INFERENCIA ESTADISTICA

Guillermo Ramirez, Adelmo Fernández y Maura Vásquez*

2012

*Escuela de Estadística y Ciencias Actuariales de la Universidad Central de Venezuela

Índice general

1. Estimación puntual **1**
1.1. Introducción 1
1.2. Métodos de obtención de estimadores 4
1.2.1. Método de los momentos 5
1.2.2. Método de máxima verosimilitud 7
1.3. Propiedades de los estimadores 10
1.3.1. Insesgamiento 11
1.3.2. Consistencia 12
1.3.3. Suficiencia 19
1.3.4. Completitud 24
1.3.5. Eficiencia 25
1.4. La clase exponencial de distribuciones 26
1.5. Estimadores insesgados de mínima varianza 30
1.5.1. Cota de Cramer-Rao 31
1.5.2. Suficiencia y completitud 34
1.6. Propiedad asintótica de los estimadores máximo verosímiles . . 37

Capítulo 1

Estimación puntual

1.1. Introducción

Uno de los más importantes problemas a los que se enfrenta la Estadística es a la obtención de conclusiones acerca de la naturaleza de una población sobre la base del estudio de la información proporcionada por una muestra. Es este tipo de generalizaciónes lo que se conoce como inferencia estadística. A lo largo de este curso se ampliarán y profundizarán los conocimientos sobre esta temática, que han sido introducidos en cursos anteriores.

En este primer tema se desarrollan conceptos fundamentales de estimación puntual, haciéndose énfasis en las propiedades de los estimadores y en los métodos de obtención de estimadores insesgados de varianza mínima.

Comencemos preguntándonos: qué es la inferencia estadística?

El problema general de la inferencia estadística puede resumirse muy brevemente en la siguiente forma:

Se desea estudiar un colectivo (población) de elementos (individuos) que resulta muy extenso o complejo de examinar. Se selecciona y estudia una parte (muestra) de la población, y de acuerdo con ciertos procedimientos, se intenta generalizar los resultados obtenidos hacia la población total. La inferencia estadística es el conjunto de técnicas y procedimientos que permiten efectuar, con cierto grado de incertidumbre, la generalización anterior.

Resulta conveniente en este momento realizar algunas precisiones sobre los conceptos de población y muestra. En general se entiende por población el conjunto de todos los elementos bajo estudio. Ahora bien, cuando estudiamos una determinada característica, originamos de hecho una función real a la cual denominamos variable poblacional. Matemáticamente se trata de una función X que asigna a cada elemento w de la población Ω un número real $X(w)$, de acuerdo con un determinado criterio. Esta función X es realmente una variable aleatoria: variable porque cada elemento de la población tendrá un valor particular, que no necesariamente es igual a los del resto; y aleatoria en el sentido de que al seleccionar un elemento al azar no se puede conocer de antemano su valor asociado. Por ser X una variable aleatoria, tendrá una función de densidad, que puede considerarse como un modelo matemático de la distribución de frecuencias obtenida en la situación idealizada de conocer los valores asociados a todos y cada uno de los elementos de la población. Los valores de las cuales depende la función de densidad de X se denominan parámetros poblacionales, se denotarán por lo general mediante letras griegas $(\mu, \sigma, \theta, \dots)$, y por lo general son desconocidos.

Si seleccionamos al azar n individuos de la población, estamos de hecho generando n variables aleatorias $X_1, X_2, \dots X_n$, donde X_i denota el valor de la variable X correspondiente al individuo i. En el caso en que la muestra se obtenga con reemplazo, las n variables anteriores van a tener la misma distribución que X y además van a ser independientes. Es muy importante aclarar aquí que el término población será utilizado en este curso tanto para referirnos al colectivo de individuos estudiados, como a una determinada variable poblacional X (más en el segundo sentido que en el primero).

