Test de hipótesis (Contraste de hipótesis)

Consideremos el siguiente contexto:

Sea X una variable aleatoria (posiblemente multidimensional: a valores en $\mathbb{R}^m$) que sigue una distribución determinada por una densidad $f(x,\theta)$, con $\theta \in \Theta$ donde Θ, para fijar ideas, podemos considerar que es un subconjunto de $\mathbb{R}^k$, aunque podríamos, con más generalidad, suponer que es un "espacio" de otra naturaleza, e incluso infinito dimensional. Supondremos además que parámetros distintos tienen asociados leyes probabilísticas distintas.

NOTA: el modelo estadístico puede también determinarse en términos de la función de distribución de X, o de la función característica.

Disponemos además de una muestra aleatoria simple de tamaño n de X, con sus correspondientes variables aleatorias muestrales $X_1, \dots, X_n$ i.i.d X. Llamamos Ω al espacio muestral de éstas variables aleatorias (Ω podemos identificarlo con $\mathbb{R}^{m\times n}$ en muchos casos)

Un problema de test de hipótesis (problema de contraste de hipótesis) consiste en establecer dos hipótesis sobre el verdadero valor del parámetro, que pueden formularse especificando un subconjunto propio del espacio de parámetros, Θ_0, a saber: (subconjunto propio: $\Theta_0 \neq \Theta$)

$H_0: \theta \in \Theta_0$ ← hipótesis nula

$H_1: \theta \in \Theta \setminus \Theta_0$ ← hipótesis alternativa ($\Theta \setminus \Theta_0 = \Theta \cap \Theta_0^c$)

hipótesis que llamaremos hipótesis nula y alternativa, respectivamente.

Una hipótesis diremos que es simple si el cardinal del subconjunto del espacio de parámetros correspondiente es igual a uno contiene un único elemento. Así H_0 será simple si $\text{card}\, \Theta_0 = 1$ y H_1 será simple si $\text{card}(\Theta \setminus \Theta_0) = 1$.

Una hipótesis será compuesta si el cardinal del subconjunto del espacio de parámetros correspondiente es mayor que 1.

Un test (o contraste) de hipótesis es una regla de decisión que asigna a cada elemento del espacio muestral una de las hipótesis.

Distinguiremos los tests puros y los tests aleatorizados.

Los tests o contrastes de hipótesis puros pueden caracterizarse mediante una aplicación

$$\delta : \Omega \rightarrow \{\Theta_0, \Theta \setminus \Theta_0\}$$

con la condición de que $\delta^{-1}(\Theta_0)$ y por consiguiente $\delta^{-1}(\Theta \setminus \Theta_0)$ sean sucesos (elementos del álgebra de sucesos de Ω).

Al suceso (subconjunto del espacio muestral)

$$W = \delta^{-1}(\Theta \setminus \Theta_0)$$

le denominaremos región crítica (del test puro).

Corresponde al subconjunto del espacio muestral formado por todos los resultados $x \in \Omega$ que llevaran al rechazo de la hipótesis nula. La regla de decisión podrá formularse como:

$$\left.\begin{array}{l} x \in W \longrightarrow \text{rechazaremos } H_0 \text{ (aceptaremos } H_1) \\ x \notin W \longrightarrow \text{aceptaremos } H_0 \text{ (rechazaremos } H_1) \end{array}\right\}$$

Se define la función de potencia de un test puro como:

$$\alpha : \Theta \longrightarrow \mathbb{R}$$

$$\theta \longrightarrow \alpha(\theta) = P_\theta(W)$$

es decir la función que asigna a cada θ el valor de la probabilidad del suceso W calculado con P_θ (tomando θ como el verdadero parámetro).

El error de primera especie (de tipo I) es igual a la probabilidad de equivocarnos al rechazar la hipótesis nula siendo ésta cierta.

Dicho error puede expresarse como:

$$\boxed{\alpha(\theta) = P_\theta(W) \qquad \theta \in \Theta_0}$$ (ya que rechazaremos H_0 cuando $x \in W$)

El error de segunda especie (de tipo II) es igual a la probabilidad de equivocarnos cuando aceptamos H_0 siendo esta falsa: siendo cierta H_1; podemos expresarlo como:

$$\boxed{\beta(\theta) = P_\theta(W^c) = 1 - P_\theta(W) \qquad \theta \in \Theta \setminus \Theta_0}$$ (ya que aceptamos H_0 cuando $x \in W^c$)

El nivel de significación, α, de un test se definirá como:

$$\boxed{\alpha \equiv \sup_{\theta \in \Theta_0} \alpha(\theta) = \sup_{\theta \in \Theta_0} P_\theta(W)}$$

el error "más grande" de la regla de decisión si es cierta Ho

es una medida del "mayor" error de tipo I posible (probabilidad "máxima" de error) cuando rechazamos Ho siendo ésta cierta

La probabilidad de error, sin más, de un test no puede calcularse salvo que introduzcamos técnicas Bayesianas, donde se hará uso de información "a priori" sobre θ.
En efecto, en el caso de disponer de una distribución "a priori" $g(\theta)$, la probabilidad de error podría expresarse como:

$$P(\text{Error}) = \int_{\Theta_0} g(\theta)\, \underbrace{\alpha(\theta)}_{\text{error tipo I}}\, d\theta + \int_{\Theta \setminus \Theta_0} g(\theta)\, \underbrace{(1-\alpha(\theta))}_{\text{error tipo II}}\, d\theta$$

pero, salvo en situaciones particulares, no resulta claro que pueda considerarse una "a priori" razonable.

