

# Teoría de verificación de las hipótesis

En los §§ 1—3, II se expone la teoría de verificación de un número finito (en particular, dos) de hipótesis simples.

Los §§ 4—12 están dedicados a los métodos de construcción de criterios óptimos para verificar dos hipótesis compuestas. En particular, se examinan los criterios bayesianos y minimax (los §§ 4 y 9) y se utilizan los principios de suficiencia, de carácter no desplazable y de invariación para construir los criterios uniformemente más potentes.

En los §§ 13—17 se estudian los métodos de construcción de criterios asintóticamente óptimos.

### § 1. Verificación de un número finito de hipótesis simples

**1. Planteamiento del problema. Concepto de criterio estadístico. Criterio más potente.** En este capítulo se tratará de la verificación de cualesquiera suposiciones (hipótesis) respecto a la distribución  $\mathbf{P}$  de la cual se ha extraído la muestra  $X$ . Aquí, al igual que en la teoría de las estimaciones, no existiría tal problema, si la distribución  $\mathbf{P}$ , de la cual se extrae la muestra  $X$ , fuera conocida.

La decisión de que es cierta o no la hipótesis dada  $H$  debe basarse exclusivamente en el conocimiento de la muestra  $X \in \mathbf{P}$  extraída y, posiblemente, también en el conocimiento de la información a priori respecto a  $\mathbf{P}$  si disponemos de ella.

Ahora bien, para determinar el procedimiento de toma de decisión basándonos en la muestra  $X$ , debemos establecer, de una u otra forma, la aplicación del espacio muestral  $\mathcal{Q}^n$  en el conjunto de hipótesis que se examinan. Tal aplicación suele llamarse criterio estadístico. Las definiciones exactas para diferentes situaciones concretas se darán más adelante.

Comencemos por el problema más simple: verificación de un número finito de hipótesis simples.

**Definición 1.** Llamaremos *hipótesis simple* cualquier suposición que define unívocamente la distribución de la muestra  $X$ .

Supongamos que se dan  $r$  distribuciones  $P_1, \dots, P_r$ , y supongamos que sabemos que  $X$  es la muestra de una de estas distribuciones. El problema consiste en determinar a qué  $P_j$  precisamente,  $j = 1, 2, \dots, r$ , pertenece  $X$ . Cada  $r$  hipótesis

$$H_j = \{X \in P_j\} \quad (1)$$

será simple y, por consiguiente, se tratará de la verificación de  $r$  hipótesis simples.

En este capítulo, al igual que en el capítulo 2, examinaremos con frecuencia el caso paramétrico cuando la muestra  $X$  se ha extraído de la distribución  $P_\theta \in \mathcal{P} = \{P_\theta\}_{\theta \in \Theta}$ . En este caso, al cumplirse las condiciones  $(A_0)$ , las hipótesis simples se escribirán en la forma:  $H_j = \{X \in P_{\theta_j}\}$ , donde  $\theta_1, \dots, \theta_r$  son los puntos fijos de  $\Theta$ . El caso (1) también puede considerarse como paramétrico con un conjunto finito  $\Theta = \{1, \dots, r\}$ .

Estos razonamientos muestran que no hay una diferencia de principio entre el problema de estimación de los parámetros y el problema de verificación de las hipótesis: en ambos casos determinamos el valor desconocido de  $\theta$ . Sin embargo, existe cierta diferencia y ésta consiste en que en el problema de verificación de las hipótesis, los valores posibles de  $\theta$  son discretos, y los enfoques relacionados con la comparación, digamos, de las desviaciones estándar, desarrollados en el capítulo 2, aquí son inaplicables. En este caso escogeremos otros criterios para comparar las reglas de aceptación de unas u otras hipótesis, basándonos en la muestra  $X$ .

Con el carácter discreto del conjunto de los posibles valores de  $\theta$  también está relacionada otra nueva cualidad que aparece aquí: ahora podemos, con una probabilidad no nula, indicar exactamente el valor desconocido de  $\theta_i$  (o la distribución  $P_{\theta_i}$ ), mientras que en los problemas de estimación de los parámetros, la probabilidad de tal suceso es, por regla general, igual a cero.

**Definición 2.** Se llama *criterio estadístico* para verificar  $r$  hipótesis  $H_1, \dots, H_r$  toda aplicación medible  $\delta: \mathcal{X}^n \rightarrow \{H_1, \dots, H_r\}$ .

En otros términos,  $\delta(X)$  es una "variable" aleatoria que toma los valores  $H_1, H_2, \dots, H_r$ : si  $\delta(X) = H_k$ , entonces aceptamos la hipótesis  $H_k$  (o sea, consideramos que  $\theta = \theta_k$  en el caso paramétrico).

La aplicación  $\delta(\cdot)$  se llama, a veces, *regla de decisión* o *función de decisión*. Claro está que la asignación de la regla de decisión es equivalente a la partición del espacio  $\mathcal{X}^n$  en  $r$  conjuntos borelianos  $\Omega_1, \Omega_2, \dots, \Omega_r$ , disjuntos, en los cuales se aceptan las hipótesis  $H_1, H_2, \dots, H_r$ , respectivamente.

La calidad del criterio se caracteriza, con más frecuencia, por el conjunto de probabilidades de decisiones erróneas:

$$\alpha_i = \alpha_i(\delta) = \mathbf{P}_i(X \notin \Omega_i) = \mathbf{P}_i(\delta(X) \neq H_i).$$

El número  $\alpha_i$  es la probabilidad de rechazar la hipótesis  $H_i$  cuando ésta es cierta. Este número se denomina *probabilidad del error de  $i$ -ésimo género del criterio  $\delta$* .

Si logramos escoger el criterio  $\delta$  de modo que todos los números  $\alpha_i$  sean pequeños, entonces, según nuestro principio fundamental mencionado en el § 2.31, consideraremos que en una sola prueba el error es prácticamente imposible y declararemos que es cierta la hipótesis  $H_k$  si  $\delta(X) = H_k$ . En este caso nos equivocaremos, aproximadamente, en parte de los casos  $\alpha_i = \mathbf{P}_i(\delta(X) \neq H_i)$  si en realidad es cierta  $H_i$ .

Es deseable, desde luego, efectuar la verificación de las hipótesis de modo que se reduzca al mínimo la probabilidad de todos los errores. No obstante, si se establece el volumen de la muestra  $X$ , entonces no podremos dirigir simultáneamente todas las probabilidades de los errores. Se puede sólo, fijando algunas de las probabilidades de errores, tratar de minimizar las demás.

Aquí llegamos a la cuestión de cómo comparar entre sí diferentes criterios. Introduzcamos en el conjunto de todos los criterios, para verificar las hipótesis  $H_1, \dots, H_r$ , un orden parcial.

**Definición 3.** El criterio  $\delta_1$  es mejor que el  $\delta_2$  si para todos  $i = 1, 2, \dots, r$

$$\alpha_i(\delta_1) \leq \alpha_i(\delta_2)$$

y al menos para un  $i$  tiene lugar la desigualdad estricta.

Sin embargo, los criterios  $\delta_1$  y  $\delta_2$  no siempre, ni mucho menos, pueden compararse desde este punto de vista. Al igual que pueden ser incomparables dos estimaciones  $\theta_1^*$  y  $\theta_2^*$  desde el punto de vista del enfoque estándar, cuando en calidad de criterio tomamos  $\mathbf{M}_\theta(\theta^* - \theta)^2$ . Para tener la posibilidad de comparar los criterios es necesario contraer el conjunto de las reglas de decisión que se examinan. Para esto examinemos las clases

$$K_{\alpha_1, \dots, \alpha_{r-1}} = \{\delta: \alpha_j(\delta) = \alpha_j; j = 1, 2, \dots, r-1\}.$$

En las clases  $K_{\alpha_1, \dots, \alpha_{r-1}}$  ya se puede establecer la relación de orden entre los criterios en cuanto a la magnitud  $\alpha_r$ : cuanto menor sea  $\alpha_r(\delta)$ , tanto mejor será el criterio.

**Definición 4.** El criterio  $\delta_0 \in K_{\alpha_1, \dots, \alpha_{r-1}}$  se llama *criterio más potente (c.m.p.) en la clase  $K_{\alpha_1, \dots, \alpha_{r-1}}$*  si para cualquier  $\delta \in K_{\alpha_1, \dots, \alpha_{r-1}}$ ,

$$\alpha_r(\delta_0) \leq \alpha_r(\delta).$$

Recordemos que hemos hecho algo semejante en el capítulo 2 al compa-

rar las estimaciones. Allí hemos destacado, por ejemplo, las clases  $K_b$  de estimaciones con desplazamiento registrado.

A la par con el enfoque recién introducido en la teoría de verificación de las hipótesis, al igual que en la teoría de estimaciones, existen otros dos enfoques que permiten ordenar el conjunto de todas las reglas de decisión con ayuda de una sola característica numérica: son los enfoques bayesiano y minimax.

Antes de estudiar los métodos de construcción de los criterios más potentes en las clases  $K_{\alpha_1, \dots, \alpha_r}$ , examinemos estos dos enfoques.

**2. Enfoque bayesiano.** Este enfoque supone que la distribución  $P_j$  de la que fue extraída la muestra  $X$  se ha elegido aleatoriamente. En este caso las hipótesis  $H_j = \{X \in P_j\}$ ,  $j = 1, \dots, r$  serán sucesos aleatorios, y designaremos las probabilidades de estos sucesos por

$$Q(H_j) = q(j),$$

así que  $Q$  es una distribución a priori en el conjunto de las hipótesis  $\{H_1, \dots, H_r\}$ , y  $q(j)$  son las probabilidades a priori de dichas hipótesis (compárese con el § 2.11). En este caso es más fácil comparar los criterios, puesto que aquí podemos determinar la probabilidad media  $\alpha_Q(\delta)$  del error del criterio  $\delta$ :

$$\alpha_Q(\delta) = \sum_{j=1}^r Q(H_j) P_j(\delta(X) \neq H_j) = \sum_{j=1}^r q(j) \alpha_j(\delta), \quad (2)$$

y de este modo ordenar por completo el conjunto de criterios en cuanto a la magnitud  $\alpha_Q(\delta)$ .

**Definición 5.** El criterio  $\delta = \delta_Q$  que minimiza la probabilidad del error  $\alpha_Q(\delta)$  se denomina *criterio bayesiano correspondiente a la distribución a priori  $Q$* .

Supongamos que se cumple la condición  $(A_\mu)$ , o sea, las distribuciones  $P_j$  tienen densidades  $f_j(x)$  respecto a cierta  $\sigma$ -finita medida  $\mu$ . Al igual que antes, la función  $f_j(X) = \prod_{j=1}^n f_j(x_j)$  se llamará *función de verosimilitud*.

La función  $f(x) = \sum q(j) f_j(x)$  es la densidad incondicional de la distribución de  $X$  respecto a la medida  $\mu^n$ , y  $q(j) f_j(x)$  es la densidad de la distribución compatible del par  $(\theta, X)$  en el que el número  $\theta$  de la hipótesis se elige al azar.

Ahora bien, si se da la muestra  $X$ , entonces, en el caso bayesiano se puede construir la distribución a posteriori  $Q_x$  de las hipótesis  $H_j$  (la medida  $\lambda$  que figura en el § 2.11, aquí es una medida de cálculo) la cual se determina por la fórmula de Bayes:

$$Q_x(H_k) = q(k/X) = \frac{q(k) f_k(X)}{f(X)}. \quad (3)$$

Esta es la *distribución condicional de  $\theta$  respecto a  $X$* .

Por  $\mathbf{M}$  designaremos la esperanza matemática incondicional que corresponde a la distribución  $\mathbf{P}$  del par  $(\theta, X)$ .

**Teorema 1.** 1) *La probabilidad del error  $\alpha_Q(\delta)$  de cualquier criterio  $\delta$  satisface la desigualdad*

$$\alpha_Q(\delta) \geq 1 - \mathbf{M} \max_j q(j/X). \quad (4)$$

2) *Para que el criterio  $\delta = \delta_Q$  sea bayesiano para la distribución a priori  $\mathbf{Q}$ , es necesario y suficiente que para  $\mathbf{P}$  de casi todos los valores de  $X$ , este criterio satisfaga las relaciones*

$$\delta(X) = H_k \quad \text{si} \quad q(k/X) = \max_j q(j/X). \quad (5)$$

Para  $\delta = \delta_Q$  en la desigualdad (4) se alcanza la igualdad.

Nótese que el segundo miembro en (4) no depende de  $\delta$ .

**Demostración.** Supongamos que se da el criterio  $\delta$ . Examinemos el suceso  $D_\delta$  que consiste en que el criterio  $\delta$  conduce a la decisión errónea:

$$D_\delta = \bigcup_{j=1}^r \{\theta = j, \delta(X) \neq H_j\}.$$

Entonces, evidentemente que  $\alpha_Q(\delta) = \mathbf{P}(D_\delta)$  y la notación (2) será el resultado obtenido al promediar sucesivamente: primero respecto a  $X$  al ser registrado  $\theta = j$ , y luego respecto a  $\theta$ . Pero también podemos escribir  $\alpha_Q(\delta)$  de otro modo: primero promediar respecto a  $\theta$ , siendo registrado  $(X)$ , y luego respecto a  $X$ :

$$\begin{aligned} \alpha_Q(\delta) &= \int \mathbf{P}(D_\delta/X = x) f(x) \mu(dx) = \\ &= \mathbf{M} \mathbf{P}(D_\delta/X) = \mathbf{M} \sum_{j=1}^r \mathbf{P}(\theta = j, \delta(X) \neq H_j/X). \end{aligned}$$

Pero  $\delta(X)$  es medible respecto a  $X$ , por eso

$$\mathbf{P}(\theta = j, \delta(X) \neq H_j/X) = I_{\{\delta(X) \neq H_j\}} \mathbf{P}(\theta = j/X) = (1 - I_{\{\delta(X) = H_j\}}) q(j/X).$$

De aquí obtenemos

$$\alpha_Q(\delta) = 1 - \mathbf{M} \sum_{j=1}^r q(j/X) I_{\{\delta(X) = H_j\}} \geq 1 - \mathbf{M} \max_j q(j/X).$$

La primera afirmación del teorema queda demostrada.

La suficiencia de la segunda afirmación del teorema se deduce con evidencia de la primera, ya que la frontera inferior establecida para  $\alpha_Q(\delta)$  se alcanza para el criterio  $\delta_Q$  definido en (5). La modificación de  $\delta_Q(X)$  en el conjunto de  $\mathbf{P}$ -probabilidad nula, por lo visto no modifica  $\alpha_Q(\delta_Q)$ .

La necesidad de la segunda afirmación se demuestra de manera igualmente sencilla. En efecto, supongamos que  $\delta = \delta_Q$  es el criterio bayesiano y que  $\delta(X) = H_k$ ,  $q(k/X) < q(l/X) = \max_j q(j/X)$  para  $X \in A$ ,  $P(A) > 0$ .

Entonces, para el criterio  $\delta_1(X)$ , que se distingue de  $\delta(X)$  sólo en el conjunto  $A$ :  $\delta_1(X) = H_l$  para  $X \in A$ , obtenemos

$$\begin{aligned} P(D_{\delta_1}; A) &= P(A) - \mathbf{M} \left[ \sum_j q(j/X) I_{\{\delta_1(X) = H_j\}}; A \right] = \\ &= P(A) - \mathbf{M}[q(l/X); A] < P(A) - \mathbf{M}[q(k/X); A] = P(D_{\delta}; A); \\ P(D_{\delta_1}) &< P(D_{\delta}) = P(D_{\delta_Q}). \end{aligned}$$

Hemos obtenido la contradicción.  $\Delta$

Cabe señalar que la notación (5) aún no define por completo el criterio  $\delta_Q$ : ella no aclara bien qué hipótesis deben aceptarse cuando resultaron máximos dos o más valores de  $q(j/X)$ . Se trata, evidentemente, de la definición de la función  $\delta_Q(X)$  en las fronteras

$$\Gamma_k = \{x \in \mathcal{X}^n: q(k)f_k(x) = \max_{j \neq k} q(j)f_j(x)\}$$

de los conjuntos

$$\tilde{\Omega}_k^Q = \{x \in \mathcal{X}^n: q(k)f_k(x) > \max_{j \neq k} q(j)f_j(x)\} \quad (6)$$

en los cuales, según (5), como criterio  $\delta_Q$  se toma la hipótesis  $H_k$ .

Por consiguiente,  $\tilde{\Omega}_k^Q$  es el "interior" de la región

$$\tilde{\Omega}_k^Q = \{x \in \mathcal{X}^n: \delta_Q(x) = H_k\}$$

de aceptación de la hipótesis  $H_k$  y necesitamos, en adición a (6), determinar tan sólo qué puntos de la frontera  $\Gamma_k$  pertenecen y no pertenecen a  $\tilde{\Omega}_k^Q$ . Pero este problema, como se deduce de los razonamientos citados, puede ser resuelto muy sencillamente: podemos asociar los puntos de  $\Gamma_k$  a cualquiera de las regiones "adyacentes"  $\tilde{\Omega}_j^Q$ , en este caso obtenemos el mismo valor de  $\alpha_Q(\delta)$ , puesto que (5) será cumplida. Mejor dicho, si  $A \subset \subset \Gamma_{k_1} \cap \dots \cap \Gamma_{k_r}$ , entonces para  $X \in A$ , según el criterio bayesiano, no importa cuál de las hipótesis  $H_{k_1}, \dots, H_{k_r}$  será aceptada. Podemos incluso tomar la decisión al azar, o sea, con probabilidad  $p_k$ , elegir la hipótesis  $H_{k_i}$ ,  $i = 1, \dots, r$ ,  $\sum_{i=1}^r p_{k_i} = 1$ . En este caso el valor de  $\alpha_Q(\delta)$  no variará.

Aquí llegamos a un concepto más general del criterio estadístico randomizado (de la palabra inglesa *random*) que resulta muy útil.

**Definición 6.** Se llama *criterio estadístico randomizado*, para comprobar las hipótesis  $H_1, \dots, H_r$ , cualquier aplicación medible  $\pi: \mathcal{X}^n \rightarrow R^{(r)}$ , donde  $R^{(r)}$  es el conjunto de vectores  $(\pi_1, \dots, \pi_r)$ ,  $\pi_i \geq 0$ ,  $\sum_{i=1}^r \pi_i = 1$ .

El criterio randomizado, a cada  $x \in \mathcal{Q}^n$  le pone en correspondencia con la distribución de las probabilidades  $\pi(x) = (\pi_1(x), \dots, \pi_r(x))$  en el conjunto  $\{H_1, \dots, H_r\}$ , y la decisión final acerca de la aceptación de la hipótesis "se sortea" al azar con esta distribución ya independientemente de  $X$ , después de haber determinado  $\pi_i(X)$ .

El criterio estadístico ordinario es, evidentemente, un caso particular del randomizado, cuando todos  $\pi_i$  equivalen a 0 y sólo uno es igual a 1. Tales criterios adquirieron el nombre de criterios *no randomizados*.

El error de  $i$ -ésimo género  $\alpha_i(\pi)$  para el criterio randomizado se determina análogamente:

$$\alpha_i(\pi) = P_i(\text{no aceptar } H_i) = 1 - M_i \pi_i(X).$$

En el caso bayesiano, el problema de minimización

$$\alpha_Q(\pi) = \sum_{j=1}^r q(j) \alpha_j(\pi)$$

se examina de manera absolutamente semejante. Si, como antes, designamos por  $\theta$  el número de la hipótesis elegida al azar, con una distribución a priori  $Q$ , de modo que  $Q(\theta = j) = q(j)$ , y por  $M$ , también como antes, designamos el símbolo de la esperanza matemática incondicional, entonces

$$\begin{aligned} \alpha_Q(\pi) &= 1 - \sum_{j=1}^r q(j) M_i \pi_i(X) = 1 - M \pi_\theta(X) = 1 - M M(\pi_\theta(X)/X) = \\ &= 1 - M \sum_{j=1}^r q(j/X) \pi_j(X) \geq 1 - M \max_j q(j/X). \end{aligned}$$

Así pues, hemos obtenido la misma frontera inferior tanto para  $\alpha_Q(\pi)$  como para los criterios no randomizados. Esto significa que ampliando la clase de criterios, en nuestro caso no podemos mejorar el valor de  $\alpha_Q(\delta)$ . Es más, el valor mínimo se alcanza en el criterio no randomizado  $\delta_Q$ . Sin embargo, en este caso el número de criterios *randomizados* bayesianos  $\pi^Q$ , o sea, de criterios para los cuales  $\alpha_Q(\pi^Q) = \alpha_Q(\delta_Q)$ , será mucho mayor que los no randomizados, ya que en el conjunto

$$\Gamma_{k_1, \dots, k_l} = \bigcap_{i=1}^l \Gamma_{k_i} \bigcap_{j \neq k_1, \dots, k_l} \bar{\Gamma}_j,$$

donde  $\bar{\Gamma} = \mathcal{Q}^n \setminus \Gamma$ , podemos tomar, en calidad de  $\pi^Q(x)$ , cualquier vector del subconjunto  $R_{k_1, \dots, k_l} \subset R^{(r)}$  compuesto de vectores  $\pi$  en los cuales sólo se diferencian del cero las coordenadas con números  $k_1, \dots, k_l$ . Es evidente que  $R_k$  se compone del único vector  $e_k$  en el que la  $k$ -ésima coordenada es igual a 1, y las demás, a cero, y debemos poner

$$\pi^Q(x) = e_k \quad \text{cuando} \quad x \in \bar{\Omega}_k^Q.$$

Como las relaciones expuestas con una exactitud de hasta los valores de  $\pi^Q(x)$  en el conjunto de  $\mathbf{P}$ -medida 0, son necesarias y suficientes para que

$$\alpha_Q(\pi^Q) = \alpha_Q(\delta_Q) = 1 - \mathbf{M} \max_j q(j/X),$$

podemos, a la par con el teorema 1, enunciar la afirmación siguiente:

**Teorema 1A.** 1) Para cualquier criterio randomizado,

$$\alpha_Q(\pi) \geq 1 - \mathbf{M} \max_j q(j/X).$$

2) Para que el criterio  $\pi^Q$  sea bayesiano es necesario y suficiente el cumplimiento de las relaciones

$$\begin{aligned} \pi^Q(x) &= e_k \quad \text{cuando } x \in \bar{\Omega}_k^Q, \\ \pi^Q(x) &\in R_{k_1, \dots, k_r} \quad \text{cuando } x \in \Gamma_{k_1, \dots, k_r} \end{aligned} \quad (7)$$

para  $\mathbf{P}$  de casi todos los valores de  $x$ .

3) Para todos  $g_j \geq 0, j = 1, \dots, r; \sum_{j=1}^r g_j = 1$  es válida la desigualdad

$$\alpha_Q(\pi^Q) = \sum_{j=1}^r q(j) \alpha_j(\pi^Q) \leq \sum_{j=1}^r q(j)(1 - g_j). \quad (8)$$

Si  $\min_j q_j > 0$  y no todos  $f_i(x)$  coinciden, o sea, si existen los valores  $k, j$  y el conjunto  $A, \mathbf{P}(A) > 0$  en el que  $f_k(x) \neq f_j(x)$ , entonces el signo en la desigualdad (8) será estricto.

**Observación 1.** De (8) se deduce que

$$\alpha_Q(\pi^Q) \leq 1 - \max_j q(j). \quad (9)$$

Aquí en el segundo miembro figura la probabilidad del error del criterio que elige  $H_k$  si  $q(k) = \max_j q(j)$  (éste es el criterio bayesiano entre todos los criterios no dependientes de la muestra  $X$ ).

**Demostración del teorema 1A.** Ya hemos demostrado las dos primeras afirmaciones. Para demostrar la última afirmación es suficiente comparar el criterio bayesiano  $\pi^Q$  con el criterio  $\pi^0(X) = g = (g_1, \dots, g_r)$  no dependiente de  $X$  y para el cual, como es evidente,  $\alpha_j(\pi^0) = 1 - g_j$ ,

$$\alpha_Q(\pi^0) = \sum_{j=1}^r q(j)(1 - g_j) \geq \alpha_Q(\pi^Q).$$

Si en (8) tiene lugar la desigualdad, entonces el criterio  $\pi^0(X) = g = \text{const}$  será bayesiano. Según la segunda afirmación del teorema, esto es posible únicamente en el caso cuando  $q(1/X) = \dots = q(r/X)$   $\mathbf{P}$  casi por doquier. Esto, a su vez, es posible únicamente cuando  $f_1(X) = \dots = f_r(X)$   $\mathbf{P}$  casi por doquier,  $q(1) = \dots = q(r)$ .  $\triangleleft$



Así pues, la introducción de los criterios randomizados no permite disminuir la probabilidad del error de  $\alpha_Q$ , pero aumenta la propia variedad de los criterios y, en particular, el número de criterios bayesianos  $\pi^Q$ . Esta circunstancia resulta, a veces, útil.

En lo sucesivo, por criterio estadístico entenderemos, por regla general, el criterio randomizado  $\pi$ .

**3. Enfoque minimax.** Mientras en el caso bayesiano hemos medido la calidad del criterio según la magnitud media  $\alpha_Q(\pi) = \sum q(j)\alpha_j(\pi)$ , ahora compararemos los valores máximos

$$\alpha(\pi) = \max_j \alpha_j(\pi) = \max_Q \alpha_Q(\pi).$$

Es evidente que esto también permite ordenar el conjunto de todos los criterios.

**Definición 7.** El criterio  $\pi = \bar{\pi}$  para el cual

$$\alpha(\bar{\pi}) = \min_{\pi} \alpha(\pi)$$

se llama criterio *minimax*.

La siguiente afirmación es el análogo completo del teorema 2.11.2.

**Teorema 2.** *Supongamos que existe el criterio bayesiano  $\bar{\pi}$  (correspondiente a cierta distribución a priori  $\bar{Q}$ ) para el cual*

$$\alpha_1(\bar{\pi}) = \dots = \alpha_r(\bar{\pi}). \quad (10)$$

*Entonces  $\bar{\pi}$  es el criterio minimax.*

**Demostración.** Designemos por  $\bar{q}(j)$  las distribuciones a priori correspondientes a  $\bar{Q}$ . Entonces para cualquier criterio  $\pi$  tenemos

$$\alpha(\pi) \geq \sum_{j=1}^r \bar{q}(j)\alpha_j(\pi) \geq \sum_{j=1}^r \bar{q}(j)\alpha_j(\bar{\pi}) = \max_j \alpha_j(\bar{\pi}) = \alpha(\bar{\pi}). \quad \triangleleft$$

La distribución  $\bar{Q} = \{\bar{q}(j)\}$  correspondiente al criterio  $\bar{\pi}$  se llama criterio *peor* (o criterio *menos favorable*, compárese con el § 2.11). Esto está relacionado con el hecho de que para  $Q = \bar{Q}$  se alcanza

$$\max_Q \alpha_Q(\pi^Q) = \max_Q \min_{\pi} \alpha_Q(\pi),$$

así que el criterio minimax (10) es el criterio bayesiano que posee la mayor probabilidad de equivocarse. La demostración de este hecho se puede hallar en los capítulos posteriores, donde también mostraremos que la peor distribución y el criterio minimax siempre existen.

Sin embargo, es preciso señalar que a distinción de los criterios bayesianos, los criterios minimax no randomizados existen no siempre, ni mucho menos. El asunto consiste en que las fronteras separadoras  $\Gamma_k$  de

los conjuntos  $\tilde{\Omega}_k^Q$  (véase (6)) pueden tener una probabilidad no nula  $P_k(X \in \Gamma_k) > 0$  y, por lo tanto, los valores de  $\alpha_k(\delta_Q)$ , al modificarse continuamente  $Q$ , pueden variar a saltos. Esto quiere decir, a su vez, que  $r - 1$  ecuaciones  $\alpha_1(\delta_Q) = \dots = \alpha_r(\delta_Q)$  para  $r - 1$  desconocidas  $q(1), \dots, q(r - 1)$   $\left( q(r) = 1 - \sum_{j=1}^{r-1} q(j) \right)$  pueden no tener solución. No obstante, en la clase de criterios bayesianos randomizados, el criterio minimax existe siempre. En calidad de ilustración examinaremos detalladamente esta cuestión (para el caso  $r = 2$ ) en el párrafo siguiente.

Así pues, hemos hallado la forma explícita de los criterios bayesianos y hemos establecido que con su ayuda se pueden construir los criterios minimax. Resulta que de manera análoga también se pueden construir los criterios más potentes en las clases  $K_{\alpha_1, \dots, \alpha_{r-1}}$  introducidas en el punto 1.

**4. Criterios más potentes.** La definición del c.m.p. no randomizado fue dada en el punto 1. Aquí será cómodo extender esta definición a la clase de criterios randomizados. Supongamos que, análogamente al punto 1,  $K_{\alpha_1, \dots, \alpha_{r-1}}$  significa la clase de criterios *randomizados* con valores registrados de las probabilidades de los errores de  $j$ -ésimo género,  $j = 1, \dots, r - 1$ :

$$K_{\alpha_1, \dots, \alpha_{r-1}} = \{ \pi: \alpha_j(\pi) = \alpha_j; j = 1, \dots, r - 1 \}.$$

**Definición 8.** El criterio  $\pi_0 \in K_{\alpha_1, \dots, \alpha_{r-1}}$  se llama *c.m.p. en  $K_{\alpha_1, \dots, \alpha_{r-1}}$*  si para cualquier  $\pi \in K_{\alpha_1, \dots, \alpha_{r-1}}$

$$\alpha_r(\pi_0) \leq \alpha_r(\pi).$$

**Teorema 3.** *Supongamos que existe una distribución  $Q = \{q(1), \dots, q(r)\}$ , tal, que*

$$\alpha_j(\pi^Q) \equiv 1 - M_j \pi_j^Q(X) = \alpha_j, \quad j = 1, \dots, r - 1 \quad (11)$$

(en realidad, aquí tenemos  $r - 1$  ecuaciones para los valores desconocidos de  $q(1), \dots, q(r - 1)$ ). *Entonces el criterio bayesiano  $\pi^Q$ , definido en (6) y (7), será el más potente en la clase  $K_{\alpha_1, \dots, \alpha_{r-1}}$ .*

**Demostración.** Según la definición del criterio bayesiano,

$$\alpha_Q(\pi^Q) \leq \alpha_Q(\pi).$$

Esto significa que para  $\pi \in K_{\alpha_1, \dots, \alpha_{r-1}}$ , tendremos

$$\sum_{j=1}^r q(j) \alpha_j(\pi^Q) \leq \sum_{j=1}^{r-1} q(j) \alpha_j + q(r) \alpha_r(\pi).$$

Pero  $\alpha_j(\pi^Q) = \alpha_j$  para  $j \leq r - 1$  y, por consiguiente,  $\alpha_r(\pi^Q) \leq \alpha_r(\pi)$ .  $\triangleleft$

Aquí, por la misma causa que al hallar los criterios minimax, las ecuaciones (11) en la clase de los criterios no randomizados  $\delta$  no siempre

son resolubles. En la clase de criterios randomizados, la situación cambia considerablemente. Esta circunstancia será ilustrada en el párrafo siguiente.

Ahora citemos el ejemplo de un problema real muy difundido, acerca de la verificación de un número finito de hipótesis simples.

**Ejemplo 1.** Supongamos que la hipótesis  $H_1$  significa que un paciente que vino para ser reconocido por el médico, está sano, mientras que  $H_k$  significa que el paciente padece de cierta enfermedad  $A_k$ ,  $k \geq 2$ . La tarea del médico consiste en aceptar una de las hipótesis  $H_j$ , basándose en las observaciones (que pueden ser escritas en forma del vector  $x_1 = (x_{11}, x_{12}, \dots, x_{1s})$  que es de por sí la muestra multidimensional  $X$  de volumen unitario). Fijaremos las enfermedades  $A_k$  para que las hipótesis  $H_k$  sean simples y asimismo determinen por completo la distribución de la muestra  $X$ . Si el médico acepta la hipótesis  $H_k$ ,  $k \geq 2$ , mientras que en realidad es cierta la hipótesis  $H_1$ , entonces cometerá un error de un tipo. Pero si, al contrario, reconoce sano ( $H_1$ ) al enfermo ( $H_k$ ), entonces cometerá un error de otro género. No es difícil comprender que los "efectos" producidos por los errores de estos dos tipos pueden ser muy diferentes.

De los resultados expuestos anteriormente deducimos que para construir la mejor regla de decisión, debemos saber las distribuciones del vector de las magnitudes observables  $(x_{11}, \dots, x_{1s})$  para individuos sanos y para individuos que padecen de la enfermedad  $A_k$  (para ello necesitamos muchos datos estadísticos de exámenes médicos). Por supuesto que una gran parte del problema aquí consiste en la propia elección de  $s$  y de las observaciones  $(x_{11}, x_{12}, \dots, x_{1s})$ . Precisamente en esto se manifiesta principalmente el arte y la experiencia de los médicos.

Si el vector  $(x_{11}, \dots, x_{1s})$  se ha elegido de manera bastante argumentada, los teoremas 1—3 nos indicarán la vía directa para algoritmizar los problemas de la diagnosis de las enfermedades.

## § 2. Verificación de las hipótesis simples

En este párrafo examinaremos un poco más detalladamente un caso particular, cuando se verifican  $r = 2$  hipótesis simples.

En los problemas de verificación de las hipótesis, estas últimas desempeñan a menudo un papel asimétrico, como ocurrió, digamos, en el ejemplo 1.1. Por eso, una de las hipótesis, por ejemplo  $H_1$ , suele llamarse *fundamental* y las demás, *alternativas*. En este caso, la probabilidad del error de primer género  $\alpha_1(\delta)$  del criterio  $\delta$  también se denomina *dimensión*, y el número  $1 - \alpha_1(\delta)$ , *nivel* del criterio. El número  $\beta(\delta) = 1 - \alpha_2(\delta)$  se llama *potencia* del criterio.

La región  $\Omega_2 \subset \mathcal{X}^n$  de aceptación de la hipótesis  $H_2$  por el criterio no randomizado  $\delta$ , en el caso de  $r = 2$  se denomina *región crítica*. La probabili-

dad  $P_2(X \in \Omega_2)$  de caer en esta región, cuando es cierta  $H_2$ , equivale a la potencia del criterio  $\beta(\delta)$ . De aquí procede la denominación de "criterio más potente" para el criterio  $\delta$  con el que  $\beta(\delta)$  alcanza su máximo para un nivel registrado del criterio  $\delta$ .

Señalemos ahora, que en el caso de  $r = 2$ , cualquier criterio, incluso el randomizado, puede caracterizarse por una función numérica. En efecto, el criterio randomizado arbitrario  $\pi(x)$  se define totalmente por el valor de su  $r$  coordenadas  $\pi_1(x), \dots, \pi_r(x)$ . Pero como  $\sum \pi_j(x) = 1$ , en caso de  $r = 2$  es suficiente designar una función, digamos,  $\pi_2(x)$ . Esta función determina la probabilidad de que se acepte la alternativa  $H_2$ . Designémosla por  $\pi(x)$  y llamémosla *función crítica* del criterio  $\pi$  que designaremos con la misma letra  $\pi$ . Es evidente que para los criterios no randomizados,  $\pi(x)$  sólo adopta los valores de 0 y 1; en el caso general  $0 \leq \pi(x) \leq 1$ .

La dimensión  $\alpha_1(\pi)$  del criterio  $\pi$  (o  $\delta$ ) y su potencia  $\beta(\pi)$  se expresan a través de  $\pi(x)$  del modo siguiente:

$$\alpha_1(\pi) = M_1 \pi(X), \quad \beta(\pi) = 1 - \alpha_2(\pi) = M_2 \pi(X).$$

Designemos por  $Z$  la relación de verosimilitud

$$Z = Z(x) = f_2(x)/f_1(x)$$

que examinaremos sólo para los valores de  $x$ , con los cuales ella está definida, o sea, para  $x$  cuando  $f_1(x) + f_2(x) > 0$ .

**Teorema 1.** 1) *Supongamos que  $c = q(1)/q(2)$ , donde  $Q = (q(1), q(2))$ , y que  $q(2) = 1 - q(1)$  es una distribución a priori dada. Entonces el criterio  $\pi_{c,p}$  con la función crítica*

$$\pi_{c,p}(x) = \begin{cases} 1, & \text{si } Z(x) > c, \\ p(x), & \text{si } Z(x) = c, \\ 0, & \text{si } Z(x) < c, \end{cases} \quad (1)$$

para cualquier función medible  $p(x)$ ,  $0 \leq p(x) \leq 1$ , es bayesiano para la distribución  $Q$ :  $\pi_{c,p} = \pi^Q$ .

Los parámetros  $\alpha_1(\pi_{c,p})$  y  $\alpha_2(\pi_{c,p})$  del criterio  $\pi_{c,p}$  satisfacen la desigualdad

$$\sum_{j=1}^2 q(j) \alpha_j(\pi_{c,p}) \leq \sum_{j=1}^2 q(j)(1 - g_j) \quad (2)$$

para todos  $g_j \geq 0$ ,  $g_1 + g_2 = 1$ .

2) Para  $\varepsilon > 0$  dado, tal que  $P_1(Z > 0) \geq \varepsilon$ , existen  $c > 0$  y  $p(x) \equiv p = \text{const}$  tales que  $\pi_{c,p} \in K_\varepsilon = \{\pi: \alpha_1(\pi) = \varepsilon\}$ , y  $\pi_{c,p}$  es el c.m.p. en  $K_\varepsilon$ . Los números  $c$  y  $p$  se definen como la solución de la ecuación

$$\alpha_1(\pi_{c,p}) \equiv M_1 \pi_{c,p}(X) \equiv P_1(Z(X) > c) + pP_1(Z(X) = c) = \varepsilon. \quad (3)$$

En este caso la potencia del criterio  $\beta(\pi_{c,p}) = 1 - \alpha_2(\pi_{c,p})$  satisface la desigualdad

$$\beta(\pi_{c,p}) \geq \varepsilon. \quad (4)$$

Si no se cumple la relación  $f_2(x) = f_1(x)$  c.d.  $[\mu]$ , entonces, las desigualdades (4) y (2) para  $0 < q_1 < 1$  son estrictas.

El criterio  $\pi_{c,p}$  minimiza la probabilidad del error de primer género  $\alpha_1(\pi)$  en la clase  $K$  de todos los criterios  $\pi$  con una probabilidad fija del error de segundo género:  $K = \{\pi: \alpha_2(\pi) = \alpha_2(\pi_{c,p})\}$ .

3) Existen  $c > 0$  y  $p(x) \equiv p = \text{const}$  tales, que el criterio  $\pi_{c,p}$  será *minimax*. Los números  $c$  y  $p$  se determinan de la ecuación  $\alpha_1(\pi_{c,p}) = \alpha_2(\pi_{c,p})$  o bien, que es lo mismo, de la ecuación

$$P_1(Z(X) > c) + P_2(Z(X) > c) + p[P_1(Z(X) = c) + P_2(Z(X) = c)] = 1. (5)$$

Es evidente que si la  $P_1$ -distribución de  $Z(X)$  es continua, o sea, si  $P_1(Z(X) = c) = 0$  para todos  $c \geq 0$ , entonces, en las dos últimas afirmaciones del teorema podemos poner  $p \equiv 1$  ó  $p \equiv 0$ .

Nótese también que

$$\begin{aligned} P_1(Z(X) = c) &= \\ &= \int_{Z(X)=c} f_1(x) \mu^n(dx) = \int_{Z(X)=c} \frac{f_2(x)}{c} \mu^n(dx) = \frac{1}{c} P_2(Z(X) = c), \end{aligned}$$

así que la continuidad en  $(0, \infty)$  de la  $P_1$ -distribución de  $Z$  conduce a la continuidad de la  $P_2$ -distribución de  $Z$ .

El criterio  $\pi_{c,p}$ , basado en la relación de verosimilitud  $Z$ , se llama *criterio de la relación de verosimilitud*.

El teorema 1 muestra que todos los *criterios óptimos son criterios de la relación de verosimilitud*.

La segunda afirmación del teorema 1 lleva el nombre de *lema de Neyman — Pearson*. Si en esta afirmación, la condición  $P_1(Z > 0) \geq \varepsilon$  no se cumple, o sea, si  $P_1(Z = 0) = 1 - \delta$ ,  $\delta < \varepsilon$ , entonces el c.m.p.  $\pi(x) = I_{\{Z(x) > 0\}}$  tendrá potencia 1 y dimensión  $\delta < \varepsilon$ . Por ejemplo, si los portadores de las distribuciones  $P_1$  y  $P_2$  son disjuntos, entonces  $Z = 0$  en el conjunto donde  $f_1(x) > 0$  y, por lo tanto,  $P_1(Z > 0) = 0$ . En este caso, las hipótesis  $H_1$  y  $H_2$  se distinguen por una observación, con probabilidades de errores iguales a cero, o sea, se distinguen de un modo determinado.

**Demostración del teorema 1.** La primera afirmación del teorema es el corolario directo del teorema 1.1A.

Para demostrar la segunda afirmación se puede hacer uso del teorema 1.3. Mostremos primeramente que la ecuación (3) es siempre resoluble respecto a  $c$  y  $p$ . Es evidente que la función  $\varphi(c) = P_1(Z > c)$  no crece en  $[0, \infty)$ . La variable aleatoria  $Z$  es propia con respecto a la distribución  $P_1$ ,

o sea,

$$\begin{aligned}\varphi(c) &= \mathbf{P}_1(Z > c) = \\ &= \int_{Z(x) > c} f_1(x) \mu^n(dx) < \frac{1}{c} \int_{Z(x) > c} f_2(x) \mu^n(dx) = \frac{1}{c} \mathbf{P}_2(Z > c) \rightarrow 0\end{aligned}$$

cuando  $c \rightarrow \infty$ . Como, según la condición,  $\varphi(0) \geq \varepsilon$ , entonces existirá  $c_\varepsilon \in (0, \infty)$  tal, que  $(\varphi(c))$  será continua a la derecha)

$$\varphi(c_\varepsilon - 0) \geq \varepsilon, \quad \varphi(c_\varepsilon) \leq \varepsilon. \quad (6)$$

Si en (3) suponemos que  $c = c_\varepsilon$ , y designamos  $\Delta_\varepsilon = \varphi(c_\varepsilon - 0) - \varphi(c_\varepsilon)$ , obtendremos

$$\alpha_1(\pi_{c_\varepsilon, p}) = \varphi(c_\varepsilon) + p\Delta_\varepsilon.$$

Es evidente que aquí, en virtud de (6), siempre se puede escoger  $p \in [0, 1]$  de modo<sup>\*)</sup> que  $\varphi(c_\varepsilon) + p\Delta_\varepsilon = \varepsilon$ .

Ahora podemos proceder igualmente que en la demostración del teorema 1.3. Pongamos  $q(1) = q_\varepsilon = c_\varepsilon / (c_\varepsilon + 1)$  y fijemos el  $p$  que hemos elegido. Entonces, el criterio  $\pi_{c_\varepsilon, p}$  será bayesiano, correspondiente a la distribución  $\mathbf{Q}_\varepsilon = (q_\varepsilon, 1 - q_\varepsilon)$  y al mismo tiempo  $\alpha_1(\pi_{c_\varepsilon, p}) = \varepsilon$ . Esto significa, en virtud del teorema 1.3, que  $\pi_{c_\varepsilon, p}$  es el c.m.p. en  $K_\varepsilon$ .

Si tomamos el criterio  $\pi(x) \equiv \varepsilon$ , obtenemos

$$\pi \in K_\varepsilon, \quad \alpha_2(\pi_{c_\varepsilon, p}) \leq \alpha_2(\pi) \equiv 1 - \varepsilon, \quad \beta(\pi_{c_\varepsilon, p}) \geq \varepsilon.$$

No es otra cosa sino la desigualdad (2) ((1.8)) para  $g_2 = \varepsilon$ . Por consiguiente, si la relación  $f_2(x) = f_1(x)$  c.d.  $[\mu]$  no se cumple, entonces estas desigualdades serán estrictas. La afirmación del teorema acerca de la minimización de  $\alpha_1(\pi)$  en el criterio  $\pi_{c, p}$  de la clase  $K = \{\pi: \alpha_2(\pi) = \alpha_2(\pi_{c, p})\}$  se deduce de los razonamientos anteriormente aducidos y de la simetría con respecto a las hipótesis  $H_1$  y  $H_2$  del planteamiento del problema en la primera afirmación del teorema.

A fin de demostrar la tercera afirmación del teorema 1 conviene valerse del teorema 1.2. Para esto sólo necesitamos comprobar si la ecuación  $\alpha_1(\pi_{c, p}) = \alpha_2(\pi_{c, p})$  es resoluble respecto a  $c$  y  $p$ . Esta ecuación se puede escribir en la forma

$$\mathbf{M}_1 \pi_{c, p}(X) = 1 - \mathbf{M}_2 \pi_{c, p}(X)$$

o bien, que es lo mismo, en la forma de (5). Su solubilidad se deter-

<sup>\*)</sup> Está claro que si  $\varphi(c)$  es continua en el punto  $c_\varepsilon$ , el problema de resolución de (3) se reduce a la determinación de la cuantila de distribución de  $Z$  de orden  $1 - \varepsilon$ .

mina al igual que la solubilidad de la ecuación (3). Sólo es necesario señalar que siempre  $P_1(Z > 0) + P_2(Z > 0) \geq 1$ , ya que  $P_2(Z > 0) = \int_{f_2(x) > 0} f_2(x) \mu^n(dx) = 1$ .  $\triangleleft$

Hemos visto una vez más que el objetivo de la introducción de los criterios bayesianos randomizados consiste en asegurar la variación "continua" de los parámetros de dichos criterios (los posibles valores de las dimensiones de los criterios  $\pi_{c,p}$  llenan todo el intervalo (0, 1)). La falta de tal variación continua de los parámetros, relacionada con el hecho de que en el conjunto de la  $P_1$ -probabilidad positiva es posible la igualdad  $f_1(x) = cf_2(x)$ , constituye el principal obstáculo al hallar los criterios de un nivel dado o los minimax en la clase de criterios no randomizados. Este cuadro también se conserva por completo en el caso de verificación de un número mayor de hipótesis.

También es importante señalar que dos tipos de criterios óptimos — los más potentes y los minimax — resultan bayesianos en unas u otras distribuciones a priori. Tampoco es difícil notar que la clase de todos los criterios más potentes coincide, desde cierto punto de vista, con la clase de todos los criterios bayesianos. Tal situación, en la que en calidad de base para la elección de los criterios óptimos puede utilizarse el enfoque bayesiano, también se conservará en mucho posteriormente.

**Ejemplo 1.** Examinemos el ejemplo 2 citado en la introducción. En este ejemplo, las hipótesis  $H_1$  y  $H_2$  tienen la forma  $H_1 = \{x_i \in F(x)\}$ ,  $H_2 = \{x_i \in F(x - a)\}$ , donde  $F(x)$  es una función dada de distribución, y  $a$ , un número dado. Supongamos que  $F(x)$  tiene densidad  $f(x)$  y que la variable aleatoria  $f(x_1 - a)/f(x_1)$  tiene una distribución continua. Entonces, según el lema de Neyman — Pearson (punto 2 del teorema 1), entre todos los criterios de nivel  $1 - \varepsilon$ , el criterio

$$\prod_{i=1}^n \frac{f(x_i - a)}{f(x_i)} \geq c_\varepsilon$$

será el más potente en el problema sujeto a examen, dedicado a la verificación de la hipótesis  $H_1$  (falta el objeto), frente a la hipótesis  $H_2$  (el objeto está presente). El número  $c_\varepsilon$  se determina de la condición

$$P_1 \left( \sum_{i=1}^n \ln \frac{f(x_i - a)}{f(x_i)} > \ln c_\varepsilon \right) = \varepsilon.$$

Si  $n$  son grandes, para el cálculo de esta probabilidad podemos, evidentemente, hacer uso del teorema central del límite.

### § 3\*. Dos enfoques asintóticos del cálculo de los criterios. Comparación numérica

**1. Observaciones preliminares.** En los §§ 1 y 2 hemos hallado la forma de los criterios óptimos para verificar las hipótesis simples. El término "cálculo de los criterios" que hemos usado en el encabezamiento significará el cálculo de los parámetros que caracterizan el criterio. En el problema del c.m.p. esto es, en caso de  $r = 2$ , la búsqueda de las magnitudes  $c_\varepsilon$  y  $p$  para  $\varepsilon > 0$  dado la determinación de la probabilidad del error de segundo género  $\alpha_2(\pi_{c_\varepsilon, p})$  o bien, que es lo mismo, de la potencia del criterio  $\beta(\pi_{c_\varepsilon, p}) = 1 - \alpha_2(\pi_{c_\varepsilon, p})$ . La cuestión también puede ser planteada de una manera algo distinta. Hemos visto que en caso de  $r = 2$  todos los criterios óptimos tienen la forma de las funciones  $\pi_{c, p}$  representadas en (2.1). Supongamos que se da el criterio  $\pi_{c, p}$ . ¿Cómo determinar para él las probabilidades de los errores  $\alpha_1(\pi_{c, p})$ ?

Esta misma pregunta también surge, por supuesto, en el caso general de  $r > 2$  para el criterio (1.7), pero en este párrafo nos limitaremos, para abreviar, al caso de dos hipótesis simples.

Más abajo se examinan los enfoques asintóticos que permiten resolver *aproximadamente* (con grandes  $n$ ) tales problemas. Esos mismos enfoques también pueden utilizarse para calcular los criterios que se examinarán en adelante.

Así pues, supongamos que se da el criterio (2.1) y que la distribución de  $Z(X)$  es, para abreviar, continua, así que podemos poner  $p \equiv 1$ . Entonces, el criterio (2.1) se volverá no randomizado (designémoslo por  $\delta_c$ ) y necesitaremos hallar sus valores:

$$\begin{aligned}\alpha_1(\delta_c) &= \mathbf{P}_1 \left( \frac{f_2(X)}{f_1(X)} \geq c \right), \\ \alpha_2(\delta_c) &= \mathbf{P}_2 \left( \frac{f_2(X)}{f_1(X)} < c \right).\end{aligned}\tag{1}$$

Como  $f_j(X) = \prod_{i=1}^n f_j(x_i)$ , el suceso que se encuentra bajo el signo de probabilidad en (1) puede ser escrito en la forma

$$\sum_{i=1}^n \ln \frac{f_2(x_i)}{f_1(x_i)} \geq \ln c,$$

donde los sumandos

$$\eta_i = \ln \frac{f_2(x_i)}{f_1(x_i)}$$



son, evidentemente, variables aleatorias independientes, igualmente distribuidas en cada uno de los casos  $X \in \mathbf{P}_j$ ,  $j = 1, 2$ .

Ahora bien, el hecho se reduce al estudio de las distribuciones de las sumas  $\sum_{i=1}^n \eta_i$  de las variables aleatorias  $\eta_i$ .

En lo sucesivo supondremos que el volumen  $n$  de la muestra  $X$  crece indefinidamente. En esto caso, por criterio entendremos, en realidad, la sucesión de los criterios definidos para cada  $n$  (hemos utilizado ese mismo acuerdo para las estimaciones en el capítulo 2).

**2. Hipótesis fijas.** En este apartado supondremos que las distribuciones  $\mathbf{P}_i$  están fijas, o sea, no dependen del volumen  $n \rightarrow \infty$  de la muestra  $X_n = [X_\infty]_n$ . Examinemos el problema de cálculo del c.m.p. de nivel fijo  $1 - \varepsilon$ . Tenemos

$$\mathbf{M}_1 \eta_i \equiv -a = \int f_1(x) \ln \frac{f_2(x)}{f_1(x)} \mu(dx) = -\varrho_1(\mathbf{P}_1, \mathbf{P}_2) < 0,$$

$$\mathbf{M}_2 \eta_i \equiv b = \int f_2(x) \ln \frac{f_2(x)}{f_1(x)} \mu(dx) = \varrho_1(\mathbf{P}_2, \mathbf{P}_1) > 0,$$

donde  $\varrho_1$  es la distancia de Kullback — Leibler (véase el § 2.21). Esto significa que, en virtud de la ley de los grandes números, la  $\mathbf{P}_1$ -distribución de  $\frac{1}{n} \sum_{i=1}^n \eta_i$  permanecerá concentrada en el entorno del punto  $-a$ , y la  $\mathbf{P}_2$ -distribución, en el entorno del punto  $b$ . Y esta “separación” de las distribuciones será la mejor desde el punto de vista del lema de Neyman — Pearson. Designemos  $\sigma_j^2 = \mathbf{D}_j \eta_i$  y supongamos que  $\sigma_j^2 < \infty$ . Entonces

$$\alpha_1(\delta_c) = \mathbf{P}_1 \left( \sum_{i=1}^n \eta_i \geq \ln c \right) = \mathbf{P}_1 \left( \frac{1}{\sigma_1 \sqrt{n}} \sum_{i=1}^n (\eta_i + a) \geq \frac{\ln c + an}{\sigma_1 \sqrt{n}} \right). \quad (2)$$

Escojamos en calidad de  $c = c(n)$  toda sucesión para la cual

$$\frac{\ln c + an}{\sigma_1 \sqrt{n}} \rightarrow \lambda_\varepsilon,$$

donde  $\lambda_\varepsilon$  es, como antes, la cuantila de la distribución normal de nivel  $1 - \varepsilon$ . Entonces, de (2) y del teorema central del límite resulta

$$\alpha_1(\delta_c) \sim 1 - \Phi \left( \frac{\ln c + an}{\sigma_1 \sqrt{n}} \right) \rightarrow \varepsilon. \quad (3)$$

**Definición 1.** El criterio  $\pi$  que satisface la relación

$$\lim_{n \rightarrow \infty} \alpha_1(\pi) = \lim_{n \rightarrow \infty} \mathbf{M}_1 \pi(X) = \varepsilon$$

se llama *criterio de nivel asintótico*  $1 - \varepsilon$  (o de dimensión asintótica  $\varepsilon$ ).

Por lo tanto, para

$$\ln c = -an + \lambda_e \sigma_1 \sqrt{n} + o(\sqrt{n}), \quad (4)$$

el criterio  $\delta_c$  tendrá el nivel asintótico  $1 - \varepsilon$ .

La relación (4) puede considerarse como la solución aproximada de la ecuación del número  $c_\varepsilon$  para el cual  $\alpha_1(\delta_{c_\varepsilon}) = \varepsilon$ .

Pongamos, para precisar,  $\ln c = -an + \lambda_e \sigma_1 \sqrt{n}$  y hallemos, para el  $c$  elegido, el comportamiento asintótico de la probabilidad del error de segundo género:

$$\begin{aligned} \alpha_2(\delta_c) &= \mathbf{P}_2 \left( \sum_{i=1}^n \eta_i < \ln c \right) = \mathbf{P}_2 \left( \sum_{i=1}^n \eta_i < -an + \lambda_e \sigma_1 \sqrt{n} \right) = \\ &= \mathbf{P}_2 \left( \frac{1}{\sigma_2 \sqrt{n}} \sum_{i=1}^n (\eta_i - b) < (a + b)\sqrt{n}/\sigma_2 + \lambda_e \sigma_1/\sigma_2 \right). \end{aligned} \quad (5)$$

Como  $-(a + b)\sqrt{n}/\sigma_2 + \lambda_e \sigma_1/\sigma_2 \rightarrow -\infty$  cuando  $n \rightarrow \infty$ , aquí la aplicación del teorema central del límite sólo nos da que  $\alpha_2(\delta_c) \rightarrow 0$ .

*El problema de cálculo del comportamiento asintóticamente exacto del segundo miembro en (5) conduce al problema de las probabilidades de grandes desviaciones para las sumas de variables aleatorias  $\eta_j$ .*

Presentemos aquí los resultados de las probabilidades de grandes desviaciones, expuestos en el § 5 del capítulo 7 [11]. Supongamos que es necesario calcular el comportamiento asintótico  $\mathbf{P} \left( \sum_{i=1}^n \xi_i > x \right)$  cuando  $n \rightarrow \infty$ ,  $x \rightarrow \infty$ , donde  $\xi_i$  son independientes y están igual distribuidas. Admitamos que la distribución  $\xi_i$  tiene una componente absolutamente continua y que

$$\psi(\lambda) = \mathbf{M}e^{\lambda \xi} < \infty$$

para ciertos  $\lambda > 0$ . Supongamos, además, que

$$\begin{aligned} \lambda_+ &= \sup \{ \lambda: \psi(\lambda) < \infty \}, \\ \Lambda(\alpha) &= -\inf_{\lambda} \{ -\alpha \lambda + \ln \psi(\lambda) \}, \end{aligned} \quad (6)$$

y que  $\lambda(\alpha)$  es el valor de  $\lambda$  con el que se alcanza este  $\inf \{ \cdot \}$ .

Entonces, es válida la afirmación siguiente. (Véanse los teoremas 9 y 10 del § 5 del capítulo 7 [11]. Las condiciones  $\mathbf{D}\xi_i = 1$  y  $\mathbf{M}\xi_i = 0$  que figuran en estos teoremas no desempeñan ningún papel).

**Teorema 1.** *Supongamos que  $\frac{x - n\mathbf{M}\xi_1}{\sqrt{n}} \rightarrow \infty$  de modo que*

$$\limsup_{n \rightarrow \infty} \frac{x}{n} < \alpha_+ = \frac{\psi'(\lambda_+)}{\psi(\lambda_+)}.$$

Entonces la ecuación

$$\alpha\psi(\lambda) = \psi'(\lambda) \quad (7)$$

para el punto  $\lambda(\alpha)$  tiene, cuando  $\alpha < \alpha_+$ , la única solución,

$$P \left( \sum_{i=1}^n \xi_i > x \right) \sim \frac{1}{\sigma(\alpha) |\lambda(\alpha)| \sqrt{2\pi n}} \exp \{ -n\Lambda(\alpha) \}, \quad (8)$$

donde

$$\alpha = \frac{x}{n}, \quad \sigma^2(\alpha) = \frac{\psi''(\lambda(\alpha))}{\psi(\lambda(\alpha))} - \alpha^2.$$

Además, son válidas las relaciones

$$\Lambda(M\xi_1) = 0, \quad \Lambda'(\alpha) = \lambda(\alpha),$$

$$\Lambda''(\alpha) = \lambda'(\alpha) = \frac{\psi(\lambda(\alpha))}{\psi''(\lambda(\alpha)) - \alpha^2 \psi(\lambda(\alpha))}.$$

Volvamos ahora al cálculo del comportamiento asintótico de la magnitud  $\alpha_2(\delta_c)$  definida en (5) e igual a

$$P_2 \left( - \sum_{i=1}^n \eta_i > an - y\sqrt{n} \right) = P_2 \left( \sum_{i=1}^n (-\eta_i + b) > (a+b)n - y\sqrt{n} \right)$$

cuando  $y = \lambda_c \sigma_1$ . Para hacer uso del teorema expuesto es necesario poner

$$\xi_i = -\eta_i = \ln \frac{f_1(x_i)}{f_2(x_i)}, \quad x = an - y\sqrt{n}.$$

Entonces, cuando  $0 \leq \lambda \leq 1$ , obtenemos

$$\begin{aligned} \psi(\lambda) &= M_2 e^{-\lambda \eta_i} = \int f_2(x) (f_1(x)/f_2(x))^\lambda \mu(dx) = \\ &= \int f_1^\lambda(x) f_2^{1-\lambda}(x) \mu(dx) \leq \left( \int f_1(x) \mu(dx) \right)^\lambda \left( \int f_2(x) \mu(dx) \right)^{1-\lambda} = 1. \end{aligned}$$

De aquí asimismo se deduce que  $\psi(\lambda)$  también será finito en cierto entorno del punto  $\lambda = 1$  si

$$\int f_1(x) (f_1(x)/f_2(x))^\gamma \mu(dx) < \infty \quad (9)$$

para cualquier  $\gamma > 0$ . Luego, la ecuación para el punto  $\lambda(\alpha)$  tendrá la forma

$$-\alpha + \frac{\psi'(\lambda)}{\psi(\lambda)} = 0,$$

o bien

$$\begin{aligned} \psi'(\lambda) &= \int f_2(x) (f_1(x)/f_2(x))^\lambda \ln \frac{f_1(x)}{f_2(x)} \mu(dx) = \\ &= \alpha \int f_2(x) (f_1(x)/f_2(x))^\lambda \mu(dx). \end{aligned} \quad (10)$$

Si  $\alpha = a = Q_1(\mathbf{P}_1, \mathbf{P}_2) = \int f_1(x) \ln \frac{f_1(x)}{f_2(x)} \mu(dx)$ , entonces (10) será satisfecha cuando  $\lambda = 1$ . Esto quiere decir que

$$\lambda(a) = 1, \quad \psi(\lambda(a)) = \psi(1) = 1.$$

De aquí se desprende que

$$\Lambda(a) = a\lambda(a) - \ln \psi(\lambda(a)) = a,$$

$$\psi''(\lambda(a)) = \psi''(1) = \int f_1(x) \left( \ln \frac{f_1(x)}{f_2(x)} \right)^2 \mu(dx),$$

$$\sigma^2(a) = \psi''(1) - a^2 = \sigma_1^2,$$

$$\Lambda'(a) = \lambda(a) = 1, \quad \Lambda''(a) = \sigma_1^{-2}.$$

Las condiciones del referido teorema se cumplirán si

1) la  $\mathbf{P}_2$ -distribución de  $\ln \frac{f_1(x_1)}{f_2(x_1)}$  tiene una componente absolutamente continua,

2)  $f_1(x)f_1(x)/f_2(x)^\gamma \mu(dx) < \infty$  para cualquier  $\gamma > 0$ .

Teniendo en cuenta que en nuestro caso las funciones  $\sigma(\alpha)$ ,  $\lambda(\alpha)$ ,  $\Lambda''(\alpha)$  son continuas en el entorno del punto  $\alpha_2 = a$  y que  $\alpha = x/n = a - y/\sqrt{n}$ ,

obtenemos  $\Lambda(\alpha) = a - \frac{y}{\sqrt{n}} + \frac{y}{2\sigma_1^2 n} + o\left(\frac{1}{n}\right)$ .

Por lo tanto, ahora podemos enunciar el siguiente corolario del teorema citado.

**Corolario 1.** *Supongamos que se cumple la condición (9) que la  $\mathbf{P}_2$ -distribución de  $\ln \frac{f_1(x_1)}{f_2(x_1)}$  tiene una componente absolutamente continua. Entonces, cuando  $n \rightarrow \infty$ ,*

$$\begin{aligned} \alpha_2(\delta_c) &= \mathbf{P}_2 \left( \sum_{i=1}^n \eta_i > an - y\sqrt{n} \right) \sim \\ &\sim \frac{1}{\sigma_1 \sqrt{2\pi n}} \exp \{ -na + y\sqrt{n} - y^2/(2\sigma_1^2) \} = \\ &= \frac{1}{\sigma_1 \sqrt{2\pi n}} \exp \{ -nQ_1(\mathbf{P}_1, \mathbf{P}_2) + \lambda_c \sigma_1 \sqrt{n} - \lambda_c^2/2 \}. \quad (11) \end{aligned}$$

Ahora bien,  $\alpha_2(\delta_c)$  decrece exponencialmente<sup>\*)</sup> cuando  $n \rightarrow \infty$ .

No es difícil ver que si tomamos un  $c$  registrado en (1), ambas probabilidades  $\alpha_1(\delta_c)$  y  $\alpha_2(\delta_c)$  decrecerán exponencialmente, al igual que el valor de  $\alpha_Q(\delta_Q)$  para cualquier  $Q$  registrado. Como

$$\mathbf{M}_1 e^{\lambda \eta} = \int f_1(x) \left( \frac{f_2(x)}{f_1(x)} \right)^\lambda \mu(dx) = \psi(1 - \lambda),$$

$$\min_{\lambda} \psi(\lambda) = \min_{\lambda} \psi(1 - \lambda),$$

entonces  $\alpha_1(\delta_c)$  y  $\alpha_2(\delta_c)$  decrecerán con igual velocidad (su dependencia de  $n$  será la misma). Esto quiere decir que el criterio minimax corresponderá a cierto  $c$  registrado, cuyo valor aproximado se determina fácilmente resolviendo la ecuación  $\alpha_1(\delta_c) = \alpha_2(\delta_c)$  y utilizando el análisis asintótico del segundo miembro (8) cuando  $\alpha = c/n$ ,  $n \rightarrow \infty$ .

La aproximación exponencial (11) actúa bastante bien con grandes  $n$  siempre que la desviación normalizada

$$\frac{x + n\mathbf{M}_2\eta}{\sigma_2\sqrt{n}} = \frac{\sqrt{n}}{\sigma_2} (\varrho_1(\mathbf{P}_1, \mathbf{P}_2) + \varrho_1(\mathbf{P}_2, \mathbf{P}_1)) - \frac{\lambda_c\sigma_1}{\sigma_2} \quad (13)$$

también sea grande (véase la enunciaci3n del teorema).

En los problemas aplicados, donde el número  $n$  está limitado por valores del orden de 100, esta condici3n se cumple rara vez y el valor de (13) a menudo resulta comparable con 1. Esto dificulta la utilizaci3n del referido enfoque del c3lculo de  $\alpha_2(\delta_c)$  y corresponde a la situaci3n en que el valor de  $\alpha_2(\delta_c)$ , junto con  $\alpha_1(\delta_c)$ , no es muy peque1o (tiene una magnitud comparable, digamos, con 0,1). Al mismo tiempo, los valores de  $n$  del orden de 100 son completamente suficientes para la aplicaci3n satisfactoria del teorema central del l3mite en la zona de "desviaciones normales".

