

0.20	1.06	0.98	2.25	3.67
1.43	4.47	1.91	0.06	1.91
5.02	7.01	1.00	1.66	0.95
0.91	5.59	6.56	3.11	0.96
1.25	9.07	2.08	1.21	9.34
1.04	0.26	2.11	6.51	7.19
6.58	0.47	1.91	2.72	4.22
0.28	5.15	5.38	2.76	20.84
3.50	0.61	0.25	0.99	0.56
9.76	8.52	1.37	16.01	1.11

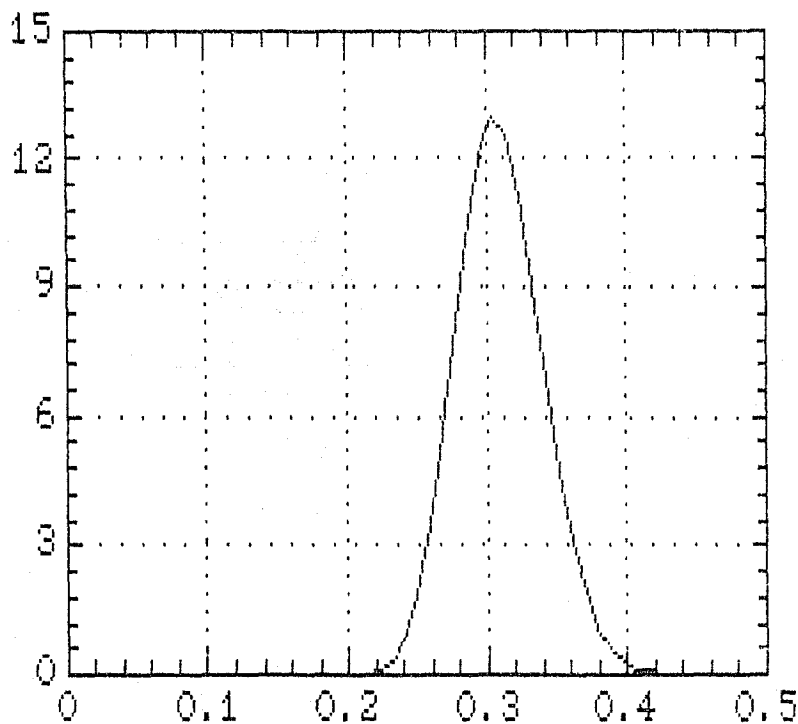
0.38	3.03	2.28	4.56	0.15
0.05	0.10	0.76	0.01	4.54
4.58	7.37	7.56	1.46	5.28
2.74	0.62	6.45	2.64	0.99
0.54	1.71	4.12	5.67	4.03
0.41	2.52	1.33	0.69	1.24
0.67	1.78	0.64	0.70	15.33
0.40	5.71	1.06	0.21	0.24
1.59	11.64	2.24	1.12	0.60
1.16	6.20	1.01	3.28	1.20

$$\sum_{i=1}^{100} x_i = 327.65$$

Distribucion inicial: Gama(0.4348, 1.0).



Distribución posterior:  $\text{Gamma}(100.4348, 0.003)$ .



## Apéndice C

PROGRAMA: T S T  
TURBO PASCAL V. 3.0

Andrés Christen.

Este programa se divide en dos procedimientos, el primero ("grafica") es un graficador de densidades  $t$  de Student generalizadas y el segundo ("calculap") calcula probabilidades como  $P\{X \leq a\}$  ó  $P\{|X-p| \leq a\}$  para  $X$  con distribución  $t$  generalizadas y también calcula intervalos de máxima densidad a cualquier probabilidad.

El procedimiento "grafica" utiliza el modo gráfico de turbo pascal para realizar las graficaciones. Todas las densidades se grafican centradas en su media ("med") fijando el punto máximo de estas en un lugar constante en la pantalla. Para imprimir los gráficos se utiliza la utilidad de MSDOS "Graphics". En Turbo pascal V.3 no se puede combinar modo gráfico con modo de texto es por esto, que para imprimir los valores de la escala y demás parámetros, es necesario hacerlo mediante el archivo de salida "lst".