NOTA: Observar que los errores de tipo I y tipo II son probabilidades condicionadas de error.

La estrategia habitual será: fijar el error de tipo I, al que mantendremos bajo y, tratando de elegir el test de forma que sea lo más pequeño posible el error de tipo II, como veremos más adelante.

Los tests o contrastes de hipótesis aleatorizados los definimos en términos de una función (medible)

$$\begin{aligned} \phi: \Omega &\longrightarrow [0,1] \\ x &\longrightarrow \phi(x) \end{aligned}$$

donde el valor $\phi(x)$ es igual a la probabilidad de rechazar Ho dada la muestra x: es decir dado "x" nos quedamos con H_1 con probabilidad $\phi(x)$ y Ho con probabilidad $1-\phi(x)$ ("sorteamos" la decisión final).

Un test puro también puede verse como un caso particular de test aleatorizado:

Un test puro con región crítica W es equivalente a un test aleatorizado con:

$$\phi(x) = \mathbb{1}_W(x)$$

donde $\mathbb{1}_W(x)$ es la función indicadora de W: $\mathbb{1}_W(x) = \begin{cases} 1 & x \in W \\ 0 & x \notin W \end{cases}$

En el caso de test aleatorizados también puede definirse la función de potencia, así como los errores de tipo I y II y el nivel de significación, como sigue:

* La función de potencia correspondiente a ϕ:

$$\alpha_\phi(\theta) = E_\theta(\phi(X))$$ (donde X representa la muestra conjunta)

* El error de tipo I será:

$$\alpha_\phi(\theta) = E_\theta(\phi(X)) \qquad \theta \in \Theta_0$$

* mientras que el error de tipo II será:

$$\beta_\phi(\theta) = E_\theta(1-\phi(X)) = 1-\alpha_\phi(\theta) \qquad \theta \in \Theta \setminus \Theta_0$$

* y el nivel de significación será:

$$\alpha \equiv \sup_{\theta \in \Theta_0} \alpha_\phi(\theta)$$

---

* Principio de Neyman:

Sean dos tests de hipótesis (puros o aleatorizados), Φ_1 y Φ_2 con nivel de significación menor o igual que α, diremos que Φ_1 es preferible que Φ_2, y escribiremos $\Phi_1 \succcurlyeq \Phi_2$ si y sólo si $\beta_{\Phi_1}(\theta) \leq \beta_{\Phi_2}(\theta) \quad \forall \theta \in \Theta \setminus \Theta_0$

es decir, para errores de tipo I menores o iguales que α, preferiremos el test cuyo error de tipo II sea menor, o equivalentemente mayor su potencia: $\alpha_{\Phi_1}(\theta) \geq \alpha_{\Phi_2}(\theta)$ para $\theta \in \Theta \setminus \Theta_0$.

Consideremos a continuación el caso de contrastes de hipótesis nula simple frente alternativa simple. Sean f_1 y f_0 la función de densidad conjunta, de la muestra que disponemos, bajo las hipótesis alternativa y nula respectivamente.

Sea $\overline{\mathbb{R}^+}$ al conjunto formado por los números reales positivos, más el cero y además $+\infty$. Adoptaremos la convención:

$$0 \times \infty = 0$$

* Un test de Neyman es un test (posiblemente aleatorizado) que satisface, para un cierto $k \in \overline{\mathbb{R}^+}$:

$$\phi(x) = 1 \qquad \text{para } x \in \Omega \text{ tales que } f_1(x) > k f_0(x)$$

y

$$\phi(x) = 0 \qquad \text{para } x \in \Omega \text{ tales que } f_1(x) < k f_0(x)$$

Nota: observemos que $\phi(x)$ puede tomar cualquier valor para los $x \in \Omega$ tales que $f_1(x) = k f_0(x)$.