Definicion 1.1 (Población y muestra aleatoria)**.**
Una población es una variable aleatoria X con una cierta función de densidad $f(x;\theta)$. La constante θ se denomina parámetro. Al conjunto de valores posibles de θ lo denominaremos espacio paramétrico y lo denotaremos por Θ. Se dice que las variables aleatorias $X_1, X_2, \dots X_n$ constituyen una muestra aleatoria de la población X, si:
i) son independientes, y
ii) están idénticamente distribuidas con función de densidad $f(x;\theta)$.

Definicion 1.2 (Distribución muestral).
Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población X con función de densidad $f(x;\theta)$. Se denomina distribución muestral a la distribución conjunta de las variables $X_1, X_2, \ldots X_n$. Por la definición anterior, se tiene que esta distribución viene dada por:

$$f_X(x_1, x_2, \ldots x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta) \tag{1.1}$$

Ejemplo 1.1. *Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con distribución de poisson $P(\theta)$. Halle la distribución muestral.*

La función de densidad poblacional:

$$f(x;\theta) = e^{-\theta}\ \frac{\theta^x}{x!}\ I_{\{0,1\ldots\}}(x)$$

La función de densidad muestral:

$$\begin{aligned} f_X(x_1, x_2, \ldots x_n; \theta) &= \prod_{i=1}^{n} f(x_i; \theta) = \prod_{i=1}^{n} (e^{-\theta}\ \frac{\theta^{x_i}}{x_i!} I_{\{0,1\ldots\}}(x_i)) \\ &= e^{-n\theta}\ \frac{\theta^{\sum x_i}}{\prod x_i!}\ \prod_{i=1}^{n} I_{\{0,1\ldots\}}(x_i) \end{aligned}$$

Ejemplo 1.2. *Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con distribución exponencial $E(\theta)$. Halle la distribución muestral.*

La función de densidad poblacional:

$$f(x;\theta) = \theta\ e^{-\theta x}\ I_{(0,\infty)}(x)$$

La función de densidad muestral:

$$\begin{aligned} f_X(x_1, x_2, \ldots x_n; \theta) &= \prod_{i=1}^{n} f(x_i; \theta) = \prod_{i=1}^{n} (\theta\ e^{-\theta x_i} I_{(0,\infty)}(x_i)) \\ &= \theta^n\ e^{-\theta \sum x_i}\ \prod_{i=1}^{n} I_{(0,\infty)}(x_i) \end{aligned}$$

Ejemplo 1.3. *Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con distribución uniforme $U(0,\theta)$. Halle la distribución muestral.*

La función de densidad poblacional:

$$f(x;\theta) = \frac{1}{\theta}\, I_{(0,\theta)}(x)$$

La función de densidad muestral:

$$\begin{aligned} f_X(x_1, x_2, \ldots x_n; \theta) &= \prod_{i=1}^{n} f(x_i;\theta) = \prod_{i=1}^{n} \Big(\frac{1}{\theta} I_{(0,\theta)}(x_i) \Big) \\ &= \Big(\frac{1}{\theta} \Big)^n \prod_{i=1}^{n} I_{(0,\theta)}(x_i) \end{aligned}$$

Definicion 1.3 (Estadístico)**.**
Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población X con función de densidad $f(x;\theta)$. Un estadístico muestral, o simplemente estadístico, es una función $T = t(X_1, X_2, \ldots X_n)$ de las variables que constituyen la muestra, que no depende de ningún parámetro desconocido.

Una vez seleccionada una muestra, puede ser obtenido el valor concreto de un estadístico T, digamos t_0, correspondiente a esa muestra. Se dice que t_0 es el "valor observado"de T. Además, como un estadístico es una función de variables aleatorias, también es una variable aleatoria.

Definicion 1.4 (Estimador)**.**
Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población X con función de densidad $f(x;\theta)$. Se define como estimador de θ a cualquier estadístico T tal que $rango(T) = \Theta$.

Por ser un estadístico, un estimador es también una variable aleatoria y por lo tanto tiene una función de densidad. Esta función de densidad se denotará por $f_T(t;\theta)$, ya que muy probablemente dependerá de la distribución poblacional y del parámetro θ.

