Universidad de Chile
Facultad de Ciencias Físicas y Matemáticas
Departamento de Ingeniería Matemática
MA3403 - Probabilidades y Estadística

Test De Hipótesis

Alberto Vera Azócar, albvera@ing.uchile.cl

1. Motivación

Estudiaremos un caso particular de test de hipótesis, el que corresponde a test paramétrico, es decir, tenemos una v.a. X y sabemos que $X \sim F(\theta)$, donde el parámetro θ es desconocido, y lo que nos gustaría es tomar una decisión sobre como se comporta θ.

Pensemos que tenemos una moneda y a priori no sabemos si está cargada o no, luego uno se pregunta "¿cuál es la probabilidad de que me equivoque si afirmo que $p > \frac{1}{2}$?". Digamos que p es la probabilidad de cara y que en 100 lanzamientos nunca salió cruz, uno pensaría que es correcto entonces afirmar que $p > \frac{1}{2}$; digamos ahora que de los 100 lanzamientos salieron 60 caras ¿es evidencia suficiente para afirmar que $p > \frac{1}{2}$? y si afirmamos que $p > \frac{1}{2}$ ¿cuál es la probabilidad que nos estemos equivocando?. Para responder estas preguntas necesitamos hacer un test de hipótesis sobre p.

2. Nociones Básicas

De ahora en más, cuando digamos que $X \sim F(\theta)$ se entenderá que en realidad tenemos un modelo paramétrico $\{f(\theta)|\theta \in \Theta\}$ de la v.a. y que θ es desconocido.

Definición 2.1 (Hipótesis Paramétrica). *Sea $X \sim F(\theta)$, diremos que H_0 y H_1 son hipótesis paramétricas si podemos escribir*

$$\begin{aligned} H_0 &: \theta \in \Theta_0 \\ H_1 &: \theta \in \Theta_1 \end{aligned}$$

donde $\Theta_0, \Theta_1 \subset \Theta$ y $\Theta_0 \cap \Theta_1 = \varnothing$. Si la intersección no es vacía, entonces las hipótesis no están bien planteadas. A H_0 la llamamos hipótesis nula y a H_1 hipótesis alternativa. Es costumbre referirse a las hipótesis como "H_0 contra H_1".

Definición 2.2 (Regla de Decisión). *Sea $X_1, \cdots, X_n$ un m.a.s. de $X \sim F(\theta)$, consideremos las hipótesis H_0 contra H_1. Una regla de decisión es cualquier función $\phi : \mathbb{R}^n \to \{0,1\}$ de los valores muestrales. Interpretamos que cuando ϕ toma el valor 1 se rechaza H_0 y cuando ϕ toma el valor 0 no se rechaza H_0.*

Definición 2.3 (Región Crítica). *Sea $X_1, \cdots, X_n$ un m.a.s. de $X \sim F(\theta)$ y ϕ una regla de decisión para las hipótesis H_0 contra H_1. Se define la región crítica del test como*

$$R := \{(X_1, \cdots, X_n) \in \mathbb{R}^n | \quad \phi(X_1, \cdots, X_n) = 1\}$$

En otras palabras, la región crítica es una parte del espacio tal que si el muestreo cae en R se debe rechazar H_0.

Ejemplo 2.1. *Consideremos el problema de decidir si una moneda está cargada o no. Claramente modelamos como $X \sim Bernoulli(p)$, sea $X_1, \cdots X_n$ un m.a.s. y definamos las hipótesis*

$$\begin{aligned} H_0 &: p = 1/2 \\ H_1 &: p \neq 1/2 \end{aligned}$$

Obtenga la región crítica asociada a la regla de decisión

$$\phi(X_1, \cdots, X_n) = \begin{cases} 1 & si\ \overline{X} \notin [0.45, 0.55] \\ 0 & \sim \end{cases}$$

Solución: En este caso

$$\begin{aligned} R &\triangleq \{(X_1, \cdots, X_n) \in \mathbb{R}^n | \quad \phi(X_1, \cdots, X_n) = 1\} \\ &= \{(X_1, \cdots, X_n) \in \mathbb{R}^n | \quad \overline{X} \notin [0.45, 0.55]\} \\ &= \{(X_1, \cdots, X_n) \in \mathbb{R}^n | \quad \overline{X} > 0.55 \vee \overline{X} < 0.45\} \end{aligned}$$

Observamos que R tiene bastante sentido, ya que se rechazará H_0 si $\overline{X}$ se aleja mucho de $\frac{1}{2}$.