Teorema (Neyman-Pearson) $H_0: \theta = \theta_0$ $H_1: \theta = \theta_1$ (simple versus simple)

1) $\forall \alpha \in [0,1]$ (nivel de significación) existe un test de Neyman ϕ tal que $E_{\theta_0}(\phi(X)) = \alpha$ y además podemos elegirlo de forma que ϕ sea constante en $\{x \in \Omega \ / \ f_1(x) = k f_0(x)\}$ para un cierto $k \in \overline{\mathbb{R}^+}$

2) Una condición necesaria y suficiente para que un test sea de Neyman es que:

$$\forall \tilde{\phi} \ / \ E_{\theta_0}(\tilde{\phi}) \leq E_{\theta_0}(\phi) \implies E_{\theta_1}(\phi) \geq E_{\theta_1}(\tilde{\phi})$$

NOTA: Observar que dicho teorema garantiza, para cualquier nivel de significación, la existencia de un test de Neyman para dicho nivel de significación. El apartado 2, nos indica que cualquier otro test con nivel de significación menor o igual a α tendrá una potencia inferior al anteriormente citado test de Neyman, y por tanto un error de tipo II superior.

NOTA: En el caso simple contra simple, entendemos por potencia, al valor de la función de potencia bajo H_1.

Por tanto, cuando consideramos problemas de contraste de hipótesis nula simple frente alternativa simple el teorema de Neyman-Pearson garantiza que podremos encontrar siempre un test de potencia máxima, fijado un determinado nivel de significación.

Ejemplo 1

$X \sim N(\mu, 1)$ $\mu \in \{\mu_0, \mu_1\}$ donde μ_0 y μ_1 son conocidas, con $\mu_0 < \mu_1$. Sea $X_1, \ldots, X_n$ i.i.d X. Consideremos el problema de contraste de hipótesis:

$H_0: \mu = \mu_0$ (simple)

$H_1: \mu \neq \mu_1$ (simple)

Determinar un test de potencia máxima (mínimo error de tipo II) para el nivel de significación $\alpha = 0.05$

Vamos a determinar en primer lugar al subconjunto, del espacio muestral siguiente:

$$W = \{ x \in \Omega \ / \ f_1(x) > k f_0(x) \}$$

aquí $x = (x_1 - - x_n)$, $\Omega = \mathbb{R}^n$

$$f_1(x) = \prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_i - \mu_1)^2} \right\}$$

$$= \left(\frac{1}{2\pi} \right)^{\frac{n}{2}} e^{-\frac{1}{2} \sum_{i=1}^{n} (x_i - \mu_1)^2}$$

$$f_0(x) = \prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_i - \mu_0)^2} \right\}$$

$$= \left(\frac{1}{2\pi} \right)^{\frac{n}{2}} e^{-\frac{1}{2} \sum_{i=1}^{n} (x_i - \mu_0)^2}$$

por tanto la desigualdad $f_1(x) > k f_0(x)$ se convierte en:

$$\left(\frac{1}{2\pi} \right)^{\frac{n}{2}} e^{-\frac{1}{2} \sum_{i=1}^{n} (x_i - \mu_1)^2} > k \left(\frac{1}{2\pi} \right)^{\frac{n}{2}} e^{-\frac{1}{2} \sum_{i=1}^{n} (x_i - \mu_0)^2}$$

Equivalente a:

$$e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu_1)^2+\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu_0)^2} > K$$

desarrollando los cuadrados, e introduciendo:

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$$

resulta:

$$e^{-\frac{1}{2}n\mu_1^2+\frac{1}{2}n\mu_0^2+(\mu_1-\mu_0)\sum_{i=1}^{n}x_i} > K$$

tomando logaritmos (ya que $K>0$, puesto que de lo contrario $W=\Omega$ y el nivel de significación del test sería $\alpha=1$) resulta:

$$-\frac{1}{2}n\mu_1^2+\frac{1}{2}n\mu_0^2+(\mu_1-\mu_0)\sum_{i=1}^{n}x_i > \ln K$$

y al ser $\mu_1>\mu_0$, $\mu_1-\mu_0>0$ y dividiendo por $\mu_1-\mu_0$ obtenemos:

(*) $$\sum_{i=1}^{n} x_i > \frac{\ln K}{\mu_1-\mu_0}+\left(\frac{1}{2}n\mu_1^2-\frac{1}{2}n\mu_0^2\right)\Big/(\mu_1-\mu_0)$$

igual a:

$$\sum_{i=1}^{n} x_i > \underbrace{\frac{\ln K}{\mu_1-\mu_0}+\frac{1}{2}n(\mu_1+\mu_0)}_{c}$$

c una cierta constante, en el sentido que no depende de $x_1 \ldots x_n$

Por tanto

$$W=\left\{ x\in\mathbb{R}^n \Big/ \sum_{i=1}^{n} x_i > c\right\}$$

Obsérvese que podemos elegir c de forma que

$$P(W\mid H_0)=\alpha$$

ya que

$$P(W\mid H_0) = P\left(\sum_{i=1}^{n}X_i > c \mid H_0\right)$$ y $\sum_{i=1}^{n}X_i$ bajo H_0

sigue una distribución normal:

$$\text{bajo } H_0, \quad \sum_{i=1}^{n} X_i \sim N(n\mu_0, \text{var}=n)$$

Por tanto,

$$Z = \frac{\sum_{i=1}^{n} X_i - n\mu_0}{\sqrt{n}} \quad , \text{ bajo } H_0, \text{ sigue una } N(0,1)$$

