Capítulo 1

Introducción

A continuación recordamos una serie de conceptos de la asignatura de Teoría de la Probabilidad necesarios para abordar el curso.

1.1. Convergencia de Variables Aleatorias

Definición 1.1.1. *Una sucesión de variables aleatorias $\{X_n\}$ definida sobre un espacio probabilístico $(\Omega, \mathscr{A}, P)$ converge de manera casi segura a X, lo cual notaremos $X_n \xrightarrow{CS} X$, si se verifica que:*

$$P\left(\{\omega \in \Omega \mid X_n(\omega) \longrightarrow X(\omega)\}\right) = 1$$

Definición 1.1.2. *Una sucesión de variables aleatorias $\{X_n\}$ definida sobre un espacio probabilístico $(\Omega, \mathscr{A}, P)$ converge en probabilidad a X, lo cual notaremos $X_n \xrightarrow{P} X$ si se verifica alguna de las siguientes condiciones, las cuales son trivialmente equivalentes:*

$$P\left(\{\omega \in \Omega \mid |X_n(\omega) - X(\omega)| \leq \varepsilon\}\right) = 1$$
$$P\left(\{\omega \in \Omega \mid |X_n(\omega) - X(\omega)| > \varepsilon\}\right) = 0$$

Definición 1.1.3. *Una sucesión de variables aleatorias $\{X_n\}$ definida sobre un espacio probabilístico $(\Omega, \mathscr{A}, P)$ con funciones de distribución F_{X_n} converge de manera en Ley a X, lo cual notaremos $X_n \xrightarrow{L} X$ si se verifica que:*

$$F_{X_n}(x) \longrightarrow F_X(x) \ \forall x \text{ punto de continuidad de } F_X$$

1.1.1. Propiedades

Proposición 1.1.4. *La convergencia casi segura implica convergencia en probabilidad, la cual a su vez implica convergencia en ley.*

Proposición 1.1.5. *Si* $X_n \xrightarrow{L} c \in \mathbb{R}$ *entonces se verifica que* $X_n \xrightarrow{P} c$.

Álgebra de convergencia

Proposición 1.1.6. *Sean* $X_n \xrightarrow{CS} X$, $Y_n \xrightarrow{CS} Y$ *y* g *función continua. Se verifica que:*

- $X_n + Y_n \xrightarrow{CS} X + Y$
- $X_n Y_n \xrightarrow{CS} XY$
- $g(X_n) \xrightarrow{CS} g(X)$

Proposición 1.1.7. *El resultado anterior es igualmente cierto para la convergencia en probabilidad.*

Proposición 1.1.8. *Si* $X_n \xrightarrow{P} X$ *e* $Y_n \xrightarrow{L} c \in \mathbb{R}$ *entonces el resultado anterior es cierto para la convergencia en ley.*

1.1.2. Algunos resultados de convergencia

Proposición 1.1.9. *Sea* $\{X_n\}$ *una sucesión de variables aleatorias absolutamente continuas y sea* f_{X_n} *la función de densidad de* X_n. *Supongamos que existe una función de densidad* f_X *de forma que:*

$$f_{X_n}(x) \longrightarrow f_X(x) \; \forall x \in \mathbb{R}$$

Entonces X*, la variable cuya función de densidad es* f_X *verifica que* $X_n \xrightarrow{L} X$.

Proposición 1.1.10. *Sea* $\{X_n\}$ *una sucesión de variables aleatorias discretas con valores en* $\mathbb{N}$ *y sea* P_{X_n} *su función de probabilidad. Supongamos que existe* P_X *función de probabilidad de forma que:*

$$P[X_n = k] \longrightarrow P[X = k] \; \forall k \in \mathbb{N}$$

Entonces X*, la variable cuya función de probabilidad es* P_X *verifica que* $X_n \xrightarrow{L} X$.

Proposición 1.1.11. *Sea $\{X_n\}$ una sucesión de variables aleatorias con función generatriz de momentos M_{X_n} definidas en un entorno I de 0. Supongamos que existe una función generatriz de momentos tal que:*

$$M_{X_n}(t) \longrightarrow M_X(t)\ \forall x \in I$$

Entonces X, la variable función generatriz es M_X verifica que $X_n \xrightarrow{L} X$.

1.2. Teoremas Importantes

Teorema 1.2.1 (Ley fuerte de los grandes números)**.** *Sea $\{X_n\}$ una sucesión de variables aleatorias i.i.d con primer momento finito. Sea $\mu = E[X_n]$ entonces:*

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{CS} \mu$$

Teorema 1.2.2 (Teorema central del límite)**.** *Sea $\{X_n\}$ una sucesión de variables aleatorias i.i.d con segundo momento finito. Sea $S_n = \sum_{k=1}^{n} X_k$, $\mu = E[X_n]$ y $\sigma^2 = V[X_n]$ entonces:*

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{L} N(0,1)$$

Nota 1.2.3. *Podemos considerar las siguientes igualdades:*

$$\frac{\sum_{k=1}^{n} X_k - E\left[\sum_{k=1}^{n} X_k\right]}{\sqrt{V[\sum_{k=1}^{n} X_k]}} = \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n\sigma^2}} = \frac{\frac{1}{n}\sum_{k=1}^{n} X_k - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Demostración (Prop 1.1).

$$F_X(x) = F_{X_1,\dots,X_n}(x_1,\dots,x_n) = P(X_1 \le x_1,\dots,X_n \le x_n) \overset{Indep}{=} \tag{1.2.1}$$

$$= \prod_{i=1}^{n} P(X_i \le x_i) = \prod_{i=1}^{n} F_{X_i}(x_i) = \prod_{i=1}^{n} F(x_i) \tag{1.2.2}$$

Demostración (Teorema 1.2). *Veamos las demostraciones de cada apartado.*

1. *ε_i es por definición una $Be(p)$ con p la probabilidad de que $X_i \le x$, es decir $F_{X_i}(x) = F(x)$. Además, son identéndicamentes distribuidas por definición y son independientes por ser estar construidas mediante variables aleatorias independientes; por lo que es una m.a.*

2. *$nF_n^*(x)$ es una suma de $Be(F(x))$, lo cual es bien sabido es una $Bi(n, F(x))$.*

3. *Veamos el cálculo*

$$E\left[F_n^*(x)\right] = E\left[\frac{1}{n}\sum_{k=1}^{n}\varepsilon_i(x)\right] = \frac{1}{n}\sum_{k=1}^{n} E\left[\varepsilon_i\right] = \frac{1}{n}\sum_{k=1}^{n} F(x) = F(x) \tag{1.2.3}$$

4. *Basta realizar un cálculo análogo al anterior.*

5. *Es consecuencia inmediata del la Ley Fuerte de los Grandes Números.*

6. *Es consecuencia inmediata del Teorema Central del Límite.*

Demostración (Prop 3.2). *Veamos la primera igualdad:*

$$ECM_\theta(T) = E_\theta[(T - h(\theta))^2] = E_\theta[(T - E_\theta[T] + E_\theta[T] - h(\theta))^2] = \tag{1.2.4}$$

$$E_\theta\left[(T - E_\theta\left[T\right])^2 + (h(\theta) - E_\theta\left[T\right])^2 - 2(T - E_\theta\left[T\right])(h(\theta) - E_\theta\left[T\right])\right] = \tag{1.2.5}$$

$$V_\theta[T] + E_\theta\left[B_\theta(T)^2\right] = V_\theta[T] + B_\theta(T)^2 \tag{1.2.6}$$

La segunda igualdad se deduce de la definición de ser insesgado.

