Asignatura	:	INFERENCIA ESTADÍSTICA
Titulación	:	LICENCIADO EN ESTADÍSTICA
Profesor	:	ISABEL MOLINA PERALTA
Capítulo 3	:	ESTIMACIÓN PUNTUAL

3.1. INTRODUCCIÓN

X v.a. con distribución $F(x;\theta)$
$\theta \in \Theta$ parámetro desconocido, Θ espacio paramétrico
$(X_1,\ldots,X_n)$ m.a.s. de X

Estimar: Inferir o intentar averiguar el valor de uno o varios parámetros poblacionales a partir de una muestra.

Estimador ($\widehat{\theta}$): Estadístico con rango contenido en el espacio paramétrico Θ.

Ejemplo 1 En la siguiente tabla aparecen parámetros comunes y sus estimadores habituales

Parámetro	p	μ	σ^2
Estimador	$\widehat{p}=Y/n$	$\bar{X}$	S^2, S'^2

3.2. ESTIMADORES INSESGADOS

Estimador insesgado: $E(\widehat{\theta}) = \theta$

Sesgo: $B(\widehat{\theta}) = E(\widehat{\theta}) - \theta$

Ejemplo 2 Considera una población normal y una m.a.s. de dicha población. ¿Es la varianza muestral S^2 un estimador insesgado de la varianza poblacional? ¿Y la cuasivarianza muestral S'^2?

Ejemplo 3 Sea X una v.a. $U(0, \theta)$, y sea $(X_1, \ldots, X_n)$ una m.a.s. de X. ¿Es $\widehat{\theta} = X_{(n)}$ un estimador insesgado de θ? Si no lo es, ¿cuál es su sesgo? ¿Cómo puedes transformarlo para obtener un estimador insesgado?

Ejemplo 4 Sea $X \sim Exp(1/\theta)$. Encontrar un estimador insesgado de θ.

Nota 1 Si disponemos de dos estimadores insesgados $\widehat{\theta}$ y $\tilde{\theta}$, una combinación convexa de ellos también lo es,

$$E[c\tilde{\theta} + (1-c)\widehat{\theta}] = \theta.$$

Prueba:

3.3. ERROR DE ESTIMACIÓN

Proposición 1 Sean $(X_1, \ldots, X_{n_1})$ e $(Y_1, \ldots, Y_{n_2})$ m.a.s. independientes. Sean $\widehat{\theta}_1$ y $\widehat{\theta}_2$ estimadores centrados de θ, obtenidos de $(X_1, \ldots, X_{n_1})$ e $(Y_1, \ldots, Y_{n_2})$ respectivamente. El estimador $\widehat{\theta}_c = c\widehat{\theta}_1 + (1-c)\widehat{\theta}_2$ con mayor precisión es

$$c = \frac{V[\widehat{\theta}_2]}{V[\widehat{\theta}_2] + V[\widehat{\theta}_1]}.$$

Prueba:

Para comparar estimadores insesgados se utiliza la varianza, o su inverso, la eficiencia.

Eficiencia: La eficiencia (o precisión) de un estimador $\widehat{\theta}$ de θ es

$$Ef(\widehat{\theta}) = 1/V(\widehat{\theta}).$$

Eficiencia relativa: La eficiencia relativa de un estimador $\widehat{\theta}$ respecto de otro $\widehat{\theta}'$ de θ es

$$ER(\widehat{\theta}'/\widehat{\theta}) = \frac{Ef(\widehat{\theta}')}{Ef(\widehat{\theta})} = \frac{V(\widehat{\theta})}{V(\widehat{\theta}')}$$

Ejemplo 5 Sea $(X_1, \ldots, X_n)$ una m.a.s. de una v.a. $N(\mu, \sigma^2)$. La mediana muestral, $\text{Med}(X_1, \ldots, X_n)$, es un estimador insesgado para μ, con varianza asintótica

$$V[\text{Med}(X_1, \ldots, X_n)] = \pi\sigma^2/(2n).$$

Calcular la eficiencia relativa de la Mediana muestral con respecto a la media muestral.

✓ No siempre existe estimador insesgado.