y resultará:

$$\alpha = P\left(\sum_{i=1}^{n} X_i > C \mid H_0\right) = P\left(Z > \frac{C-n\mu_0}{\sqrt{n}}\right) = 1-\Phi\left(\frac{C-n\mu_0}{\sqrt{n}}\right)$$

donde

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt$$

(la función de distribución de una normal estandarizada)

$$\Phi\left(\frac{C-n\mu_0}{\sqrt{n}}\right) = 1-\alpha \quad \Longrightarrow \quad \frac{C-n\mu_0}{\sqrt{n}} = \Phi^{-1}(1-\alpha)$$

$$\boxed{C = n\mu_0 + \sqrt{n}\,\Phi^{-1}(1-\alpha)}$$

Existe pues, para cada α (nivel de significación) un test de Neyman puro, aquel cuya región crítica es W, de este nivel de significación; si hacemos $\boxed{\phi(x) = \mathbb{1}_W(x)}$ resulta:

$$E_{\mu_0}(\phi(x)) = P_{\mu_0}(W) = P(W \mid H_0) = \alpha$$

y por el Teorema de Neyman-Pearson tendrá el menor error de tipo II posible, o equivalentemente la máxima potencia. Dicha potencia será:

$$\boxed{\alpha(\mu_1) = E_{\mu_1}(\phi(x)) = P(W \mid H_1) = P\left(\sum_{i=1}^{n} X_i > n\mu_0 + \sqrt{n}\,\Phi^{-1}(1-\alpha) \,\Big|\, H_1\right)}$$

como ahora $\sum_{i=1}^{n} X_i \sim N(n\mu_1, \text{var}=n)$, resultará que

$$Z = \frac{\sum_{i=1}^{n} X_i - n\mu_1}{\sqrt{n}} \sim N(0,1) \quad \text{y tendremos}$$

$$\alpha(\theta_1) = P\left(Z > \frac{n(\mu_0-\mu_1) + \sqrt{n}\,\Phi^{-1}(1-\alpha)}{\sqrt{n}} \right)$$

$$= P\left(Z > \sqrt{n}(\mu_0-\mu_1) + \Phi^{-1}(1-\alpha) \right)$$ y finalmente:

$$\boxed{\alpha(\theta_1) = 1 - \Phi\left(\sqrt{n}(\mu_0-\mu_1) + \Phi^{-1}(1-\alpha) \right)}$$

potencia del test (puro) obtenido.

Nota: Obsérvese que si $\mu_0 \to \mu_1$ entonces $\alpha(\theta_1) \to \alpha$.
Obsérvese también que si $n \to \infty$ entonces $\alpha(\theta_1) \to 1$.

Nota: Si hubiéramos tratado de resolver el problema de test de hipótesis $H_0: \mu = \mu_0$ $H_1: \mu = \mu_1$ pero con $\mu_1 < \mu_0$ (y ambos conocidos) tendríamos que proceder de la misma forma, pero en (*) de la página 7 habríamos obtenido

$$\sum_{i=1}^{n} x_i < \frac{K}{\mu_1-\mu_0} \left(\frac{1}{2} n \mu_1^2 - \frac{1}{2} n \mu_0^2 \right)$$

puesto que ahora $\mu_1-\mu_0$ sería negativo. La región crítica sería pues:

$$W = \left\{ x \in \mathbb{R}^n \;/\; \sum_{i=1}^{n} x_i < C \right\}$$

y el valor de C que garantizaría un nivel de significación α sería:

$$C = n\mu_0 - \sqrt{n}\,\Phi^{-1}(1-\alpha)$$

Finalmente, la potencia sería:

$$\alpha(\theta_1) = 1 - \Phi\left(\sqrt{n}(\mu_1-\mu_0) + \Phi^{-1}(1-\alpha) \right)$$

con comportamiento similar al caso anterior.

Nota: si llamamos $\partial W = \{ x \in \mathbb{R}^n \mid f_1(x) = K f_0(x) \}$ (para un cierto K) entonces: $P(\partial W \mid H_0) = 0$.

Ejemplo 2

Consideremos ahora $X \sim N(\mu, 1)$ $\mu \in [\mu_0, \infty)$ (μ_0 conocido). Sea $X_1, \dots, X_n$ i.i.d X. Consideremos el problema de contraste de hipótesis:

$H_0: \mu = \mu_0$ (hipótesis simple)

$H_1: \mu > \mu_0$ (hipótesis compuesta)

Determinar, si existe, un test UMP, para un nivel de significación $\alpha = 0.05$.