El procedimiento "calculap" utiliza la regla de Simpson para hacer la integración numérica. Primero calcula la constante de integración, integrado el núcleo de la distribución  $t$  correspondiente, de la media hasta que la función sea muy cercana a cero y multiplicando el resultado por dos. Para calcular los intervalos de máxima densidad se desarrolla una búsqueda binaria para encontrar el cuantil que acumule la probabilidad  $1-(1-p)/2$  si la probabilidad del intervalo es  $p$  y este cuantil, junto con su simétrico con respecto a la media, forman el intervalo de máxima densidad deseado.

```
program t (input,output);
const sub=20;
var
    Iton,k:integer;
    prob,t0,med,t0:real;
    op:char;
{ $I graf.esp } (* Utileria para el procedimiento "Linea" *)
procedure leapar; (* Este procedimiento captura los parámetros *)
var (* de la distribución t *)
    k1:integer;
begin
repeat
writeLn;
writeLn(' Los grados de libertad es un entero mayor que cero');
writeLn('(si esta es cero se toman los parámetros que se usaron)');
writeLn('la ultima vez');
writeLn;
```

```

write('Desea la suma de libertad ');
readln(k);
if k<0 then
begin
k:=1;
writeln;
writeln(' La media es cualquier real. ');
writeln;
write('Dame la media ');
readln(med);
writeln;
writeln(' La precisión es un real mayor que cero. ');
writeln;
write('Dame la precisión ');
readln(tao);
end;
until ((k=0) and (tao>0));
writeln;
end;

procedure presentae; (* Presentación inicial *)
begin;
clrscr;
graphcolormede;
writeln;writeln;writeln;writeln;
writeln(' T S T ');
writeln;writeln;writeln;
writeln(' v 3.3 ');
writeln;writeln;writeln;
writeln(' Por Andres Christen ');
writeln;writeln;writeln;writeln;writeln;
writeln(' 1 9 3 9 ');
repeat until keypressed;
end;

function f(x:real):real; (* Esta funcion calcula la densidad *)
(* t en el punto x *)
var
aux:real;

begin
aux:=1+(tao/k)*(sqrt(x-med));
sux:=1/aux;
f:=exp(-((k+1)/2)*ln(sux));
end;

procedure grafica; (* GRAFICA *)
var
i:integer;
lax,y,y1,lon,l.l.s;
c,e,sux,x:real;
esp:string[12];

```

intervalo,sim,marca:boolean;

```
procedure marca(color:integer);      (* Dibuja los ejes *)
var
  i:integer;
begin
  linea(155,0,155,190,color);
  linea(0,sub,320,sub,color);
  for i:=0 to 6 do
    linea(20+45*i,0,20+45*i,30,color);
end;
```

```
procedure escala;      (* Escribe los valores de la escala *)
begin
  writeln('T student:',k,',',med:4:3,',',tao:4:3,');
  writeln;
  write(' ');
  for i:=0 to 6 do
    write(3+45*i*h:4:3,esp);
  writeln;
end;
```

```
begin
  clrscr;
  writeln;
  writeln;
  writeln('Despues que aparezca el "#" en la gráfica, presione cualquier ');
  writeln('tecla para continuar.');
```

```
repeat
  writeln;
  writeln;
  leepar;
  writeln('Dame lon≠0 (longitud de la media al extremo de la gráfica)');
  writeln('Si esta es cero, se toma lon=cinco veces la desviacion estandar');
  readln(lon);
  if lon=0 then lon:=5*sqrt(1/tao);
  writeln('Quieres algún valor especial que se marque en la gráfica?');
  writeln('(NO=,1234)? ');
  readln(c);
  marca:=((c=0,1234);
  if marca then
  begin
    write('Quieres que se marque su simétrico (1=si) ');
    readln(aux);
    sim:=(aux=1);
  end;
  writeln;
  write('Quieres que se marque el último intervalo de MO calculado (1=si) ');
  readln(aux);
  intervalo:=(aux=1);

while lon≠0 do
begin
```

```

linea(2*trunc(v2/2));
atrad:=1;
repeat
write('Presione cualquier tecla para continuar...');
readln(v1);
esca1a;
write('Esta bien (1=si) ');
readln(v);
until (v=1);
v1:=v2;
v1:=t(a);
graphcol:=mord;
marca(2);
for i:=0 to 290 do      (* Graficacion de la densidad *)
begin
x:=a+(i-20)*h;
y2:=f(x);
linea(i-1, trunc(a*(1+sub.i, trunc(ef*y2)+sub.3));
v1:=v2;
end;      (* Marcus *)

if marca then
linea(20+trunc((c-a)/h),0,20+trunc((c-a)/h),30,3);
if sim then
linea(20+trunc(((2*med-c)-a)/h),0,20+trunc(((2*med-c)-a)/h),30,3);
if intervalo then
begin
linea(20+trunc((t0-a)/h),0,20+trunc((t0-a)/h),30,1);
linea(20+trunc(((2*med-t0)-a)/h),0,20+trunc(((2*med-t0)-a)/h),30,1);
end;
write('# ');
repeat until keypressed;
clrscr;
esca1a;
write('Quieres imprimir (1=si) ');
readln(v2);
if v2=1 then
begin
writeln(lst);      (* Impresión mediante lst *)
writeln(' ');
for i:=0 to F do
write(lst,chr(57)*i*4:2,esp);
writeln(lst);
writeln(lst,'E-student: ',t,',' ,med:4:1,' ,fac:4:1,' ');
writeln(lst);
if marca then writeln(lst,' Marca : ',c:4:3);
writeln(lst);
if sim then writeln(lst,' Simétrico: ',2*med-c:4:3);
writeln(lst);
if intervalo then
begin
writeln(lst,' Intervalo de MD : ( ',2*med-t0:4:1,' , ',2*med-t0:4:1,' ');
writeln(lst);
writeln(lst,' Probabilidad : ',1-2*prob:4:3);