<sup>\*)</sup> A la vez hemos obtenido la posibilidad de dar una definici3n m3s de la distancia de Kullback — Leibler:

$$\varrho_1(\mathbf{P}_1, \mathbf{P}_2) = - \lim_{n \rightarrow \infty} \frac{1}{n} \ln \alpha_2(\delta_c) = - \lim_{n \rightarrow \infty} \frac{1}{n} \inf_{\delta \in K_c} \ln \alpha_2(\delta).$$

Con arreglo a esto se puede se1alar que ese mismo orden de peque1ez  $\exp \{-n\varrho_1(\mathbf{P}_1, \mathbf{P}_2)\}$  es propio de la  $\mathbf{P}_2$ -probabilidad de que la funci3n em p3rica de distribuci3n  $F_n^*$  vaya a parar al entorno de la funci3n de distribuci3n  $F_1$  correspondiente a  $\mathbf{P}_1$ . Mejor dicho, si  $\delta = \delta(n) \rightarrow 0$  bastante lentamente, entonces

$$- \lim_{n \rightarrow \infty} \frac{1}{n} \ln \mathbf{P}_2(\sup_x |F_n^*(x) - F_1(x)| < \delta) = \varrho_1(\mathbf{P}_1, \mathbf{P}_2) \quad (12)$$

(teorema de Sanov). Por consiguiente, la distancia  $\varrho_1(\mathbf{P}_1, \mathbf{P}_2)$  tiene un sentido probabil3stico profundo. Superando ciertas dificultades, el lector puede obtener (del teorema 6, § 2, cap3tulo V en [11]) la demostraci3n de la relaci3n (12).

Ahora bien, la cuestión que nos interesa consiste en saber cuando podemos usar las aproximaciones normales

$$\alpha_1(\delta_c) = P_1 \left( \sum_{i=1}^n \eta_i \geq \ln c \right) \approx 1 - \Phi \left( \frac{\ln c - nM_1\eta_1}{\sigma_1\sqrt{n}} \right),$$

$$\alpha_2(\delta_c) = P_2 \left( \sum_{i=1}^n \eta_i < \ln c \right) \approx \Phi \left( \frac{\ln c - nM_2\eta_1}{\sigma_2\sqrt{n}} \right) \quad (14)$$

a fin de calcular ambos valores de  $\alpha_1(\delta_c)$  y  $\alpha_2(\delta_c)$ .

Para fundamentar las fórmulas (14) surge otro enfoque basado en la suposición de que las hipótesis  $H_1$  y  $H_2$  son próximas.

**3. Hipótesis próximas.** Aquí examinaremos la muestra  $X$  en el esquema de series y estimaremos que las distribuciones  $P_1$  y  $P_2$  dependen de  $n$  de modo que

$$q_1(P_1, P_2) + q_2(P_2, P_1) \rightarrow 0 \quad (15)$$

cuando  $n \rightarrow \infty$ , y la sucesión (13) converge hacia el límite positivo finito.

Para facilitar los razonamientos y hacerlos útiles en la exposición ulterior, aquí nos limitaremos al caso paramétrico cuando  $X \in P_\theta$ .

$$H_1 = \{\theta = \theta_1\}, \quad H_2 = \{\theta = \theta_2\},$$

y la familia  $\{P_\theta\}$  satisface las condiciones de regularidad (RR) (véase el § 2.24).

Hagamos primeramente algunas observaciones no formales que expliquen la esencia de la cuestión. Examinamos las hipótesis próximas, o sea, supongamos que  $\theta_2 = \theta_1 + \delta$ , donde  $\delta$  es pequeño. En este caso, el logaritmo de la relación de verosimilitud, a base del cual se construye el c.m.p., puede representarse en la forma<sup>\*)</sup>

$$\ln \frac{f_{\theta_2}(X)}{f_{\theta_1}(X)} \sim \delta L'(X, \theta_1). \quad (16)$$

La estadística  $U = L'(X, \theta_1)$ , es decir, la parte principal en (16), es llamada, a veces, *aporte eficiente*. Si la hipótesis  $H_1$  es cierta, entonces

$$M_{\theta_1} U = 0, \quad D_{\theta_1} U = nI(\theta_1).$$

Como

$L'(X, \theta_1) - L'(X, \theta_2) \sim \delta L''(X, \theta_2)$ ,  $M_{\theta_2} L''(X, \theta_2) = -nI(\theta_2)$ , entonces

$$M_{\theta_2} U \sim \delta nI(\theta_2) \sim \delta nI(\theta_1),$$

$$D_{\theta_2} U \sim nI(\theta_2) \sim nI(\theta_1).$$

<sup>\*)</sup> El signo  $\sim$ , aquí utilizado, significa la equivalencia asintótica cuando  $\delta \rightarrow 0$ .

Esto quiere decir que las distribuciones de  $U$  para las hipótesis  $H_1$  y  $H_2$  y para grandes  $n$  serán distinguibles siempre que la magnitud  $\mathbf{M}_{\theta_1} U - \mathbf{M}_{\theta_2} U \sim \delta n I(\theta_1)$  sea mucho mayor que  $\sqrt{\mathbf{D}_{\theta_1} U} \sim \sqrt{n I(\theta_1)}$  o comparable con ésta. En otros términos, debe cumplirse la igualdad  $\delta n = v\sqrt{n}$ ,  $v \neq 0$ , o bien, que es lo mismo,  $\delta = v/\sqrt{n}$ .

Así pues, pasando a una exposición más exacta, supongamos que

$$\theta_2 = \theta_1 + v/\sqrt{n}, \quad (17)$$

donde consideraremos registradas las magnitudes  $\theta_1$  y  $v$ .

Siguendo las designaciones del capítulo 2, pongamos

$$Z_i(t) = \frac{f_{\theta_1+t}(X)}{f_{\theta_1}(X)}, \quad Y_i(v) = \ln Z_i \left( \frac{v}{\sqrt{n}} \right).$$

Entonces

$$\sum_{i=1}^n \eta_i = \ln \frac{f_{\theta_1}(X)}{f_{\theta_1}(X)} = Y_1(v) = -Y_2(-v). \quad (18)$$

En virtud del teorema 2.29.3 para  $X \in \mathbf{P}_{\theta_1}$ , tenemos

$$Y_1(v) = \xi_n v - \frac{1}{2} v^2 (I(\theta_1) + \varepsilon_n), \quad (19)$$

donde  $\varepsilon_n \xrightarrow{\mathbf{P}_{\theta_1}} 0$ ,  $\xi_n I^{-1/2}(\theta_1) \in \Phi_{0,1}$ . Análogamente, para  $X \in \mathbf{P}_{\theta_2}$ ,

$$-Y_2(-v) = \xi_n v + \frac{1}{2} v^2 (I(\theta_2) + \varepsilon_n),$$

donde  $\varepsilon_n \xrightarrow{\mathbf{P}_{\theta_2}} 0$ ,  $\xi_n I^{-1/2}(\theta_2) \in \Phi_{0,1}$ .

Como  $I(\theta_2) \rightarrow I(\theta_1)$ , obtenemos que para la hipótesis  $H_j$ ,  $j = 1, 2$ ,

$$\sum_{i=1}^n \eta_i \Rightarrow \xi |v| \sqrt{I(\theta_1)} + (-1)^j \frac{v^2}{2} I(\theta_1), \quad \xi \in \Phi_{0,1}.$$

Esto significa que del teorema 2.29.3 se deduce el

**Corolario 2.** *Supongamos que se cumplen las condiciones (RR), (17). Entonces, para cualquier  $c$  registrado son válidas las fórmulas (14) o bien, más exactamente,*

$$\alpha_1(\delta_c) = \mathbf{P}_{\theta_1} \left( \sum_{i=1}^n \eta_i \geq \ln c \right) \rightarrow 1 - \Phi \left( \frac{\frac{v^2}{2} I(\theta_1) + \ln c}{|v| \sqrt{I(\theta_1)}} \right), \quad (20)$$

$$\alpha_2(\delta_c) = \mathbf{P}_{\theta_2} \left( \sum_{i=1}^n \eta_i < \ln c \right) \rightarrow \Phi \left( \frac{-\frac{v^2}{2} I(\theta_1) + \ln c}{|v| \sqrt{I(\theta_1)}} \right).$$

**Definición 2.** Los criterios  $\pi_1$  y  $\pi_2$  se llaman *equivalentes asintóticamente* si

$$\limsup_{n \rightarrow \infty} |\alpha_j(\pi_1) - \alpha_j(\pi_2)| = 0, \quad j = 1, 2.$$

El criterio  $\pi$  se llama *criterio asintóticamente más potente (c.a.m.p.)* si el mismo es asintóticamente equivalente al c.m.p.

En vista de que en las representaciones (18) y (19),  $\xi_n = L'(X, \theta_1)n^{-1/2}$ , de éstas se deduce que el criterio  $\delta$ , con la región crítica

$$\frac{vL'(X, \theta_1)}{\sqrt{nI(\theta_1)}} > vd, \quad d = \frac{v^2 I(\theta) + 2 \ln c}{2|v|\sqrt{I(\theta_1)}},$$

(aquí tiene importancia el signo de  $v$ ) tendrá los mismos valores límites  $\alpha_j(\delta)$  que el criterio  $\delta_c$  y por consiguiente, será el c.a.m.p.

Además, en virtud de los resultados del § 2.29,

$$\xi_n = L'(X, \theta_1)/\sqrt{n} = (\hat{\theta}^* - \theta_1)\sqrt{nI(\theta_1)}(1 + \varepsilon_n(X, \theta_1)),$$

$\varepsilon_n(X, \theta_1) \xrightarrow{P_{\theta_1}} 0$ . De aquí resulta que el criterio con la región crítica

$$v(\hat{\theta}^* - \theta_1)\sqrt{nI(\theta_1)} > vd, \quad (21)$$

también será el c.a.m.p.

Para obtener el c.m.p.  $\delta_c$  de nivel asintótico  $1 - \varepsilon$ , es suficiente en (20) poner  $d = \lambda_\varepsilon$ . La probabilidad del error de segundo género  $\alpha_2(\delta_c)$  convergerá hacia  $\Phi(-v\sqrt{I(\theta_1)} + \lambda_\varepsilon)$ .

Para  $c = 1$  ambos límites en (20) tendrán el mismo valor:

$$\lim_{n \rightarrow \infty} \alpha_j(\delta_c) = \Phi(-v\sqrt{I(\theta_1)}/2).$$

En este caso, el criterio  $\delta_c$  (compárese con el teorema 1.2) es natural llamarlo *asintóticamente minimax*.

**4. Comparación de los enfoques asintóticos. Ejemplo numérico.** En los apartados 2 y 3 hemos examinado dos enfoques asintóticos (cada uno de los cuales está justificado en determinadas condiciones) que permiten indicar los valores aproximados de las probabilidades de los errores de primero y segundo género del c.m.p.<sup>\*)</sup> En el caso de *hipótesis registradas*, estas fórmulas se dan en (3) y (11), y en el caso de *hipótesis próximas*, en (14) y (20). Las fórmulas (11) y (20) son una aproximación secundaria en compa-

<sup>\*)</sup> Nótese que a la par con los dos enfoques propuestos se puede examinar un espectro entero de casos intermedios, los cuales en el lenguaje paramétrico pueden representarse en la forma (compárese con (17))  $\theta_2 = \theta_1 + zn^{-\gamma}$ ,  $0 \leq \gamma \leq 1/2$ . Las hipótesis próximas de tal género representan interés al seleccionar las fórmulas aproximadas que reflejan lo más exactamente una situación concreta dada.



ración con (8) y (14), por eso es necesario, en la medida de lo posible, dar preferencia a estas últimas.

Ya hemos señalado que para pequeños valores de  $\alpha_1(\delta)$ ,  $\alpha_2(\delta)$  (digamos, del orden de 0,01 y menos) conviene más utilizar el enfoque relacionado con las hipótesis registradas. Aquí es importante tener una precisión relativa de aproximación bastante buena, la cual es asegurada por las fórmulas (8) y no es garantizada por el teorema central del límite. No obstante, si  $\alpha_1(\delta)$  y  $\alpha_2(\delta)$  son comparables con 0,1 (digamos,  $\geq 0,1$ ), se puede recomendar el segundo enfoque, considerando la segunda hipótesis dada  $H_2 = \{\theta = \theta_2\}$  como un elemento de la sucesión de las hipótesis próximas  $H_{2,n} = \{\theta = \theta_1 + v/\sqrt{n}\}$ , donde, evidentemente, es necesario, para  $\theta_1$  y  $\theta_2$  dados, poner  $v = \sqrt{n}(\theta_2 - \theta_1)$ . Como los valores  $\alpha_1(\delta)$  y  $\alpha_2(\delta)$  esperados no son muy pequeños, el valor absoluto de  $v/\sqrt{I(\theta_1)}$  no debe ser grande.

**Ejemplo 1.** Citemos ahora un ejemplo numérico que ilustra, en cierta medida, la relación existente entre los dos métodos de aproximación propuestos anteriormente.

Supongamos que  $X \in \Gamma_{\theta,1}$ , o sea, que  $x_i$  tienen una densidad

$$f_{\theta}(x) = \theta e^{-\theta x}, \quad x \geq 0,$$

y la hipótesis fundamental  $H_1$  tienen la forma  $H_1 = \{\theta = 1\}$ . En calidad de alternativas examinemos las hipótesis simples  $H_2^{(1)} = \{\theta = 0,5\}$ ,  $H_2^{(2)} = \{\theta = 0,8\}$ ,  $H_2^{(3)} = \{\theta = 0,9\}$ .

Basándose en la muestra  $X$ , la hipótesis  $H_1$  se verificará frente a una de las hipótesis  $H_2^{(j)}$ ,  $j = 1, 2, 3$ . Ahora bien, aquí  $\theta_1 = 1$ , y para  $\theta_2$  hay tres variantes:  $\theta_2 = 0,5$ ,  $\theta_2 = 0,8$  y  $\theta_2 = 0,9$ , las dos últimas de las cuales trataremos de examinarlas como correspondientes a las hipótesis "próximas" a  $H_1$ . Realicemos el cálculo de los criterios para las muestras de volúmenes  $n = 30, 100, 300, 1000$ .

En nuestro caso

$$\eta_i = \ln \frac{f_{\theta_2}(x_i)}{f_{\theta_1}(x_i)} = \ln \theta_2 - (\theta_2 - 1)x_i, \quad (22)$$

$$l'(\theta_1, x_i) = 1 - x_i, \quad (23)$$

$$\hat{\theta}^* = 1/\bar{x}.$$

De aquí resulta que el c.m.p.  $\delta_c$ , así como ambos c.a.m.p. examinados anteriormente (con regiones críticas en forma de

$$\sum l'(x_i, \theta_1) < d_1 \quad \text{y} \quad \hat{\theta}^* - \theta_1 < d_1/(nI(\theta_1)), \quad d_1 = d\sqrt{nI(\theta_1)},$$

tendrán el aspecto:  $\delta_c(X) = H_2^{(j)}$  si

$$\sum_{i=1}^n (x_i - 1) > d_1. \quad (24)$$

Si  $X \in \Gamma_{1,1}$  (hipótesis  $H_1$ ), entonces

$$\mathbf{M}_1 X_1 = 1, \quad \mathbf{D}_1 X_1 = 1 \Rightarrow J(1) = \mathbf{M}_1 [I'(X_1, 1)]^2.$$

Por lo tanto, si ponemos  $d_1 = 2\sqrt{n}$ , entonces (compárese con (14))

$$\begin{aligned} \alpha_1(\delta_c) &= \mathbf{P}_1 \left( \sum_{i=1}^n (x_i - 1) > d_1 \right) = \\ &= \mathbf{P}_1 \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n (x_i - 1) > 2 \right) \rightarrow 1 - \Phi(2) \approx 0,023 \quad (25) \end{aligned}$$

cuando  $n \rightarrow \infty$ . Como en nuestro caso  $\sum_{i=1}^n \eta_i = n \ln \theta_2 + (1 - \theta_2) \sum_{i=1}^n x_i$ , entonces  $\ln c$  en (14) (o en (20)) está ligado a  $d_1$  mediante la relación

$$\ln c = n(\ln \theta_2 + 1 - \theta_2) + (1 - \theta_2)d_1.$$

A continuación presentamos tres tablas. En todas  $d_1$  se supone elegido de modo que se cumple (25) (o sea,  $d_1 = 2\sqrt{n}$ ). En la primera tabla se comparan los valores verdaderos de  $\alpha_1(\delta_c)$  con la aproximación (25). En la segunda tabla se dan los valores verdaderos de la probabilidad del error de segundo género  $\alpha_2(\delta_c)$  y de la aproximación para  $\alpha_2(\delta_c)$ , obtenidos por las fórmulas de las grandes desviaciones (8). En la tercera tabla se comparan los valores verdaderos de  $\alpha_2(\delta_c)$  con las aproximaciones obtenidas por las fórmulas de las hipótesis próximas (14). Nótese que aquí utilizamos las aproximaciones (8) y (14) sin hacer uso de las aproximaciones secundarias (11) y (20) que contienen errores adicionales. Todos los cálculos necesarios se exponen más adelante.

Los números en las tablas 1—3 se dan con una exactitud de hasta dos cifras significativas después de la coma.

Tabla 1. Valores de  $\alpha_1(\delta_c)$ . Renglón superior: valores verdaderos; renglón inferior: valores aproximaciones (14)

$n$	30	100	300	1000
	0,031	0,028	0,026	0,024
	0,023	0,023	0,023	0,023

Tabla 2. Valores de  $\alpha_2(\delta_c)$ . Renglón superior: valores verdaderos; renglón inferior: valores aproximaciones (8) o (26) (grandes desviaciones)

$n$	30	100	300	1000
$\theta_2$	30	100	300	1000
0,5	0,028 0,033	$15 \cdot 10^{-7}$ $15 \cdot 10^{-7}$	$19 \cdot 10^{-19}$ $19 \cdot 10^{-19}$	$18 \cdot 10^{-72}$ $18 \cdot 10^{-72}$
0,8	0,71	0,35	0,028	$33 \cdot 10^{-8}$
—	—	—	0,033	$34 \cdot 10^{-8}$
0,9	0,89	0,79	0,53	0,085
—	—	—	—	0,11

La comparación de las tablas 2 y 3 muestra que de acuerdo con las observaciones hechas anteriormente, la aproximación basada en grandes desviaciones actúa mejor en la parte derecha superior de la tabla (donde  $(\theta_1 - \theta_2)\sqrt{n} = (1 - \theta_2)\sqrt{n} > 3$ ), mientras que la aproximación basada en hipótesis próximas actúa mejor en la parte izquierda inferior de la tabla (donde  $(1 - \theta_2)\sqrt{n} < 3$ ). Las rayas en las tablas están puestas allí donde la aplicación del referido enfoque no tiene sentido (en la tabla 2, por ejemplo, la aproximación (8) no se aplica en todos los casos cuando  $\alpha_2(\delta_c) > 0,1$ ). El cálculo de  $\alpha_2(\delta_c)$ , cuando este valor es, digamos, menor de  $10^{-6}$ , rara vez tiene sentido práctico. En la tabla 2 hemos calculado valores muy pequeños de  $\alpha_2(\delta_c)$ , cuando  $\theta_2 = 0,5$ ,  $n = 300, 1000$ , únicamente a fin de comparar los resultados de los cálculos.

Tabla 3. Valores de  $\alpha_2(\delta_c)$ . Renglón superior: valores verdaderos; renglón inferior: valores aproximaciones (14) (hipótesis semejantes)

$\theta_2 \backslash n$	30	100	300	1000
0,5	0,028	$15 \cdot 10^{-7}$	$19 \cdot 10^{-19}$	$18 \cdot 10^{-72}$
	0,041	$31 \cdot 10^{-6}$	—	—
	0,71	0,35	0,028	$33 \cdot 10^{-8}$
0,8	0,69	0,35	0,031	$12 \cdot 10^{-7}$
	0,89	0,79	0,53	0,085
0,9	0,89	0,79	0,52	0,086

Para acabar con los comentarios dedicados a las tablas, es preciso explicar cómo hemos calculado los valores verdaderos  $\alpha_i(\delta_c)$ ,  $i = 1, 2$  y en qué se transforman las aproximaciones (8) y (14) en nuestro caso concreto.

El valor de  $\alpha_2(\delta_c)$  es igual a

$$\alpha_2(\delta_c) = P_{\theta_2} \left( \sum_{i=1}^n (x_i - 1) < 2\sqrt{n} \right).$$

Como  $M_{\theta_2, x_i} = 1/\theta_2 D_{\theta_2, x_i} = 1/\theta_2^2$ , la aproximación normal (14) para  $\alpha_2(\delta_c)$  tiene la forma

$$\Phi \left( \frac{\theta_2}{\sqrt{n}} \left[ \left( 1 - \frac{1}{\theta_2} \right) n + 2\sqrt{n} \right] \right) = \Phi((\theta_2 - 1)\sqrt{n} + 2\theta_2).$$

Examinemos ahora la fórmula (8) en la que en nuestro caso es necesario poner  $\xi_i = x_i$ ,  $x = -n - 2\sqrt{n}$ . Aquí, la condición del teorema 1,

$$\frac{x - nM\xi_1}{\sqrt{n}} - \frac{-n - 2\sqrt{n} + n/\theta_2}{\sqrt{n}} = \sqrt{n} \left( \frac{1 - \theta_2}{\theta_2} \right) - 2 \rightarrow \infty,$$

se cumple. Seguidamente,

$$\psi(\lambda) = M_{\theta_2} e^{-\lambda x} = \theta_2 \int_0^{\infty} e^{-\lambda x - \theta_2 x^2} dx = \frac{\theta_2}{\lambda + \theta_2},$$

$$\lambda_+ = \infty, \quad \alpha_+ = \lim_{\lambda \rightarrow \infty} \frac{\psi'(\lambda)}{\psi(\lambda)} = 0,$$

$$\alpha = \frac{x}{n} = -1 - \frac{2}{\sqrt{n}}.$$

Como  $\lim_{n \rightarrow \infty} \alpha = -1 < 0$ , la condición  $\limsup_{n \rightarrow \infty} \frac{x}{n} < \alpha_+$  también se cumple.

En nuestro caso la ecuación (7) tiene la forma

$$\frac{\alpha \theta_2}{\lambda + \theta_2} = -\frac{\theta_2}{(\lambda + \theta_2)^2},$$

y su solución es  $\lambda(\alpha) = -1/\alpha - \theta_2$ . De aquí hallamos

$$\Lambda(\alpha) = -\ln(-\alpha \theta_2) - 1 - \alpha \theta_2, \quad \sigma^2(\alpha) = 1/\lambda'(\alpha) = \alpha^2.$$

Ahora bien, en virtud de (8) obtenemos

$$\alpha_2(\delta_c) = P_{\theta_2} \left( \sum_{i=1}^n \xi_i > x \right) = P_{\theta_2} \left( \sum_{i=1}^n (x_i - 1) < 2\sqrt{n} \right) - \frac{1}{(1 + \alpha \theta_2)\sqrt{2\pi n}} \exp \{ n[\ln(-\alpha \theta_2) + 1 + \alpha \theta_2] \}. \quad (26)$$

Suponiendo aquí  $\alpha = -1 - 2/\sqrt{n}$ , obtenemos las fórmulas con las que hemos calculado los valores de  $\alpha_2(\delta_c)$  en la tabla 2 (renglón inferior).

Señalemos, para comparar, que el segundo miembro de (11) en nuestro caso se transforma en la expresión

$$\frac{1}{(1 - \theta_2)\sqrt{2\pi n}} \exp \{ n[\ln \theta_2 + 1 - \theta_2] + 2(1 - \theta_2)\sqrt{n} - 2 \} \quad (27)$$

que puede ser obtenida de (26), substituyendo allí  $\alpha = -1 - 2/\sqrt{n}$  y eliminando, después del desarrollo en serie, los términos del orden de  $1/\sqrt{n}$  y superiores.

En el denominador (26), el primer factor igual a  $\sigma(\alpha)|\lambda(\alpha)| = 1 + \alpha \theta_2 = 1 - \theta_2 - 2\theta_2/\sqrt{n}$  se substituye en (27) por  $\sigma_1 = 1 - \theta_2$ . Si  $\theta_2$  es próximo a 1, el error relativo, relacionado con el sumando de corrección  $-2\theta_2/\sqrt{n}$ , puede resultar considerable. Por ejemplo, para  $\theta_2 = 0,8$ ,  $n = 100$  obtenemos  $2\theta_2/\sqrt{n} = 0,16$ ,  $\sigma_1 = 1 - \theta_2 = 0,2$ ,  $\sigma(\alpha)|\lambda(\alpha)| = 0,2 - 0,16 = 0,04$ , así que el primer factor en (27) es 5 veces (!) mayor que en (26). Este ejemplo muestra que en el caso de hipótesis semejantes, cuando el factor  $\sigma_1$  en (11) es pequeño, las aproximaciones (11) o (27) deben utilizarse con mucho cuidado.

Para calcular los valores verdaderos de  $\alpha_2(\delta_c)$  hemos usado el hecho siguiente. Sea  $\eta(t)$  el proceso de reconstrucción (véase [1]) para errar a saltos  $x_1, x_2, \dots$ , o sea,

$$\eta(t) = \min \left\{ k: \sum_{i=1}^k x_i \geq t \right\}.$$

En este caso, si  $x_i \in \Gamma_{\theta,1}$ , entonces, como hemos mostrado en el § 4 del capítulo 13

[11], el proceso  $\xi(t) = \eta(t) - 1$  es, para  $t > 0$ , el proceso de Poisson con parámetro  $\theta$ , o sea,

$$P(\eta(t) - 1 = k) = e^{-\theta t} \frac{(\theta t)^k}{k!}.$$

Ahora señalemos que  $\left\{ \sum_{i=1}^n x_i \geq t \right\} = \{\eta(t) \leq n\}$  y, por consiguiente,

$$P_{\theta} \left( \sum_{i=1}^n x_i \geq t \right) = \sum_{k=0}^{n-1} e^{-\theta t} \frac{(\theta t)^k}{k!}. \quad (28)$$

Por eso cuando  $t = n + 2\sqrt{n}$ ,

$$\alpha_1(\delta_c) = P_1 \left( \sum_{i=1}^n x_i \geq t \right) = \sum_{k=0}^{n-1} e^{-t} \frac{t^k}{k!},$$

$$\alpha_2(\delta_c) = P_{\theta_2} \left( \sum_{i=1}^n x_i < t \right) = 1 - \sum_{k=0}^{k \leq \frac{n}{2}} e^{-\theta_2 t} \frac{(\theta_2 t)^k}{k!}.$$

Precisamente estas igualdades fueron utilizadas para calcular los valores exactos de  $\alpha_i(\delta_c)$ ,  $i = 1, 2$ .

Nótese que a la par con (28) también se pueden escribir otras fórmulas para la distribución de  $\sum_{i=1}^n x_i$ , basadas en el hecho de que  $\sum_{i=1}^n x_i \in \Gamma_{\theta, n}$ .

**5. Relación entre el c.m.p. y la eficacia asintótica de la e.v.m.** Utilizando los cálculos realizados y los resultados de los §§ 1 y 2, ahora podemos demostrar el teorema 2.25.3 de la eficacia asintótica de la e.v.m.  $\theta^*$  en la clase  $\mathcal{K}^0$  de estimaciones asintóticamente centrales (la pertenencia de  $\theta^* \in \mathcal{K}^0$  ha sido establecida en el apartado (2.29.3)).

**Demostración del teorema 2.25.3.** Admitamos lo contrario, es decir, el hecho de que existe una estimación asintóticamente normal  $\theta^*$  tal que, para cualquier  $\theta_1$ ,

$$\lim_{n \rightarrow \infty} M_{\theta_1, n}(\theta^* - \theta_1)^2 = \sigma^2(\theta_1) < I^{-1}(\theta_1) = \lim_{n \rightarrow \infty} M_{\theta_1, n}(\hat{\theta}^* - \theta_1)^2.$$

Examinemos el problema de verificación de la hipótesis  $H_1 = \{X \in P_{\theta_1}\}$  frente a  $H_2 = \{X \in P_{\theta}, \theta = \theta_1 + \nu n^{-1/2}\}$  y construyamos para esto el criterio  $\delta$  que tiene la forma siguiente:

$$\delta(X) = \begin{cases} H_1 & \text{si } \theta^* \leq \theta_1 + \nu n^{-1/2}, \\ H_2 & \text{si } \theta^* > \theta_1 + \nu n^{-1/2}, \end{cases}$$

donde hemos tomado, para precisar, que  $\nu > 0$ . Entonces

$$\alpha_1(\delta) = P_{\theta_1}(\theta^* > \theta_1 + \nu n^{-1/2}) = P_{\theta_1} \left( \frac{(\theta^* - \theta_1)\sqrt{n}}{\sigma(\theta_1)} > \frac{\nu}{\sigma(\theta_1)} \right) \rightarrow 1 - \Phi \left( \frac{\nu}{\sigma(\theta_1)} \right).$$

A continuación, la pertenencia de  $\theta^* \in \mathcal{K}^0$  significa que

$$\alpha_2(\delta) = P_{\theta}(\theta^* \leq \theta_1 + \nu n^{-1/2}) = P_{\theta}(\theta^* \leq \theta) \rightarrow 1/2.$$

Examinemos ahora otro criterio  $\delta_0(X)$  con la región crítica

$$\hat{\theta}^* - \theta_1 > (\nu + \gamma)/\sqrt{n}, \quad \gamma > 0,$$

que, como hemos establecido, será el c.a.m.p. (véase (21)). En vista de que cuando es bastante pequeña  $\gamma > 0$ ,

$$(v + \gamma)\sqrt{I(\theta_1)} < v/\sigma(\theta_1),$$

para este criterio,

$$\lim_{n \rightarrow \infty} \alpha_1(\delta_0) = 1 - \Phi((v + \gamma)\sqrt{I(\theta_1)}) > 1 - \Phi(v/\sigma(\theta_1)),$$

$$\lim_{n \rightarrow \infty} \alpha_2(\delta_0) = \lim_{n \rightarrow \infty} P_\theta(\theta^* \leq \theta + \gamma/\sqrt{n}) > 1/2.$$

Esto significa que a partir de cierto  $n$ , el criterio  $\delta$  será mejor que el c.m.p. La contradicción obtenida demuestra el teorema.  $\triangleleft$

## § 4. Verificación de las hipótesis compuestas.

### Clases de criterios óptimos.

**1. Planteamiento del problema y conceptos principales.** En los §§ 1 y 2 hemos examinado los problemas menos complejos de verificación de las hipótesis cuando estas últimas son simples. Sin embargo, a menudo las hipótesis sujetas a verificación tienen una naturaleza más compleja. En el caso paramétrico, por ejemplo, la hipótesis puede tener la forma  $\{X \in P_\theta; \theta \in \Theta_1\}$ , donde  $\Theta_1$  es un subconjunto dado del conjunto  $\Theta$ . Evidentemente, tal hipótesis ya no define de manera unívoca la distribución de la muestra.

Llamaremos *compuesta* toda hipótesis  $H$  que no sea simple.

Por ejemplo, las hipótesis  $\{X \in \Phi_{0,\sigma^2}; \sigma \geq 0\}$ ,  $\{X \in \Phi_{\alpha,1}; \alpha \geq 0\}$  son compuestas.

Posteriormente en este capítulo examinaremos siempre los problemas relacionados con la verificación de *dos* hipótesis que designaremos por  $H_1$  y  $H_2$ . Además, en los párrafos inmediatos nos limitamos a estudiar el *caso paramétrico*  $X \in P_\theta$ ,  $\theta \in \Theta$ . En este caso, las hipótesis  $H_i$  se pueden escribir de la forma siguiente:

$$H_i = \{X \in P_\theta; \theta \in \Theta_i\}, \quad \Theta_i \subset \Theta, \quad \Theta_1 \cap \Theta_2 = \emptyset.$$

Como los demás valores de  $\theta$  que no pertenecen a  $\Theta_1 \cup \Theta_2$  no se examinan en general, entonces, sin limitar la generalidad, podemos considerar que  $\Theta = \Theta_1 \cup \Theta_2$ , y que  $H_2$  es una hipótesis *adicional* (o *contraria*) a  $H_1$ , así que la hipótesis  $H_2$  también puede ser escrita en forma de  $H_2 = \{H_1 \text{ no es cierta}\}$ . Al igual que en el § 2, una de las hipótesis será llamada *fundamental* (en este caso es  $H_1$ ), y las hipótesis simples  $H_\theta = \{X \in P_\theta\}$ ,  $\theta \in \Theta_2$  se llamarán *alternativas*.

La separación de una hipótesis fundamental entre todas las demás, a menudo refleja la actitud del investigador hacia el objeto de estudio. La hipótesis fundamental suele corresponder a cierta concepción, y la alternativa, a las desviaciones de ésta, cuya presencia ha de ser demostrada o

rechazada. Por regla general sólo hay una o un pequeño número de hipótesis fundamentales y una gran cantidad de hipótesis alternativas.

El procedimiento de admisión de las hipótesis se basa en el criterio estadístico. Como sólo examinamos dos hipótesis, entonces, al igual que en el § 2, todo criterio (randomizado)  $\pi$  será unívocamente definido por la función medible  $\pi(x)$ ,  $0 \leq \pi(x) \leq 1$ , la cual determina la probabilidad de aprobación  $\pi(X)$  de la hipótesis  $H_2$  para cada muestra  $X$  (la realización de la elección aleatoria con probabilidad  $\pi(X)$  debe llevarse a cabo con ayuda de un dispositivo adicional). Al igual que en el § 2, la función  $\pi(x)$  se llama *crítica*. Para el criterio no randomizado  $\delta$ , la función  $\pi(x) = \delta(x)$  sólo adopta dos valores: 0 y 1; la región  $\Omega_2$  del espacio  $\mathcal{X}^n$ , en la que  $\delta(x) = 1$  (región de admisión de  $H_2$ ), en este caso se denomina *región crítica* y a menudo se identifica con el criterio  $\delta$ .

**Definición 1.** Se llama *dimensión* o *probabilidad del error de primer género* del criterio  $\pi$  el número

$$\alpha_1(\pi) = \sup_{\theta \in \Theta_1} \mathbf{M}_\theta \pi(X).$$

Es evidente que para los criterios no randomizados,

$$\alpha_1(\delta) = \sup_{\theta \in \Theta_1} \mathbf{P}_\theta(X \in \Omega_2).$$

Esta es la máxima probabilidad (respecto a  $\theta \in \Theta_1$ ) de rechazar la hipótesis  $H_1$  cuando ella es verdaderamente cierta. Por lo general, para facilitar las búsquedas de los criterios óptimos se examinan los criterios  $\pi$  que satisfacen la condición

$$\alpha_1(\pi) = \varepsilon \quad (\text{o } \alpha_1(\pi) \leq \varepsilon).$$

Designemos por  $K_\varepsilon$  la clase de tales criterios.

Llamaremos *nivel (de significación)*<sup>\*)</sup> del criterio  $\pi$  el número  $1 - \alpha_1(\pi) = 1 - \varepsilon$ .

La utilización del criterio  $\delta \in K_\varepsilon$ , estadísticamente significa que en una larga serie de experimentos para verificar la hipótesis  $H_1$  con ayuda del criterio  $\delta \in K_\varepsilon$ , no nos equivocaremos más a menudo que en una porción de casos  $\varepsilon$ , si realmente era cierta la hipótesis  $H_1$ .

---

<sup>\*)</sup> Con frecuencia se llama nivel de significación el número  $\varepsilon$ , y no el  $1 - \varepsilon$ . Pero esto es algo perverso: pues es natural considerar que cuando más alto sea el nivel de significación, tanto más "significativo" será el criterio. Partiendo precisamente de estas consideraciones hemos definido el nivel de significación (o de confianza) para los intervalos confidenciales. Como entre los criterios estadísticos y los intervalos confidenciales existe una relación directa (véase el § 8), no sería razonable cambiar esta terminología (al pasar a los criterios) por una contraria.

La elección del nivel de significación del criterio es, en gran medida, arbitraria. En calidad de  $\varepsilon$  se elige, de ordinario, uno de los valores estándar, tales como 0,005, 0,01, 0,05, 0,1. Esta estandarización tiene la ventaja de que permite reducir el volumen de las tablas que el estadista utiliza en su trabajo. No hay ninguna otra causa especial para escoger precisamente estos valores. Eligiendo el nivel de significación del criterio  $\pi$ , es necesario prestar atención a la *potencia* del criterio

$$\beta_{\pi} = \mathbf{M}_{\theta} \pi(X), \quad \theta \in \Theta_2.$$

Si ésta resulta demasiado pequeña, conviene, tal vez, sustituir el nivel  $1 - \varepsilon$  por uno menor.

Nuestra actitud hacia la hipótesis antes de realizar el experimento es una circunstancia importante que puede influir en la elección del nivel de significación. Si creemos firmemente en la veracidad de la hipótesis (la probabilidad a priori  $Q(H_1)$  en el planteamiento bayesiano del problema es grande), se necesitarán pruebas convincentes contra ella para que renunciemos a nuestra seguridad. En estas condiciones hacen falta criterios de alto nivel, y  $\varepsilon$  se elige muy pequeño (entonces, la toma de un valor perteneciente a  $\Omega_2$  será demasiado inverosímil si es cierta  $H_1$ ).

Aquí se utiliza la misma concepción que hemos expuesto al construir los intervalos confidenciales. La misma consiste en lo siguiente: si la probabilidad  $\varepsilon$  de cierto suceso  $A$  es pequeña, consideraremos prácticamente imposible el hecho de que este suceso ocurra al realizar una sola prueba.

Entre algunos especialistas de estadística matemática también existe otro punto de vista, el cual radica en que no hay necesidad de asignar un nivel de significación fijo y que para su elección preliminar no hay una regla razonable. Ellos consideran la verificación de las hipótesis no como un procedimiento que conduce obligatoriamente a la aprobación de una de dos hipótesis, sino como cierto proceso que se desarrolla en la conciencia del investigador y que determina la actitud de éste hacia las hipótesis. Desde este punto de vista, al número de significación registrado se le puede anteponer el nivel "realmente alcanzable" que se determina del modo siguiente. Examinemos la familia de criterios no randomizados  $\delta$  de nivel  $1 - \varepsilon$  cuando  $\varepsilon$  recorre los valores del intervalo  $(0, 1)$ , y designemos por  $\Omega_{2,\varepsilon}$  la región crítica  $\delta$ , suponiendo que  $\Omega_{2,\varepsilon_2} \subset \Omega_{2,\varepsilon_1}$  cuando  $\varepsilon_2 < \varepsilon_1$ .

**Definición 2.** Llámase *nivel realmente alcanzable* de la familia de criterios  $\delta$  en la muestra  $X$ , la variable aleatoria  $1 - \varepsilon(X)$ , donde

$$\varepsilon(X) = \inf \{ \varepsilon : X \in \Omega_{2,\varepsilon} \}.$$

Cuanto mayor es  $1 - \varepsilon(X)$  tanto más fuertemente testimonia la muestra contra la hipótesis  $H_1$ .

El valor  $\varepsilon(X)$  da la posibilidad de aceptar o rechazar la hipótesis para cualquiera que sea el nivel  $1 - \varepsilon$  dado de antemano, mediante la simple comparación de  $\varepsilon(X)$  con  $\varepsilon$ .



**Ejemplo 1.** En el párrafo anterior hemos construido el c.m.p. para verificar la hipótesis  $H_1 = \{X \in \Gamma_{1,1}\}$  frente a la hipótesis  $H_2 = \{X \in \Gamma_{1/2,1}\}$ . Este criterio tiene la siguiente región crítica:

$$\Omega_2 = \left\{ x \in \mathcal{X}^n : \sum_{i=1}^n (x_i - 1) > d_1 \right\}.$$

Supongamos que para la muestra  $X$  de volumen  $n = 10$  ha resultado  $\sum_{i=1}^{10} x_i = 18$ . Como para la hipótesis  $H_1 \sum_{i=1}^n x_i \in \Gamma_{1,n}$  y  $\Gamma_{1,n}((a, b)) = H_{2n}((2a, 2b))$ , entonces  $\Gamma_{1,10}((18, \infty)) = H_2((36, \infty)) = 0,0154$  (véanse las tablas III ó [8], y el nivel que en este caso se alcanza realmente será igual a  $1 - \varepsilon(X) = 1 - 0,0154 = 0,9846$ , así que la hipótesis  $H_1$  será rechazada por el c.m.p. de nivel  $1 - \varepsilon = 0,98$  y no será rechazada por el c.m.p. de nivel  $1 - \varepsilon = 0,99$ .

**2. Criterios uniformemente más potentes.** Volvamos a examinar los criterios randomizados arbitrarios  $\pi$  que hemos acordado designarlos por la función crítica  $\pi(x)$ ,  $x \in \mathcal{X}^n$  (La función  $\pi(x)$  también se puede llamar función estadística (randomizada) de decisión).

Si existe una estadística suficiente  $S(X)$ , entonces es posible limitarse a los criterios  $\pi(X)$  que dependen de  $X$  sólo por la estadística suficiente  $S(X)$ , o sea, por los criterios representables en la forma  $\pi(X) = \varphi(S(X))$ . Pues sabemos que toda la información sobre el parámetro desconocido está concentrada en  $S$ , y la utilización de otras estadísticas (otra información sobre la muestra  $X$ ) no tiene sentido.

Como ya hemos señalado, para determinar los criterios óptimos, se reduce, de ordinario, el conjunto de criterios que se examinan, hasta la clase  $K_\varepsilon$  de los criterios de nivel registrado. Entre ellos se puede tratar de hallar un criterio tal, para el que la potencia

$$\beta_\pi(\theta) = M_\theta \pi(X)$$

en la región  $\Theta_2$  sea máxima (es decir, la probabilidad del error de segundo género  $1 - \beta_\pi(\theta)$  debe ser mínima). Con otras palabras, ha de ser máxima la probabilidad de aceptar la hipótesis  $H_2$  cuando ésta es cierta.

La función  $\beta_\pi(\theta) = M_\theta \pi(X)$  también suele llamarse *función de potencia* del criterio  $\pi$ .

**Definición 3.** El criterio  $\pi^\circ \in K_\varepsilon$  se denomina *criterio uniformemente más potente* (c.u.m.p.) en  $K_\varepsilon$ , si para cualquier  $\pi \in K_\varepsilon$

$$\beta_{\pi^\circ}(\theta) \geq \beta_\pi(\theta) \quad \text{para todos } \theta \in \Theta_2. \quad (1)$$

Claro está que c.u.m.p. existe no siempre, ni mucho menos. Si tal criterio  $\pi^\circ$  existiera, la función de potencia  $\beta_{\pi^\circ}(\theta)$  para él en el gráfico permanecería más alta que cualquier otra función  $\beta_\pi(\theta)$  en la región  $\Theta_2$  a condición

de que ambas funciones no excedan el valor  $\varepsilon$  en la región  $\Theta_1$  (pues  $\alpha_1(\pi) = \sup_{\theta \in \Theta_1} \beta_{\pi}(\theta)$ ), así que  $\beta_{\pi}(\theta)$  es la envolvente de la familia  $\{\beta_{\pi}(\theta)\}$  en la región  $\Theta_2$ .

Supongamos que  $\Theta_1 = \{\theta_1\}$ ,  $M_{\theta_1} \pi^{\circ}(X) = \varepsilon$ . Entonces el c.u.m.p.  $\pi^{\circ}$  será, evidentemente, el c.m.p. de nivel  $1 - \varepsilon$  para verificar la hipótesis  $\{\theta = \theta_1\}$  frente a la alternativa  $\{\theta = \theta_2\}$  con cualquier  $\theta_2 \in \Theta_2$ . Como conocemos la forma del c.m.p., de aquí surge el siguiente procedimiento natural de búsqueda del c.u.m.p.: lo encontraremos si resulta que en el problema antes planteado, acerca de la verificación de las hipótesis  $\{\theta = \theta_1\}$  y  $\{\theta = \theta_2\}$ , el c.m.p. no depende de  $\theta_2$ .

También es cierto lo contrario: si el c.m.p. de  $K_{\varepsilon}$  para verificar la hipótesis  $\{\theta = \theta_1\}$  frente a  $\{\theta = \theta_2\}$ ,  $\theta_2 \in \Theta_2$  depende considerablemente de  $\theta_2$ , esto significará que el c.u.m.p. para verificar  $\{\theta = \theta_1\}$  frente a  $\theta \in \Theta_2$  no existe.

Si la hipótesis  $H_2$  es simple ( $\Theta_2$  consta de un solo punto  $\theta_2$ ), el concepto de c.u.m.p. pierde parcialmente su sentido y se transforma en concepto de c.m.p. ordinario, o sea, en un criterio para el que en la clase  $K_{\varepsilon}$  se maximiza  $M_{\theta_2} \pi(X)$ .

Definamos ahora los criterios bayesianos y minimax para comprobar las hipótesis compuestas.

**3. Criterios bayesianos.** Al comprobar las hipótesis compuestas distinguiremos dos enfoques bayesianos.

a) *Enfoque bayesiano completo.* Consiste en la suposición de que las hipótesis  $H_{\theta} = \{X \in P_{\theta}\}$ ,  $\theta \in \Theta$  se escogen al azar, con una distribución a priori  $\mathbf{Q}$ . Con otras palabras, en  $\Theta = \Theta_1 \cup \Theta_2$  se registra cierta  $\sigma$ -álgebra de los subconjuntos  $\mathfrak{S}$ ,  $\Theta_1 \in \mathfrak{S}$ ,  $\Theta_2 \in \mathfrak{S}$ , y  $\theta$  se considera como variable aleatoria en el espacio muestral  $(\Theta, \mathfrak{S}, \mathbf{Q})$ .

La distribución  $\mathbf{Q}$  induce la distribución  $\mathbf{Q}_i$  en  $\Theta_i$ ,  $i = 1, 2$  y las probabilidades  $q_i = \mathbf{Q}(\theta \in \Theta_i)$ , así que  $\mathbf{Q} = q_1 \mathbf{Q}_1 + q_2 \mathbf{Q}_2$ . La hipótesis de que  $\theta \in \Theta_i$  se elige al azar, con una distribución  $\mathbf{Q}_i$ , la designaremos por  $H_{Q_i}$ .

**Definición 4.** El criterio  $\pi_{\mathbf{Q}}$  se llama *bayesiano* si es un criterio bayesiano correspondiente a la distribución a priori  $(q_1, q_2)$  para verificar dos hipótesis simples  $H_{Q_1}$  y  $H_{Q_2}$  (véase el § 1).

b) *Enfoque parcialmente bayesiano.* Aquí se supone que han sido dadas las distribuciones a priori  $\mathbf{Q}_i$  en  $\Theta_i$ , pero que faltan las probabilidades a priori  $q_1, q_2$ . En este caso se trata de la verificación de dos hipótesis simples  $H_{Q_1}$  y  $H_{Q_2}$ .

Designemos, como antes,

$$K_{\varepsilon} = \{ \pi : \sup_{\theta \in \Theta_1} M_{\theta} \pi(X) \leq \varepsilon \}$$

y pongamos

$$K_{\varepsilon}^{Q_i} = \{ \pi : M_{Q_i} \pi(X) \leq \varepsilon \},$$

donde  $M_{Q_i}$  designa la esperanza matemática incondicional de la distribución en  $\Theta_i \times \mathcal{Q}^n$ , engendrada por  $Q_i$  y  $P_\theta$ .

**Definición 5.** El criterio  $\pi_{Q_1, Q_2}$  se llama *bayesiano* en  $K_\varepsilon^{Q_i}$  si es el c.m.p. de nivel  $1 - \varepsilon$  para la verificación de dos hipótesis simples  $H_{Q_1}$  y  $H_{Q_2}$ .

Si una de las hipótesis  $H_i$  degenera en hipótesis simple ( $\Theta_1$  ó  $\Theta_2$  unipuntualmente), también degenera la distribución respectiva. En este caso acortaremos el índice en la designación  $\pi_{Q_1, Q_2}$  y escribiremos  $\pi_{Q_i}$  en vez de  $\pi_{Q_1, Q_2}$  si  $\Theta_2 = \{\theta_2\}$  unipuntualmente.

La construcción de los criterios  $\pi_{Q_1, Q_2}$  no presenta dificultades. Utilizaremos estos criterios como medio auxiliar para construir los c.u.m.p. y los minimax.

#### 4. Criterios minimax.

**Definición 6.** El criterio  $\bar{\pi}$  para verificar  $H_1 = \{\theta \in \Theta_1\}$  frente a  $H_2 = \{\theta \in \Theta_2\}$  se llama *minimax* en  $K_\varepsilon$  (en  $K_\varepsilon^{Q_i}$ ) si  $\bar{\pi} \in K_\varepsilon(\bar{\pi} \in K_\varepsilon^{Q_i})$ , y para él se maximiza

$$\inf_{\theta \in \Theta_2} M_\theta \pi(X) = \inf_{\theta \in \Theta_2} \beta_{\bar{\pi}}(\theta).$$

Sería más correcto llamar este criterio *maximín* (se maximiza el mínimo). Sin embargo, a pesar de todo utilizaremos el término único "minimax", ya que el mismo conserva su sentido aún cuando se trata no de la potencia, sino de las probabilidades de segundo género.

Los criterios bayesianos y minimax se examinan más detalladamente en el § 9. Los párrafos están dedicados a la aclaración de las condiciones en las que es posible construir los c.u.m.p.

### § 5. Criterios uniformemente más potentes

En este párrafo examinaremos dos importantes casos particulares, referentes al parámetro unidimensional  $\theta$  cuando se logra construir el c.u.m.p. También obtendremos un resultado útil en cuanto a la construcción del c.m.p.

**1. Alternativas unilaterales. Relación monótona de verosimilitud.** Supongamos que la hipótesis fundamental  $H_1$  consiste en que  $\theta \leq \theta_1$ , y la hipótesis alternativa  $H_2$ , en que  $\theta > \theta_1$ . Llamaremos *unilateral* tal hipótesis  $H_2$ , a distinción, digamos, de la hipótesis  $H_2 = \{\theta \neq \theta_1\}$  (adicional a  $H_1 = \{\theta = \theta_1\}$ ), la cual es bilateral, puesto que admite desviaciones respecto a  $\theta_1$  en ambas direcciones.

Nuestra otra suposición consiste en lo siguiente. Supongamos que se cumple la condición  $(A_0)$  y que existe una función  $T(x)$  tal, que para todos  $\theta, \theta_0, \theta > \theta_0$ , la relación de verosimilitud

$$\frac{f_\theta(x)}{f_{\theta_0}(x)} \tag{1}$$

es una función no decreciente (o no creciente) de  $T(x)$ . En este caso se dice que la familia  $\{\mathbf{P}_\theta\}$  tiene una *relación de verosimilitud monótona*.

En vista de que  $T$  es una estadística suficiente, entonces  $f(x) = \psi(T(x), \theta)h(x)$ , y la condición enunciada corresponderá a la relación  $\psi(T, \theta)/\psi(T, \theta_0)$ . Esta condición significa que para todos  $\theta > \theta_0$  y para cualquier  $d > 0$ , la desigualdad  $f_\theta(x)/f_{\theta_0}(x) \geq d$  será resoluble en la forma  $T(x) \geq c_n(\theta, \theta_0, d)$  (o bien  $T(x) \geq c_n(\theta, \theta_0, d)$ ).

Por ejemplo, las familias  $\{\Phi_{\alpha,1}\}$  y  $\{\Phi_{0,\sigma^2}\}$  tienen una relación de verosimilitud monótona, ya que

$$\frac{f_{\alpha,1}(X)}{f_{\alpha_0,1}(X)} = \exp \left\{ (\alpha - \alpha_0)n\bar{x} - \frac{n}{2}(\alpha^2 - \alpha_0^2) \right\},$$

$$\frac{f_{0,\sigma^2}(X)}{f_{0,\sigma_0^2}(X)} = \exp \left\{ -\frac{1}{2} \left( \frac{1}{\sigma^2} - \frac{1}{\sigma_0^2} \right) \sum_{i=1}^n x_i^2 \right\},$$

y las desigualdades respectivas tendrán la forma ( $\alpha > \alpha_0$ ,  $\sigma > \sigma_0$ )

$$\bar{x} \geq c_n(\alpha, \alpha_0, d) = \frac{1}{2}(\alpha + \alpha_0) + \frac{\ln d}{n(\alpha - \alpha_0)} \quad (T(X) = \bar{x}),$$

$$\sum_{i=1}^n x_i^2 \leq c_n(\sigma, \sigma_0, d) = \frac{2(\sigma\sigma_0)^2}{\sigma^2 - \sigma_0^2} \ln d \quad \left( T(X) = \sum_{i=1}^n x_i^2 \right).$$

Muchas familias paramétricas del § 2.2 también tienen una relación de verosimilitud monótona. En lo sucesivo, para precisar, consideramos que (61) es una función *no decreciente*  $T(x)$ .

**Teorema 1.** Sea  $\theta$  un parámetro unidimensional y supongamos que  $\{\mathbf{P}_\theta\}$  tiene una relación de verosimilitud monótona. Entonces

1) En  $K_\varepsilon$  existe c.u.m.p. para verificar la hipótesis  $H_1 = \{\theta \leq \theta_1\}$  frente a la alternativa  $H_2 = \{\theta > \theta_1\}$ , el cual tiene la forma siguiente:

$$\pi^\circ(X) = \begin{cases} 1, & \text{si } T(X) > c, \\ p, & \text{si } T(X) = c, \\ 0, & \text{si } T(X) < c, \end{cases} \quad (2)$$

donde  $c$  y  $p$  se deducen de la condición

$$\mathbf{M}_\theta \pi^\circ(X) = \mathbf{P}_\theta(T(X) > c) + p\mathbf{P}_\theta(T(X) = c) = \varepsilon. \quad (3)$$

2) La función de potencia  $\beta^\circ(\theta) = \mathbf{M}_\theta \pi^\circ(X)$  crece estrictamente en  $\theta$  con todos  $\theta$  para los cuales  $\beta^\circ(\theta) < 1$ .

3) Con todos  $\theta_0$  el criterio (2) es el c.u.m.p. en la clase  $K_{\beta^\circ(\theta_0)}$  para verificar la hipótesis  $H_1^\circ = \{\theta \leq \theta_0\}$  frente a  $H_2^\circ = \{\theta > \theta_0\}$ .

4) Para cualquier  $\theta < \theta_1$ , nuestro criterio minimiza  $\beta(\theta) = \mathbf{M}_\theta \pi(X)$  en la clase  $K_\varepsilon$ .

**Demostración.** Examinemos primeramente las hipótesis simples  $\{\theta = \theta_1\}$  y  $\{\theta = \theta_2\}$ ,  $\theta_2 > \theta_1$ . El c.m.p. para verificar estas hipótesis en la clase de criterios  $\pi$ , para los cuales  $M_\theta \pi(X) = \varepsilon$ , tiene, según el teorema 2.1, la forma (2), ya que la desigualdad  $Z(X) > d$  equivale a  $T(X) > c$  (en caso de la debida correspondencia entre  $c$  y  $d$ ), donde las constantes  $c$  y  $p$  se deducen de (3) (compárese con (2.3)). Como los números  $c$  y  $p$  de la ecuación de forma (3) se determinan de un modo único, entonces también obtenemos que el criterio (2) será el c.m.p. en  $K_{\beta^\circ(\theta_0)}$  para verificar la hipótesis  $\{\theta = \theta_0\}$  frente a  $\{\theta = \theta_2\}$ ,  $\theta_2 > \theta_0$ . De aquí y del teorema 2.1 (véase (2.4)) resulta que  $\beta^\circ(\theta_2) > \beta^\circ(\theta_0)$ .

Como  $\beta^\circ(\theta)$  no decrece, entonces

$$M_\theta \pi^\circ(X) \leq \varepsilon \quad \text{cuando} \quad \theta \leq \theta_1. \quad (4)$$

La clase  $K_\varepsilon$  de los criterios  $\pi$  que satisfacen (4) está presente en la clase  $\{\pi: M_\theta \pi(X) = \varepsilon\}$ . En vista de que el criterio (2) maximiza  $\beta(\theta_2)$  en esta última clase, también maximizará  $\beta(\theta_2)$  en  $K_\varepsilon$ . Queda señalar que el criterio (2) no depende de ningún modo de  $\theta_2$  y, por consiguiente, las conclusiones sacadas son válidas para cualquier  $\theta_2 > \theta_1$ . Aquí pues, han sido demostradas las primeras tres afirmaciones del teorema.

La cuarta afirmación se deduce de las tres primeras si éstas se aplican al problema de la verificación de la hipótesis  $H'_1 = \{\theta \geq \theta_1\}$  frente a  $H'_2 = \{\theta < \theta_1\}$ , para la cual el c.u.m.p. en la clase  $\{\Pi(X): M_\theta \Pi(X) \leq 1 - \varepsilon, \theta \geq \theta_1\}$  tendrá la forma  $\Pi^\circ(X) = 1 - \pi^\circ(X)$ , y la función  $1 - \beta^\circ(\theta) = M_\theta \Pi^\circ(X)$  en máxima función de potencia cuando  $\theta < \theta_1$ .  $\triangleleft$

Una importante clase de familias de distribuciones que admiten la relación de verosimilitud monótona es formada por la familia exponencial monoparamétrica (véase el § 2.15) cuando la densidad  $f_\theta(x)$  es representable en la forma

$$f_\theta(x) = h(x) \exp \{a(\theta)U(x) + V(\theta)\}. \quad (5)$$

En efecto, en este caso

$$\frac{f_\theta(x)}{f_{\theta_0}(x)} = \exp \left\{ (a(\theta) - a(\theta_0)) \sum_{i=1}^n U(x_i) + n(V(\theta) - V(\theta_0)) \right\}$$

y la relación de verosimilitud dependerá monótonamente de  $T(x) = \sum_{i=1}^n U(x_i)$  si  $a(\theta) - a(\theta_0)$  conserva el signo en todos  $\theta$ ,  $\theta_0$ ,  $\theta > \theta_0$ .

**Corolario 1.** Supongamos que  $f_\theta(x)$  tiene la forma (5), donde  $a(\theta)$  es una función monótona. Entonces existe el c.u.m.p.  $\pi^\circ$  en la clase  $K_\varepsilon$  para la verificación de la hipótesis  $H_1 = \{\theta \leq \theta_1\}$  frente a  $H_2 = \{\theta > \theta_1\}$ . Si  $a(\theta)$  crece, este criterio tiene la forma (2) y (3). Si  $a(\theta)$  decrece, las desigualdades en (2) y (3) se sustituyen por las contrarias.

Nótese que si se verifica la alternativa bilateral, por ejemplo, la hipótesis  $H_1 = \{\theta = \theta_1\}$  frente a  $H_2 = \{\theta \neq \theta_1\}$ , entonces el c.u.m.p. para la familia exponencial (5) ya no existe. En efecto, admitamos, para abreviar, que  $a(\theta)$  crece y que  $P_\theta$ -distribución de  $T(X) = \sum_{i=1}^n U(x_i)$  para todos  $\theta$  es absolutamente continua. Entonces, de acuerdo con el teorema 2.1, el c.m.p. para la verificación de  $\{\theta = \theta_1\}$  frente a  $\{\theta = \theta_2\}$  será no randomizado y tendrá la región crítica  $T(X) \geq c$  si  $\theta_2 > \theta_1$ . No obstante, si  $\theta_2 < \theta_1$ , la región crítica tendrá la forma  $T(X) < c$ . Vemos que la potencia máxima en el punto  $\theta_2$  se alcanzará con criterios muy diferentes en función del signo de diferencia de  $\theta_2 - \theta_1$ . Del teorema 1 se deduce que si tomamos cualquiera de estos criterios, por ejemplo, aquél para el cual  $\pi(X) = 1$  cuando  $T(X) \geq c$ , entonces éste será el c.u.m.p. para todos  $\theta_2 > \theta_1$  y a ciencia cierta no será tal para  $\theta_2 < \theta_1$ .

Ya hemos señalado que la situación de dos hipótesis simples en el teorema 2.1. del c.m.p. es, en cierto sentido, simétrica (el c.m.p. minimiza la probabilidad del error de segundo género  $\alpha_2(\pi)$  si ha sido registrado el valor de  $\alpha_1(\pi)$  y, al contrario, minimiza  $\alpha_1(\pi)$  si se ha registrado  $\alpha_2(\pi)$ ). En el planteamiento del problema de la verificación de las hipótesis compuestas no existe tal simetría. Con esta circunstancia está vinculado el siguiente hecho interesante. Acabamos de ver que para una familia exponencial no existe el c.u.m.p. destinado a verificar la hipótesis  $H_1 = \{\theta = \theta_1\}$  frente a  $H_2 = \{\theta \neq \theta_1\}$ . De las investigaciones realizadas es fácil comprender que no existe tampoco el c.u.m.p. para la verificación de la hipótesis  $\{\theta_1 < \theta < \theta_2\}$  frente a la alternativa  $\{\theta \notin (\theta_1, \theta_2)\}$ . No obstante, si examinamos ahora, en calidad de hipótesis fundamental  $H_1$ , la  $H_1 = \{\theta \notin (\theta_1, \theta_2)\}$ , y en calidad de alternativa, la hipótesis  $H_2 = \{\theta \in (\theta_1, \theta_2)\}$ , entonces el c.u.m.p. en  $K_\varepsilon$  ya existirá. Así pues, vamos examinar ahora la segunda posibilidad cuando se logra construir el c.u.m.p.

## 2. Hipótesis fundamental bilateral. Familia exponencial.

**Teorema 2.** *Supongamos que  $f_\theta(x)$  se define por la igualdad (5) y que se verifica la hipótesis  $H_1 = \{\theta \notin (\theta_1, \theta_2)\}$ ,  $\theta_1 < \theta_2$ , frente a la alternativa  $H_2 = \{\theta \in (\theta_1, \theta_2)\}$ . En este caso si la función  $a(\theta)$  es monótona,*

1) *en la clase  $K_\varepsilon = \{\pi: \sup_{\theta \in (\theta_1, \theta_2)} \mathbf{M}_\theta \pi(X) \leq \varepsilon\}$  existe un c.u.m.p.  $\pi^\circ$  que tiene la forma*

$$\pi^\circ(x) = \begin{cases} 1, & \text{si } c_1 < T(x) < c_2, \\ p_1, & \text{si } T(x) = c_i, \quad i = 1, 2, \\ 0, & \text{si } T(x) \notin [c_1, c_2], \end{cases} \quad (6)$$

donde  $T(x) = \sum_{i=1}^n U(x_i)$  y las constantes  $c_i, p_i$  se deducen de las condiciones

$$\mathbf{M}_{\theta_1} \pi^\circ(X) = \mathbf{M}_{\theta_2} \pi^\circ(X) = \varepsilon. \quad (7)$$

2) Este criterio maximiza la función de potencia  $\beta(\theta) = M_{\theta}\pi(X)$  a condición de (7) dentro del intervalo  $(\theta_1, \theta_2)$ , y la minimiza fuera de este intervalo (véase la fig. 4).

3) Cuando  $0 < \varepsilon < 1$ , la función  $\beta^{\circ}(\theta)$  tiene el máximo en cierto punto  $\theta_0 \in (\theta_1, \theta_2)$  y decrece estrictamente al alejarse  $\theta$  de  $\theta_0$  a la derecha o a la izquierda. Además excluimos el caso cuando la distribución de  $T(X)$  está concentrada en dos puntos, o sea, cuando existen tales  $t_1, t_2$  que

$$P_{\theta}(T(X) = t_1) + P_{\theta}(T(X) = t_2) = 1 \quad \text{para todos } \theta. \quad (8)$$

En las investigaciones que se realizan también es útil la afirmación siguiente.



Fig. 4. Forma de la función de potencia  $\beta^{\circ}(\theta) = M_{\theta}\pi^{\circ}(X)$  y  $\beta(\theta) = M_{\theta}\pi(X)$  para el criterio arbitrario  $\pi \in K$ .

**Lema 1.** Las ecuaciones (7) para  $0 < \varepsilon < 1$  son siempre resolubles con respecto a  $c_i$  y  $p_i$ ,  $i = 1, 2$ .

La demostración de este lema se dará más tarde.

**Demostración del teorema 2.** Escribamos la función de verosimilitud en la forma

$$f_{\theta}(x) = c(\theta)e^{a(\theta)T(x)}h(x), \quad (9)$$

donde, supondremos, con el fin de precisar, que  $a(\theta)$  crece estrictamente.

Examinemos el siguiente planteamiento bayesiano del problema. Admitamos que se verifica la hipótesis fundamental "mixta"  $H$ , la cual consiste en que  $\{\theta = \theta_1\}$  con probabilidad  $q$ , y  $\{\theta = \theta_2\}$  con probabilidad  $1 - q$  frente a la alternativa  $H_0 = \{\theta = \theta_0\}$ ,  $\theta_0 \in (\theta_1, \theta_2)$ . Supongamos después, que las probabilidades a priori de las hipótesis  $H$  y  $H_0$  son iguales a  $r$  y  $1 - r$ , respectivamente. Como las hipótesis  $H$  y  $H_0$  determinan por completo la distribución de la muestra, ellas pueden considerarse simples y podemos hacer uso de los resultados del § 2. En este caso el criterio bayesiano (designémoslo por  $\pi^{\circ}$ ) tendrá la forma

$$\pi^{\circ}(X) = \begin{cases} 1, & \text{si } R(X) \equiv \frac{f_{\theta_0}(X)}{qf_{\theta_1}(X) + (1-q)f_{\theta_2}(X)} > \frac{r}{1-r}, \\ p(X), & \text{si } R(X) = \frac{r}{1-r} \\ 0, & \text{si } R(X) < \frac{r}{1-r}. \end{cases} \quad (10)$$

En virtud de (9) la desigualdad  $R(X) > r/(1-r)$  es equivalente a la desigualdad

$$q \frac{c(\theta_1)}{c(\theta_0)} e^{(a(\theta_1) - a(\theta_0))T} + (1-q) \frac{c(\theta_2)}{c(\theta_0)} e^{(a(\theta_2) - a(\theta_0))T} < \frac{1-r}{r}. \quad (11)$$

Como  $a(\theta_1) - a(\theta_0) < 0$ ,  $a(\theta_2) - a(\theta_0) > 0$ , aquí el primer miembro es una función convexa de  $T$ . Esto quiere decir que (11) se puede escribir en la forma

$$c_1 < T < c_2,$$

donde  $c_i = c_i(q, r)$ ; los números  $c_1 < c_2$  recorren, al variar  $q$  y  $r$ , todos los valores posibles. La función  $p(X)$  en (10) se supone igual a  $p_1$  si  $T(X) = c_1$  y  $p_2$  si  $T(X) = c_2$ .