Es siempre recomendable expresar R en términos de un estadístico conocido, como lo es el promedio, la suma de las observaciones, el máximo de las observaciones, etc. Esto es porque al expresar R así, uno ve cosas interesantes. Otra razón más técnica es que así nos permite calcular probabilidades.

Definición 2.4 (Potencia). *Sea $X_1, \cdots, X_n$ un m.a.s. de $X \sim F(\theta)$ y consideremos las hipótesis H_0 contra H_1, definimos la función de potencia del test $\pi : \Theta \to [0, 1]$ como*

$$\pi(\theta) := \mathbb{P}((X_1, \cdots, X_n) \in R | \theta)$$

Algunos autores restringen el dominio de π a Θ_1, esto en realidad no es relevante.

La función de potencia se debe leer como $\pi(\theta) = \mathbb{P}(\text{rechazar } H_0 \text{ dado } \theta)$, en el ejemplo de la moneda $\pi(p)$ será alto por ejemplo si p es cercano a 1, ya que en este caso es muy probable que el promedio no esté en el intervalo $[0.45, 0.55]$ y por lo tanto se rechazará H_0, asímismo $\pi(p)$ será alto también cuando p es cercano a 0. Esta discusión motiva a introducir los errores en el test, ya que uno pensaría que es bueno que $\pi(p)$ sea alto cuando H_0 es falsa.

Ejemplo 2.2. *Sea un m.a.s. $X_1, \cdots, X_n$ del modelo normal con media μ desconocida y varianza σ^2 conocida. Considere las hipótesis*

$$\begin{aligned} H_0 &: \mu = 0 \\ H_1 &: \mu > 0 \end{aligned}$$

Considere la regla de decisión que rechaza H_0 cuando $\overline{X} > \alpha$, donde $\alpha > 0$ dado. Calcule la potencia del test.

Solución: Claramente la región crítica es

$$R = \{(X_1, \cdots, X_n) : \overline{X} > \alpha\}$$

Sabemos que $\overline{X} \sim \mathcal{N}(\mu, \sigma^2/n)$*, calculemos entonces por definición*

$$\begin{aligned}
\pi(\mu) &\triangleq \mathbb{P}((X_1, \cdots, X_n) \in R|\mu) \\
&= \mathbb{P}(\overline{X} > \alpha|\mu) \\
&= \mathbb{P}\left(\frac{\overline{X} - \mu}{\sqrt{\sigma^2/n}} > \frac{\alpha - \mu}{\sqrt{\sigma^2/n}} \middle| \mu \right) \\
&= 1 - \Phi\left(\frac{\alpha - \mu}{\sqrt{\sigma^2/n}} \right)
\end{aligned}$$

Ahora interpretemos, si μ *es grande (es decir, es falsa* H_0*), entonces* $\alpha - \mu < 0$*, luego*

$$\lim_{n\to\infty} \Phi\left(\frac{\alpha - \mu}{\sqrt{\sigma^2/n}} \right) = 0$$

sigue que para μ *grande* $\mathbb{P}(\textit{rechazar } H_0|\mu) = \pi(\mu) \to 1$ *cuando* $n \to \infty$*, que es lo que uno esperaría.*

Observar que $\pi(\mu)$ *depende explícitamente de* α*, que es en el fondo la regla de decisión ¿cuál será el mejor* α*?*

Propuesto: Encuentre en función de α el límite cuando $n \to \infty$ de $\pi(\mu)$ en el caso que $\mu < \alpha$, concluya que el buen valor de α es 0.