Ejemplo 6 Se nos ofrece el siguiente juego, en el que hay que pagar 6 euros para participar: nos dan 12 euros si obtenemos dos caras en dos lanzamientos de una moneda con probabilidad de cara π, y nos dan 0 euros si no obtenemos el resultado deseado. Se nos permite realizar un lanzamiento previo de la moneda para estimar el valor de la probabilidad de éxito $\theta = \pi^2$.

En el lanzamiento previo se observará el valor de la v.a.

$$X_1 = \begin{cases} 1 & \text{si se obtiene cara,} \\ 0 & \text{si se obtiene cruz.} \end{cases}$$

Sabemos que $X_1 \sim Bern(\pi)$. Sea $\widehat{\theta} = \widehat{\theta}(X_1)$ un estimador de $\theta = \pi^2$. Su esperanza vendría dada por

$$E(\widehat{\theta}) = \widehat{\theta}(1)\pi + \widehat{\theta}(0)(1 - \pi) \neq \pi^2,$$

para cualquier estimador $\widehat{\theta}$. Por tanto, no existe estimador insesgado de π^2 con la muestra dada X_1.

Error absoluto medio: $EAM(\hat{\theta}) = E[|\hat{\theta} - \theta|]$

Error cuadrático medio: $ECM(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$

$$ECM(\hat{\theta}) = B^2(\hat{\theta}) + V(\hat{\theta})$$

Reglas de selección de estimadores:

- Seleccionar el que tenga menor ECM;
- De entre los estimadores con menor sesgo, elegir el que tenga menor varianza.

Ejemplo 7 Para una población normal y una m.a.s. de dicha población, ¿cual de los estimadores S^2 o S'^2 de σ^2 tiene menor ECM?

Nota 2 Cuando n es elevado, es conveniente usar estimadores insesgados, dado que la varianza suele ser pequeña.
Cuando n es pequeño, la varianza puede ser grande comparada con el sesgo, con lo cual se puede conseguir una mejor reducción del ECM disminuyendo la varianza.

3.4. ESTIMADORES INVARIANTES

Ejemplo 8 El fabricante de un producto afirma que los envases tienen al menos θ gramos del producto. Si esto es cierto, la cantidad de producto que contine el envase se distribuye $U(\theta, \theta + 100)$. Para comprobar si se cumple la especificación del fabricante, se toma una m.a.s. $(x_1, \ldots, x_n)$ de contenidos de n envases, en gramos, y se calcula un estimador $\widehat{\theta}(x_1, \ldots, x_n)$. Después se descubre que la balanza que se usaba pesaba sisteméticamente c gramos menos. ¿Podemos simplemente corregir nuestro estimador de la forma $\widehat{\theta}(x_1, \ldots, x_n) + c$?

Estimadores invariantes por traslaciones:

$$\widehat{\theta}(X_1 + c, \ldots, X_n + c) = \widehat{\theta}(X_1, \ldots, X_n) + c, \quad \forall c \in \mathbb{R}.$$

Ejemplo 9 Comprobar que $X_{(1)}$, $\bar{X}$ y $(X_{(1)} + X_{(n)})/2$ son invariantes por traslaciones, pero no lo son la media geométrica $(\prod_{i=1}^n X_i)^{1/n}$ ni la media armónica $n/\sum_{i=1}^n X_i^{-1}$.

Ejemplo 10 Una persona llega a la parada del autobús siempre a la misma hora, y desea estimar su tiempo máximo de espera del autobús. Se sabe que el tiempo de espera está distribuido $U(0,\theta)$. Cronometra su tiempo de espera durante n días, y obtiene una m.a.s. $(x_1,\ldots,x_n)$ expresada en segundos, con la que calcula un estimador del tiempo máximo de espera, $\widehat{\theta}(x_1,\ldots,x_n)$ en segundos. Si lo quiere pasar a minutos, ¿Puede hacer simplemente $\widehat{\theta}(x_1,\ldots,x_n)/60$?

Estimador invariante por cambios de escala:

$$\widehat{\theta}(cX_1,\ldots,cX_n) = c\,\widehat{\theta}(X_1,\ldots,X_n), \quad \forall c \in \mathbb{R}.$$

Ejemplo 11 Comprobar si los estimadores $\bar{X}$, $X_{(n)}$, $\log((1/n)\sum_{i=1}^{n} \exp(X_i))$ y $X_{(n)}/X_{(1)}$ son invariantes por cambios de escala.