Según el lema 1, habrá  $c_i$ ,  $i = 1, 2$  (o bien, que es lo mismo,  $q$  y  $r$ ) y  $p_i$  tales que (7) sea cumplida. Mostremos ahora, que la función  $\pi^\circ(X)$  definida en (10) o, que es lo mismo, en 6, poseerá todas las propiedades enunciadas en el teorema 2. Lo dicho significa que ahora consideramos  $\pi^\circ$  simultáneamente como función de decisión para la verificación de  $H_1$  frente a  $H_2$ . Como el criterio  $\pi^\circ$  es bayesiano (para la verificación de  $H$  frente a  $H_0$ ), entonces, para cualquier otro criterio  $\pi$ ,

$$\begin{aligned} r[qM_{\theta_1}\pi^\circ + (1-q)M_{\theta_2}\pi^\circ] + (1-r)M_{\theta_0}(1-\pi^\circ) &\leq \\ &\leq r[qM_{\theta_1}\pi + (1-q)M_{\theta_2}\pi] + (1-r)M_{\theta_0}(1-\pi). \end{aligned} \quad (12)$$

Por consiguiente, si el criterio  $\pi$ , a la par con  $\pi^\circ$ , satisface (7), entonces

$$M_{\theta_0}\pi^\circ \geq M_{\theta_0}\pi.$$

Esto significa que en cada punto interior  $\theta_0 \in (\theta_1, \theta_2)$ , el criterio  $\pi^\circ$  maximiza la función de potencia  $\beta(\theta) = M_\theta\pi$  en la clase de criterios  $\pi$  que satisface (7). Pero las condiciones (7) destacan una clase de criterios que es más amplia que  $K_\varepsilon$ . Por lo tanto,  $\pi^\circ$  también maximizará  $\beta(\theta)$  en  $K_\varepsilon$ . En vista de que el criterio  $\pi^\circ$  no depende de  $\theta_0$ , el mismo será el c.u.m.p. en  $K_\varepsilon$ .

También cabe señalar que, en virtud del teorema 2.1,

$$\beta^\circ(\theta_0) = M_{\theta_0}\pi^\circ \geq \varepsilon$$

y aquí la igualdad sólo es posible en el caso de que

$$qf_{\theta_1}(x) + (1-q)f_{\theta_2}(x) = f_{\theta_0}(x) \quad (13)$$

$\mu^n$  casi por doquier.

De un modo absolutamente análogo podemos convencernos, con ayuda de (12), de que  $\pi^\circ$  minimizará  $M_\theta\pi$  para  $M_{\theta_1}\pi$ ,  $M_{\theta_2}\pi$  registradas (aquí utilizamos las mismas consideraciones que en la demostración de los teoremas del § 1).



Mostremos ahora, que  $\pi^\circ$  minimiza  $\beta(\theta)$  fuera de  $(\theta_1, \theta_2)$ . Sea  $\theta^\circ < \theta_1$ . Sustituyamos en las investigaciones precedentes, los tres puntos  $(\theta_1, \theta_0, \theta_2)$  por los tres puntos  $(\theta^\circ, \theta_1, \theta_2)$  y notemos que para el nuevo problema, el criterio  $\pi^\circ$  volverá a ser bayesiano (pues su forma no depende de la elección de los puntos  $\theta_i$ ,  $i = 0, 1, 2$ ) en la clase de criterios  $\pi$  para los cuales  $\mathbf{M}_{\theta^\circ} \pi = \beta^\circ(\theta^\circ)$ ,  $\mathbf{M}_{\theta_2} \pi = \varepsilon$ . Pero, según la observación hecha anteriormente,  $\pi^\circ$  minimizará  $\mathbf{M}_{\theta^\circ} \pi$  para  $\mathbf{M}_{\theta_1} \pi$  y  $\mathbf{M}_{\theta_2} \pi$  registradas. Las primeras dos afirmaciones del teorema quedan demostradas.

Demostremos la tercera afirmación. Nótese previamente que, utilizando la sustitución de las variables de integración, podemos escribir

$$\mathbf{P}_\theta(T \in A) = c(\theta) \int_{\{x: T(x) \in A\}} e^{a(\theta)T(x)} h(x) \mu^n(dx) = c(\theta) \int_{t \in A} e^{a(\theta)t} \nu(dt),$$

donde la medida  $\nu$  se define por la relación

$$\nu(A) = \int_{\{x: T(x) \in A\}} h(x) \mu^n(dx).$$

Esto quiere decir que la distribución  $T$  respecto a la medida  $\nu$  tiene densidad (véase también el lema 2.15.1)  $g_\theta(t) = c(\theta)e^{a(\theta)t}$  y, por consiguiente, también pertenece a la familia exponencial. Luego, en virtud de la monotonía de  $a(\theta)$  se puede introducir un nuevo parámetro  $\beta = a(\theta)$  sin modificar absolutamente el problema y sus condiciones. Por consiguiente, podemos considerar, sin limitar la generalidad, que  $a(\theta) = \theta$ . En este caso las funciones  $c(\theta) = \left[ \int e^{\theta t} \nu(dt) \right]^{-1}$  y  $\beta^\circ(\theta) = \mathbf{M}_\theta \pi^\circ(X)$  serán, evidentemente, continuas.

Admitamos ahora que la afirmación del teorema acerca del carácter del comportamiento de  $\beta^\circ(\theta)$  no es cierta. Entonces habrá tres puntos  $\theta' < \theta'' < \theta'''$  para los cuales

$$\beta^\circ(\theta') = \beta^\circ(\theta'') = \beta^\circ(\theta''') = \alpha \in (0, 1). \quad (14)$$

Hemos visto que  $\pi^\circ$  maximiza  $\beta(\theta'')$  para las condiciones  $\beta(\theta') = \beta(\theta'') = \alpha$ , con la particularidad de que si no se cumple la condición que tiene la forma (13), entonces  $\beta^\circ(\theta'') > \alpha$ . Pero en nuestro caso la igualdad (13) quiere decir que

$$q \frac{f_{\theta'}}{f_{\theta''}} + (1 - q) \frac{f_{\theta''}}{f_{\theta''}} \equiv q \frac{c(\theta')}{c(\theta'')} e^{(\theta' - \theta'')T} + (1 - q) \frac{c(\theta''')}{c(\theta'')} e^{(\theta'' - \theta'')T} = 1$$

$\nu$ -casi por doquier. En virtud de la convexidad del primer miembro respecto a  $T$ , esta igualdad es posible no más que para dos valores de  $T$ . Por lo tanto, si (8) se excluye, entonces  $\beta^\circ(\theta'') > \beta^\circ(\theta') = \alpha$ , y (14) es imposible.  $\triangleleft$

La demostración del lema 1 se llevará a efecto suponiendo simplemente que la distribución  $T(X)$  es continua, o sea, que  $\mathbf{P}_\theta(T = c) = 0$  para todos  $\theta$  y  $c$ . Esto nos liberará de complicaciones poco importantes. En este caso, en virtud de las observaciones hechas al final de la demostración del teorema 2, podemos escribir

$$\mathbf{M}_\theta \pi^\circ(X) = \mathbf{P}_\theta(T \in (c_1, c_2)) = \int_{c_1}^{c_2} g_\theta(t) \nu(dt) = c(\theta) \int_{c_1}^{c_2} e^{\theta t} \nu(dt).$$

Esta será una función continua de  $\theta$ ,  $c_1$ ,  $c_2$ .

Designemos por  $c_+$  el valor de  $c$  para el cual  $\mathbf{P}_{\theta_1}(T \leq c_+) = 1 - \varepsilon$ . Entonces, en  $(-\infty, c_+)$  estará definida una función  $d(c)$  tal, que

$$\mathbf{P}_{\theta_1}(T \in (c, d(c))) = \int_c^{d(c)} g_{\theta_1}(t) \nu(dt) = \varepsilon.$$

Naturalmente que  $d(c)$  es una función continua creciente.

Demostremos la afirmación requerida si nos convencemos de que la función

$$\psi(c) = \mathbf{P}_{\theta_2}(T \in (c, d(c))) = \int_c^{d(c)} g_{\theta_2}(t) \nu(dt)$$

crece continuamente,  $\psi(-\infty) < \varepsilon$ ,  $\psi(c_+) > \varepsilon$ . En este caso existirá un valor de  $c_0$  tal, que  $\psi(c_0) = \varepsilon$  y, por lo tanto  $\mathbf{P}_{\theta_2}(c_0, d(c_0)) = \varepsilon$ ,  $i = 1, 2$ .

La continuidad de  $\psi(c)$  es evidente. Demostremos ahora la monotonía. Escribamos  $\psi(c)$  en la forma

$$\psi(c) = \int_c^{d(c)} g_{\theta_2}(t) r(t) \nu(dt), \quad (15)$$

donde  $r(t)$  es la densidad de la  $\mathbf{P}_{\theta_2}$ -distribución de  $T$  respecto a la  $\mathbf{P}_{\theta_1}$ -distribución:

$$r(t) = \frac{c(\theta_2)}{c(\theta_1)} e^{(\theta_2 - \theta_1)t}.$$

Supongamos, para precisar, que  $\Delta$  es tal, que  $c + \Delta < d(c)$ . En este caso, como

$$\int_c^{c+\Delta} g_{\theta_2}(t) \nu(dt) = \int_c^{d(c+\Delta)} g_{\theta_2}(t) \nu(dt), \quad (16)$$

entonces

$$\begin{aligned} \psi(c + \Delta) - \psi(c) &= \int_c^{d(c+\Delta)} g_{\theta_2}(t) r(t) \nu(dt) - \int_c^{c+\Delta} g_{\theta_2}(t) r(t) \nu(dt) \geq \\ &\geq [r(d(c)) - r(c + \Delta)] \lambda \geq 0, \end{aligned}$$

donde  $\lambda$  es el valor general de la integral (16).

Ahora nos convencemos de que  $\psi(-\infty) < \varepsilon$ . Designemos por  $t_0$  la solución de la ecuación  $r(t) = 1$ . Si  $d(-\infty) \leq t_0$ , entonces  $r(t) < 1$  en el intervalo  $(-\infty, d(-\infty))$ , y la igualdad requerida es, en virtud de (15), evidente. Si  $d(-\infty) > t_0$ , entonces, de un modo análogo obtenemos

$$\begin{aligned} \psi(-\infty) &= 1 - P_{\theta_1}(T \in (d(-\infty), \infty)) < \\ &< 1 - P_{\theta_1}(T \in (d(-\infty), \infty)) = P_{\theta_1}(T \in (-\infty, d(-\infty))) = \varepsilon. \end{aligned}$$

Exactamente igual se establece que  $\psi(c_+) > \varepsilon$ .  $\triangleleft$

**Observación 1.** Le dejamos al lector que el mismo se convenza de que para  $\theta_1 < \theta_2$  la afirmación del teorema 2 y todas las investigaciones realizadas serán válidas si sustituimos el intervalo  $(\theta_1, \theta_2)$  por el segmento  $[\theta_1, \theta_2]$ , o sea, si verificamos la hipótesis  $H_1 = \{\theta \in [\theta_1, \theta_2]\}$  frente a  $H_2 = \{\theta \in \mathbb{R} \setminus [\theta_1, \theta_2]\}$ .

**Observación 2.** La exigencia del carácter exponencial de la familia  $\{P_{\theta}\}$ , como se deduce de la demostración del teorema, puede ser debilitada hasta la condición de convexidad de la relación

$$q \frac{f_{\theta_1}(X)}{f_{\theta_0}(X)} + (1 - q) \frac{f_{\theta_2}(X)}{f_{\theta_0}(X)}$$

con respecto a cierta estadística  $T$  (compárese con (10) y (11)).

**Observación 3.** Prestemos atención una vez más en que si la hipótesis principal fuera  $H_2 = \{\theta \in (\theta_1, \theta_2)\}$ , y la alternativa  $H_1 = \{\theta \in (\theta_1, \theta_2)\}$ , entonces, el c.u.m.p. no existiría, ya que en este caso, los criterios "unilaterales" que tienen la forma  $T > c$  o  $T < c$  para las alternativas  $\theta > \theta_2$  y  $\theta < \theta_1$ , respectivamente, resultarían más potentes que el criterio de forma  $T \notin (c_1, c_2)$ . Por ejemplo, para las alternativas  $\theta > \theta_2$  existirá el c.u.m.p. de forma  $T > c$ , y la condición  $\pi \in K_\varepsilon$  conducirá a la única limitación  $M_{\theta_1} \pi \leq \varepsilon$  (véanse las observaciones al final del punto 2).

No obstante, resultará que si la clase  $K_\varepsilon$  se reduce un poco adicionalmente, procediendo de un modo natural (véanse los §§ 6 y 7), entonces el c.u.m.p. también existirá en este problema.

**3. Otro enfoque de los problemas sujetos a examen.** La esencia matemática de la afirmación principal del teorema 2, así como de los teoremas en los §4 1 y 2, es muy simple y merece la pena que hablemos de ella especialmente. Por ejemplo, en el teorema 2, la misma consiste en el siguiente problema variacional. En la clase de funciones  $\pi$  que satisfacen las condiciones

$$\int \pi(x) f_{\theta_i}(x) \mu^n(dx) = \varepsilon, \quad i = 1, 2$$

buscamos el elemento  $\pi^\circ$  para el cual se maximiza

$$\int \pi(x) f_{\theta_0}(x) \mu^n(dx).$$

La siguiente afirmación suele llamarse *generalización del lema fundamental de Neumann — Pearson*.

**Lema 2.** Sean  $f_1, \dots, f_{m+1}$  las funciones reales definidas en  $\mathcal{X}^n$  e integrables respecto a la medida  $\mu^n$ . Supongamos que las funciones críticas  $\pi$  son tales, que

$$\int \pi(x) f_i(x) \mu^n(dx) = \varepsilon_i, \quad i = 1, \dots, m. \quad (17)$$

Entonces, el elemento  $\pi^\circ$ , en el que  $\int \pi(x) f_{m+1}(x) \mu^n(dx)$  alcanza el máximo, tiene la forma

$$\pi^\circ(x) = \begin{cases} 1, & \text{si } f_{m+1}(x) > \sum_{i=1}^m k_i f_i(x), \\ 0, & \text{si } f_{m+1}(x) < \sum_{i=1}^m k_i f_i(x), \end{cases}$$

donde  $k_1, \dots, k_m$  se determinan de las condiciones (17).

**Demostración.** Designemos  $F_i(\pi) = \int \pi(x) f_i(x) \mu^n(dx)$ ,  $i = 1, \dots, m+1$ . El elemento  $\pi$  que satisface las condiciones  $F_i(\pi) = \varepsilon_i$ ,  $i = 1, \dots, m$ , maximiza  $F_{m+1}(\pi)$  si y sólo si maximiza  $F_{m+1}(\pi) - \sum_{i=1}^m k_i F_i(\pi)$  para cualesquiera  $k_1, \dots, k_m$  (pues el valor de la suma aquí está registrado). Por consiguiente, es suficiente que  $\pi$  maximice

$$\int \left( f_{m+1}(x) - \sum_{i=1}^m k_i f_i(x) \right) \pi(x) \mu^n(dx).$$

Pero esta expresión se vuelve máxima si se supone que  $\pi(x) = 1$  allí donde  $f_{m+1}(x) - \sum_{i=1}^m k_i f_i(x) > 0$ , y  $\pi(x) = 0$  allí donde esta expresión es negativa. Las constantes  $k_i$ , de las cuales depende este  $\pi$ , así como los valores "libres" de  $\pi$  en el conjunto  $\left\{ f_{m+1}(x) = \sum_{i=1}^m k_i f_i(x) \right\}$ , deben escogerse de modo que se cumpla (17).  $\triangleleft$

**4. Enfoque bayesiano y distribuciones a priori menos favorables al construir el c.m.p. y el c.u.m.p.** El lema 2 aclara la esencia matemática de las construcciones que hemos realizado en este párrafo. En el apartado presente también se tratará de la esencia de estas investigaciones, pero desde un punto de vista algo diferente. El hecho consiste en que al demostrar el teorema 2 hemos utilizado, implícitamente, el enfoque relacionado con la construcción de los criterios minimax a base de los criterios bayesianos (compárese con el teorema 1.2). Este enfoque se examina más detalladamente en la exposición sucesiva. Aquí obtendremos una afirmación general, útil para construir el c.u.m.p. en el caso general, y explicaremos su relación con el enfoque minimax.

Supongamos que se verifica la hipótesis fundamental  $H_1 = \{\theta \in \Theta_1\}$  frente a la alternativa simple  $H_2 = \{\theta = \theta_2\}$ ,  $\theta_2 \notin \Theta_1$ . En calidad de  $H_2$  aquí también se puede tomar la alternativa arbitraria  $\{X \in G\}$ , donde  $G$  tiene una densidad  $g$  respecto a  $\mu$  y no está de ningún modo relacionada con la familia  $\{P_\theta\}$ . El problema consiste en determinar el c.m.p.  $\pi$  de nivel  $1 - \varepsilon$  para verificar  $H_1$  frente a  $H_2$ . Con otras palabras, es necesario hallar la función  $\pi$  de  $K_\varepsilon$ ,

$$K_\varepsilon = \{ \pi : \sup_{\theta \in \Theta_1} \mathbf{M}_\theta \pi(X) \leq \varepsilon \} \quad (18)$$

que minimiza  $\beta(\theta_2) = \mathbf{M}_{\theta_2} \pi(X)$ . En las investigaciones precedentes hemos observado varias veces cierta dualidad en el planteamiento del problema: la maximización de la potencia, al ser registrada la probabilidad del error de primer género, equivale a la minimización de este último al ser registrada la potencia. Pero con tal inversión llegamos, en nuestra tarea, a la cuestión de minimización (18), que es precisamente el problema de construcción del criterio minimax (este problema se examina más detalladamente en el § 9). Ello explica, en cierta medida, la semejanza de la afirmación (que se demostrará más abajo) con el teorema 1.2.

Así pues, examinemos el planteamiento parcialmente bayesiano del problema, en virtud del cual el parámetro  $\theta$  en el conjunto  $\Theta_1$  se elige al azar, con una distribución  $\mathbf{Q}_1$ . En este caso, la hipótesis compuesta  $H_1$  se sustituye por la hipótesis simple  $H_{\mathbf{Q}_1}$ , según la cual la densidad de  $X$  se define como el valor promediado respecto a la medida  $\mathbf{Q}_1$ :

$$f_{\mathbf{Q}_1}(x) = \int_{\Theta_1} f_\theta(x) \mathbf{Q}_1(d\theta).$$

Para verificar  $H_{\mathbf{Q}_1}$  frente a  $H_2$  en la clase  $K_\varepsilon^{\mathbf{Q}_1} = \{ \pi : \mathbf{M}_{\mathbf{Q}_1} \pi(X) \leq \varepsilon \}$  de los criterios de nivel  $1 - \varepsilon$  existe el c.m.p.  $\pi_{\mathbf{Q}_1}$ , que tiene la forma ( $\pi_{\mathbf{Q}_1}$  es el criterio  $\pi_{\mathbf{Q}_1, \mathbf{Q}_2}$  en las designaciones del § 4, donde  $\mathbf{Q}_2$  es la distribución degenerada en el punto  $\theta_2$ ):

$$\pi_{\mathbf{Q}_1}(x) = \begin{cases} 1, & \text{si } g(x) > c f_{\mathbf{Q}_1}(x), \\ 0, & \text{si } g(x) < c f_{\mathbf{Q}_1}(x) \end{cases} \quad (19)$$

(aquí  $g(x) = f_{\theta_2}(x)$  en el caso paramétrico).

**Teorema 3.** *Supongamos que existe tal distribución  $\mathbf{Q}_1$ , concentrada en el subconjunto  $\Theta_1^? \subset \Theta_1$  ( $\mathbf{Q}_1(\Theta_1^?) = 1$ ), para la cual*

$$1) \quad \pi_{\mathbf{Q}_1} \in K_\varepsilon^{\mathbf{Q}_1} \quad (20)$$

$$2) \quad \mathbf{M}_\theta \pi_{\mathbf{Q}_1}(X) = \text{const} = \sup_{\theta \in \Theta_1} \mathbf{M}_\theta \pi_{\mathbf{Q}_1}(X) \quad (21)$$

para todos  $\theta \in \Theta_1^?$ .

Entonces el criterio  $\pi_{\mathbf{Q}_1} \in K_\varepsilon$  es precisamente el c.m.p. para la verificación de  $H_1$  frente a  $H_2$ .

**Demostración.** Comprobemos primeramente la pertenencia de  $\pi_{Q_1} \in K_\varepsilon$ . En virtud de las condiciones del teorema,

$$\sup_{\theta \in \Theta_1} \mathbf{M}_\theta \pi_{Q_1}(X) = \int_{\Theta_1} \mathbf{M}_\theta \pi_{Q_1}(X) \mathbf{Q}_1(d\theta) = \mathbf{M}_{Q_1} \pi_{Q_1}(X) \leq \varepsilon. \quad (22)$$

Sea ahora  $\pi$  cualquier otro criterio de  $K_\varepsilon$ , o sea, el criterio de nivel  $1 - \varepsilon$  para verificar  $H_1$  frente a  $H_2$ . Entonces

$$\mathbf{M}_{Q_1} \pi(X) = \int \pi(x) f_{Q_1}(x) \mu^n(dx) = \int_{\Theta_1} \mathbf{M}_\theta \pi(X) \mathbf{Q}_1(d\theta) \leq \varepsilon$$

y, por lo tanto,  $\pi \in K_\varepsilon^{Q_1}$ . Pero entonces, en virtud de la definición de  $\pi_{Q_1}$ ,

$$\mathbf{M}_{\theta_2} \pi_{Q_1}(X) \geq \mathbf{M}_{\theta_2} \pi(X),$$

que es lo que se necesitaba demostrar.  $\triangleleft$

La distribución  $\mathbf{Q}_1$  que figura en el teorema se llama *distribución menos favorable*. Esto está relacionado con la circunstancia siguiente. La magnitud  $\beta_{Q_1}(\theta_2) = \mathbf{M}_{\theta_2} \pi_{Q_1}(X)$  es el mayor valor de potencia que puede ser alcanzado en  $K_\varepsilon^{Q_1}$  con la distribución "a priori"  $\mathbf{Q}_1$  en  $\Theta_1$ . Si tomamos ahora cualquier otra distribución  $\mathbf{Q}'$  en  $\Theta_1$ , obtenemos

$$\beta_{Q'}(\theta_2) \geq \beta_{Q_1}(\theta_2), \quad \beta_{Q_1}(\theta_2) = \inf_{Q'} \beta_{Q'}(\theta_2)$$

(esto es precisamente el sentido del término "la peor distribución"). En efecto, en virtud de (22)  $\pi_{Q_1}$  pertenece a  $K_\varepsilon$  y, por lo tanto, a  $K_\varepsilon^{Q'}$ . Esto quiere decir que su potencia  $\beta_{Q_1}(\theta_2) = \mathbf{M}_{\theta_2} \pi_{Q_1}(X)$  no superará la potencia del c.m.p. en  $K_\varepsilon^{Q'}$  que, por definición, es igual a  $\beta_{Q'}(\theta_2)$ .

Ahora, con ayuda del teorema 3 podríamos demostrar los teoremas 1 y 2. El conjunto  $\Theta_1^*$ , en el que está concentrada la distribución menos favorable, en los teoremas 1 y 2 consta de un solo  $\{\theta_1\}$  y de los puntos  $\{\theta_1, \theta_2\}$ , respectivamente. Las condiciones (20) y (21) se transforman, respectivamente, en condiciones (3) y (7).

Análogamente ha de utilizarse el teorema 3 para construir el c.u.m.p. en otros casos: si el criterio construido  $\pi_{Q_1}$  no depende de  $\theta_2 \in \Theta_2$ , entonces él será el c.u.m.p. para verificar  $H_1 = \{\theta \in \Theta_1\}$  frente a  $H_2 = \{\theta \in \Theta_2\}$  en la clase  $K_\varepsilon$ .

La distribución menos favorable  $\mathbf{Q}_1$ , que satisface las condiciones del teorema 3, existe para suposiciones muy amplias que suelen cumplirse en los problemas reales. Es suficiente exigir la compacticidad de  $\Theta_1$  y la continuidad de  $f_\theta(x)$  respecto a  $\theta$  para  $x$  c.d. (véase [57] y los capítulos posteriores).

La investigación ulterior de las relaciones entre los enfoques bayesiano y minimax véase en el § 9.

### § 6\*. Criterios no desplazados

En este párrafo y en el siguiente utilizaremos los principios de no desplazamiento y de invariación para la reducción natural de la clase de criterios que se examinan. El objetivo de tal reducción consiste en determinar los criterios óptimos.

**1. Definiciones y c.u.m.p. no desplazados.** Al igual que en el párrafo anterior, examinaremos la verificación de la hipótesis compuesta  $H_1 = \{\theta \in \Theta_1\}$  frente a  $H_2 = \{\theta \in \Theta_2\}$ , basándonos en la muestra  $X \in \mathbf{P}_\theta$ ,  $\theta \in \Theta = \Theta_1 \cup \Theta_2$ . Examinemos primero los criterios  $\pi$  de la clase  $K_\varepsilon = \{\pi: \sup_{\theta \in \Theta_1} M_\theta \pi \leq \varepsilon\}$ .

Si, por ejemplo,  $\Theta_1$  comprende un solo punto  $\theta_1$ ,  $M_{\theta_1} \pi = \varepsilon$ , entonces  $\varepsilon$  es la probabilidad de que se rechace  $H_1$  cuando  $H_1$  es cierta. La exigencia natural respecto al criterio  $\pi$  consiste en que la *probabilidad de rechazar  $H_1$ , cuando  $H_1$  no es cierta, ha de ser mayor que  $\varepsilon$* . Si no es así, entonces habrá alternativas con las que la aceptación de  $H_1$  será más probable que en los casos cuando  $H_1$  es cierta. Tal situación es indeseable. Llegamos a la necesidad de destacar la siguiente clase importante de criterios.

**Definición 1.** El criterio  $\pi$  se llama *no desplazado* si para él

$$\inf_{\theta \in \Theta_2} M_\theta \pi(x) \geq \sup_{\theta \in \Theta_1} M_\theta \pi(X). \quad (1)$$

Ahora bien, el criterio  $\pi \in K_\varepsilon$  (para el cual  $\sup_{\theta \in \Theta_1} M_\theta \pi = \varepsilon$ ) no estaría desplazado si  $\beta_\pi(\theta) \geq \varepsilon$  cuando  $\theta \in \Theta_2$ . La clase de criterios no desplazados de nivel  $1 - \varepsilon$  se designa por  $\tilde{K}_\varepsilon$ .

El criterio unilateral  $\pi$  con región crítica  $T > c$  (o  $T < c$ ) para familias exponenciales, mencionado en el párrafo anterior, no puede permanecer sin desplazamiento al verificar  $H_1 = \{X \in \mathbf{P}_{\theta_1}\}$  frente a  $H_2 = \{X \in \mathbf{P}_{\theta_1}, \theta \neq \theta_1\}$ , ya que aquí  $\Theta_2 = \{\theta: \theta \neq \theta_1\}$ ,  $M_\theta \pi < \varepsilon$  para  $\theta < \Theta_1$  si  $M_{\theta_1} \pi = \varepsilon$  (véase el teorema 5.1).

Al contrario, los c.u.m.p., si existen, con la necesidad pueden no estar desplazados, ya que para ellos la potencia  $\beta(\theta)$ , cuando  $\theta \in \Theta_2$ , no puede ser menor que la potencia del criterio  $\pi(X) = \varepsilon$ .

El principio de no desplazamiento<sup>9)</sup> reviste interés especial, puesto que permite reducir naturalmente la clase de criterios. Esto nos permite construir los c.u.m.p. en las clases  $K_\varepsilon$  cuando los c.u.m.p. no existen en la clase  $K_\varepsilon$ .

<sup>9)</sup> El término "no desplazamiento" también se utilizó con arreglo a las estimaciones. Desde cierto punto de vista la propiedad de no desplazamiento de la estimación es análoga a la propiedad de no desplazamiento del criterio: si la estimación  $\theta^*$  no está desplazada, entonces  $M_{\theta^*} \theta^* \neq \theta_0$  y habrá otros valores del parámetro  $\theta \neq \theta_0$  con los cuales el valor medio  $M_\theta \theta^*$  será igual a  $\theta_0$ .

Como veremos, esto se refiere, en particular, al problema de verificación de la hipótesis  $H_1 = \{\theta \in [\theta_1, \theta_2]\}$ ,  $\theta_1 \leq \theta_2$ , frente a la alternativa bilateral  $H_2 = \{\theta \notin [\theta_1, \theta_2]\}$  (compárese con el apartado 2 del § 5).

La determinación de los criterios no desplazados y uniformemente más potentes puede ser bastante reducida al uso de los procedimientos ya empleados, cuya esencia se expone en el lema 5.2. En este caso puede ser útil la siguiente afirmación.

Supongamos que existe una frontera común no vacía  $\Gamma$  de los conjuntos  $\Theta_1$  y  $\Theta_2$  de  $R^k$ :

$$\Gamma = \partial\Theta_1 \cap \partial\Theta_2$$

( $\partial\Theta_1$  designa la frontera de  $\Theta_1$ ), o sea, un conjunto de puntos límites para  $\Theta_1$  y  $\Theta_2$ . Supongamos además, que para todos  $\pi \in K_\varepsilon$ ,

$$\beta_\pi(\theta) = M_\theta \pi(X) = \varepsilon \quad \text{cuando} \quad \text{todos } \theta \in \Gamma. \quad (2)$$

Es evidente que esta propiedad siempre se cumplirá si  $\beta_\pi(\theta)$  depende continuamente de  $\theta$  para cualquier criterio  $\pi$  de  $K_\varepsilon$ .

Como

$$\beta_\pi(\theta) = \int \pi(x) f_\theta(x) \mu^n(dx), \quad 0 \leq \pi(x) \leq 1,$$

entonces la continuidad de  $\beta_\pi(\theta)$  tendrá lugar si la función  $f_\theta(x)$  es continua respecto a  $\theta$  para c.t.  $\mu^n$  de  $x$ . Esto se deduce del corolario 1 del Suplemento VI.

Designemos por  $\bar{K}_\varepsilon$  la clase de todos los criterios  $\pi$  que satisfacen (2).

**Lema 1.** *Supongamos que  $K_\varepsilon \subset \bar{K}_\varepsilon$  (o sea, que se cumple (2)). En este caso, si  $\tilde{\pi}$  es el c.u.m.p. en  $\bar{K}_\varepsilon \cap K_\varepsilon$ , entonces  $\tilde{\pi}$  es el c.u.m.p. en  $K_\varepsilon$ .*

**Demostración.** Es suficiente convencerse que  $\tilde{\pi} \in K_\varepsilon$  y que  $K_\varepsilon \subset \bar{K}_\varepsilon \cap K_\varepsilon$ . La segunda de estas relaciones se desprende de la suposición de que  $K_\varepsilon \subset \bar{K}_\varepsilon$ . La primera se deduce del hecho de que el criterio  $\pi \equiv \varepsilon$  pertenece a  $\bar{K}_\varepsilon \cap K_\varepsilon$  y, por lo tanto,  $\inf_{\theta \in \Theta_2} M_\theta \tilde{\pi}(X) \geq \inf_{\theta \in \Theta_2} M_\theta \pi = \varepsilon$ .  $\triangleleft$

Ahora bien, el lema 1 permite reducir la búsqueda de los criterios no desplazados y uniformemente más potentes, a la búsqueda de los c.u.m.p. ordinarios, pero al disponer de condiciones de frontera (2). Si el número de puntos de la frontera  $\Gamma$  es finito, resultaremos en las condiciones del lema 5.2 donde nos quedará verificar que la función crítica óptima obtenida  $\pi$  no dependa del valor  $\theta \in \Theta_2$  para el cual se ha maximizado la funcional  $M_\theta \pi(X)$ . Esto significará precisamente la potencia uniforme máxima.

Nótese ahora la siguiente circunstancia, relacionada con la degeneración de las condiciones (2), la cual se aclara fácilmente para el caso unidimensional. Si  $\Theta_1 = [\theta_1, \theta_2]$  y  $\Theta_2$  es el complemento de  $\Theta_1$ , entonces las condi-



ciones (2) serán dos ecuaciones  $M_{\theta_i} \pi(X) = \varepsilon$ ,  $i = 1, 2$ . Sin embargo, en el caso límite  $\theta_1 = \theta_2$ , estas ecuaciones se transforman en una sola. Pero en virtud del no desplazamiento del criterio  $\pi$ , su potencia  $\beta_{\pi}(\theta)$  debe alcanzar su mínimo en el punto  $\theta_1$  (véase (1)). Por consiguiente, si  $\beta_{\pi}(\theta)$  es derivable, entonces, el papel de las ecuaciones (2) en el caso de las  $\theta_1 = \theta_2$  lo desempeñarán las igualdades

$$\beta_{\pi}(\theta_1) = \varepsilon, \quad \beta'_{\pi}(\theta_1) = 0. \quad (3)$$

Las condiciones de derivabilidad de  $\int f_{\theta}(x) \mu(dx)$  y, por consiguiente, también de  $\beta_{\pi}(\theta) = M_{\theta} \pi(X)$ , son aclaradas en el Suplemento VI. Si se cumplen estas condiciones, entonces

$$\begin{aligned} \beta'_{\pi}(\theta) &= \int \pi(x) f'_{\theta}(x) \mu^n(dx) = \\ &= \int \pi(x) L'(x, \theta) f_{\theta}(x) \mu^n(dx) = M_{\theta} \pi(X) L'(X, \theta). \end{aligned}$$

Esto significa que las condiciones (3) pueden escribirse de nuevo en términos integrales:

$$M_{\theta_1} \pi(X) = \varepsilon, \quad M_{\theta_1} \pi(X) L'(x, \theta_1) = 0. \quad (4)$$

Por ejemplo, para la familia exponencial (5.9),

$$L'(x, \theta) = c'(\theta)/c(\theta) + a'(\theta)T(x).$$

Como  $M_{\theta} L'(x, \theta) = 0$ , entonces  $c'(\theta)/c(\theta) = -a'(\theta)M_{\theta} T(X)$ ,

$$M_{\theta} \pi(X) L'(X, \theta) = -a'(\theta)M_{\theta} T(X) \cdot M_{\theta} \pi(X) + a'(\theta)M_{\theta} \pi(X) T(X),$$

y las ecuaciones (4) adoptan la forma

$$M_{\theta_1}(\pi(X) - \varepsilon) = 0, \quad M_{\theta_1}(\pi(X) - \varepsilon)T(X) = 0.$$

En calidad de ejemplo ilustremos un caso para cuyo examen, de hecho, ya todo está preparado.

## 2. Alternativas bilaterales. Familia exponencial.

**Teorema 1.** *Supongamos que  $f_{\theta}(x)$  se define por la igualdad (5.9), y que se verifica la hipótesis  $H_1 = \{\theta \in [\theta_1, \theta_2]\}$ ,  $\theta_1 \leq \theta_2$ , frente a la alternativa  $H_2 = \{\theta \notin [\theta_1, \theta_2]\}$ . Entonces, si la función  $a(\theta)$  es monótona,*

1) *en la clase  $K_{\varepsilon}$  de criterios no desplazados de nivel  $1 - \varepsilon$  existe un c.u.m.p.  $\tilde{\pi}$  que tiene la forma siguiente:*

$$\tilde{\pi}(x) = \begin{cases} 0 & \text{si } c_1 < T(x) < c_2, \\ p_i & \text{si } T(x) = c_i, \quad i = 1, 2, \\ 1 & \text{si } T(x) \notin [c_1, c_2], \end{cases} \quad (5)$$

donde  $T(x) = \sum_{i=1}^n U(x_i)$ , y las constantes  $c_i, p_i, i = 1, 2$  se deducen de las

condiciones

$$\mathbf{M}_{\theta_i} \tilde{\pi}(X) = \varepsilon, \quad i = 1, 2, \quad (6)$$

si  $\theta_1 < \theta_2$ , y de las condiciones

$$\mathbf{M}_{\theta} \tilde{\pi}(X) = \varepsilon, \quad \mathbf{M}_{\theta}(\tilde{\pi}(X) - \varepsilon)T(X) = 0, \quad (7)$$

si  $\theta_1 = \theta_2$ .

2) El criterio  $\tilde{\pi}$  minimiza la función  $\beta_x(\theta) = \mathbf{M}_{\theta} \pi(X)$  en las condiciones (6) dentro del segmento  $[\theta_1, \theta_2]$ , y la maximiza fuera de  $[\theta_1, \theta_2]$  en las condiciones (6) ó (7) (esto último sucede cuando  $\theta_1 = \theta_2$ ).

3) cuando  $0 < \varepsilon < 1$  y  $\theta_1 < \theta_2$ , la función  $\tilde{\beta}(\theta) = \mathbf{M}_{\theta} \tilde{\pi}(X)$  alcanza su valor mínimo en cierto punto  $\theta_0 \in (\theta_1, \theta_2)$  y crece estrictamente al alejarse  $\theta$  de  $\theta_0$  a la derecha o a la izquierda. Además, excluimos el caso (5.8).

No es difícil ver que la enunciación de este teorema casi repite la afirmación del teorema 5.2. La única diferencia consistente en que las propias afirmaciones tienen, a veces, carácter "contrario" y no se excluye la igualdad  $\theta_1 = \theta_2$ .

**Demostración.** En el caso de  $\theta_1 < \theta_2$ , ésta es absolutamente análoga a la demostración del teorema 5.2. En la nota 1 adjunta a este teorema hemos dicho que para  $\theta_1 < \theta_2$  todos los razonamientos del referido teorema conservan su validez en el caso cuando se verifica la hipótesis  $\{\theta \notin [\theta_1, \theta_2]\}$  frente a  $\{\theta \in [\theta_1, \theta_2]\}$ , o sea, a los símbolos de este párrafo: la hipótesis  $H_2$  frente a la  $H_1$ . Pongamos  $\tilde{\pi}(x) = 1 - \pi^\circ(x)$ , donde  $\pi^\circ$  es la función definida en (5.6) para las condiciones  $\mathbf{M}_{\theta_i} \pi^\circ(X) = 1 - \varepsilon$ ,  $i = 1, 2$ , en vez de (5.7). Entonces, las afirmaciones 2) y 3) serán, evidentemente, los corolarios directos de las respectivas afirmaciones del teorema 5.2.

La primera afirmación del teorema resulta de la segunda, ya que la clase de criterios  $\pi$  que satisfacen (6) es más amplia que  $K_\varepsilon$  y, por consiguiente,  $\tilde{\pi}$  maximizará  $\mathbf{M}_{\theta} \pi(x)$  en la clase  $K_\varepsilon$  en cualquier punto  $\theta$  fuera de  $[\theta_1, \theta_2]$ . Esto significa que  $\pi$  es el criterio no desplazado uniformemente más potente.

Nos queda examinar el caso  $\theta_1 = \theta_2$ . Aquí es más simple, por lo visto, hacer uso del lema 5.2. Tomemos cualquier  $\theta \neq \theta_1$  y examinemos el problema de maximización de  $\mathbf{M}_{\theta} \pi(X)$  para las condiciones

$$\mathbf{M}_{\theta} \pi(x) = \varepsilon, \quad \mathbf{M}_{\theta} \pi(X)T(X) = \varepsilon \mathbf{M}_{\theta} T(X).$$

Es evidente que nos encontraremos en condiciones del lema 5.2 si ponemos  $m = 2$ ,  $f_1 = f_{\theta_1}$ ,  $f_2 = T f_{\theta_1}$ ,  $f_3 = f_{\theta}$ ,  $\varepsilon_1 = \varepsilon$ ,  $\varepsilon_2 = \varepsilon \mathbf{M}_{\theta} T(X)$ . Según este lema, el máximo  $\mathbf{M}_{\theta} \pi$  se alcanzará en la función

$$\tilde{\pi}(x) = \begin{cases} 1, & \text{si } f_{\theta}(x) > k_1 f_{\theta_1}(x) + k_2 T(x) f_{\theta_1}(x), \\ 0, & \text{si } f_{\theta}(x) < k_1 f_{\theta_1}(x) + k_2 T(x) f_{\theta_1}(x). \end{cases}$$

Examinemos la última desigualdad, que puede ser escrita en la forma

$$\frac{c(\theta)}{c(\theta_1)} e^{(a(\theta) - a(\theta_1))T(x)} < k_1 + k_2 T(x).$$

Está claro que para todos  $c_1 < c_2$  siempre se puede escoger  $k_1, k_2$  de modo que esta desigualdad equivalga a

$$c_1 < T < c_2.$$

Esto demuestra que el criterio de forma (5) maximiza  $\mathbf{M}_\theta \pi(X)$  en las condiciones (7) siempre que  $c_i$  y  $p_i$ ,  $i = 1, 2$  puedan escogerse en (5) de modo que se satisfaga (7) (u (8)). Este criterio será, evidentemente, el criterio no desplazado uniformemente más potente, ya que la clase de criterios  $\pi$  que satisfacen (8) es más amplia que  $K_\varepsilon$  y, por lo tanto,  $\bar{\pi}$  también maximizará  $\mathbf{M}_\theta \pi(X)$  en  $K_\varepsilon$ . Así pues, para demostrar el teorema queda demostrar que es válido el

**Lema 2.** *La ecuación (7) cuando  $0 < \varepsilon < 1$  es resoluble respecto a  $c_i$  y  $p_i$ ,  $i = 1, 2$ .*

La demostración de este lema, al igual que la del lema 5.1, será expuesta suponiendo simplemente que la  $\mathbf{P}_{\theta_1}$ -distribución de  $T(X)$  es continua, es decir,  $\mathbf{P}_{\theta_1}(T(X) = c) = 0$  para todos  $c$ .

Recordemos que la densidad de la distribución  $T$  respecto a cierta medida  $\nu$  puede considerarse igual a (véase el § 5)  $g_\theta(t) = c(\theta)e^{\theta t}$ . Entonces, las ecuaciones (7) y (8) serán equivalentes a las relaciones

$$\mathbf{M}_{\theta_1}(1 - \pi(X)) = c(\theta_1) \int_{c_1}^{c_2} e^{\theta_1 t} \nu(dt) = 1 - \varepsilon, \quad (9)$$

$$\mathbf{M}_{\theta_1}(1 - \pi(X))T(X) = c(\theta_1) \int_{c_1}^{c_2} te^{\theta_1 t} \nu(dt) = (1 - \varepsilon)c(\theta_1) \int_{c_1}^{c_2} te^{\theta_1 t} \nu(dt).$$

Designando  $r(t) = t$ ,  $m = \mathbf{M}_{\theta_1} T(X) = c(\theta_1) \int_{c_1}^{c_2} te^{\theta_1 t} \nu(dt)$ , podemos escribir las ecuaciones (9) en la forma

$$c(\theta_1) \int_{c_1}^{c_2} e^{\theta_1 t} \nu(dt) = 1 - \varepsilon,$$

$$c(\theta_1) \int_{c_1}^{c_2} r(t) e^{\theta_1 t} \nu(dt) = (1 - \varepsilon)m. \quad (10)$$

Hemos llegado al problema que coincide con el problema examinado en el lema 5.1. La única diferencia consistente en que la distribución con densidad  $r(t)g_\theta(t)$  puede ser generalizada (o sea, también puede adoptar valores negativos). En estas nuevas condiciones conviene poner  $t_0 = m$ . En lo demás, los razonamientos del lema 5.1 no cambian.  $\triangleleft$

§ 7\*. Criterios invariantes.

En este párrafo examinaremos otra manera de reducir la clase de todos los criterios, basada, esta vez, en las consideraciones de invariación.

Supongamos que  $X \in \{P_\theta\}$  y que  $\{P_\theta\}$  es una familia invariante. Recordemos las designaciones necesarias y los conceptos respectivos (véase el § 2.19). Supongamos asimismo, que se ha dado un grupo  $G$  de transformaciones medibles  $g$  del espacio  $\mathcal{D}^n$  en sí. La familia  $\{P_\theta\}$  será invariante respecto a  $G$ , si para cada  $g \in G$  y cada  $\theta \in \Theta$  hay un elemento  $\theta_g \in \Theta$  tal, que

$$P_{\theta_g}(X \in A) = P_\theta(gX \in A)$$

para cualquier  $A \in \mathfrak{B}_{\mathcal{D}^n}$ .

Las transformaciones  $\bar{g}$  del espacio  $\Theta$ , definidas por la igualdad  $\bar{g}\theta = \theta_g$ , forman, al cumplirse las condiciones  $A_0$ , el grupo  $\bar{G}$  (véase el § 2.19).

**Definición 1.** Diremos que el problema de verificación de la hipótesis  $H_1 = \{\theta \in \Theta_1\}$  frente a  $H_2 = \{\theta \in \Theta_2\}$ ,  $\Theta_1 \cup \Theta_2 = \Theta$  es invariante siempre que se cumplan las dos condiciones siguientes:

- 1) La familia  $\{P_\theta\}$  es invariante respecto a  $G$ .
- 2) Los conjuntos  $\Theta_1$  y  $\Theta_2$  son invariantes respecto a  $\bar{g} \in \bar{G}$ , o sea,  $\bar{g}\Theta_i = \Theta_i$ ,  $i = 1, 2$ .

Si el problema de verificación de las hipótesis es invariante, es natural que para su solución se haga uso del criterio invariante.

**Definición 2.** El criterio  $\pi$  se llama *invariante* cuando  $\pi(x)$  es estadística invariante respecto a  $g^*$ ):

$$\pi(gx) = \pi(x) \quad \text{para todos } x \in \mathcal{D}^n. \quad g \in G.$$

Si  $\pi$  es un criterio no randomizado y  $\Omega_j$  es la región de aceptación de la hipótesis  $H_j$ , entonces, la invariación de  $\pi$  significará que  $g\Omega_j = \Omega_j$ ,  $j = 1, 2$ .

La utilización natural de los criterios invariantes se puede comprender, por lo visto, con más facilidad, a base de ejemplos. La investigación general, relacionada con la interpretación de  $g$  como la sustitución de las coordenadas y la insensibilidad de las estadísticas respectivas a esta sustitución, está contenida en el § 2.19.

**Ejemplo 1.** Los ejemplos más simples se refieren al caso cuando el grupo  $\bar{G}$  es trivial, o sea, cuando  $\bar{g}$  para todo  $g$  es la transformación idéntica  $\bar{e}$  del espacio  $\Theta$ .

Supongamos que  $X \in \Phi_{0, \sigma^2}$ ; se verifica la hipótesis  $H_1 = \{\sigma_1 \leq \sigma \leq \leq \sigma_2\}$  frente a la alternativa adicional  $H_2$ . En este caso

$$f_{\sigma^2}(X) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 \right\}.$$

<sup>\*)</sup> Véase la nota en la pág. 195.

Es evidente que la familia  $\Phi_{0,\sigma^2}$  es invariante respecto al grupo  $G$  de transformaciones ortogonales  $g$  (revoluciones) del espacio  $\mathcal{R}^n$ , con la particularidad de que  $\bar{g} = \bar{e}$  para cualquier  $g$ . Por eso es natural examinar los criterios  $\pi$  que dependen exclusivamente de la estadística  $T(X) = \sum_{i=1}^n x_i^2$ . En vista de que  $\sigma^{-2}T(X) \in \Gamma_{1/2, n/2} = H_n$ , entonces  $T(X) \in \Gamma_{\alpha, n/2}$  para  $\alpha = 1/(2\sigma^2)$  y llegamos al problema de verificación de la hipótesis  $H_1 = \{\alpha_1 \leq \alpha \leq \alpha_2\}$ ,  $\alpha_1 = 1/(2\sigma_1^2)$ ,  $\alpha_2 = 1/(2\sigma_2^2)$ , según la observación  $T(X)$  que tiene la distribución  $\Gamma_{\alpha, n/2}$  de una familia exponencial. Con ayuda de los resultados de los párrafos precedentes podemos construir el criterio no desplazado y uniformemente más potente, de nivel  $1 - \varepsilon$ , que acepta  $H_1$  cuando

$$c_1 \leq T(X) \leq c_2, \quad (I)$$

donde  $c_i$  se elige de modo que  $\Gamma_{\alpha_1, n/2}(R \setminus [c_1, c_2]) = \Gamma_{\alpha_2, n/2}(R \setminus [c_1, c_2]) = \varepsilon$ .

Nótese que en este ejemplo podríamos construir el criterio de la forma (I) partiendo también de otras consideraciones, o sea, basándonos en el principio de insuficiencia, ya que la estadística  $T$  es suficiente. Pues sabemos que toda la información acerca del parámetro  $\sigma^2$  está concentrada en  $T$  y no vale la pena utilizar otras estadísticas (o sea, otra información relacionada con la muestra).

En lo sucesivo, allí donde sea posible, reduciremos inmediatamente este problema al problema de distribución de las estadísticas suficientes.

**Ejemplo 2.** Supongamos que  $X \in \Phi_{\alpha, \sigma^2}$ ,  $H_1 = \{\sigma_1 \leq \sigma \leq \sigma_2\}$ . En este caso  $\theta = (\alpha, \sigma^2)$  y la transformación de desplazamiento  $gX = X + c = (x_1 + c, \dots, x_n + c)$  induce la transformación  $\bar{g}\alpha = \alpha + c$  que mantiene invariable la hipótesis  $H_1$ . Si nos limitamos a investigar las estadísticas suficientes

$$T_1 = \bar{x}, \quad T_2 = \sum_{i=1}^n (x_i - \bar{x})^2,$$

entonces, la transformación  $g$  proporcionará

$$T_1(gX) = \bar{x} + c, \quad T_2(gT) = T_2(X).$$

Ahora bien, la estadística  $T_2$  es invariante respecto a  $G$ . Es decir, el criterio invariante  $\pi$ , basado en las estadísticas suficientes, debe ser una función de  $T_2$ . (Más adelante veremos que *cualquier* criterio invariante  $\pi$  debe ser una función de  $T_2$ ). En virtud del § 2.32,  $\sigma^{-2}T_2 \in \Gamma_{1/2, (n-1)/2}$  y llegamos al problema examinado en el ejemplo precedente. El criterio invariante no desplazado y uniformemente más potente tendrá la forma  $c_1 \leq T_2 \leq c_2$ .

**Ejemplo 3.** Los dos ejemplos examinados más arriba se referían a la distribución normal. Con arreglo a la distribución de la muestra  $X$ , la misma era una distribución normal multidimensional con una matriz diagonal de segundos momentos. Para la exposición posterior es útil notar que la familia de distribuciones normales multidimensionales arbitrarias  $\Phi_{\alpha, \sigma^2}$ ,  $\alpha \in R^m$ ,  $\sigma^2 = \|\sigma_{ij}\|$ ,  $i, j = 1, \dots, m$  es invariante respecto al grupo  $G$  de transformaciones no degeneradas lineales

$$gx = (x - a)C,$$

donde  $C$  es una matriz inversa. En efecto, debemos convencernos que, con cierta transformación  $\bar{g}$ , se cumple  $P_{\bar{g}\theta}(A) = P_{\theta}(g^{-1}A)$ , donde  $P_{\theta} \equiv \Phi_{\alpha, \sigma^2}$ ,  $\theta = (\alpha, \sigma^2)$ ,  $g^{-1}A$  significa, por lo común, el conjunto  $g^{-1}A = \{x \in R^m: gx \in A\}$ . Tenemos ( $\sigma = \sqrt{|\sigma^2|}$ )

$$\Phi_{\alpha, \sigma^2}(g^{-1}A) = \frac{1}{(2\pi)^{m/2}\sigma} \int_{g^{-1}A} \exp \left\{ -\frac{1}{2} (x - \alpha)\sigma^{-2}(x - \alpha)^T \right\} dx.$$

Después de sustituir  $y = gx$ , obtenemos

$$\Phi_{\alpha, \sigma^2}(g^{-1}A) = \frac{1}{(2\pi)^{m/2}\sigma|C|} \int_A \exp \left\{ -\frac{1}{2} (g^{-1}y - \alpha)\sigma^{-2}(g^{-1}y - \alpha)^T \right\} dy.$$

Teniendo en cuenta que  $g^{-1}y = yC^{-1} + a$ , en la última integral podemos escribir el exponente de la forma siguiente:

$$(y - (\alpha - a)C^{-1}\sigma^{-2}(C^{-1})^T(y - (\alpha - a)C)^T).$$

Por consiguiente, si se pone

$$\bar{g}\theta = \bar{g}(\alpha, \sigma^2) = (g\alpha, C^T\sigma^2C) = ((\alpha - a)C, C^T\sigma^2C), \quad (2)$$

obtenemos

$$\Phi_{\alpha, \sigma^2}(g^{-1}A) = \Phi_{\bar{g}(\alpha, \sigma^2)}(A). \quad (3)$$

**Ejemplo 4.** Supongamos que las hipótesis  $H_j$  tienen la forma siguiente:  $H_j = \{X \in P_{j, \alpha}\}$ ,  $\alpha \in \mathcal{X}$ ,  $j = 1, 2$ , donde  $P_{j, \alpha}$  son las distribuciones con densidades  $f_j(x - \alpha)$ ,  $j = 1, 2$ . Con otras palabras, nos interesa a cuál de dos tipos de distribuciones le pertenece, con una exactitud de hasta el desplazamiento, la muestra  $X$ . Aquí conviene poner  $\theta = (\nu, \alpha)$ ,  $\nu = 1, 2$ ,  $\alpha \in \mathcal{X}$  y examinar la transformación  $gX = X + c$  que en el espacio paramétrico induce la transformación  $\bar{g}(\theta) = (\nu, \alpha + c)$ . Está claro que las hipótesis  $H_j = \{\nu = j\}$ ,  $j = 1, 2$  son invariantes respecto a  $\bar{g}$  y, por lo tanto, el problema de verificación de estas hipótesis también es invariante. La estadística

$$Y = (x_1 - x_n, \dots, x_{n-1} - x_n)$$

será invariante respecto a  $g$  (compárese con el § 2.18). La distribución de esta estadística en el punto  $y = (y_1, \dots, y_{n-1})$ , en caso de la hipótesis  $H_j$ , tiene la densidad siguiente:

$$f_j^Y(y) = \int \prod_{i=1}^{n-1} f_j(y_i + z) f_j(z) \mu(dz). \quad (4)$$

De aquí se deduce que para la observación  $Y$ , las hipótesis  $H_j$  se transforman en hipótesis simples, conforme a las cuales las densidades  $f_j^Y$  para  $Y$  tienen la forma (4). En estas condiciones podemos hacer uso del lema de Neumann — Pearson y construir el c.m.p.  $\pi$  que acepta la hipótesis  $H_2$  si

$$f_2^Y(Y)/f_1^Y(Y) > c. \quad (5)$$

Como este criterio no depende de  $\alpha$ , el mismo será el c.u.m.p. para verificar  $H_1$  frente a  $H_2$  entre todos los criterios invariantes basados en la estadística  $Y$ .

Con arreglo a los ejemplos examinados es conveniente estar seguro de que los demás criterios invariantes en estos problemas también son funciones de las estadísticas invariantes escogidas por nosotros. Esto se refiere especialmente al último ejemplo, puesto que en los dos ejemplos anteriores, la elección de los criterios también se basaba en las consideraciones de suficiencia.

Para aclarar las relaciones mutuas entre los invariantes, introduzcamos algunos conceptos. Dos puntos  $x$  y  $x'$  de  $\mathcal{X}^n$  se llamarán *equivalentes* respecto al grupo  $G$  si existe  $g \in G$  tal, que  $x' = gx$ . Como  $G$  es un grupo, entonces todo el espacio  $\mathcal{X}^n$  se divide en clases disjuntas de equivalencia, que en el § 2.19 hemos llamado *órbitas*. Para obtener cierta órbita es suficiente tomar un punto cualquiera  $x_0$  de la misma y aplicar a éste todas las transformaciones  $g$  de  $G$ . Por ejemplo, para las transformaciones ortogonales del ejemplo 1, las órbitas forman esferas cuyos centros coinciden con el origen de coordenadas.

La invariación de la estadística  $T$  respecto a  $G$  es unívoca al hecho de que  $T$  es constante en cada órbita.

**Definición 3.** La estadística  $T$  se denomina *invariante máximo* si la misma es invariante, y de  $T(x') = T(x)$  se deduce  $x' = gx$  para cierto  $g \in G$ .

Esto significa que el invariante máximo adopta distintos valores en órbitas diferentes.

**Teorema 1.** Sea  $T$  el invariante máximo. La estadística  $S$  es invariante si y sólo si  $S$  depende de  $X$  a través de  $t$ , o sea, si existe una función  $\varphi$  tal, que  $S(X) = \varphi(T(X))$ .

Para simplificar la exposición, aquí no tratamos una cuestión importante, relacionada con la mensurabilidad de  $\varphi$ . Nótese solamente que en los ejemplos examinados en este párrafo, tal mensurabilidad tendrá lugar<sup>\*)</sup>.

**Demostración.** Si  $S(x) = \varphi(t(x))$ , entonces  $S(gx) = \varphi T(gx) = \varphi(T(x)) = S(x)$  y, por lo tanto,  $S$  es invariante. Para demostrar la afirmación inversa debemos convencernos de que de  $T(x) = T(x')$  resulta  $S(x) = S(x')$ . Pero esto es así en virtud del hecho de que  $T(x) = T(x')$  provoca la existencia de una  $g$  tal, que  $x' = gx$ . Pero como  $S$  es un invariante,  $S(x) = S(x')$ . <

A título de ejemplo examinemos el grupo  $G$  de desplazamientos

$$gx = x + c = (x_1 + c, \dots, x_n + c).$$

Como ya hemos señalado, la estadística  $Y(x) = (x_1 - x_n, \dots, x_{n-1} - x_n)$  es un invariante. Mostremos que éste es el invariante máximo. En efecto, de  $Y(x) = Y(x') \equiv (x'_1 - x'_n, \dots, x'_{n-1} - x'_n)$  se desprende que  $x_i - x_n = x'_i - x'_n$  para todos  $i = 1, \dots, n-1$ . Poniendo  $x'_n - x_n = c$ , obtenemos  $x'_i = x_i + c$ ,  $i = 1, \dots, n$ ,  $x' = x + c = gx$ , lo que precisamente significa la equivalencia necesaria de  $x'$  y  $x$ .

Ahora podemos volver al ejemplo 3 y afirmar que el criterio (5) es el c.u.m.p. entre todos los criterios invariantes, puesto que según el teorema 1 todos los criterios invariantes son funciones de  $Y$  y, por consiguiente, la suposición de que exista un criterio invariante más potente que (5) será contradictoria.

Por analogía a lo expuesto anteriormente, el lector puede convencerse de que la estadística  $\sum_{i=1}^n x_i^2$  en el ejemplo 1 también es un invariante máximo.

Si existen estadísticas suficientes, al principio suele ser conveniente reducir el problema inicial al problema respecto a la distribución de las estadísticas suficientes y luego emplear las consideraciones de invariación así como se hizo en el ejemplo 2, donde la estadística  $T_2 = \sum_{i=1}^n (x_i - \bar{x})^2$  es, evidentemente, el máximo invariante en la observación  $(\bar{x}, T_2)$ .

En conclusión de este párrafo es preciso señalar una vez más, que la esencia del enfoque relacionado con la invariación consiste en que los problemas sometidos a examen y destinados a la verificación de las hipótesis, deben reducirse a problemas más simples, referentes a la distribución de los invariantes máximos. En estas nuevas condiciones, que son más simples, resulta posible, en varios casos, construir el c.m.p. o el c.u.m.p.

<sup>\*)</sup> Véanse, por ejemplo, [57] y [95].



En este sentido, el "principio de invariación" se asemeja a los "principios" de suficiencia y de no desplazamiento, de acuerdo con los cuales el problema inicial se reduce a un problema en términos de estadística suficiente o de estadística no desplazada.

### § 8\*. Enlace con los conjuntos confidenciales.

**1. Enlace de los criterios estadísticos y los conjuntos confidenciales. Enlace de las propiedades de optimización.** Los conceptos de conjunto confidencial y de criterio estadístico están estrechamente ligados entre sí. En el § 2.31 hemos dado la definición del conjunto confidencial. Recordémosla.

Sea  $X \in \mathbf{P}_\theta$ ,  $\theta \in \Theta$ .

**Definición 1.** El subconjunto aleatorio  $\Theta^* = \Theta^*(x, \varepsilon)$  del espacio estadístico  $\Theta$  se llama *conjunto confidencial de nivel  $1 - \varepsilon$* , si

$$\mathbf{P}_\theta(\Theta^*(X, \varepsilon) \ni \theta) \geq 1 - \varepsilon \quad (1)$$

para todos  $\theta \in \Theta$ .

Evidentemente, el intervalo confidencial es un caso particular del conjunto confidencial. Este último tiene el mismo sentido: con una probabilidad  $\geq 1 - \varepsilon$  recubre el valor verdadero del parámetro.

Designemos

$$\Omega(\theta, \varepsilon) = \{x \in \mathcal{X}^n: \theta \in \Theta^*(x, \varepsilon)\}. \quad (2)$$

Entonces, las relaciones

$$\theta \in \Theta^*(x, \varepsilon) \quad \text{y} \quad x \in \Omega(\theta, \varepsilon) \quad (3)$$

serán equivalentes.

La definición del conjunto confidencial supone que el conjunto  $\Omega(\theta, \varepsilon)$  en (2) es medible, así que la probabilidad en (1) tiene sentido y es igual a  $\mathbf{P}_\theta(X \in \Omega(\theta, \varepsilon))$ .

Los conjuntos confidenciales y los criterios estadísticos para verificar la hipótesis  $H_1 = \{\theta = \theta_1\}$  frente a la alternativa adicional  $H_2 = \{\theta \in \Theta_2\}$ ,  $\theta_1 \notin \Theta_2$ , están enlazados entre sí del modo siguiente. Supongamos que para cada  $\theta_1$  ha sido definido su conjunto  $\Theta_2 = \Theta_2(\theta_1) \not\ni \theta_1$ .

**Teorema 1.** 1) *Examinemos para cada  $\theta_1$  el criterio no randomizado  $\pi = \delta$  de nivel  $1 - \varepsilon$  para verificar la hipótesis  $H_1$  frente a  $H_2$ , y designemos por  $\Omega(\theta_1, \varepsilon)$  su región de aceptación de la hipótesis  $H_1$ . Entonces, el conjunto*

$$\Theta^*(X, \varepsilon) = \{\theta \in \Theta; X \in \Omega(\theta, \varepsilon)\}$$

*será un conjunto confidencial de nivel  $1 - \varepsilon$ .*

Al contrario, si  $\Theta^*(X, \varepsilon)$  es un conjunto confidencial de nivel  $1 - \varepsilon$ , entonces el conjunto  $\Omega(\theta_1, \varepsilon) \subset \mathcal{L}^n$ , definido en (2) y adoptado como región de aceptación de  $H_1$ , determinará el criterio para verificar  $H_1 = \{\theta = \theta_1\}$  frente a  $H_2 = \{\theta \in \Theta_2(\theta_1)\}$  de nivel  $1 - \varepsilon$  para cualquier  $\Theta_2(\theta_1)$ ,  $\theta_1 \notin \Theta_2(\theta_1)$ .

2) Si el criterio  $\pi$  con la región de aceptación  $\Omega(\theta_1, \varepsilon)$  de la hipótesis  $H_1$  es el c.u.m.p., entonces, el conjunto respectivo  $\Theta^*(X, \varepsilon)$  minimizará la probabilidad

$$P_\theta(\theta' \in \Theta^*(X, \varepsilon)) \text{ para todos } \theta, \theta', \theta \in \Theta_2(\alpha') \quad (4)$$

en la clase de todos los conjuntos confidenciales de nivel  $1 - \varepsilon$ .

También es cierta la afirmación contraria: La minimalidad (4) significa que el conjunto respectivo  $\Omega(\theta, \varepsilon)$  engendrará el c.u.m.p.

Para el parámetro unidimensional se usan principalmente los casos

$$\Theta_2(\theta') = \{\theta: \theta \neq \theta'\} \text{ y } \Theta_2(\theta') = \{\theta: \theta > \theta'\} \text{ (o bien } \{\theta: \theta < \theta'\}).$$

En el primero de ellos en (4) tendrá lugar la minimización para todos  $\theta' \neq \theta$ , y en el segundo, para todos  $\theta' < \theta$ .

Así pues, en (4), el teorema afirma que para  $\Theta^*$ , la probabilidad  $P_\theta$  se minimiza de que todo otro valor de  $\theta' \neq \theta$ , tal que  $\theta \in \Theta_2(\theta')$ , pertenezca a un conjunto confidencial. Esta es una de las maneras de separar los intervalos confidenciales óptimos.

**Definición 2.** Los conjuntos confidenciales para los cuales se minimiza (4) a condición (1) se llaman *conjuntos confidenciales más exactos* (de nivel  $1 - \varepsilon$ ) respecto a las alternativas  $\theta'$  tales que  $\theta \in \Theta_2(\theta')$ .

Más adelante expondremos cierta argumentación adicional para tal entendimiento del intervalo confidencial óptimo.

Ahora bien, el teorema 1, establece que la "inversión" del conjunto  $\Omega(\theta_1, \varepsilon)$  para el c.u.m.p. da el conjunto confidencial más exacto. En este caso es importante señalar que el referido procedimiento de construcción de los conjuntos confidenciales no está de ningún modo relacionado con la dimensión de  $\theta$ . Incluso se pueden examinar los parámetros de dimensión infinita  $\theta$  e identificar  $\theta$  con la propia distribución  $\mathbf{P}$  de la muestra  $X$ . Entonces, las relaciones de equivalencia (3), donde  $\Omega(\theta, \varepsilon) = \Omega(\mathbf{P}, \varepsilon)$  es la región de aceptación de la hipótesis  $\{X \in \mathbf{P}\}$  frente a la alternativa  $\{X \in \mathbf{P}_1 \neq \mathbf{P}\}$ , permiten construir el conjunto confidencial para  $\mathbf{P}$ . Por ejemplo, en el § 1.6 hemos visto que la distribución de la estadística  $D_n = \sqrt{n} \sup_t |F_n^*(t) - F(t)|$ , a condición de que  $X \in \mathbf{P}$ , donde  $F$  es una función continua de la distribución correspondiente a  $\mathbf{P}$ , no depende de  $F$  y puede ser determinada. Por consiguiente, podemos hallar tal  $d = d(\varepsilon)$ , que

$P(D_n \leq d(\varepsilon)) = 1 - \varepsilon$ . Ahora bien, la desigualdad

$$\sqrt{n} \sup_f |F_n^*(f) - F(f)| \leq d$$

define la región de aceptación de la hipótesis  $\{X \in P\}$  para el criterio de nivel  $1 - \varepsilon$ .