Demostración (Prop 3.3). *Realicemos la prueba por reducción al absurdo. Supongamos que $\exists T^*$. Dado θ_0, consisderamos el estimador $T = h(\theta_0)$. Entonces:*

$$ECM_\theta(T) = (h(\theta) - h(\theta_0))^2 \Rightarrow ECM_{\theta_0}(T) = 0 \Rightarrow ECM_{\theta_0}(T^*) = 0 \tag{1.2.7}$$

$$ECM_\theta(T^*) = 0\ \forall\theta \Rightarrow P_\theta[T^* = h(\theta)] = 1\ \forall\theta \Rightarrow T^* = h(\theta) \tag{1.2.8}$$

Demostración (Prop 3.5)**.** *Vamos a demostrar que* $T_n \xrightarrow{P} h(\theta)$. *Esto ocurre si* $\forall\varepsilon$ $\lim_{n\to\infty} P_\theta[|T_n - h(\theta) > \varepsilon] = 0$. *Consideramos:*

$$0 \leq P_\theta[|T_n - h(\theta)| > \varepsilon] \overset{Cheb.}{\leq} \frac{E_\theta[(T_n - h(\theta))^2]}{\varepsilon^2} = \frac{V_\theta[T] + B_\theta(T)^2}{\varepsilon^2} \to 0 \qquad (1.2.9)$$

Deducimos por tanto que: $P_\theta[|T_n - h_\theta| > \varepsilon] \to 0$.

Demostración 1.2.4 (Teorema 3.1)**.** *Observamos que* $\max_\theta L(x,\theta) = \max_\lambda M(x,\lambda)$. *Luego:*

$$M(x, h(\theta_{mv}) = \max_{\theta | h(\theta) = h(\theta_{MV})} L(x,\theta) \geq L(x, \theta_{MV}) = \max_\theta L(x,\theta) = \max_\lambda M(x,\lambda) \Rightarrow$$
$$M(x, h(\theta_{mv}) = \max_\lambda M(x,\lambda) \Leftrightarrow h(\theta_{mv}) \ EMV \ de \ h(\theta)$$

Demostración 1.2.5 (Teorema)**.** *Haremos la demostración para el caso discreto. Supongamos que* S *es suficiente. Sean* x_0 *arbitrario y sea* $s_0 = S(x_0)$.

$$P_\theta[X = x_0] = P_\theta[X = x_0, S = s_0] = P_\theta[X = x_0 | S = s_0] P_\theta[S = s_0]$$
$$g(S(x),\theta) = P_\theta[S = s_0] \quad h(x) = P_\theta[X = x_0 | S = s_0]$$

Para el recíproco, veamos que S *es suficiente.*

$$P_\theta[X = x_0 | S = s_0] = \frac{P_\theta[X = x_0, S = s_0]}{P_\theta[S = s_0]} = \begin{cases} 0 & A \cap B = \emptyset \\ \dfrac{P_\theta[X = x_0]}{\sum P_\theta[X = y]} & A \subset B \end{cases}$$
$$\frac{P_\theta[X = x_0]}{\sum_{y|S(y)=s_0} P_\theta[X = y]} = \frac{h(x) g(S(x),\theta)}{\sum_{y|S(y)=s_0} h(y) g(S(y),\theta)} = \frac{h(x)}{\sum_{y|S(y)=s_0} h(y)}$$

Demostración 1.2.6 (I.C. Chebychev)**.**

$$\sigma^* \geq \sup_\theta \sqrt{V_\theta[T]} \qquad P[|X-\mu| \geq a] \leq \frac{\sigma^2}{a^2}$$

$$P[|T(X)-h(\theta)| \leq k\sigma^*] \geq 1 - \frac{V[T(X)]}{(\sigma^* k)^2} \geq 1 - \frac{1}{k^2}$$

Demostración 1.2.7. *Si* $T_n \xrightarrow{L} N(\theta)$, *entonces* $\dfrac{T_n-\theta}{\sqrt{V[T_n]}} \xrightarrow{L} N(0,1)$. *Dado que* $V[T_n] \xrightarrow{P} V(\theta)$, *se tiene que:*

$$H_n = \frac{T_n-\theta}{\sqrt{V_n}} \xrightarrow{L} N(0,1)$$

$$P[-z_{1-\frac{\alpha}{2}} \leq H_n \leq z_{1-\frac{\alpha}{2}}] \geq P[-z_{1-\frac{\alpha}{2}} < H_n \leq z_{1-\frac{\alpha}{2}}] =$$

$$F_{H_n}(z_{1-\frac{\alpha}{2}}) - F_{H_n}(-z_{1-\frac{\alpha}{2}}) \to F_Z(z_{1-\frac{\alpha}{2}}) - F_Z(-z_{1-\frac{\alpha}{2}}) = 1-\alpha$$

Proposición 1.2.8. *Supongamos que* $\exists E[X]=\mu$, $\exists V[X]=\sigma^2$. *Entonces:*

$$\overline{X} \pm z_{1-\frac{\alpha}{2}} \frac{S_c}{\sqrt{n}}$$

Demostración 1.2.9.

$$\frac{\overline{X}-\mu}{\sigma}\sqrt{n} \xrightarrow{L} Z\ N(0,1) \qquad S_c^2 \xrightarrow{P} \sigma^2$$

$$\frac{\overline{X}-\mu}{S_c}\sqrt{n} \xrightarrow{L} N(0,1)$$

Tema 1

Introducción a la Inferencia Estadística

Contenido

1.1 Introducción

- Variable aleatoria: Herramienta para modelizar la variabilidad asociada a la característica objeto de estudio.
- Objetivo de la inferencia estadística: Dar respuestas "razonables" a cuestiones de interés relacionadas con la variabilidad de la característica objeto de estudio.
- Técnicas básicas de inferencia estadística: Estimación puntual, estimación por intervalos y contrastes de hipótesis.
- Muestra aleatoria: Modelización de las observaciones.

1.2 Muestra aleatoria

Definición 1.1 *Sea X una variable aleatoria (v.a.) con función de distribución (FdD) F.*

Las v.a. $X_1, ..., X_n$ se dirá que constituyen una muestra aleatoria (m.a.) de tamaño n de X (o de F) si son independientes y están idénticamente distribuidas (i.i.d.) a X.

Proposición 1.1 *Sea $X_1, ..., X_n$ una m.a. de una v.a. X y sea F su FdD.*

La función de distribución asociada a $\underline{X} = (X_1, ..., X_n)$ viene dada por

$$F_{\underline{X}}(\underline{x}) = F_{X_1,...,X_n}(x_1, ..., x_n) = \prod_{i=1}^{n} F(x_i).$$

Si X es absolutamente continua con función de densidad (fdd) f, entonces las fdd asociada a $\underline{X}$ viene dada por

$$f_{\underline{X}}(\underline{x}) = f_{X_1,...,X_n}(x_1, ..., x_n) = \prod_{i=1}^{n} f(x_i).$$

Si X es discreta con función de probabilidad $P(\cdot)$, la función de probabilidad asociada a $\underline{X}$ viene dada por

$$\Pr(\underline{X} = \underline{x}) = \Pr(X_i = x_i,\ i = 1, ..., n) = \prod_{i=1}^{n} P(X = x_i).$$

1.3 Estadísticos muestrales. Propiedades

Definición 1.2 *Sea $X_1, ..., X_n$ una m.a.de una v.a. X y sea F su FdD.*

Un estadístico es una función de la muestra aleatoria $T = T(\underline{X})$ siendo $T : \mathbb{R}^n \longrightarrow \mathbb{R}^k$ una función medible.