3.5. ESTIMADORES CONSISTENTES

3.5.1 Consistencia débil y en media cuadrática.

¿Qué ocurre con un estimador $\widehat{\theta}$ al aumentar el tamaño muestral?¿Se aproxima al verdadero valor θ?

Si la probabilidad de que se aleje del verdadero valor converge a cero, el estimador se llamará *consistente*.

Ejemplo 12 Sea $X \sim \mathcal{N}(\mu, \sigma^2 = 4)$ y sea $(X_1, \ldots, X_n)$ una m.a.s. de X. ¿Cómo varía la distribución de $\bar{X}_n - \mu$ al aumentar n?

n	$\bar{X}_n$	$P\left(\lvert\bar{X}_n - \mu\rvert > 1\right)$	$P\left(\lvert\bar{X}_n - \mu\rvert > 0{,}5\right)$
1			
2			
3			
10			
20			

Obsérvese que $P(|\bar{X}_n - \mu| > \varepsilon)$ converge a cero cuando n converge a infinito, $\forall \varepsilon > 0$. Equivalentemente, la probabilidad $P(|\bar{X}_n - \mu| \leq \varepsilon)$ converge a uno.

La distribución del estimador $\bar{X}_n$ se va concentrando alrededor de μ al aumentar n.

Consistencia (débil): Sea X una v.a. con función de probabilidad inducida $P(\cdot\,;\theta)$ y sea $(X_1,\ldots,X_n)$ una m.a.s. de X. El estimador $\widehat{\theta}_n = \widehat{\theta}_n(X_1,\ldots,X_n)$ es *consistente* para θ, si

$$\forall \epsilon > 0, \qquad \lim_{n\to\infty} P\{|\widehat{\theta}_n - \theta| > \epsilon\} = 0.$$

Ejemplo 13 Sea $X \sim U(0,\theta)$, y sea $(X_1,\ldots,X_n)$ una m.a.s. de X. Demostrar que el estimador $\widehat{\theta}_n = X_{(n)}$ es consistente para θ.

Convergencia en probabilidad: Una sucesión $\{X_n,\ n \in \mathbb{N}\}$ de variables aleatorias definidas sobre el mismo espacio de probabilidad $(\Omega, \mathcal{A}, P)$, se dice que *converge en probabilidad* a la v.a. X ($X_n \xrightarrow{P} X$) si se verifica

$$\forall \epsilon > 0, \quad \lim_{n\to\infty} P\{|X_n - X| > \epsilon\} = 0.$$

La siguiente definición de consistencia es más fuerte que la anterior.

Consistencia en media cuadrática: Un estimador $\hat{\theta}_n$ es *consistente en media cuadrática* para θ si verifica

$$\lim_{n\to\infty} ECM(\hat{\theta}_n) = 0,$$

o, equivalentemente, si

$$\lim_{n\to\infty} B(\hat{\theta}_n) = 0 \quad \text{y} \quad \lim_{n\to\infty} V(\hat{\theta}_n) = 0.$$

Ejemplo 14 En la situación del Ejemplo 13, ¿es $\widehat{\theta}_n = X_{(n)}$ consistente en media cuadrática para θ?

Proposición 2 Un estimador $\widehat{\theta}_n$ de θ insesgado para todo $n \in \mathbb{N}$ es consistente en media cuadrática para θ si y solo si

$$\lim_{n \to \infty} V(\widehat{\theta}_n) = 0.$$

3.5.2 Desigualdades útiles para acotar probabilidades

Teorema 1 (Desigualdad de Markov)
Para cualquier variable aleatoria X, se verifica

$$P(|X| > k) \leq \frac{E[X^2]}{k^2}, \quad k \geq 0.$$

Prueba:

Teorema 2 (Desigualdad de Tchebyshev)
Sea X una v.a. con $E(X) = \mu$ y $V(X) = \sigma^2$. Entonces,

$$P\left(|X - \mu| \geq k\sigma\right) \leq \frac{1}{k^2}, \quad k \geq 0.$$

Prueba:

Ejemplo 15 Sea $\widehat{\theta}$ un estimador insesgado de θ con distribución desconocida. ¿Cuál es la probabilidad de que $\widehat{\theta}$ se diferencie de θ en menos de 2 desviaciones típicas $\sigma_{\widehat{\theta}}$? ¿y la probabilidad de que se diferencie en menos de 3 desviaciones típicas? Si supiésemos que la distribución de $\widehat{\theta}$ es una $N(\theta, \sigma^2_{\widehat{\theta}})$, ¿qué valor tomarían las probabilidades anteriores? Rellena la tabla siguiente

k	Tchebyshev	Exacto
2		
3		

Ejemplo 16 La experiencia ha determinado que el tiempo X en minutos que lleva una inspección periódica de mantenimiento de un dictáfono tiene una distribución $\gamma(\alpha = 3, \beta = 2)$. Supóngase que un nuevo reparador necesita 19 minutos para inspeccionar la máquina ¿Concuerda este tiempo con la experiencia previa?