Pero esta misma desigualdad también define el conjunto confidencial para  $F$ : simplemente debido a la simetría de esta desigualdad respecto a  $F$  y  $F_n^*$ , aquí no se necesita ningún procedimiento especial de "inversión".

La demostración del teorema 1 es casi evidente. La misma se basa en la equivalencia (3), en virtud de la cual

$$P_\theta(\theta \in \Theta^*(X, \varepsilon)) = P_\theta(X \in \Omega(\theta, \varepsilon)) \geq 1 - \varepsilon.$$

Esto demuestra la primera afirmación. Para demostrar la segunda examinemos cualquier otro conjunto confidencial  $\tilde{\Theta}^*(X, \varepsilon)$ , y sea  $\tilde{\Omega}(\theta, \varepsilon)$  el subconjunto correspondiente en  $\mathcal{X}^n$ .

Entonces,

$$\begin{aligned} P_\theta(X \in \tilde{\Omega}(\theta, \varepsilon)) &= P_\theta(\theta \in \tilde{\Theta}^*(X, \varepsilon)) \geq 1 - \varepsilon, \\ P_\theta(X \in \tilde{\Omega}(\theta_1, \varepsilon)) &\geq P_\theta(X \in \Omega(\theta_1, \varepsilon)) \end{aligned}$$

para todos  $\theta \in \Theta_2(\theta_1)$  y, por lo tanto,

$$P_\theta(\theta_1 \in \tilde{\Theta}^*(X, \varepsilon)) \geq P_\theta(\theta_1 \in \Theta^*(X, \varepsilon)). \quad \triangleleft$$

Examinemos ahora un importante caso particular relacionado con el parámetro unidimensional  $\theta$ .

## 2. Intervalos confidenciales más exactos.

**Teorema 2.** *Supongamos que el conjunto  $\Omega(\theta, \varepsilon)$  del c.u.m.p. examinado en el teorema 1 tiene la forma*

$$c_1(\theta, \varepsilon) \leq T(x) \leq c_2(\theta, \varepsilon),$$

donde  $c_i(\theta, \varepsilon)$  dependen monótona y continuamente\* de  $\theta$ . Supongamos, para precisar, que  $c_i(\theta, \varepsilon)$  crecen. Entonces, el conjunto confidencial más exacto (de nivel  $1 - \varepsilon$ ) respecto a las alternativas  $\theta'$  tales, que  $\theta \in \Theta_2(\theta')$ , tendrá la forma de intervalo

$$c_2^{-1}(T, \varepsilon) \leq \theta \leq c_1^{-1}(T, \varepsilon),$$

donde  $T = T(X)$ ,  $c_i^{-1}(t, \varepsilon)$  son las soluciones de las ecuaciones  $c_i(\theta, \varepsilon) = t$  respecto a  $\theta$ .

\* Las propiedades de monotonía y de continuidad de  $c_i(\theta, \varepsilon)$  se deducen, por lo general, de las mismas propiedades de la función de distribución  $P_\theta T(X) < c$ . En las designaciones del § 2.31,  $c_1(\theta, \varepsilon) = G_\theta^{-1}(\varepsilon_1)$ ,  $c_2(\theta, \varepsilon) = G_\theta^{-1}(1 - \varepsilon_2)$ , donde  $G_\theta$  es la función de distribución  $T(X)$ ,  $\varepsilon_1 + \varepsilon_2 = \varepsilon$ .

Ahora bien, vemos que el procedimiento de construcción del intervalo confidencial es aquí, de hecho, el mismo que en el § 2.31 con la única particularidad de que en calidad de estadística  $S$  aquí se utiliza la estadística  $T$  del criterio uniformemente más potente.

La demostración del teorema es evidente y se la dejamos al lector.

Ahora examinemos más detalladamente los *intervalos confidenciales unilaterales para  $\theta$  escalar*. Estos intervalos se utilizan allí donde reviste mayor interés una sola cota para estimar el parámetro. Tales situaciones surgen cuando se estima la probabilidad de que se produzca cualquier suceso indeseable o, digamos, cuando se estima el esfuerzo de rotura de una nueva aleación.

Debido a la simetría es posible reducirse al examen de la frontera confidencial inferior  $\theta^-(X, \varepsilon)$  para la cual

$$\mathbf{P}_\theta(\theta^-(X, \varepsilon) \leq \theta) \geq 1 - \varepsilon. \quad (5)$$

**Definición 3.** La frontera  $\theta^- = \theta^-(X, \varepsilon)$  para la cual  $\mathbf{P}_\theta(\theta^- \leq \theta')$  es mínima con todos  $\theta' < \theta$  se llama *frontera confidencial inferior más exacta de nivel  $1 - \varepsilon$* .

Supongamos que  $w(\theta^-, \theta)$  es cualquier medida de pérdidas que surgen debido a la "subestimación" de  $\theta$ :  $w(\theta^-, \theta) = 0$  cuando  $\theta^- \geq \theta$  y  $w(\theta^-, \theta) \geq 0$  cuando  $\theta^- < \theta$ ; en este caso  $w(\theta^-, \theta)$  crece continuamente al alejarse  $\theta^-$  de  $\theta$ ,  $\mathbf{M}_\theta w(\theta^-, \theta) < \infty$ .

La siguiente afirmación aclara, en cierta medida, el sentido de la definición 3.

**Lema 1.** *La frontera inferior más exacta  $\theta^-$  minimiza el valor  $\mathbf{M}_\theta w(\theta^-, \theta)$  para la condición (5) y para cualquier función  $w$  que posea las propiedades enunciadas anteriormente.*

**Demostración.** Sea  $\tilde{\theta}^-$  otra frontera inferior. Entonces, como los incrementos  $d_u w(u, \theta)$  respecto a  $u$  en la región  $u < \theta$  son negativos,

$$\begin{aligned} \mathbf{M}_\theta w(\theta^-, \theta) &= \int_{-\infty}^{\theta} w(u, \theta) d_u \mathbf{P}_\theta(\theta^- < u) = - \int_{-\infty}^{\theta} \mathbf{P}_\theta(\theta^- < u) d_u w(u, \theta) \leq \\ &\leq - \int_{-\infty}^{\theta} \mathbf{P}_\theta(\tilde{\theta}^- < u) d_u w(u, \theta) = \mathbf{M}_\theta w(\tilde{\theta}^-, \theta) \quad \triangleleft \end{aligned}$$

Así pues, vemos que el enfoque de la definición de los conjuntos confidenciales más exactos en caso de los conjuntos unilaterales es muy natural. Ahora, con ayuda de los teoremas 1 y 2 y los resultados del § 5 se pueden construir explícitamente los intervalos confidenciales unilaterales para el caso cuando la relación de verosimilitud es monótona.

**Teorema 3.** *Supongamos que  $X \in \mathbf{P}_\theta$  y que la familia  $\{\mathbf{P}_\theta\}$  tiene relación de verosimilitud monótona respecto a la estadística  $T(X)$  cuya  $\mathbf{P}_\theta$ -distribución  $G_\theta(t) = \mathbf{P}_\theta(T(X) < t)$  es continua respecto a  $\theta$  y  $t$ . Enton-*

ces, la estadística  $T$  de la distribución depende monótona y continuamente de  $\theta$ , (o sea,  $G_\theta(t)$  decrece continuamente con el crecimiento de  $\theta$ , véase la definición 2.31.3). Si  $b(t, \gamma)$  es la solución de la ecuación  $G_\theta(t) = \gamma$  respecto a  $\theta$ , entonces, la frontera inferior más exacta  $\theta^-(X, \varepsilon)$  de nivel  $1 - \varepsilon$  es igual a

$$\theta^-(X, \varepsilon) = b(T(X), 1 - \varepsilon).$$

Con otras palabras, en la afirmación del teorema 2.31.1 obtendremos la frontera confidencial inferior más exacta si utilizamos en calidad de  $S$  la estadística  $T$ .

**Demostración.** En nuestro caso, en condiciones de los teoremas 1 y 2 es necesario poner  $\Theta_2(\theta) = \{t: t > \theta\}$ . En virtud del teorema 5.1 existe un c.u.m.p. no randomizado para verificar  $H_1 = \{\theta = \theta_1\}$  frente a  $H_2 = \{\theta > \theta_1\}$  con la región  $\Omega(\theta_1, \varepsilon) = \{X: T(X) < c\}$  de aceptación de  $H_1$ , donde  $c = c(\theta_1, 1 - \varepsilon) = G_{\theta_1}^{-1}(1 - \varepsilon)$  se deduce de la condición

$$P_{\theta_1}(T(X) < c(\theta_1, 1 - \varepsilon)) = 1 - \varepsilon.$$

En este caso

$$P_\theta(T(X) \geq c) > \varepsilon = P_{\theta_1}(T(X) \geq c)$$

cuando  $\theta > \theta_1$ . Esto último quiere decir que  $c(\theta_1, 1 - \varepsilon) < c(\theta, 1 - \varepsilon)$  cuando  $\theta_1 < \theta$ , o sea, la función  $c(\theta, 1 - \varepsilon)$  crece respecto a  $\theta$ . La continuidad de  $c(\theta, 1 - \varepsilon) = G_\theta^{-1}(1 - \varepsilon)$  respecto a  $\theta$  se deduce de la continuidad de  $G_\theta$ .

Vemos que las condiciones de los teoremas 1 y 2 se cumplen por completo cuando  $c_2(\theta, \varepsilon) = c(\theta, 1 - \varepsilon)$  y, por lo tanto, el conjunto confidencial más exacto tiene la forma del semiintervalo  $(c^{-1}(T(X), 1 - \varepsilon), \infty)$ , donde, como hemos visto en el teorema 2.31.1,  $c^{-1}(T, 1 - \varepsilon) = b(T, 1 - \varepsilon)$ .  $\triangleleft$

De un modo exactamente igual se puede construir la frontera superior más exacta  $\theta^+(X, \varepsilon)$ .

Ahora supongamos que  $\theta^-(X, \varepsilon_1) < \theta^+(X, \varepsilon_2)$  designan las fronteras confidenciales superior e inferior de los niveles  $1 - \varepsilon_1$  y  $1 - \varepsilon_2$ , respectivamente. Como los sucesos  $\{\theta^-(X, \varepsilon_1) > \theta\}$  y  $\{\theta^+(X, \varepsilon_2) < \theta\}$  son disjuntos, entonces

$$P_\theta(\theta^-(X, \varepsilon_1) < \theta < \theta^+(X, \varepsilon_2)) = 1 - \varepsilon_1 - \varepsilon_2,$$

y  $(\theta^-(X, \varepsilon_1), \theta^+(X, \varepsilon_2))$  es el intervalo confidencial de nivel  $1 - \varepsilon_1 - \varepsilon_2$ .

Sean  $w_1(\theta^-, \theta)$  y  $w_2(\theta^+, \theta)$  las funciones de pérdidas para las fronteras  $\theta^\mp$  que poseen las propiedades descritas en la enunciación del lema 1.

**Lema 2.** Sea  $w(\theta^-, \theta^+, \theta) = w_1(\theta^-, \theta) + w_2(\theta^+, \theta)$ . Entonces, el intervalo confidencial  $(\theta^-, \theta^+)$ , formado por las fronteras superiores e inferiores más exactas, minimiza  $M_\theta w(\theta^-, \theta^+, \theta)$  para las condiciones

$$P_\theta(\theta^- > \theta) \leq \varepsilon_1, \quad P_\theta(\theta^+ < \theta) \leq \varepsilon_2.$$

Este lema es el corolario evidente del lema 1. El mismo muestra que el intervalo confidencial construido con ayuda de las fronteras inferior exacta y superior exacta también poseerá propiedades de optimización.

El teorema 3 da la posibilidad de construir explícitamente tales intervalos para las familias paramétricas que tienen monótonas las relaciones de verosimilitud.

Le proponemos al lector que el mismo se cerciore, a base de las observaciones efectuadas, de que los intervalos confidenciales, construidos en el § 2.32 para la media y la varianza de la distribución normal, tendrán las fronteras superiores e inferiores más exactas.

En el teorema 1 y en las investigaciones posteriores figuraba la condición de que el c.u.m.p. no es randomizado. Sin embargo, esta limitación no es importante. Cualquier criterio randomizado  $\pi$  puede ser representado como criterio no randomizado, si en la investigación se introduce una observación adicional  $Y$  que sea independiente de  $X$  y que esté uniformemente distribuida en  $[0, 1]$ . En efecto, examinemos, para la nueva muestra  $(X, Y)$ , la región crítica

$$\Omega = \{(x, y): \pi(x) \geq y\},$$

o sea, supongamos que  $\delta(X, Y) = 1$  si  $(X, Y) \in \Omega$ , y que  $\delta(X, Y) = 0$  en el caso contrario. Entonces, para toda distribución de  $X$ ,

$$P(\delta(X, Y) = 1) = P(\pi(X) \geq Y) = \int_0^1 P(\pi(X) \geq y) dy = M\pi(x),$$

y, por consiguiente, el criterio  $\delta$  es equivalente (según sus parámetros) a  $\pi$ . ¿Cómo aprovechar esta circunstancia para construir los intervalos confidenciales en condiciones del teorema 3? Supongamos, para abreviar, que la estadística  $T(X)$  es de números enteros (como hemos visto, la falta de los c.u.m.p. sólo puede ser provocada por el carácter discreto de la distribución  $T$ ). Entonces, la observación  $S(X, Y) = T(X) + Y$ ,  $Y \in U_{0,1}$  conserva toda la información contenida en  $T(X)$ , ya que  $T(X)$  es una parte entera de  $S(X, Y)$ . Eligiendo  $c(\theta, \varepsilon)$  entero, al c.u.m.p. de nivel  $1 - \varepsilon$  se le puede conferir la forma siguiente: se acepta la hipótesis  $H_1$  si

$$S(X, Y) \leq c(\theta_1, 1 - \varepsilon).$$

Así pues, hemos construido los conjuntos requeridos  $\Omega(\theta, \varepsilon)$  y sólo queda "invertirlos" usando el mismo procedimiento que antes. Obtendremos la frontera inferior

$$\theta^-(X, Y, \varepsilon) = c^{-1}(T(X) + Y, 1 - \varepsilon),$$

donde  $c^{-1}$  es la función inversa a  $c$  con arreglo al primer argumento. Aquí, de la propia escritura se deduce que para definir  $\theta^-$  es necesario realizar una observación adicional  $Y$ .

**Ejemplo 1.** Sea  $X \in \mathbf{B}_p$ , y nos interesa la frontera confidencial superior  $p^+$  de nivel  $1 - \varepsilon$  para la probabilidad  $p = \mathbf{P}(x_i = 1) = 1 - \mathbf{P}(x_i = 0)$ . La familia de distribuciones  $\{\mathbf{B}_p\}$  es exponencial y satisface las condiciones del teorema 3, donde conviene poner  $T(X) = \sum_{i=1}^n x_i$ . Examinemos la observación

$$S = \sum_{i=1}^n x_i + Y, \quad Y \in U_{0,1}.$$

Esta tiene en el punto  $t$ ,  $0 \leq t \leq n + 1$ , la densidad  $C_n^{[t]} p^{[t]} (1-p)^{n-[t]}$ . Designemos por  $G_p^{(t)}$  la función de distribución con esta densidad. Entonces  $p^+$  será la solución de la ecuación  $G_p(t) = \varepsilon$ .

**3. Conjuntos confidenciales no desplazados.** Volvamos a la cuestión acerca de los conjuntos confidenciales más exactos. Con ayuda del teorema 3 podemos construir las fronteras superiores e inferiores más exactas basándose en el hecho de que para las alternativas unilaterales  $\{\theta > \theta_1\}$ ,  $\{\theta < \theta_1\}$  de las hipótesis  $\{\theta = \theta_1\}$ , en una serie de casos existe el c.u.m.p. Si tratamos de utilizar los teoremas 1 y 2 directamente para construir los intervalos confidenciales más exactos, necesitaremos la existencia de c.u.m.p. para verificar la hipótesis  $\{\theta = \theta_1\}$  frente a  $\{\theta \neq \theta_1\}$ , lo cual ocurre muy raramente. La salida de esta posición consiste en la reducción natural de la clase de intervalos confidenciales sujetos a investigación, procediendo del mismo modo que cuando reducimos las clases de criterios examinados en el § 6.7, es decir, introduciendo los conceptos de conjuntos confidenciales no desplazados e invariantes.

Supongamos que, como antes, a cada  $\theta$  le corresponde el conjunto  $\Theta_2(\theta)$ ,  $\theta \in \Theta_2(\theta)$ .

**Definición 4.** El conjunto confidencial  $\Theta^*(X, \varepsilon)$  para  $\theta$  de nivel  $1 - \varepsilon$  se considera *no desplazado respecto a las alternativas  $\theta'$* , tales que  $\theta \in \Theta_2(\theta')$  si

$$\mathbf{P}_\theta(\theta' \in \Theta^*(X, \varepsilon)) \leq 1 - \varepsilon \text{ para todos } \theta, \theta', \theta \in \Theta_2(\theta'). \quad (6)$$

El conjunto  $\Theta^*(X, \varepsilon)$  se considera simplemente *no desplazado* si (6) es válida para todos  $\theta' \neq \theta$ .

El no desplazamiento del conjunto confidencial significa que la *probabilidad de que éste recubra el valor falso de  $\theta'$  no es mayor que la probabilidad de que el mismo recubra el valor verdadero*.

**Definición 5.** Los conjuntos confidenciales para los cuales se minimiza (4) en condiciones (1) y (6) se llaman *conjuntos confidenciales no desplazados más exactos* ((de nivel  $1 - \varepsilon$ ) respecto a las alternativas para las cuales  $\theta \in \Theta_2(\theta')$ ).

**Teorema 4.** 1) *Los criterios no randomizados y no desplazados engendran, en virtud de la equivalencia (3), conjuntos confidenciales no desplazados, y al contrario.*

2) Si  $\Omega(\theta_1, \varepsilon)$  para cada  $\theta_1 \in \Theta$  es la región de aceptación de la hipótesis  $\{\theta = \theta_1\}$  del criterio uniformemente más potente no desplazado y no randomizado, con una alternativa  $\{\theta \in \Theta_2(\theta_1)\}$ , entonces, el conjunto respectivo  $\Theta^*(X, \varepsilon)$  será el conjunto confidencial no desplazado más exacto, y al contrario.

La demostración del teorema repite por completo los razonamientos del teorema 1, a los cuales sólo es necesario añadir que la propiedad de desplazamiento se conserva al pasar de los criterios a los conjuntos confidenciales y al contrario. En efecto, las relaciones (1) y (6) son equivalentes a

$$\sup_{\theta \in \Theta_2(\theta_1)} \mathbf{P}_\theta(X \in \Omega(\theta_1, \varepsilon)) \leq 1 - \varepsilon \leq \mathbf{P}_{\theta_1}(X \in \Omega(\theta_1, \varepsilon)).$$

Si  $\pi(X)$  es la función crítica de los criterios no randomizados que figuran en el teorema  $\pi(X) = 0$  para  $X \in \Omega(\theta_1, \varepsilon)$ , entonces obtenemos

$$\begin{aligned} \mathbf{M}_\theta \pi(X) &= 1 - \mathbf{P}_\theta(X \in \Omega(\theta_1, \varepsilon)), \\ \inf_{\theta \in \Theta_2(\theta_1)} \mathbf{M}_\theta \pi(X) &\geq \varepsilon \geq \mathbf{M}_{\theta_1} \pi(X). \end{aligned}$$

Esta es, precisamente, la propiedad de no desplazamiento que equivale a (6).  $\triangleleft$

Si utilizamos los resultados del § 6 y construimos el conjunto confidencial no desplazado y más exacto para el parámetro  $\theta$  de una familia exponencial, obtendremos el mismo intervalo confidencial  $(\theta^-, \theta^+)$  que hemos construido utilizando la monotonía de la relación de verosimilitud, o sea, el intervalo en el cual  $\theta^-$  y  $\theta^+$  son las fronteras inferior y superior más exactas, respectivamente, de niveles  $1 - \varepsilon/2$ .

**4. Conjuntos confidenciales invariantes.** La siguiente definición utiliza las designaciones y los conceptos del párrafo precedente. Sea  $\{\mathbf{P}_\theta\}$  una familia invariante respecto a  $G$ .

**Definición 6.** El conjunto confidencial  $\Theta^*(X, \varepsilon)$  se llama *invariante*<sup>\*)</sup> respecto al grupo  $G$  si

$$\Theta^*(gX, \varepsilon) = \bar{g}\Theta^*(X, \varepsilon) \quad (7)$$

para todos  $g \in G$ .

El sentido de este concepto es análogo al de la estimación equivariante (§ 2.19). Si las transformaciones  $g$  y  $\bar{g}$  se interpretan como la sustitución del sistema de coordenadas que conserva la distribución, entonces (7) significará que el conjunto confidencial no depende del sistema de coordenadas en el que se expresan los datos iniciales.

<sup>\*)</sup> Ateniéndose a la observación expuesta en la p. 195 del § 2.19, sería más natural llamar el conjunto confidencial con propiedad (7), conjunto equivariante.



**Definición 7.** El conjunto confidencial  $\Theta^*(X, \varepsilon)$  se denomina *conjunto confidencial invariante más exacto* de nivel  $1 - \varepsilon$ , si en él se minimiza  $P_\theta(\theta' \in \Theta^*(X, \varepsilon))$  para todos  $\theta' \neq \theta$  en la clase de todos los conjuntos  $\Theta^*$  que satisfacen (7) y la condición  $P_\theta(\theta \in \Theta^*(X, \varepsilon)) = 1 - \varepsilon$ . Sea  $\Omega(\theta_1, \varepsilon)$  la región de aceptación de la hipótesis  $H_1 = \{\theta = \theta_1\}$  cuando la alternativa constituye  $\{\theta \neq \theta_1\}$  para el criterio invariante de nivel  $1 - \varepsilon$ . Nótese que hay una diferencia esencial en las definiciones del criterio invariante y del conjunto confidencial invariante (esta diferencia no existiría si se necesitara el cumplimiento de la igualdad  $g\Omega(\theta, \varepsilon) = \Omega(g\theta, \varepsilon)$  y no de la igualdad  $g\Omega(\theta, \varepsilon) = \Omega(\theta, \varepsilon)$ ). Con este hecho está relacionada la circunstancia de que la correspondencia entre los criterios invariantes uniformemente más potentes y los intervalos confidenciales invariantes más exactos tiene un aspecto más complejo que en los teoremas precedentes.

Examinemos el grupo de transformaciones  $G$  y supongamos que para cada  $\theta$  en este grupo hay un subgrupo  $G[\theta_1]$  que deja invariante el problema de verificación de la hipótesis  $H_1 = \{\theta = \theta_1\}$ . Con otras palabras,  $\bar{g}\theta_1 = \theta_1$  cuando  $g \in G[\theta_1]$ .

**Teorema 5.** Sea  $\Theta^*(X, \varepsilon)$  un conjunto confidencial de nivel  $1 - \varepsilon$  invariante respecto a  $G$ . Entonces

1) La región  $\Omega(\theta, \varepsilon) = \{x: \theta \in \Theta^*(x, \varepsilon)\}$  será invariante respecto a  $G[\theta]$  para cada  $\theta$ .

2) Si la región  $\Omega(\theta_1, \varepsilon)$ , correspondiente a  $\Theta^*(X, \varepsilon)$ , es la región de aceptación de  $H_1$  cuando la alternativa constituye  $\{\theta \neq \theta_1\}$  para el criterio invariante uniformemente más potente de nivel  $1 - \varepsilon$ , entonces  $\Theta^*(X, \varepsilon)$  será el conjunto invariante confidencial más exacto.

**Demostración.** 1) Supongamos que  $g \in G[\theta]$ . Entonces  $\bar{g}\theta = \theta$ ,

$$\begin{aligned} g\Omega(\theta, \varepsilon) &= \{gx: \theta \in \Theta^*(x, \varepsilon)\} = \{x: \theta \in \Theta^*(g^{-1}x, \varepsilon)\} = \\ &= \{x: \theta \in \bar{g}^{-1}\Theta^*(x, \varepsilon)\} = \{x: \bar{g}\theta \in \Theta^*(x, \varepsilon)\} = \\ &= \{x: \theta \in \Theta^*(x, \varepsilon)\} = \Omega(\theta, \varepsilon). \end{aligned}$$

2) Sea  $\tilde{\Theta}^*$  cualquier otro conjunto confidencial invariante de nivel  $1 - \varepsilon$ . Según la primera afirmación, a él le corresponde el criterio invariante de nivel  $1 - \varepsilon$  con la región  $\tilde{\Omega}(\theta_1, \varepsilon)$  de aceptación de  $H_1$ .

Como, por suposición,

$$P_\theta(X \in \Omega(\theta_1, \varepsilon)) \geq P_\theta(X \in \tilde{\Omega}(\theta_1, \varepsilon)),$$

entonces

$$P_\theta(\theta_1 \in \Theta^*(X, \varepsilon)) \geq P_\theta(\theta_1 \in \tilde{\Theta}^*(X, \varepsilon)).$$

cuando  $\theta_1 \neq \theta$ . Que es lo que se necesitaba demostrar.  $\triangleleft$

**Ejemplo 2.** Supongamos que  $X \in \Phi_{\alpha, \sigma^2}$ . Se necesita construir el conjunto confidencial más exacto para el parámetro  $\sigma^2$ , siendo desconocido  $\alpha$ . En el ejemplo 2 del párrafo precedente hemos visto que la familia  $\Phi_{\alpha, \sigma^2}$  es invariante respecto a las transformaciones de desplazamiento  $gX = X + c$  si  $\bar{g}(\alpha, \sigma^2) = (\alpha + c, \sigma^2)$ . La estadística  $S_0^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$  es el máximo invariante construido según la esta-

dística suficiente. Además, la hipótesis  $H_1 = \{\sigma = \sigma_1\}$  es invariante respecto a  $G$ . Conforme al ejemplo 7.2, el criterio uniformemente más potente invariante y no desplazado para verificar  $H_1$  tiene la forma

$$h_{1,\varepsilon} \sigma_1^2 < (n-1)S_0^2 < h_{2,\varepsilon} \sigma_1^2, \quad (8)$$

donde  $h_{i,\varepsilon}$  se deduce de las condiciones (véase la condición (6.7) del teorema 6.1):

$$\begin{aligned} \mathbf{P}(h_{1,\varepsilon} < \chi_{n-1}^2 < h_{2,\varepsilon}) &= 1 - \varepsilon, \\ \mathbf{M}(\chi_{n-1}^2; h_{1,\varepsilon} < \chi_{n-1}^2 < h_{2,\varepsilon}) &= (1 - \varepsilon) \mathbf{M}\chi_{n-1}^2, \\ \chi_{n-1}^2 &\in H_{n-1}. \end{aligned}$$

El conjunto confidencial  $\Theta^*(X, \varepsilon)$  correspondiente a (8) tiene la forma del intervalo

$$(n-1)S_0^2/h_{2,\varepsilon} < \sigma^2 < (n-1)S_0^2/h_{1,\varepsilon}. \quad (9)$$

Este intervalo es, evidentemente, invariante respecto a  $g$ , al igual que el criterio (8) (en este ejemplo  $G|\sigma_1 \equiv G$  para cualquier  $\sigma_1$ ). Por lo tanto, en virtud de las segundas afirmaciones de los teoremas 4 y 5, el intervalo (9) es el conjunto confidencial no desplazado e invariante más exacto de nivel  $1 - \varepsilon$ .

**Ejemplo 3.** Supongamos que  $X \in \Phi_{\alpha, \sigma^2}$ . Es necesario construir el conjunto confidencial más exacto para el parámetro  $\alpha$  cuando se desconoce  $\sigma$ . Aquí

$$f_{\alpha, \sigma^2}(X) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \alpha)^2 \right\}.$$

La familia  $\Phi_{\alpha, \sigma^2}$  será invariante respecto al grupo  $G$  de las transformaciones lineales  $gX = aX + b$  si se pone  $\bar{g}(\alpha, \sigma) = (a\alpha + b, a\sigma)$ . El par de observaciones  $(\bar{x}, S_0^2)$  forma una estadística suficiente. Es fácil ver que con su ayuda no se puede construir una estadística que sea invariante respecto a  $G$ . No obstante, para cada  $\alpha_1$  se puede separar un subgrupo  $G[\alpha_1]$  de transformaciones  $gX = a(X - \alpha_1) + \alpha_1$  respecto al cual la estadística  $(\bar{x} - \alpha_1)/S_0$  será el máximo invariante. La hipótesis  $H_1 = \{\alpha = \alpha_1\}$  queda invariante respecto a  $G[\alpha_1]$ . Investigando la densidad  $(\bar{x} - \alpha_1)/S_0$  se puede

mostrar, con ayuda de los métodos del § 7 (omitimos estas consideraciones puesto que son muy complicadas\*), que para cada  $\sigma$ , el criterio uniformemente más potente no desplazado e invariante para verificar la hipótesis  $H_1$  frente a  $\{\alpha \neq \alpha_1\}$  existe y tiene una región de aceptación de  $H_1$  en forma de

$$\sqrt{n}|\bar{x} - \alpha_1|/S_0 < \tau_\varepsilon, \quad (10)$$

donde  $\tau_\varepsilon$  se determina de la condición  $\mathbf{P}(|t_{n-1}| \geq \tau_\varepsilon) = \varepsilon$ ,  $t_{n-1} \in T_{n-1}$ .

El conjunto confidencial respectivo  $\Theta^*$  tiene la forma

$$\bar{x} - \tau_\varepsilon S_0/\sqrt{n} < \alpha < \bar{x} + \tau_\varepsilon S_0/\sqrt{n}. \quad (11)$$

Es fácil ver que este intervalo confidencial es invariante ( $\Theta^*(gX, \varepsilon) = \bar{g}\Theta^*(X, \varepsilon)$ ). Según la primera afirmación del teorema 5, el criterio (10) será invariante respecto a  $G[\alpha_1]$ . De acuerdo con la segunda afirmación, el intervalo confidencial (11) será el criterio confidencial más exacto (uniformemente respecto a  $\sigma$ ) no desplazado e invariante de nivel  $1 - \varepsilon$ .

Ahora bien, en este párrafo hemos establecido que todos los intervalos confidenciales construidos en el § 2.32 son, en cierto sentido, óptimos.

## § 9. Enfoques bayesiano y minimax de la verificación de las hipótesis compuestas

**1. Criterios bayesianos y minimax.** En el § 4 hemos descrito los enfoques bayesiano y minimax. Allí mismo hemos dado las definiciones respectivas que recordaremos en la exposición posterior.

Supongamos, como antes, que se verifica la hipótesis  $H_1 = \{\theta \in \Theta_1\}$  frente a  $H_2 = \{\theta \in \Theta_2\}$ , basándose en la muestra  $X \in \mathbf{P}_\theta$ .

El *enfoque bayesiano completo* supone que  $\Theta$  se elige al azar con la distribución a priori  $\mathbf{Q}$  en  $\Theta = \Theta_1 \cup \Theta_2$ . La distribución  $\mathbf{Q}$  induce las distribuciones  $\mathbf{Q}_i$  en  $\Theta_i$ ,  $i = 1, 2$  y las probabilidades  $q(i) = \mathbf{Q}(\theta \in \Theta_i)$ , así que  $\mathbf{Q} = q(1)\mathbf{Q}_1 + q(2)\mathbf{Q}_2$ . Designemos por  $H_{Q_i}$  la hipótesis de que  $\theta \in \Theta_i$  se elige al azar, con la distribución  $\mathbf{Q}_i$ . Según esta hipótesis,  $X$  tiene la densidad

$$f_{Q_i}(x) = \int f_\theta(x)\mathbf{Q}_i(dx).$$

Se entiende, por supuesto (véase el § 4), que en  $\Theta_i$  están definidas las  $\sigma$ -álgebras de  $\sigma_b$  a base de las cuales se eligen  $\mathbf{Q}_b$  y que  $f_\theta(x)$  es medible respecto a  $\mathfrak{S}_1 \times \mathfrak{B}^n \mathfrak{X}$ .

De los resultados del § 1, 2 se deduce que el criterio bayesiano  $\pi_{\mathbf{Q}}$  para verificar  $H_{Q_1}$  frente a  $H_{Q_2}$  en el problema descrito anteriormente tendrá

\* Esto se expone más detalladamente en [57], p. 312.

la forma

$$\pi_{Q_1}(X) = \begin{cases} 1, & \text{si } f_{Q_1}(X) > cf_{Q_2}(X), \\ p, & \text{si } f_{Q_1}(X) = cf_{Q_2}(X), \\ 0, & \text{si } f_{Q_1}(X) < cf_{Q_2}(X), \end{cases} \quad (1)$$

donde  $c = q(1)/q(2)$ ,  $p \in [0, 1]$  es arbitrario.

El enfoque parcialmente bayesiano está relacionado con la verificación de la hipótesis  $H_{Q_1}$  frente a  $H_{Q_2}$  en el caso cuando falta la distribución a priori entre  $H_{Q_1}$  y  $H_{Q_2}$  (que se define por las probabilidades  $q(1)$  y  $q(2)$ ). Pongamos

$$K_\varepsilon^{Q_1} = \{ \pi: \mathbf{M}_{Q_1} \pi(X) \leq \varepsilon \}.$$

Entonces el criterio  $\pi_{Q_1, Q_2}$  se llama bayesiano en  $K_\varepsilon^{Q_1}$  si éste es el c.m.p. de nivel  $1 - \varepsilon$  para verificar  $H_{Q_1}$  frente a  $H_{Q_2}$ . El criterio  $\pi_{Q_1, Q_2}$  tendrá la misma forma (1), donde  $c$  y  $p$  se eligen de la condición  $\mathbf{M}_{Q_1} \pi_{Q_1, Q_2}(X) = \varepsilon$ .

En vez de  $\pi_{Q_1, Q_2}$  escribiremos  $\pi_{Q_1}$  y  $\pi_{Q_2}$  si uno de los conjuntos  $\Theta_1$  o  $\Theta_2$  se degenera en conjunto de un punto  $\{\theta_1\}$  o  $\{\theta_2\}$ .

En las aplicaciones rara vez se encuentran problemas en las que las distribuciones  $Q_i$  son completamente conocidas. Sin embargo, ya hemos visto repetidas veces que la utilidad del enfoque bayesiano no se limita exclusivamente a la posibilidad de aplicarlo directamente. Este enfoque permite construir los c.u.m.p., y también los minimax (compárese con los §§ 1, 5 y 6). Posteriormente utilizaremos el enfoque bayesiano también para construir los criterios asintóticamente óptimos. Sea, como antes,

$$K_\varepsilon = \{ \pi: \sup_{\theta \in \Theta_1} \mathbf{M}_\theta \pi(X) \leq \varepsilon \}. \quad (2)$$

Entonces el criterio,  $\bar{\pi}$  se denomina minimax en  $K_\varepsilon$  (en  $K_\varepsilon^{Q_1}$ ) si  $\bar{\pi} \in K_\varepsilon$  ( $\bar{\pi} \in K_\varepsilon^{Q_1}$ ), y para él se minimiza

$$\inf_{\theta \in \Theta_2} \mathbf{M}_\theta \pi(X) = \inf_{\theta \in \Theta_2} \beta(\theta). \quad (3)$$

Cabe señalar que si las funciones de potencia  $\beta(\theta) = \mathbf{M}_\theta \pi(X)$  son continuas y los conjuntos  $\Theta_1$  y  $\Theta_2$  se tocan, entonces

$$\beta = \sup_{\pi \in K_\varepsilon} \inf_{\theta \in \Theta_2} \beta(\theta) \leq \varepsilon \quad (4)$$

y la desigualdad  $\beta > \varepsilon$  no puede cumplirse. Por eso, si se desea que la potencia garantizada (3) sea suficientemente grande (en todo caso, mayor que  $\varepsilon$ ), conviene examinar los conjuntos "separados"  $\Theta_1$  y  $\Theta_2$ . Con otras palabras, es necesario eliminar la zona de los valores de  $\theta$ , donde  $\beta(\theta)$  es próxima a  $\varepsilon$  como zona de "indiferencia" de los criterios, y examinar, en calidad de  $\Theta_2$ , el conjunto que no toca  $\Theta_1$ .

No obstante, si los conjuntos se tocan, *todo criterio no desplazado en  $K_\varepsilon$  será minimax*. En efecto, para los criterios no desplazados  $\beta(\theta) = M_\theta \pi(X) \geq \varepsilon$ ,  $\theta \in \Theta_2$  y, por lo tanto,  $\beta = \inf_{\theta \in \Theta_2} \beta(\theta) \geq \varepsilon$  alcanza, en virtud de (4), su valor máximo.

La afirmación inversa es cierta en el caso general: *el criterio minimax, si existe, no está desplazado*. Esto se desprende del hecho de que

$$\beta = \sup_{\pi \in K_\varepsilon} \inf_{\theta \in \Theta_2} \beta(\theta) \geq \varepsilon$$

(podemos tomar  $\pi(X) \equiv \varepsilon$ ) y del hecho de que para el criterio minimax

$$\inf_{\theta \in \Theta_2} \beta(\theta) = \beta.$$

*El criterio uniformemente más potente no desplazado  $\check{\pi}$  en la clase  $K_\varepsilon$  de todos los criterios no desplazados, es minimax en  $K_\varepsilon$* . En efecto, sea  $\check{\beta}(\theta)$  la función de potencia del criterio  $\check{\pi}$ . Entonces, para cualesquiera  $\pi \in K_\varepsilon$ ,  $\theta \in \Theta_2$ ,

$$\begin{aligned} \check{\beta}(\theta) &\geq \beta(\theta), \quad \inf_{\theta \in \Theta_2} \beta(\theta) \geq \inf_{\theta \in \Theta_2} \check{\beta}(\theta), \\ \inf_{\theta \in \Theta_2} \check{\beta}(\theta) &= \sup_{\pi \in K_\varepsilon} \inf_{\theta \in \Theta_2} \beta(\theta) = \sup_{\pi \in K_\varepsilon} \inf_{\theta \in \Theta_2} \check{\beta}(\theta). \end{aligned} \quad (5)$$

La última igualdad se explica por el hecho de que la adición a  $K_\varepsilon$  de los criterios de  $K_\varepsilon$ , para los cuales  $\inf_{\theta \in \Theta_2} \beta(\theta) < \varepsilon$ , no cambia la magnitud  $\sup_{\pi \in K_\varepsilon}$  en (5).  $\triangleleft$

En el teorema 5.3 hemos utilizado los criterios bayesianos para determinar el c.u.m.p. La siguiente afirmación es cierto "desarrollo" del teorema 5.3. La misma también es el análogo de los teoremas 1.2 y 2.11.2 y establece que los criterios minimax han de buscarse en la clase de criterios (1) cuya forma explícita conocemos.

**Teorema 1.** *Supongamos que existen las distribuciones  $Q_i$  concentradas, respectivamente, en los conjuntos  $\Theta_i^\circ \subset \Theta_i$ ,  $i = 1, 2$ , y las constantes  $c$  y  $p$  tales, que el criterio  $\pi_{Q_1, Q_2}$ , definido en (1), poseen las propiedades*

$$\begin{aligned} 1) \quad &\pi_{Q_1, Q_2} \in K_\varepsilon^{Q_1}, \\ 2) \quad &M_\theta \pi_{Q_1, Q_2}(X) = \sup_{\theta \in \Theta_1} M_\theta \pi_{Q_1, Q_2}(X) \end{aligned} \quad (6)$$

para todos  $\theta \in \Theta_1^\circ$ ,

$$3) \quad M_\theta \pi_{Q_1, Q_2}(X) = \inf_{\theta \in \Theta_2} M_\theta \pi_{Q_1, Q_2}(X) \quad (7)$$

para todos  $\theta \in \Theta_2^\circ$ .

Entonces  $\pi_{Q_1, Q_2} \in K_\varepsilon$  es precisamente el criterio minimax en  $K_\varepsilon$  para verificar  $H_1$  frente a  $H_2$ .

El par de distribuciones  $Q_1$  y  $Q_2$  que posee las propiedades 2) y 3) es el menos favorable en el sentido de que para cualesquiera dos otras distribuciones  $Q'_1$  y  $Q'_2$ ,

$$\inf_{\theta \in \Theta_2} M_\theta \pi_{Q_1, Q_2} \leq \inf_{\theta \in \Theta_2} M_\theta \pi_{Q'_1, Q'_2},$$

donde  $\pi_{Q'_1, Q'_2}$  es el criterio de forma (1) de  $K_\varepsilon$ .

La última afirmación significa que entre todos los criterios bayesianos (1), el criterio  $\pi_{Q_1, Q_2}$  posee la potencia menos garantizada.

**Demostración.** Como

$$\sup_{\theta \in \Theta_1} M_\theta \pi_{Q_1, Q_2}(X) = \int_{\Theta_1} M_\theta \pi_{Q_1, Q_2} Q_1(d\theta) = M_{Q_1} \pi_{Q_1, Q_2} = \varepsilon,$$

entonces  $\pi_{Q_1, Q_2} \in K_\varepsilon$ . La potencia garantizada  $\pi_{Q_1, Q_2}$  es igual a (véase (7))

$$\inf_{\theta \in \Theta_2} M_\theta \pi_{Q_1, Q_2}(X) = \int_{\Theta_1} M_\theta \pi_{Q_1, Q_2} Q_2(d\theta) = M_{Q_2} \pi_{Q_1, Q_2} \equiv \beta_{Q_1, Q_2}. \quad (8)$$

Sea ahora  $\pi$  cualquier otro criterio de  $K_\varepsilon$  para verificar  $H_1$  frente a  $H_2$ . Entonces  $\pi$  será simultáneamente el criterio de  $K_\varepsilon^{Q_1}$  para verificar  $H_{Q_1}$  frente a  $H_{Q_2}$ , ya que

$$M_{Q_1} \pi(X) = \int_{\Theta_1} M_\theta \pi(X) Q_1(d\theta) \leq \sup_{\theta \in \Theta_1} M_\theta \pi(X) \leq \varepsilon. \quad (9)$$

Pero el criterio  $\pi_{Q_1, Q_2}$  es el c.m.p. en  $K_\varepsilon^{Q_1}$  para verificar  $H_{Q_1}$  frente a  $H_{Q_2}$ . Por consiguiente, en virtud de (8),

$$\inf_{\theta \in \Theta_2} M_\theta \pi_{Q_1, Q_2}(X) = \beta_{Q_1, Q_2} \geq M_{Q_1} \pi(X) \geq \inf_{\theta \in \Theta_2} M_\theta \pi(X). \quad (10)$$

La primera afirmación del teorema queda demostrada. Sean ahora  $Q'_1$  y  $Q'_2$  cualesquiera dos otras distribuciones en  $\Theta_1$  y  $\Theta_2$  respectivamente. El criterio  $\pi_{Q'_1, Q'_2}$ , al igual que  $\pi_{Q_1, Q_2}$ , será el criterio de  $K_\varepsilon^{Q'_1}$  para verificar  $H_{Q'_1}$  frente a  $H_{Q'_2}$ , ya que

$$M_{Q'_1} \pi_{Q'_1, Q'_2}(X) = \int_{\Theta_1} M_\theta \pi_{Q'_1, Q'_2}(X) Q'_1(d\theta) \leq \sup_{\theta \in \Theta_1} M_\theta \pi_{Q'_1, Q'_2}(X) \leq \varepsilon.$$

Pero el criterio  $\pi_{Q_1, Q_2}$  es el c.m.p. para estas hipótesis, por eso, en virtud de (8),

$$\begin{aligned} \beta_{Q_1, Q_2} \equiv M_{Q'_2} \pi_{Q_1, Q_2}(X) &\geq M_{Q'_2} \pi_{Q'_1, Q'_2}(X) = \\ &= \int_{\Theta_2} M_\theta \pi_{Q_1, Q_2}(X) Q'_2(d\theta) \geq \inf_{\theta \in \Theta_2} M_\theta \pi_{Q_1, Q_2}(X) = \beta_{Q_1, Q_2}. \quad \triangleleft \end{aligned}$$

La principal dificultad en la aplicación del teorema 1 a los problemas reales consiste en buscar (o adivinar) las distribuciones menos favorables  $Q_1$  y  $Q_2$ . En este caso a veces pueden resultar útiles las consideraciones de invariación, así como ocurre en los ejemplos del apartado siguiente. Estos ejemplos tienen interés autónomo y se utilizarán posteriormente.

## 2. Criterios minimax para el parámetro $\alpha$ de distribuciones normales.

**Ejemplo 1.** Supongamos que  $X = x_1 \in \Phi_{\alpha, E}$  es una muestra de volumen  $n = 1$  de una distribución normal  $m$ -dimensional con media  $\alpha = (\alpha_1, \dots, \alpha_m)$  y con matriz unidad de segundos momentos. Designemos  $|\alpha|^2 = \sum_{i=1}^m \alpha_i^2$  y examinemos el problema de verificación de la hipótesis

$H_1 = \{|\alpha| \leq a\}$  frente a  $H_2 = \{|\alpha| \geq b\}$ ,  $b > a$  (aquí hay una zona "separadora"  $a < |\alpha| < b$ ).

Si, por ejemplo  $X$  determina (en un canal de comunicación) las amplitudes de la señal vectorial compuesta por el "ruido"  $X_0 \in \Phi_{0,1}$  y por la señal útil  $\alpha$ ,  $|\alpha| \geq b$ , las hipótesis  $H_i$  se pueden considerar, para  $a = 0$ , como hipótesis de la presencia de la señal útil.

En vista de que el ejemplo sujeto a examen se utilizará repetidas veces posteriormente, la afirmación referente a la forma del criterio minimax será enunciada en forma de teorema.

**Teorema 2.** El criterio minimax  $\bar{\pi} \in K_\varepsilon$  para verificar  $H_1 = \{|\alpha| \leq a\}$  frente a  $H_2 = \{|\alpha| \geq b\}$ ,  $a < b$ , según la observación  $X \in \Phi_{\alpha, E}$ , tiene la forma

$$\pi(X) = \begin{cases} 1, & \text{si } |X| > c_\varepsilon, \\ 0, & \text{si } |X| \leq c_\varepsilon, \end{cases}$$

donde  $c_\varepsilon$  se elige de la condición  $p_c(a) = \varepsilon$ , la potencia garantizada  $\bar{\pi}$  es igual a  $p_{c_\varepsilon}(b)$ ,

$$p_c(t) = \mathbf{P}((\xi_1 - t)^2 + \xi_2^2 + \dots + \xi_m^2 > c^2),$$

$\xi_i \in \Phi_{0,1}$  son independientes.

**Demostración.** Comencemos por consideraciones sugestivas. En nuestro caso, para  $x = (x^{(1)}, \dots, x^{(m)})$  tenemos

$$f_\alpha(x) = \frac{1}{(2\pi)^{m/2}} \exp \left\{ -\frac{1}{2} (x - \alpha)(x - \alpha)^T \right\},$$

donde  $x^T$  es el vector columna. De aquí se deduce que la familia de distribuciones espuesta a examen es invariante respecto a la transformación ortogonal  $gx = xC$ , donde  $C$  es la matriz de la transformación ortogonal en  $R^m$ . En este caso hay que poner  $\bar{g}\alpha = \alpha C$ . Las hipótesis  $H_i$  serán invariantes respecto a  $\bar{g}$ .

Supongamos, para abreviar, que  $a = 0$ . Si la distribución  $Q_2$  en  $\Theta_2 = \{\alpha: |\alpha| \geq b\}$  no manifestara invariación respecto a  $\bar{g}$  (así sucederá, por ejemplo, cuando la misma se halle concentrada en el entorno de cualquier punto  $\alpha_0$ ), entonces, esta asimetría podría utilizarse, de una u otra manera, para resolver tal problema (con la suposición que acabamos de hacer estaríamos próximos al problema de verificación de dos hipótesis simples

$\{\alpha = 0\}$  y  $\{\alpha = \alpha_0\}$  y en este caso obtendríamos un criterio de gran potencia). Por lo tanto, dicha distribución no puede ser la menos favorable. Esta debe ser la distribución  $Q_2$ , invariante respecto a  $\bar{g}$ . Además, está claro que obtendremos la peor variante si toda la distribución permanece concentrada en la frontera  $\Theta_2$  (cuanto más semejantes sean las hipótesis, tanto más difícil será distinguir las). Se pueden citar razonamientos sugestivos análogos respecto a  $Q_1$ , si  $a \neq 0$ .

Así pues, es natural que en nuestro ejemplo las distribuciones menos favorables  $Q_1$  y  $Q_2$  sean distribuciones uniformes en las esferas  $\Theta_1 = \{\alpha: |\alpha| = a\}$  y  $\Theta_2^\circ = \{\alpha: |\alpha| = b\}$ . En este caso, de acuerdo con el teorema 1 el criterio minimax  $\bar{\pi}$  tendrá la forma  $\bar{\pi}(x) = \pi_{Q_1, Q_2}(x)$ , donde  $\pi_{Q_1, Q_2}(x) = 1$  si

$$\int_{\Theta_2} \exp \left\{ -\frac{1}{2} (x - v)(x - v)^T \right\} \frac{dV(v)}{V_2} > c \int_{\Theta_1} \exp \left\{ -\frac{1}{2} (x - v)(x - v)^T \right\} \frac{dV(v)}{V_1} \quad (11)$$

y  $\pi_{Q_1, Q_2}(x) = 0$  en el caso contrario. Aquí  $dV(v)$  significa el área del elemento de la esfera correspondiente,  $V_i = \text{mes } \Theta_i$ ,  $i = 1, 2$ .

Examinemos cualquiera de estas integrales, por ejemplo, la derecha, y notemos que ésta puede ser escrita en la forma

$$\exp \left\{ -\frac{1}{2} x x^T - a^2 \right\} \cdot \int_{\Theta_1} \exp \{x v^T\} \frac{dV(v)}{V_1}.$$

Aquí la integral es igual a

$$\int_{\Theta^\circ} \exp \{ |x| a e_x v^T \} dV(v) / V, \quad V = \text{mes } \Theta^\circ,$$

donde  $\Theta^\circ$  es la superficie de una esfera unitaria,  $e_x = x/|x|$ . Por consiguiente, si designamos

$$\psi(t) = \int_{\Theta^\circ} \exp \{ t e_x v^T \} dV(v), \quad (12)$$

entonces, la región (11) de aceptación de  $H_2$  tendrá la forma

$$\psi(|x|b) > c \psi(|x|a) \quad (13)$$

(aquí, por  $c$  designamos las constantes que no coinciden obligatoriamente con el valor en (11)). Pero, evidentemente,  $\psi(t)$  no depende de  $x$ , puesto que el valor de la integral (12) no depende del sentido de dirección del vector unitario  $e_x$ . Por eso

$$\psi(t) = \int_{\Theta^\circ} \exp \{ t v_1 \} dV(v),$$

donde  $v_1$  es la primera coordenada del vector  $v$ .



Como  $\psi'(0) = 0$ ,  $\psi''(t) > 0$  cuando  $t > 0$ , entonces  $\psi(t)$  es una función convexa creciente en  $[0, \infty)$ . De aquí resulta que la desigualdad (13) u (11) equivale a

$$|x| > c. \quad (14)$$

Esto es, evidentemente, un criterio invariante. Comprobemos para él el cumplimiento de las condiciones 1—3 del teorema 1 y establezcamos asimismo que ello es el criterio minimax.

Tenemos

$$M_{\alpha} \pi_{Q_1, Q_2}(X) = P_{\alpha}(|X| > c) = \Phi_{0, E}(\{x: |x - a| > c\}).$$

Está claro que el traslado del punto  $\alpha$  en la esfera  $|\alpha| = \text{const}$  no modifica dicha probabilidad. Por lo tanto, esta última sólo depende de  $|\alpha|$  y, por consiguiente,

$$\begin{aligned} M_{\alpha} \pi_{Q_1, Q_2} &= P(|\xi - \alpha|^2 > c^2) = \\ &= P\left(\sum_{i=1}^m (\xi_i - \alpha_i)^2 > c^2\right) = P((\xi_1 - |\alpha|)^2 + \xi_2^2 + \dots + \xi_m^2 > c^2), \end{aligned}$$

donde  $\xi_i \in \Phi_{0,1}$  son las coordenadas independientes del vector  $\xi$ .

**Lema 1.** La función  $p_c(t) = P((\xi_1 - t)^2 + \xi_2^2 + \dots + \xi_m^2 > c^2)$  es para cada  $c$  la función creciente  $|t|$ .

De este lema se desprende que

$$M_{\alpha} \pi_{Q_1, Q_2}(X) = p_c(|\alpha|) \leq p_c(a) \text{ cuando } |\alpha| \leq a,$$

$$M_{\alpha} \pi_{Q_1, Q_2}(X) = p_c(|\alpha|) \geq p_c(b) \text{ cuando } |\alpha| \geq b.$$

Estas relaciones equivalen a las condiciones 2) y 3) del teorema 1. Para que el criterio  $\pi_{Q_1, Q_2}$  sea el criterio de nivel  $1 - \varepsilon$ , debemos suponer que  $c$  es igual a la solución  $c_{\varepsilon}$  de la ecuación  $p_c(a) = \varepsilon$ . Ahora bien,  $\pi_{Q_1, Q_2}$  es el criterio minimax de nivel  $1 - \varepsilon$  y su potencia garantizada es igual a  $p_{c_{\varepsilon}}(b)$ .  $\triangleleft$

**Demostración del lema 1.** Como  $p_c(t) = p_c(-t)$ , podemos limitarnos a examinar los valores de  $t \geq 0$ .

Examinemos primeramente el caso de  $m = 1$ . Designemos en este caso la función  $p_c(t)$  por  $p(t)$ . Tenemos

$$p(t) = P(|\xi_1 - t|^2 > c^2) = \Phi(t - c) + 1 - \Phi(t + c).$$

Por consiguiente, la derivada respecto a  $t$  es igual a

$$\begin{aligned} p'(t) &= \frac{1}{\sqrt{2\pi}} [e^{-(t-c)^2/2} - e^{-(t+c)^2/2}] = \\ &= \frac{1}{\sqrt{2\pi}} e^{-(c^2+t^2)/2} [e^{ct} - e^{-ct}] \geq 0 \end{aligned}$$

y la función  $p(t)$  crece cuando  $t \geq 0$ .

Cuando  $m > 1$  la función  $p_c(t)$  es la convolución de la función  $p(t) = p(t, c^2)$  con la distribución  $\chi^2$  de  $m - 1$  grados de libertad:

$$p_c(t) = \int_0^{\infty} p(t, c^2 - u) dH_{m-1}(u).$$

Evidentemente, ésta también es una función creciente de  $t$  para  $t \geq 0$ .  $\triangleleft$

En lo que se refiere al teorema 2 se puede señalar lo siguiente. Supongamos, para abreviar, que  $a = 0$ . Entonces, la hipótesis  $H_1 = \{\alpha = 0\}$  será simple. Si construimos el c.m.p. para cada alternativa  $\alpha \in \Theta_2$ , obtendremos el criterio que tiene la forma

$$x_\alpha^T > c.$$

Esto significa que cada sentido de dirección de  $\alpha = \alpha_0 t$ ,  $\alpha_0 \in \Theta_2^0$ ,  $t \geq 1$  tendrá su propio criterio más potente de nivel  $1 - \varepsilon$

$$x\alpha_0^T > c_\varepsilon, \quad (15)$$

donde  $c_\varepsilon$  depende únicamente de  $\varepsilon$ , ya que  $\mathbf{M}_0(X\alpha_0^T) = 0$ .  $\mathbf{D}_0(X\alpha_0^T) = |\alpha_0|^2 = b$ . Pero la región crítica del criterio minimax (invariante) debe ser igualmente sensible respecto a todas las alternativas. En concordancia con esto, la misma tiene forma de unión de los semiespacios (15), que no es otra cosa sino el exterior de la esfera.

**Ejemplo 2.** Ahora supongamos que  $X = x_1 \in \Phi_{\alpha, \sigma^2}$ , donde  $\sigma^2 = \|\sigma_{ij}\|$  es una matriz arbitraria de segundos momentos, definida positivamente. Examinemos el problema de verificación de la hipótesis  $H_1 = \{\alpha\sigma^{-2}\alpha^T \leq a^2\} = \{|\alpha\sigma^{-1}| \leq a\}$  frente a  $H_2 = \{\alpha\sigma^{-2}\alpha^T \geq b^2\} = \{|\alpha\sigma^{-1}| \geq b\}$ ,  $a < b$ . Del teorema 2 se deduce el

**Teorema 2A.** *El conjunto crítico del criterio minimax de nivel  $1 - \varepsilon$  para verificar  $H_1$  frente a  $H_2$  tiene la forma*

$$x\sigma^{-2}x^T > c_\varepsilon^2$$

y la potencia garantizada  $p_{c_\varepsilon}(b)$ , donde  $c_\varepsilon$  es, como antes, la solución de la ecuación  $p_c(a) = \varepsilon$ .

**Demostración.** Pongamos  $gx = x\sigma$  y notemos que, en virtud de (7.3),

$$\Phi_{\alpha, E}(A) = \Phi_{\bar{g}(\alpha, E)}(gA),$$

donde  $\bar{g}(\alpha, E) = (\alpha, \sigma^2)$ . Para la esfera  $A = \{x: |x| < c\}$  tendremos

$$gA = \{y = x\sigma: xx^T < c^2\} = \{y: y\sigma^{-2}y^T < c^2\}, \quad (16)$$

$$\Phi_{\alpha, E}(A) = \Phi_{\alpha, \sigma^2}(\{x: x\sigma^{-2}x^T < c^2\}).$$

El conjunto  $\{\alpha: |\alpha| \leq a\}$  pasa, después de la transformación  $\bar{g}$ , al conjunto  $\{\beta = \alpha\sigma: \alpha\alpha^T \leq a^2\} = \{\beta: \beta\sigma^{-2}\beta^T \leq a^2\}$ .

Ahora bien, todas las relaciones establecidas en el ejemplo 1 para  $\Phi_{\alpha, E}(A)$  cuando  $|\alpha| \leq a$  o cuando  $|\alpha| \geq b$  serán válidas para  $\Phi_{\beta, \sigma^2}(\{x: x\sigma^{-2}x^T < c^2\})$  cuando  $|\beta\sigma^{-1}| \leq a$  o bien  $|\beta\sigma^{-1}| \geq b$ , respectivamente.

Esto demuestra el teorema 2A.  $\triangleleft$

**Ejemplo 3.** Volvamos a examinar la muestra de la distribución normal  $\Phi_{\alpha, E}$  con una matriz unidad de segundos momentos. Sin embargo, a distinción del ejemplo 1, las hipótesis  $H_i$  sometidas a comprobación sólo tocarán una parte de las coordenadas del vector  $\alpha$ . Representemos  $\alpha$  en forma de un conjunto de dos vectores  $\alpha = (\alpha', \alpha'')$ , donde  $\alpha' = (\alpha_1, \dots, \alpha_l)$ ,  $\alpha'' = (\alpha_{l+1}, \dots, \alpha_m)$ , y examinemos el problema de verificación de la hipótesis  $H_1 = \{|\alpha''| \leq a\}$  frente a  $H_2 = \{|\alpha''| \geq b\}$ , conforme a la muestra  $X = x_1 = (x_{1,1}, \dots, x_{1,m})$  de volumen  $n = 1$ . Para cada una de las hipótesis, la magnitud  $\alpha'$  puede adoptar un valor arbitrario. Procedamos del mismo modo que en el ejemplo 1, pero en calidad de  $Q_1$  y  $Q_2$  escojamos las distribuciones uniformes en las "esferas"  $\Theta_1^0 = \{\alpha: |\alpha''| = a, \alpha' = \alpha'_0\}$ ,  $\Theta_2^0 = \{\alpha: |\alpha''| = b, \alpha' = \alpha'_0\}$ , donde  $\alpha'_0$  es un punto registrado cualquiera. Si designamos  $x_1' = (x_{1,1}, \dots, x_{1,l})$ ,  $x_1'' = (x_{1,l+1}, \dots, x_{1,m})$ , obtendremos como resultado el criterio minimax

$$|x_1''| > c_\varepsilon,$$

donde  $c_\varepsilon$  es la solución de la ecuación

$$P((\xi_1 - a)^2 + \xi_2^2 + \dots + \xi_{m-l}^2 > c^2) = \varepsilon \quad (17)$$

(los factores  $\exp\left\{-\frac{1}{2}(x' - \alpha'_0)(x' - \alpha'_0)^T\right\}$  en la desigualdad  $f_{Q_1}(X) + f_{Q_2}(X) > c$  serán eliminados, y ésta se convertirá en una igualdad del tipo (11)). Este resultado es completamente natural, ya que en nuestro caso las coordenadas  $x_{j,j}$  son independientes y, por lo tanto, el subvector  $x_1'$  no lleva en sí ninguna información respecto a  $\alpha''$ . Por eso, de toda la muestra  $X = x_1$  sólo es suficiente examinar el subvector  $x_1''$  y, en este caso, el problema se reduce al ejemplo 1.

La verificación de las hipótesis en el ejemplo 3 pertenece a la clase de problemas en que existe el llamado parámetro "obstaculizador". En nuestro caso, en calidad de tal parámetro servía el vector  $\alpha'$ . En virtud de las causas mencionadas anteriormente, éste en realidad no obstaculizaba la construcción del criterio minimax, el cual automáticamente resultaba independiente de  $\alpha'$ .

De manera algo diferente ocurre en el ejemplo siguiente, más general, cuando las coordenadas  $x_{ij}$  son dependientes.

**Ejemplo 4.** Supongamos que  $X = x_1 \in \Phi_{\alpha, \sigma^2}$ . Examinemos el problema de verificación de la hipótesis

$$H_1 = \{\alpha d^{-2} \alpha^T \leq a^2\} \text{ frente a } H_2 = \{\alpha d^{-2} \alpha^T \geq b^2\}, \quad (18)$$

donde  $d^{-2}$  es una matriz definida no negativamente de rango  $m - l < m$ , obtenida de  $\sigma^{-2}$  a base de sustituir por ceros los elementos de cualesquiera  $l$  renglones y  $l$  columnas (con los mismos números de orden). Para facilitar la exposición podemos considerar que, para la matriz definida positivamente  $\sigma_2^{-2}$  de orden  $(m - l) \times (m - l)$ , inversa a la matriz

$$\sigma_2^2 = \mathbf{M}_{\alpha, \sigma^2} (x_1'' - \alpha'')^T (x_1'' - \alpha''),$$

formada por las últimas  $m - l$  columnas y renglones de la matriz  $\sigma^2 = \|\sigma_{ij}\|$ , se verifica la hipótesis  $H_1 = \{\alpha'' \sigma_2^{-2} \alpha''^T \leq a^2\}$  frente a  $H_2 = \{\alpha'' \sigma_2^{-2} \alpha''^T \geq b^2\}$ , donde  $x_1''$ ,  $\alpha''$  designan, al igual que en el ejemplo anterior, los mismos subvectores de los vectores  $x_1$  y  $\alpha$ . En cada una de las hipótesis  $H_i$ , el parámetro obstaculizador  $\alpha'$  puede ser arbitrario.

Hablando en general, en este ejemplo, la distribución de  $x_1'$  depende de  $\alpha''$ . Hagamos la siguiente transformación para convertir  $x_1$  en vector con coordenadas "ortonormalizadas". Pongamos

$$Y = x_1 \Lambda, \quad (19)$$

donde  $\Lambda = \|a_{ij}\|$  es una matriz triangular con elementos  $a_{ij} = 0 \ j > i$ . Los restantes elementos se eligen de la condición  $y \in \Phi_{\beta, E}$ , donde  $\beta = (\beta_1, \dots, \beta_m) = \alpha \Lambda$ . Esto siempre se puede hacer, ya que de (19) obtenemos

$$\begin{aligned} y_m &= x_{1,m} a_{m,m}, \\ y_{m-1} &= x_{1,m} a_{m,m-1} + x_{1,m-1} a_{m-1,m-1}, \\ &\dots \end{aligned}$$

De aquí y de las condiciones

$$\begin{aligned} \mathbf{M}_{\alpha, \sigma^2} (y_i - \beta_i)^2 &= 1, \\ \mathbf{M}_{\alpha, \sigma^2} (y_i - \beta_i)(y_j - \beta_j) &= 0, \quad i \neq j, \end{aligned}$$

se determinan uno tras otro los valores

$$\begin{aligned} a_{m,m}^2 &= 1/\sigma_{m,m}, \\ \sigma_{m,m} a_{m,m-1} + \sigma_{m-1,m} a_{m-1,m-1} &= 0, \\ \sigma_{m,m} a_{m,m-1} + 2\sigma_{m,m-1} a_{m,m-1} a_{m-1,m-1} + \sigma_{m-1,m-1} a_{m-1,m-1}^2 &= 1, \\ &\dots \end{aligned}$$

Ahora bien, la matriz triangular  $\Lambda$  es tal, que

$$\mathbf{M}_{\alpha, \sigma^2} (y - \beta)^T (y - \beta) = \mathbf{M}_{\alpha, \sigma^2} \Lambda^T (x_1 - \alpha)^T (x_1 - \alpha) \Lambda = \Lambda^T \sigma^2 \Lambda = E.$$

Del carácter triangular de  $\Lambda$  se deduce que el vector  $\beta'' = (\beta_{l+1}, \dots, \beta_m)$

depende únicamente de  $\alpha''$ , y al contrario. Si designamos por  $\Lambda_2$  la matriz triangular de orden  $(m-l) \times (m-l)$ , obtenida de los últimos  $m-l$  renglones y columnas de la matriz  $\Lambda$  entonces, obtenemos, evidentemente  $\beta'' = \alpha'' \Lambda_2$ ,  $\Lambda_2^T \sigma_2^2 \Lambda_2 = E$ . El conjunto  $\Theta_1 = \{\alpha: \alpha'' \sigma_2^{-2} \alpha''^T \leq a^2\}$  se convertirá en el conjunto

$$\{\beta: \beta = \alpha \Lambda, \alpha'' \sigma_2^{-2} \alpha''^T \leq a^2\} = \{\beta: \beta'' \Lambda_2^{-1} \sigma_2^{-2} \Lambda_2^{-1T} \beta''^T \leq a^2\} = \\ = \{\beta: \beta'' \beta''^T \leq a^2\} = \{\beta: |\beta| \leq a\}.$$

El "subparámetro"  $\beta'$  puede ser arbitrario si es arbitrario  $\alpha'$ .