1.2. Métodos de obtención de estimadores

En este apartado presentaremos dos métodos para encontrar estimadores de un determinado parámetro θ. Ambos procedimientos pueden generalizarse al caso de dos o más parámetros.

1.2.1. Método de los momentos

Se trata de un procedimiento bastante intuitivo cuya formulación se desprende de los dos resultados siguientes:

$$E(M'_r) = \mu'_r \tag{1.2}$$

$$V(M'_r) = \frac{1}{n}(\mu'_{2r} - \mu'^2_r) \tag{1.3}$$

siendo M'_r el momento muestral de orden r y μ'_r el momento poblacional de orden r. Lo que este método plantea es una simple igualación de los momentos ordinarios con los momentos poblacionales correspondientes.

Definicion 1.5 (Estimador de los momentos)**.**
Se define como estimador de los momentos del parámetro θ *a la función* $\tilde{\theta} = \tilde{\theta}(X_1, X_2, \dots X_n)$ *obtenida como solución de la ecuación:*

$$M'_r = \mu'_r(\theta)$$

siendo r *el mínimo valor para el cual existe* μ'_j *y depende de* θ*.*

Recuérdese que:

$$\mu'_r = \text{Momento poblacional de orden r} = E(X^r) \quad r = 1, 2...$$

$$M'_r = \text{Momento muestral de orden r} = \frac{1}{n}\sum X_i^r \quad r = 1, 2...$$

En caso de más de un parámetro, definimos el vector $\theta = (\theta_1, \theta_2, , \dots \theta_k)$, y hallamos el estimador de los momentos $\tilde{\theta} = (\tilde{\theta}_1, \tilde{\theta}_2, \dots \tilde{\theta}_k)$ resolviendo el sistema de ecuaciones:

$$M'_j = \mu'_j(\theta_1, \theta_2, \dots \theta_k) \quad \text{para j = 1,2... k}$$

Este procedimiento presenta un conjunto de desventajas:

i) No necesariamente existen estos estimadores ya que no se garantiza que el sistema de ecuaciones tenga solución.

ii) No siempre μ'_j dependerá de $\theta_1, \theta_2, \dots \theta_k$, en cuyo caso habría que escoger aquellos momentos que realmente dependan de todos los parámetros

iii) No son únicos en el sentido de que obtendríamos diferentes estimadores si considerásemos, por ejemplo, los momentos centrales en lugar de los momentos ordinarios.

iv) No son únicos en el sentido de que obtendríamos diferentes estimadores si considerásemos el sistema de ecuaciones para $j = 1, 2 \dots k$ o el sistema para $j = 2, 3 \dots k + 1$

v) No hay garantía de que los estimadores obtenidos con este método tengan buenas propiedades.

Ejemplo 1.4. *Sea $X_1, X_2, \dots X_n$ una muestra aleatoria de una población con distribución de poisson $P(\theta)$. Halle el estimador de θ por el método de los momentos.*

Para hallar $\tilde{\theta}$ resolvemos la ecuación:

$$M_1' = \mu_1'(\theta)$$

Como en este caso se cumple que $\mu_1'(\theta) = \theta$, entonces la ecuación anterior queda:

$$\bar{X} = \theta$$

y resolviendo para θ:

$$\tilde{\theta} = \bar{X}$$

Ejemplo 1.5. *Sea $X_1, X_2, \dots X_n$ una muestra aleatoria de una población con distribución exponencial $E(\theta)$. Halle el estimador de θ por el método de los momentos.*

Para hallar $\tilde{\theta}$ resolvemos la ecuación:

$$M_1' = \mu_1'(\theta)$$

Como en este caso se cumple que $\mu_1'(\theta) = 1/\theta$, entonces la ecuación anterior queda:

$$\bar{X} = 1/\theta$$

de donde:

$$\tilde{\theta} = 1/\bar{X}$$

Ejemplo 1.6. *Sea $X_1, X_2, \dots X_n$ una muestra aleatoria de una población con distribución uniforme $U(0, \theta)$. Halle el estimador de θ por el método de los momentos.*

Para hallar $\tilde{\theta}$ resolvemos la ecuación:

$$M_1' = \mu_1'(\theta)$$

En este caso se tiene que $\mu_1'(\theta) = \theta/2$, así que la ecuación de partida es:

$$\bar{X} = \theta/2$$

obteniéndose:

$$\tilde{\theta} = 2\bar{X}$$

1.2.2. Método de máxima verosimilitud

Este método se sustenta en un razonamiento de tipo probabilístico y requiere la introducción del concepto de función de verosimilitud[1].