Algunos estadísticos de especial interés:

- Media muestral
$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$
- Momento muestral de orden k
$$m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$
- Varianza muestral
$$S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = m_2 - \overline{X}^2$$
- Cuasivarianza muestral
$$S_c^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{n}{n-1}S^2$$
- Función de distribución empírica
$$F_n^*(x) = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i(x)$$
con $\varepsilon_i(x) = 1$ si $X_i \leq x$ y $\varepsilon_i(x) = 0$ en otro caso.

Teorema 1.1 *(Propiedades de la media y de la cuasivarianza muestral)*

Supuesto que existen $E(X) = \mu$ y $var(X) = \sigma^2$, se verifica

1. $E\left(\overline{X}\right) = \mu$ *y* $var\left(\overline{X}\right) = \sigma^2/n$
2. $E\left(S^2\right) = \dfrac{n-1}{n}\sigma^2$ *y* $E\left(S_c^2\right) = \sigma^2$
3. $\overline{X}_n \xrightarrow{c.s.} \mu$
4. $\dfrac{\overline{X}_n - \mu}{\sigma}\sqrt{n} \xrightarrow{\mathcal{L}} Z \sim N(0,1)$

Teorema 1.2 *(Propiedades de la función de distribución empírica)*

1. $\varepsilon_1(x), ..., \varepsilon_n(x)$ *forman una muestra aleatoria de una distribución* $Be(F(x))$

2. $nF_n^*(x) \sim B(n, F(x))$
3. $E(F_n^*(x)) = F(x)$
4. $var(F_n^*(x)) = F(x)(1 - F(x))/n$
5. $F_n^*(x) \xrightarrow{c.s.} F(x)$
6. $\dfrac{F_n^*(x) - F(x)}{\sqrt{F(x)(1 - F(x))}}\sqrt{n} \xrightarrow{\mathcal{L}} Z \sim N(0,1)$

Teorema 1.3 *Teorema de Glivenko-Cantelli*

$$\forall \varepsilon > 0, \qquad \lim_{n \longrightarrow \infty} P\left(\sup_x |F_n^*(x) - F(x)| > \varepsilon\right) = 0.$$

Tema 3

Estimación puntual paramétrica. Propiedades de los estimadores. Métodos de estimación

Contenido

3.1 Introducción

Sea X la variable aleatoria que modeliza a variabilidad de la característica objeto de interés sobre la población. Sea F la función de distribución (FdF) de X

En el marco de la Inferencia Paramétrica se supone que la distribución de X está perfectamente determinada excepto un número finito de parámetros (θ)

$$F \in \left\{ F_\theta; \ \ \theta \in \Theta \subseteq \mathbb{R}^k \right\}.$$

θ se considera un valor fijo desconocido. A Θ se le denomina espacio paramétrico.

El objetivo básico que se aborda en el tema es la estimación de $h(\theta)$ con $h : \Theta \longrightarrow \mathbb{R}$.

La información disponible para realizar la estimación es la suministrada por una muestra aleatoria (m.a.) de X, $\underline{X} = (X_1, ..., X_n)$.

Se denomina espacio muestral, y se denotará por Ξ, al conjunto de todas las posibles realizaciones muestrales.

Definición 3.1 *Un estimador de $h(\theta)$ es un estadístico muestral, $T = T(\underline{X})$ con*

$$T : \Xi \to h(\Theta).$$

Nota 3.1 *T no puede depender de parámetros desconocidos.*

Nota 3.2 *La definición anterior se suele relajar considerando estimador cualquier estadístico muestral.*

3.2 Propiedades de los estimadores

En lo que sigue se supone que $\underline{X} = (X_1, ..., X_n)$ es una m.a de una v.a. X con FdD $F \in \{F_\theta, \theta \in \Theta\}$, $h : \Theta \longrightarrow \mathbb{R}$, la función paramétrica de interés y $T = T(\underline{X})$ un estimador de $h(\theta)$.

Definición 3.2 *Se define el error de estimación de T respecto de $h(\theta)$ como*

$$T - h(\theta).$$

El error de estimación es una variable aleatoria.

Definición 3.3 *Se define el sesgo de T respecto de $h(\theta)$ como*

$$B_\theta(T) = E_\theta(T) - h(\theta).$$

$B_\theta(T)$ es una función de θ.

Definición 3.4 *Se dirá que T es un estimador insesgado de $h(\theta)$ si*

$$B_\theta(T) = 0, \ \ \forall\theta.$$

Es decir, T es un estimador insesgado de $h(\theta)$ si

$$E_\theta(T) = h(\theta), \ \ \forall\theta.$$

Proposición 3.1 *En cualquier distribución:*

a) $\bar{X}$ es un estimador insesgado de la media poblacional μ.

b) S_c^2 es un estimador insesgado de la varianza poblacional σ^2.

c) $F_n^*(x)$ es un estimador insesgado de $F(x)$.

Lema 3.1 *a) La clase de estimadores insesgados de una función paramétrica dada será vacía, contendrá un único estimador o a un número infinito de estimadores.*

b) La insesgadez es una propiedad que se conserva por transformaciones lineales.

Definición 3.5 *Se define el Error Cuadrático Medio (ECM) de T respecto de $h(\theta)$ como*

$$ECM_\theta(T) = E_\theta\left[(T - h(\theta))^2\right].$$

Proposición 3.2

$$ECM_\theta(T) = Var_\theta(T) + [B_\theta(T)]^2.$$

y en el caso particular en que T sea insesgado

$$ECM_\theta(T) = Var_\theta(T).$$

Definición 3.6 *Sean T_1 y T_2 estimadores de $h(\theta)$. Se dirá que T_1 es más eficiente que T_2 si*

$$ECM_\theta(T_1) \leq ECM_\theta(T_2), \forall\theta$$

y

$$\exists\theta^* \ \text{de forma que} \ ECM_{\theta^*}(T_1) < ECM_{\theta^*}(T_2).$$

El ECM es uno de los criterios más importante para evaluar a un estimador y para comparar estimadores. Sin embargo, el ECM no define una relación de orden total en la clase de estimadores.

Proposición 3.3

$$\nexists T^* / \quad ECM_\theta(T^*) \leq ECM_\theta(T), \forall T, \forall\theta.$$

El hecho de que $\nexists T^*$ se debe a que la clase de estimadores es demasido amplia. Por esta razón podría plantearse el problema de determinar el estimador más eficiente en una clase más restringida.

Uno de los criterios para restringir la clase de estimadores es el de insesgadez. En la clase de estimadores insesgados es posible, en ciertas ocasiones, encontrar el estimador más eficiente (estimador insesgado uniformemente de mínima varianza). Este problema se abordará posteriormente.

Definición 3.7 *Sea $T_n = T(X_1, ..., X_n)$ un estimador de $h(\theta)$.*

Se dira que la T_n es débilmente consistente (fuertemente consistente) si

$$\widehat{T}_n \underset{(c.s.)}{\xrightarrow{P}} h(\theta), \ \forall\theta.$$

Proposición 3.4 *En cualquier distribución:*

a) $\bar{X}$ *es un estimador consistente de la media poblacional* μ*.*

b) S_c^2 *es un estimador consistente de la varianza poblacional* σ^2*.*

c) $F_n^*(x)$ *es un estimador consistente de* $F(x)$*.*

Proposición 3.5 *Sea $T_n = T(\underline{X})$ un estimador de $h(\theta)$ para que el que se verifica*

a)

$$\lim_n Var_\theta(T_n) = 0, \forall\theta,$$

b)(insesgadez asintótica)

$$\lim_n E_\theta(T_n) = h(\theta), \forall\theta.$$

Entonces, T_n es débilmente consistente de $h(\theta)$.