Hemos llegado al problema del ejemplo 3. El criterio minimax de nivel  $1 - \varepsilon$  para verificar  $H_1$  frente a  $H_2$  tiene, por consiguiente, la forma  $y'' y''^T > c_\varepsilon$  o bien  $(\Lambda_2 \Lambda_2^T = \sigma_2^{-2})$

$$x_1'' \sigma_2^{-2} x_1''^T > c_\varepsilon,$$

donde  $c_\varepsilon$  es la solución de la ecuación (17).

El último ejemplo es el más general entre los ejemplos 1—4. El mismo resume el contenido de estos ejemplos de la manera siguiente.

**Teorema 2B.** Si a base de la muestra  $X = x_1 \in \Phi_{\alpha, \sigma^2}$  se verifican las hipótesis (18) relacionadas con el valor  $\alpha, d^{-2} \alpha^T$ , entonces el criterio minimax de nivel  $1 - \varepsilon$  tendrá la forma

$$x_1 d^{-2} x_1^T > c_\varepsilon, \quad (20)$$

donde  $c_\varepsilon$  se define en (17), y  $m-l$  es el rango  $d^{-2}$ .

La potencia garantizada del criterio (20) es igual a

$$P((\xi_1 - b)^2 + \xi_2 + \dots + y_{m-1} > c_\varepsilon^2, \xi_i \in \Phi_{0,1}).$$

Si la muestra  $X$  tiene volumen  $n$ , entonces  $\bar{x} \in \Phi_{\alpha, \sigma^2/n}$  tendrá la forma

$$\bar{x} d^{-2} \bar{x}^T > c_\varepsilon/n.$$

El siguiente ejemplo tiene, en cierta medida, otro carácter.

**Ejemplo 5.** Supongamos, al igual que en el ejemplo 1, que  $X = x_1 \in \Phi_{\alpha, E}$  es una muestra de volumen  $n = 1$  de una distribución normal  $m$ -dimensional de media  $\alpha = (\alpha_1, \dots, \alpha_m)$ . Supongamos también, que  $H_1 = \{\alpha = 0\}$  y que la hipótesis  $H_2$  consiste en que  $\alpha$  pertenece a cierto conjunto  $\Theta_2$  que no contiene los puntos  $\alpha \in \Theta_2$ . Designemos por  $\bar{\Theta}_2$  la clausura convexa del conjunto  $\Theta_2$  (conjunto cerrado convexo mínimo que contiene  $\Theta_2$ ), y sea  $\beta$  el punto de  $\bar{\Theta}_2$  más próximo al origen de coordenadas. Entonces, si  $\beta \in \Theta_2$ , la distribución  $Q_2$  concentrada en el punto  $\beta$  será la menos favorable, y el criterio minimax  $\bar{\pi}$  tendrá la forma  $\bar{\pi}(X) = 1$  si

$$(X - \beta)(X - \beta)^T < XX^T + c_1$$

o bien, que es lo mismo, si

$$X\beta^T/|\beta| > c_2,$$

donde  $c_2$  se elige de la condición  $\bar{\pi} \in K_\varepsilon$ .

En efecto, es suficiente comprobar la condición (7). Tenemos

$$\mathbf{M}_\alpha \bar{\pi}(X) = \mathbf{P}_\alpha(X\beta^T/|\beta| > c_2),$$

donde  $X\beta^T/|\beta| \in \Phi_{\alpha\beta^T/|\beta|, 1}$ , así que

$$\mathbf{M}_\alpha \bar{\pi}(X) = 1 - \Phi(c_2 - \alpha\beta^T/|\beta|).$$

Esto significa que el mínimo  $\mathbf{M}_\alpha \bar{\pi}(X)$ ,  $\alpha \in \Theta_2$  se alcanza para  $\alpha$  que minimiza la función  $\alpha\beta^T/|\beta|$ . Pero es evidente que  $\alpha\beta^T \geq \beta\beta^T = |\beta|^2$  para todos  $\alpha \in \Theta_2$ , así que

$$\mathbf{M}_\beta \bar{\pi}(X) = \inf_{\alpha \in \Theta_2} \mathbf{M}_\alpha \bar{\pi}(X). \quad \triangleleft$$

Le proponemos al lector que construya el criterio minimax conforme a ese mismo problema, es decir, cuando  $X \in \Phi_{\alpha, \sigma^2}$ ,  $\sigma^2$  es una matriz arbitraria de segundos momentos.

**3. Distribuciones degeneradas menos favorables para las hipótesis unilaterales.** Supongamos que  $X \in \mathbf{P}_\theta$ , donde  $\theta$  y los elementos  $x_i$  de la muestra  $X$  son reales.

Supongamos además, que verificamos la hipótesis unilateral  $H_1 = \{\theta \leq \theta_1\}$  frente a  $H_2 = \{\theta \geq \theta_2\}$  siempre que haya una "zona de indiferencia" no vacía  $\theta_1 < \theta < \theta_2$ . ¿A qué condiciones las distribuciones menos favorables quedarán concentradas en los puntos  $\theta_1$  y  $\theta_2$ ? Pues en este caso el criterio minimax  $\bar{\pi}$  de nivel  $1 - \varepsilon$  tendría una forma muy simple:

$$\bar{\pi}(X) = \begin{cases} 1, & \text{si } f_{\theta_2}(X) > cf_{\theta_1}(X) \\ p, & \text{si } f_{\theta_2}(x) = cf_{\theta_1}(X), \\ 0, & \text{si } f_{\theta_2}(X) < cf_{\theta_1}(X), \end{cases} \quad (21)$$

donde  $p$  y  $c$  se definen por la igualdad  $\mathbf{M}_{\theta_1} \bar{\pi}(X) = \varepsilon$ .

Ya sabemos que si la relación de verosimilitud es monótona, tal criterio será el c.u.m.p. y, por consiguiente, también será minimax. La siguiente afirmación ofrece otra condición suficiente para que el criterio sea minimax.

**Teorema 3.** *Supongamos que la densidad  $f_\theta(x)$  posee la propiedad de que la relación  $f_{\theta'}(x)/f_\theta(x)$  no decrece respecto a  $x$  para cualesquiera  $\theta' > \theta$ . Entonces las distribuciones  $\mathbf{Q}_1$  y  $\mathbf{Q}_2$  menos favorables estarán concentradas en los puntos  $\theta_1$  y  $\theta_2$ , respectivamente, y, por lo tanto, el criterio (21) será minimax.*

**Demostración.** Supongamos primeramente que  $n = 1$ . Según las condiciones del teorema, habrá  $a \leq b$  tales, que  $f_{\theta'}(x)/f_{\theta}(x) \leq 1$  cuando  $x \in (-\infty, a]$ ,  $f_{\theta'}(x)/f_{\theta}(x) = 1$  cuando  $x \in (a, b)$  y  $f_{\theta'}(x)/f_{\theta}(x) \geq 1$  cuando  $x \in [b, \infty)$ . Como  $\bar{\pi}(x)$  no decrece, entonces  $\bar{\pi}(b) \geq \bar{\pi}(a)$  y

$$\begin{aligned} M_{\theta'} \bar{\pi}(X) - M_{\theta} \bar{\pi}(X) &\geq \\ &\geq \bar{\pi}(a) \int_{-\infty}^a (f_{\theta'}(x) - f_{\theta}(x)) \mu(dx) + \bar{\pi}(b) \int_b^{\infty} (f_{\theta'}(x) - f_{\theta}(x)) \mu(dx) = \\ &= (\bar{\pi}(b) - \bar{\pi}(a)) \int_b^{\infty} (f_{\theta'}(x) - f_{\theta}(x)) \mu(dx) \geq 0. \end{aligned}$$

Si  $n > 1$ , para obtener esta misma desigualdad es necesario valerse de la integración sucesiva (primero respecto a  $x_1$ , luego respecto a  $x_2$ , etc.) y del hecho de que  $\bar{\pi}(X)$  no decrece con arreglo a cada uno de sus argumentos.

Ahora bien, hemos establecido que la potencia  $\beta(\theta) = M_{\theta} \pi(X)$  es una función no decreciente.

De aquí se deduce que el nivel de  $\bar{\pi}$  es igual a  $1 - \varepsilon$  y que  $\beta(\theta_1) = \sup_{\theta < \theta_1} \beta(\theta)$  y  $\beta(\theta_2) = \inf_{\theta > \theta_2} \beta(\theta)$ . Esto significa que se cumplen todas las condiciones del teorema 1. El teorema 3 queda demostrado.  $\triangleleft$

Si  $\theta$  es el parámetro de desplazamiento:  $f_{\theta}(x) = f(x - \theta)$ , se puede mostrar que  $f_{\theta'}(x)/f_{\theta}(x)$  será monótona respecto a  $x$  si y sólo si la función  $-\ln f(x)$  es convexa (véase [57]).

## § 10. Criterio de la relación de verosimilitud

En los párrafos anteriores hemos obtenido varios resultados concernientes a la construcción de todo género de criterios óptimos. Una deducción importante que se puede sacar de las consideraciones citadas consiste en que estos criterios óptimos sólo existen en condiciones bastante limitadas. En la teoría de la estimación hemos tenido, aproximadamente, la misma situación: las estimaciones eficientes también existen únicamente en condiciones limitadas. No obstante, en el capítulo 2 hemos visto que si se examina no la propiedad exacta de eficacia, sino la propiedad asintótica, entonces las estimaciones que poseen esta propiedad ya existen muy a menudo en condiciones relativamente amplias, relacionadas casi siempre con la regularidad de la familia  $\{P_{\theta}\}$ . Tales condiciones son las e.v.m.

Otra expresión de la optimización asintótica de la e.v.m. consiste como hemos visto, en que las e.v.m. son asintóticamente equivalentes a las estimaciones bayesianas para cualquier distribución a priori suave registrada.

En la teoría de verificación de las hipótesis, cierto análogo de la e.v.m. es el llamado *criterio de la relación de verosimilitud* (c.r.v.). En caso de

amplias suposiciones, el referido criterio coincide con los criterios óptimos, si tales existen, y resulta asintóticamente equivalente al criterio bayesiano cuando  $\Theta_1 = \{\theta_1\}$  para cualquier distribución a priori suave registrada  $Q_2$  en  $\Theta_2$ . Esta propiedad y una serie de otras propiedades asintóticas del c.r.v. serán establecidas en los párrafos inmediatos.

Demos la definición del c.r.v. Supongamos que en el caso paramétrico, cuando  $X \in P_\theta$ , se verifica la hipótesis  $H_1 = \{\theta \in \Theta_1\}$  frente a la hipótesis  $H_2 = \{\theta \in \Theta_2\}$ .

**Definición 1.** El criterio  $\hat{\pi}(X)$  con la región crítica

$$R(X) \equiv \frac{\sup_{\theta \in \Theta_2} f_\theta(X)}{\sup_{\theta \in \Theta_1} f_\theta(X)} > c \quad (1)$$

se llama *criterio de la relación de verosimilitud* (c.r.v.) para verificar la hipótesis  $H_1$  frente a  $H_2$ .

La constante  $c$  suele elegirse de la condición

$$\sup_{\theta \in \Theta_1} P_\theta(R(X) > c) = \varepsilon, \quad (2)$$

para la cual el c.r.v. tendrá un nivel de  $1 - \varepsilon$ .

A la par con el criterio (1) a menudo se examina un criterio que, de hecho, equivale al primero (también llamado c.r.v) y que tiene la forma siguiente:

$$R_1(X) \equiv \frac{\sup_{\theta \in \Theta} f_\theta(X)}{\sup_{\theta \in \Theta_1} f_\theta(X)} = \frac{f_{\hat{\theta}^*}(X)}{\sup_{\theta \in \Theta_1} f_\theta(X)} > c. \quad (3)$$

La semejanza de estos criterios se desprende del hecho de que cuando  $\Theta = \Theta_1 \cup \Theta_2$ ,

$$f_{\hat{\theta}^*}(X) = \max\left\{\sup_{\theta \in \Theta_1} f_\theta(X), \sup_{\theta \in \Theta_2} f_\theta(X)\right\}$$

y, por lo tanto,  $R_1(X) = \max\{1, R(X)\}$ .

Si la hipótesis  $H_1$  es simple:  $\Theta_1 = \{\theta_1\}$ ,  $H_2 = \{\theta \neq \theta_1\}$ , así que  $\Theta_2 = \Theta \setminus \{\theta_1\}$ , entonces para  $f_\theta(x)$ , continuas respecto a  $\theta$ , tendremos

$$R(X) = R_1(X) = f_{\hat{\theta}^*}(x)/f_{\theta_1}(X).$$

Según su forma, el criterio (1) generaliza de un modo natural el c.m.p. para verificar las hipótesis simples en el lema de Neumann—Pearson. Y aunque en el caso general este criterio no tiene, por lo visto, *exactas* propiedades de optimización, a menudo resulta ser el mejor asintóticamente (véanse los §§ 13—16).

Muchos criterios invariantes y minimax no desplazados, examinados más arriba, son los c.r.v. En calidad de ilustración examinemos los ejemplos



9.1—9.4 donde se construyeron los criterios minimax para el parámetro  $\alpha$  de poblaciones normales. En todos estos ejemplos, los criterios minimax son los c.r.v. Demostremoslo. Los problemas de los ejemplos 9.2 y 9.4 se han reducido, con una exactitud de hasta las transformaciones lineales del parámetro, a los problemas de los ejemplos 9.1 y 9.3. En vista de que la relación de verosimilitud (1) no depende de tales sustituciones (al variar respectivamente las regiones  $\Theta_i$ ), es suficiente examinar tan sólo los ejemplos 9.1 y 9.3.

En el ejemplo 9.1, a base de una muestra  $X \in \Phi_{\alpha, E}$  de volumen unitario y procedente de una población normal multidimensional con una matriz unidad  $E$  de segundos momentos, hemos verificado la hipótesis  $H_1 = \{|\alpha| \leq a\}$  frente a  $H_2 = \{|\alpha| \geq b\}$ ,  $a < b$ . Resultó que el criterio minimax tiene la forma

$$|X| > c. \quad (4)$$

En nuestro caso,  $\sup_{\theta \in \Theta_1} f_{\theta}(X)$  se define por el valor

$$\inf_{\alpha \in \Theta_1} (X - \alpha)(X - \alpha)^T = \inf_{\theta \in \Theta_2} |X - \alpha|^2,$$

así que para la estadística  $R(X)$  en (1) tendremos

$$\ln R(X) = \begin{cases} -\frac{1}{2} (|X| - b)^2, & \text{si } |X| \leq a, \\ -\frac{1}{2} (|X| - b)^2 + \frac{1}{2} (|X| - a)^2, & \text{si } a < |X| < b, \\ -\frac{1}{2} (|X| - a)^2, & \text{si } |X| \geq b. \end{cases}$$

Esta es una función creciente continua de  $|X|$ . Por eso las regiones (1) y (4) coinciden para valores convenientes de  $c$ .

Le proponemos al lector que él mismo se cerciore de que en este ejemplo el criterio (3) también tiene la forma (4).

En el ejemplo 9.3, a base de la muestra  $X \in \Phi_{\alpha, E}$  de volumen unitario, hemos verificado la hipótesis  $H_1 = \{|\alpha''| \leq a\}$  frente a  $H_2 = \{|\alpha''| \geq b\}$ , donde  $\alpha'' = (\alpha_{l+1}, \dots, \alpha_m)$  es un subvector del vector  $\alpha$  constituido por sus últimas  $m - l$  coordenadas. El criterio minimax tiene la forma

$$|X''| > c, \quad (5)$$

donde  $X''$  está constituido por las últimas  $m - l$  coordenadas del vector  $X$ . Pero en este caso

$$\inf_{\alpha \in \Theta_1} (X - \alpha)(X - \alpha)^T = \inf_{\alpha'' : |\alpha''| \leq a} (X'' - \alpha'')(X'' - \alpha'')^T.$$

La desigualdad análoga es válida para  $\Theta_2$ . Por eso todo se reduce a las consideraciones del ejemplo 9.1, y los c.r.v. (1) y (3) coincidirán con (5).

En condiciones del § 5, los c.u.m.p. allí construidos para las familias exponenciales

$$f_{\theta}(x) = c(\theta)e^{\theta T(x)}h(x) \quad (6)$$

también coincidirán con los c.r.v. El lector puede comprobar esto personalmente, notando que la función

$$\varphi(\theta) = \ln c(\theta) = -\ln \left( \int e^{\theta T(x)}h(x)\mu^n(dx) \right)$$

es convexa, puesto que  $\varphi'(\theta) = -\mathbf{M}_{\theta}T$ ,  $\varphi''(\theta) = -\mathbf{D}_{\theta}T < 0$ . De la convexidad de  $\varphi$  se deduce la solubilidad unívoca de la ecuación

$$\varphi'(\theta) + T(X) = 0$$

para la c.v.m.  $\hat{\theta}^* = \psi(T)$  y la monotonía de la función  $\varphi$ . En este caso, uno de los  $\sup_{\theta \in \Theta_i} f_{\theta}(X)$  se alcanzará en el punto  $\hat{\theta}^*$ , y el otro, en los puntos  $\theta_1$  o  $\theta_2$ .

La verificación de la referida afirmación para las familias normales  $\Phi_{\alpha, E}$ , que son un caso particular de (6), se expone en el § 15.

Es algo diferente el asunto examinado en el ejemplo 9.5, donde, de acuerdo con la muestra  $X \in \Phi_{\alpha, E}$  hemos verificado la hipótesis  $H_1 = \{\alpha = 0\}$  frente a  $H_2 = \{\alpha \in \Theta_2\}$ . Se supone que el conjunto  $\Theta_2$  y su clausura convexa  $\bar{\Theta}_2$  no contienen puntos  $\alpha = 0$ . Si el punto  $\beta$  más próximo al origen de coordenadas del conjunto  $\bar{\Theta}_2$  pertenece a  $\Theta_2$ , entonces el criterio minimax existe y tiene la forma siguiente:

$$X\beta^T > c. \quad (7)$$

Este criterio no es invariante respecto a cualquier grupo de transformaciones. Le proponemos al lector que él mismo se cerciore de que en este caso el c.r.v. es distinto de (7) y tiene la forma

$$\varrho^2(X, \Theta_2) - \varrho^2(X, 0) < c,$$

donde  $\varrho(X, \Theta_2) = \inf_{\alpha \in \Theta_2} |X - \alpha|$ ,  $\varrho(X, 0) = |X|$ .

Ahora demostraremos que cuando se cumplen ciertas suposiciones, el criterio de la relación de verosimilitud posee propiedades de invariación. Sea  $G$  cualquier grupo de transformaciones en  $\mathcal{X}^n$ , respecto al cual el problema de verificación de las hipótesis  $H_1$  y  $H_2$  es invariante, y sea  $\bar{G}$  el grupo respectivo de transformaciones  $\bar{g}$  en  $\Theta$ .

**Teorema 1.** Si  $f_{\theta}(x)$  posee la propiedad

$$f_{\theta}(gx) = c(g, x)f_{\bar{g}\theta}(x), \quad (8)$$

entonces el criterio de la relación de verosimilitud es invariante respecto a  $G$ .

En cuanto a la condición (8) diremos que la misma siempre se cumple cuando  $\mu$  es la medida de Lebesgue, y  $g$ , la transformación que conserva esa medida (desplazamiento y giro). En este caso  $c(g, x) \equiv 1$ . Para las transformaciones de contracción,  $c(g, x) \equiv \text{const.}$

**Demostración del teorema 1.** En virtud de que  $\bar{g}\Theta_i = \Theta_i$ ,  $i = 1, 2$ , tendremos

$$R(gx) = \frac{\sup_{\theta \in \Theta_2} f_{\theta}(gx)}{\sup_{\theta \in \Theta_1} f_{\theta}(gx)} = \frac{\sup_{\theta \in \Theta_2} c(g, x) f_{\bar{g}\theta}(x)}{\sup_{\theta \in \Theta_1} c(g, x) f_{\bar{g}\theta}(x)} = \frac{\sup_{\theta \in \bar{g}\Theta_2} f_{\theta}(x)}{\sup_{\theta \in \bar{g}\Theta_1} f_{\theta}(x)} = R(x). \quad \triangleleft$$

Otras propiedades del c.r.v. véanse en los §§ 11, 13—16.

### § 11\*. Análisis sucesivo

**1. Observaciones preliminares.** En todos los planteamientos anteriores, el volumen  $n$  de la muestra  $X = X_n$ , de la cual disponemos, estaba registrado. En tales condiciones hemos hallado criterios que poseían unas u otras propiedades de optimización. Por ejemplo, en el caso más elemental, cuando se verificaban dos hipótesis simples  $H_i = \{X \in P_i\}$ ,  $i = 1, 2$ , resultó que existe un c.m.p.  $\pi$  de nivel  $1 - \varepsilon$ , el cual tiene la forma (véase el teorema 2.1)

$$\pi(X) = \begin{cases} 1, & \text{si } f_2(X) > cf_1(X), \\ p, & \text{si } f_2(X) = cf_1(X), \\ 0, & \text{si } f_2(X) < cf_1(X). \end{cases}$$

Aquí  $c$  y  $p$  se deducen de la condición  $M_1 \pi(X) = \varepsilon$ , y  $f_i(x)$  son las densidades de las distribuciones  $P_i$ ,  $i = 1, 2$ , respecto a cierta medida  $\mu$ .

¿Será posible mejorar ulteriormente este procedimiento estadístico? En las condiciones enunciadas claro está que no es posible. Pero si desistimos en registrar el volumen de la muestra, o sea, si procedemos a que el número de observaciones  $n$  sea una variable aleatoria dependiente de las observaciones ya realizadas, entonces los mejoramientos son posibles. Se tiene en cuenta la reducción de la cantidad de observaciones indispensables para construir los criterios a base de ciertos parámetros dados. Esta circunstancia es importante en los experimentos donde la ejecución de ensayos ofrece gastos considerables.

La posibilidad de tal mejoramiento de los criterios puede ser aclarada citando el ejemplo siguiente. Supongamos que las distribuciones  $P_1$  y  $P_2$  no son del todo recíprocamente continuas, y supongamos también, que existen conjuntos  $B_1$  y  $B_2$  de  $\mathfrak{B}_{\mathcal{X}}$  tales, que  $f_1(x) > 0$ ,  $f_2(x) = 0$  cuando  $x \in B_1$ , y  $f_1(x) = 0$ ,  $f_2(x) > 0$  cuando  $x \in B_2$ . Entonces está claro que si

$x_1 \in B_1$  ( $x_1 \in B_2$ ), podemos afirmar infaliblemente que tiene lugar la hipótesis  $H_1$  ( $H_2$ ). En este caso no hay ninguna necesidad de llevar a efecto las observaciones posteriores.

Ahora bien, si los experimentos se realizan no de una vez (en cantidad de  $n$ ), sino sucesivamente, examinando el resultado de cada nueva serie de observaciones, entonces es posible reducir el volumen general de observaciones.

La introducción del procedimiento sucesivo también es muy natural desde el punto de vista del enfoque bayesiano. En efecto, el referido enfoque, examinado en el § 2, prescribe aceptar la hipótesis  $H_2$  si la probabilidad a posteriori  $q(2/X)$  de esta hipótesis  $\geq 1/2$ . En este caso, en el conjunto crítico se encontrarán, entre otras, tanto muestras  $X$  para las cuales  $q(2/x)$  es próxima a 1 (para tales  $X$ , la aceptación de  $H_2$  es oportuna), como muestras  $X$  para las cuales  $q(2/X)$  es próxima a  $1/2$ . Estas últimas podrían considerarse como muestras "insuficientes" para tomar decisiones y las cuales requieren experimentos adicionales. Además, al igual que en el ejemplo expuesto más arriba, la probabilidad a posteriori  $q(2/X)$  puede resultar grande ya después de las primeras pruebas, y entonces se podría tomar decisiones sin efectuar pruebas posteriores (en el ejemplo mencionado,  $q(2/X) = 1$  cuando  $X = x_1 \in B_2$  para cualquier distribución a priori ( $q(1), q(2), q(2) > 0$ )).

Más abajo examinaremos el procedimiento sucesivo para verificar dos hipótesis simples, en el cual se alcanzará la reducción máxima posible de la cantidad de observaciones.

**2. Criterio sucesivo bayesiano.** Examinemos primeramente el planteamiento bayesiano del problema y designemos por  $q(1) = q$  y  $q(2) = 1 - q$  las probabilidades a priori de las hipótesis  $H_1$  y  $H_2$ . Entonces, la probabilidad a posteriori de la hipótesis  $H_i$  después de las observaciones  $X = X_n$  será igual a

$$q(i/X_n) = \frac{q(i)f_i(X_n)}{q(1)f_1(X_n) + q(2)f_2(X_n)}. \quad (1)$$

Realizaremos sucesivamente las observaciones y para cada  $n$  calcularemos los valores de  $q(2/X_n)$ ,  $n = 1, 2, \dots$  (o de  $q(1/X_n)$ ). En el plano de las variables  $(n, y)$  examinaremos la trayectoria aleatoria de las probabilidades a posteriori (quebrada aleatoria), que parte del punto  $q = q(2)$  cuando  $n = 0$  y que toma, en los puntos  $n = 1, 2, \dots$ , los valores de  $y = q(2/X_n)$ . Con ayuda de esta trayectoria se puede construir el siguiente criterio para verificar la hipótesis  $H_1$  frente a  $H_2$ : examinemos en el plano  $(n, y)$  dos fronteras rectilíneas  $y = \gamma_i$ ,  $i = 1, 2$ ;  $0 < \gamma_1 < \gamma_2 < 1$  para la variable  $q(2/X_n)$ . Se acepta la hipótesis  $H_2$  si la trayectoria  $q(2/X_n)$ ,  $n = 0, 1, \dots$ , sale por primera vez de la franja  $(\gamma_1, \gamma_2)$  a través de la frontera superior  $\gamma_2$ . Si la trayectoria  $q(2/X_n)$ ,  $n = 0, 1, \dots$ , sale de esta franja a través de

la frontera inferior  $\gamma_1$ , entonces se acepta  $H_1$ . Más adelante veremos que la  $P_i$ -probabilidad ( $i = 1, 2$ ) de que  $q(2/X_n)$  nunca saldrá de la franja  $(\gamma_1, \gamma_2)$ , o sea, la probabilidad del suceso

$$\{\gamma_1 < q(2/X_n) < \gamma_2, n = 0, 1, \dots\}$$

es igual a cero.

El número de pruebas  $\nu$  que se necesita para aceptar una de las hipótesis (o sea, para alterar las desigualdades (2)) es, evidentemente, *variable aleatoria markoviana* (momento de parada) respecto a la sucesión  $x_1, x_2, \dots$  para cada una de las distribuciones  $P_1$  y  $P_2$ . Desde este punto de vista, dicha regla de aceptación de las hipótesis es *sucesiva* y concuerda bastante bien con las reglas conforme a las cuales actúa el hombre en su actividad práctica: tomar una u otra decisión después que las observaciones permitan reducir en sumo grado la incertidumbre que tiene lugar con respecto al objeto sometido a examen.

El criterio construido depende de  $q = q(1)$  y del vector  $\gamma = (\gamma_1, \gamma_2)$ . Por eso, designémoslo por  $\delta_{q, \gamma}$ . Ahora establezcamos que el criterio  $\delta_{q, \gamma}$  es óptimo. Con este fin introduzcamos primeramente el concepto general de criterio sucesivo, cuyas características esenciales, a la par con las probabilidades de los errores de primero y segundo genero, se convierten en los valores medios  $M_1\nu$  y  $M_2\nu$  para el número de observaciones  $\nu$  necesarias para tomar decisiones.

Supongamos que en  $(\mathcal{X}^\infty, \mathfrak{B}_{\mathcal{X}^\infty}^\infty)$  se da una variable aleatoria entera arbitraria  $\nu \geq 0$  que es markoviana respecto a la sucesión  $x_1, x_2, \dots$  ( $\{\nu \geq n\} \in \sigma(x_1, \dots, x_n) = \mathfrak{B}_{\mathcal{X}^\infty}^n$ ). Designemos por  $\mathcal{X}^\nu$  el espacio de los vectores  $(n, X_n)$  tales, que  $\nu(X_\infty) = n, X_n = [X_\infty]_n$ . Introduzcamos en  $\mathcal{X}^\nu$  la  $\sigma$ -álgebra de  $\mathfrak{B}^\nu$  engendrada por los sucesos  $\{\nu = n, X_n \in B^n\}, B^n \in \mathfrak{B}_{\mathcal{X}^\nu}^n, n = 0, 1, \dots$ . Está claro que cualquier distribución en  $(\mathcal{X}, \mathfrak{B}_{\mathcal{X}})$  (o en  $(\mathcal{X}^\infty, \mathfrak{B}_{\mathcal{X}^\infty}^\infty)$ ) induce la distribución respectiva en  $(\mathcal{X}^\nu, \mathfrak{B}^\nu)$ .

**Definición 1.** Llámase *criterio sucesivo*  $\delta$  para verificar  $H_1$  frente a  $H_2$ , el par  $(\nu, \Omega)$ , donde  $\Omega \in \mathfrak{B}^\nu$  es la región de aceptación de  $H_2$  (región crítica), y la variable aleatoria  $\nu$  se supone que es propia respecto a ambas distribuciones  $P_1, P_2$  ( $P_i(\nu < \infty) = 1, i = 1, 2$ ).

En los casos cuando sea necesario señalar que  $\nu$  y  $\Omega$  pertenecen al criterio  $\delta$ , escribiremos  $\nu(\delta)$  y  $\Omega(\delta)$ .

Es natural que, de un modo equivalente, el criterio sucesivo puede ser designado con ayuda de una función biforme medible en  $\mathcal{X}^\nu$ . También está claro que el criterio sucesivo  $\delta$  puede ser designado mediante la construcción de la región crítica (volvamos a designarla por  $\Omega$ ) en todo el espacio  $\mathcal{X}^\infty$ . Sin embargo, con tal aplicación (en  $\mathcal{X}^\infty$ ) de las regiones  $\Omega$  y  $\mathcal{X}^\nu \setminus \Omega$  de aceptación de las hipótesis  $H_2$  y  $H_1$ , no obtendremos obligatoriamente todos los elementos de  $\mathcal{X}^\infty$ : en aquellos de ellos para los cuales  $\nu(X_\infty) = \infty$ ,

no se acepta ninguna hipótesis. Pero según la definición de la  $P_i$ -probabilidad, los conjuntos de tales  $X_\infty$  equivalen a cero.

El criterio no randomizado ordinario  $\delta$  es un caso particular del criterio sucesivo, cuando  $\nu(\delta) \equiv n$  es constante (si  $\nu(\delta) \equiv 0$ , entonces la decisión se toma sin realizar ensayos).

El criterio sucesivo  $\delta$ , al igual que cualquier criterio ordinario para verificar dos hipótesis simples, se caracteriza por las probabilidades  $\alpha_i(\delta)$  de errores de  $i$ -ésimo género ( $i = 1, 2$ ):

$$\alpha_i(\delta) = P_i((\nu, X_\nu) \notin \Omega_i),$$

donde  $\Omega_2 = \Omega$ ,  $\Omega_1 = \mathcal{X}^\nu \setminus \Omega_2$ . Además, como ya hemos señalado, caracterizaremos el criterio sucesivo por los valores medios  $M_{i\nu}$ ,  $i = 1, 2$ . Es evidente que para el criterio ordinario  $\delta$ , construido según la muestra  $X_n$ , se cumple  $M_{i\nu}(\delta) \equiv n$ .

Para tomar en consideración la aparición de estos nuevos factores en el planteamiento del problema (o sea, de las características relacionadas con la magnitud  $\nu$ ), supondremos que la realización de cada observación necesita gastos de valor  $a$ . También será cómodo caracterizar las pérdidas que surgen al tomar decisiones incorrectas, por medio de distintos valores de  $w_1$  y  $w_2$ . Es decir, consideraremos que las pérdidas de  $i$ -ésimo género que surgen al tomar decisiones erróneas, cuando es cierta  $H_i$ , equivalen a  $w_i$ ,  $i = 1, 2$ .

Con estos acuerdos, la esperanza matemática  $R(q, \delta)$  de las pérdidas que surgen al utilizar el criterio  $\delta$ , es igual a

$$R(q, \delta) = q[\alpha_1(\delta)w_1 + aM_{1\nu}(\delta)] + (1 - q)[\alpha_2(\delta)w_2 + aM_{2\nu}(\delta)]. \quad (3)$$

Esta expresión se denomina *riesgo bayesiano* en el problema sujeto a examen. Si aquí suponemos que  $a = 0$ ,  $w_1 = w_2 = 1$ , obtendremos la expresión para la probabilidad de una decisión errónea del criterio  $\delta$ , la cual ya hemos utilizado repetidas veces en los §§ 1, 2.

**Definición 2.** El criterio sucesivo  $\delta$  que minimiza el riesgo bayesiano (3) se denomina criterio sucesivo *bayesiano*.

La siguiente afirmación establece la optimización (carácter bayesiano) del criterio  $\delta_{q,\gamma}$  construido al principio de este párrafo.

**Teorema 1.** Para  $a$ ,  $w_1$ ,  $w_2$  dados existen  $\gamma_1$ ,  $\gamma_2$  tales, que el criterio  $\delta_{q,\gamma}$  es bayesiano.

**Demostración.** Designemos por  $\delta_i$  el criterio que acepta la hipótesis  $H_i$  sin realizar pruebas, así que  $\nu(\delta_i) = 0$ ,  $\alpha_i(\delta_i) = 0$ . Aclaremos primeramente en qué casos el criterio  $\delta$ , que minimiza  $R(q, \delta)$ , coincide con  $\delta_1$  o con  $\delta_2$ . Es evidente que

$$R(q, \delta_1) = (1 - q)w_2, \quad R(q, \delta_2) = qw_1.$$

Sea  $K$  la clase de criterios  $\{\delta = \delta(X)\}$  que dependen al menos de una observación, o sea, la clase de criterios  $\delta$  para los cuales  $\nu(\delta) \geq 1$ . Es evidente que  $R(q, \delta) \geq a$  para  $\delta \in K$ . Designemos

$$R(q) = \inf_{\delta \in K} R(q, \delta).$$

Como el criterio  $\delta$ , basado en una sola prueba ( $\nu(\delta) \equiv 1$ ), pertenece a  $K$ , entonces  $R(q) < \infty$ .

Para cualquier  $p \in (0, 1)$  tenemos, en virtud de la linealidad de  $R(q, \delta)$  como función de  $q$ :

$$\begin{aligned} R(pq_1 + (1-p)q_2) &= \inf_{\delta \in K} [pR(q_1, \delta) + (1-p)R(q_2, \delta)] \geq \\ &\geq pR(q_1) + (1-p)R(q_2). \end{aligned}$$

Esto quiere decir que  $R(q)$  es una función cóncava. En vista de que  $a < R(q) < \infty$ , de aquí se deduce que  $R(q)$  también es una función continua en  $[0, 1]$ . Comparemos ahora los riesgos de los criterios  $\delta_1$  y  $\delta_2 \in K$  en función de  $q$  (véase la fig. 5).

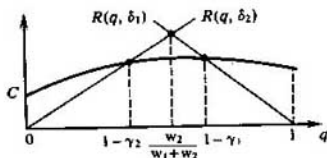


Fig. 5.

Una de dos: o bien  $R(q) \geq \min R(q, \delta_i)$  para todos  $q$  (esto corresponde al hecho de que  $R\left(\frac{w_2}{w_1 + w_2}\right) \geq \frac{w_1 w_2}{w_1 + w_2}$ , o bien existen soluciones de las ecuaciones  $R(q, \delta_1) = R(q)$ ,  $R(q, \delta_2) = R(q)$ , que designaremos  $1 - \gamma_1$ ,  $1 - \gamma_2$ ,  $1 - \gamma_1 > 1 - \gamma_2$ , respectivamente. Es evidente que  $R(q) < \min R(q, \delta_i)$  dentro del intervalo  $(1 - \gamma_2, 1 - \gamma_1)$ . Para la primera de las posibilidades mencionadas supongamos

$$1 - \gamma_1 = 1 - \gamma_2 = \frac{w_1}{w_1 + w_2},$$

así que

$$R(1 - \gamma_1, \delta_1) = R(1 - \gamma_1, \delta_2).$$

De los referidos razonamientos y de la fig. 5 se deduce la siguiente regla óptima de acciones. A base de los datos  $a$ ,  $w_1$ ,  $w_2$  calculamos  $1 - \gamma_1$ ,  $1 - \gamma_2$ . Si  $q \leq 1 - \gamma_2$  o bien, que es lo mismo,  $1 - q \geq \gamma_2$ , el menor riesgo

entre todos los criterios lo proporciona  $\delta_2$  (o sea, es necesario aceptar inmediatamente  $H_2$ ). Si  $q \geq 1 - \gamma_1$  ( $1 - q \leq \gamma_1$ ), entonces  $\delta_1$  ofrece el menor riesgo (es preciso aceptar  $H_1$ ). Y sólo en el caso de  $1 - \gamma_2 < 1 - \gamma_1$ ,  $q \in (1 - \gamma_2, 1 - \gamma_1)$  (o bien  $1 - q \in (\gamma_1, \gamma_2)$ ) es necesario utilizar el criterio de  $K$ , o sea, hay que realizar el experimento.

Ahora aprovechemos la inducción. Supongamos que se han efectuado  $n$  observaciones y que disponemos de la muestra  $X_n$ . Antes de la observación  $n + 1$  tenemos la misma alternativa: no realizar más observaciones y aceptar una de las hipótesis  $H_i$ , o bien continuar las observaciones. El hecho de que ya hemos sufrido las pérdidas  $an$  no desempeña ningún papel, ya que éstas no pueden ser eliminadas de ningún modo. Los cambios esenciales sólo están relacionados con la distribución a priori. Ahora el papel de probabilidades  $q(1) = q$  y  $q(2) = 1 - q$  deben desempeñarlos las probabilidades a posteriori  $q(1/X_n)$ ,  $q(2/X_n)$ . Con arreglo a esta nueva situación, la regla óptima ya elaborada por nosotros, dice que es necesario aceptar  $H_2$  si  $q(2/X_n) \geq \gamma_2$ , y  $H_1$  si  $q(2/X_n) \leq \gamma_1$ . Si  $q(2/X_n) \in (\gamma_1, \gamma_2)$ , entonces conviene continuar las observaciones. Pero la regla obtenida no es otra cosa sino el criterio  $\delta_{q,\gamma}$ . Ahora bien, hemos hallado  $\gamma_i = \gamma_i(a, w_1, w_2)$  que poseen la propiedad de que el criterio  $\delta_{q,\gamma}$  minimiza el riesgo  $R(q, \delta)$ .  $\triangleleft$

Nótese que los números  $\gamma_i(a, w_1, w_2)$  permanecen invariables al multiplicar  $a, w_1, w_2$  por un mismo número: esto es evidente de su definición, ya que tal operación sólo conduce a que todos los riesgos  $R(q, \delta)$  sean multiplicados por ese mismo número. Así pues, en realidad  $\gamma_i$  es una función de dos variables, por ejemplo, de  $a$  y  $w_1$  si consideramos que  $w_2 = 1 - w_1$ .

¿Qué representa en sí el criterio bayesiano  $\delta_{q,\gamma}$ ? El mismo prescribe no realizar observaciones en dos casos: cuando  $\gamma_1 = \gamma_2$  (lo cual sucede en caso de que  $a$  es grande en comparación con  $w_1, w_2$ ), o bien cuando  $q(2) \leq \gamma_1$  o cuando  $q(2) \geq \gamma_2$ . En los demás casos es preciso realizar experimentos hasta la primera alteración de las desigualdades

$$\gamma_1 < q(2/X_n) < \gamma_2$$

o bien, que es lo mismo, (véase (1)), hasta la primera alteración de las desigualdades

$$\frac{\gamma_1 q(1)}{(1 - \gamma_1)q(2)} < \frac{f_2(X_n)}{f_1(X_n)} < \frac{\gamma_2 q(1)}{(1 - \gamma_2)q(2)}. \quad (4)$$

En este caso se acepta la hipótesis  $H_2$  si por primera vez se altera la desigualdad derecha, y la hipótesis  $H_1$  si se altera la desigualdad izquierda. En tal forma, la parte "variable" del criterio  $\delta_{q,\gamma}$  ya no está relacionada con el planteamiento bayesiano del problema y podemos, designando por  $\Gamma_1, \Gamma_2$  las fronteras izquierda y derecha en (4), examinar el criterio sucesivo  $\delta_T$ ,  $\Gamma = (\Gamma_1, \Gamma_2)$  que se llama *criterio sucesivo de la relación de verosimilitud*. Fue Wald quien lo introdujo por primera vez.



### 3. Criterio sucesivo que minimiza el número medio de pruebas.

**Teorema 2.** Sea  $\Gamma_1 < 1 < \Gamma_2$ . Designemos por  $\alpha_1$  y  $\alpha_2$  las probabilidades de errores de primero y segundo género del criterio  $\delta_\Gamma$ . Entonces, entre todos los criterios sucesivos  $\delta$ , para los cuales  $\alpha_1(\delta) \leq \alpha_1$ ,  $\alpha_2(\delta) \leq \alpha_2$ , el criterio  $\delta_\Gamma$  tendrá los menores valores de  $M_{1\nu}(\delta)$  y  $M_{2\nu}(\delta)$ .

Este teorema significa, en particular, que si  $\delta$  es un criterio construido según la muestra  $X_n$  de volumen registrado, para el cual  $\alpha_1(\delta) \leq \alpha_1$ ,  $\alpha_2(\delta) \leq \alpha_2$ , entonces

$$M_{i\nu}(\delta_\Gamma) \leq n, \quad i = 1, 2.$$

**Demostración.** El criterio bayesiano  $\delta_{2,\gamma}$ , examinado en el teorema 1, se determina por el conjunto de números  $(q, a, w_1, w_2)$ . Pero, como ya hemos señalado, la multiplicación de  $a, w_1, w_2$  por un mismo número no altera las fronteras  $\gamma_i$ , así que, de hecho,  $\delta_{q,\gamma}$  se determina a base de tres parámetros, por ejemplo,  $(q, a, w)$  si se toma  $w_1 = w$  y  $w_2 = 1 - w$ .

Si partimos de este acuerdo, en el teorema 1 hemos construido, a base de los valores dados de  $(a, w)$ , los números  $\gamma_i = \gamma_i(a, w)$  para los cuales el criterio  $\delta_{q,\gamma}$  es bayesiano. Ahora necesitaremos, en cierto sentido, la afirmación inversa acerca de que para los valores dados de  $\gamma_1, \gamma_2$  existen  $a, w$  tales, que  $\gamma_i(a, w) = \gamma_i$ , o sea, tales  $a, w$ , para los cuales el criterio  $\delta_{q,\gamma}$  será bayesiano en el problema correspondiente al conjunto  $(q, a, w)$ . Esta afirmación tiene carácter técnico y se demuestra de un modo bastante complicado (véase [57]). Por eso la aceptamos como tolerable\*).

Así pues, examinemos el criterio  $\delta_\Gamma$ , y para el valor dado de  $g$  hallemos  $\gamma_i$  de las ecuaciones

$$\frac{\gamma_i q}{(1 - \gamma_i)(1 - q)} = \Gamma_i.$$

Para los valores obtenidos de  $\gamma_i = \Gamma_i(1 - q)/(\Gamma_i(1 - q) + q)$  hallemos  $a, w$  con los cuales el criterio  $\delta_{q,\gamma}$  será bayesiano en el problema que corresponde al conjunto  $(q, a, w)$ . Como  $\Gamma_1 < 1 < \Gamma_2$ , entonces  $\gamma_1 < 1 - q < \gamma_2$  y  $\nu(\delta_{q,\gamma}) \geq 1$ . Esto significa que  $\delta_{q,\gamma} = \delta_\Gamma$ .

Sea ahora  $\delta$  cualquier otro criterio para el cual  $\alpha_i(\delta) \leq \alpha_i$ . En vista de que el criterio  $\delta_{q,\gamma} = \delta_\Gamma$  minimiza el riesgo bayesiano, entonces

$$q[\alpha_1 w + aM_{1\nu}(\delta_\Gamma) + (1 - q)[\alpha_2(1 - w) + aM_{2\nu}(\delta_\Gamma)] \leq \\ \leq q[\alpha_1(\delta)w + aM_{1\nu}(\delta)] + (1 - q)[\alpha_2(\delta)(1 - w) + aM_{2\nu}(\delta)].$$

\* Aquí tampoco demostramos otra afirmación útil acerca de que para las  $P_1$ -distribuciones continuas de la magnitud  $f_2(X)/f_1(X)$ , y para todos los valores dados de  $\alpha_1, \alpha_2$  habrá  $\Gamma_1, \Gamma_2$  tales, que  $\alpha_1(\delta_\Gamma) = \alpha_1, \alpha_2(\delta_\Gamma) = \alpha_2$ . Por su esencia esta afirmación se asemeja a los lemas 6.1 y 7.1, pero su demostración es más difícil.

De aquí resulta

$$qM_1\nu(\delta_{\Gamma}) + (1 - q)M_2\nu(\delta_{\Gamma}) \leq qM_1\nu(\delta) + (1 - q)M_2\nu(\delta).$$

Como el número  $q \in (0, 1)$  aquí es arbitrario, entonces

$$M_1\nu(\delta_{\Gamma}) \leq M_1\nu(\delta), \quad M_2\nu(\delta_{\Gamma}) \leq M_2\nu(\delta). \quad \triangleleft$$

Aquí hemos utilizado, para la demostración, el mismo método de comparación con los criterios bayesianos que habíamos empleado en los §§ 1, 2, 5.

Examinemos algunas *propiedades del criterio*  $\delta_{\Gamma}$ . Designemos por  $\Omega_i^n$  los subconjuntos de  $\mathcal{X}^{\infty}$  que se definen del modo siguiente ( $X_k = [X_{\infty}]_k$ ):

$$\Omega_i^n = \left\{ X_{\infty}: \Gamma_1 < \frac{f_2(X_k)}{f_1(X_k)} < \Gamma_2, \quad k = 1, \dots, n-1, \frac{f_2(X_n)}{f_1(X_n)} \leq \Gamma_1 \right\}.$$

El conjunto  $\Omega_2^n$  se define del mismo modo, pero la última desigualdad debe sustituirse por  $f_2(X_n)/f_1(X_n) \geq \Gamma_2$ . Es evidente que  $\Omega_i^n$  son disjuntos, pues

$$\Omega_i = \bigcup_{n=1}^{\infty} \Omega_i^n \text{ es la región de aceptación de } H_i,$$

$\nu(\delta_{\Gamma}) = n$  en la región  $\{x \in \mathcal{X}^{\infty}: x \in \Omega_i^n\}$ ,

$$\begin{aligned} \alpha_1(\delta_{\Gamma}) &= \sum_{n=1}^{\infty} \mathbf{P}_1(\Omega_1^n) = \sum_{n=1}^{\infty} \int_{\mathcal{X}^n \cap \Omega_1^n} f_1(x) \mu^n(dx) \leq \\ &\leq \sum_{n=1}^{\infty} \int_{\mathcal{X}^n \cap \Omega_1^n} f_2(x) \Gamma_2^{-1} \mu^n(dx) = (1 - \alpha_2(\delta_{\Gamma}))/\Gamma_2. \end{aligned} \quad (5)$$

Análogamente se establece que

$$\alpha_2(\delta_{\Gamma}) \leq \Gamma_1(1 - \alpha_1(\delta_{\Gamma})). \quad (6)$$

Pongamos, para abreviar,  $\alpha_i(\delta_{\Gamma}) = \alpha_i$ . El grado de exactitud de las desigualdades obtenidas

$$\Gamma_2 \leq \frac{1 - \alpha_2}{\alpha_1}, \quad \Gamma_1 \geq \frac{\alpha_2}{1 - \alpha_1} \quad (7)$$

lo examinaremos más adelante. Ahora aclararemos las propiedades del criterio que obtendremos si hacemos uso de las relaciones (7) en calidad de base para determinar  $\Gamma_i$  por los valores de  $\alpha_i$  dados. Si ponemos

$$\Gamma_1' = \frac{\alpha_2}{1 - \alpha_1}, \quad \Gamma_2' = \frac{1 - \alpha_2}{\alpha_1}, \quad \alpha_i' = \alpha_i(\delta_{\Gamma'}),$$

entonces para el criterio obtenido  $\delta_{\Gamma'}$  tendremos, en virtud de (7),

$$\frac{\alpha_2}{1 - \alpha_1} \geq \frac{\alpha_2'}{1 - \alpha_1'}, \quad \frac{1 - \alpha_2}{\alpha_1} \leq \frac{1 - \alpha_2'}{\alpha_1'}, \quad (8)$$

De aquí resulta

$$\alpha'_1 \leq \frac{\alpha_1(1 - \alpha'_2)}{1 - \alpha_2} \leq \frac{\alpha_1}{1 - \alpha_2}, \quad \alpha'_2 \leq \frac{\alpha_2(1 - \alpha'_1)}{1 - \alpha_1} \leq \frac{\alpha_2}{1 - \alpha_1}.$$

Reduciendo las desigualdades (8) al denominador común y sumándolas, obtenemos asimismo

$$\alpha'_1 + \alpha'_2 \leq \alpha_1 + \alpha_2.$$

Ahora bien, si  $\alpha_i$  son pequeños, el criterio  $\delta_{\Gamma'}$  tendrá los valores de  $\alpha'_i$  cuya suma no excede  $\alpha_1 + \alpha_2$ , y cada  $\alpha'_i$  puede superar  $\alpha_i$  sólo insignificanamente y dentro de los límites que conocemos.

**Ejemplo 1.** Supongamos que  $x_i$  tiene una distribución binomial con una probabilidad de éxito  $p$ . El problema consiste en verificar la hipótesis  $H_1 = \{p = p_1\}$  frente a  $H_2 = \{p = p_2\}$ ,  $p_1 < p_2$ . En este caso

$$\frac{f_2(X)}{f_1(X)} = \frac{p_2^{\eta_n}(1 - p_2)^{n - \eta_n}}{p_1^{\eta_n}(1 - p_1)^{n - \eta_n}} = \left( \frac{p_2(1 - p_1)}{p_1(1 - p_2)} \right)^{\eta_n} \left( \frac{1 - p_2}{1 - p_1} \right)^n,$$

donde  $\eta_n$  es el número de casos favorables (éxitos) en  $n$  pruebas. Para los valores  $p_1 = 0,05$ ,  $p_2 = 0,17$ ,  $\alpha_1 = 0,05$ ,  $\alpha_2 = 0,10$  obtenemos\*)  $\Gamma'_1 = 0,105$ ,  $\Gamma'_2 = 18$ ,  $\alpha_1 = 0,031$ ,  $\alpha_2 = 0,099$ .

$$\mathbf{M}_1 \nu(\delta_{\Gamma'}) = 31,4, \quad \mathbf{M}_2 \nu(\delta_{\Gamma'}) = 30,0.$$

Por otro lado, el procedimiento con un volumen fijo de la muestra y con probabilidades de los errores de primero y segundo género correspondientes a 0,05 y 0,10, respectivamente, requiere  $n = 57$  observaciones. Ahora bien, en este ejemplo el procedimiento sucesivo reduce casi el doble el número medio de observaciones.

**4. Cálculo de los parámetros del mejor criterio sucesivo.** Las relaciones (7) y (8) dan la posibilidad de establecer cierta correspondencia entre la frontera  $\Gamma$  y las probabilidades de los errores  $\alpha_i(\delta_{\Gamma})$ . Ahora examinemos más detalladamente el problema de cálculo del criterio  $\delta_{\Gamma}$ .

a) *Fórmulas exactas.* Designemos

$$z_k = \ln \frac{f_2(x_k)}{f_1(x_k)}, \quad k = 1, 2, \dots$$

$$A_i = \ln \Gamma_i, \quad i = 1, 2.$$

En este caso el criterio  $\delta_{\Gamma}$  puede adquirir la forma siguiente: si  $A_1 < 0 < A_2$ , entonces los experimentos se realizan sucesivamente, y los valores  $z_k$  independientes e igualmente distribuidos se suman hasta que  $Z_n = \sum_{k=1}^n z_k$  toque por primera vez una de las fronteras  $A_i$ . Si es cierta la hipótesis  $H_2$ , la divagación descrita será dirigida, por término medio, hacia arriba,

\*) Los datos numéricos se han tomado de [57], p. 143.

ya que

$$M_2 z_1 = \int \ln \frac{f_2(x)}{f_1(x)} \cdot f_2(x) \mu(dx) = \varrho_1(P_2 P_1) > 0$$

(véase el lema 2.6.1). De un modo análogo se determina que  $M_1 z_1 = -\varrho_1(P_1, P_2) < 0$ .

Si las fronteras  $A_i$  se alejan a partir del origen de coordenadas, esto corresponde (compárese con (5) y (6)) a la reducción de los errores de primero y segundo género.

Los conjuntos  $\Omega_2^n$  en los términos de divagación  $\{Z_k\}$  tendrán la forma

$$\Omega_2^n = \{A_1 < Z_k < A_2, k = 1, \dots, n-1, Z_n \geq A_2\}.$$

Los conjuntos  $\Omega_2^n$  tendrán una forma análoga.

Designemos por  $\eta(t)$  la variable aleatoria igual al tiempo de la primera salida de la divagación aleatoria  $Z_0 = 0, Z_1, Z_2, \dots$  fuera de la frontera de  $t$ :

$$\eta(t) = \begin{cases} \text{mín } \{k: Z_k \geq t\} & \text{para } t > 0, \\ \text{mín } \{k: Z_k \leq t\} & \text{para } t < 0. \end{cases}$$

Es el proceso de reconstrucción que corresponde a la sucesión  $\{Z_k\}$  (véase [11], capítulo 8). Las diferencias  $\chi(A_i) = Z_{\eta(A_i)} - A_i$  serán los valores de excesos (saltos) a través de los niveles  $A_i$  en la divagación  $\{Z_k\}$  (véase [11]).

Para la probabilidad de error de primer género ahora podemos escribir

$$\begin{aligned} \alpha_1(\delta_T) &= \sum_{n=1}^{\infty} P_1(\Omega_2^n) = \sum_{n=1}^{\infty} \int_{\mathcal{Q}^n \cap \Omega_2^n} \frac{f_1(x)}{f_2(x)} f_2(x) \mu^n(dx) = \\ &= \sum_{n=1}^{\infty} M_2(e^{-Z_n}; \Omega_2^n) = \Gamma_2^{-1} M_2(e^{-x(A_2)}; \Omega_2), \quad (9) \end{aligned}$$

donde  $\Omega_2 = \bigcup_{n=1}^{\infty} \Omega_2^n$  es la región de aceptación de  $H_2$ . Análogamente

$$\alpha_2(\delta_T) = \Gamma_1 M_1(e^{x(A_1)}; \Omega_1), \quad \Omega_1 = \bigcup_{n=1}^{\infty} \Omega_1^n. \quad (10)$$

Seguidamente, para los valores de  $M_i \nu, i = 1, 2, \nu = \nu(\delta_T)$ , en virtud de la identidad de Wald, obtenemos  $M_i(Z_\nu) = M_i z_1 M_i \nu, i = 1, 2$ .

Como  $Z_\nu = A_2 + \alpha(A_2)$  en el conjunto  $\Omega_2, Z_\nu = A_1 + \chi(A_1)$  en el conjunto  $\Omega_1$ , entonces

$$\begin{aligned} M_1 \nu &= \frac{1}{M_1 z_1} [\alpha_1 A_2 + M_1(\chi(A_2); \Omega_2) + (1 - \alpha_1) A_1 + M_1(\chi(A_1); \Omega_1)], \\ M_2 \nu &= \frac{1}{M_2 z_1} [(1 - \alpha_2) A_2 + M_2(\chi(A_2); \Omega_2) + \alpha_2 A_1 + M_2(\chi(A_1); \Omega_1)]. \quad (11) \end{aligned}$$

En varios casos los segundos miembros en las fórmulas (9)—(11) pueden ser determinados de forma explícita. Estas fórmulas también resultan muy útiles en los cálculos aproximados.

b) *Fórmulas (para  $A_1$  y  $A_2$  grandes) y desigualdades aproximadas.* Ya hemos señalado que los grandes valores de  $|A_i|, i = 1, 2$  corresponden a pequeñas probabilidades de errores  $\alpha_i(\delta_T)$ . Examinemos el valor

$$\begin{aligned} \alpha_1(\delta_T) &= P_1 \left( \sup_{k < \eta(A_1)} Z_k \geq A_2 \right) = P_1 \left( \sup_{k \geq 0} Z_k \geq A_2 \right) - \\ &- P_1 \left( \sup_{k < \eta(A_1)} Z_k < A_2, \sup_{k > \eta(A_1)} Z_k \geq A_2 \right). \quad (12) \end{aligned}$$

Aquí el último sumando no supera, en virtud del carácter markoviano de la variable aleatoria  $\eta(t)$ , los valores

$$P_1 \left( \sup_{k > \eta(A_1)} (Z_k - Z_{\eta(A_1)}) \geq A_2 - Z_{\eta(A_1)} \right) \leq P_1 \left( \sup_{k \geq 0} Z_k \geq A_2 - A_1 \right).$$

Como en casi todos los casos prácticamente interesantes, la probabilidad  $u(A) = P_1(\sup_{k \geq 0} Z_k \geq A)$  decrece exponencialmente con el aumento de  $A$  (véase, por ejemplo, [32], t. 2. Esto mismo se puede deducir del capítulo 10 en [11], donde se exponen los métodos de cálculos de  $u(A)^*$ ), entonces, para  $|A_i|$  grandes, el valor de  $u(A_2 - A_1)$  tendrá un orden más alto de pequeñez que  $u(A_2)$ . Esto significa, en virtud de (12), que

$$\alpha_1(\delta_T) \approx P_1 \left( \sup_{k \geq 0} Z_k \geq A_2 \right) = u(A_2), \quad (13)$$

así que, para grandes  $A_1$  y  $A_2$  en (12), la segunda frontera puede ser omitida. Exactamente igual obtenemos la aproximación

$$\alpha_2(\delta_T) \approx P_2 \left( \inf_{k \geq 0} Z_k \leq A_1 \right). \quad (14)$$

Si  $|A_i|$  son grandes y  $\alpha_i$  pequeños, los miembros principales en (11) proporcionan

$$M_{1\nu} \approx \frac{A_1}{M_{1z_1}}, \quad M_{2\nu} \approx \frac{A_2}{M_{2z_1}}. \quad (15)$$

Estas fórmulas también se basan en la omisión de la segunda frontera (ellas también pueden obtenerse mediante las aproximaciones  $M_{i\nu} \approx M_{i\eta}(A_i) \approx A_i/M_{iz_1}$ . La última relación tiene lugar en virtud del teorema de reconstrucción ([11])).

Teniendo en cuenta los términos siguientes, según su orden de pequeñez en (11), obtenemos

$$\begin{aligned} M_{1\nu} &= \frac{1}{M_{1z_1}} (A_1 + \alpha_1(A_2 - A_1) + M_{1\chi_1}), \\ M_{2\nu} &= \frac{1}{M_{2z_1}} (A_2 + \alpha_2(A_2 - A_1) + M_{2\chi_2}), \end{aligned} \quad (16)$$

donde  $\alpha_i$  se definen por las aproximaciones (12) y (13), los valores  $M_{i\chi_i} = \lim_{|A_i| \rightarrow \infty} M_{i\chi_i}(A_i)$  pueden ser determinados por los métodos descritos en el capítulo 10 en [11].

Examinemos ahora las desigualdades (8). Como  $\chi(A_1) \leq 0$ ,  $\chi(A_2) \geq 0$ , estas igualdades se deducen de (9) y (10) si  $\chi(A_i)$  se sustituye por 0. Consiguientemente, la exactitud de tales desigualdades depende del error originado por dicha sustitución.

Si las variables aleatorias  $z_i$  están limitadas,  $b_1 \leq z_i \leq b_2$ , es evidente que  $\chi(A_2) \leq b_2$ ,  $\chi(A_1) \geq b_1$ , y además de (5) y (6) pueden escribirse las desigualdades inversas. Es decir,

$$\begin{aligned} \alpha_1(\delta_T) &= \Gamma_2^{-1} M_2 (e^{-\chi(A_2)}; \Omega_2) \geq \Gamma_2^{-1} e^{-b_2} (1 - \alpha_2), \\ \alpha_2(\delta_T) &\geq \Gamma_1 e^{b_1} (1 - \alpha_1). \end{aligned} \quad (17)$$

A fin de ilustrar las relaciones obtenidas, volvamos a examinar el ejemplo 1. Para éste,

$$Z_n = \eta_n \ln \frac{p_2(1-p_1)}{p_1(1-p_2)} + n \ln \frac{1-p_2}{1-p_1},$$

donde  $\eta_n$  es el número de casos favorables en  $n$  pruebas. Esto quiere decir que  $z_1$ , para la  $P_T$ -distribución, adopta el valor de  $b_2 = \ln(p_2/p_1) \approx 1,224$  con probabilidad  $p_1$ , y el valor de

\* Esto se expone más detalladamente en [9].

$b_1 = \ln \frac{1 - p_2}{1 - p_1} \approx -0,135$  con probabilidad  $1 - p_i$ ,  $i = 1, 2$ . De aquí obtenemos

$$M_{1z_1} = -0,067, M_{2z_1} = 0,096, e^{b_2} = 3,400, e^{b_1} = 0,874.$$

De los dos últimos valores sólo el segundo es próximo a 1, así que será relativamente exacta tan sólo la segunda igualdad de (17). Utilizando esta desigualdad en (7) para el criterio  $\delta_{r'}$ , obtenemos

$$0,102 = \frac{\alpha_2'}{1 - \alpha_1'} \leq \Gamma_1' \leq \frac{\alpha_2'}{(1 - \alpha_1')e^{b_1}} = 0,117.$$

Esto proporciona fronteras bastante exactas para el valor de  $\Gamma_1' = 0,105$ . En nuestro caso

$$A_1' = \ln \Gamma_1' \approx -2,254, A_2' = \ln \Gamma_2' \approx 2,890.$$

De aquí, utilizando las fórmulas aproximadas (15), obtenemos para  $M_{i\nu'}$ ,  $i = 1, 2$ , los valores

$$A_1'/M_{1z_1} = 33,639, A_2'/M_{2z_1} = 30,108.$$

Vemos que incluso aproximaciones que están lejos de ser precisas, tales como (15), dan una noción correcta de las magnitudes  $M_{i\nu'}$ . Los resultados serán mucho más exactos si hacemos uso de las fórmulas (16).

## § 12. Verificación de las hipótesis compuestas en el caso general

En este párrafo no vamos a suponer que la muestra pertenece a cualquier familia paramétrica.

El problema de verificación de dos hipótesis en el caso general tiene la forma siguiente. Sean  $\mathcal{P}_1$  y  $\mathcal{P}_2$  dos familias de distribuciones tales, que la distribución  $\mathbf{P}$  de la muestra  $X$  pertenece a  $\mathcal{P}_1 \cup \mathcal{P}_2$ . Se verifica la hipótesis  $H_1 = \{X \in \mathbf{P}, \mathbf{P} \in \mathcal{P}_1\}$  frente a  $H_2 = \{X \in \mathbf{P}, \mathbf{P} \in \mathcal{P}_2\}$ . El principio general de construcción del criterio (no randomizado\*)  $\pi(X) = \delta(X)$  aquí queda igual que antes, tal como fue descrito en el § 4 para el caso paramétrico. Se construye precisamente el conjunto crítico  $\Omega \subset \mathcal{X}^n$  (que a menudo se identifica con el concepto de criterio) tal, que aceptamos  $H_2$  cuando  $X \in \Omega$ , y aceptamos  $H_1$  en el caso contrario. El número

$$1 - \varepsilon \inf_{\mathbf{P} \in \mathcal{P}_1} \mathbf{P}(X \notin \Omega)$$

se llama nivel de importancia del criterio. La magnitud

$$\beta_\pi(\mathbf{P}) = \mathbf{P}(X \in \Omega), \quad \mathbf{P} \in \mathcal{P}_2,$$

es el valor de la potencia del criterio  $\pi$  en el "punto"  $\mathbf{P} \in \mathcal{P}_2$ .

Cuando el conjunto  $\mathcal{P}_2$  de alternativas  $\mathbf{P}$  es muy abundante, en estas condiciones es muy difícil o incluso imposible comparar las potencias  $\beta_\pi(\mathbf{P})$

\* Para mantener la uniformidad de las designaciones, en lo sucesivo designaremos los criterios estadísticos con el símbolo  $\pi$ , aunque dentro de los límites de este capítulo se tratará, por lo general, de criterios no randomizados

de los criterios  $\pi$  y construir los criterios óptimos. Las mínimas exigencias planteadas ante los criterios, en este caso consisten, por lo general, en que para cada  $\mathbf{P} \in \mathcal{P}_2$  registrado se cumpla

$$\lim_{n \rightarrow \infty} \beta_{\pi}(\mathbf{P}) = 1.$$

**Definición 1.** El criterio  $\pi$  que posee esta propiedad se denomina *criterio conciliable*.

La esencia de los criterios sometidos a estudio, al igual que de todos los criterios estadísticos, corresponde al principio fundamental de la estadística matemática, del cual ya hemos hablado en los párrafos 1.4 y 2.31. Si  $\varepsilon$  es pequeño, entonces, al cumplirse la hipótesis  $H_1$  y al utilizarse muchas veces el criterio construido de nivel  $1 - \varepsilon$ , nos equivocaremos (o sea, caeremos en la región crítica), por término medio, sólo en el  $100\varepsilon\%$  de todas las pruebas. Por lo tanto, en caso de cumplirse la hipótesis  $H_1$ , consideramos prácticamente imposible la caída en esa región al realizar una sola prueba. Consiguientemente, si a pesar de todo caemos en ella, eso significará que la suposición hecha no es cierta y anunciamos que la hipótesis  $H_1$  no es verdadera. En este caso se dice que los resultados del experimento no concuerdan con la hipótesis  $H_1$  desde el punto de vista del criterio de nivel  $1 - \varepsilon$ .