Definicion 1.6 (Función de verosimilitud)**.**
Sea $X_1, X_2, \dots X_n$ una muestra aleatoria de una población X con función de densidad $f(x; \theta)$. Se define como función de verosimilitud a a la función de densidad muestral, pero considerada como función de θ y no de $X_1, X_2, \dots X_n$:

$$L(\theta; x_1, x_2, \dots x_n) = \prod_{i=1}^{n} f(x_i; \theta) \tag{1.4}$$

Intuitivamente, la función de densidad muestral $f(x_1, x_2, \dots x_n; \theta)$ puede interpretarse como la probabilidad de la muestra observada dado el valor particular de θ, mientras que la función de verosimilitud $L(\theta; x_1, x_2, \dots x_n)$ puede interpretarse como la probabilidad del valor observado de θ dada la muestra obtenida. El método de máxima verosimilitud define como estimador de θ al valor más probable una vez observada la muestra, lo que se halla maximizando la función de verosimilitud.

Definicion 1.7 (Estimador máximo verosímil)**.**
Sea $X_1, X_2, \dots X_n$ una muestra aleatoria de una población X con función de densidad $f(x; \theta)$ y $L(\theta)$ la función de verosimilitud. Se define como estimador máximo verosímil de θ a la función $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots X_n)$ que maximiza la función de verosimilitud.

[1]El término verosimilitud es sinónimo de credibilidad

Cuando la función $L(\theta)$ satisface ciertas condiciones, se obtiene resolviendo para θ la ecuación:

$$\frac{d}{d\theta}L(\theta) = 0 \tag{1.5}$$

además, como las funciones $L(\theta)$ y $l(\theta) = logL(\theta)$ alcanzan su máximo para el mismo valor de θ, también puede hallarse $\hat{\theta}$ resolviendo la ecuación:

$$\frac{d}{d\theta}l(\theta) = 0 \tag{1.6}$$

Cuando la función $L(\theta)$ no satisface las condiciones, debe hallarse $\hat{\theta}$ maximizando $L(\theta)$ mediante otro procedimiento. Por ejemplo, estudiando el crecimiento o decrecimiento de la función.

Ejemplo 1.7. *Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con distribución de poisson $P(\theta)$. Halle el estimador de θ por el método de máxima verosimilitud.*

La función de verosimilitud:

$$L(\theta) = e^{-n\theta}\frac{\theta^{\sum x_i}}{\prod x_i!} \qquad \theta > 0$$

Tomando logaritmo:

$$l(\theta) = logL(\theta) = -n\theta + \sum x_i log(\theta) - log(\prod x_i!)$$

Derivando respecto de θ:

$$\frac{d}{d\theta}l(\theta) = -n + \frac{\sum X_i}{\theta}$$

Igualando a cero se obtiene el punto crítico $\hat{\theta} = \bar{X}$. Utilizando el criterio de la segunda derivada se encuentra que en dicho punto hay un máximo. En consecuencia:

$$\hat{\theta} = \bar{X}$$

es el estimador máximo verosímil de θ.