3.3 Métodos de estimación

A continuación se estudian dos métodos de estimación: el método de los momentos y el método de máxima verosimilitud.

3.3.1 Método de los momentos

Dada una muestra aleatoria $X_1, ..., X_n$ se denotará por $m_j,\ j = 1, 2, ...$ al momento muestral de orden j:

$$m_j = \frac{1}{n}\sum_{i=1}^{n} X_i^j.$$

Definición 3.8 *Sea X una v.a. con FdD $F \in \{F_\theta : \theta \in \Theta\}$ para la que $\exists E_\theta\left[X^k\right] = \mu_k(\theta)$. Sea $h : \Theta \to \mathbb{R}$ de forma que $h(\theta) = g(\mu_1(\theta), ..., \mu_k(\theta))$ siendo $g : \mathbb{R}^k \to \mathbb{R}$ una función medible.*

En estas condiciones se dirá que $g(m_1, ..., m_k)$ es un estimador por el método de los momentos de $h(\theta)$.

Proposición 3.6 *Si $g : \mathbb{R}^k \to \mathbb{R}$ es una función continua, entonces $g(m_1, ..., m_k)$ es un estimador débilmente consistente de $h(\theta) = g(\mu_1(\theta), ..., \mu_k(\theta))$.*

3.3.2 Método de máxima verosimilitud

En lo que sigue, se representará mediante $f_\theta(x)$ a la función de densidad (fdd) en el caso absolutamente continuo y la función de probabilidad en el caso discreto.

Definición 3.9 *Sea $\underline{x}$ una realización de la muestra aleatoria $\underline{X}$. Se define la función de verosimilitud asociada a $\underline{x}$ como*

$$\begin{array}{rccl} L(\cdot, \underline{x}) : & \Theta & \to & \mathbb{R} \\ & \theta & \to & L(\theta, \underline{x}) = f_\theta(\underline{x}). \end{array}$$

Nota 3.3 *Aunque formalmente la función de densidad conjunta y la función de verosimilitud pueden confundirse son funciones distintas.*

La función de densidad conjunta es una función de $\underline{x}$ y θ se considera fijo.

La función de verosimilitud es una función de θ en la que la realización muestral $\underline{x}$ está fijada.

Definición 3.10 *Se dirá que $\widehat{\theta}(\underline{X})$ es un estimador de máxima verosimilitud (EMV) de θ si*

$$L(\widehat{\theta}(\underline{x}), \underline{x}) = \max_{\theta\in\Theta} L(\theta, \underline{x}).$$

$$\left(\widehat{\theta}(\underline{x}) = arg \max_{\theta\in\Theta} L(\theta, \underline{x})\right)$$

Nota 3.4 *De forma equivalente, $\widehat{\theta}(\underline{X})$ es un estimador de máxima verosimilitud de θ si*

$$\ln L(\widehat{\theta}(\underline{x}), \underline{x}) = \max_{\theta\in\Theta} \ln L(\theta, \underline{x}).$$

En el método de los momentos se planteó el problema de estimar en general una función $h(\theta)$. En el método de máxima verosimilitud sólo se ha planteado la estimación de θ. ¿Qué se entiende por estimador de máxima verosmilitud de $h(\theta)$?

Definición 3.11 *Sea*

$$\begin{array}{rccc} h: & \Theta & \to & \Lambda & /\Lambda \subseteq \mathbb{R}^p \\ & \theta & \to & h(\theta) & \end{array}$$

Se define la función de verosimilitud inducida por h *como*

$$M(\lambda, \underline{x}) = \max_{\theta/h(\theta)=\lambda} L(\theta, \underline{x}), \;\; \lambda \in \Lambda.$$

Se dirá que $T(\underline{X})$ *es un EMV de* $h(\theta)$ *si*

$$\max_{\lambda \in \Lambda} M(\lambda, \underline{x}) = M(T(\underline{x}), \underline{x}).$$

Teorema 3.1 *(Teorema de Zenha)*

Si $\widehat{\theta}_{MV}$ *es un EMV de* θ *entonces* $h(\widehat{\theta}_{MV})$ *es un EMV de* $h(\theta)$.

Teorema 3.2 *Bajo condiciones de regularidad se tiene que si el EMV* $\widehat{\theta}_n$ *es único, entonces*

a) $\widehat{\theta}_n$ *es consistente.*

b) $\widehat{\theta}_n$ *es asintóticamente normal:*

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} N(0, 1/I(\theta)),$$

con $I(\theta)$ *la cantidad de información de Fisher*

$$I(\theta) = E_\theta \left[\left(\frac{\partial}{\partial \theta} \log f(X;\theta) \right)^2 \right].$$

3.4 Suficiencia

La muestra aleatoria proporciona información sobre el parámetro desconocido θ debido a que su distribución (su variabilidad) está caracterizada por él.

Un estadístico $T(\underline{X})$ puede considerarse como un resumen de la muestra y por tanto el estadístico proporciona un resumen de la información contenida en la muestra (su distribución también está determinada por θ).

¿Cómo es el resumen de la información que proporciona el estadístico?, ¿es un resumen adecuado para nuestros propósitos?, ¿es posible resumir la muestra sin perder información relevante sobre el parámetro?

Un estadístico es suficiente si resume la información de la muestra sin perder información relevante sobre el parámetro.

Definición 3.12 *Sea $X_1, ..., X_n$ una m.a de una v.a. X con FdD $F \in \{F_\theta, \theta \in \Theta\}$ y sea $S = S(X_1, ..., X_n)$ un estadístico muestral.*

Se dirá que S es suficiente si las distribuciones condicionadas $(X_1, ..., X_n)/S = s$ no dependen de θ (excepto a los sumo para valores $s \in A$, siendo $P_\theta(S \in A) = 0$, $\forall\theta$).

Teorema 3.3 *(Criterio de factorización de Fisher-Neyman)*

S es suficiente sii

$$f_\theta(\underline{x}) = g(S(\underline{x}), \theta) h(\underline{x}), \quad \forall \underline{x}.$$

Nota 3.5

- *Las dimensiones de S y θ no siempre son iguales. En caso de que se sean iguales, no se puede afirmar nada sobre la suficiencia componente a componente.*
- *La muestra siempre es un estadístico suficiente (pero no hace "resumen").*
- *Si S es suficiente y T es otro estadístico, entonces las distribuciones condicionadas $T/S = s$ no dependen de θ y por tanto $E(T/S)$ no depende de θ.*

Proposición 3.7

1. *$\sum_{i=1}^n T(X_i)$ es un estadístico suficiente en la Familia Exponencial Uniparamétrica.*
2. *$\left(\sum_{i=1}^n T_1(X_i), ..., \sum_{i=1}^n T_k(X_i)\right)$ es un estadístico suficiente en la Familia Exponencil k-paramétrica.*

3.5 Estimación insesgada de mínima varianza

Dada una función paramétrica $h : \Theta \to R$, y considerando como criterio de comparación de estimadores el ECM, se tiene que

$$\nexists T^*/ECM_\theta(T^*) \leq ECM_\theta(T) \ \ \forall T, \ \ \forall\theta.$$

Una de las razones por las que no existe dicho T^* es, sin duda, que la clase de estimadores es demasiado amplia ya que no se ha exigido ningún requisito. Por esta razón cabe preguntarse si para una determinada clase C es posible encontrar un estimador T^* verificando la condición anterior $\forall T \in C$. En concreto vamos a plantear este problema en el caso de que C sea la clase de los estimadores insesgados de $h(\theta)$.