Procederemos de entrada como en el Ejemplo 1, tratando de encontrar un subconjunto W del espacio muestral que permita definir un test de Neyman de nivel de significación α:

$$f_1(x) > k f_0(x)$$

equivale, como en el Ejemplo 1, a:

$$\left(\frac{1}{2\pi}\right)^{\frac{n}{2}} e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2} > K\left(\frac{1}{2\pi}\right)^{\frac{n}{2}} e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu_0)^2}$$

donde ahora $\mu \in (\mu_0, \infty)$. Procediendo como en el Ejemplo 1 obtenemos:

$$(\mu - \mu_0)\sum_{i=1}^{n} x_i > \ln K + \frac{1}{2} n (\mu + \mu_0)(\mu - \mu_0)$$

ahora bien, aunque $\mu - \mu_0$ no es conocido, podemos asegurar que es positivo, por tanto, dividiendo por $\mu - \mu_0$ obtenemos

$$\sum_{i=1}^{n} x_i > \frac{\ln K}{\mu - \mu_0} + \frac{1}{2} n (\mu + \mu_0)$$

la parte derecha es una cierta constante, en el sentido que no depende de $x_1 \dots x_n$, que habrá que elegir de forma que el nivel de significación sea α, (al elegir K lo haremos de forma que la expresión $\frac{\ln K}{\mu - \mu_0} + \frac{1}{2} n (\mu + \mu_0)$ no dependa de μ)

Si llamamos $C = \dfrac{\ln K}{\mu-\mu_0} + \frac{1}{2} n (\mu+\mu_0)$

Deberemos elegir C de forma que:

$$\alpha = P(W|H_0) = P\left(\sum_{i=1}^{n} X_i > C \;|H_0\right)$$

y ello implica que debemos elegir $C = n\mu_0 + \sqrt{n}\;\Phi^{-1}(1-\alpha)$ como en el Ejemplo 1. $\left(K = e^{C(\mu-\mu_0) - \frac{1}{2}(\mu+\mu_0)}\right)$

Observemos que, con independencia del valor de μ, podemos construir un test de Neyman puro, con nivel de significación α, y cuya región crítica será W. Dicho test de Neyman, $\boxed{\phi(x) = 1_W}$, SEGUN TEOREMA NEYMAN-PEARSON será un test de error de tipo II mínimo, dentro de la clase de test de nivel de significación α, o bien de potencia máxima, sea cual sea el valor de μ, con $\mu > \mu_0$. Por tanto, diremos que en este caso (Ejemplo 2) existe un test uniformemente de potencia máxima para resolverlo: un test UMP (Uniformly Most Powerful test)

La función de potencia será:

$$\boxed{\alpha(\mu) = 1 - \Phi\left(\sqrt{n}\,(\mu_0-\mu) + \Phi^{-1}(1-\alpha)\right)}$$

Nota:
Obsérvese que si $\mu \to \infty$, entonces $\alpha(\mu) \to 1$ y si $\mu \to \mu_0$, entonces $\alpha(\mu) = \alpha$

Por otra parte, fijado $\mu > 0$, si $n \to \infty$ entonces $\alpha(\mu) \to 1$.
Este problema ilustra el "funcionamiento" de los tests de hipótesis: controlamos el error de tipo I, mientras que el error de tipo II depende del valor del parámetro desconocido.
La única manera de rebajar el error de tipo II consiste en aumentar el tamaño muestral n. Si $\theta \in \Theta \setminus \Theta_0$ a medida que $n \to \infty$ la potencia de cualquier test "razonable" tenderá a 1, y, por tanto, el error de tipo II tenderá a cero.

El test de la Razón de Verosimilitud

No siempre existen tests UMP para un problema de contraste de hipótesis determinado. En este caso, existe una metodología general que permite construir tests razonables en muchas situaciones.

Utilizando la misma notación que antes, dado el problema de contraste de hipótesis:

$H_0: \ \theta \in \Theta_0$

$H_1: \ \theta \in \Theta \setminus \Theta_0$

definiremos el estadístico:

$$\widehat{L_x(\Theta_0)} = \sup_{\theta \in \Theta_0} L_x(\theta)$$

donde $L_x(\theta)$ es la función de verosimilitud. Definimos también:

$$\widehat{L_x(\Theta)} = \sup_{\theta \in \Theta} L_x(\theta)$$

así como:

$$\boxed{\Lambda(x) = \frac{\widehat{L_x(\Theta_0)}}{\widehat{L_x(\Theta)}}}$$

este último estadístico se denomina la razón de verosimilitud.

Obsérvese que $0 \leq \Lambda(x) \leq 1$. Si la hipótesis nula es cierta cabe esperar valores altos de $\Lambda(x)$ mientras que valores bajos son poco probables bajo H_0. Por tanto, parece razonable definir un test puro, con región crítica igual a:

$$\boxed{W = \{ x \in \Omega \ / \ \Lambda(x) < c \} \quad \text{para cierta constante } c \in (0,1)}$$

a este test le llamaremos test de la razón de verosimilitud.

En algunos casos afortunados se conocerá la distribución de $\Lambda(x)$ o de una transformación de $\Lambda(x)$. En este caso podrá determinarse c de forma que el test tenga nivel de significación α.