Ejemplo 1.8. *Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con distribución exponencial $E(\theta)$. Halle el estimador de θ por el método de máxima verosimilitud.*

La función de verosimilitud:

$$L(\theta) = \theta^n \, e^{-\theta \sum x_i} \qquad \theta > 0$$

Su logaritmo:

$$l(\theta) = logL(\theta) = nlog(\theta) - \theta \sum x_i$$

Al derivar respecto de θ queda:

$$\frac{d}{d\theta}\, l(\theta) = \frac{n}{\theta} - \sum X_i$$

Igualando a cero se obtiene el punto crítico $\hat{\theta} = 1/\bar{X}$. Mediante el criterio de la segunda derivada se concluye que en este punto se alcanza un máximo. En consecuencia, el estimador máximo verosímil de θ:

$$\hat{\theta} = \frac{1}{\bar{X}}$$

Ejemplo 1.9. *Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con distribución uniforme $U(0, \theta)$. Halle el estimador máximo versosímil de θ.*

La función de verosimilitud:

$$L(\theta) = \left(\frac{1}{\theta}\right)^n I_{(y_n, \infty)}(\theta)$$

En este caso no tiene interés maximizar su logaritmo. Si derivamos directamente la función $L(\theta)$ encontramos que no hay ningún valor de θ que anule esa derivada. Sin embargo, al analizar la función se encuentra que es decreciente y, por tanto, el valor de θ donde se alcanza el máximo de la función es el mínimo de θ. Por ello:

$$\hat{\theta} = Y_n$$

En general los estimadores máximo verosímiles tienen buenas propiedades. En el siguiente teorema nos referiremos a una de ellas.

Teorema 1.1 (Propiedad de invarianza)**.** *Si $\hat{\theta}$ es el estimador máximo verosímil de θ, entonces $\tau(\hat{\theta})$ es el estimador máximo verosímil de $\tau(\theta)$.*

Se desprende de esta propiedad que si queremos, por ejemplo, obtener el estimador máximo verosímil de θ^2 podríamos proceder de dos maneras: Maximizar la función de verosimilitud respecto de θ^2, o elevar al cuadrado el estimador máximo verosímil de θ.

Ejemplo 1.10. *Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con distribución normal $N(0, \theta)$. Halle el estimador máximo verosímil de $\sqrt{\theta}$.*

La función de densidad poblacional:

$$f(x;\theta) = \frac{1}{\sqrt{\theta}\sqrt{2\pi}}\, e^{-x^2/2\theta}$$

La función de verosimilitud:

$$\begin{aligned} L(\theta) &= \prod f(x_i;\theta) = \prod \Big(\frac{1}{\sqrt{\theta}\sqrt{2\pi}}\, e^{-x_i^2/2\theta} \Big) \\ &= \frac{1}{\theta^{n/2}(2\pi)^{n/2}} e^{-\sum x_i^2/2\theta} \qquad \theta > 0 \end{aligned}$$

Al tomar logaritmo:

$$l(\theta) = -\frac{n}{2} log(\theta) - \frac{n}{2} log(2\pi) - \frac{\sum x_i^2}{2\theta}$$

derivando respecto de θ:

$$\frac{d}{d\theta}\, l(\theta) = -\frac{n}{2\theta} + \frac{\sum x_i^2}{2\theta^2}$$

igualando a cero:

$$\frac{\sum x_i^2}{2\theta^2} = \frac{n}{2\theta}$$

de donde:

$$\hat{\theta} = \frac{\sum x_i^2}{n}$$

Por la propiedad de invarianza enunciada en el teorema 1.1, tenemos que el estimador máximo verosímil de $\sqrt{\theta}$ es:

$$(\hat{\sqrt{\theta}}) = \sqrt{\hat{\theta}} = \sqrt{\frac{\sum x_i^2}{n}}$$

1.3. Propiedades de los estimadores

Una vez que hemos presentado dos métodos para hallar estimadores, nos planteamos ahora el problema de decidir si un estimador es bueno o no, y si un estimador es mejor que otro. Para ello vamos a definir una serie de propiedades deseables en un estimador, y en principio diremos que un estimador será mejor en la medida que cumpla con más de estas propiedades.