Definición 3.13 *Sea $h : \Theta \to R$. Se dirá que $h(\theta)$ es una función estimable si $\exists T = T(\underline{X})$ tal que*

$$E_\theta[T] = h(\theta), \ \ \forall\theta.$$

Es decir, $h(\theta)$ es una función estimable si existe un estimador insesgado de $h(\theta)$.

Consideremos la clase de todos los estimadores, T, insesgados de $h(\theta)$ y con varianza finita,

$$\mathcal{U}(h(\theta)) = \{T(\underline{X}) : \quad E[T] = h(\theta), \quad Var_\theta[T] < +\infty, \quad \forall\theta \in \Theta\}.$$

Sea $\mathcal{U}_0$ a la clase de estimadores insesgados de cero (función idénticamente cero) con varianza finita,

$$\mathcal{U}_0 = \{\nu(\underline{X}) : \quad E_\theta[\nu] = 0, \quad Var_\theta[\nu] < +\infty, \quad \forall \theta \in \Theta\}.$$

Definición 3.14 *Un estimador $T^* \in \mathcal{U}(h(\theta))$ se dirá que es un estimador insesgado uniformemente de míma varianza (UMVUE) de $h(\theta)$, si*

$$Var_\theta[T^*] \leq Var_\theta[T], \quad \forall \theta \in \Theta, \quad \forall T \in \mathcal{U}(h(\theta)).$$

Teorema 3.4 *(Caracterización del UMVUE)*
Un estimador $T \in \mathcal{U}(h(\theta))$ es un UMVUE de $h(\theta)$ si y sólo si

$$E_\theta[\nu T] = 0, \qquad \forall \theta \in \Theta, \quad \forall \nu \in \mathcal{U}_0.$$

Teorema 3.5 *(Unicidad del UMVUE)*
Si existe el UMVUE de una función $h(\theta)$, éste es único.

Corolario 3.1 *(Del teorema de caracterización)*

Si T_i^* $i = 1, ..., k$ es UMVUE de $h_i(\theta) \Rightarrow \sum_{i=1}^k T_I^*$ es UMVUE de $\sum_{i=1}^k h_i(\theta)$.

Teorema 3.6 *Si T^* es el UMVUE de $h(\theta)$, T^* es una función simétrica de $X_1, ..., X_n$*

3.5.1 Construcción del UMVUE

Teorema 3.7 *(Teorema de Rao-Blackwell)*

Sea S un estadístico suficiente para θ y $T \in \mathcal{U}(h(\theta))$. Entonces $T^* = E[T/S]$ verifica:

1. $E_\theta[T^*] = h(\theta), \forall \theta \in \Theta.$
2. $Var_\theta[T^*] \leq Var_\theta[T], \quad \forall \theta \in \Theta.$

   Además,

   $Var_\theta[T^*] = Var_\theta[T]$, $\forall \theta$ *si y sólo si* $T = E[T/S]$ *(c.s).*

Nota 3.6 *El teorema de Blackwell-Rao afirma que dado $T \in U(h(\theta))$ y S suficiente entonces $E(T/S) \in U(h(\theta))$ y tiene menor varianza que T. Si a $E(T/S)$ se le aplica nuevamente el teorema se obtendrí $E(T/S)$. Es decir, el teorema sólo puede aplicarse como procedimiento de mejora en un paso, no como un procedimiento iterativo.*

Nota 3.7 *Si $T_1, T_2 \in U(h(\theta)) \Rightarrow E(T_i/S)$ mejora a T_i , $i = 1, 2$, en el sentido de la eficiencia, pero no sabemos nada sobre la relación en términos de eficiencia de $E(T_1/S)$ y $E(T_2/S)$.*

Definición 3.15 *Una familia de distribuciones, $\{F(x; \theta), \theta \in \Theta\}$, se dice completa, si*

$$E_\theta[U(X)] = 0, \quad \forall \theta \qquad \Rightarrow \quad P_\theta[U(X) = 0] = 1, \quad \forall \theta.$$

Definición 3.16 *Un estadístico T es completo, si la familia de distribuciones asociada a T es completa.*

Nótese que si T es completo y $g_1(T)$, $g_2(T)$ son funciones medibles de T verificando

$$E_\theta(g_1(T)) = E_\theta(g_2(T)), \quad \forall\theta,$$

entonces

$$g_1(T) = g_2(T) \quad (c.s).$$

Teorema 3.8 *En la familia exponencial natural de rango total:*

1. $\sum_{i=1}^n T(X_i)$ *es un estadístico suficiente y completo (caso uniparamétrico).*
2. $(\sum_{i=1}^n T_1(X_i), ..., \sum_{i=1}^n T_k(X_i))$ *es un estadístico suficiente y completo (caso k-paramétrico).*

Teorema 3.9 *(Teorema Lehmann-Scheffé)*

Sea S un estadístico suficiente y completo para θ y $T \in \mathcal{U}(h(\theta))$. Entonces $T^* = E[T/S]$ es el (único) UMVUE de $h(\theta)$.

Corolario 3.2 *Si S es suficiente y completo y $g(S) \in U(h(\theta)) \Rightarrow g(S)$ es el UMVUE de $h(\theta)$.*

Nota 3.8 *Si se dispone de un estadístico suficiente y completo S:*

- *Si se conoce $g(S) \in U(h(\theta))$, $g(S)$ es el UMVUE de $h(\theta)$.*
- *Si no se conoce $g(S) \in U(h(\theta))$, pero se dispone de $T \in U(h(\theta))$, entonces se puede construir el UMVUE haciendo uso del Teorema de Blackwell-Rao, $E(T/S)$ es el UMVUE de $h(\theta)$.*

Nota 3.9

1. *Puede existir el UMVUE aunque no exista un estadístico suficiente y completo.*
2. *El UMVUE Puede proporcionar estimaciones absurdas (fuera del espacio paramétrico).*
3. *Pueden existir estimadores que no sean insesgado pero con menor ECM que el UMVUE.*

Tema 4

Regiones de Confianza

Contenido

4.1 Regiones de confianza

En el contexto paramétrico, responder a la cuestión ¿cuánto vale $h(\theta)$? consiste básicamente en la selección de un estimador $T(\underline{X})$. Cada realización muestral $\underline{x}$ determina un punto $T(\underline{x})$ en el espacio $h(\Theta)$.

Desde la perspectiva de las regiones de confianza el problema consiste básicamente en la determinación de un estadístico $R(\underline{X})$. Cada realización muestral $\underline{x}$ determina un subconjunto $R(\underline{x})$ del espacio $h(\Theta)$.

Definición 4.1 *Sea X una v.a. con función de distribución $F \in \{F_\theta : \theta \in \Theta\}$ y $\underline{X} = (X_1, ...X_n)$ una m.a. de X.*

Se dirá que $S(\underline{X})$ es una familia de conjuntos aleatorios para θ si $S(\underline{x}) \subseteq \Theta$, $\forall \underline{x} \in \Xi$.