Teorema 3 La consistencia en media cuadrática de un estimador $\widehat{\theta}_n$ al parámetro θ implica la consistencia (débil) de $\widehat{\theta}_n$ a θ.

Prueba:

Condiciones para consistencia en m.c.:

✓ $\lim_{n \to \infty} E(\widehat{\theta}_n) = \theta$, ✓ $\lim_{n \to \infty} V(\widehat{\theta}_n) = 0$.

Ejemplo 17 Sea $(X_1, \ldots, X_n)$ una m.a.s. de una v.a. con media μ y varianza σ^2. Considérense los siguientes estimadores de μ,

$$\hat{\mu}_1 = \frac{X_1 + X_2}{2}, \quad \hat{\mu}_2 = \frac{X_1}{4} + \frac{X_2 + \ldots + X_{n-1}}{2(n-2)} + \frac{X_n}{4}, \quad \hat{\mu}_3 = \bar{X}.$$

(a) ¿Cuál de los tres es insesgado?

(b) ¿Cuál de los tres es consistente para μ?

3.5.3 Ley de los Grandes Números

Tomando muchas observaciones independientes y promediando, podemos acercarnos arbitrariamente al valor medio real μ.

Teorema 4 (Ley de los Grandes Números)
La media muestral $\bar{X}$ de una m.a.s. $(X_1, \ldots, X_n)$ de una v.a. X con media μ y varianza $\sigma^2 < \infty$ es un estimador consistente de μ:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \overset{m.c.}{\longrightarrow} E(X).$$

Prueba:

Ejemplo 18 Sea $Y \sim Bin(n, \pi)$. ¿Es $\widehat{\pi} = Y/n$ un estimador consistente de π?

Ejemplo 19 Para una v.a. X de la que se ha tomado una m.a.s. $(X_1, \ldots, X_n)$, ¿Es el momento muestral k-ésimo $a_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ un estimador consistente del momento poblacional k-ésimo $\alpha_k = E(X^k)$?

Cualquier transformación continua de un estimador consistente, es consistente para la misma transformación del parámetro.

3.5.4 Álgebra de la consistencia

Teorema 5 Se verifican las siguientes propiedades:

(a) Sea $\widehat{\theta}_n$ un estimador consistente de θ. Sea $g(x)$ una función continua en $x = \theta$. Entonces, $g(\widehat{\theta}_n)$ es consistente para $g(\theta)$.

(b) Sean $\widehat{\theta}_n$ e $\widehat{\theta}'_n$ estimadores consistentes de θ y θ' respectivamente. Sea $g(x, y)$ una función continua en el punto $(x, y) = (\theta, \theta')$. Entonces, $g(\widehat{\theta}_n, \widehat{\theta}'_n)$ es consistente para $g(\theta, \theta')$.

Prueba:

Corolario 1 (Álgebra de la consistencia)
Supóngase que $\widehat{\theta}_n$ converge en probabilidad a θ y que $\widehat{\theta}'$ converge en probabilidad a θ'. Entonces,

(a) $\widehat{\theta}_n + \widehat{\theta}'_n \xrightarrow{P} \theta + \theta'$;

(b) $\widehat{\theta}_n \widehat{\theta}'_n \xrightarrow{P} \theta\theta'$;

(c) $\widehat{\theta}_n / \widehat{\theta}'_n \xrightarrow{P} \theta/\theta'$ si $\theta' \neq 0$;

(d) $\sqrt{\widehat{\theta}_n} \xrightarrow{P} \sqrt{\theta}$ si $P(\widehat{\theta}_n \geq 0) = 1$;

(e) Sea a_n una sucesión con límite a, entonces $a_n \widehat{\theta}_n \xrightarrow{P} a\theta$.