```



```

end;
end;
writeLn('Repetir otra lon=2');
writeLn('Incidencia de la media al extremo de la grafica');
writeLn('Si esta es cero, se toma lon=5 veces la desviación estandar');
writeLn('la lon=1, termina');
readLn(lon);
if lon=0 then lon:=5*sqrt(1/tao);
end;
write('Quieres graficar otra t (t=si) ');
readLn(e);
clrscr;
until e<>'1';
end; (* FIN GRAFICA *)

```

(\* CALCULAR \*)

```

procedure calculap;

```

```

var
  piv:real;
  op:char;
  Cons:real;
function calp(b:real):real; (* Esta función realiza la *)
var (* Integración numerica *)
  j:integer;

```

```

x1,x2,x0,
  x,h,xi:real;

```

```

begin
  a:=med;
  if b>a then
  begin
    h:=(b-a)/(2*Iter);
    x0:=T(a)+T(b);
    x1:=0;
    x2:=0;
    for j:=1 to 2*Iter-1 do
    begin
      xi:=a+j*h;
      if odd(j) then x1:=x1+T(xi)
      else x2:=x2+T(xi);
    end;
    x0:=(h*(x0+2*x2+4*x1))/3;
    calp:=x0/Cons;
  end;
end;
end;

```

```

function distri(ac:real):real; (* Calcula cualquier valor de la *)
begin (* distribución t *)
  if ac=med then distri:=0.5
  else
    if ac>med then distri:=0.5+calp(ac)
    else distri:=0.5-calp(2*med-ac);
  end;
end;

```

```

end;

procedure proba;          (* Escribe los parametros *)
begin
  writeln;
  writeln(' G. L. : ',g);
  writeln(' Media : ',med:4:3);
  writeln(' Des. : ',fac:4:4);
end

procedure proba;          (* Calcula probabilidades acumuladas *)
var
  t1,orbital;
begin
  clrscr;
  writeln;
  write(' Dame el valor de la "t" : ');
  readln(t1);
  prb:=distrit(t1);
  writeln;
  writeln(' La probabilidad acumulada hasta ',t1:4:3,' es: ',prb:1:4);
  repeat until keypressed;
end;

procedure inter;          (* Calcula intervalos de Maxima Densidad *)
var
  t1,t2,alfa;
begin
  repeat
    clrscr;
    writeln;
    writeln(' ¿ Cual es la probabilidad del intervalo de maxima densidad ? ');
    writeln(' La probabilidad tiene que ser entre 0 y 1');
    readln(alfa);
  until (alfa>0) and (alfa<1);
  t1:=med*(5.4)/sqrt(2*alfa);
  t2:=med;

  repeat                    (* Busqueda binaria *)
    t0:=(t1+t2)/2;
    prob:=0.5+calp(t0);
    if prob>((1-alfa)/2) then
      t1:=t0
    else
      t2:=t0;
  until abs(prob-(1-alfa)/2)<0.00009;

  writeln;

```

```

writeIn;
writeLn(' El intervalo es : [-(' ,2*med-t0:4:3,' , ',t0:4:3,' )];');
writeIn;
writeLn(' El radio del intervalo es : ',r0:med:4:3);
writeIn;
writeLn(' El valor de la funcion de densidad en ',t0:4:3);
writeLn(' es : ',f(t0)/Const:4:3);
writeIn;
writeLn;
repeat until keypressed;
end;

procedure vecindad;      (* Calcula probabilidades alrededor de *)
var                      (* un cierto radio *)
    piv,radio:real;
begin
    clrscr;
    writeLn;
    writeLn(' El pivote de la vecindad es : ',piv:4:3);
    writeLn;
    writeLn(' Cual es el radio : ');
    readLn(radio);
    piv:=distri((piv+radio)-distri(piv-radio));
    writeLn;
    writeLn(' La vecindad es ( ',piv+radio:4:3,' , ',piv-radio:4:3,' )');
    writeLn;
    writeLn(' y acumula una probabilidad de : ',pib:1:4);
    repeat until keypressed;
end;

procedure calConst:      (* Calcula la constante de integraci3n *)
var
    i:integer;
    x:real;
begin
    writeLn;
    writeLn(' Calculando constante de integraci3n ');
    x:=med;
    i:=1;
    while f(x)>0.00009 do
        begin
            x:=med+i/sqrt(tao);
            i:=i+1;
        end;
    writeLn(i);
    Const:=2*falm(x);
end;