Definición 4.2 *Una familia de conjuntos aleatorios $S(\underline{X})$ se dirá que es una región de confianza para θ al nivel $1-\alpha$ $(0 < \alpha < 1)$, y se denotará $S(\underline{X}) = RC(\theta, 1-\alpha)$, si*

$$P_\theta \left[S(\underline{X}) \ni \theta \right] \geq 1 - \alpha, \ \forall \theta \in \Theta.$$

Se denomina coeficiente de confianza a

$$\inf_\theta P_\theta \left[S(\underline{X}) \ni \theta \right].$$

En particular, si $\Theta \subseteq \mathbb{R}$ y $S(\underline{X}) = [T_1(\underline{X}), T_2(\underline{X})]$, con $T_1(\underline{X}) \leq T_2(\underline{X})$, se dirá que $S(\underline{X})$ es un intervalo de confianza y se denotará

$$S(\underline{X}) = IC(\theta, 1-\alpha).$$

Si $S(\underline{X})$ es de la forma $(-\infty, T(\underline{X})]$ o $[T(\underline{X}), +\infty)$ se le denomina cota superior o inferior de confianza, respectivamente.

Nota 4.1 *Interpretación frecuentista.*

En un muestreo repetido, $1-\alpha$ representa la fracción del número de intervalos generados que contienen al parámetro θ: "si se seleccionan 100 muestras, aproximadamente $100(1-\alpha)$ de los intervalos correspondientes contendrán al verdadero valor del parámetro".

4.2 Métodos de construcción

4.2.1 Método de la cantidad pivotal

Definición 4.3 *Sea $X_1, ...X_n$ una m.a. de una v.a. X con función de distribución $F \in \{F_\theta : \theta \in \Theta \subseteq \mathbb{R}\}$. Sea $T(\underline{X}, \theta)$ una función real. Se dirá que $T(\underline{X}, \theta)$ es un pivot o cantidad pivotal para θ si*

a) $T(\underline{X}, \theta)$ es una función medible para cada valor de θ.

b) La distribución de $T(\underline{X}, \theta)$ no depende de θ.

Construcción de una región de confianza a partir de un pivot

Sea $T(\underline{X}, \theta)$ un pivot. Dado $\alpha \in (0, 1)$, existen $\lambda_1(\alpha)$ y $\lambda_2(\alpha)$, que no dependen de θ, de forma que:

$$P_\theta [\lambda_1(\alpha) \leq T(\underline{X}, \theta) \leq \lambda_2(\alpha)] \geq 1 - \alpha, \ \forall \theta.$$

Dado $\underline{x}$, considérese el subconjunto del espacio paramétrico:

$$R(\underline{x}, \lambda_1(\alpha), \lambda_2(\alpha)) = \{\theta \in \Theta / \lambda_1(\alpha) \leq T(\underline{x}, \theta) \leq \lambda_2(\alpha)\}.$$

Se verifica que $R(\underline{X}, \lambda_1(\alpha), \lambda_2(\alpha))$ es una región de confianza para θ al nivel $1 - \alpha$.

Nota 4.2 *Si $T(\underline{x}, \theta)$ es monótona en θ, $R(\underline{x}, \lambda_1(\alpha), \lambda_2(\alpha))$ será un intervalo.*

Nota 4.3 *Si $T(\underline{x}, \theta)$ sigue una distribución continua es posible encontrar $\lambda_1(\alpha)$ y $\lambda_2(\alpha)$ de forma que*

$$P_\theta [\lambda_1(\alpha) \leq T(\underline{X}, \theta) \leq \lambda_2(\alpha)] = 1 - \alpha, \ \forall \theta.$$

Nota 4.4 *Ya que los valores $\lambda_1(\alpha)$ y $\lambda_2(\alpha)$ no quedan determinados de forma única, y teniendo en cuenta que la información que proporciona el intervalo de confianza es más precisa cuanto menor sea su amplitud, es interesante determinar $\lambda_1^*(\alpha)$ y $\lambda_2^*(\alpha)$ de forma que se minimice la amplitud del intervalo o bien el valor esperado de la amplitud.*

La amplitud del intervalo suele ser también una función decreciente del tamaño muestral n. Un incremento de n lleva asociado un incremento de la precisión del intervalo.

4.2.2 Intervalos de confianza basados en la desigualdad de Tchebychev

Proposición 4.1 *Sea $T = T(\underline{X})$ un estimador insesgado de $h(\theta)$ y $\sigma^* \geq \sup_\theta \sqrt{Var_\theta(T)}$.*

Entonces $[T - k\sigma^*, T + k\sigma^*]$ es un intervalo de confianza al nivel $\left(1 - \frac{1}{k^2}\right)$ para $h(\theta)$.

4.2.3 Intervalos de confianza asintóticos

Definición 4.4 *Sean $T_1 = T_1(\underline{X})$ y $T_2 = T_2(\underline{X})$ dos estadísticos muestrales, $T_1 \leq T_2$. Se dirá que $[T_1, T_2]$ es un intervalo de confianza asintótico para θ al nivel $1 - \alpha$ si*

$$\lim_{n\to\infty} P_\theta\left[\theta \in [T_1, T_2]\right] \geq 1 - \alpha, \ \forall \theta.$$

Proposición 4.2 *Sean $T_n = T(\underline{X})$ y $V_n = V(\underline{X})$, verificando*

a)

$$T_n \xrightarrow{\mathcal{L}} N(\theta, V(\theta))$$

b)

$$V_n \xrightarrow{P} V(\theta).$$

Entonces

$$[T_n - Z_{1-\alpha/2}\sqrt{V_n}\ ,\ T_n + Z_{1-\alpha/2}\sqrt{V_n}]$$

es un intervalo de confianza asintótico al nivel $(1 - \alpha)$ para θ.

4.3 Estimación del tamaño muestral

Sea $T_n = T(X_1, ..., X_n)$ un estimador de $h(\theta)$. ¿Qué tamaño muestral n es necesario para que

$$P_\theta\left[|T_n - h(\theta)| \leq \delta\right] \geq 1 - \alpha, \ \forall\, \theta\ ?$$

El problema planteado es equivalente a determinar el tamaño n de forma que

$$[T_n - \delta, T_n + \delta] = IC(h(\theta), 1 - \alpha).$$

Tema 5

Contrastes de hipótesis paramétricas

Contenido

5.1 Introducción a los contrastes de hipótesis. Conceptos básicos

Muchas actividades en la Ciencia, en la Industria y en la vida cotidiana, se desarrollan respondiendo sí o no a cuestiones importantes planteadas en términos de hipótesis.

El objetivo de un contraste de hipótesis es el de decidir, a partir de los resultados muestrales obtenidos, sobre la veracidad o falsedad de una hipótesis realizada.

Definición 5.1 *Una hipótesis estadística (H) es una afirmación o conjetura sobre las distribuciones de una o más poblaciones o sobre alguna característica de las mismas.*

Una vez considerada una hipótesis, un test o contraste de hipótesis no es más que una regla de decisión para decidir a la vista de los resultados experimentales $\mathbf{x} = (x_1, ..., x_n)$ si rechazar la hipótesis (hay evidencia para rechazarla a partir de los datos experimentales, consideramos que los datos son discrepantes con la hipótesis) o no rechazarla (no hay evidencia para rechazarla).

Es decir un test de hipótesis no es más que una regla de decisión que especifica la decisión a adoptar (rechazar o no rechazar la hipótesis) para cada posible valor del espacio muestral $\underline{x} \in \Xi$. Es decir, hemos de decidir cuando consideramos que los datos datos experimentales discrepan o no de la hipótesis.