1.3.1. Insesgamiento

Definicion 1.8 (Estimador insesgado)**.**
Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de $f(x;\theta)$. Se dice que un estimador T es insesgado para $\tau(\theta)$ si se cumple que:

$$E(T) = \tau(\theta) \tag{1.7}$$

es decir, si su distribución muestral está centrada en $\tau(\theta)$. Los estimadores insesgados no siempre existen, y cuando existen no necesariamente son únicos.
En el caso en que $limE(T_n) = \tau(\theta)$ se dice que T es un estimador asintóticamente insesgado para $\tau(\theta)$. La propiedad de insesgamiento no es invariante en el sentido de que si T es insesgado para θ, no necesariamente $\tau(T)$ es un estimador insesgado para $\tau(\theta)$. El hecho por ejemplo, de que s^2 sea un estimador insesgado de σ^2 no implica que s sea un estimador insesgado de σ. Sí se cumple esta propiedad de invarianza en el caso en que $\tau(\theta)$ sea una función lineal de θ ya que si $E(\hat{\theta})) = \theta$ entonces $E(a + b\hat{\theta}) = a + b\theta$.

Definicion 1.9 (Error cuadrático medio)**.**
Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de $f(x;\theta)$ y T un estimador de $\tau(\theta)$. Se define como error cuadrático medio de T con respecto a $\tau(\theta)$, a:

$$ECM_T(\tau(\theta)) = E[(T - \tau(\theta))^2] \tag{1.8}$$

Se trata pues del valor esperado del desvío cuadrático de T con respecto a $\tau(\theta))$, lo que constituye una medida de dispersión. Como consecuencia, un estimador será mejor que otro si su error cuadrático medio es menor.
De esta definición se desprende una relación inmediata del error cuadrático medio con la varianza de un estimador.

Teorema 1.2. *Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de $f(x;\theta)$ y T un estimador de $\tau(\theta)$. Entonces:*

$$ECM_T(\tau(\theta)) = Var(T) + b_T^2(\tau(\theta)) \tag{1.9}$$

donde $b_T^2((\tau(\theta)) = (sesgo)^2 = (\tau(\theta) - E(T))^2$.

Corolario 1.2.1. *Si T es un estimador insesgado para $\tau(\theta)$, entonces el sesgo es cero y su error cuadrático medio coincide con la varianza.*

En el siguiente teorema se afirma que sea cual sea la función de densidad poblacional $f(x;\theta)$, la media y la varianza muestrales constituyen estimadores insesgados de la media y la varianza poblacionales.

Teorema 1.3. *Sea $X_1, X_2, \ldots X_n$ una muestra de una población con media μ y varianza σ^2. Entonces:*

i) La media muestral $\bar{x}$ es un estimador insesgado para μ.

ii) La varianza muestral s^2 es un estimador insesgado para σ^2.

1.3.2. Consistencia

Utilizaremos ahora algunos conceptos de convergencia de variables aleatorias para definir propiedades de los estimadores que tienen interés en el caso de muestras grandes.

Definicion 1.10 (Estimador consistente)**.**
Se dice que un estimador T es consistente para $\tau(\theta)$, si la sucesión $\{T_n\}$ converge en probabilidad a $\tau(\theta)$. Es decir, si:

$$lim_{n\to\infty}\, P(|T_n - \tau(\theta)| < \varepsilon) = 1 \quad \forall\varepsilon > 0,\ \forall\theta \tag{1.10}$$

Si un estimador es consistente, se dice que cumple con la propiedad de consistencia simple. A continuación definiremos un tipo de consistencia más fuerte que la consistencia simple.

Definicion 1.11 (Estimador consistente en error cuadrático)**.**
Se dice que un estimador T es consistente en error cuadrático para $\tau(\theta)$, si la sucesión $\{T_n\}$ converge en probabilidad a $\tau(\theta)$. Es decir, si:

$$lim_{n\to\infty}\, ECM_{T_n}(\tau(\theta)) = 0 \quad \forall\theta \tag{1.11}$$

De acuerdo con el teorema 1.2, un estimador será consistente en error cuadrático si tanto $Var(T_n)$ como $b_T^2(\tau(\theta))$ tienden a cero cuando n tiende a infinito, y de acuerdo con el corolario de ese mismo teorema, si un estimador es insesgado y su varianza tiende a cero, entonces es consistente en error cuadrático.