```

```

begin
  clrscr;
  const:1;
  writeln(' Procedimiento para calculo de probabilidades. ');
  writeln;
  writeln(' Parametros de la distribucion: ');
  writeln;
  writeln;
  l:epar;
  writeln;
  write(' Valor del pivote (Ignoralo si no quieres ninguno) : ');
  readln(piv);
  calCons;
  repeat
  clrscr;
  writeln(' Para calcular una probabilidad - P ');
  writeln;
  writeln(' Intervalo de máxima densidad - I ');
  writeln;
  writeln(' prob. de una Vecindad del pivote - V ');
  writeln;
  writeln(' Modificar el numero de iteraciones - M ');
  writeln(' (se encuentra con ',2*iter,' ');
  writeln;
  writeln(' Fin - F ');
  writeln;
  write(' -> ');
  readln(op);
  case op of
    'I','p':proba;
    'I','i':inter;
    'M','m':begin (* Modifica el numero de iteraciones en *)
      repeat (* la integración numerica *)
        clrscr;
        writeln;writeln;
        write(' Cuantas iteraciones quieres (mas de 4) ? ');
        readln(Iter);
        until Iter>4;
        Iter:=trunc(Iter/2);
        calCons;
      end;
    'V','v':vecindad;
  end;
  until (op='F') or (op='I');
end; (* FIN CALCULAR *)

```

```

begin (* PROGRAMA PRINCIPAL *)
  presenta;
  iter:=25;
  repeat
  clrscr;
  writeln;writeln;

```

```

writeLn(' Este programa grafica t de student generalizadas. ');
writeLn(' todas las grafias con centrada en la media. ');
writeLn;
writeLn(' Calcula tambien probabilidades para valores de la "t" ');
writeLn(' e intervalos de máxima densidad. ');
writeLn;writeLn;
writeLn(' Para graficar = G ');
writeLn;
writeLn(' Para calcular = C ');
writeLn;
writeLn(' F I R = F ');
writeLn;
write(' --> ');
readLn(op);
case op of
  'G', 'g': grafica;
  'C', 'c': calcula;
end;
until (op='F') or (op='f');
end. (* FIN *)

```

## Bibliografia

## Bibliografía

Bakahvalov (1977), *Numerical Methods*.

Mir, Moskow.

Berger (1980), *Statistical Decision Theory*.

Springer-Verlag, N.Y.

Box and Tiao (1973), *Bayesian Inference in Statistical Analysis*.

Addison-Wesley, Mass. USA.

Broemeling, L (1985), *Bayesian Analysis of Linear Models*.

Marcel Dekker, Inc, USA.

Brunk, H.D (1979), *Introducción a la Estadística Matemática*.

Trillas. México.

Burden y Fairies (1985), *Análisis Numérico*.

Grupo editorial Iberoamérica, Mexico.

Dalal, S.R. and Hall (1983), *Approximating Priors by Mixtures of Natural Conjugate Priors*.

Journal of the Royal Statistical Society-B V.45 N.2 p.278-286, U.K.

DeGroot (1970), *Optimal Statistical Decisions*.

McGrow-Hill, USA.

Fernandez, P. (1970), *Medida e Integração*.

IMPA, Brasil.

Ferrandiz (1985), *Bayesian Inference on Mahalonobis Distance*.

Bayesian Statistics 2 p.645-654, North-Holland, Netherlands.

Jeffreys, H. (1961), *Theory of Probability*.

Oxford University Press, U.K.

Lindley (1957), *A statistical Paradox*.

Biometrika 44 p.187-192.

Lindley (1970), *Introduction to Probability and Statistics from a Bayesian View Point*.

Cambridge University Press, U.K.

Lindley (1971), *Making Decisions*.

Cambridge University Press, U.K.

Pfeffer, D.V. (1971), *Integrals and Measures*.

Dekker, N.Y.

Phillips, L. (1973), *Bayesian Statistics for Social Scientists*.

Nelson, USA.

Winkler, R.L. (1972), *An Introduction to Bayesian Inference and Decision*.

Holt, Rinehart and Watson, USA.

Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*.

Wiley, USA.