En caso de no conocer la distribución de $\Lambda(x)$ o de alguna transformación conveniente, podemos recurrir al siguiente resultado:

Teorema de Wilks

Bajo las condiciones de regularidad habituales que garantizan las propiedades asintóticas de los estimadores máximo-verosímiles, y si $\dim \Theta - \dim \Theta_0 = k > 1$ (dimensión en el sentido de "variedades diferenciables": espacios "geométricos" con coordenadas, donde "dimensión" es el número de coordenadas correspondientes)

Entonces, bajo H_0,

converge en ley

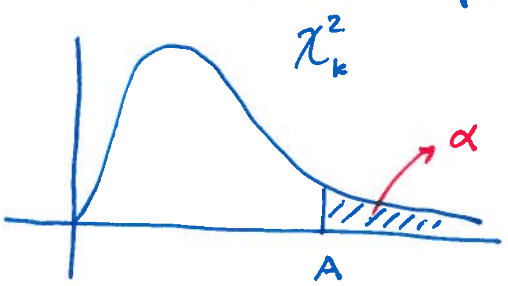
$$\boxed{-2 \ln \Lambda(x) \xrightarrow{\mathcal{L}} U \sim \chi^2_k}$$

es decir: la distribución asintótica (cuando el tamaño muestral tiende a ∞) de $-2\ln\Lambda(x)$, bajo la hipótesis nula, es una distribución χ^2 con k grados de libertad.

Entonces podemos definir al test de la Razón de verosimilitud de forma equivalente a partir de la región crítica

$$W = \{ x \in \Omega \;/\; -2\ln\Lambda(x) > \underbrace{-2\ln c}_{A} \}$$

y determinar A a partir de la distribución asintótica:



Es decir:

$$\alpha = P(U > A)$$

Entonces tendremos:

$$P(-2\ln\Lambda(x) > A \mid H_0) \simeq \alpha$$

y por tanto la región crítica W tendrá un nivel de significación aproximado α.

Ejemplo 3

$X \sim N(\mu, 1) \quad \mu \in \mathbb{R} \qquad X_1, \dots, X_n$ i.i.d X

$H_0: \mu = \mu_0$ (simple)

$H_1: \mu \neq \mu_0$ (compuesta)

A partir del ejemplo 2 vemos que en este caso no existirá un test UMP. Vamos a efectuar el test de la razón de verosimilitud.

La función de verosimilitud es:

$$L_x(\mu) = \prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_i-\mu)^2} \right\} = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

pero como:

$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$

$$\sum_{i=1}^{n}(x_i-\mu)^2 = \sum_{i=1}^{n}\left((x_i-\bar{x}_n)+(\bar{x}_n-\mu)\right)^2 =$$

$$= \sum_{i=1}^{n}(x_i-\bar{x}_n)^2 + n(\bar{x}_n-\mu)^2 + 2(\bar{x}_n-\mu)\underbrace{\sum_{i=1}^{n}(x_i-\bar{x}_n)}_{=0}$$

$$= n\, s_n^2 + n(\bar{x}_n-\mu)^2$$

$s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x}_n)^2$

por tanto:

$$L_x(\mu) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} e^{-\frac{n}{2}s_n^2 - \frac{n}{2}(\bar{x}_n-\mu)^2} =$$

$$= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} e^{-\frac{n}{2}s_n^2} e^{-\frac{n}{2}(\bar{x}_n-\mu)^2} \leq \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} e^{-\frac{n}{2}s_n^2}$$

valor que se alcanza cuando $\bar{x}_n = \mu$

por tanto:

$$L_x(\hat{\Theta}) = L_x(\bar{x}_n) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} e^{-\frac{n}{2}s_n^2}$$

y por otra parte:

$$L_x(\hat{\Theta}_0) = L_x(\mu_0) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} e^{-\frac{n}{2}s_n^2} e^{-\frac{n}{2}(\bar{x}_n-\mu_0)^2}$$

La razón de verosimilitud será:

$$\Lambda(x) = \frac{L_x(\widehat{\Theta_0})}{L_x(\widehat{\Theta})} = e^{-\frac{n}{2}(\bar{x}_n - \mu_0)^2}$$

por tanto, la región crítica correspondiente al test de la razón de verosimilitud vendrá dada por:

$$e^{-\frac{n}{2}(\bar{x}_n - \mu_0)^2} < c$$

equivalente a:

$$-\frac{n}{2}(\bar{x}_n - \mu_0)^2 < \ln c$$

$$(\bar{x}_n - \mu_0)^2 > -\frac{2}{n}\ln c$$

$$|\bar{x}_n - \mu_0| > \underbrace{\sqrt{-\frac{2}{n}\ln c}}_{= A}$$

$= A$ (una constante en el sentido de que no depende de $x_1 - x_n$)

por tanto podemos expresar la región crítica del test de la razón de verosimilitud como:

$$W = \{ (x_1 - - - x_n) \in \mathbb{R}^n \mid \quad |\bar{x}_n - \mu_0| > A \}$$