Ejemplo 20 (Consistencia de la cuasivarianza muestral)
Sea $(X_1, \ldots, X_n)$ una m.a.s. de una v.a. X con los siguientes momentos finitos,

$$E(X) = \mu, \quad E(X^2) = \alpha_2, \quad E(X^4) = \alpha_4,$$

y con varianza $V(X) = \alpha_2 - \mu^2 = \sigma^2$. ¿Es S'^2 consistente para σ^2?

Teorema 6 (Versión del Teorema de Slutsky)
Sean U_n y W_n dos variables aleatorias que verifican respectivamente

$$U_n \xrightarrow{d} N(0,1), \quad W_n \xrightarrow{P} 1.$$

Entonces se verifica que

$$U_n/W_n \xrightarrow{d} N(0,1).$$

Ejemplo 21 Sea $(X_1, \ldots, X_n)$ una m.a.s. de una v.a. X con $E[X] = \mu$ y $V[X] = \sigma^2$. ¿Qué distribución asintótica tiene el siguiente estadístico?

$$T = \frac{\bar{X} - \mu}{S'/\sqrt{n}}.$$

Ejemplo 22 Sean $X_1, \ldots, X_n$ variables aleatorias i.i.d. con distribución $Bern(\pi)$, siendo π la probabilidad de éxito. Sea $\widehat{\pi} = \sum_{i=1}^{n} X_i/n$ la proporción de éxitos. Calcula la distribución asintótica del estadístico

$$\frac{\widehat{\pi} - \pi}{\sqrt{\widehat{\pi}(1 - \widehat{\pi})/n}}.$$

3.6. ESTIMADORES CENTRADOS DE MÍNIMA VARIANZA Y ESTIMADORES EFICIENTES

En esta sección vamos a considerar estimadores $\widehat{\theta}$ de θ insesgados y con varianza finita.

Estimador ECUMV: Un estimador $\widehat{\theta}$ de θ es *centrado de uniformemente mínima varianza* (ECUMV), si es insesgado, y dentro del conjunto de estimadores insesgados con varianza finita, presenta varianza mínima para cualquier valor del parámetro $\theta \in \Theta$, es decir,

$$V[\widehat{\theta}] \leq V[\widehat{\theta}'], \qquad \forall \widehat{\theta}', \forall \theta \in \Theta.$$

Cantidad de Información de Fisher: Sea X una v.a. continua con distribución dependiente de un parámetro $\theta \in \Theta \subset I\!R$. Se define la *cantidad de información de Fisher* de X sobre θ como

$$I(\theta) = E\left[\left(\frac{\partial \log f(x;\theta)}{\partial \theta}\right)^2\right],$$

donde se toma $\partial \log f(x;\theta)/\partial\theta = 0$ si $f(x;\theta) = 0$.

Si X es discreta, entonces se reemplaza la función de densidad $f(x;\theta)$ por la función de cuantía $p(x;\theta)$.

Nota 3 Obsérvese que la cantidad

$$\frac{\partial \log f(x;\theta)}{\partial \theta} = \frac{\frac{\partial f(x;\theta)}{\partial \theta}}{f(x;\theta)} \qquad (1)$$

es la tasa de variación relativa de f al variar ligeramente θ, para la realización x de X. Por tanto, (1) representa la información que posee x para diferenciar θ de un valor próximo $\theta + h$. Al tomar esperanzas, obtenemos la información promedio de X para diferenciar θ de valores próximos.

Ejemplo 23 Calcular la cantidad de información de Fisher de una v.a. $X \sim Pois(\lambda)$.

Ejemplo 24 Si ahora se dispone de una m.a.s. $(X_1, \ldots, X_n)$ de $X \sim Pois(\lambda)$. Calcula la cantidad de información de Fisher de $(X_1, \ldots, X_n)$ sobre λ. ¿Es esta cantidad de información mayor o menor que la anterior?