Como vimos en el ejemplo, R depende explícitamente de una regla de decisión, por lo que para ser formales deberíamos escribir la función de potencia como $\pi_\phi(\theta)$ ("probabilidad de rechazar H_0 dado θ y que estamos usando la regla ϕ"), pero para no recargar la notación no lo haremos, a menos que sea estrictamente necesario para evitar ambigüedad. En general tendremos una sola regla de decisión.

Aceptaremos la siguiente definición informal, porque no vale la pena formalizar un concepto tan intuitivo:

Definición 2.5 (Errores). *Sean las hipótesis* H_0 *contra* H_1*, se define el error de tipo I como "rechazo incorrecto de hipótesis nula" y el error de tipo II como el contrario, es decir "aceptación incorrecta de hipótesis nula".*

Proposición 2.1. *Sean las hipótesis* H_0 *contra* H_1*, entonces*

$$\begin{aligned}
\mathbb{P}(\textit{Cometer error de tipo I } |\theta) &= \pi(\theta) \quad \theta \in \Theta_0 \\
\mathbb{P}(\textit{Cometer error de tipo II } |\theta) &= 1 - \pi(\theta) \quad \theta \in \Theta_1
\end{aligned}$$

Demostración. Directa □

Definición 2.6 (Tamaño de un Test). *Sean las hipótesis* H_0 *contra* H_1*, con una regla de decisión* ϕ*, se define el tamaño del test como*

$$\alpha(\phi) := \sup_{\theta \in \Theta_0} \pi(\theta)$$

Es entonces el tamaño del test la mayor probabilidad de cometer error de tipo I, en otras palabras, si tenemos una regla de decisión ϕ que te garantiza tamaño 5 %, entonces cada vez que rechacemos H_0 (usando la regla ϕ) con probabilidad 95 % no nos estaremos equivocando. Esto suena muy bien y es en realidad con lo que uno se contenta: fijamos una cota máxima para el error de tipo I (es decir, exigimos cierto tamaño de test) y luego hacemos lo mejor que podemos con el error de tipo II[1].

3. El p-valor

Consideremos dos enfoques distintos, primero el que ya revisamos, que nos dice que dado un nivel α encontremos una región de rechazo que respete ese nivel, luego se debe rechazar si los datos caen en esa región, esto lo muestra el siguiente ejemplo.

Ejemplo 3.1. *Consideremos el tiempo que dura una ampolleta $T \sim exp(\lambda)$ con λ desconocido. La hipótesis nula es $H_0 : \lambda = 1$, se tiene sólo un dato y se propone rechazar la hipótesis nula si $T < r$. Encuentre un valor de r que garantice que la probabilidad de cometer error de tipo I es a lo más α.*

Solución: En este caso podemos calcular directamente la probabilidad de cometer error de tipo I, consideremos que r está fijo:

$$\begin{aligned}\mathbb{P}(\textit{error de tipo I}) &= \mathbb{P}(\textit{rechazar } H_0 | H_0 \textit{ cierta})\\ &= \mathbb{P}(T < r | H_0 \textit{ cierta})\\ &= \mathbb{P}(T < r | \lambda = 1)\\ &= 1 - e^{-r}\end{aligned}$$

Imponemos que el error de tipo I es α y obtenemos $\alpha = 1 - e^{-r}$ o bien $r = -\ln(1-\alpha)$.

Finalmente, la regla es la siguiente: "si el valor observado es menor a $-\ln(1-\alpha)$, rechaza H_0 y te estarás equivocando con probabilidad α.

Propuesto: El valor de r dado en el ejemplo anterior ¿es una cota superior o inferior del valor de r?

Ahora consideremos otro enfoque, dada una regla de decisión que nos da la región de rechazo R, nos preguntamos ¿cuál es la probabilidad de caer en la región R suponiendo que es cierta H_0?, a ese valor le llamamos p-valor.

Antes de digerir la definición formal, veamos un ejemplo donde calculamos el p-valor.