Ahora bien, bajo H_0 $\quad \bar{X}_n \sim N(\mu_0, var = \frac{1}{n})$

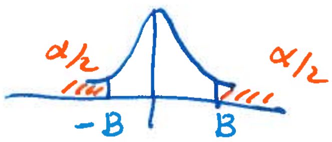
por tanto:

$$\alpha = P(\; |\bar{X}_n - \mu_0| > A \mid H_0) =$$

$$= P\Big(\; |Z| > \underbrace{\frac{A}{\frac{1}{\sqrt{n}}}}_{B} \Big) \qquad \text{donde } Z \sim N(0,1)$$

por tanto

$$\alpha = P(\,|Z| > \underbrace{\sqrt{n}\,A}_{B})$$



(por simetría)

$$\Phi(\sqrt{n}\,A) = 1 - \alpha/2 \quad \Longrightarrow \quad \sqrt{n}\,A = \Phi^{-1}(1 - \alpha/2)$$

$$\boxed{A = \frac{1}{\sqrt{n}}\,\Phi^{-1}(1 - \alpha/2)}$$

de esta forma se garantiza que la región crítica tiene nivel de significación α:

$$\boxed{W = \left\{(x_1 - x_n) \in \mathbb{R}^n \;\middle|\; |\bar{x}_n - \mu| > \frac{1}{\sqrt{n}}\, \Phi^{-1}(1-\alpha/2)\right\}}$$

donde Φ es la función de distribución de una normal $N(0,1)$ y Φ^{-1} es su inversa.

Ejemplo 4

$X \sim$ Exponencial (α), $\alpha \in \mathbb{R}^+$ $\qquad X_1 - - X_n$ i.i.d X

$H_0: \alpha = 1$ (simple)

$H_1: \alpha \neq 1$ (compuesta)

Resolver el problema de contraste de hipótesis mediante el test de la razón de verosimilitud.

NOTA: Si tratáramos de obtener un test UMP veríamos que no es posible.

La función de verosimilitud es: (con probabilidad 1)

$$L_X(\alpha) = \prod_{i=1}^{n} \left(\alpha\, e^{-\alpha x_i}\right) = \alpha^n e^{-\alpha \sum_{i=1}^n x_i}$$

$$L_X(\widehat{\Theta_0}) = L_X(1) = e^{-\sum_{i=1}^n x_i}$$

$$L_X(\widehat{\Theta}) = L_X(\alpha^*) = \left(\frac{1}{\bar{x}_n}\right)^n e^{-\frac{1}{\bar{x}_n}\sum_{i=1}^n x_i} = \left(\frac{1}{\bar{x}_n}\right)^n e^{-n}$$

ya que $\alpha^* = \dfrac{1}{\bar{x}_n}$ (MLE) $\qquad$ donde $\bar{x}_n = \dfrac{1}{n}\sum_{i=1}^n x_i$

por tanto:

$$\Lambda(X) = \frac{L_X(\widehat{\Theta_0})}{L_X(\widehat{\Theta})} = \frac{e^{-n\bar{x}_n}}{\left(\frac{1}{\bar{x}_n}\right)^n e^{-n}} = (\bar{x}_n)^n e^{-n(\bar{x}_n - 1)}$$

La región crítica obtenida mediante el test de la razón de verosimilitud es:

$$W = \left\{ (x_1 - x_n) \in \mathbb{R}^n \;/\; (\bar{x}_n)^n e^{-n(\bar{x}_n - 1)} < c \right\}$$

Para determinar C de forma que el nivel de significación aproximado sea α, podemos utilizar el Teorema de Wilks:

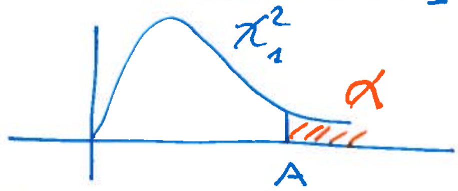
$$W = \left\{ (x_1 - x_n) \in \mathbb{R}^n \;\middle|\; -2n\ln(\bar{x}_n) + 2n(\bar{x}_n - 1) > \underbrace{-2\ln c}_{A} \right\}$$

determinando A de forma que

$$\alpha = P(\; U > A) \qquad \text{donde } U \sim \chi^2_1$$

ya que bajo Ho $-2\ln\Lambda \xrightarrow{\mathcal{L}} U \sim \chi^2_1$

siendo U una variable con distribución χ^2_1



$$1 - \alpha = F_{\chi^2_1}(A)$$

Si $F_{\chi^2_1}$ es la función de distribución de una χ^2 con 1 grado de libertad, entonces

$$A = F^{-1}_{\chi^2_1}(1-\alpha)$$

y el test queda bien determinado, con nivel de significación α

$$\boxed{W = \left\{ (x_1 - x_n) \in \mathbb{R}^n \;/\; -2n\ln(\bar{x}_n) + 2n(\bar{x}_n - 1) > F^{-1}_{\chi^2_1}(1-\alpha) \right\}}$$