Proposición 3 Para una m.a.s. $(X_1, \ldots, X_n)$, la cantidad de información de Fisher es aditiva. Es decir, la cantidad de información de

Fisher de $(X_1, \ldots, X_n)$ sobre θ es

$$\mathcal{I}_n(\theta) = E\left[\left(\frac{\partial \log f(x_1, \ldots, x_n; \theta)}{\partial \theta}\right)^2\right] = nI(\theta).$$

Teorema 7 (Desigualdad de Frechet-Crámer-Rao)
Sea $(X_1, \ldots, X_n)$ una m.a.s. de una v.a. con función de densidad $f(x;\theta)$, y sea $\mathcal{I}_n(\theta)$ la cantidad de información de Fisher de la muestra $(X_1, \ldots, X_n)$ sobre θ. Si $\widehat{\theta} = \widehat{\theta}(X_1, \ldots, X_n)$ es una estimador insesgado de θ, entonces bajo ciertas condiciones generales, se verifica

$$V[\widehat{\theta}] \geq \mathcal{I}_n(\theta)^{-1}.$$

Estimador eficiente: Un estimador insesgado $\widehat{\theta}$ de θ que verifica $V[\widehat{\theta}] = \mathcal{I}_n^{-1}(\theta)$ se denomina *eficiente*.

Ejemplo 25 ¿Es un estimador eficiente ECUMV?. ¿Y al contrario?

Ejemplo 26 Demostrar que para una v.a. $Pois(\lambda)$, el estimador $\widehat{\lambda} = \bar{X}$ es eficiente.

Ejemplo 27 Demostrar que para una m.a.s. $(X_1, \ldots, X_n)$ de una v.a. $\mathcal{N}(\mu, \sigma^2)$, con σ^2 conocida, el estimador $\widehat{\mu} = \bar{X}$ es eficiente.

3.7 ESTADÍSTICOS SUFICIENTES

Un estadístico es *suficiente* para un parámetro θ si recoge la información que la muestra contiene para estimar θ.

En otras palabras, si para una muestra determinada, se conoce el valor del estadístico (que resume la información de la muestra sobre θ), entonces la muestra ya no contiene ninguna información extra sobre θ.

Ejemplo 28 Un experimento con posibles resultados *éxito* (con probabilidad p), y *fracaso* se repite n veces, siendo $X_1, \ldots, X_n$ los resultados. Sea $Y = \sum_{i=1}^{n} X_i$ el número de éxitos obtenidos en las n pruebas. Si se han observado $Y = y$ éxitos ¿se puede sacar más información de la muestra sobre p aparte de y? Para contestar a esta pregunta, calcula la probabilidad de una muestra, condicionada al valor observado del estadístico, y comprueba si el resultado depende de p.

⋮

Estadístico suficiente: Sea $(X_1, \ldots, X_n)$ una m.a.s. de una v.a. X con distribución dependiente de un parámetro desconocido θ. Un estadístico $T = T(X_1, \ldots, X_n)$ es *suficiente* para θ si la distribución de la muestra $(X_1, \ldots, X_n)$, condicionada a T no depende de θ.

Ejemplo 29 Sea T es un estadístico suficiente para θ:

(a) ¿Crees que lo es también para $g(\theta)$, siendo $g(\cdot)$ una función inyectiva cualquiera?.

(b) Sea $T' = kT$, donde k es un número real conocido. ¿Crees que T' es suficiente para θ?.

Sea θ un valor concreto del parámetro desconocido. Si se observa una muestra $(x_1, \ldots, x_n)$ de una v.a. discreta, la *verosimilitud* de θ va a ser la probabilidad de observar $(x_1, \ldots, x_n)$ si el verdadero valor del parámetro fuese θ; es decir, es la credibilidad que tiene el valor θ si se ha observado $(x_1, \ldots, x_n)$.

Ejemplo 30 Sea la variable X : Número de caras obtenidas en dos lanzamientos de una moneda con probabilidad de cara $\theta \in \{0{,}2, 0{,}8\}$.

$X \sim \text{Bin}(n = 2, \theta)$,

$L(\theta; x) = p_X(x; \theta) = \binom{2}{x} \theta^x (1-\theta)^{2-x}$, $x \in \{0, 1, 2\}$.

θ	$x = 0$	$x = 1$	$x = 2$
0.2	0.64	0.32	0.04
0.8	0.04	0.32	0.64

Si se observan $x = 0$ caras, ¿qué valor de θ es el más creible?
¿Y si se observan $x = 2$ caras?