En otras palabras, lo que se pretende es dividir el espacio muestral Ξ en dos subconjuntos excluyentes, R_C y R_A, de modo que si $\mathbf{x} \in R_C$ rechazamos H y en caso contrario no la rechazamos

$$\begin{aligned} R_C &= \{\ \mathbf{x}/ \text{ Rechazo } H \text{ si la realización muestral es } \mathbf{x}\} \\ R_A &= \{\ \mathbf{x}/ \text{ No rechazo } H \text{ si la realización muestral es } \mathbf{x}\} \end{aligned}$$

A R_c se denomina región crítica y a R_A región de aceptación.

Hipótesis paramétricas - Hipótesis no paramétricas

Cuando el problema de inferencia se aborde en el contexto paramétrico, suponemos que la F.d.D. de X pertenece a una familia paramétrica $F \in \{F_\theta,\ \theta \in \Theta\}$, las hipótesis serán del tipo $H : \theta \in \Theta_0 \subset \Theta$, y se dirá que es una hipótesis paramétrica. En el caso particular en que $\Theta_0 = \{\theta_0\}$ se dirá que es una hipótesis simple, y en otro caso que es una hipótesis compuesta.

Cuando la inferencia no se aborde en el contexto paramétrico, se dirá que las hipótesis son no paramétricas.

Hipótesis nula - Hipótesis alternativa

En los problemas de contrastes de hipótesis se considerarán dos hipótesis: H_0 y H_1 (en general la complementaria de H_0) que se denominan hipótesis nula e hipótesis alternativa, respectivamente. El papel que representa cada una de estas hipótesis se comentará más adelante.

En el caso paramétrico se tendrá

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

Un contraste de hipótesis será un criterio para decidir si rechazar o no H_0.

Por tanto

$$R_C = \{ \ \mathbf{x}/ \text{ Rechazo } H_0 \text{ si la realización muestral es } \ \mathbf{x}\}$$

y

$$R_A = \{ \ \mathbf{x}/ \text{ No rechazo } H_0 \text{ si la realización muestral es } \ \mathbf{x}\} .$$

Error de Tipo I - Error de Tipo II

Se han de definir algunos criterios en base a los cuales evaluar las reglas de decisión y que nos permitan seleccionar entre varias opciones posibles.

Analicemos las distintas situaciones que se pueden presentar dependiendo de la decisión que se adopte y lo que realmente ocurre.

Formalmente, al realizar un contraste de hipótesis, pueden darse las siguientes circunstancias:

		Realidad	
		H_0 cierta	H_0 falsa
Decisión	Rechazar H_0	Error Tipo I	Decisión Correcta
	No rechazar H_0	Decisión Correcta	Error Tipo II

Definición 5.2 *Se denomina Error de Tipo I al error que se comete si se rechaza H_0 siendo cierta.*

Se denomina Error de Tipo II al error que se comete si se acepta H_0 siendo falsa.

Así pues, cualquiera que sea la decisión que se adopte ésta podría ser errónea.

Parece razonable cuantificar probabilísticamente los errores

$$\begin{aligned} P\,(\text{ETI}) &= P_{H_0}\,(\text{Rechazar H}_0) \\ P\,(\text{ETII}) &= P_{H_1}\,(\text{No rechazar H}_0) \end{aligned}$$

e intentar que sean mínimos.

5.2 Función test y función potencia

En este apartado formalizamos los conceptos introducidos en la sección anterior.

Analizamos a continuación la teoría de los contraste de hipótesis paramétricos para la búsqueda de test óptimos. Esta teoría fue introducida por Neyman y Pearson (1928, 1933) y ha sido extendida y generalizada.

Definición 5.3 *Sea X una variable aleatoria con función de distribución $F \in \{F_\theta,\ \theta \in \Theta\}$ y consideremos las hipótesis*

$$H_0 : \theta \in \Theta_0 \subset \Theta$$
$$H_1 : \theta \in \Theta \setminus \Theta_0$$

Sea $\mathbf{X} = (X_1, ..., X_n)$ una muestra aleatoria de X y sea $\varphi(\mathbf{X})$ un estadístico muestral.

Se dirá que $\varphi(\mathbf{X})$ es una función test si $\varphi(\mathbf{x}) \in [0,1] \quad \forall \mathbf{x}$, siendo $\varphi(\mathbf{x})$ la probabilidad de rechazar H_0 cuando la realización muestral es $\mathbf{x}$.

Se dirá que $\varphi(\mathbf{X})$ es un test no aleatorizado si $\varphi(\mathbf{x}) \in \{0,1\} \quad \forall \mathbf{x}$.

Nota 5.1 *Si $\varphi(\mathbf{X})$ es un test no aleatorizado*

$$\varphi(\mathbf{x}) = \begin{cases} 1 & si \quad \mathbf{x} \in A \\ 0 & si \quad \mathbf{x} \notin A \end{cases}$$

Los valores en los que $\varphi(\mathbf{x}) = 1$ son aquellos para los que se rechaza H_0 y los valores en los que $\varphi(\mathbf{x}) = 0$ son aquellos para los que no se rechaza H_0. Por tanto A es la región crítica del test.

Un test no aleatorizado se puede expresar como $\varphi(\mathbf{x}) = I_{R_c}(\mathbf{x})$.

En la práctica es normal utilizar tests no aleatorizados. La definición general enriquece la teoría de los tests de hipótesis.

Definición 5.4 *Sea $\varphi = \varphi(\mathbf{X})$ un test para contrastar las hipótesis*

$$H_0 : \theta \in \Theta_0 \subset \Theta$$
$$H_1 : \theta \in \Theta \setminus \Theta_0$$

1. *Se define la función potencia de φ como*

$$\begin{aligned} \beta_\varphi : \quad \Theta &\longrightarrow \quad [0,1] \\ \theta &\longrightarrow \quad \beta_\varphi(\theta) = E_\theta(\varphi(\mathbf{X})) \end{aligned}$$

2. *Se dirá que el test es de nivel α si*

$$\beta_\varphi(\theta) \leq \alpha, \quad \forall \theta \in \Theta_0$$

3. *Se define el tamaño del test como*

$$\sup_{\theta\in\Theta_0} \beta_\varphi(\theta)$$

Nota 5.2 *Interpretemos el significado de la función potencia sobre un test no aleatorizado.*

1. $\beta_\varphi(\theta) = P_\theta\left(Rechazar\ H_0\right).$
2. *Si consideramos* $\theta \in \Theta_0,\quad \beta_\varphi(\theta) = P_{H_0}\left(Rechazar\ H_0\right)$ *guarda relación con la* $P(ETI)$. *Para un test de nivel* α *se tiene que* $P\left(ETI\right) \leq \alpha$.
3. *Si consideramos* $\theta \in \Theta - \Theta_0$, $\beta_\varphi(\theta) = 1 - P_{H_1}\left(No\ rechazar\ H_0\right)$ *guarda relación con la P(ETII).*
4. *Si* H_0 *es una hipotesis simple,* $\Theta_0 = \{\theta_0\}$, *el tamaño del test es* $\beta_\varphi(\theta_0)$.

5.3 Optimalidad

Denotemos por ϕ_α la clase de tests de nivel α para contrastar las hipótesis

$$\begin{aligned} H_0 &: \theta \in \Theta_0 \\ H_1 &: \theta \in \Theta - \Theta_0, \end{aligned}$$

es decir

$$\phi_\alpha = \{\varphi \ / \ \beta_\varphi(\theta) \leq \alpha \ \ \forall\theta \in \Theta_0\}.$$

El objetivo que nos plantameos, a continuación, es, fijado un nivel de significación α, $0 < \alpha < 1$, encontrar un test $\varphi \in \phi_\alpha$ que cumpla unas determinadas condiciones de optimalidad.