Otro análisis más exacto:

Sea $\quad h(x) = x^n e^{-n(x-1)}$ para $x > 0$

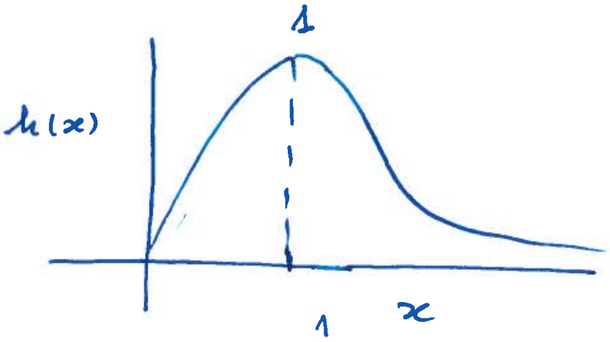
$h(x) > 0 \quad \forall x > 0$ y además:

$$h'(x) = n x^{n-1} e^{-n(x-1)} + x^n e^{-n(x-1)}(-n)$$

$$= \underbrace{n x^{n-1} e^{-n(x-1)}}_{>0} \{1 - x\}$$

por tanto $h(x)$ es creciente para $1-x>0$, es decir $0<x<1$ y decreciente para $1-x<0$, es decir si $x>1$, presentando un máximo absoluto en $x=1$.

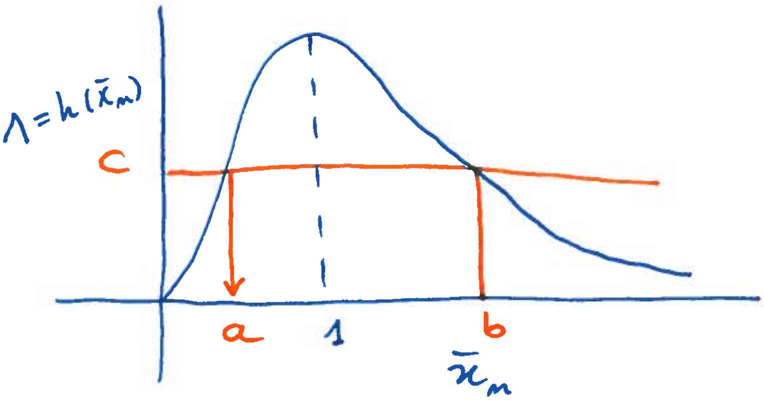
obsérvese la gráfica aproximada



Ahora bien la desigualdad $(\bar{x}_n)^n e^{-n(\bar{x}_n-1)} < c$ puede expresarse como:

$$h(\bar{x}_n) < c \qquad (\text{porque } \Lambda(x_1-x_n) = h(\bar{x}_n))$$

es decir, en términos gráficos:



$$h(\bar{x}_n) < c \iff \bar{x}_n < a \text{ o bien } \bar{x}_n > b$$

donde a y b son números tales que:

$h(a) = h(b)$, es decir:

$$a^n e^{-n(a-1)} = b^n e^{-n(b-1)} \qquad \text{equivalente a}$$

$$a^n e^{-na} e^n = b^n e^{-nb} e^n \qquad \text{equivalente a}$$

$$a e^{-a} = b e^{-b} \qquad (\text{ya que } a^n e^{-na} = (a e^{-a})^n \text{ etc..})$$

o bien:
$$\boxed{\frac{a}{b}\, e^{\,b-a} = 1}$$

Además:

$$P(\bar{X}_n < a \mid H_0) + P(\bar{X}_n > b \mid H_0) = \alpha$$

equivalente a

$$P(a \leq \bar{X}_n \leq b \mid H_0) = 1-\alpha$$

$$P\left(na \leq \sum_{i=1}^{n} X_i \leq nb \mid H_0\right) = 1-\alpha$$

pero bajo H_0 $\sum_{i=1}^{n} X_i \sim \text{Gamma}(1, n)$

y, multiplicando por 2, bajo H_0, $\boxed{2\sum_{i=1}^{n} X_i \sim \text{Gamma}\left(\frac{1}{2}, n\right)}$ $= \chi^2_{2n}$

es decir

$$P(2na \leq U \leq 2nb) = 1-\alpha \qquad \text{donde } U \sim \chi^2_{2n}$$

Por tanto a y b pueden determinarse resolviendo (numéricamente) el sistema:

$$\boxed{\begin{array}{c} F_{\chi^2_{2n}}(2nb) - F_{\chi^2_{2n}}(2na) = 1-\alpha \\[2ex] \dfrac{a}{b}\, e^{\,b-a} = 1 \end{array}}$$

donde $F_{\chi^2_{2n}}$ es la función de distribución de una χ^2 con $2n$ grados de libertad