Están muy difundidos los criterios de verificación de la hipótesis simple  $H_1 = \{X \in \mathbf{P}_1\}$  frente a la hipótesis alternativa compuesta  $H_2 = \{X \in \mathbf{P} \neq \mathbf{P}_1\}$ ; la hipótesis  $H_2$  significa que  $X$  es una muestra de la distribución arbitraria  $\mathbf{P} \neq \mathbf{P}_1$ .

La construcción de los criterios para verificar la hipótesis simple  $H_1 = \{X \in \mathbf{P}_1\}$  suele basarse en el "alejamiento" de la distribución empírica  $\mathbf{P}_n^*$  respecto a la distribución  $\mathbf{P}_1$  desde el punto de vista de cierta "distancia"  $d(\mathbf{P}, \mathbf{Q})$ . La propiedad deseable de esta distancia consiste en reducir  $(\mathbf{P}, \mathbf{Q})$  a cero sólo cuando  $\mathbf{Q} = \mathbf{P}$ , y en transformar la continuidad  $d(\mathbf{P}, \mathbf{Q})$  en el "entorno" del punto  $\mathbf{Q} = \mathbf{P}$ , por ejemplo, en la métrica uniforme (de lo contrario las pequeñas desviaciones de  $\mathbf{Q}$  respecto a  $\mathbf{P}$  pueden conducir a grandes valores de la distancia  $d$ ). Recordemos que en el caso paramétrico hemos utilizado consideraciones análogas al construir las estimaciones del parámetro desconocido aplicando el método de distancia mínima.

Así pues, sea  $d(\mathbf{P}, \mathbf{Q})$  cierta distancia (no obligatoriamente métrica) en el espacio de distribuciones. Supongamos que a partir de  $\varepsilon > 0$  dado se puede hallar tal  $c > 0$ , para el cual

$$\mathbf{P}_1(d(\mathbf{P}_1, \mathbf{P}_n^*) > c) = \varepsilon.$$

Entonces el criterio se construye del modo siguiente:

$$\pi(X) = \begin{cases} 0, & \text{si } d(\mathbf{P}_1, \mathbf{P}_n^*) \leq c, \\ 1, & \text{en el caso contrario.} \end{cases}$$

Evidentemente,  $\pi$  es un criterio de nivel  $1 - \varepsilon$ .

Al igual que en el § 3, se puede introducir un criterio de *nivel asintótico*  $1 - \varepsilon$  para el cual

$$\lim_{n \rightarrow \infty} P_1(d(\mathbf{P}_1, \mathbf{P}_n^*) > c) = \varepsilon. \quad (2)$$

Los criterios descritos suelen llamarse *criterios de aceptación* (suponiendo que  $\{X \in \mathbf{P}_1\}$ ). Análogamente, su estructura también puede ser representada de una forma algo diferente. Supongamos que tenemos una funcional  $G(\mathbf{P})$  (o una sucesión de funcionales  $G_n(\mathbf{P})$ ) tal, que  $G(\mathbf{P}) \neq G(\mathbf{P}_1)$  cuando  $\mathbf{P} \neq \mathbf{P}_1$ . Entonces podemos poner  $\pi(X) = 1$  si  $|G(\mathbf{P}_n^*) - G(\mathbf{P}_1)| > c$ , y  $\pi(X) = 0$  en el caso contrario, donde  $c$  se elige partiendo de las mismas consideraciones que en (1) y (2). No es difícil comprobar que este segundo enfoque es equivalente al primero, puesto que a partir de la funcional  $G$  se puede construir la distancia  $d(\mathbf{P}, \mathbf{P}_1) = |G(\mathbf{P}) - G(\mathbf{P}_1)|$  (compárese con el principio de sustitución en la teoría de estimación), y al contrario, a partir de la distancia  $d(\mathbf{P}, \mathbf{P}_1)$  se puede construir la funcional  $G(\mathbf{P}) = d(\mathbf{P}, \mathbf{P}_1)$  ( $G(\mathbf{P}_1) = 0$ ) que satisface las propiedades requeridas.

Si en la estructura descrita, la funcional  $G$  posee, además, la propiedad  $G(\mathbf{P}_n^*) \xrightarrow{p} G(\mathbf{P})$  cuando  $X \in \mathbf{P}$  (esto siempre es así cuando  $G$  es una función de primero o segundo tipo (véase el § 1.3)), entonces *el criterio construido será conciliable*. En efecto, en este caso el número  $c = c(n)$  que asegura la igualdad (2) debe convergir a cero ( $P_1(|G(\mathbf{P}_n^*) - G(\mathbf{P}_1)| > \varepsilon) \rightarrow 0$  para cualquier  $\varepsilon > 0$ ) y, por lo tanto, tendremos  $G(\mathbf{P}_n^*) \xrightarrow{p} G(\mathbf{P})$ ,  $P(|G(\mathbf{P}_n^*) - G(\mathbf{P}_1)| > c(n)) \rightarrow 0$  para cada  $\mathbf{P} \neq \mathbf{P}_1$  registrado.

Examinemos ahora algunos criterios de aceptación bien conocidos que son la realización del enfoque descrito anteriormente.

a) *Criterio de Kolmogórov*. Examinemos la estadística (distancia)

$$D(\mathbf{P}_1, \mathbf{P}_n^*) = \sup_t |F_n^*(t) - F(t)|,$$

donde  $F_n^*(t)$  y  $F(t)$  son las funciones de distribución que corresponden a las medidas  $\mathbf{P}_n^*$  y  $\mathbf{P}_1$ . En el § 1.8 hemos establecido que si  $F(t)$  es continua,  $X \in \mathbf{P}_1$ , entonces

$$d_K(\mathbf{P}_1, \mathbf{P}_n^*) \equiv \sqrt{n} D(\mathbf{P}_1, \mathbf{P}_n^*) = \sup_{0 \leq t < 1} |w^0(t)|,$$

donde  $w^0(t)$  es el puente browniano. De aquí se deduce el

**Teorema 1** (A.N. Kolmogórov). *Si  $F(t)$  es continua, entonces existe*

$$\lim_{n \rightarrow \infty} P_1(d_K(\mathbf{P}_1, \mathbf{P}_n^*) < x) = K(x) \equiv P \sup_{0 \leq t < 1} |w^0(t)| < x.$$

La función  $K(x)$  se puede hallar en forma explícita. La misma es igual a

$$K(x) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 x^2}.$$



Con ayuda de este teorema se pueden construir los criterios de nivel asintótico  $1 - \varepsilon$ . La función  $K(x)$  está tabulada en muchos manuales de estadística matemática. Por eso, para  $\varepsilon$  dado podemos, mediante tablas, hallar una constante  $c = c_\varepsilon$  para la cual  $K(c) = 1 - \varepsilon$ . Poniendo  $\pi(X) = 1$  cuando  $d_k(\mathbf{P}_1, \mathbf{P}_n^*) > c_\varepsilon$ , obtenemos el criterio de aceptación de nivel asintótico  $1 - \varepsilon$ . Es fácil notar que el criterio obtenido es conciliable, ya que la funcional  $G(\mathbf{P}) = \sup_t |F_P(t) - F(t)|$  (aquí  $F_P(t) = \mathbf{P}((-\infty, t))$ ), con

cuya ayuda se ha construido el criterio de Kolmogórov, es continua respecto a  $F_P$  en la métrica uniforme y, por consiguiente, es una funcional del tipo II (véase el capítulo para la cual  $G(\mathbf{P}_n^*) \rightarrow G(\mathbf{P})$  cuando  $X \in \mathbf{P}$ ). Queda hacer uso de las observaciones hechas anteriormente sobre las condiciones de conciliabilidad de los criterios de aceptación.

Con ayuda de los resultados del capítulo 1 podemos determinar el comportamiento asintótico de la potencia del criterio de Kolmogórov respecto a alternativas semejantes (véase el § 3). Supongamos que  $X \in \mathbf{P}$ , donde la distribución  $\mathbf{P}$  tiene la función de distribución

$$F_P(x) = F(x) + p(x)n^{-1/2}.$$

Supondremos, para abreviar, que  $p(x)$  es continua, y que  $F(x)$  es continua y estrictamente monótona. La potencia  $\beta(\mathbf{P})$  del criterio de Kolmogórov en el "punto"  $\mathbf{P}$  será igual a

$$\begin{aligned} \beta(\mathbf{P}) = \mathbf{P}(d_k(\mathbf{P}_1, \mathbf{P}_n^*) > c) &= \mathbf{P}\left(\sup_t |F(t) - F_n^*(t)|\sqrt{n} > c\right) \\ &= \mathbf{P}\left(\sup_t |F_P(t) - p(t)n^{-1/2} - F_n^*(t)|\sqrt{n} > c\right). \end{aligned}$$

Si sustituimos  $t = F_P^{-1}(u)$ , donde  $F_P^{-1}$  es una función inversa a  $F_P$ , entonces obtenemos la expresión

$$\mathbf{P}\left(\sup_{0 \leq u \leq 1} |u - p(F_P^{-1}(u))n^{-1/2} - F_n^*(F_P^{-1}(u))|\sqrt{n} - c\right). \quad (4)$$

Aquí  $U_n^*(u) = F_n^*(F_P^{-1}(u))$  es una función empírica que corresponde a la distribución  $U_{0,1}$  uniforme en  $[0, 1]$ , así que (4) es igual a

$$\mathbf{P}\left(\sup_{0 \leq u \leq 1} |u - U_n^*(u) - p(F_P^{-1}(u))n^{-1/2}|\sqrt{n} - c\right).$$

Además  $F_P^{-1}(u) \rightarrow F^{-1}(u)$  en virtud de la estricta monotonía de  $F$ . De aquí y de la continuidad de  $p$  se desprende que

$$\lim_{n \rightarrow \infty} \beta(\mathbf{P}) = \mathbf{P}\left(\sup_{0 \leq t \leq 1} |w^0(t) - a(t)| > c\right), \text{ donde } a(t) = p(F^{-1}(t)). \quad (5)$$

Se puede mostrar que esta expresión es mínima cuando  $a(t) \equiv 0$  ( $p = 0$ ). En este sentido el criterio de Kolmogórov es un criterio no desplazado asintóticamente.

b) *Criterio de Mises—Smirnov (criterio  $\omega^2$ )*. Examinemos, en calidad de distancia entre  $\mathbf{P}_1$  y  $\mathbf{P}_n^*$ , la estadística

$$\omega_n^2 = d_{\omega^2}(\mathbf{P}_1, \mathbf{P}_n^*) = n \int (F(x) - F_n^*(x))^2 dF(x),$$

con cuya ayuda también es posible construir el criterio de aceptación de un nivel dado. En el capítulo 1 hemos demostrado que aquí, al igual que en el caso precedente, es válido el

**Teorema 2.** *Existe la distribución límite*

$$\lim_{n \rightarrow \infty} P_1(\omega_n^2 < x) = \Omega(x) \equiv P\left(\int_0^1 (w^\circ(t))^2 dt < x\right).$$

La función  $\Omega(x)$  tiene una forma muy compleja (véase [8]) y aquí no la mostraremos.

Como la funcional

$$G(\mathbf{P}) = \int (F(t) - F_P(t))^2 dF(t)$$

es una funcional del tipo II (§ 1.3), entonces, conforme a las mismas consideraciones que en el punto a), el criterio  $\omega^2$  es conciliable.

Siguiendo los razonamientos del punto anterior, también se puede establecer el comportamiento asintótico de la potencia  $\beta(\mathbf{P})$  del criterio  $\omega^2$  para las alternativas semejantes de  $\mathbf{P}$  de forma (3). De un modo absolutamente análogo obtenemos que

$$\beta(\mathbf{P}) = P(\omega_n^2 > c) \rightarrow P\left(\int_0^1 (w^\circ(t) - a(t))^2 dt > c\right),$$

donde  $a(t)$  está definida en (5). El valor límite obtenido es, al igual que en (5), mínimo para  $a(t) \equiv 0$ , así que el criterio  $\omega^2$  también es un criterio no desplazado asintóticamente.

Los dos criterios examinados, al igual que otros criterios de aceptación de la hipótesis  $H_1 = \{X \in P_1\}$ , contruidos con ayuda de las distancias  $d(\mathbf{P}, \mathbf{Q})$ , permiten obtener inmediatamente *conjuntos confidenciales* para la función desconocida de distribución  $F(x)$  o para la distribución desconocida  $P_1$  de la muestra  $X$ . En efecto, la relación (1) (ó (2)) también puede ser interpretada así: la probabilidad de que el  $c$ -entorno del "punto"  $P_n^*$  (en sentido de la distancia  $d$ ) recubra el "punto"  $P_1$  es igual a  $1 - \varepsilon$ . (Para (2) obtendremos la variante asintótica de esta afirmación). Ello significa (véase el § 8) que el  $c$ -entorno del punto  $P_n^*$  no es más que un conjunto confidencial de nivel  $1 - \varepsilon$  para la distribución desconocida  $P_1$ ,  $X \in P_1$ . El criterio de Kolmogórov, por ejemplo, determina tal entorno en términos de las funciones de distribución: el mismo es el conjunto de todas  $F(x)$  para las cuales

$$\sup_t |F(t) - F_n^*(t)| \leq c_\varepsilon / \sqrt{n},$$

donde  $c_\varepsilon$  se deduce de (1).

Volvamos a examinar los criterios. Ya hemos señalado que en los niveles asintóticos de significación podemos confiar únicamente cuando son grandes los valores de  $n$ . Pero si el volumen de la muestra no es grande, entonces, al construir el criterio (mejor dicho, al determinar  $c = c_\varepsilon$ ) es necesario utilizar las fórmulas exactas para la distribución de  $d(P_1, P_n^*)$ . No obstante, la obtención de tales fórmulas choca, por lo general, con grandes dificultades. En este sentido desempeñan un papel muy importante los llamados criterios *no paramétricos*, basados en estadísticas cuya distribución no de-

pende de la distribución verdadera  $\mathbf{P}_1$  (o no depende del parámetro  $\theta$  cuando  $X \in \mathbf{P}_\theta$ ).

En este caso, las probabilidades  $\mathbf{P}_1(d(\mathbf{P}_1, \mathbf{P}_n^*) < x)$  no dependen de  $\mathbf{P}_1$  y, por consiguiente, es posible realizarlos una sola vez, hacer las tablas y utilizarlas posteriormente para cualesquiera  $\mathbf{P}_1$ .

*El criterio de Kolmogórov y el criterio  $\omega^2$  no son paramétricos.* Este hecho fue establecido en el § 1.6.

Los criterios no paramétricos también surgen al verificar dos hipótesis compuestas.

c) *Criterio de signos.* Supongamos que  $F(x)$  es la función de distribución para  $\mathbf{P}_1$ , y que la hipótesis  $H_1$  consiste en que  $F(a) = p$  para un punto  $a$  dado. Esta es, evidentemente, una hipótesis compuesta. La hipótesis  $H_2$  es suplementaria:  $H_2 = \{X \in \mathbf{P}, F_p(a) \neq p\}$ . En este caso es natural hacer uso de la estadística siguiente: designemos por  $\nu(X)$  el número de observaciones  $x_i$  para las cuales el signo de diferencia  $x_i - a$  es negativo. En calidad del conjunto crítico  $\Omega$  examinaremos todas las muestras  $X$  para las cuales

$$\nu(X) \notin (c_1, c_2)$$

con ciertos  $c_1 < c_2$ .

Si la hipótesis  $H_1$  es verdadera, entonces

$$\mathbf{P}_1(\nu(X) = k) = C_n^k p^k (1-p)^{n-k}.$$

Así pues, para el caso de la hipótesis  $H_1$ , la distribución  $\nu(X)$  no depende de  $\mathbf{P}_1$ , ya que nuestro criterio no es paramétrico. Los números  $c_i$  han de elegirse de modo que

$$\mathbf{P}(\nu(X) \in (c_1, c_2)) \geq 1 - \varepsilon$$

(debido al carácter discreto de  $\nu(X)$ , aquí puede ser que no se alcance el signo de igualdad). La heterogeneidad en la elección de  $c_i$  se puede eliminar exigiendo el no desplazamiento respecto a los cambios de  $c$ . En general, este problema es equivalente a la verificación de la hipótesis acerca de que la probabilidad de éxito en el esquema de Bernoulli es igual a  $p$ . Análogamente se pueden construir los criterios "unilaterales" para verificar las hipótesis de que  $F(a) \leq p$ .

Si en calidad de generalización del problema examinado verificamos la hipótesis  $F(a_i) = p_i$ ,  $i = 1, \dots, r$  para los valores dados de  $a_i$  y  $p_i$ , llegaremos al criterio  $\chi^2$  que hemos examinado detalladamente en el § 16.

d) *Criterio de Morán.* Así se llama el siguiente criterio para verificar la hipótesis de que  $X \in \mathbf{P}_1$ . Sea  $x_{(1)}, \dots, x_{(n)}$  una serie variacional construida según la muestra  $X$ . Supongamos que  $\mathbf{P}_1$  tiene una función continua de distribución  $F$ , establezcamos la estadística

$$M_n = \sum_{k=0}^n [F(x_{(k+1)}) - F(x_{(k)})]^2, \quad (6)$$

donde se adopta  $F(x_{(0)}) = 0, F(x_{(n+1)}) = 1$ . El criterio de Morán rechaza la hipótesis  $\{X \in P_1\}$  si  $M_n > c$ .

Evidentemente, este parámetro no es paramétrico, ya que  $F(x_k) \in U_{0,1}$ . Por lo tanto es suficiente examinar el criterio  $M_n > c$  basado en la estadística

$$M_n = \sum_{k=0}^n (x_{(k+1)} - x_{(k)})^2$$

y destinado a verificar la uniformidad de la distribución de  $X$ . En este caso, la utilización de la estadística  $M_n$  es natural, ya que la magnitud  $\sum_{i=1}^n y_i^2$  alcanza su mínimo a condición de que  $\sum_{i=1}^n y_i = 1$  en el punto  $y_1 = \dots = y_n = 1/n$ .

Para calcular el nivel asintótico del criterio de Morán puede servir la afirmación siguiente:

**Teorema 3.** Si  $X \in P_1$ , entonces

$$\sqrt{n}(nM_n/2 - 1) \in \Phi_{0,1}.$$

**Demostración.** Supongamos que  $\xi_j \in \Gamma_{\alpha, j}, j = 1, 2, \dots$ . Entonces  $\zeta_k = \sum_{j=1}^k \xi_j \in \Pi_{\alpha, k}$  y, en virtud del corolario 1.6.2, la distribución compatible de las diferencias

$$X_{(1)}, X_{(2)} - X_{(1)}, \dots, X_{(n)} - X_{(n-1)}, 1 - X_{(n)}$$

coincide con la distribución compatible

$$\frac{\xi_1}{\zeta_{n+1}}, \frac{\xi_2}{\zeta_{n+1}}, \dots, \frac{\xi_{n+1}}{\zeta_{n+1}},$$

así que<sup>\*)</sup>

$$M_n = \sum_{j=1}^{n+1} \frac{\xi_j^2}{\zeta_{n+1}^2}$$

La distribución de  $M_n$  no depende de  $\alpha$ , y se puede poner  $\alpha = 1$ . Entonces (véase el § 2.2)

$$M\xi_j^k = \Gamma(k+1) = k!, \quad D\xi_j = 1, \quad D\xi_j^2 = 20,$$

$$\varrho_n \equiv \frac{1}{\sqrt{n}} \sum_{j=1}^n (\xi_j - 1) \in \Phi_{0,1},$$

$$\eta_n \equiv \frac{1}{\sqrt{n}} \sum_{j=1}^n (\xi_j^2 - 2) \in \Phi_{0,20},$$

<sup>\*)</sup> El signo  $\underset{d}{=}$  significa la coincidencia de las distribuciones.

Tenemos

$$nM_{n-1} = \frac{n \left[ 2n + \sum_{j=1}^n (\xi_j^2 - 2) \right]}{\left[ n + \sum_{j=1}^n (\xi_j - 1) \right]^2} = \frac{n(2n + \eta_n \sqrt{n})}{(n + \varrho_n \sqrt{n})^2} = \frac{2 + \eta_n n^{-1/2}}{(1 + \varrho_n n^{-1/2})^2},$$

$$(nM_{n-1} - 2)\sqrt{n} = \frac{\eta_n - 4\varrho_n - 2\varrho_n^2 n^{-1/2}}{(1 + \varrho_n n^{-1/2})^2}. \quad (7)$$

Aquí

$$\eta_n - 4\varrho_n = \frac{1}{\sqrt{n}} \sum_{j=1}^n (\xi_j^2 + 2), \quad \xi_j^2 = \xi_j^2 - 4\xi_j,$$

$$\mathbf{M}\xi_j' = -2, \quad \mathbf{D}\xi_j' = \mathbf{M}(\xi_j^4 - 8\xi_j^3 + 16\xi_j^2) - 4 = 4.$$

Por lo tanto,  $\eta_n - 4\varrho_n \in \Phi_{0,4}$ , así que, en virtud de los teoremas de continuidad de (7), obtenemos

$$\sqrt{n}(nM_{n-1}/2 - 1) \in \Phi_{0,1}.$$

Esto equivale a la afirmación del teorema.  $\triangleleft$ .

Citemos ahora las consideraciones que muestran que el criterio de Morán es conciliable. Examinemos la estadística (6) para  $X \in \mathbf{P}$ , donde  $\mathbf{P}$  se distingue de  $\mathbf{P}_1$ . Una de las distribuciones  $\mathbf{P}_1$  o  $\mathbf{P}$  puede considerarse, sin limitar la generalidad, uniforme. Supongamos que ella será  $\mathbf{P}$ . Con respecto a  $F$  podemos suponer, para abreviar, que existe una densidad continua  $f(t) = F'(t)$  concentrada en  $[0, 1]$ . Entonces, para  $X \in U_{0,1}$ , la parte principal de  $nM_n$  será igual a

$$n \sum_{k=0}^n [f(x_{(k+1)}(x_{(k+1)} - x_{(k)})]^2 = \frac{n+1}{n} \sum_{k=1}^{n+1} [f(\xi_k/\xi_{n+1})\xi_k/\xi_{n+1}]^2. \quad (8)$$

Según la ley fuerte de los grandes números,  $k^{-1}\xi_k \xrightarrow{c.s.} 1$  cuando  $k \rightarrow \infty$ . Por eso, a su vez, la parte principal de (8) será igual a

$$\sum_{k=1}^n f^2(k/n) \xi_k^2/n.$$

Volviendo a utilizar la ley de los grandes números (o la desigualdad de Chébishev), obtenemos que esta expresión converge, en probabilidad, hacia

$$2 \int_0^1 f^2(t) dt \geq 2 \left( \int_0^1 f(t) dt \right)^2 = 2.$$

Aquí el signo de desigualdad es estricto cuando  $f(t) \neq 1$ . Esto quiere decir que cuando  $X \in \mathbf{P} = U_{0,1} \neq \mathbf{P}_1$  y cuando  $n \rightarrow \infty$ ,

$$\sqrt{n}(nM_n/2 - 1) \xrightarrow{\mathbf{P}} \infty,$$

lo cual conduce, en virtud del teorema 3, a la conciliabilidad del criterio de Morán de cualquier nivel registrado  $1 - \varepsilon$ .  $\triangleleft$

Siendo conciliable, el criterio de Morán *no distingue*, sin embargo, *las hipótesis afines*. Supongamos que  $X \in \mathbf{P} = U_{0,1}$ ,

$$\begin{aligned} F(t) &= t + p(t)n^{-1/2}, \quad t \in [0, 1], \\ p(0) &= p(1) = 0, \end{aligned} \tag{10}$$

y que la función  $p(t)$  es continuamente derivable. Entonces

$$\begin{aligned} n^{3/2}M_n &= n^{3/2} \sum_{k=0}^n (x_{(k+1)} - x_{(k)})^2 + 2n \sum_{k=0}^n (x_{(k+1)} - x_{(k)}) \times \\ &\times (p(x_{(k+1)}) - p(x_{(k)})) + \sqrt{n} \sum_{k=0}^n p(x_{(k+1)}) - p(x_{(k)})^2. \end{aligned} \tag{11}$$

La parte principal de la segunda suma aquí es igual a  $2n \sum_{k=0}^n p'(x_{(k)})(x_{(k+1)} - x_{(k)})^2$ , o bien, en virtud de las mismas consideraciones que en (9),

$$2 \sum_{k=1}^n p'(k/n) \xi_k^2/n \xrightarrow{P} 4 \int_0^1 p'(t) dt = 0.$$

El último sumando en (11) también converge (en probabilidad) a cero, ya que su parte principal coincide (en distribución) con

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n [p'(k/n)]^2 \xi_k^2/n,$$

o con  $\frac{2}{\sqrt{n}} \int_0^1 [p'(t)]^2 dt \rightarrow 0$ . Lo dicho significa que para la función  $F$  en forma de (10), la

estadística  $n^{3/2}/M_n/2 - \sqrt{n}$  tendrá la misma distribución límite de  $\Phi_{0,1}$  que para  $F(t) = t$ .  $\triangleleft$

Conviene señalar que de este hecho no se deben sacar conclusiones apresuradas de que el criterio de Morán es malo. La cosa consiste en que, sin distinguir las hipótesis afines de forma (10), el criterio de Morán distingue otras hipótesis (que son, en cierto sentido, también afines) las cuales no pueden ser distinguidas por otros criterios examinados en este párrafo. Se trata de las *hipótesis para las densidades*.

Examinemos la hipótesis  $H_2 = \{X \in \mathbf{P}\}$ , donde la distribución  $\mathbf{P}$  tiene una densidad de

$$f(t) = \begin{cases} 2 & \text{cuando } 2k\Delta_n \leq t < (2k+1)\Delta_n, \\ 0 & \text{cuando } (2k+1)\Delta_n \leq t < (2k+2)\Delta_n, \end{cases} \quad k = 0, 1, \dots, N-1,$$

donde  $\Delta_n = \frac{1}{2N}$ ,  $N = N_n > 0$  es un número entero. Entonces, para  $\Delta_n = 0(n^{-1/2})$ , la función de distribución  $F_{\mathbf{P}}(t)$ , correspondiente a la distribución  $\mathbf{P}$ , poseerá la propiedad

$$\sup_t |F_{\mathbf{P}}(t) - t| = 0(n^{-1/2}).$$

Esto quiere decir que la hipótesis  $H_2$  como hipótesis para la función de distribución será tan próxima a  $H_1 = \{X \in U_{0,1}\}$ , que los criterios de Kolmogórov y  $\omega^2$  no las distinguirán (el valor límite de la potencia en el punto  $\mathbf{P}$  coincidirá con el nivel límite del criterio). No obstante, como hipótesis para las *densidades*, las hipótesis  $H_1$  y  $H_2$  se distinguen considerablemente, ya que  $\sup |f(t) - 1| = 1$ . Como  $x_{(0)} = 0$ ,  $x_{(n+1)} = 1$ , para  $X \in \mathbf{P}$  la estadística  $M_n$  superará la magnitud  $\Delta_n^2 N = \Delta_n/2$ . Por consiguiente, si  $n/N = 2n\Delta_n \rightarrow \infty$  cuando la

P-probabilidad es igual a 1, tendremos

$$nM_n \rightarrow \infty.$$

Fijando el conjunto crítico  $\Omega_2 = \{nM_n > 3\}$  obtendremos  $P_1(\Omega_2) \rightarrow 0$ . Esto significa que cuando  $\Delta_n = O(n^{-1/2})$ ,  $\Delta_n n \rightarrow \infty$ , el criterio de Morán distinguirá las hipótesis  $H_1$  y  $H_2$  con una probabilidad próxima a 1. Con otras palabras, la estadística  $M_n$  es sensible a las desviaciones relacionadas con la densidad, y el propio criterio de Morán puede ser recomendado como criterio para verificar las hipótesis referentes a las densidades. Por otro lado, del § 1.10 sabemos que la velocidad con que las densidades empíricas se aproximan a la densidad verdadera es inferior a  $n^{-1/2}$ . Por eso, la "indistinguibilidad" de las hipótesis de las densidades que difieren una de otra en orden de  $n^{-1/2}$  (véase (10)) no debe causar sorpresa.

De acuerdo con el criterio de Morán y con algunos otros criterios examinados anteriormente, se puede hacer una observación general. Si se comparan dos criterios de un mismo nivel registrado, el primero de los cuales está destinado al rechazamiento de mayor número de alternativas que el segundo, la potencia del primer criterio para cada alternativa registrada rechazada por ambos criterios) será, por lo general, menor que la potencia del segundo. A título de ejemplo elemental que ilustra esta circunstancia, el lector puede examinar los criterios  $|x_1| > \lambda_{\varepsilon/2}$  y  $x_1 > \lambda_{\varepsilon}$  destinados a verificar, respectivamente, las hipótesis  $\{\alpha \neq 0\}$  y  $\{\alpha > 0\}$  frente a  $\{\alpha = 0\}$ , basándose en la observación  $x_1 \in \Phi_{\alpha,1}$ . Aquí  $\lambda_{\varepsilon}$  es la cuantía de distribución  $\Phi_{0,1}$  de orden  $1 - \varepsilon$ . Las potencias en el punto  $\alpha > 0$  serán iguales a

$$1 - \Phi_{0,1}(-\lambda_{\varepsilon+2} - \alpha), \quad \lambda_{\varepsilon+2} - \alpha < 1 - \Phi(\lambda_{\varepsilon} - \alpha),$$

respectivamente.

### § 13. Criterios asintóticamente óptimos.

#### Criterio de la relación de verosimilitud como criterio asintóticamente bayesiano para verificar una hipótesis simple frente a otra compuesta

**1. Propiedades asintóticas del c.r.v. y del criterio bayesiano.** Examinemos el problema de verificación de una hipótesis simple  $H_1 = \{X \in P_{\theta_1}\}$  frente a la hipótesis alternativa  $H_2 = \{X \in P_{\theta}; \theta \neq \theta_1, \theta \in \Theta\}$ . En los párrafos precedentes hemos visto, en ejemplos, que en este caso el c.u.m.p. no existe, por lo general.

Vamos a examinar el planteamiento "parcialmente bayesiano" del problema que hemos descrito en los §§ 4 y 9. El mismo consiste en la suposición de que  $\theta$  es escogido en  $\Theta_2 = \Theta \setminus \{\theta_1\}$  al azar, con una distribución  $Q_2 = Q$ . Se puede considerar que  $Q$  se da en  $\Theta$ ,  $Q(\{\theta_1\}) = 0$ . En este caso la distribución de la muestra  $X$  se definirá por la densidad "mediada"

$$f_{Q_2}(x) = \int f_i(x) Q(dt). \quad (1)$$

Ahora bien, si se conoce  $Q$ , entonces la hipótesis  $H_{Q_2} = H_Q$ , en virtud de la cual  $X$  tiene una distribución de densidad (1), puede considerarse, junto con  $H_1$ , como hipótesis simple, y para la construcción del criterio más potente se puede utilizar el lema de Neumann — Pearson.

Resulta que en este caso para "casi todas" las  $Q$  suaves, los criterios más potentes coincidirán asintóticamente con el criterio de la relación de

verosimilitud

$$R(X) = \frac{\sup_{\theta \in \Theta} f_{\theta}(X)}{f_{\theta_1}(X)} = \frac{f_{\hat{\theta}}(X)}{f_{\theta_1}(X)} > c \quad (2)$$

y, por consiguiente, no dependerán de  $\mathbf{Q}$ . Este hecho permite considerar como óptimo el criterio hallado al menos en los casos en que se puede suponer que  $\theta$  en  $\Theta_2$  se escoge aleatoriamente, pero desconocemos su distribución  $\mathbf{Q}$ .

Antes de enunciar el teorema respectivo recordemos algunos resultados que necesitamos y demostremos una afirmación auxiliar. En ella desempeñarán un papel muy importante las propiedades asintóticas conocidas de la relación de verosimilitud. Vamos a examinar inmediatamente el caso del parámetro multidimensional; todo lo necesario para esto se contiene en los §§ 2.28 y 2.29.

Así pues, supongamos que  $\theta \in \Theta \subset R^k$ ,  $k \geq 1$ , y que se cumplen las condiciones de regularidad (RR) cuya enunciación se da en el § 2.28. Supongamos, además, que  $\mathbf{Q}$  tiene una densidad  $q(t)$  respecto a la medida de Lebesgue  $\lambda(dt) = dt$ .

Según el lema de Neumann — Pearson, el criterio no randomizado más potente  $\pi_{Q_2} = \pi_Q$  para verificar  $H_1$  frente a  $H_Q$  tendrá la forma siguiente:  $\pi_Q(X) = 1$  si

$$X \in \Omega(c) = \left\{ x : \frac{f_Q(x)}{f_{\theta_1}(x)} > c \right\}, \quad f_Q(x) = \int q(u) f_u(x) du, \quad (3)$$

donde escogeremos  $c = c_n$  más tarde, según el nivel dado del criterio.

Los criterios bayesianos para verificar  $H_1$  frente a  $H_Q$  también tendrán la misma forma.

Las probabilidades de los errores de primero y segundo género son iguales a

$$\alpha_1(\pi_Q) = \mathbf{P}_{\theta_1} \left( \frac{f_Q(X)}{f_{\theta_1}(X)} > c \right), \quad 1 - \beta(\pi_Q) = \int q(t) \mathbf{P}_t \left( \frac{f_Q(X)}{f_{\theta_1}(X)} \leq c \right) dt, \quad (4)$$

respectivamente, donde  $\beta(\pi_Q) = \int_{\{f_Q(x) \leq cf_{\theta_1}(x)\}} f_Q(x) \mu^n(dx)$  es la potencia del criterio más potente.

Podemos escribir las expresiones análogas para el c.r.v.  $\pi$  que acepta  $H_Q$  si se cumple (2):

$$\alpha_1(\hat{\pi}) = \mathbf{P}_{\theta_1} \left( \frac{f_{\hat{\theta}}(X)}{f_{\theta_1}(X)} > c \right),$$

$$\alpha_2(\pi) = \int q(t) \mathbf{P}_t \left( \frac{f_{\hat{\theta}}(X)}{f_{\theta_1}(X)} \leq c \right) dt = \int_{\{f_{\hat{\theta}}(x) \leq cf_{\theta_1}(x)\}} f_Q(x) \mu^n(dx). \quad (5)$$



Pongamos  $I = I(\theta_1)$  (el valor de la matriz de información de Fisher en el punto  $\theta_1$ )

$$\frac{f_Q(X)}{f_{\theta_1}(X)} = \left(\frac{2\pi}{n}\right)^{k/2} \frac{q(\theta_1)}{\sqrt{|I|}} e^{T(X)},$$

$$\frac{f_{\hat{\theta}}(X)}{f_{\theta_1}(X)} = e^{T(X)}. \quad (6)$$

Entonces las regiones críticas de los criterios  $\pi_Q$  y  $\hat{\pi}$  (véanse (3) y (2)) pueden escribirse, respectivamente, en la forma

$$T(X) > c_Q, \quad \hat{T}(X) > \hat{c}. \quad (7)$$

**Lema 1.** *Supongamos que se cumplen las condiciones (RR) del § 2.28,  $X \in \mathbf{P}_{\theta_1}$ , y que  $\theta_1$  es el punto interior de  $\Theta$ . Entonces*

$$2T(X) = 2\hat{T}(X)(1 + \varepsilon_n(X)) \in H_k, \quad \varepsilon_n(X) \xrightarrow{c.s.} 0.$$

**Demostración.** La afirmación del lema es el corolario evidente de los teoremas 2.28.4 y 2.28.5. Sólo debemos señalar que  $\hat{T}(X)$  en las designaciones del teorema 2.28.4 no es otra cosa sino  $Y(u^*)$  (cuando  $\theta = \theta_1$ ). <

## 2. Carácter asintóticamente bayesiano del c.r.v.

Pasemos a enunciar la afirmación fundamental. Recordemos que cuando estudiamos las propiedades asintóticas de los criterios, en realidad tenemos presente no uno sino toda la sucesión de los criterios  $\pi = \pi_n$ , donde  $\pi_n$  es el criterio basado en la muestra  $X_n$ . Teníamos la misma situación al examinar las propiedades asintóticas de las estimaciones. Ahora bien, aquí y en lo sucesivo, siempre que esto sea necesario, por criterio  $\pi$  entendemos la sucesión de las funciones  $\pi_n(X_n)$  definidas para cada  $n$  y  $X_n = [X_\infty]_n$ .

**Definición 1.** *El criterio  $\pi$  para verificar la hipótesis  $H_1 = \{\theta \in \Theta_1\}$  frente a  $H_2 = \{\theta \in \Theta_2\}$  pertenece a la clase  $\tilde{K}_\varepsilon$  de los criterios de nivel asintótico  $1 - \varepsilon$  si*

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in \Theta_1} \mathbf{M}_\theta \pi(X) \leq \varepsilon. \quad (8)$$

En nuestro caso, cuando la hipótesis  $H_1$  es simple y  $\Theta_1 = \{\theta_1\}$ , la relación (8) se transforma en desigualdad:

$$\limsup_{n \rightarrow \infty} \mathbf{M}_{\theta_1} \pi(X) \leq \varepsilon.$$

Sea  $h_\varepsilon$  una cuantila de orden  $1 - \varepsilon$  de la distribución  $\lambda^2$  de  $k$  grados de libertad ( $\mathbf{H}_k((h_\varepsilon, \infty)) = \varepsilon$ ). Entonces, del lema 1 se desprende que  $\pi_Q \in \tilde{K}_\varepsilon$ ,  $\hat{\pi} \in \tilde{K}_\varepsilon$  si  $c_Q = \hat{c} = h_\varepsilon/2$ .

**Definición 2.** Pongamos  $c_Q = h_\varepsilon/2$ , de modo que  $\pi_Q \in \tilde{K}_\varepsilon$ . El criterio  $\pi \in \tilde{K}_\varepsilon$  se denomina *criterio asintóticamente bayesiano (c.a.b.)* en  $\tilde{K}_\varepsilon$  para verificar la hipótesis  $H_1 = \{\theta = \theta_1\}$  frente a  $H_Q$  si para las probabilidades

de los errores de segundo género, calculadas para la hipótesis  $H_Q$ , es válida la relación

$$\limsup_{n \rightarrow \infty} \frac{\alpha_2(\pi)}{\alpha_2(\pi_Q)} = \limsup_{n \rightarrow \infty} \frac{1 - \beta(\pi)}{1 - \beta(\pi_Q)} = \limsup_{n \rightarrow \infty} \frac{M_Q(1 - \pi(X))}{M_Q(1 - \pi_Q(X))} = 1.$$

En esta definición hemos utilizado la relación (y no la diferencia) de las probabilidades de los errores de segundo género, ya que  $\alpha_2(\pi_Q) \rightarrow 0$  cuando  $n \rightarrow \infty$ .

**Teorema 1.** *Supongamos que se cumplen las condiciones (RR) y que el punto  $\theta_1$  es un punto interior de  $\Theta$ . Entonces el criterio de la relación de verosimilitud  $\hat{\pi}$  (véanse (2) y (7)) para  $\hat{c} = h_e/2$  pertenece a  $\tilde{K}_e$  y es el c.a.b. en  $\tilde{K}_e$  para verificar  $H_1$  frente a  $H_Q$ , cualquiera que sea la distribución  $Q$  cuya densidad  $q(t)$  es continua y positiva en  $\Theta$ . En este caso*

$$\alpha_2(\hat{\pi}) \sim \alpha_2(\pi_Q) \sim \frac{q(\theta_1)}{n^{k/2} \sqrt{|I|}} V_k h_e^{k/2},$$

donde  $I = I(\theta_1)$ ,  $V_k$  es el volumen de la esfera unitaria en  $R^k$ .

**Demostración.** Ya hemos demostrado la pertenencia de  $\hat{\pi} \in \tilde{K}_e$  cuando  $\hat{c} = h_e/2$ . Examinemos ahora los errores de segundo género. En virtud de (4) y (7) tenemos

$$\begin{aligned} \alpha_2(\pi_Q) &= \int_{\{T(X) \leq c_Q\}} f_Q(x) \mu^n(dx) = M_{\theta_1} \left\{ \frac{f_Q(X)}{f_{\theta_1}(X)}; 2T(X) \leq h_e \right\} = \\ &= \left( \frac{2\pi}{n} \right)^{k/2} \frac{q(\theta_1)}{\sqrt{|I|}} M_{\theta_1} \{ e^{T(X)}; 2T(X) \leq h_e \}. \end{aligned}$$

Aquí, bajo el signo de esperanza matemática se encuentra la función limitada de  $2T$  que es casi por doquier continua respecto a la distribución límite ( $H_k$ ). Por eso, cuando  $n \rightarrow \infty$ ,  $\chi_k^2 \in H_k$ ,

$$\begin{aligned} M_{\theta_1}(e^{T(X)}; 2T(X) \leq h_e) &\rightarrow M \left\{ e^{\frac{1}{2}\chi_k^2}; \chi_k^2 \leq h_e \right\} = \\ &= (2\pi)^{-k/2} \int_{\{|y|^2 \leq h_e\}} e^{\frac{1}{2}|y|^2 - \frac{1}{2}|y|^2} dy_1 \dots dy_k = (2\pi)^{-k/2} h_e^{k/2} V_k. \end{aligned}$$

Determinemos ahora el comportamiento asintótico de  $\alpha_2(\hat{\pi})$ . Designemos  $A_n = \{X: \pi_Q \neq \hat{\pi}\}$ . En virtud del lema 1  $P_{\theta_1}(A_n) \rightarrow 0$ . Por eso, del teorema 2.29.5 se deduce que para cualquier  $N$  registrado,

$$\sup_{|n| \leq N} P_{\theta + \mu\sqrt{n}}(A_n) \rightarrow 0. \tag{9}$$

Hagamos uso de la representación (véase (5))

$$\begin{aligned} \alpha_2(\hat{\pi}) &= \int q(t) \mathbf{P}_t(\hat{T}(X) \leq c) dt = \\ &= \int_{|t-\theta_1| \leq N/\sqrt{n}} + \int_{|t-\theta_1| > N/\sqrt{n}} \leq \int q(t) \mathbf{P}_t(T(X) \leq c) dt + \\ &\quad + \int_{|t-\theta_1| \leq N/\sqrt{n}} q(t) \mathbf{P}_t(A_n) dt + \int_{|t-\theta_1| > N/\sqrt{n}} q(t) \mathbf{P}_t(\hat{T}(X) \leq c) dt. \end{aligned}$$

En virtud de (9) obtenemos

$$\begin{aligned} \limsup_n n^{k/2} \alpha_2(\hat{\pi}) &\leq \lim_{n \rightarrow \infty} n^{k/2} \alpha_2(\pi_Q) + \\ &\quad + \max_t q(t) \cdot \limsup_{n \rightarrow \infty} \int_{|t-\theta_1| > N/\sqrt{n}} \mathbf{P}_t\left(\frac{f_{\theta_1}(X)}{f_t(X)} \leq c^\varepsilon\right) dt. \end{aligned}$$

Pero la probabilidad bajo el signo integral no excede

$$\mathbf{P}_t\left(\frac{f_{\theta_1}(X)}{f_t(X)} \geq e^{-\varepsilon}\right) \leq \exp\{\varepsilon/2 - |t - \theta_1|^2 n g/2\}. \quad (10)$$

Aquí hemos utilizado el teorema 2.28.1. Por consiguiente, la propia integral no excede

$$e^{\varepsilon/2} \int_{|u| > N} e^{-|u|^2 g/2} du \rightarrow 0$$

cuando  $N \rightarrow \infty$ . De aquí se deduce que

$$\limsup_{n \rightarrow \infty} n^{k/2} \alpha_2(\hat{\pi}) \leq \lim_{n \rightarrow \infty} n^{k/2} \alpha_2(\pi_Q). \quad (11)$$

Es evidente que esto equivale a que  $\hat{\pi}$  es el c.a.b.

Sólo queda determinar que  $\alpha_2(\hat{\pi}) \sim \alpha_2(\pi_Q)$  o que, también en virtud de (11),

$$\liminf_{n \rightarrow \infty} n^{k/2} \alpha_2(\hat{\pi}) \geq \lim_{n \rightarrow \infty} n^{k/2} \alpha_2(\pi_Q). \quad (12)$$

Para esto, nótese que el criterio  $\pi_Q$  construido es bayesiano y corresponde a la probabilidad a priori  $q_1$  de la hipótesis  $H_1$  que se define por la ecuación (compárense (3) y (6))

$$\frac{q_1}{1 - q_1} = \left(\frac{2\pi}{n}\right)^{k/2} \frac{q(\theta_1)}{\sqrt{|I|}} e^\varepsilon.$$

Esto quiere decir que la probabilidad del error  $\pi_Q$  se comportará asintóticamente como

$$\varepsilon q_1 + (1 - q_1) \alpha_2(\pi_Q) \sim \varepsilon q_1 + \alpha_2(\pi_Q).$$

Si admitimos que (12) no es cierta, obtenemos el criterio  $\hat{\pi}$  para el cual la probabilidad del error será menor. Como esto no es posible, (12) queda demostrada. El teorema está demostrado por completo.  $\triangleleft$

De los razonamientos citados se deduce que en las probabilidades de los errores de segundo género hacen el aporte principal los valores aleatorios de  $\theta$  que entran en el entorno  $n^{-1/2}$  del punto  $\theta_1$  (con ello se explica el orden de pequeñez  $n^{-k/2}$  de estas probabilidades).

Las modificaciones insuficientes de los razonamientos para la demostración del teorema 1 también permiten obtener la afirmación siguiente.

**Teorema 2.** *Los criterios  $\pi'$  y  $\pi''$  con las regiones críticas*

$$\begin{aligned} \Omega' &= \{x \in \mathcal{X}^n: n(\hat{\theta}^* - \theta_1)I(\theta_1)(\hat{\theta}^* - \theta_1)^T > h_\varepsilon\}, \\ \Omega'' &= \{x \in \mathcal{X}^n: L'(X, \theta_1)I^{-1}(\theta_1)(L'(X, \theta_1))^T > h_\varepsilon\} \end{aligned} \quad (13)$$

son, a la par con  $\hat{\pi}$ , los c.a.b. en  $\hat{K}_\varepsilon$ . Esta propiedad se conserva si  $I(\theta_1)$  en (13) se sustituye por  $I(\hat{\theta}^*)$ .

Los criterios (13) se obtienen si se utiliza el desarrollo

$$\ln \frac{f_{\hat{\theta}^*}(X)}{f_{\theta_1}(X)} = L(X, \hat{\theta}^*) - L(X, \theta_1)$$

en serie cerca del punto  $\hat{\theta}^*$  (véase el teorema 2.28.4). La forma del criterio  $\hat{\pi}$  es, en cierto sentido, más cómoda, ya que no está relacionada con la dimensión.

La demostración del teorema 2 se la concedemos al lector.

En el caso unidimensional, el conjunto crítico  $\Omega'$  (al sustituir  $I(\theta_1)$  por  $I(\hat{\theta}^*)$ ) tiene la forma

$$\Omega' = \left\{ |\hat{\theta}^* - \theta_1| > \left[ \frac{h_\varepsilon}{n|I(\hat{\theta}^*)|} \right]^{1/2} \right\}, \quad (14)$$

donde, evidentemente,  $h_\varepsilon = \lambda_{\varepsilon/2}^2$ ,  $\Phi_{0,1}((-\lambda_{\varepsilon/2}, \lambda_{\varepsilon/2})) = 1 - \varepsilon$ . Vemos que el criterio  $\pi'$  respectivo (14), que equivale asintóticamente a  $\hat{\pi}$ , puede interpretarse así:  $\pi'(X) = 1$  si  $\theta_1$  no ha caído en el intervalo confidencial de nivel asintótico  $1 - \varepsilon$  para el parámetro  $\theta$ , construido con ayuda de la e.v.m.  $\hat{\theta}^*$ .

Esa misma interpretación también se conservará, evidentemente, en el caso multidimensional; además, los conjuntos confidenciales tendrán forma de elipsoides:

$$(\hat{\theta}^* - \theta)I(\hat{\theta}^*)(\hat{\theta}^* - \theta)^T \leq n^{-1}h_\varepsilon.$$

Así pues, vemos que la e.v.m. está estrechamente relacionada con el c.a.b.

**Ejemplo 1.** Supongamos que  $X \in \Pi_\lambda$  y que se verifica la hipótesis  $H_1 = \{\lambda = \lambda_1\}$  frente a  $H_2 = \{\lambda \neq \lambda_1\}$ . En este caso  $\hat{\lambda}^* = \bar{x}^*$ ,  $I(\lambda) = \lambda^{-1}$  y el c.a.b. tendrá la forma

$$(\bar{x} - \lambda_1)^2 > h_\varepsilon \lambda_1 / n,$$

donde  $H_1((h_\varepsilon, \infty)) = \varepsilon$ .

**Ejemplo 2.** Supongamos que  $X \in \Phi_{\alpha, \sigma^2}$  y que se verifica la hipótesis  $H_1 = \{(\alpha, \sigma^2) = (\alpha_1^*, \sigma_1^{*2})\}$  frente a la alternativa adicional. Aquí  $\hat{\alpha}^* = \bar{x}$ ,

$$\hat{\sigma}^{*2} = S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2, \quad I(\alpha, \sigma^2) = \begin{pmatrix} \sigma^{-2} & 0 \\ 0 & \sigma^{-4/2} \end{pmatrix} \quad (\text{véase el § 2.16}).$$

Por eso el c.a.b. tiene la forma

$$\frac{(\bar{x} - \alpha_1)^2}{\sigma_1^2} + \frac{(S^2 - \sigma_1^2)^2}{2\sigma_1^4} > \frac{h_\varepsilon}{n},$$

donde  $H_2((h_\varepsilon, \infty)) = \varepsilon$ .

**3. Carácter de no desplazamiento asintótico del c.r.v.** Concluyendo este párrafo estableceremos que el c.r.v. (2) no está asintóticamente desplazado. Recordemos previamente que el criterio  $\pi$  para verificar  $H_1 = \{\theta \in \Theta_1\}$  frente a  $H_2 = \{\theta \in \Theta_2\}$  se llama criterio no desplazado si

$$\inf_{\theta \in \Theta_2} \mathbf{M}_\theta \pi - \sup_{\theta \in \Theta_1} \mathbf{M}_\theta \pi \geq 0.$$

**Definición 3.** El criterio  $\pi$  se denomina criterio *asintóticamente no desplazado* si

$$\liminf_{n \rightarrow \infty} (\inf_{\theta \in \Theta_2} \mathbf{M}_\theta \pi - \sup_{\theta \in \Theta_1} \mathbf{M}_\theta \pi) \geq 0.$$

**Teorema 3.** El c.r.v.  $\hat{\pi}$  (véase (2), (6) y (7)) para verificar  $H_1 = \{\theta = \theta_1\}$  frente a  $H_2 = \{\theta \neq \theta_1\}$  es un criterio asintóticamente no desplazado.

**Demostración.** Como en nuestro caso  $\Theta_1 = \{\theta_1\}$  y  $\lim_{n \rightarrow \infty} \mathbf{M}_{\theta_1} \hat{\pi} = \varepsilon$ , es suficiente cerciorarse de que

$$\liminf_{n \rightarrow \infty} \inf_{t \in \Theta} \mathbf{M}_t \hat{\pi} = \liminf_{n \rightarrow \infty} \inf_{t \in \Theta} \mathbf{P}_t \left( \frac{f_{\hat{\theta}^*}(X)}{f_{\theta_1}(X)} > e^{\hat{c}} \right) \geq \varepsilon, \quad (15)$$

donde  $\hat{c} = h_\varepsilon/2$ .

De la estimación (10) resulta que existe  $N > 0$  tal, que

$$\inf_{|t - \theta_1| \geq N/\sqrt{n}} \mathbf{P}_t \left( \frac{f_{\hat{\theta}^*}(X)}{f_{\theta_1}(X)} > e^{\hat{c}} \right) > \varepsilon.$$

Queda demostrar que  $\inf_{|t - \theta_1| \leq N/\sqrt{n}} \mathbf{M}_t \hat{\pi} \rightarrow \varepsilon$ .

Pero, en virtud de los teoremas 2.28.4 y 2.29.3, cuando  $X \in \mathbf{P}_t$  uniformemente respecto a  $u$ ,  $|u| \leq N$ ,  $u = \sqrt{n}(t - \theta_1)$ ,

$$\hat{T}(X) = \frac{1}{2}(\xi - u)I(\xi - u)^T, \quad \xi \in \Phi_{0,1}^{-1},$$

$$\mathbf{M}_t(\hat{\pi}) \rightarrow \mathbf{P}_t \left( \frac{1}{2}(\xi - u)I(\xi - u)^T > \hat{c} = h_\varepsilon/2 \right).$$

El segundo miembro aquí alcanza su valor mínimo cuando  $u = 0$ . Este valor es igual a  $\mathbf{P}(\xi/\xi^T > h_\varepsilon) = \varepsilon$ .  $\triangleleft$

### § 14. Criterios asintóticamente óptimos para verificar las hipótesis compuestas semejantes

**1. Planteamiento del problema y definiciones.** En el § 3 hemos estudiado dos enfoques asintóticos del problema de verificación de dos hipótesis simples  $H_1$  y  $H_2$ . Si consideramos estas hipótesis fijas, o sea, invariables para el volumen creciente  $n$  de la muestra  $X_n$ , entonces, al calcular las probabilidades de los errores, llegaremos al problema de las probabilidades de grandes desviaciones, de modo que la probabilidad de uno de los errores, como mínimo, convergerá a cero. De acuerdo con otro enfoque, las hipótesis  $H_1$  y  $H_2$  se consideran como elementos de la sucesión de hipótesis "que se aproximan", en este caso la velocidad de aproximación se escoge de manera que las probabilidades de los errores de primero y segundo género converjan hacia sus propios límites (distintos de 0 y 1). Hemos visto que en el caso paramétrico, los valores del parámetro  $\theta_1$  y  $\theta_2$ , correspondientes a las hipótesis  $H_1$  y  $H_2$ , deben distinguirse en orden de  $n^{-1/2}$ . Cada uno de estos enfoques puede ser justificado conforme a las condiciones concretas.

En el párrafo precedente hemos examinado la distribución  $Q$ , no dependiente de  $n$ , para el valor alternativo de  $\theta$  y, como era natural de esperar, hemos obtenido que la probabilidad de un error de segundo género converge a cero como  $n^{-k/2}$ . Esto se debe al hecho de que a esta probabilidad contribuyen principalmente las hipótesis semejantes para las cuales  $\theta$  está alejado de  $\theta_1$  a una distancia del orden de  $n^{-1/2}$  (el volumen de la región que contiene tales  $\theta$  tendrá precisamente un orden de pequeñez de  $n^{-k/2}$ ).

En este párrafo examinaremos el problema de verificación de las *hipótesis compuestas semejantes*, cuando los valores alternativos del parámetro se aproximan cuando  $n \rightarrow \infty$ . Resulta que en este caso, el problema de verificación de las hipótesis se puede reducir, en cierto sentido, a un problema mucho más simple para la distribución normal.

Pasemos a enunciaciones más exactas. Supongamos que a base de la muestra  $X \in P_\theta$  se comprueba la hipótesis  $H_1 = \{\theta \in \Theta_1\}$  frente a  $H_2 = \{\theta \in \Theta_2\}$ . Fijemos cualquier punto interior  $\theta_1$  del conjunto  $\Theta$  y pongamos

$$\theta = \theta_1 + \gamma n^{-1/2}. \quad (1)$$

Ahora supongamos que el conjunto  $\Theta_i$  tiene la forma

$$\Theta_i = \theta_1 + \Gamma_i n^{-1/2}, \quad (2)$$

donde  $\Gamma_i$  no dependen de  $n$ . La notación (2) significa que  $\theta \in \Theta_i$  sí y sólo sí en (1)  $\gamma \in \Gamma_i$ . Las hipótesis  $H_i = \{\theta \in \Theta_i\}$  para la condición (1) serán llamadas, al igual que en el § 3, hipótesis *semejantes* (en realidad son una sucesión de hipótesis propias de cada  $n$ ).

El problema de verificación de las hipótesis semejantes  $H_i$  a base de la muestra  $X \in \mathbf{P}_\theta$  se llamará *problema A*.

Examinemos ahora otro problema. Sea  $Y \in \Phi_{\gamma, I^{-1}}$  una muestra de volumen unitario de la población normal  $\Phi_{\gamma, I^{-1}}$  con un vector de valores medios  $\gamma$  y con una matriz de segundos momentos  $I^{-1} = I^{-1}(\theta_1)$ , donde  $I(\theta_1)$  es la matriz de información de Fisher para el problema *A* en el punto  $\theta_1$ . Designemos por  $h_i$  las hipótesis  $\{\gamma \in \Gamma_i\}$ . El problema de verificación de las hipótesis  $h_i$  a base de una sola observación  $Y \in \Phi_{\gamma, I^{-1}}$  se denominará *problema B*.

El hecho extraordinario que permite realizar la reducción antes mencionada consiste, aproximadamente, en lo siguiente. Sea  $\pi(Y)$  el criterio óptimo en uno u otro sentido (el c.u.m.p., el criterio bayesiano o el criterio minimax) para verificar  $h_1$  frente a  $h_2$  en el problema *B*. Y sea  $\hat{\theta}^*$ , como siempre, la e.v.m. en el problema *A*,  $\gamma^* = (\hat{\theta}^* - \theta_1)\sqrt{n}$ . Entonces el criterio  $\pi(\gamma^*)$  para verificar  $H_1$  frente a  $H_2$  en el problema *A* poseerá asintóticamente las mismas propiedades que el criterio  $\pi(Y)$  en el problema *B*.

Ahora bien, para hallar el criterio asintóticamente óptimo en el problema *A*, debemos examinar el problema *B*, que es más simple, y encontrar en éste (si es posible) el criterio  $\pi$  dotado de la propiedad de optimización necesaria. Si ahora tomamos, en calidad de la observación  $Y$ , el valor de  $\gamma^*$  y lo sustituimos en  $\pi$ , obtendremos el criterio buscado en el problema *A*.

Este hecho podría llamarse *indicio límite de optimización*. Su sentido es bastante sencillo. Pues sabemos, de los resultados del capítulo 2, que cuando  $X \in \mathbf{P}_\theta$ ,

$$\sqrt{n}(\hat{\theta}^* - \theta)I^{1/2}(\theta) \in \Phi_{0, E}$$

uniformemente respecto a  $\theta$ . Por consiguiente, para  $\theta = \theta_1 + \gamma n^{-1/2}$ ,

$$\sqrt{n}(\hat{\theta}^* - \theta_1) - \gamma \in \Phi_{0, I^{-1}(\theta_1)}$$

o bien, que es lo mismo,

$$\gamma^* \in \Phi_{\gamma, I^{-1}}.$$

Así, pues,  $\Phi_{\gamma, I^{-1}}$ , o sea, la distribución presente en el problema *B* no es otra cosa sino la distribución límite para  $\gamma^*$ . Por eso, el indicio límite de optimización es muy natural: reduce el problema de verificación de las hipótesis a un problema "límite". Lo interesante en todo esto es el hecho de que con tal reducción no ocurre ninguna pérdida considerable de información respecto a  $\theta$ : el criterio óptimo en el problema *B* también conserva esta optimalidad con arreglo al problema *A*.

Para conferir a lo dicho un sentido exacto, introduzcamos ahora los principales conceptos de optimización asintótica de los criterios para verificar las hipótesis semejantes en el problema *A*.

En el párrafo precedente hemos dado la definición de la clase  $\tilde{K}_\varepsilon$  de los criterios  $\pi$  de nivel asintótico  $1 - \varepsilon$  (definición 2). Para  $\pi \in \tilde{K}_\varepsilon$  es válida

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in \Theta_1} M_\theta \pi(X) \leq \varepsilon.$$

**Definición 1.** El criterio  $\pi_1 \in \tilde{K}_\varepsilon$  se llama *criterio asintóticamente más uniforme y más potente (c.a.u.m.p.)* en  $\tilde{K}_\varepsilon$  si para cualquier  $\gamma \in \Gamma_2$  y para cualquier  $\pi \in \tilde{K}_\varepsilon$

$$\liminf_{n \rightarrow \infty} (M_\theta \pi_1(X) - M_\theta \pi(X)) \geq 0,$$

donde  $\theta = \theta_1 + \gamma n^{-1/2} \in \Theta_2$  cuando  $\gamma \in \Gamma_2$ .

Supongamos que en  $\Gamma_1$  se dan las distribuciones  $\Pi_i$  que inducen en  $\Theta_1$  algunas otras distribuciones (concentradas en el entorno  $n^{-1/2}$  del punto  $\theta_1$ ) que designaremos por  $Q_i$ ,  $i = 1, 2$ . Las hipótesis de que  $\theta$  se elige al azar con la distribución  $Q_i$ , las designaremos, como antes, por  $H_{Q_i}$ .

Por  $\tilde{K}_\varepsilon^{Q_i}$  designaremos la clase de criterios  $\pi$  para los cuales

$$\limsup_{n \rightarrow \infty} M_{Q_i} \pi(X) \leq \varepsilon,$$

donde  $M_{Q_i}$  significa la esperanza matemática incondicional de la distribución compatible de  $\theta$  y  $X$ ,  $\theta \in Q_i$ ,  $X \in P_\theta$ . Es evidente que  $\tilde{K}_\varepsilon \subset \tilde{K}_\varepsilon^{Q_i}$  para cualquier  $Q_i$ .

**Definición 2.** El criterio  $\pi_1 \in \tilde{K}_\varepsilon^{Q_1}$  para verificar  $H_{Q_1}$  frente a  $H_{Q_2}$ , se denomina *criterio asintóticamente bayesiano (c.a.b.)* en  $\tilde{K}_\varepsilon^{Q_1}$  si para cualquier otro criterio  $\pi \in \tilde{K}_\varepsilon^{Q_1}$ ,

$$\liminf_{n \rightarrow \infty} (M_{Q_2} \pi_1(X) - M_{Q_2} \pi(X)) \geq 0. \quad (3)$$

Se puede dar una definición equivalente del carácter bayesiano en la cual en vez de (3) se exige que

$$\liminf_{n \rightarrow \infty} (M_{Q_2} \pi_1(X) - M_{Q_2} \pi_{Q_1, Q_2}(X)) \geq 0, \quad (4)$$

donde  $\pi_{Q_1, Q_2}$  es el criterio bayesiano de  $\tilde{K}_\varepsilon^{Q_1}$  para verificar las hipótesis  $H_{Q_1}$  y  $H_{Q_2}$  (o, que es lo mismo, el criterio más potente para verificar  $H_{Q_1}$  frente a  $H_{Q_2}$  de nivel asintótico  $1 - \varepsilon$ ).

Cabe señalar que la definición 2 se distingue algo de la del c.a.b. que hemos dado en el párrafo anterior (véase la definición 13.2. Allí figura la relación de las probabilidades de los errores, y no su diferencia). Desde el punto de vista de la exposición ulterior, estas definiciones son equivalentes, pero la última de ellas será la más conveniente para nosotros.

**Definición 3.** El criterio  $\pi_1 \in \tilde{K}_\varepsilon$  se llama *criterio asintóticamente minímax* en  $\tilde{K}_\varepsilon$  para verificar  $H_1$  frente a  $H_2$  si para cualquier otro criterio  $\pi \in \tilde{K}_\varepsilon$  se cumple

$$\liminf_{n \rightarrow \infty} (\inf_{\theta \in \Theta_2} M_\theta \pi_1(X) - \inf_{\theta \in \Theta_2} M_\theta \pi(X)) \geq 0. \quad (5)$$



Al igual que al examinar los criterios minimax ordinarios (véase el § 9), para evitar consideraciones poco importantes, es cómodo separar los conjuntos  $\Theta_1$  y  $\Theta_2$  por medio de cierta zona intermedia, de modo que ellos no se toquen. De lo contrario ambos límites inferiores en (5) pueden resultar iguales a  $\varepsilon$  para cualquier criterio no desplazado asintóticamente  $\pi$ .

De las definiciones citadas se deduce que la propiedad de una u otra optimización asintótica se distingue de la propiedad corriente de esa misma optimización tan sólo por el hecho de que ante la respectiva diferencia aparece el signo  $\liminf$ .

A la par con los criterios asintóticamente bayesianos y minimax, en las clases  $\tilde{K}_\varepsilon$  y  $\tilde{K}_\varepsilon^Q$ , se puede estudiar las clases asintóticamente bayesianas y minimax ordinarias. Supongamos que en  $\Theta = \Theta_1 \cup \Theta_2$  tenemos la distribución  $Q = q(1)Q_1 + q(2)Q_2$ ,  $q(1) + q(2) = 1$ . Entonces, el criterio  $\pi_1$  se denomina *asintóticamente bayesiano para la distribución a priori Q*, si para cualquier otro criterio  $\pi$ ,

$$\liminf_{\pi \rightarrow \infty} [q(1)M_{Q_1} \pi_1(X) + q(2)M_{Q_2}(1 - \pi_1(X)) - q(1)M_{Q_1} \pi(X) - q(2)M_{Q_2}(1 - \pi(X))] \leq 0. \quad (6)$$

La probabilidad de error del criterio  $\pi$  promediado respecto a  $Q$ , presente en esta desigualdad, puede ser escrita mediante la probabilidad  $\alpha(\pi, \theta)$  de error en el punto  $\theta$ , en forma de  $M_Q \alpha(\pi, \theta)$ , donde

$$\alpha(\pi, \theta) = \begin{cases} M_\theta \pi(X) & \text{cuando } \theta \in \Theta_1, \\ M_\theta (1 - \pi(X)) & \text{cuando } \theta \in \Theta_2. \end{cases}$$

Entonces, la desigualdad (6) adopta la forma

$$\liminf_{\pi \rightarrow \infty} M_Q [\alpha(\pi_1(X), \theta) - \alpha(\pi(X), \theta)] \leq 0.$$

El criterio  $\pi_1$  será *asintóticamente minimax* si

$$\liminf_{\pi \rightarrow \infty} [\sup_{\theta \in \Theta} \alpha(\pi_1, \theta) - \sup_{\theta \in \Theta} \alpha(\pi, \theta)] \leq 0$$

para cualquier otro criterio  $\pi$ .