Ejemplo 3.2. *Se tiene una moneda que sale cara con probabilidad p desconocida. Se cree que la moneda está cargada y se plantean las siguientes hipótesis:*

$$\begin{aligned}H_0 : p = \frac{1}{2}\\ H_1 : p > \frac{1}{2}\end{aligned}$$

Se lanza 3 veces, si las 3 veces resulta cara entonces se rechaza H_0. Calcule el p-valor del test.

Solución: Según nuestra definición informal, el p-valor es la probabilidad de caer en la región de rechazo dado que es cierta H_0, vale decir, dado que $p = \frac{1}{2}$ que las tres monedas sean cara. Escrito así, trivialmente el p-valor es $(\frac{1}{2})^3$.

[1]Este es el enfoque inicial, hay porsupuesto otros más elaborados, como por ejemplo minimizar una suma ponderada de los errores.

Del ejemplo anterior se interpreta que si el p-valor es pequeño **hay que rechazar** H_0, en efecto, el p-valor es la probabilidad de observar a una moneda equilibrada que sale sólo en caras.

Definición 3.1. *Sea una regla de decisión con región de rechazo R y sean $X_1, \cdots, X_n$ los valores muestrales. Sea φ la función que nos dice si rechazar dado un nivel de confianza α:*

$$\varphi(\alpha) = \begin{cases} 1 & si\ \mathbb{P}((X_1, \cdots, X_n) \in R | H_0) \leq \alpha \\ 0 & \sim \end{cases}$$

entonces el p-valor viene dado por $\inf\{\alpha \in [0,1] : \varphi(\alpha) = 1\}$.

4. Lema De Neyman-Pearson

Definición 4.1. *Sean las hipótesis H_0 contra H_1, diremos que H_0 es simple (resp. H_1) si Θ_0 (resp. Θ_1) es un síngleton. En caso contrario diremos que la hipótesis es compuesta.*

El lema de Neyman-Pearson nos da una regla de decisión para hipótesis simple contra simple que para un tamaño dado α es u.m.p., es decir, de entre todos los test de tamaño α es el que aminora (uniformemente) el error de tipo II.

Teorema 4.1. *Sea X v.a. con densidad (o ley) $f(x|\theta)$ y $X_1, \cdots, X_n$ un m.a.s. de la v.a., consideremos las hipótesis*

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_1 &: \theta = \theta_1 \end{aligned}$$

entoces dado $\alpha \in (0,1)$, el mejor test de tamaño α (en el sentido de la potencia) está dado por

$$\begin{aligned} \frac{f_1}{f_0} > K &\implies \textit{Rechaza } H_0 \\ \frac{f_1}{f_0} < K &\implies \textit{No rechaza } H_0 \end{aligned}$$

donde

$$f_1 = \prod_{i=1}^{n} f(X_i|\theta_1) \qquad f_0 = \prod_{i=1}^{n} f(X_i|\theta_0)$$

5. Test De Razón De Verosimilitud

Definición 5.1. *Sea X v.a. con densidad (o ley) $f(x|\theta)$ y $X_1, \cdots, X_n$ un m.a.s. de la v.a., consideremos las hipótesis*

$$\begin{aligned} H_0 &: \theta \in \Theta_0 \\ H_1 &: \theta \in \Theta_1 \end{aligned}$$

entoces dado $\alpha \in (0,1)$, la regla de razón de verosimilitud está dada por

$$\begin{aligned} \lambda(X_1, \cdots, X_n) > K &\implies \textit{Rechaza } H_0 \\ \lambda(X_1, \cdots, X_n) < K &\implies \textit{No rechaza } H_0 \end{aligned}$$

donde

$$\lambda(X_1, \cdots, X_n) := \frac{\sup_{\theta \in \Theta} f(X_1, \cdots, X_n|\theta)}{\sup_{\theta \in \Theta_0} f(X_1, \cdots, X_n|\theta)}$$

A diferencia del lema de Neyman-Pearson, la razón de verosimilitud no ofrece ninguna garantía de optimalidad, sin embargo en general entrega reglas de decisión muy razonables.