Definición 5.5 *Un test* $\varphi \in \phi_\alpha$ *diremos que es un test uniformemente de máxima potencia (U.M.P.) si*

$$\beta_\varphi(\theta) \geq \beta_{\varphi^*}(\theta), \quad \forall\theta \in \Theta - \Theta_0,\ \forall\varphi^* \in \phi_\alpha.$$

Nota 5.3 *El test UMP de nivel* α *es el de menor P(ETII) entre todos los que tienen P(ETI)* $\leq \alpha$.

Es decir, limitamos P(ETI) y buscamos el que tiene menor P(ET(II). Vemos como la consideración de los dos tipos de error es diferente.

La idea es similar a la del UMVUE: limitamos la clase de estimadores y buscamos el óptimo dentro de la clase.

En general, no se tiene garantizada la existencia de tests U.M.P., sin embargo para determinadas hipótesis y/o distribuciones diferentes resultados proporcionan la construcción de estos tests óptimos.

5.4 Métodos de construcción

A continuación se recogen diferentes procedimientos para la construcción, desde un punto de vista práctico, de tests de hipótesis. Si bien, en general, no se tiene garantizado que sean óptimos, en la mayoría de las situaciones se obtienen tests con buenas propiedades.

5.4.1 Test de la Razón de Verosimilitudes

Sea $X_1, ..., X_n$ una muestra aleatoria de X con F.d.D. $F \in \left\{F_\theta,\ \theta \in \Theta \subseteq R^k\right\}$ y consideremos el contraste

$$\begin{aligned} &H_0 : \theta \in \Theta_0 \\ &H_1 : \theta \in \Theta - \Theta_0. \end{aligned}$$

El estadístico de razón de verosimilitudes asociado a las hipótesis considerada se define como

$$\lambda\left(\mathbf{x}\right) = \frac{\sup\limits_{\theta \in \Theta_0} L(\mathbf{x}, \theta)}{\sup\limits_{\theta \in \Theta} L(\mathbf{x}, \theta)}$$

siendo $L(\mathbf{x}, \theta)$ la función de verosimilitud asociada a la realización muestral $\mathbf{x}$.

Cualquier test de la forma

$$\varphi\left(\mathbf{x}\right) = \begin{cases} 1 & \text{si} \quad \lambda\left(\mathbf{x}\right) \leq c \\ 0 & \qquad c.c. \end{cases}$$

se dirá que es un test de razón de verosimilitudes.

La constante c se determina a partir de la restricción del nivel de significación, es decir:

$$c \ / \ \sup_{\theta \in \Theta_0} P_\theta \left\{\ \lambda\left(\mathbf{X}\right) \leq c\ \right\} \leq \alpha$$

Nota 5.4

$$\lambda\left(\mathbf{x}\right) = \frac{\sup\limits_{\theta \in \Theta_0} L(\mathbf{x}, \theta)}{\sup\limits_{\theta \in \Theta} L(\mathbf{x}, \theta)} = \frac{L(\mathbf{x}, \widehat{\theta}_{0,MV})}{L(\mathbf{x}, \widehat{\theta}_{MV})}$$

siendo $\widehat{\theta}_{MV}$ *el estimador de máxima verosimilitud y* $\widehat{\theta}_{0,MV}$ *el EMV restringido*

$$\begin{aligned} \widehat{\theta}_{MV} &= \arg \max_{\theta \in \Theta} L(\mathbf{x}, \theta) \\ \widehat{\theta}_{0,MV} &= \arg \max_{\theta \in \Theta_0} L(\mathbf{x}, \theta) \end{aligned}$$

$L(\mathbf{x}, \widehat{\theta}_{0,MV})$ *representa la mejor explicación posible bajo* H_0 *a los datos observados.*

$L(\mathbf{x}, \widehat{\theta}_{MV})$ *representa la mejor explicación posible a los datos observados.*

Si $\lambda\left(\mathbf{x}\right)$ *es "muy pequeño" significa que existe un valor de* $\theta \notin \Theta_0$ *bajo el que los datos observados son mucho más probables y por tanto rechazamos* H_0.

Teorema 5.1 *Bajo condiciones de regularidad, la variable aleatoria*

$$-2\ln\lambda(\mathbf{X}) \underset{n\longrightarrow\infty}{\longrightarrow} \chi^2$$

siendo los grados de libertad de la misma la diferencia entre el número de parámetros independientes y desconocidos en Θ y el número bajo H_0.

5.4.2 Relación entre tests y regiones de confianza

A continuación, se recogen algunos resultados que relacionan los tests de hipótesis y las regiones de confianza. Estos resultados proporcionan otro procedimiento para la construcción de tests de hipótesis.

Teorema 5.2 *Sea $\mathbf{X}$=($X_1...X_n$) una muestra aleatoria de X con función de distribución $F \in \{F(x,\theta), \theta \in \Theta\}$. Sea $C(\mathbf{X})$ un conjunto de confianza, al nivel de confianza 1-α, para θ.*

Entonces el test no aleatorizado $\varphi(\mathbf{X}) = \begin{cases} 1 \text{ si } \theta_0 \notin C(\mathbf{X}) \\ 0 \text{ c.c.} \end{cases}$ *es un test, a un nivel de significación α para contrastar $H_0 : \theta = \theta_0$.*

5.4.3 Procedimiento general

Hasta el momento, hemos visto dos procedimientos para la construcción de tests de hipótesis, ambos dentro del contexto paramétrico:el test de razón de verosimilitudes y a partir de una región de confianza.

Desde un punto de vista práctico, para construir un test de hipótesis, tanto en los contrastes paramétricos como no paramétricos, es suficiente con especificar una región critica, es decir: $\{\mathbf{x}$ / Rechazo H_0 *si observo* $\mathbf{x}\}$.

En general, podemos decir que la región crítica de un test no aleatorizado de nivel α será de alguna de las formas siguientes:

a) $T(X_1 \ldots X_n) \leq a_\alpha$

b) $T(X_1 \ldots X_n) \geq b_\alpha$

c) $T(X_1 \ldots X_n) \leq c_\alpha$ ó $T(X_1 \ldots X_n) \geq d_\alpha$

d) $e_\alpha \leq T(X_1 \ldots X_n) \leq f_\alpha$

En general, podemos decir que la construcción de la región crítica, R_c, se hará a partir de un estadístico $T(X_1 \ldots X_n)$ que evalúe la concordancia de la muestra $(X_1 \ldots X_n)$ con lo establecido en H_0, de forma se rechazará H_0 si "la realización muestral no está de acuerdo con lo establecido en H_0 (lo observado es poco probable bajo la hipótesis nula)".

Por estas razones, de forma general, podemos decir que para construir un test de hipótesis hemos de:

1.- Definir $T(X_1 \ldots X_n)$.

2.- Establecer R_c en función de qué representa $T(X_1 \ldots X_n)$ y lo especificado en H_1.

3.- Estudiar la distribución de $T(X_1 \ldots X_n)$ bajo H_0. Esta etapa es necesaria para determinar el nivel o el tamaño del test.

4.- Estudiar la distribución de $T(X_1 \ldots X_n)$ bajo H_1. Esta etapa es de interés para determinar la potencia del test.

5.- Determinar los puntos críticos asociados al nivel de significación fijado.

6.- Cálculo del p-valor