El estudio de los criterios asintóticamente bayesianos (en  $\tilde{K}_\varepsilon^Q$ ) y asintóticamente minimax (en  $\tilde{K}_\varepsilon$ ), y simplemente el estudio de los criterios asintóticamente bayesianos y minimax es, de hecho, una misma cosa. Por ejemplo, el criterio bayesiano de  $\tilde{K}_\varepsilon^Q$  es un criterio bayesiano ordinario para  $q(1)$  correspondiente. En este párrafo estudiaremos los criterios de las clases  $\tilde{K}_\varepsilon$  y  $\tilde{K}_\varepsilon^Q$ , en tanto que los criterios asintóticamente bayesianos y minimax ordinarios serán examinados en los capítulos ulteriores al investigar un planteamiento más general del problema.

**2. Afirmaciones principales.** Para simplificar al máximo la exposición posterior, introduciremos una suposición que de ningún modo está relacionada con la esencia de la cuestión y que, si se desea, puede ser retirada,

ya que para ello existen todos los resultados necesarios. Es decir, supondremos que los conjuntos  $\Gamma_i$  están limitados, o sea, existe  $N > 0$  tal, que  $\Gamma_i \subset \{\gamma: |\gamma| \leq N\}$ .

**Definición 4.** Los criterios  $\pi_1$  y  $\pi_2$  para verificar las hipótesis semejantes  $H_1 = \{\theta \in \Theta_1\}$  y  $H_2 = \{\theta \in \Theta_2\}$  a base de la muestra  $X$ , se denominan criterios *asintóticamente equivalentes* si

$$\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta_1 \cup \Theta_2} |M_\theta \pi_1(X) - M_\theta \pi_2(X)| = 0. \quad (7)$$

Después de tal suposición podemos poner la región  $|\theta - \theta_1| \leq N/\sqrt{n}$  bajo el signo sup en (7).

Los criterios asintóticamente equivalentes  $\pi_1$  y  $\pi_2$  poseen las propiedades siguientes:

1) Si  $\pi_1 \in \tilde{K}_\varepsilon$  (o  $\tilde{K}_\varepsilon^{Q_1}$ ), entonces  $\pi_2 \in \tilde{K}_\varepsilon$  ( $\tilde{K}_\varepsilon^{Q_1}$ ).

2) Si  $\pi_1$  posee una de las propiedades de la optimización asintótica en las definiciones 1—3, el criterio  $\pi_2$  poseerá esa misma propiedad.

La primera afirmación se deduce de (7) y de la desigualdad

$$\sup_{\theta \in \Theta_1} M_\theta \pi_2(X) \leq \sup_{\theta \in \Theta_1} M_\theta \pi_1(X) + \sup_{\theta \in \Theta_1} |M_\theta(\pi_2 - \pi_1)|.$$

La segunda afirmación se demuestra análogamente. Si, por ejemplo,  $\pi_1$  es asintóticamente minimax, el carácter asintóticamente minimax de  $\pi_2$  será el corolario de (7) y de la desigualdad

$$\inf_{\theta \in \Theta_2} M_\theta \pi_2(X) \geq \inf_{\theta \in \Theta_2} M_\theta \pi_1(X) - \sup_{\theta \in \Theta_2} |M_\theta(\pi_2 - \pi_1)|. \quad \triangleleft$$

Las condiciones de la equivalencia asintótica de los criterios son establecidas por el

**Lema 1.** Supongamos que en el entorno del punto  $\theta_1$  se cumplen las condiciones (RR),  $\pi_i(X) = I_{\{T_n(X) + \varepsilon_{ni}(X) > c\}}$ ,  $i = 1, 2$ , donde para  $X \in P_\theta$ , tienen lugar las relaciones  $\varepsilon_{ni}(X) \xrightarrow{P_\theta} 0$ ,  $T_n(X) \in G$ , y la distribución  $G$  es continua. Entonces, los criterios  $\pi_1$  y  $\pi_2$  son asintóticamente equivalentes.

**Demostración.**  $|M_\theta \pi_1(X) - M_\theta \pi_2(X)| \leq P_\theta(A_n)$ , donde para el suceso  $A_n = \{\pi_1(X) \neq \pi_2(X)\}$  se cumple  $P_{\theta_1}(A_n) = P_{\theta_1}(T_n(X) + \varepsilon_{n1}(X) > c, T_n(X) + \varepsilon_{n2}(X) \leq c) + P_{\theta_1}(T_n(X) + \varepsilon_{n1}(X) \leq c, T_n(X) + \varepsilon_{n2}(X) > c) \rightarrow 0$  cuando  $n \rightarrow \infty$ , ya que la distribución límite  $T_n$  es continua. Por consiguiente, en virtud del teorema 2.29.5,  $\sup_{|\theta - \theta_1| \leq N/\sqrt{n}} P_\theta(A_n) \rightarrow 0. \quad \triangleleft$

El criterio bayesiano de nivel  $1 - \varepsilon$  en el problema  $B$  para verificar las hipótesis  $H_{\Pi_i}$ , de que  $\gamma$  se elige al azar con la distribución  $\Pi_i$  en  $\Gamma_i$ ,  $i = 1, 2$ , lo designaremos por  $\pi_{\Pi_1, \Pi_2}(Y)$ . Este criterio tiene la forma

$$r(Y) \equiv \frac{\int \exp \left\{ -\frac{1}{2}(Y - u)I(Y - u)^T \right\} \Pi_2(du)}{\int \exp \left\{ -\frac{1}{2}(Y - u)I(Y - u)^T \right\} \Pi_1(du)} > c, \quad (8)$$

donde  $c = c_\varepsilon$  se elige de la condición

$$\int \varphi(\gamma, c) \Pi_1(d\gamma) = \varepsilon, \varphi(\gamma, c) = \mathbf{P}(r(Y) > c), Y \in \Phi_{\gamma, I-1}. \quad (9)$$

Estas relaciones significan, evidentemente, que  $\mathbf{M}_{\Pi_1, \pi_{\Pi_1, \Pi_2}}(Y) = \varepsilon$ .

Nótese que  $r(y)$  es una función analítica de  $y$ . En virtud de su analiticidad, esta función no puede adquirir un valor constante en el conjunto de la medida positiva de Lebesgue o de la medida  $\Phi_{\gamma, I-1}$  (de lo contrario sería constante en todas partes, lo cual sólo es posible cuando  $\Pi_1 = \Pi_2$ ). Por lo tanto,  $\mathbf{P}(r(Y) = c) = 0$  para cualquier  $c$ , y la distribución de  $r(Y)$  es continua.

Supongamos, como antes, que  $\pi_{Q_1, Q_2}(X)$  designa el criterio bayesiano de nivel asintótico  $1 - \varepsilon$  en el problema  $A$ .

**Teorema 1.** *Supongamos que las condiciones (RR) se cumplen en el entorno del punto  $\theta_1$ . Entonces, el criterio  $\pi(X) = \pi_{\Pi_1, \Pi_2}(\gamma^*)$ ,  $\gamma^* = (\hat{\theta}^* - \theta_1)\sqrt{n}$  es asintóticamente equivalente al criterio  $\pi_{Q_1, Q_2}$  y, por consiguiente, es asintóticamente bayesiano.*

Además,

$$\sup_{|\gamma| \leq N} |\mathbf{M}_{\theta_1 + \gamma/\sqrt{n}} \pi(X) - \varphi(\gamma, c)| \rightarrow 0 \quad (10)$$

cuando  $n \rightarrow \infty$ , donde  $\varphi(\gamma, c) = \mathbf{M}_{\gamma} \pi_{\Pi_1, \Pi_2}(Y)$  está definida en (9).

**Demostración.** Examinemos el criterio bayesiano  $\pi_{Q_1, Q_2}$  en el problema  $A$ . Este criterio tiene la forma

$$T(X) = \frac{\int_{\theta_1 + u/\sqrt{n}} f_{\theta_1 + u/\sqrt{n}}(X) \Pi_2(du)}{\int_{\theta_1 + u/\sqrt{n}} f_{\theta_1 + u/\sqrt{n}}(X) \Pi_1(du)} > c.$$

Si  $X \in \mathbf{P}_{\theta_1}$ , entonces, en virtud del teorema 2.28.5,

$$T(X) = r(\gamma^*)(1 + \varepsilon(X, \theta_1))$$

( $\gamma^* = u^*$  cuando  $\theta = \theta_1$ ). Como la distribución de  $r(Y)$  es continua,  $\gamma^* = Y \in \Phi_{0, I-1}$ , y como el criterio  $\pi$  tiene la forma  $r(\gamma^*) > c$ , en virtud del lema 1 queda demostrada la primera afirmación del teorema.

La relación (10) se deduce de la representación

$$\mathbf{M}_{\theta_1 + \gamma/\sqrt{n}} \pi(X) = \mathbf{M}_{\theta_1 + \gamma/\sqrt{n}} I_{\{r(\gamma^*) > c\}} \rightarrow \mathbf{P}(r(Y) > c),$$

$Y \in \Phi_{\gamma, I}$  y del teorema 2.29.4.  $\triangleleft$

**Teorema 2.** *Supongamos que en el entorno del punto  $\theta_1$  se cumplen las condiciones (RR),  $\gamma^* = (\hat{\theta}^* - \theta_1)\sqrt{n}$ .*

*Supongamos, además, que existe el criterio minimax  $\pi_1(Y)$  de nivel  $1 - \varepsilon$  para verificar  $k_1$  frente a  $k_2$  en el problema  $B$ , y que este criterio es bayesiano*

$$\pi_1(Y) = \pi_{\Pi_1, \Pi_2}(Y) \quad (11)$$

para las distribuciones a priori  $\Pi_1$  y  $\Pi_2$  que satisfacen las condiciones

$$M_{\Pi_1} \pi_1(Y) = \sup_{\gamma \in \Gamma_1} M_{\gamma} \pi(Y), \tag{12}$$

$$M_{\Pi_2} \pi_1(Y) = \sup_{\gamma \in \Gamma_2} M_{\gamma} \pi(Y), \quad Y \in \Phi_{\gamma, I-1}$$

(compárense con las condiciones 9.1). Entonces, el criterio  $\pi(X) = \pi_{\Pi_1, \Pi_2}(\gamma^*)$  será asintóticamente minimax en la clase  $\tilde{K}_\varepsilon$  de los criterios para verificar  $H_1$  frente a  $H_2$  en el problema inicial  $A$ .

**Demostración.** Como  $\pi_1$  es un criterio de nivel  $1 - \varepsilon$ , entonces

$$\sup_{\gamma \in \Gamma_1} M_{\gamma} \pi_1(Y) = M_{\Pi_1} \pi(Y) = \varepsilon.$$

De aquí, en virtud de (10) y (12), obtenemos

$$\lim_{n \rightarrow \infty} \sup_{\gamma \in \Gamma_1} M_{\theta_1 + \gamma/\sqrt{n}} \pi_{Q_1, Q_2}(X) = \lim_{n \rightarrow \infty} M_{Q_1} \pi_{Q_1, Q_2}(X) = \varepsilon.$$

Esto significa que  $\pi_{Q_1, Q_2} \in \tilde{K}_\varepsilon$ ,  $\pi_{Q_1, Q_2} \in \tilde{K}_\varepsilon^{Q_1}$ .

Ahora es necesario demostrar que para cualquier criterio  $\pi^* \in \tilde{K}_\varepsilon$ ,

$$\lim_{n \rightarrow \infty} \inf_{\theta \in \Theta_2} \left( \inf_{\theta \in \Theta_2} M_{\theta} \pi(X) - \inf_{\theta \in \Theta_2} M_{\theta} \pi^*(X) \right) \geq 0.$$

Tenemos

$$\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta_2} \inf M_{\theta} \pi^*(X) \leq \lim_{n \rightarrow \infty} \sup M_{\theta} \pi^*(X) \leq \lim_{n \rightarrow \infty} \sup M_{Q_2} \pi_{Q_1, Q_2}(X). \tag{13}$$

La última desigualdad es válida en virtud del carácter bayesiano de  $\pi_{Q_1, Q_2}$  (o sea, de la minimización de  $q_1 M_{Q_1} \pi_{Q_1, Q_2} + (1 - q_1) M_{Q_2} (1 - \pi_{Q_1, Q_2})$  para  $q_1$  correspondiente) y en virtud del hecho de que  $\lim_{n \rightarrow \infty} \sup M_{Q_1} \pi^*(X) \leq \varepsilon$ ,  $\lim_{n \rightarrow \infty} M_{Q_1} \pi_{Q_1, Q_2} = \varepsilon$ .

Seguidamente, en virtud de (10) y (12) y del teorema 1, el segundo miembro en (13) es igual a

$$\begin{aligned} \lim_{n \rightarrow \infty} M_{Q_2} \pi_1(\gamma^*) &= M_{\Pi_2} \pi_{\Pi_1, \Pi_2}(Y) = \inf_{\gamma \in \Gamma_2} M_{\gamma} \pi_{\Pi_1, \Pi_2}(Y) = \\ &= \lim_{n \rightarrow \infty} \inf_{\gamma \in \Gamma} M_{\theta_1 + \gamma/\sqrt{n}} \pi_{Q_1, Q_2}(X). \quad \triangleleft \end{aligned}$$

**Teorema 3.** Supongamos que existe un c.u.m.p.  $\pi_1(Y)$  de nivel  $1 - \varepsilon$  para verificar  $h_1$  frente a  $h_2$  en el problema  $B$ . Supongamos, además, que para cualquier  $\gamma_2 \in \Gamma_2$  existe una distribución  $\Pi_1$  en  $\Gamma_1$  tal, que

$$\pi_1(Y) = \pi_{\Pi_1, \Pi_2}(Y) \tag{14}$$

es el criterio bayesiano para verificar  $h_{\Pi_1}$  frente a  $h_{\Pi_2}$  (aquí  $\Pi_2$  está concentrada en el punto  $\gamma_2$ ). Entonces, el criterio  $\pi(X) = \pi_1(\gamma^*)$  es el c.a.u.m.p. (de nivel asintótico  $1 - \varepsilon$ ) para verificar  $H_1$  frente a  $H_2$  en el problema inicial  $A$ .

Nótese que para los problemas de los §§ 5—7 siempre se cumple la condición (14). Esto se deduce de la propia construcción del c.u.m.p. en estos párrafos.

**Demostración del teorema 3.** La pertenencia de  $\pi_1(\gamma^*) \in \tilde{K}_\varepsilon$  se deduce del teorema 1, ya que

$$\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta_1} \mathbf{M}_\theta \pi_1(\gamma^*) = \sup_{\theta \in \Theta_1} \lim_{n \rightarrow \infty} \mathbf{M}_\theta \pi_1(\gamma^*) = \sup_{\gamma \in \Gamma_2} \varphi(\gamma, c) \leq \varepsilon.$$

Sea ahora  $\pi^*$  cualquier otro criterio de  $\tilde{K}_\varepsilon$ . Entonces

$$\lim_{n \rightarrow \infty} \sup \mathbf{M}_{Q_1} \pi^*(X) \leq \lim_{n \rightarrow \infty} \sup \sup_{\theta \in \Theta_1} \mathbf{M}_\theta \pi^*(X) \leq \varepsilon$$

y, por consiguiente,  $\pi^*$  también se puede considerar como criterio de  $\tilde{K}_\varepsilon^{Q_1}$  para verificar  $H_{Q_1}$  frente a  $H_{Q_2}$ , donde  $Q_1$  está inducida por la distribución  $\Pi_1$  (véase la enunciación del teorema), y  $Q_2$  está concentrada en el punto  $\theta_2 = \theta_1 + \gamma_2 n^{-1/2}$ . Si  $\pi_{Q_1, Q_2}$  es un criterio bayesiano de nivel asintótico  $1 - \varepsilon$  para estas distribuciones, entonces

$$\lim_{n \rightarrow \infty} \mathbf{M}_{\theta_2} \pi_{Q_1, Q_2}(X) \geq \lim_{n \rightarrow \infty} \sup \mathbf{M}_{\theta_2} \pi^*(X).$$

Pero el primer miembro de esta desigualdad coincide, en virtud del teorema 1, con el valor

$$\lim_{n \rightarrow \infty} \mathbf{M}_{\theta_2} \pi_{\Pi_1, \Pi_2}(\gamma^*) = \lim_{n \rightarrow \infty} \mathbf{M}_{\theta_2} \pi_1(\gamma^*). \quad \triangleleft$$

De un modo análogo se puede buscar el c.a.u.m.p. en la clase de los criterios no desplazados asintóticamente.

**Observación 1.** Si las distribuciones  $\Pi_1$  y  $\Pi_2$  están concentradas en los puntos  $\gamma_1$  y  $\gamma_2$ , respectivamente, entonces

$$r(Y) = \frac{\exp \left\{ -\frac{1}{2}(Y - \gamma_2)I(Y - \gamma_2)^T \right\}}{\exp \left\{ -\frac{1}{2}(Y - \gamma_1)I(Y - \gamma_1)^T \right\}}.$$

Por lo tanto, la región crítica  $\pi_{\Pi_1, \Pi_2}(Y)$  tendrá la forma

$$YI(\gamma_2 - \gamma_1)^T = (YI, (\gamma_2 - \gamma_1)) > c.$$

En el caso unidimensional, de aquí obtenemos el c.a.m.p. (3.21) que hemos estudiado en el § 3.

**Observación 2.** Si la distribución  $\Pi_1$  está concentrada en el punto  $u = 0$ , y la distribución  $\Pi_2$  es uniforme en la esfera  $|u| \leq N$ , el denominador de la función  $r(Y)$  será igual a  $\exp \left\{ -\frac{1}{2}YIY^T \right\}$ , y el denominador para grandes  $N$  y  $|\gamma| < N - \sqrt{N}$  será próximo a  $\sqrt{|I|} (2\pi)^{k/2}$ . Por consiguiente, la región crítica para  $\pi_{\Pi_1, \Pi_2}$  con tales  $\Pi_1$  y  $\Pi_2$  será próxima al aspecto exterior del elipsoide

$$YIY^T > c,$$

y la región crítica del criterio asintóticamente bayesiano  $\pi_{\Pi, \Pi_2}(\gamma^*)$  será próximo a

$$\gamma^* I \gamma^{*T} > c.$$

Esto no es otra cosa sino la forma asintótica del c.r.v. que hemos estudiado en el párrafo anterior (compárese con el teorema 13.2).

**Observación 3.** En los teoremas 2 y 3 están presentes las condiciones consistentes en que el criterio minimax (teorema 2) o el c.u.m.p. (teorema 3) para el problema B son bayesianos en caso de algunas distribuciones  $\Pi_i$  en  $\Gamma_i$ . En los capítulos posteriores veremos que estas condiciones son inútiles: la clase de todos los criterios bayesianos comprende todos los criterios "inmejorables", incluso los c.u.m.p. y los minimax.

### § 15. Propiedades de la optimización asintótica del criterio de relación de verosimilitud que se deducen del indicio límite de optimización

En este párrafo examinaremos algunas consecuencias de los resultados del § 14, vinculadas con el criterio de relación de verosimilitud. Estableceremos, en particular, la potencia máxima uniforme asintótica y el carácter minimax asintótico del c.r.v. para algunos problemas importantes concretos, relacionados con la verificación de las hipótesis próximas.

En lo sucesivo siempre estimaremos que en el entorno del punto  $\theta_1$  se cumplen las condiciones (RR). Para simplificar los cálculos será conveniente, al igual que en el párrafo anterior, considerar, donde sea necesario, que los conjuntos  $\Gamma_i$  están limitados.

**1. C.a.u.m.p. para hipótesis semejantes con alternativas unilaterales.** Supongamos que el parámetro  $\theta$  es unidimensional y que se verifica la hipótesis unilateral  $H_1 = \{\theta \leq \theta_1 + \gamma_1 n^{-1/2}\}$  frente a la hipótesis  $H_2 = \{\theta > \theta_2 = \theta_1 + \gamma_2 n^{-1/2}\}$ ,  $\gamma_1 \leq \gamma_2$ .

**Teorema 1.** *El criterio de relación de verosimilitud  $\hat{\pi}(X)$  con la región crítica*

$$R(X) \equiv \frac{\sup_{\theta \in \Theta_2} f_{\theta}(X)}{\sup_{\theta \in \Theta_1} f_{\theta}(X)} > c, \quad (1)$$

cuando  $\Theta_1 = \{\theta: \theta \leq \theta_1 + \gamma_1 n^{-1/2}\}$ ,  $\Theta_2 = \{\theta: \theta \geq \theta_1 + \gamma_2 n^{-1/2}\}$  y con un valor conveniente de  $c$ , es asintóticamente equivalente al criterio

$$\gamma^* = (\hat{\theta}^* - \theta_1)\sqrt{n} > c_{\varepsilon} = \lambda_{\varepsilon} I^{-1/2} + \gamma_1, \quad \Phi_{0,1}(\lambda_{\varepsilon}) = 1 - \varepsilon \quad (2)$$

y es el c.a.u.m.p. de nivel asintótico  $1 - \varepsilon$  para verificar la hipótesis  $H_1 = \{\theta \leq \theta_1 + \gamma_1 n^{-1/2}\}$  frente a  $H_2 = \{\theta > \theta_1 + \gamma_2 n^{-1/2}\}$ . En las fórmulas (2),  $I$  designa la información de Fisher  $I(\theta_1)$  en el punto  $\theta_1$  para la familia  $f_{\theta}$ .

**Demostración.** Del § 5 se deduce que para una muestra  $Y \in \Phi_{\gamma, I^{-1}}$  de volumen unitario, procedente de una población normal de varianza conocida  $I^{-1}$ , existe un c.u.m.p. para verificar la hipótesis  $k_1 = \{\gamma \leq \gamma_1\}$  frente a  $k_2 = \{\gamma > \gamma_2\}$  de forma  $Y > c_\varepsilon$ , donde  $c$  está definida en (2). Así mismo será, evidentemente, el criterio bayesiano para las distribuciones degeneradas concentradas en los puntos  $\gamma_1$  y  $\gamma_2$  (o en los puntos  $\gamma_1$  y  $\gamma > \gamma_1$  si  $\gamma_1 = \gamma_2$ ). A base de esto, del teorema 14.3 se deduce que existe el c.a.u.m.p. de nivel asintótico  $1 - \varepsilon$  para verificar  $H_1$  frente a  $H_2$  y que el mismo tiene la forma (2).

Queda demostrar que los criterios (1) y (2) son asintóticamente equivalentes. De acuerdo con el teorema 2.28.4, suponiendo que  $Z_1(t) =$

$$= \frac{f_{\theta_1} + t^{(X)}}{f_{\theta_1}(X)}, \text{ tendremos, cuando } X \in P_{\theta_1},$$

$$R(X) = \frac{\sup_{u > \gamma_2} Z_1(un^{-1/2})}{\sup_{u \leq \gamma_1} Z_1(un^{-1/2})} = \frac{\sup_{u > \gamma_2} \exp \left\{ -\frac{1}{2}(\gamma^* - u)^2 I + \varepsilon_n^{(2)}(X) \right\}}{\sup_{u \leq \gamma_1} \exp \left\{ -\frac{1}{2}(\gamma^* - u)^2 I + \varepsilon_n^{(1)}(X) \right\}} = T_n(X) + \varepsilon_n^{(3)}(X),$$

donde  $\varepsilon_n^{(i)}(X) \xrightarrow{P_{\theta_1}} 0$ ,  $i = 1, 2, 3$ ,

$$T_n(X) = r(\gamma^*) = \frac{\sup_{u > \gamma_2} \exp \left\{ -\frac{1}{2}(\gamma^* - u)^2 I \right\}}{\sup_{u \leq \gamma_1} \exp \left\{ -\frac{1}{2}(\gamma^* - u)^2 I \right\}} =$$

$$= \begin{cases} \exp \left\{ -\frac{1}{2}(\gamma^* - \gamma_2)^2 I \right\} & \text{cuando } \gamma^* \leq \gamma_1, \\ \exp \left\{ -\frac{1}{2}(\gamma^* - \gamma_2)^2 I + \frac{1}{2}(\gamma^* - \gamma_1)^2 I \right\} & \text{cuando } \gamma_1 < \gamma^* < \gamma_2, \\ \exp \left\{ \frac{1}{2}(\gamma^* - \gamma_1)^2 I \right\} & \text{cuando } \gamma^* \geq \gamma_2. \end{cases}$$

Esta es una función continua monótonamente creciente de  $\gamma^*$ . Por consiguiente, la desigualdad  $T_n(X) > c$  equivale a la desigualdad  $\gamma^* > c'$  para cierta  $c'$ . Además, como  $\gamma^* = Y \in \Phi_{0,1-1}$ , entonces la distribución  $r(Y)$  es absolutamente continua. Las condiciones del lema 14.1 para los criterios (1) y (2) se cumplen.  $\triangleleft$

**2. C.a.u.m.p. para alternativas bilaterales.** Supongamos que el parámetro  $\theta$  es, como antes, unidimensional, y que el problema  $A$  consiste

en verificar la hipótesis  $H_1 = \{(\theta - \theta_1)\sqrt{n}\phi(\gamma_1, \gamma_2)\}$  frente a  $H_2 = \{(\theta - \theta_1)\sqrt{n} \in (\gamma_1, \gamma_2)\}$ ,  $\gamma_2 > \gamma_1$ . Designemos

$$\bar{\gamma} = \frac{\gamma_1 + \gamma_2}{2}, \quad \Delta = \frac{\gamma_2 - \gamma_1}{2}.$$

**Teorema 2.** El criterio de relación de verosimilitud  $\hat{\pi}(X)$ , definido en (1) para el valor correspondiente de  $c$  y para  $\Theta_1 = \{\theta: (\theta - \theta_1)\sqrt{n}\phi(\gamma_1, \gamma_2)\}$ ,  $\Theta_2 = \{\theta: (\theta - \theta_1)\sqrt{n} \in (\gamma_1, \gamma_2)\}$ , al igual que el criterio

$$|\gamma^* - \bar{\gamma}| = |(\hat{\theta}^* - \theta_1)\sqrt{n} - \bar{\gamma}| < c_\varepsilon, \quad (3)$$

donde  $c_\varepsilon$  se determina de la ecuación  $\Phi_{0, I-1}(-c - \Delta, c - \Delta) = \varepsilon$ , son los c.a.u.m.p. de nivel asintótico  $1 - \varepsilon$  para verificar  $H_1 = \{(\theta - \theta_1)\sqrt{n}\phi(\gamma_1, \gamma_2)\}$  frente a  $H_2 = \{(\theta - \theta_1)\sqrt{n} \in (\gamma_1, \gamma_2)\}$ .

La demostración de este teorema es bastante parecida a la del teorema anterior. Del § 5 resulta que para el problema  $B$  destinado a verificar, a base de la observación  $Y \in \Phi_{\gamma, I-1}$ , la hipótesis  $k_1 = \{\gamma\phi(\gamma_1, \gamma_2)\}$  frente a  $k_2 = \{\gamma \in (\gamma_1, \gamma_2)\}$ , existe un c.a.u.m.p. en forma de  $c' < Y < c''$ , donde  $c'$  y  $c''$  se eligen de modo que

$$\Phi_{\gamma_1, I-1}((c', c'')) = \Phi_{\gamma_2, I-1}((c', c'')) = \varepsilon.$$

Es fácil notar que podremos satisfacer estas relaciones si ponemos  $c' = \bar{\gamma} - c_\varepsilon$ ,  $c'' = \bar{\gamma} + c_\varepsilon$ , ya que

$$\Phi_{\gamma_1, I-1}((\bar{\gamma} - c_\varepsilon, \bar{\gamma} + c_\varepsilon)) = \Phi_{0, I-1}((-c_\varepsilon + \Delta, c_\varepsilon + \Delta)) = \varepsilon,$$

$$\Phi_{\gamma_2, I-1}((\bar{\gamma} - c_\varepsilon, \bar{\gamma} + c_\varepsilon)) = \Phi_{0, I-1}((-c_\varepsilon - \Delta, c_\varepsilon - \Delta)) = \varepsilon.$$

Además, en el § 5 hemos visto que para cualquier  $\gamma_0 \in (\gamma_1, \gamma_2)$  existe  $q \in (0, 1)$  tal, que el criterio bayesiano  $\pi_{\Pi_1, \Pi_2}$  al verificar la hipótesis  $k_{\Pi_1}$  para la distribución  $\Pi_1: \Pi_1(\{\gamma_1\}) = q$ ,  $\Pi_1(\{\gamma_2\}) = 1 - q$  frente a la hipótesis  $k_{\Pi_2} = \{\gamma = \gamma_0\}$ , tendrá la forma

$$c' < Y < c''.$$

Esto significa que las condiciones del teorema 14.3 serán cumplidas y que el criterio (3) será el c.a.u.m.p. para verificar  $H_1$  frente a  $H_2$ .

Examinemos ahora el c.r.v. (1) para las regiones  $\Theta_i$  definidas en el teorema y mostremos que el mismo equivale asintóticamente a (3). Al igual que en la demostración del teorema 1, del teorema 2.28.4 obtenemos que, para  $X \in \mathbf{P}_{\theta_1}$ ,

$$\frac{\sup_{u \in \gamma_1, \gamma_2} Z_1(un^{-1/2})}{\sup_{u \notin \gamma_1, \gamma_2} Z_1(un^{-1/2})} = \frac{\sup_{u \in \gamma_1, \gamma_2} \exp \left\{ -\frac{1}{2}(\gamma^* - u)^2 I + \varepsilon_n^{(1)}(X) \right\}}{\sup_{u \notin \gamma_1, \gamma_2} \exp \left\{ -\frac{1}{2}(\gamma^* - u)^2 I + \varepsilon_n^{(2)}(X) \right\}} = T_n(X) + \varepsilon_n^{(3)}(X),$$



donde  $\varepsilon_n^{(i)}(X) \xrightarrow{P_{\theta_1}} 0$ ,  $i = 1, 2, 3$ ,

$$T(X) = r(\gamma^*) = \frac{\sup_{u \in \gamma_1, \gamma_2} \exp \left\{ -\frac{1}{2}(\gamma^* - u)^2 I \right\}}{\sup_{u \in \gamma_1, \gamma_2} \exp \left\{ -\frac{1}{2}(\gamma^* - u)^2 I \right\}} =$$

$$= \begin{cases} \exp \left\{ -\frac{1}{2}(\gamma^* - \gamma_1)^2 I \right\} & \text{cuando } \gamma^* \leq \gamma_1, \\ \exp \left\{ \frac{1}{2}(\gamma^* - \gamma_1)^2 I \right\} & \text{cuando } \gamma_1 < \gamma^* \leq \bar{\gamma}, \\ \exp \left\{ \frac{1}{2}(\gamma^* - \gamma_2)^2 I \right\} & \text{cuando } \bar{\gamma} < \gamma^* \leq \gamma_2, \\ \exp \left\{ -\frac{1}{2}(\gamma^* - \gamma_2)^2 I \right\} & \text{cuando } \gamma_2 < \gamma^*. \end{cases}$$

De estas igualdades se deduce que  $r(\gamma^*)$  es una función continua monótonamente decreciente de  $|\gamma^* - \bar{\gamma}|$  (ella es simétrica respecto al punto  $\gamma^* = \bar{\gamma}$ ). Por eso la desigualdad  $r(\gamma^*) > c$  equivale a la desigualdad  $|\gamma^* - \bar{\gamma}| < c'$ . Como  $\gamma^* = Y \in \Phi_{0, I^{-1}}$ , entonces se cumplen las condiciones del lema 14.1.  $\triangleleft$

**3. Criterio asintóticamente minimax para hipótesis semejantes referentes a un parámetro multidimensional.** Examinemos ahora el parámetro multidimensional  $\theta$ . En este caso, el c.a.u.m.p. para verificar la hipótesis  $H_1 = \{\theta \in \Theta_1\}$  frente a  $H_2 = \{\theta \in \Theta_2\}$ , por lo general, no existe, y examinaremos el problema de construcción de los criterios asintóticamente minimax.

Al principio es necesario exponer una observación general para simplificar los razonamientos posteriores. Dicha observación consiste en que el referido problema de verificación de las hipótesis siempre se puede "reparametrizar" (o sea, introducir un nuevo parámetro) de modo que la matriz de información  $I = I(\theta_1)$  en el punto  $\theta_1$  se convierta en matriz unidad. Para esto es suficiente (véase el § 2.1) efectuar una transformación lineal e introducir un nuevo parámetro  $\beta$  mediante la igualdad

$$\theta = \beta I^{-1/2}.$$

Entonces, la matriz de información de Fisher  $J(\beta)$  para la familia paramétrica  $\mathbf{P}_{\beta, I^{-1/2}}$  será igual, en el punto  $\beta_1 = \theta_1 I^{1/2}$ , a

$$J(\beta_1) = I^{-1/2} I I^{-1/2} = E.$$

En este apartado nos será más fácil examinar el parámetro  $\beta$ . Siempre podremos volver al parámetro inicial con ayuda de la transformación lineal inversa.

Así pues, supongamos que  $I = I(\theta_1) = E$ , y examinemos el problema  $A$  de verificación de la hipótesis

$$H_1 = \{|\theta - \theta_1| \leq an^{-1/2}\} \text{ frente a } H_2 = \{|\theta - \theta_1| \geq bn^{-1/2}\}, \quad a < b \quad (4)$$

a base de la muestra  $X \in \mathbf{P}_\theta$ .

**Teorema 3.** *El criterio de relación de verosimilitud  $\hat{\pi}$  definido en (1) para el valor correspondiente de  $c$  y para  $\Theta_1 = \{\theta: |\theta - \theta_1| \leq an^{-1/2}\}$   $\Theta_2 = \{\theta: |\theta - \theta_1| \geq bn^{-1/2}\}$  es asintóticamente equivalente, para cualesquiera  $0 \leq a < b < \infty$ , a los criterios*

$$\frac{f_{\theta_1}(X)}{f_{\theta_2}(X)} > c, \quad (5)$$

$$|\gamma^*| \equiv |(\hat{\theta}^* - \theta_1)\sqrt{n}| > c_\varepsilon, \quad (6)$$

donde  $c_\varepsilon$  es la solución, respecto a  $c$ , de la ecuación

$$p_c(a) \equiv \mathbf{P}(\xi_1 + a)^2 + \xi_2^2 + \dots + \xi_k^2 > c^2) = \varepsilon, \quad (7)$$

y es el criterio asintóticamente minimax de nivel asintótico  $1 - \varepsilon$  para verificar las hipótesis  $H_1$  y  $H_2$  definidas en (4). Las variables aleatorias  $\xi_i$  en (7) son independientes,  $\xi_i \in \Phi_{0,1}$ , la potencia límite garantizada de los criterios  $\pi$ , (5), (6) es igual a  $p_c(b)$ .

**Demostración.** Aquí el problema  $B$  consistirá en verificar, valiéndose de la observación  $Y \in \Phi_{\gamma, E}$ , la hipótesis  $k_1 = \{|\gamma| \leq a\}$  frente a  $k_2 = \{|\gamma| \geq b\}$ . En el ejemplo 9.1 hemos visto que en este problema existe un criterio minimax de nivel  $1 - \varepsilon$  que tiene la forma

$$|Y| > c_\varepsilon.$$

Para construir este criterio hemos utilizado el teorema 9.1. Esto significa que las condiciones del criterio 14.2 se cumplen. Por consiguiente, el criterio

$$|\gamma^*| > c_\varepsilon$$

será un criterio asintóticamente minimax de nivel asintótico  $1 - \varepsilon$  para el problema  $A$ .

El criterio de relación de verosimilitud (1) aquí tendrá la forma

$$R(X) = \frac{\sup_{|u| \geq b} Z_1(un^{-1/2})}{\sup_{|u| \leq a} Z_1(un^{-1/2})} > c. \quad (8)$$

Observando exactamente los razonamientos utilizados en las demostraciones de los teoremas 1 y 2, obtendremos que  $R(X) = T_n(X) + \varepsilon_n(X)$ ,  $\varepsilon_n(X) \xrightarrow{P_\theta} 0$ , donde

$$T_n(X) = r(\gamma^*) = \frac{\sup_{|u| \geq b} \exp\left\{-\frac{1}{2}|\gamma^* - u|^2\right\}}{\sup_{|u| \leq a} \exp\left\{-\frac{1}{2}|\gamma^* - u|^2\right\}}.$$

De aquí, como antes, se deduce la continuidad absoluta de la distribución  $r(Y)$  y la equivalencia asintótica de los criterios  $R(X) > c$  y  $T(X) > c$ . Este último equivale al criterio

$$|\gamma^*| > c',$$

el cual, cuando  $c' = c_\varepsilon$ , será un criterio de nivel  $1 - \varepsilon$ . Según el teorema 14.2 (véase (14.10)), éste tendrá una potencia límite garantizada igual a  $p_\varepsilon(b)$  (véase el teorema 9.2).  $\triangleleft$

**Observación 1.** Si volvemos al parámetro inicial (hasta la reparametrización que transforma  $I(\theta_1)$  en una matriz unidad), obtendremos que la afirmación del teorema será válida respecto a las hipótesis  $H_i = \{\theta \in \Theta_i\}$ , donde (compárese con el ejemplo 9.2 cuando  $\sigma^2 = I^{-1}$ )

$$\Theta_1 = \{\theta: (\theta - \theta_1)I(\theta_1)(\theta - \theta_1)^T \leq a^2 n^{-1}\},$$

$$\Theta_2 = \{\theta: (\theta - \theta_1)I(\theta_1)(\theta - \theta_1)^T \geq b^2 n^{-1}\}.$$

El criterio (6) adoptará la forma

$$(\hat{\theta}^* - \theta_1)I(\theta_1)(\hat{\theta}^* - \theta_1)^T n > c_\varepsilon^2$$

o bien (véase el teorema 13.2)

$$L'(X, \theta_1)I^{-1}(\theta_1)(L'(X, \theta_1))^T > c_\varepsilon^2. \quad (9)$$

El criterio de relación de verosimilitud no variará, evidentemente, ya que el valor máximo de  $f_\theta(X)$  en la región  $\Theta_i$  no depende de la sustitución de las variables (después de la transformación correspondiente de las regiones de  $\Theta_i$ ).

También cabe señalar que la forma del criterio (9) es, a veces, más cómoda que la del (5) y el (6), puesto que no está relacionada con los cálculos de  $\hat{\theta}^*$ . Sustituciones análogas pueden hacerse con arreglo a los criterios (2) y (3) en los teoremas 1 y 2. Le dejamos al lector que las haga él mismo.

**Observación 2.** De un modo absolutamente análogo al teorema 3 se puede construir el criterio asintóticamente minimax para los problemas  $A$  que pueden ser reducidos al problema  $B$  examinado en el ejemplo 9.5.

**Observación 3.** En el § 13 hemos construido el criterio asintóticamente bayesiano para verificar la hipótesis  $\{\theta = \theta_1\}$  frente a  $\{\theta \neq \theta_1\}$ , el cual tiene la forma del c.r.v.

$$\frac{f_{\hat{\theta}^*}(X)}{f_{\theta_1}(X)} > c.$$

Ahora bien, este criterio, siendo el c.a.b., también posee propiedad asintóticamente minimax al verificar la hipótesis  $\{\theta = \theta_1\}$  frente a  $\{(\theta - \theta_1)I(\theta_1) \times (\theta - \theta_1)^T \geq b^2 n^{-1}\}$  para cualquier  $b > 0$ .

**4. Criterio asintóticamente minimax de pertenencia de la muestra a una subfamilia paramétrica.** Ahora examinaremos el c.r.v. en un problema más complejo de verificación de la hipótesis  $H_1 = \{\theta \in \Theta_1\}$  frente a  $H_2 = \{\theta \in \Theta_2\}$  cuando la dimensión  $l$  del subconjunto  $\Theta_1$  es positiva pero me-

nor que  $k > 1$ . Supongamos que tenemos la función suave  $\theta = g(\alpha)$  del parámetro  $l$ -dimensional ( $l < k$ )  $\alpha \in A_1 \subset R^l$ . La imagen del conjunto  $A_1$  en  $\Theta$ , engendrada por la aplicación de  $g$ , podemos designarla por  $\Theta_1$ . El problema consiste en verificar la hipótesis  $H_1 = \{\theta \in \Theta_1\}$  de que el parámetro  $\theta$  pertenece a la "curva"  $\Theta_1$  (o bien de que  $X \in P_{g(\alpha)}$  para cierto  $\alpha \in A_1$ ) frente a la alternativa adicional  $\{X \in P_g; \theta \notin \Theta_1\}$ , así que en este caso  $\Theta_2 = \Theta \setminus \Theta_1$ . Con otras palabras, éste es el problema de verificación de la pertenencia de la muestra  $X$  a la subfamilia paramétrica de distribuciones  $\{P_{g(\alpha)}; \alpha \in A_1\}$ .

A esta clase de problemas pertenecen, por ejemplo, los problemas ya conocidos de verificación de la hipótesis  $\{X \in \Phi_{\alpha_0, \sigma^2}\}$  frente a  $\{X \in \Phi_{\alpha, \sigma^2}; \alpha \neq \alpha_0\}$  para un valor de  $\alpha_0$  dado y un valor de  $\sigma^2$  desconocido, o los problemas de verificación de la hipótesis  $\{X \in \Phi_{\alpha, \sigma_0}\}$  frente a  $\{X \in \Phi_{\alpha, \sigma^2}; \sigma = \sigma_0\}$  para un valor de  $\sigma_0$  dado y un valor de  $\alpha$  desconocido, y otros.

En cuanto a la curva  $\theta = g(\alpha)$  en  $\Theta$ , supondremos que la misma es dos veces continuamente derivable, y que la matriz  $G = \|\partial g_i(\alpha) / \partial \alpha_j\|$  ( $i = 1, \dots, k; j = 1, \dots, l; g_i(\alpha)$  y  $\alpha_i$  son las coordenadas de  $g(\alpha)$  y  $\alpha$ , respectivamente) tiene el rango  $l$ . Esto quiere decir que podemos realizar la sustitución biunívoca derivable del parámetro (la reparametrización del problema) de modo que las primeras  $l$  coordenadas (sin limitar la generalidad se puede suponer que las mismas constituyen  $\alpha = (\alpha_1, \dots, \alpha_l)$ ) determinen la posición del punto  $\theta$  en la curva  $\Theta_1$ , y las demás (designémoslas por  $\beta = (\beta_1, \dots, \beta_{k-l})$ ) que determinen la posición de  $\theta$  en el "plano" (subespacio), digamos, ortogonal (pero no obligatoriamente) a la "curva"  $g(\alpha)$  en el punto  $\alpha$ . Entonces, el problema se reduce a la verificación de la hipótesis  $\{\beta = 0\}$  frente a  $\{\beta \neq 0\}$  siempre que exista el subparámetro "obstaculizador" desconocido  $\alpha$ .

En este caso examinaremos las hipótesis semejantes, suponiendo que  $\beta = \gamma'' n^{-1/2}$ , y comprobaremos la hipótesis  $\{\gamma'' = 0\}$  frente a  $\{\gamma'' \neq 0\}$ , o frente a

$$\{\gamma'' M_2(\alpha) \gamma''^T \geq b^2\} \quad (10)$$

para  $b > 0$  y para cierta matriz definida positivamente  $M_2(\alpha)$ .

En las coordenadas iniciales, el último problema corresponderá a la verificación de la hipótesis  $H_1 = \{\theta \in \Theta_1\}$  frente a las alternativas semejantes, cuando el parámetro  $\theta$  se sitúe en el entorno  $n^{-1/2}$  de la curva  $\Theta_1$  y permanezca fuera de cierto "tubo" que contiene  $\Theta_1$  y corresponde al conjunto (10). También es posible otra variante de planteamiento del problema de verificación de las hipótesis semejantes, la cual parte del hecho de que el parámetro  $\theta$  está "localizado" y sabemos que el mismo se halla en el entorno de cierto punto  $\theta_0 = g(\alpha^0)$ ,  $\alpha^0 \in A_1$ . Entonces, el nuevo parámetro  $\tau = (\beta, \alpha - \alpha^0)$  será localizado cerca del punto  $\tau_0 = (0, 0)$ . Pongamos

$\alpha - \alpha^0 = \gamma' n^{-1/2}$ ,  $\beta = \gamma'' n^{-1/2}$  y comprobemos la hipótesis  $\{\gamma'' = 0\}$  frente a  $\{\gamma'' \neq 0\}$  o frente a  $\{\gamma'' M_2(\alpha^0) \gamma''^T \geq b^2\}$  al disponer del parámetro localizador  $\gamma'$ .

Los resultados que nos interesan en estos dos planteamientos de los problemas coinciden prácticamente. Sin embargo, es más cómodo investigar el segundo planteamiento, puesto que en este caso disponemos de todos los resultados previos necesarios. La suposición acerca de la localización del parámetro  $\theta$  tiene carácter convencional, y la forma de las afirmaciones obtenidas más abajo no dependerá de  $\theta_0$ .

Así pues, consideraremos que el nuevo parámetro  $\tau = (\alpha - \alpha^0, \beta)$  tiene la forma

$$\tau = \gamma n^{-1/2}, \quad \gamma = (\gamma', \gamma''),$$

y comprobaremos la hipótesis  $H_1 = \{\gamma'' = 0\}$  frente a  $H_2 = \{\gamma'' M_2 \gamma''^T \geq b^2\}$ , donde en calidad de  $M_2 = M_2(\alpha^0)$  tomaremos la matriz de información de Fisher para la familia paramétrica  $\{P_{\theta(0, \beta)}\}$  en el punto  $\beta = 0$ , donde  $\theta(\tau) = \theta((\alpha - \alpha^0, \beta))$  es la función que reconstruye  $\theta$  según el valor de  $\tau = (\tau', \tau'')$ .

**Teorema 4.** *Supongamos que  $\theta_0 = g(\alpha^0)$  es un punto interior de  $\theta$ , y que en el entorno de este punto se cumplen las condiciones (RR). Supongamos también, que la función  $g(\alpha)$  es dos veces continuamente derivable en el punto  $\alpha^0$  y que la matriz  $G = \|\partial g_i(\alpha) / \partial \alpha_j\|_{\alpha=\alpha^0}$  tiene el rango 1. Entonces, para  $\Theta_1$  y  $\Theta_2$  definidas anteriormente, así como para  $c$  correspondiente, el criterio de relación de verosimilitud equivale asintóticamente a los criterios*

$$R_1(X) \equiv \frac{f_{\hat{\theta}^*}(X)}{f_{g(\hat{\alpha}^*)}(X)} > e^{h_c/2}, \quad (11)$$

$$\begin{aligned} (\hat{\theta}^* - g(\hat{\alpha}^*)) I(g(\hat{\alpha}^*)) (\hat{\theta}^* - g(\hat{\alpha}^*))^T &> h_c n^{-1}, \\ (\hat{\theta}^* - g(\hat{\alpha}^*)) I(\hat{\theta}^*) (\hat{\theta}^* - g(\hat{\alpha}^*))^T &> h_c n^{-1} \end{aligned} \quad (12)$$

y es el criterio asintóticamente minimax de nivel asintótico  $1 - \varepsilon$  para verificar la hipótesis  $H_1 = \{\theta \in \Theta_1\} = \{\gamma'' = 0\}$  frente a  $H_2 = \{\gamma'' M_2 \gamma''^T \geq b^2\}$ .

La distribución de la estadística  $2 \ln R_1(X)$  para  $X \in P_{g(\alpha^0)}$  (o sea, para la hipótesis  $H_1$ ) converge, cuando  $n \rightarrow \infty$ , hacia la distribución  $\chi^2$  de  $k - 1$  grados de libertad (y, por consiguiente, no depende de  $f_{\theta}$  y  $\alpha^0$ ). De acuerdo con esto,  $h_c$  en (11) y (12) significa la cuantila de orden  $1 - \varepsilon$  de la distribución  $H_{k-1}$ .

La potencia asintótica garantizada del c.r.v. es igual a  $P((\xi_1 + b)^2 + \xi_2^2 + \dots + \xi_{k-1}^2 > h_c)$ , donde  $\xi_i \in \Phi_{0,1}$  y son independientes.

Vemos que los criterios asintóticamente minimax (11) y (12) no están de ningún modo relacionados con  $\alpha^0$ .

**Observación 4.** La hipótesis  $H_2$ , en términos del parámetro inicial  $\theta$  puede ser escrita de la forma siguiente:

$$H_2 = \{ \inf_{\gamma'} (\theta - g(\alpha^\circ + \gamma' n^{-1/2}) I(g(\alpha^\circ)) (\theta - g(\alpha^\circ + \gamma' n^{-1/2}))^T \geq b^2 n^{-1} \}.$$

Recordemos que consideramos limitado el conjunto  $\Gamma_i$ , ya que aquí  $(\theta - \theta_0) \leq N n^{-1/2}$ ,  $|\gamma'| \leq N$  para cierto  $N > 0$ .

**Observación 5.** Como veremos de la demostración, la afirmación del teorema conservará por completo su validez si la hipótesis  $H_1 = \{\gamma'' = 0\}$  es sustituida por  $H_1 = \{\gamma'' M_2 \gamma''^T \leq a^2\}$ ,  $a < b$ , con la sustitución respectiva del conjunto  $\Theta_1$ .

**Demostración del teorema 4.** En calidad de criterio "principal" aquí examinaremos el criterio (11) equivalente a (1) y más cómodo en cuanto a su forma. Además estableceremos la equivalencia asintótica del mismo respecto al criterio asintóticamente minimax, y luego, su equivalencia asintótica a (12).

Examinemos las distribuciones  $P_\theta$  y  $P_{g(\theta)}$  como dependientes de los parámetros  $\tau = (\tau', \tau'')$  y  $\alpha = \tau' + \alpha^\circ$ , respectivamente. Pongamos  $\tau = \gamma n^{-1/2}$ ,  $\gamma = (\gamma', \gamma'')$ , de modo que  $\tau' = \gamma' n^{-1/2}$ ,  $\tau'' = \gamma'' n^{-1/2}$ , y comprobemos la hipótesis  $H_1 = \{\gamma'' = 0\}$  frente a  $H_2 = \{\gamma'' M_2 \gamma''^T \geq b^2\}$ , donde  $M_2$  es la matriz de información de Fisher para la familia  $P_{g(\theta)}$  en el punto  $\alpha^\circ$ . Efectuemos ahora una transformación más del parámetro, semejante a la realizada en el ejemplo 9.4 y la cual convierte las matrices de información en matrices de unidad. Supongamos que  $\varrho = \tau \Lambda$  y que, respectivamente,  $\delta = \gamma \Lambda$  ( $\varrho = \delta n^{-1/2}$ ), donde  $\Lambda$  es una matriz triangular, semejante a la descrita en el ejemplo 9.4 y la cual posee las propiedades siguientes:

$$J^{-1} = \Lambda^T M^{-1} \Lambda = E, \quad J_2^{-1} = \Lambda_2^T M_2^{-1} \Lambda_2 = E,$$

donde  $J, M, J_2, M_2$  son matrices de información en el punto  $\theta_0$  para  $\varrho, \tau, \varrho'', \tau''$ , respectivamente (las tildes superiores y las designaciones tienen el mismo sentido que en  $\tau', \tau'', \gamma', \gamma''$ ),  $\Lambda_2$  es la matriz del orden  $(k-l) \times (k-l)$ , formada por los últimos  $k-l$  renglones y columnas de la matriz  $\Lambda$ , de modo que  $\varrho'' = \tau'' \Lambda_2$ ,  $\delta'' = \gamma'' \Lambda_2$ .

En nuevos parámetros las hipótesis  $H_1$  y  $H_2$  se escribirán de la forma siguiente:

$$H_1 = \{\delta'' = 0\}, \quad H_2 = \{|\delta''| \geq b\}.$$

De las propiedades de las transformaciones realizadas se deduce que  $\theta = \theta_0$  es una función biunívoca de  $\varrho$  y que todas las familias paramétricas examinadas (incluso con parámetros  $\varrho', \varrho''$ ) satisfacen las condiciones (RR). Pongamos  $\varrho_0 = \theta^{-1}(\theta_0)$  (ésta es la solución de la ecuación  $\theta(\varrho) = \theta_0$ ),

$$Z_0(t) = f_{\theta_0}(\varrho_0 + t(X)/f_{\theta_0}(X), \quad Y_0(u) = \ln Z_0(u n^{-1/2}).$$

Hagamos uso del teorema 2.29.3. Para  $|u| \leq \delta_n \sqrt{n}$ ,  $X \in P_{\theta_0}$  obtenemos

$$\varrho = \varrho_0 + \delta n^{-1/2},$$

$$Y_0(u) = (\xi_n + \delta, u) - \frac{1}{2}(u, u) + (|u|^2 + |\delta^2|) \varepsilon_n(X, u, \delta), \quad (13)$$

donde  $|\varepsilon_n(X, u, \delta)| \leq \varepsilon_n(X) \rightarrow 0$  uniformemente respecto a  $\delta$  para  $|\delta| \leq \delta_n \sqrt{n}$ , donde  $\delta_n$  es

una sucesión arbitraria que converge a cero. En estas igualdades hemos utilizado el hecho de que la matriz de información para el parámetro  $\varrho$  es una matriz unidad. El vector  $\xi_n$  es

el vector de las funciones derivadas  $n^{-1/2}L(X, \theta(q))$  respecto a  $q_j$  en el punto  $q = q_0 + \delta\pi^{-1/2}$ , de modo que  $\xi_n \in \Phi_{0,E}$  uniformemente respecto a  $q$  (respecto a  $\delta$ ) cuando  $|\delta| \leq \delta_n \sqrt{n}$ . (En vista de la suposición de que  $(\theta - \theta_0) \sqrt{n}$  está limitada, aquí y más adelante es suficiente establecer la uniformidad de convergencia para  $|\delta| \leq N$ , cuando  $N$  se ha registrado arbitrariamente. Sin embargo, nada nos molesta establecer también la uniformidad necesaria en una región más amplia  $|\delta| \leq \delta_n \sqrt{n} \rightarrow \infty$ .)

Ahora supongamos que  $u = (u', u'')$ ,  $u'' = 0$  en (13). Entonces, según el acuerdo anterior respecto a los símbolos con tildes, podemos escribir

$$Y_0(u', 0) = (\xi_n' + \delta', u') - \frac{1}{2}(u', u') + (|u'|^2 + |\delta|^2)e_n(X, u', \delta). \quad (14)$$

De (13) y (14) se deduce que los valores máximos de  $Y_0(u)$  y  $Y_0(u', 0)$  se alcanzan, respectivamente, para

$$\begin{aligned} u &= (\xi_n + \delta)(E + \varepsilon_n(X, \delta)), \\ u' &= (\xi_n' + \delta')(E + \varepsilon_n^{(1)}(X, \delta)), \end{aligned} \quad (15)$$

donde  $\varepsilon_n(X, \delta) \rightarrow 0$ ,  $\varepsilon_n^{(1)}(X, \delta) \rightarrow 0$  uniformemente en  $\delta$ ,  $|\delta| < \delta_n \sqrt{n}/2$ . Tan sólo es necesario notar que la probabilidad de grandes valores de  $|\xi_n + \delta|$  es uniformemente pequeña, ya que  $\xi_n + \delta \in \Phi_{\delta,E}$  uniformemente en  $\delta$ ,  $|\delta| < \delta_n \sqrt{n}$  y  $P_\theta(|\xi_n + \delta| > \delta_n \sqrt{n}) \rightarrow 0$  uniformemente en  $\delta$ ,  $|\delta| < \delta_n \sqrt{n}/2$ .

Volvamos ahora a examinar el c.r.v. Para  $\theta = \theta_{(q)}$ ,  $X \in P_\theta$ ,  $q = q_0 + \delta\pi^{-1/2}$  tenemos

$$\begin{aligned} R_1(X) &= \frac{\sup_{\theta} f_\theta(X)}{\sup_{\alpha} f_{g(\alpha)}(X)} = \frac{\sup_u e^{Y_0(u)}}{\sup_{u'} e^{Y_0(u', 0)}} = \\ &= \frac{\exp \left\{ \frac{1}{2} |\xi_n + \delta|^2 + \varepsilon_n(X, \delta) \right\}}{\exp \left\{ \frac{1}{2} |\xi_n' + \delta'|^2 + \varepsilon_n^{(1)}(X, \delta) \right\}} = \exp \left\{ \frac{1}{2} |\xi_n'' + \delta''|^2 + \varepsilon_n''(X, \delta) \right\}, \end{aligned} \quad (16)$$

donde la función  $\varepsilon_n$  con diferentes índices converge a cero en  $P_\theta$ -probabilidad uniformemente cuando  $|\delta| < \delta_n \sqrt{n}$ :

$$2 \ln R_1(X) \rightarrow |Y'' + \delta''|^2, \quad Y \in \Phi_{0,E}, \quad (17)$$

uniformemente en  $\delta$ .

En vista de que para  $\theta = g(\alpha)$  con la necesidad de  $\delta'' = 0$ , de aquí resulta la afirmación del teorema respecto a la estadística  $2 \ln R_1(X)$ .

Recordemos ahora que (véase el teorema 2.29.3)  $\xi_n = u^*(E + \varepsilon_n(X, \delta))$ , donde  $u^* = (\hat{q}^* - q_0) \sqrt{n}$ ,  $\hat{q}^*$  es la e.v.m. para el parámetro  $q$ . De aquí y de la igualdad  $q_0 = 0$ , suponiendo  $\delta^* = (\hat{q}^* - q_0) \sqrt{n}$ , obtenemos

$$\begin{aligned} \xi_n + \delta &= \sqrt{n}(\hat{q}^* - q_0) + u^* \varepsilon_n(X, \delta) = \sqrt{n}(\hat{q}^* - q_0) + \\ &\quad + u^* \varepsilon_n(X, \delta) = \delta^* + u^* \varepsilon_n(X, \delta) \in \Phi_{\delta,E}, \\ \xi_n' + \delta' &= (\delta^*)' + (u^* \varepsilon_n(X, \delta))'. \end{aligned}$$

Por lo tanto, el segundo miembro en (16) también puede ser escrito en la forma  $\exp \left\{ \frac{1}{2} |(\delta^*)''|^2 + \varepsilon_n''(X, \delta) \right\}$ ,  $\varepsilon_n''(X, \delta) \rightarrow 0$ . Esto quiere decir que el criterio

$$|(\delta^*)''|^2 > h_\varepsilon \quad (18)$$

y el c.r.v. son asintóticamente equivalentes, o sea,

$$\limsup_{n \rightarrow \infty} \alpha \mathbf{P}_{g(\alpha)}(R_1(X) > e^{h_\varepsilon/2}) = \limsup_{n \rightarrow \infty} \alpha \mathbf{P}_{g(\alpha)}(|\delta^*|'' > h_\varepsilon) = \varepsilon,$$

$$\begin{aligned} \limsup_{n \rightarrow \infty} \sup_{\theta \in \Theta_2} P_\theta(R_1(X) > e^{h_\varepsilon/2}) &= \limsup_{n \rightarrow \infty} \sup_{\theta \in \Theta_2} P_\theta(|\hat{\delta}^*|^2 > h_\varepsilon) = \\ &= \sup_{|\delta^*| > b^2} P(|Y^n + \delta^*|^2 > h_\varepsilon) = P((y_1 + b)^2 + y_2^2 + \dots + y_{k-1}^2 > h_\varepsilon), \end{aligned}$$

donde  $y_i \in \Phi_{0,1}$  son independientes.

Demostremos ahora que el criterio (18) es un criterio asintóticamente minimax de nivel asintótico  $1 - \varepsilon$ . Hagamos uso del teorema 14.2. En nuestro caso,  $\delta^* = (\hat{\rho}^* - \rho_0)\sqrt{n} \in \Phi_{\delta, \varepsilon}$ . El problema  $B$  para  $Y \in \Phi_{\delta, \varepsilon}$  se ha examinado en los ejemplos 9.3 y 9.4. Allí hemos establecido que el criterio

$$|Y^n|^2 > h_\varepsilon$$

es minimax y de nivel  $1 - \varepsilon$ . Por consiguiente, de acuerdo con el teorema 14.2, el criterio (18) es asintóticamente minimax.

Para terminar la demostración nos queda establecer la equivalencia asintótica de (11) y (12). Esta equivalencia se deduce fácilmente de los resultados del § 2.29 y del lema 14.1.  $\triangleleft$

**Ejemplo 1.** Supongamos que  $X \in \Phi_{\lambda, \sigma^2}$ , donde  $\lambda$  y  $\sigma^2$  son parámetros escalares. (Aquí utilizaremos el símbolo  $\lambda$  en vez del  $\alpha$  tradicional para que no haya confusión con el argumento de la función  $g(\alpha)$ ). Es necesario verificar la hipótesis  $\{\lambda = \lambda_0\}$  frente a  $\{\lambda \neq \lambda_0\}$  o frente a  $\{|\lambda - \lambda_0| \geq b n^{-1/2}\}$ ,  $b > 0$ , cuando  $\sigma$  se desconoce. Sabemos que en este caso las e.v.m. tienen la forma siguiente. Si ambas componentes  $\lambda$  y  $\sigma^2$  del vector  $\theta = (\lambda, \sigma^2)$  se desconocen, entonces la e.v.m. para  $\theta$  es

$$\hat{\theta}^* = (\lambda, \sigma^2)^* = (\bar{x}, S^2), \quad S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Si  $\lambda = \lambda_0$ , la e.v.m. para  $\sigma^2$  tiene la forma  $(\sigma^2)^* = S_1^2 = \frac{1}{n} \sum (x_i - \lambda_0)^2$ , así que  $g(\hat{\alpha}^*) = (\lambda_0, S_1^2)$ . Como

$$f_\theta(X) = (\sqrt{2\pi\sigma})^{-n} \exp\left\{-\frac{1}{2\sigma^2} \sum (x_i - \lambda)^2\right\},$$

el criterio de la relación de verosimilitud (11) tiene la forma

$$S_1^2/S^2 > c.$$

En virtud de la igualdad  $S_1^2 = S^2 + (\bar{x} - \lambda_0)^2$ , este criterio equivale al criterio

$$|\bar{x} - \lambda_0|/S > c_1. \quad (19)$$

Pero éste es el conocido criterio de Student que hemos examinado anteriormente (las propiedades óptimas de este criterio se exponen en el § 7).

Es fácil comprobar que el criterio (12) tendrá esa misma forma. En efecto, en el § 2.16 hemos visto que la matriz  $I(\theta)$  para la familia  $\Phi_{\lambda, \sigma^2}$  tiene la forma

$$I(\theta) = \begin{pmatrix} \sigma^{-2} & 0 \\ 0 & (2\sigma^4)^{-1} \end{pmatrix}.$$



En nuestro caso  $\hat{\theta}^* - g(\hat{\alpha}^*) = (\bar{x} - \lambda_0, S^2 - S_1^2) = (\bar{x} - \lambda_0, n(\bar{x} - \lambda_0)^2)$ ,

$$I^{1/2}(\hat{\theta}^*) = \begin{pmatrix} S^{-1} & 0 \\ 0 & (\sqrt{2} S^2)^{-1} \end{pmatrix}.$$

Como en el primer miembro (12) figura el cuadrado de la norma  $|g(\hat{\alpha}^*) - \hat{\theta}^*| I^{1/2}(\hat{\theta}^*)|^2$ , el criterio (12) tendrá la forma

$$\frac{(\bar{x} - \lambda_0)^2}{S^2} + \frac{(\bar{x} - \lambda_0)^4}{2S^4} > c_2$$

que, evidentemente, equivale a (19).

Si en vez de  $I(\hat{\theta}^*)$  aquí utilizamos  $I(g(\hat{\alpha}^*))$ , obtendremos el criterio asintóticamente equivalente

$$|\bar{x} - \lambda_0|/S_1 > c_1.$$

**Ejemplo 2.** Supongamos que  $X \in \Phi_{\lambda, \sigma^2}$ . Se necesita verificar la hipótesis  $\{\sigma = \sigma_0\}$  frente a  $\{|\sigma^2 - \sigma_0^2| \geq bn^{-1/2}\}$  cuando se desconoce  $\lambda$ . Aquí, la e.v.m.  $\hat{\theta}^*$  para  $\theta = (\lambda, \sigma^2)$  será, evidentemente, la misma que en el ejemplo precedente. Si  $\sigma = \sigma_0$ , entonces  $\hat{\lambda}^* = \bar{x}$ , de modo que  $g(\hat{\alpha}^*) = (\bar{x}, \sigma_0^2)$ ,  $\hat{\theta}^* - g(\hat{\alpha}^*) = (0, \sigma_0^2 - S^2)$ .

Los criterios (11) (o, que es lo mismo, el criterio de relación de verosimilitud) tienen la forma

$$(S^2 - \sigma_0^2)^2/\sigma_0^4 > 2h_e n^{-1},$$

que equivale, evidentemente, a

$$|S^2/\sigma_0^2 - 1| > \sqrt{2h_e n^{-1}},$$

donde  $\Phi_{0,1}((h_e^{1/2}, \infty)) = \varepsilon/2$ . Este criterio también ya fue examinado en el § 7.