Verosimilitud: Sea $(X_1, \ldots, X_n)$ un conjunto de variables aleatorias con distribución dependiente de un parámetro θ, y sea $(x_1, \ldots, x_n)$ una realización (valor observado) de $(X_1, \ldots, X_n)$. Si las variables son discretas, la *verosimilitud* (credibilidad) que $(x_1, \ldots, x_n)$ le da al valor θ del parámetro es

$$L(\theta; x_1, \ldots, x_n) = p(X_1 = x_1, \ldots, X_n = x_n; \theta) = p_{(X_1, \ldots, X_n)}(x_1, \ldots, x_n; \theta).$$

Si las variables son continuas, entonces dicha verosimilitud es

$$L(\theta; x_1, \ldots, x_n) = f_{(X_1, \ldots, X_n)}(x_1, \ldots, x_n; \theta).$$

¡Ojo! $L(\theta; x_1, \ldots, x_n)$ es una función de θ, ya que la muestra $(x_1, \ldots, x_n)$ es fija.

Ejemplo 31 Si $(X_1, \ldots, X_n)$ son independientes, ¿qué valor toma la verosimilitud de θ?

Teorema 8 (Criterio de factorización de Fisher y Neyman) Sea $(X_1, \ldots, X_n)$ una m.a.s. de una v.a. con distribución dependiente de un parámetro desconocido θ. El estadístico $T = T(X_1, \ldots, X_n)$ es suficiente para θ si y solo si la verosimilitud se puede factorizar en dos funciones no negativas de la forma

$$L(\theta; x_1, \ldots, x_n) = g(t, \theta) h(x_1, \ldots, x_n),$$

donde $g(t, \theta)$ depende de la muestra solo a través del valor del estadístico t, y $h(x_1, \ldots, x_n)$ no depende de θ.

Prueba:

Ejemplo 32 Para el Ejemplo 28, comprueba que $T = \sum_{i=1}^{n} X_i$ es suficiente para p utilizando el Teorema de Factorización.

Ejemplo 33 Sea $(X_1, \ldots, X_n)$ una m.a.s. de una $Exp(\alpha)$. ¿Es $\bar{X}$ suficiente para α?.

Ejemplo 34 Sea $(X_1, \ldots, X_n)$ una m.a.s. de una v.a. $U(\theta_1, \theta_2)$, con $\theta_1 < \theta_2$, siendo θ_1 y θ_2 desconocidos. Obtener un estadístico suficiente para (θ_1, θ_2).

Obtén un estadístico suficiente para θ_2, si θ_1 es conocido.

Ejemplo 35 Sea $(X_1, \ldots, X_n)$ una m.a.s. de una v.a. con distribución $\mathcal{N}(\mu, \sigma^2)$. Obténgase estadísticos suficientes para

(a) σ^2, si μ es conocido;

(b) μ, si σ^2 es conocido;

(c) μ y σ^2.

Familia Exponencial ($\mathcal{F}_{Exp}$): Una v.a. X pertenece a $\mathcal{F}_{Exp}$ si su función de cuantía o de densidad $f(x;\theta)$ se puede expresar como

$$f(x;\theta) = c(\theta)h(x)\,\mathrm{exp}\{w(\theta)t(x)\}.$$

Ejemplo 36 Comprobar que una v.a. $X \sim Bin(n,\theta)$ pertenece a la Familia Exponencial:

Ejemplo 37 Comprobar que una v.a. $\gamma(\theta, 3)$ pertenece a la Familia Exponencial:

Ejemplo 38 Una $X \sim U(0, \theta)$ no pertenece a la Familia Exponencial:

$$f(x;\theta) = \left\{ \begin{array}{ll} 1/\theta, & \text{si } x \in [0,\theta]; \\ 0, & \text{si } x \notin [0,\theta]. \end{array} \right\} = \frac{1}{\theta} I_{[0,\theta]}(x)$$

El indicador es función de x y de θ, y no se puede expresar en forma de una función exponencial: X no pertenece a la Familia Exponencial.

Proposición 4 Sea $(X_1, \ldots, X_n)$ una m.a.s. de una v.a. $X \in \mathcal{F}_{Exp}$. Entonces, el estadístico

$$T(X_1, \ldots, X_n) = \sum_{i=1}^{n} t(X_i)$$

es suficiente.

Prueba:

Ejemplo 39 Obtener un estadístico suficiente para θ en el Ejemplo 36.

Ejemplo 40 Obtener un estadístico suficiente para θ en el Ejemplo 37.