## § 16. Criterio $\chi^2$ . Verificación de las hipótesis por los datos agrupados

**1. Criterio  $\chi^2$ . Propiedades de optimización asintótica.** El criterio  $\chi^2$  como tal se destina a verificar, basándose en la muestra  $X$  de la distribución

polinomial  $B_\theta$ ,  $\theta = (\theta_1, \dots, \theta_r)$ ,  $\sum_{i=1}^r \theta_i = 1$ , la hipótesis simple  $H_1 = \{\theta = p\}$

frente a la alternativa adicional  $H_2 = \{\theta \neq p\}$ ,  $p = (p_1, \dots, p_r)$ . La distribución polinomial  $B_\theta$  se describe por las probabilidades  $\theta_i = \mathbf{P}(A_i)$ ,  $i = 1, \dots, r$ , de que se produzca, en cada prueba aislada, uno de los  $r$  sucesos disjuntos  $A_1, \dots, A_r$ . El elemento  $x_j$  de la muestra  $X$  de esta distribución puede representarse como uno de los vectores  $e_1, \dots, e_r$  con  $r$  coordenadas. La coordenada del vector  $e_k(r-1)$  es igual a cero, y la coordenada del número  $k$  es igual a 1. En este caso  $x_j = e_k$  si se ha producido el suceso

$A_k$ . Designemos por  $\nu_k$  el número de veces que se produce el suceso  $A_k$  en  $n$  pruebas independientes. Entonces  $\nu = (\nu_1, \dots, \nu_r) = \sum_{i=1}^n x_i$  es una estadística suficiente para  $\theta$ , ya que la función de verosimilitud  $f_\theta(X)$  tiene la forma

$$f_\theta(X) = \prod_{i=1}^r \theta_i^{\nu_i}. \quad (1)$$

La estadística  $\chi^2$  es, por definición,

$$\chi^2(X) = \sum_{i=1}^r \frac{(\nu_i - n p_i)^2}{n p_i},$$

y el conjunto crítico del criterio  $\chi^2$  (la región de aceptación de  $H_2$ ) tiene la forma

$$\chi^2(X) \geq c,$$

donde  $c$  se elige según el nivel de significación establecido.

Ahora examinemos más detalladamente el problema antes enunciado acerca de la verificación de la hipótesis  $H_1 = \{\theta = p\}$  frente a  $H_2 = \{\theta \neq p\}$ .

Está claro que las distribuciones  $\{B_\theta\}$  forman una familia paramétrica que no depende del parámetro  $k = (r - 1)$ -dimensional  $(\theta_1, \dots, \theta_{r-1})$ ; el valor de  $\theta_r$  se define por la igualdad  $\theta_r = 1 - \sum_{i=1}^{r-1} \theta_i$ . El vector  $(\theta_1, \dots, \theta_{r-1})$ ,

al igual que el  $(\theta_1, \dots, \theta_r)$ , será designado con la letra  $\theta$ . Esto no provocará equivocaciones. La región  $\theta$  no es otra cosa sino el simplex  $\theta_i \geq 0$ ,

$i = 1, \dots, r - 1$ .  $\sum_{i=1}^{r-1} \theta_i \leq 1$ . La función logarítmica de verosimilitud  $L(X, \theta)$  es igual a

$$L(X, \theta) = \sum_{k=1}^r \nu_k \ln \theta_k = \sum_{i=1}^n l(x_i, \theta). \quad (2)$$

La familia  $\{B_\theta\}$  satisface las condiciones  $(A_0)$ ,  $(A_n)$ ,  $(A_c)$ , y también las condiciones de regularidad  $(RR)$  en cualquier punto interior de  $\Theta$ , o sea, en cualquier punto  $\theta$  para el cual todos  $\theta_i > 0$ . Efectivamente, en nuestro caso

$$l(x_1, \theta) = \ln \theta_j \quad \text{para } x_1 = e_j; \\ \frac{\partial l(x_1, \theta)}{\partial \theta_j} = \begin{cases} \theta_j^{-1} & \text{para } x_1 = e_j, \\ -\theta_r^{-1} & \text{para } x_1 = e_r, \\ 0 & \text{para } x_1 \neq e_j, x_1 \neq e_r, \end{cases} \quad (3)$$

$$\frac{\partial^2 l(x_1, \theta)}{\partial \theta_1 \partial \theta_j} = \begin{cases} \frac{\delta_{ij}}{\theta_i \theta_j} & \text{para } x_1 = e_j, \\ -\theta_r^{-2} & \text{para } x_1 = e_r, \\ 0 & \text{para } x_1 \neq e_j, x_1 \neq e_r, \end{cases} \quad (4)$$

donde  $\delta_{ij}$  es el símbolo de Kronecker. De estas fórmulas se deduce que

$$\frac{\partial^2 l(x_1, \theta)}{\partial \theta_i \partial \theta_j} = -\frac{\partial l(x_1, \theta)}{\partial \theta_i} \cdot \frac{\partial l(x_1, \theta)}{\partial \theta_j}, \quad i, j \leq r-1.$$

Parte de las condiciones (RR) relacionadas con la existencia de las esperanzas matemáticas, aquí se cumplen evidentemente, ya que en nuestro caso el conjunto  $\chi$  es finito.

De (3) o (4) se deduce

$$I(\theta) = \|I_{ij}(\theta)\| = -\left\| \mathbf{M}_\theta \frac{\partial^2 l(x_1, \theta)}{\partial \theta_i \partial \theta_j} \right\| = \left\| \frac{\delta_{ij}}{\theta_i} + \frac{1}{\theta_r} \right\|, \quad (5)$$

$$i, j = 1, \dots, r-1.$$

Si en esta matriz sustraemos la primera fila de todas las demás y luego utilizamos el desarrollo en elementos de la primera fila, obtenemos

$$|I(\theta)| = \left(1 + \sum_{j=1}^{r-1} \frac{\theta_j}{\theta_r}\right) \prod_{j=1}^{r-1} \theta_j^{-1} = \left(\prod_{j=1}^r \theta_j\right)^{-1}.$$

Así pues,  $0 < |I(\theta)| < \infty$  si  $\prod_{k=1}^r \theta_k > 0$ , o sea, si el punto  $\theta$  es el punto interior del simplex  $\Theta$ .

Por lo tanto, vemos que podemos utilizar los resultados de los §§ 13 y 14 en los criterios asintóticamente óptimos. De estos resultados se desprende que para verificar la hipótesis  $H_1 = \{\theta = p\}$  frente a  $H_2 = \{\theta \neq p\}$  existe un c.a.b. que coincide con el criterio de relación de verosimilitud

$$\frac{f_{\hat{\theta}}(X)}{f_p(X)} > c. \quad (6)$$

Este mismo criterio será asintóticamente minimax para verificar  $H_1$  frente a la hipótesis  $\{(\theta - p)I(\theta)(\theta - p)^T > b^2 n^{-1}\}$  (véase el teorema 15.3).

Para hallar de una forma más cómoda la región crítica (6), es necesario calcular el valor de  $f_{\hat{\theta}}(X)$ . Derivando (2) respecto a  $\theta_1, \dots, \theta_{r-1}$ , obtenemos

$$\frac{\partial L(X, \theta)}{\partial \theta_i} = \frac{v_i}{\theta_i} - \frac{v_r}{\theta_r}, \quad i = 1, \dots, r-1.$$

Igualando a cero estas derivadas, obtenemos que la e.v.m. equivale a

$$\hat{\theta}^* = n^{-1} \nu,$$

Así que  $\hat{\theta}_i^* = n^{-1} \nu_i$ .

Ahora bien, pasando a los logaritmos, el criterio (6) se puede escribir de la forma siguiente:

$$\psi^2(X) = \sum_{i=1}^r \nu_i \ln \frac{\nu_i}{np_i} > c_1. \quad (7)$$

De acuerdo con el teorema 13.1 (véase también el lema 13.1), la estadística  $2\psi^2(X)$  para la hipótesis  $H_1$  tiene una distribución límite  $\chi^2$  con  $r - 1$  grados de libertad. Por eso obtendremos el criterio de nivel asintótico  $1 - \varepsilon$  si ponemos  $c_1 = h_\varepsilon/2$ , donde  $h_\varepsilon$  es la cuantila de la distribución  $H_{r-1}$  del orden de  $1 - \varepsilon$ .

¿Qué representa en nuestras condiciones el criterio  $\pi'$  asintóticamente equivalente a (6), obtenido en el teorema 13.2 y que tiene la forma

$$n(\hat{\theta}^* - p)I(p)(\hat{\theta}^* - p)^T > h_\varepsilon? \quad (8)$$

Para  $t = (t_1, \dots, t_{r-1})$ ,  $s = \sum_{i=1}^{r-1} t_i$ , obtenemos

$$tI(p) = \left( \frac{t_1}{p_1} + \frac{s}{p_r}, \dots, \frac{t_{r-1}}{p_{r-1}} + \frac{s}{p_r} \right),$$

$$tI(p)t^T = \sum_{i=1}^{r-1} \frac{t_i^2}{p_i} + \frac{s^2}{p_r} = \sum_{i=1}^r \frac{t_i^2}{p_i}, \quad (9)$$

donde

$$t_r = -s, \quad \sum_{i=1}^r t_i = 0. \quad (10)$$

Suponiendo  $t = \hat{\theta}^* - p$  y notando que la condición (10) está cumplida, en calidad de (8) obtenemos

$$\sum_{i=1}^r \frac{(\nu_i - np_i)^2}{np_i} > h_\varepsilon. \quad (11)$$

Esto no es otra cosa sino el criterio  $\chi^2$ . De las afirmaciones citadas se deduce que  $\chi^2(X) \in H_{r-1}$ .

El criterio  $\pi''$  en el teorema 13.2 equivale asintóticamente a (7) y (11) y tendrá la forma

$$\sum_{i=1}^r \frac{(v_i - np_i)^2}{v_i} > h_\epsilon. \quad (12)$$

Teniendo también en cuenta el teorema 15.3 y la observación 15.1, podemos resumir lo dicho en la forma de la afirmación siguiente.

**Teorema 1.** *El criterio (7) para  $c_1 = h_\epsilon/2$ , así como el criterio  $\chi^2$  (11) y el criterio (12) tienen un nivel asintótico  $1 - \epsilon$  y son los c.a.b. para verificar, basándose en la muestra  $X \in B_\theta$ , la hipótesis  $\{\theta = p\}$  frente a  $\{\theta \neq p\}$ . Estos son, a su vez, los criterios asintóticamente minimax para verificar la hipótesis  $\{\theta = p\}$  frente a la alternativa  $\left\{ \sum_{i=1}^r (\theta_i - p_i)^2/p_i > b^2/n \right\}$  para cualquier  $b > 0$ .*

La equivalencia asintótica de los criterios (7), (11) y (12) también podría ser establecida directamente, utilizando el desarrollo en serie de  $\ln \frac{v_i}{np_i} = \ln \left( 1 + \frac{v_i - np_i}{np_i} \right)$  en (7).

Estos criterios son asintóticamente no paramétricos, ya que la distribución límite de las estadísticas que se utilizan en ellos es "absoluta", o sea, no está de ningún modo relacionada con la naturaleza de la distribución inicial.

**2. Aplicaciones del criterio  $\chi^2$ . Verificación de las hipótesis por los datos agrupados.** El criterio  $\chi^2$  está ampliamente difundido y su importancia sale fuera de los límites del problema examinado en el apartado anterior.

Volvamos a examinar el problema general concerniente a la hipótesis  $H_1 = \{X \in P_1\}$  frente a  $H_2 = \{X \in P, P \neq P_1\}$  que hemos estudiado en el § 12. Puesto que la teoría de los criterios óptimos se ha desarrollado, en cierta medida, sólo en el caso paramétrico, es natural que se trate de "parametrizar" de algún modo este problema<sup>\*)</sup>.

En el caso general, la manera más simple y natural de hacer esto es la *agrupación de los datos*, que consiste en lo siguiente. El campo de los valores posibles de las magnitudes sujetas a observación (o sea, el espacio  $\mathcal{X}$ ) se divide en  $r$  regiones disjuntas  $\Delta_1, \dots, \Delta_r$ , y en vez de la observación  $x_j$  sólo se indica el intervalo  $\Delta_k$  donde esta observación ha ido a parar.

<sup>\*)</sup> Se tiene en cuenta un parámetro de dimensión finita. Cualquier problema puede considerarse paramétrica si se admite un parámetro de dimensión infinita, ya que éste puede ser identificado con la distribución  $P, X \in P$ .

Con otras palabras, reducimos la precisión de las observaciones, y los  $x_j$  que cayeron en  $\Delta_k$  pueden ser sustituidos por un solo valor  $z_k \in \Delta_k$ . Claro está que eligiendo una división bastante completa, podemos aproximar la observación  $x_j$  mediante  $z_k$  tan exactamente como se quiera.

Así pues, la agrupación conduce a que la observación  $x_j$  es sustituida por el vector  $e_k$  si se ha producido el suceso  $A_k = \{x_j \in \Delta_k\}$  (los vectores  $e_k$  han sido definidos al principio del apartado anterior). Pero la nueva muestra obtenida como resultado de tal operación, evidentemente, no es otra cosa sino la muestra de  $B_\theta$ ,  $\theta_k = P(x_j \in \Delta_k)$ . Ya sabemos que en este caso, el vector  $\nu = (\nu_1, \dots, \nu_r)$  de frecuencias de caídas en los intervalos  $\Delta_1, \dots, \Delta_r$  será una estadística suficiente.

La reducción realizada de la muestra  $X$  al vector  $\nu$  es precisamente la llamada agrupación de los datos.

Por supuesto que tal agrupación está relacionada con cierto "empobrecimiento" de la muestra  $X$  y con una pérdida parcial de información.

La parametrización realizada también puede ser considerada desde otro punto de vista. Supongamos, para evidenciar, que  $\mathcal{R} = R$  y que todas las distribuciones que han de ser estudiadas, están concentradas en un intervalo finito y tienen densidad, o sea, satisfacen la condición  $(A_\mu)$ , donde  $\mu$  es la medida de Lebesgue. Con la partición  $\Delta_1, \dots, \Delta_r$  establecida, examinemos, a la par con la densidad  $f(x)$ , la densidad constante a trozos

$$f_\theta(x) = \frac{P(\Delta_i)}{\Delta_i} = \frac{1}{\Delta_i} \int_{\Delta_i} f(x) dx = \frac{\theta_i}{\Delta_i} \text{ para } x \in \Delta_i. \quad (13)$$

Donde  $\Delta_i$  también designa la longitud del intervalo  $\Delta_i$ . Esta es la familia paramétrica de las distribuciones  $P_\theta$ ,  $P_\theta(B) = \int_B f_\theta(x) dx$ .

La muestra  $Y$  de  $P_\theta$  podrá ser obtenida si para cada  $k$  recogemos todas las observaciones de  $X \in P$  que han ido a parar a  $\Delta_k$  y luego las "dispersamos" por  $\Delta_k$  uniformemente y al azar. En realidad esto es lo mismo que hemos hecho antes, ya que los datos que indican en qué punto del intervalo  $\Delta_k$  se encuentra la observación  $y_i$ , no contienen ninguna información acerca del parámetro  $\theta$ : la función de verosimilitud  $f_\theta(Y)$  no cambia después del "desplazamiento" de las observaciones dentro de los límites de sus intervalos. Por lo tanto, sólo es suficiente saber las cantidades  $\nu_1, \dots, \nu_r$  de observaciones que fueron a parar a  $\Delta_1, \dots, \Delta_r$ .

Está claro que si  $f(x)$  es una función suave,  $f_\theta(x)$  aproximará bien  $f(x)$  siempre que la partición de  $\{\Delta_1, \dots, \Delta_r\}$  sea bastante "menuda".

Las relaciones (13) significan otro método de parametrización, equivalente al primero. Tal equivalencia resulta de la coincidencia de las funciones de verosimilitud, con una exactitud de hasta un factor que no depende del

parámetro. Para la distribución (13), dicha equivalencia es igual a

$$f_{\theta}(Y) = \prod_{i=1}^r \theta_i^{y_i} \prod_{i=1}^r \Delta_i^{-y_i},$$

donde el primer factor es la función de verosimilitud de la muestra de  $\mathbf{B}_{\theta}$  (véase (4)).

Cabe señalar que la agrupación de las observaciones a menudo también surge por sí misma no para fines de parametrización, sino simplemente como un método cómodo y económico de anotación de la información que contiene la muestra. Si, por ejemplo,  $n = 10^4$  y la precisión de las mediciones de los valores observados en  $[0, 1]$  es comparable con 0,1, entonces claro está que prácticamente no merece la pena conocer todas la  $10^4$  observaciones y es suficiente indicar 10 frecuencias  $\nu_1, \dots, \nu_{10}$  de caída en los intervalos  $\Delta_i = ((i-1)/10, i/10)$ ,  $i = 1, \dots, 10$ , o sea, basta conocer tan sólo el histograma de la muestra.

Volvamos al problema de verificación de la hipótesis  $H_1 = \{X \in \mathbf{P}_1\}$  frente a  $H_2 = \{X \in \mathbf{P} \neq \mathbf{P}_1\}$ . Supondremos que la referida agrupación de observaciones es tal que la desviación (importante para nosotros) de la distribución  $\mathbf{P}$  de la muestra  $X$  respecto a  $\mathbf{P}_1$  se reflejará obligatoriamente en las distribuciones de los datos agrupados. Entonces, nuestro problema se puede considerar como un problema de verificación de la hipótesis  $\{\theta = p\}$ , donde  $p_i = \mathbf{P}_1(\Delta_i)$ , frente a  $\{\theta \neq p\}$ , para las familias paramétricas  $\mathbf{B}_{\theta}$  o (13). Como ya sabemos, *en este problema, el criterio  $\chi^2$  (al igual que los criterios (7) y (12)) será asintóticamente óptimo desde el punto de vista enunciado en el teorema 1.*

Además, el criterio  $\chi^2$  *no es asintóticamente paramétrico*, ya que, para la hipótesis  $H_1$ , la distribución límite de la estadística  $\chi^2(X)$  no depende de la distribución inicial de la muestra  $X$ .

En este caso cabe señalar que la verificación de la hipótesis  $\{\theta = p\}$  para las familias (13) o  $\mathbf{B}_{\theta}$  no es, a pesar de todo, equivalente a la verificación de la hipótesis  $\{X \in \mathbf{P}_1\}$ , aunque, con una partición abundante de  $\{\Delta_1, \dots, \Delta_r\}$ , ella pueda ser próxima a esta última. En efecto, para la muestra  $X$  se verifica la hipótesis  $X \in \mathbf{P}$ ,  $\mathbf{P}(\Delta_i) = p_i = \mathbf{P}_1(\Delta_i)$ . Esto contribuye a que el criterio  $\chi^2$  sea inconciliable respecto a las alternativas  $\mathbf{P} \neq \mathbf{P}_1$  para las cuales  $\theta_i = \mathbf{P}(\Delta_i) = \mathbf{P}_1(\Delta_i) = p_i$ . Por eso indicaremos una vez más, que el criterio  $\chi^2$  es un criterio que posee una serie de propiedades de optimización asintótica, pero que actúa exclusivamente contra las alternativas que modifican el vector  $\theta$ , o sea, contra las alternativas para las cuales  $\{\mathbf{P}(\Delta_i)\} \neq \{\mathbf{P}_1(\Delta_i)\} = \{p_i\}$ .

Hagamos algunas observaciones concernientes a las aplicaciones de los criterios  $\chi^2$ , (7) y (12). En este caso hablaremos fundamentalmente tan sólo del criterio  $\chi^2$ , ya que, por un lado, dichos criterios se asemejan unos a

otros y, por otro lado, el criterio  $\chi^2$  históricamente (en parte, debido a su evidencia) adquirió una aplicación mucho más amplia.

El nivel de importancia del criterio  $\chi^2(X) > h_\varepsilon$  es igual a  $1 - \varepsilon$  únicamente en el "límite". La experiencia muestra que para  $\varepsilon \geq 0,01$ , el verdadero nivel de importancia de este criterio se aproxima satisfactoriamente, mediante el valor de  $1 - \varepsilon$ , sólo cuando  $np_i \geq 8$ ,  $i = 1, \dots, r$ .

Si el número de grupos  $r$  es grande, digamos, cuando  $n > r > 30$ , se puede utilizar la aproximación normal tanto para la distribución  $\frac{1}{\sqrt{2r}} (\chi^2 - r)$ ,  $\chi^2 \in \mathbf{H}_r$  (véase el § 2.2), como también, en caso de la hipótesis  $H_1$ , para la distribución de la estadística  $\chi^2(X)$  normalizada por los momentos

$$M\chi^2(X) = r - 1,$$

$$D\chi^2(X) = 2(r - 1) + \frac{1}{n} \left( \sum_{i=1}^r p_i^{-1} - r^2 - 2r + 2 \right).$$

Con frecuencia también se utiliza la aproximación normal  $\Phi_{0,1}$  para distribuir la variable aleatoria (véase el § 2.2)  $\sqrt{2}\chi^2 - \sqrt{2r - 1}$ ,  $\chi^2 \in \mathbf{H}_r$ .

También debemos señalar que al aumentar el número de grupos mejora la aproximación de la densidad  $f(x)$  mediante una función escalonada construida según los valores de  $P_1(\Delta_i) = \int_{\Delta_i} f(x) dx$ . Esto significa que aumenta

el número de alternativas que no concuerdan con  $H_1$ , y que el criterio  $\chi^2$  se transforma cada vez más en criterio de verificación de las hipótesis acerca de la densidad. De acuerdo con esto, al aumentar el número de grupos, la potencia de los criterios  $\chi^2$  de nivel registrado disminuirá (compárese con las observaciones del párrafo anterior acerca del criterio de Morán. Esto se analiza más detalladamente en [12] y [21]).

Como defecto del criterio  $\chi^2$  debe considerarse el hecho de que en una serie de casos de partición  $\{\Delta_1, \dots, \Delta_r\}$  hay que establecer la estadística. Aquí es necesario tener cuidado, ya que en este caso se introduce un elemento de subjetivismo en el "empobrecimiento" de la muestra  $X$ . Además, a veces esta partición se elige en función de la muestra  $X$ , lo cual, hablando en general, no siempre es admisible, ya que, a su vez,  $\Delta_i$  se vuelven aleatorias (esto se examina más detalladamente en [49], p. 575).

**Ejemplo 1<sup>o</sup>** En la ciudad  $N$ , un individuo observó las indicaciones de 500 relojes expuestos en las vitrinas de distintas relojerías. Los resultados de las observaciones fueron divididos en 12 grupos (conforme a la posición del horario en la esfera). He aquí la tabla de las observaciones obtenidas:

<sup>o</sup> Este ejemplo se ha tomado de [25].



Intervalos en la esfera	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10	10-11	11-12
Número de observaciones	41	34	54	39	49	45	41	33	37	41	47	39

Se verifica una hipótesis simple:  $H_1 = \{ \text{la distribución de la posición del horario en la esfera según los grupos de horas es uniforme} \}$  frente a la alternativa adicional compuesta.

En este ejemplo,  $n = 500$ ,  $p_i = 1/12$ ,  $i = 1, \dots, 12$ ,  $np_i \approx 41,67$ . A base del teorema 1 podemos considerar que  $\chi^2(X) \in H_{11}$  aproximadamente. Sin embargo, en nuestro ejemplo, mediante el cálculo directo nos convencemos de que  $\chi^2(X) \approx 10$ , y el nivel realmente alcanzado del criterio  $\chi^2$  es aproximadamente igual a  $1 - H_{11}((10, \infty)) \approx 0,47$  (véase la tabla III). Esto significa que los resultados del experimento concuerdan con la hipótesis  $H_1$  desde el punto de vista del criterio  $\chi^2$  de cualquier nivel  $1 - \epsilon$  situado entre 0,47 y 1.

Ya hemos señalado que el criterio  $\chi^2$  está muy difundido. Además, la esfera de su aplicación consiste no sólo en verificar las hipótesis simples. Uno de tales ejemplos será examinado en el párrafo siguiente.

### § 17. Verificación de las hipótesis de pertenencia de la muestra a una familia paramétrica

Examinemos el problema de verificación de la hipótesis compuesta  $H_1 = \{ X \in \mathbf{P}_\alpha, \alpha \in A \}$  de que la distribución de la muestra pertenece a la familia paramétrica  $\{ \mathbf{P}_\alpha \}_{\alpha \in A}$  frente a la alternativa adicional  $H_2 = \{ X \in \mathbf{P}, \mathbf{P} \notin \{ \mathbf{P}_\alpha \}_{\alpha \in A} \}$ . Como ejemplo de tal género de hipótesis puede servir la afirmación de que  $X$  es la muestra de cualquier población normal (hipótesis  $H_1$ ), así como la afirmación adicional a la mencionada (hipótesis  $H_2$ ).

Como un segundo ejemplo puede servir la verificación de la hipótesis de que  $X \in \mathbf{B}_{\theta(\alpha)}$ , donde la dimensión de  $\alpha$  es menor que la de  $\theta$ . Este problema también puede ser interpretado como el problema de verificación de la hipótesis de pertenencia de  $X$  a una subfamilia paramétrica (véase el § 15). No obstante, la primera interpretación también será cierta, puesto que en el caso en que como resultado del experimento sólo acontezca un número finito de sucesos posibles (véase la definición de  $\mathbf{B}_\theta$  en el § 2.2), la familia  $\mathbf{B}_\theta$  comprenderá todas las distribuciones posibles de la muestra.

En el apartado siguiente examinaremos el problema de verificación de la hipótesis  $X \in \mathbf{B}_{\theta(\alpha)}$  y mostraremos que el problema general de pertenencia a la familia paramétrica puede ser reducido al primer problema mediante la agrupación de los datos.

**1. Verificación de la hipótesis  $X \in \mathbf{B}_{\theta(\alpha)}$ . Agrupación de los datos.** Examinemos primeramente el problema general enunciado al principio del párrafo y destinado al espacio arbitrario  $\mathcal{X}$ . Dividamos el espacio  $\mathcal{X}$  en regiones ("intervalos")  $\{\Delta_1, \dots, \Delta_r\}$  de tal modo que el número de "intervalos"  $r$  sea mayor que  $l + 1$ , donde  $l$  es la dimensión del parámetro  $\alpha$ . Realicemos la agrupación de las observaciones en estos intervalos. Si la hipótesis  $H_1 = \{X \in \mathbf{P}_\alpha\}$  es cierta, las probabilidades de que las observaciones caigan en los intervalos  $\Delta_i$  serán iguales a

$$p_i(\alpha) = \mathbf{P}_\alpha(\Delta_i).$$

Esto significa que en este caso el vector  $\theta = (\theta_1, \dots, \theta_r)$  de las probabilidades de que las observaciones caigan en  $\Delta_i$  debe situarse en la curva  $\theta = p(\alpha) = (p_1(\alpha), \dots, p_r(\alpha))$ .

Ahora bien, a base de la muestra  $Y \in \mathbf{B}_\theta$  obtenida en la agrupación, debemos verificar la hipótesis  $H_1$  acerca de la pertenencia de  $Y$  a la subfamilia paramétrica  $\mathbf{B}_{\theta(\alpha)}$  frente a la alternativa  $\{Y \in \mathbf{B}_\theta\}$ , donde  $\theta$  no se sitúa en la curva  $\theta = p(\alpha)$ ,  $\alpha \in A$ . Este problema fue examinado en el § 15, donde hemos hallado el criterio asintóticamente minimax para verificar  $H_1$  frente a la alternativa semejante

$$H_2 = \{Y \in \mathbf{B}_\theta, \inf_{\alpha} |\theta - p(\alpha_0 + \gamma n^{-1/2})|^{1/2} (p(\alpha_0 + \gamma n^{-1/2})) > bn^{-1/2}\} \quad (1)$$

(véase la aclaración 15.3 al teorema 15.4. El punto  $\alpha_0$  significa el valor "localizado" del parámetro, tal que las alternativas se disponen en el entorno del punto  $\theta_0 = p(\alpha_0)$ ). En nuestro caso, el criterio de la relación de verosimilitud (15.11) tiene la forma

$$\ln R_1(X) = \max_{\theta} \sum_{i=1}^r v_i \ln \theta_i - \max_{\alpha} \sum v_i \ln p_i(\alpha) > h_\varepsilon/2,$$

o bien, que es lo mismo,

$$\sum_{i=1}^r v_i \ln \frac{v_i}{np_i(\hat{\alpha}^*)} > h_\varepsilon/2,$$

donde  $\hat{\alpha}^*$  es la e.v.m. del parámetro  $\alpha$  según la muestra  $Y$  o según el vector  $\nu = (v_1, \dots, v_r)$ . Este criterio equivale asintóticamente (véase el teorema 15.4) al criterio

$$(p(\hat{\alpha}^*) - \nu n^{-1}) I(p(\hat{\alpha}^*)) (p(\hat{\alpha}^*) - \nu n^{-1})' > h_\varepsilon.$$

Como la forma de la matriz  $I(\theta)$  es conocida (véase (16.5), entonces, utilizando (16.9), del teorema 15.4 obtenemos el

**Corolario 1.** Si  $r - 1 > l$  y la función  $p(\alpha)$  satisface las condiciones del teorema 15.4, entonces el criterio de la relación de verosimilitud de nivel asintótico  $1 - \varepsilon$  para verificar, basándose en los datos agrupados, la hipótesis  $H_1 = \{X \in \mathbf{P}_\alpha, \mathbf{P}_\alpha \in \{\mathbf{P}_\alpha\}_{\alpha \in A}\}$  frente a la alternativa adicional  $H_2$ , es

asintóticamente minimax (para verificar la hipótesis  $H_1$  frente a (1)) y tiene la forma

$$\sum_{i=1}^r \nu_i \ln \frac{\nu_i}{np_i(\hat{\alpha}^*)} > h_\varepsilon/2, \quad (2)$$

donde  $h_\varepsilon$  es una cuantila del orden de  $1 - \varepsilon$  de la distribución  $\chi^2$  con  $r - l - 1$  grados de libertad. Este criterio equivale asintóticamente al criterio

$$\hat{\chi}^2(X) = \sum_{i=1}^r \frac{(\nu_i - np_i(\hat{\alpha}^*))^2}{np_i(\hat{\alpha}^*)} > h_\varepsilon. \quad (3)$$

Este último criterio también se llama *criterio  $\chi^2$* , pero en caso de que los parámetros "obstaculizadores" desconocidos se estimen con arreglo a la muestra. Como se deduce del corolario 1, la distribución de la estadística  $\hat{\chi}^2(X)$  converge, siempre que se trate de la hipótesis  $H_1$ , a la distribución  $\chi^2$  con  $r - l - 1$  grados de libertad (el número de grados de libertad  $r - 1$  en la distribución límite de la estadística  $\chi^2(X)$  ha disminuido en el número de parámetros escalares  $\alpha_1, \dots, \alpha_l$  que se estiman por la muestra).

**Ejemplo 1.** En el ejemplo 2.26.3 hemos descrito el mecanismo de herencia de los grupos de sangre 0 (cero), A, B y AB. Este mecanismo es controlado por los genes de tres tipos A, B y 0. Las probabilidades de aparición de estos genes en una población dada designémoslas por  $p, q, r = 1 - p - q$ . En el ejemplo 2.26.3 hemos hallado y en la tabla 1 del § 26 hemos escrito las probabilidades  $p_i(\alpha)$  de que una persona tenga el  $i$ -ésimo grupo de sangre.

Disponemos de la muestra  $X$  con las frecuencias  $\nu_i, i = 1, 2, 3, 4$  (véase la tabla 1) de aparición del  $i$ -ésimo grupo de sangre, obtenida como resultado del examen de  $n = 353$  personas. En el ejemplo 2.26.3 hemos hallado, para esta muestra, los valores aproximados de la e.v.m.  $\hat{\alpha}^* = (p^*, q^*) = (0,246, 0,173)$ . Esto nos proporcionó los valores de  $p_i(\hat{\alpha}^*)$  expuestos en la tabla 1.

Tabla 1. Distribución de las personas según los grupos de sangre

	0	A	B	AB	En total
$\nu_i$	121	120	79	33	353
$p_i^* = \nu_i n^{-1}$	0,343	0,340	0,224	0,093	1
$p_i(\hat{\alpha}^*)$	0,337	0,347	0,231	0,085	1

Hemos recibido la posibilidad de utilizar el corolario I para verificar la hipótesis acerca de que tiene lugar el mecanismo de herencia de la sangre, descrito anteriormente. Con ayuda de la tabla 1 hallamos que, en nuestro caso, la estadística  $\chi^2(X)$  (véase (3)) es igual a 0,44, aproximadamente. Esto concuerda bien con la hipótesis, ya que el valor crítico  $h_\varepsilon$ , correspondiente a la distribución  $\chi^2$  con un grado de libertad y al valor de  $\varepsilon = 0,2$ , es igual a  $h_{0,2} \approx 1,64$ .

**Ejemplo 2. Problema acerca de los indicios conjugados.** Supongamos que la muestra  $X$  es el resultado de la investigación de ciertos objetos, cada uno de los cuales se caracteriza por dos indicios  $A$  y  $B$ . El primero puede adoptar los valores  $A_1, \dots, A_s$ , y el segundo,  $B_1, \dots, B_t$ . Se pregunta, ¿serán esos indicios dependientes o no? Por ejemplo, podemos realizar cierto experimento  $G$ , obteniendo resultados  $B_1, \dots, B_t$  en condiciones  $A_1, \dots, A_s$  diferentes. El problema consiste en aclarar si dichos resultados dependen o no de las condiciones en que se realiza el experimento.

Este problema también puede considerarse como el problema de verificación de la independencia de dos variables aleatorias  $\xi$  y  $\eta$  según las observaciones agrupadas en el par  $(\xi, \eta)$ .

En nuestro ejemplo, los resultados de los experimentos son una matriz de valores  $\{v_{ij}\}$ , donde  $v_{ij}$  es el número de aparición de resultados con indicios  $A_i$  y  $B_j$  en la muestra  $X$  de volumen  $n$  (cada elemento de la muestra es un par de indicios del objeto que se examina).

$$\text{Designemos } p_{ij} = \mathbf{P}(A_i B_j), p_{i.} = \sum_{j=1}^t p_{ij}, p_{.j} = \sum_{i=1}^s p_{ij}.$$

Entonces, la hipótesis  $H_1$  de independencia de los indicios tendrá la forma  $H_1 = \{p_{ij} = p_{i.} p_{.j}\}$ . No es difícil notar que ésta es la hipótesis de pertenencia de la distribución de la muestra a una familia paramétrica, donde el papel de parámetro  $\alpha$  lo desempeña el vector  $\alpha = (p_{1.}, \dots, p_{s-1.}, p_{.1}, \dots, p_{.t-1})$  de  $s + t - 2$  dimensiones (los valores de  $p_{s.}$  y  $p_{.t}$  se deducen de las igualdades  $p_{s.} = 1 - \sum_{i=1}^{s-1} p_{i.}$ ,  $p_{.t} = 1 - \sum_{j=1}^{t-1} p_{.j}$ ).

La función de verosimilitud de la muestra  $X$ , siempre que se trate de la hipótesis  $H_1$ , es igual a

$$\prod_{i,j} p_{ij}^{v_{ij}} = \prod_i p_{i.}^{v_{i.}} \prod_j p_{.j}^{v_{.j}}, v_{i.} = \sum_{j=1}^t v_{ij}, v_{.j} = \sum_{i=1}^s v_{ij}.$$

De los resultados del § 16 (compárense con los del apartado (16.1)) se deduce que la e.v.m.  $\hat{\alpha}^*$  para tal función de verosimilitud tiene la forma

$$\hat{p}_{i.}^* = v_{i.}/n, \hat{p}_{.j}^* = v_{.j}/n.$$

Así pues, en nuestro caso, el criterio  $\chi^2$  adquiere la forma

$$\hat{\chi}^2(X) = \sum_{i,j} \frac{(v_{ij} - n\hat{p}_i \cdot \hat{p}_{\cdot j})^2}{n\hat{p}_i \cdot \hat{p}_{\cdot j}} = n \sum_{i,j} \frac{(v_{ij} - n^{-1}v_i \cdot v_{\cdot j})^2}{v_i \cdot v_{\cdot j}} > h_\varepsilon,$$

donde  $h_\varepsilon$  es una cuantila del orden de  $1 - \varepsilon$  de la distribución  $\chi^2$  con un número de grados de libertad de  $st - 1 - (s + t - 2) = (s - 1)(t - 1)$ .

Se pueden señalar muchos problemas aplicados, donde se utiliza el criterio de conjugación de los indicios que hemos construido. A título de ejemplo examinaremos uno de ellos: el problema de investigación sociológica de la relación entre los ingresos de las familias y la cantidad de niños en ellas (véase [25], p. 481).

**Ejemplo 2A.** Supongamos que el indicio  $A$  significa la cantidad de niños y adopta los valores 0, 1, 2, 3,  $\geq 4$ . El indicio  $B$  indica a qué intervalo (0 - 1), (1 - 2), (2 - 3), ( $\geq 3$ ) (por unidad se han adoptado 1000 coronas suecas) pertenece el salario. Según los resultados de  $n = 25\ 263$  investigaciones se han obtenido los datos expuestos en la tabla 2.

Tabla 2

$B$					
$A$	0-1	1-2	2-3	$\geq 3$	$v_i$
0	2161	3577	2184	1636	9558
1	2755	5081	2222	1052	11110
2	936	1753	640	306	3635
3	225	419	96	38	778
$\geq 4$	39	98	31	14	182
$v_{\cdot j}$	6116	10928	5173	3016	25263

En este ejemplo,  $\hat{\chi}^2(X) = 568,5$ , lo cual supera en mucho el valor crítico de  $h_\varepsilon$  para la distribución  $\chi^2$  de  $(5 - 1)(4 - 1) = 12$  grados de libertad, incluso con valores de  $\varepsilon$  bastante pequeños. Así que debemos reconocer la inconciliabilidad de la hipótesis  $H_1 = \{A \text{ y } B \text{ son independientes (inconjugados)}\}$ .

No obstante, debemos señalar que un análisis más minucioso ha demostrado la existencia de una dependencia muy débil entre los indicios  $A$  y  $B$ .

**2. Caso general.** El criterio  $\chi^2$  aplicado al problema de este párrafo posee los mismos defectos que los indicados con arreglo a los problemas del párrafo anterior.

El problema de verificación de la hipótesis  $\{X \in \mathbf{P}_\theta\}$  acerca de la pertenencia de  $X$  a la familia paramétrica  $\{\mathbf{P}_\theta\}_{\theta \in \Theta}$  también admite, por supues-

to, un enfoque más amplio, análogo al expuesto en el § 12. Elijamos cierta distancia  $d(\mathbf{P}, \mathbf{Q})$  en el espacio de distribuciones. Luego, hallemos el punto  $\mathbf{P}_\theta$  de  $\{\mathbf{P}_\theta\}$ , inmediato a  $\mathbf{P}_n^*$  desde el punto de vista de la distancia  $d$ . En calidad de  $\mathbf{P}_\theta$ , también se puede tomar  $\mathbf{P}_{\hat{\theta}^*}$ , donde  $\hat{\theta}^*$  es la e.v.m. (véase el § 2.5) o cualquier otra estimación razonable. Si la hipótesis  $H_1$  es cierta, entonces  $d(\mathbf{P}_{\hat{\theta}^*}, \mathbf{P}_n^*)$  no debe ser grande y, al contrario, si es cierta  $H_2$ , entonces  $d(\mathbf{P}_{\hat{\theta}^*}, \mathbf{P}_n^*)$  será considerable. Esta consideración nos ofrece la siguiente estructura del criterio: rechazaremos la hipótesis  $H_1$  si  $d(\mathbf{P}_{\hat{\theta}^*}, \mathbf{P}_n^*) > c$ , y la aceptaremos en el caso contrario.

El número  $c$  debe elegirse de modo que

$$\sup_{\theta \in \Theta} \mathbf{P}_\theta(d(\mathbf{P}_{\hat{\theta}^*}, \mathbf{P}_n^*) > c) \leq \varepsilon,$$

o de modo que esta relación se cumpla asintóticamente. El corolario 1 propone que en calidad de distancia  $d(\mathbf{P}_{\hat{\theta}^*}, \mathbf{P}_n^*)$  se adopten las estadísticas en (2) y (3). Entre otras, estas últimas también poseen la ventaja de que asintóticamente no son paramétricas: en el caso de la hipótesis  $H_1 = \{X \in \mathbf{P}_\theta\}$ , la distribución límite  $\hat{\chi}^2(X)$  no depende, por ejemplo, de  $\theta$ .

Examinemos la realización del enfoque general expuesto anteriormente en dos casos particulares importantes, cuando las familias paramétricas están formadas por parámetros de desplazamiento y de escala.

1) Supongamos que se verifica la hipótesis  $X \in \mathbf{P}_\theta$ ,  $\theta \in R$ , donde  $\mathbf{P}_\theta(A) = \mathbf{P}(A - \theta)$ ,  $A \subset R$ . Designemos por  $F(x)$  la función de distribución correspondiente a  $\mathbf{P}$  y pongamos  $F_\theta(x) = F(x - \theta)$ . En calidad de  $d$  adoptaremos la distancia que hemos utilizado en el criterio de Kolmogórov.

**Teorema 1.** *Supongamos que  $X \in \mathbf{P}_\theta$ ,  $F_\theta(x) = F(x - \theta)$  y que la función  $F(x)$  tiene una densidad uniformemente continua limitada igual a  $f(x) = F'(x)$ ,  $\int x^2 f(x) dx < \infty$ . Si designamos  $\int x f(x) dx = a$ ,  $\theta^* = \bar{x} - a$ , entonces, cualquier  $\theta$*

$$\lim_{n \rightarrow \infty} \mathbf{P}_\theta(\sup_x \sqrt{n} |F_n^*(x) - F_{\theta^*}(x)| > c) = \mathbf{P}(\sup_x |w^\circ F(x) + f(x) \int w^\circ(F(t)) dt| > c),$$

donde  $w^\circ$  es el puente browniano estándar.

En esta relación, el segundo miembro no depende de  $\theta$ . Calculándolo para un valor dado de  $F$  y escogiendo  $c = c_\varepsilon$  de modo que sea igual a  $\varepsilon$ , obtenemos el criterio

$$D_n = \sup_x \sqrt{n} |F_n^*(x) - F(x - \theta^*)| > c_\varepsilon$$

de nivel asintótico  $1 - \varepsilon$  para verificar la hipótesis  $H_1$  de pertenencia de la muestra  $X$  a la familia paramétrica  $\{\mathbf{P}_\theta\}$ , donde  $\theta$  es el parámetro de desplazamiento.

**Demostración del teorema 1.** Examinemos el proceso

$$W_n(x) = \sqrt{n}(F_n^*(x) - F_{\theta^*}(x)) = w_n(x) - \sqrt{n}(F_{\theta^*}(x) - F_\theta(x)),$$

donde  $w_n(x) = \sqrt{n}(F_n^*(x) - F_\theta(x))$ . Para  $t \rightarrow \theta$  tenemos

$$F_t(x) - F_\theta(x) = -(t - \theta)(f(x - \theta) + \varepsilon(t, \theta, x)),$$

$$|\varepsilon(t, \theta, x)| \leq \omega_{|t - \theta|},$$

donde  $\omega_\Delta$  es el módulo de continuidad de la función  $f$ , el cual no depende de  $x$ ,  $\omega_\Delta \rightarrow 0$  cuando  $\Delta \rightarrow 0$ . Como  $\theta^* \rightarrow \theta$ , entonces, poniendo  $t = \theta^*$  y adoptando, sin limitar la generali-

dad,  $\alpha = 0$ , obtenemos

$$\begin{aligned}\sqrt{n}(F_{\theta^*}(x) - F_{\theta}(x)) &= -f(x - \theta) \int td[\sqrt{n}(F_n^*(t) - F_{\theta}(t))] + \varepsilon(\theta^*, \theta, x) = \\ &= -f(x - \theta) \int tdw_n(t) + \varepsilon(\theta^*, \theta, x), \\ |\varepsilon(\theta^*, \theta, x)| &\leq \omega(\theta^* - \theta) = \sqrt{n}|\theta^* - \theta| \omega_{|\theta^* - \theta|} \rightarrow 0.\end{aligned}$$

Seguidamente, la funcional

$$H_N(w_n) = \sup_x |w_n(x) - f(x - \theta) \int_{-N}^N w_n(t) dt|,$$

para cualquier  $N > 0$ , es continua en la métrica uniforme. La sustitución de la variable  $x$  por  $F_{\theta}^{-1}(y) = \theta + F^{-1}(y)$ , cuya ejecución natural para la aplicación del teorema 1.6.3, no modifica este hecho. Por eso, en virtud del teorema mencionado,

$$H_N(w_n) = \sup_x \left| w^{\circ}(F(x) - \theta) + f(x - \theta) \int_{-N}^N w^{\circ}(F(t - \theta)) dt \right|.$$

Para demostrar la relación requerida,

$$D_n \rightarrow \sup_x |w^{\circ}(F(x - \theta)) + f(x - \theta) \int w^{\circ}(F(t - \theta)) dt|$$

(el desplazamiento en  $\theta$  del argumento en el segundo miembro no modifica el valor de este último) y, en virtud de las relaciones

$$\begin{aligned}|D_n - H_N(w_n)| &\leq \omega(\theta^* - \theta) + c \int_{|t| \geq N} w_n(t) dt \\ \omega(\theta^* - \theta) &\rightarrow 0,\end{aligned} \quad (4)$$

sólo queda convencernos que la integral en (4), juntamente con la integral  $\int_{|t| \geq N} w^{\circ}(F(t)) dt$

(pongamos, para abreviar,  $\theta = 0$ ), convergen, de modo probable, a cero cuando  $n \rightarrow \infty$ ,  $N \rightarrow \infty$ . Por lo visto, el método más simple de estimar ambas integrales consiste en demostrar la pequeñez de sus dispersiones utilizando la desigualdad de Chébishev. En vista de que los primeros dos momentos de las expresiones subintegrales en ambas integrales se comportan del mismo modo, podemos limitarnos a estimar tan sólo una de estas últimas. Examinemos, por ejemplo,

$$\int_{-\infty}^{-N} w^{\circ}(F(t)) dt.$$

En virtud de las relaciones  $Mw^{\circ}(s)w^{\circ}(u) = \min(s, u) + su \leq 2 \min(s, u)$  cuando  $s \leq 1$  y  $u \leq 1$ , tenemos

$$\begin{aligned}M \left( \int_{-\infty}^{-N} w^{\circ}(F(t)) dt \right)^2 &\leq 2 \int_{-\infty}^{-N} \int_{-\infty}^{-N} \min(F(t), F(s)) dt ds = \\ &= 4 \int_{-\infty}^{-N} (-t - N)F(t) dt \leq -8 \int_{-\infty}^{-N} tF(t) dt \rightarrow 0\end{aligned}$$

cuando  $N \rightarrow \infty$ , ya que  $\int t^2 dF(t) < \infty$ . Análogamente se examinan los demás intervalos.  $\triangleleft$

2) Supongamos que ahora se verifica la hipótesis  $X \in P_\theta$ ,  $\theta \in R$ ,  $\theta > 0$ , donde  $P_\theta(A) = P(A/\theta)$ ,  $A \subset R$ . Volvamos a designar por  $F$  la función de distribución correspondiente a  $P$ , y pongamos

$$\sigma^2 = M_1 x_1^2 = \int x^2 P(dx), \quad \theta^* = \frac{1}{\sigma} \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}.$$

**Teorema 2.** Supongamos que  $X \in P_\theta$ ,  $F_\theta(x) = F(x/\theta)$  y que existe una densidad continua limitada  $f(x) = F'(x)$  tal, que

$$\sup_x |xf(x)| < \infty, \quad \int x^4 f(x) dx < \infty. \quad (5)$$

Entonces, para cualquier  $\theta$ ,

$$\begin{aligned} \lim_{n \rightarrow \infty} P_\theta(\sup_x |\sqrt{n}(F_n^*(x) - F(x/\theta^*))| > c) = \\ = P\left(\sup_x |w^\circ(F(x)) + xf(x) \int tw^\circ(F(t)) dt| > c\right). \end{aligned}$$

La demostración de este teorema es absolutamente análoga a la del teorema 1. Tenemos

$$W_n(x) = \sqrt{n}(F_n^*(x) - F(x/\theta^*)) = w_n(x) - \sqrt{n}(F(x/\theta^*) - F(x/\theta)),$$

$$w_n(x) = \sqrt{n}(F_n^*(x) - F(x/\theta)).$$

Cuando  $t \rightarrow \theta$

$$F_t(x) - F_\theta(x) = \left(\frac{x}{t} - \frac{x}{\theta}\right) \cdot \left(f\left(\frac{x}{\theta}\right)\right) + \varepsilon(t, \theta, x),$$

donde, en virtud de la relación  $f(x) < c/|x|$  y de la continuidad uniforme de  $f$  en cualquier intervalo finito, se cumple  $\sup_x |\varepsilon(t, \theta, x)| \leq \omega_{|t-\theta|} \rightarrow 0$ . Poniendo  $t = \theta^* \rightarrow \theta$ , obtenemos

$$\begin{aligned} \sqrt{n} \left( F\left(\frac{x}{\theta^*}\right) - F\left(\frac{x}{\theta}\right) \right) = \\ = \sqrt{n} \left( \frac{x}{\theta^*} - \frac{x}{\theta} \right) f\left(\frac{x}{\theta}\right) - \frac{\sqrt{n}(\theta^* - \theta)}{\theta^*} \cdot \frac{x}{\theta} \cdot f\left(\frac{x}{\theta}\right) \varepsilon(\theta^*, \theta, x), \end{aligned}$$

donde  $\sup_x$  del segundo sumando converge a cero respecto a la  $P_\theta$ -probabilidad. Sólo nos

queda utilizar los razonamientos del teorema anterior (la pequeñez de las integrales  $\int_{|t|>N} tw^\circ(F(t)) dt$  y  $\int_{|t|>N} tw_n(t) dt$  es asegurada por la condición (5)) y señalar que la parte principal  $W_n(x)$  es igual a (adoptemos, sin limitar la generalidad,  $\sigma^2 = 1$ )

$$\begin{aligned} w_n(x) - \frac{\sqrt{n} x(\theta^{*2} - \theta^2)}{\theta\theta^*(\theta + \theta^*)} f(x/\theta) = w_n(x) - \frac{xf(x/\theta)}{\theta\theta^*(\theta + \theta^*)} \int t^2 dw_n(t) = \\ = w_n(x) - \frac{2xf(x/\theta)}{\theta\theta^*(\theta + \theta^*)} \int tw_n(t) dt, \end{aligned}$$



donde  $\theta^*(\theta + \theta^*) \xrightarrow{P_\theta} 2\theta^2$ . Por eso

$$\begin{aligned} \sup_x |W_n(x)| &= \sup_x \left| w^\circ \left( F \left( \frac{x}{\theta} \right) \right) + \theta^{-3} x f \left( \frac{x}{\theta} \right) \int t w^\circ \left( F \left( \frac{t}{\theta} \right) \right) dt \right| = \\ &= \sup_x \left| w^\circ \left( F \left( \frac{x}{\theta} \right) \right) + \frac{x}{\theta} f \left( \frac{x}{\theta} \right) \int t w^\circ(F(t)) dt \right|. \end{aligned}$$

Como la transformación de la contracción respecto a  $x$  debajo del signo  $\sup_x$  no modifica nada, el teorema 2 queda demostrado.

El lector también puede obtener resultados análogos para las estadísticas  $\int (F_n^*(x) - F_{\theta^*}(x))^2 dF_n^*(x)$ .

### § 18. Estabilidad de las decisiones estadísticas

Al construir distintos procedimientos estadísticos en los párrafos anteriores — en los problemas de estimación o de verificación de las hipótesis — cada vez partíamos de cierto conjunto de condiciones. Estas últimas se referían, en particular, a la independencia de las observaciones y a su igual distribución, así como a las suposiciones acerca del carácter de distribución  $\mathbf{P}$  de los elementos de una muestra. El incumplimiento de tales condiciones significaría que las afirmaciones respectivas (por ejemplo, acerca del carácter de distribución límite o acerca de la optimización de una u otra estadística) son, hablando en general, inciertas.

Por otro lado, en la práctica, las referidas condiciones son, como regla, el resultado de la aproximación y la idealización inevitable. Por consiguiente, dichas condiciones suelen no cumplirse de manera exacta y surgen dudas acerca de la validez de las recomendaciones basadas en uno u otro procedimiento estadístico elegido.

Por lo tanto, al igual que en cualquier otra rama de las matemáticas, referente a las aplicaciones, aquí es necesario (en la última etapa, antes de aplicar los métodos elaborados) aclarar cuán grandes deben ser las divergencias de las condiciones adoptadas, para que este hecho nos obligue a modificar las conclusiones enunciadas.

Desde el punto de vista matemático, tal procedimiento constituye un problema muy parecido al problema de la estabilidad. En los libros editados en inglés, para este tipo de problemas se ha adoptado el término “robustness”<sup>\*)</sup>. Por eso en los manuales editados en ruso, a la par con el término “estabilidad” también se utiliza la palabra “robusticidad”.

Las divergencias más difundidas de las condiciones antes mencionadas consisten en lo siguiente.

<sup>\*)</sup> Robustez o robusticidad.

1) En la serie de observaciones  $X$  está presente una pequeña porción de "desechos", o sea, de observaciones provocadas por graves errores de medición o de registro, o engendradas por cualquier otro mecanismo "obscurizador", distinto del sistema sujeto a investigación. Por lo general, la separación de dichas observaciones es imposible. En vez de esto se buscan procedimientos que sean poco sensibles a tal "ensuciamiento" de la muestra.

2) La distribución de  $x_i$  no equivale con exactitud a  $P$ , sino que tan sólo aproximadamente.

3) Los elementos de la muestra  $X$  no son independientes, sino que tan sólo débilmente dependientes.

La tarea consiste en construir las reglas resolutorias para los problemas principales de la estadística matemática, que sean semejantes, por su eficacia, a las reglas óptimas y que al mismo tiempo sean insensibles a las referidas divergencias de las condiciones adoptadas o, al menos, a aquellas de ellas que para nosotros no tienen importancia. Esta tarea, difícilísima y no siempre planteada con exactitud, aún no está estudiada del todo. Aquí los resultados tienen un carácter muy heterogéneo. Por eso sólo nos detendremos en algunos ejemplos típicos.

1. **Estimación de la media para las distribuciones simétricas.** Supongamos que  $X \in P$  y que la distribución en la recta  $P$  tiene una densidad de  $f(t - \alpha)$  respecto a la medida de Lebesgue,  $f(t) = f(-t)$ . Examinemos las dos estimaciones siguientes del parámetro  $\alpha = Mx_1$ . Una de ellas es

$$\alpha^* = \bar{x},$$

y la otra,  $\alpha^{**}$ , que se basa en las cuantilas muestrales:

$$\alpha^{**} = \frac{1}{r-1} \sum_{k=1}^{r-1} \zeta^* k p, \quad (1)$$

donde  $0 < p < 1$ ,  $r = 1/p$  es un número entero. Cuando  $p = 1/2$ , la estimación  $\alpha^{**}$  se transforma en la mediana muestral  $\zeta^* = \zeta_{1/2}^*$ .

Limitémonos por ahora al caso de  $p = 1/2$ . Cuando  $n \rightarrow \infty$  tenemos

$$(\alpha^* - \alpha)\sqrt{n} \in \Phi_{0, \sigma_1^2}, \quad \sigma_1^2 = \int t^2 f(t) dt. \quad (2)$$

Además, en el corolario 2.2.1 hemos establecido que para  $n \rightarrow \infty$

$$(\alpha^{**} - \alpha)\sqrt{n} \in \Phi_{0, \sigma_2^2}, \quad \sigma_2^2 = \frac{1}{4f^2(0)}. \quad (3)$$

Analizando la demostración de este corolario es fácil establecer que junto con  $\alpha^{**} = \zeta^* = x_{(k_0)}$ ,  $k_0 = [(n+1)/2]$ , esa misma distribución límite se observará en el término de la serie variacional  $x_{(k)}$  para cualquier valor registrado de la diferencia  $k - k_0$ .

De aquí se deduce que la estimación  $\alpha^{**} = \zeta^*$  es insensible (desde el punto de vista de sus propiedades asintóticas) al hecho de que a la muestra  $X$  se agregue cualquier número finito de "desechos". En efecto, si tenemos  $l$  "desechos" cualesquiera en la muestra  $X$ , entonces  $\alpha^{**}$  se situará entre los valores  $y_{(k_1)}$  e  $y_{(k_2)}$ , donde  $k_1 = k_0 - l$ ,  $k_2 = k_0 + l$  e  $y_{(k)}$ ,  $k = 1, \dots, n - l$  forman la serie variacional de la muestra  $Y \in \mathbf{P}$  de volumen  $n - l$ . Pero las propiedades asintóticas de  $y_{(k_1)}$  e  $y_{(k_2)}$  son iguales y coinciden con las de la mediana muestral.

Así pues, cualesquiera que sean los "desechos", la estimación  $\alpha^{**}$  será insensible a ellos. Eso no se puede decir de la estimación  $\alpha^* = \bar{x}$ , donde los referidos desechos pueden influir considerablemente (por ejemplo, si son comparables, en cuanto a su magnitud, con  $n$ ). Es fácil comprender que la propiedad de estabilidad de  $\alpha^{**}$  también se conservará para pequeñas muestras, si el número de desechos  $l$  es pequeño respecto a  $n$ . Asimismo esta propiedad se conservará en el caso en que en vez de  $\zeta^*$  se utilice una estadística (1) de una forma más general.

Por otro lado, para un caso particular importante, cuando  $\mathbf{P} = \Phi_{\alpha, \sigma_1^2}$ , hay una ley normal: el valor de  $\sigma_2^2 = \sigma_1^2 \pi / 2 (f(0) = (\sigma_1 \sqrt{2\pi})^{-1})$  excede la dispersión  $\sigma_1^2$  de la estimación eficiente  $\alpha^* = \bar{x}$  solamente  $\pi/2$  veces. Esta diferencia entre la eficacia de  $\alpha^{**}$  y  $\alpha^*$  puede disminuir aún más si las estimaciones (1) se examinan cuando  $r = 3, 4$ , etc. Entonces obtendremos una estimación  $\alpha^{**}$  casi tan eficiente como  $\bar{x}$  (al carecer de desechos) y al mismo tiempo estable respecto a los desechos. Además de (1) se puede tomar la media truncada

$$\alpha^{**} = \frac{1}{n - 2np} \sum_{k=np+1}^{n-np} x_{(k)}, \quad (4)$$

cuya dispersión también se aproxima con pequeños valores de  $p$ ) a la dispersión  $\sigma_1^2$  de la estimación  $\alpha^*$ .

Señalemos a continuación, que las propiedades de la estimación  $\alpha^* = \bar{x}$  dependen poco de las variaciones de  $\mathbf{P}$ , que conservan la varianza  $\sigma_1^2 = \int t^2 f(t) dt$  y, en particular, de las variaciones locales de  $f(t)$  en el punto  $t = 0$ . En este sentido dicha estimación es estable. Pero su *propiedad de optimización*, que tiene lugar para  $\mathbf{P} = \Phi_{\alpha, \sigma_1^2}$ , es inestable. En efecto, supongamos que para un valor pequeño de  $\varepsilon > 0$ ,

$$\mathbf{P} = (1 - \varepsilon)\Phi_{\alpha, 1} + \varepsilon U_{\alpha - \varepsilon, \alpha + \varepsilon}$$

Entonces  $f(0) = (1 - \varepsilon)/\sqrt{2\pi} + 1/2 > 1/2$  y, como muestran las relaciones (2) y (3), la estimación  $\alpha^{**} = \zeta^*$  será mucho mejor (el valor de  $\varepsilon$  debe ser pequeño, pero no menor de  $(1/\sqrt{n})$ ).

Por otro lado, la estimación  $\alpha^{**} = \zeta^*$  es estable (se tiene en cuenta

su distribución) respecto a las variaciones de  $\mathbf{P}$  que no afecten el valor de  $f(0)$ .

Las observaciones expuestas también pueden enunciarse de otro modo: con arreglo a los criterios estadísticos, por ejemplo, a los c.u.m.p. no desplazados  $|\bar{x} - \alpha_0| > c$  para verificar, a partir de la muestra  $X \in \Phi_{\alpha,1}$ , la hipótesis  $H_1 = \{\alpha = \alpha_0\}$  frente a  $H_2 = \{|\alpha - \alpha_0| > d > 0\}$ .

2. **Estadística de Student y  $S_0^2$ .** Examinemos ahora la cuestión concerniente a la estabilidad de los procedimientos estadísticos (estimación y verificación de las hipótesis) relacionados con las estadísticas

$$t = \frac{(\bar{x} - \alpha)\sqrt{n}}{S_0}, \quad S_0^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Como sabemos (véase los §§ 3.7 y 3.8), en estas estadísticas se basan los criterios óptimos para verificar, correspondientemente, las hipótesis respecto a la media  $\alpha$  y a la varianza  $\sigma^2$  de las poblaciones normales en el caso cuando se desconoce el segundo parámetro ( $\sigma^2$  o  $\alpha$ ) de la distribución  $\Phi_{\alpha, \sigma^2}$ .

Las estadísticas  $t$  y  $S_0^2$  se comportan de manera diferente con arreglo a las alteraciones de las condiciones  $X \in \Phi_{\alpha, \sigma^2}$ . Supongamos que  $n$  es grande y  $X \in \mathbf{P}$ , donde  $\mathbf{P}$  es cualquier distribución, con  $\alpha$  media y con varianza finita. Entonces, la distribución  $t$ , al igual que en el caso  $X \in \Phi_{\alpha, \sigma^2}$ , se aproximará a la distribución normal  $\Phi_{0,1}$ . Esto se deduce de los teoremas de continuidad (§ 1.5) y del hecho de que

$$(\bar{x} - \alpha)\sqrt{n}/\sqrt{Dx_1} \in \Phi_{0,1}, \quad S_0^2 \xrightarrow{\mathbf{P}} D_1.$$

Lo dicho significa que la dimensión del criterio de Student se diferenciará poco, para grandes valores de  $n$ , de la dimensión dada, si incluso la distribución  $\mathbf{P}$  de la muestra  $X$  se diferencia considerablemente de la distribución normal.

Esto no se puede decir con arreglo a los criterios construidos a base de la estadística  $S_0^2$ . Esta circunstancia se debe al hecho de que la distribución límite  $S_0^2$  depende del valor  $Mx_1^4$ . En efecto, de las consideraciones del capítulo 1 resulta

$$(S_0^2 - \sigma^2)\sqrt{n} \in \Phi_{0,d^2}, \quad d^2 = M(x_1^2 - \sigma^2)^2 = Dx_1^2.$$

Por consiguiente, la dimensión del criterio construido a base de la estadística  $S_0^2$  para una población normal puede diferenciarse considerablemente de la dimensión dada, si  $X \in \mathbf{P}$  y  $\mathbf{P}$  se diferencian de  $\Phi_{\alpha, \sigma^2}$  (pero si coinciden los cuartos momentos de  $\mathbf{P}$  y  $\Phi_{\alpha, \sigma^2}$ , entonces no habrá diferencia).

Ambas estadísticas  $t$  y  $S_0^2$  son sensibles al rechazamiento de la suposición acerca de la independencia de las observaciones en la muestra  $X$ . Si,

por ejemplo, todas las observaciones en la muestra están relacionadas unas con otras, y el coeficiente de correlación es igual a  $\rho$ , entonces, adoptando  $\alpha = 0$  sin limitar la generalidad, obtenemos

$$\begin{aligned} MS_0^2 &= \frac{1}{n-1} \mathbf{M} \left[ \sum_{i=1}^n x_i^2 - n(\bar{x})^2 \right] = \\ &= \frac{1}{n-1} \left[ n\sigma^2 - \frac{1}{n} \mathbf{M} \left( \sum_{i=1}^n x_i \right)^2 \right] = \\ &= \frac{1}{n-1} [n\sigma^2 - \sigma^2(1-\rho) - n\sigma^2\rho] = \sigma^2(1-\rho). \end{aligned}$$

Ahora bien, aquí se altera incluso la propiedad de no desplazamiento de  $S_0^2$ , aunque para pequeños valores de  $\rho$  la divergencia será pequeña. El establecimiento de las distribuciones de  $t$  y  $S_0^2$  suele chocar con grandes dificultades al aparecer cierta dependencia.

**3. Criterio de relación de verosimilitud.** Este criterio suele ser muy sensible a la existencia de desechos e incluso de pequeñas divergencias en las suposiciones acerca de la distribución de  $X$ . Supongamos, por ejemplo, que se verifican dos hipótesis simples  $H_1 = \{X \in \Phi_{0,1}\}$  y  $H_2 = \{X \in U_{-1,1}\}$ . Está claro que, al utilizar el criterio más potente de Neyman — Pearson, la aparición incluso de una sola observación  $x$  fuera del segmento  $[-1, 1]$ , siempre que las demás observaciones correspondan idealmente a la distribución  $U_{-1,1}$ , nos obligará (¡con una probabilidad nula de equivocación!) a reconocer la hipótesis  $H_1$ . Esto significa que la presencia de un solo desecho o la aparición incluso de pequeñas divergencias de la distribución  $U_{-1,1}$  pueden obligarnos a tomar una decisión falsa.

En este sentido, el criterio de Kolmogórov es, por ejemplo, mucho más estable (aunque también menos potente respecto a  $H_2$ ). En general, los criterios no paramétricos, como era de esperar, son mucho más estables que los criterios "individuales" dotados de propiedades de optimización en uno u otro problema concreto.

En cuanto al referido problema de verificación de la normalidad ( $H_1$ ) frente a la uniformidad ( $H_2$ ) de la muestra  $X$ , el establecimiento de criterios potentes y al mismo tiempo estables respecto a los desechos, se puede realizar utilizando, como antes, la relación de verosimilitud, pero para muestras "truncadas" (compárese con (4)). También se puede ir por la vía de elección de otro criterio cualquiera. En este sentido, la existencia de una reserva bastante grande de criterios y estimaciones diferentes es muy útil. A esto a menudo se acude no sólo por razones de estabilidad, sino también por cuestiones de comodidad de los cálculos.