En el siguiente teorema se establece cuándo un estimador asintóticamente insesgado es consistente en error cuadrático.

Teorema 1.4. *Si T es un estimador asintóticamente insesgado y su varianza tiende a cero, entonces es consistente en error cuadrático medio.*

Demostración.
Si T es asintóticamente insesgado:

$$lim\ E(T_n) = \tau(\theta) \quad \Rightarrow \quad lim\ b_T^2(\tau(\theta)) = 0$$

Por el teorema 1.2:

$$ECM_T(\tau(\theta)) = Var(T_n) + b_T^2(\tau(\theta))$$

tomando límite:

$$lim\ ECM_T(\tau(\theta)) = lim\ Var(T_n) = 0$$

así que T es consistente en error cuadrático.

□

Veremos a continuación que la consistencia en error cuadrático implica la consistencia simple.

Teorema 1.5. *Si T es un estimador consistente en error cuadrático medio para $\tau(\theta)$), entonces es consistente para $\tau(\theta)$).*

Demostración.
Si T es consistente en error cuadrático entonces $lim\ ECM_T(\tau(\theta)) = 0$.
Por otro lado, según la desigualdad general de Tchebycheff:

$$P(g(T_n) < k) \geq 1 - \frac{E[g(T_n)]}{k} \qquad \forall k > 0 \tag{1.12}$$

Si hacemos:
$g(T_n) = (T_n - \tau(\theta))^2$ y $k = \varepsilon^2$, queda:

$$P(|T_n - \tau(\theta))| < \varepsilon) \geq 1 - \frac{ECM_{T_n}(\tau(\theta))}{\varepsilon^2}$$

y al tomar límite:

$$lim\ P(|T_n - \tau(\theta))| < \varepsilon) = 1$$

lo que demuestra que T es consistente para $\tau(\theta)$. □

Teorema 1.6. *Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con media μ y varianza σ^2. Entonces:*

i) La media muestral $\bar{x}$ es un estimador consistente para la media poblacional μ.

ii) La varianza muestral s^2 es un estimador consistente para la varianza poblacional σ^2.

Demostración.

i) Como $\bar{x}$ es insesgado para μ y $lim\ Var(\bar{x}) = 0$, entonces es consistente en error cuadrático, y por tanto consistente.

ii) Como s^2 es insesgado para σ^2 y $lim\ Var(s^2) = 0$, entonces es consistente en error cuadrático, y por tanto consistente.

□

Conviene recordar aquí el teorema central del límite, que establece que para n grande la distribución de la media muestral se aproxima a la normal.

Teorema 1.7 (Teorema Central del Límite)**.** *Sea $\{X_n\}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas tales que $E(X_n) = \mu$ y $Var(X_n) = \sigma^2$. Entonces, la sucesión de variables aleatorias $\{Z_n\}$ definidas por:*

$$Z_n = \frac{\bar{x}_n - \mu}{\sigma/\sqrt{n}}$$

converge en distribución al modelo normal $N(0, 1)$.

Este teorema nos permite afirmar (asumiendo n grande) que:
$P[-\varepsilon < \bar{x} - \mu < \varepsilon] = 1 - \alpha$ siendo n la parte entera de $(\frac{k_\alpha \sigma}{\varepsilon})^2$, con $0 < \alpha < 1$, y $k_\alpha = Z_{1-\alpha/2}$.

Por ejemplo, si $\sigma = 46$, $\alpha = 0{,}05$ y $\varepsilon = 5$, entonces $k_\alpha = 1{,}96$ y $n = 326$. En palabras, necesitamos una muestra de tamaño 326 o más para que la probabilidad de que la media muestral $\bar{x}$ no se "aleje"de la media poblacional más de 5 unidades, sea igual a 0.95. O también: Con una muestra de tamaño 326 o mayor, tenemos una "confianza"del 95 % de que el error de estimación es menor de 5.

Ejercicios 1.1

1. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de cada una de las siguientes poblaciones. Halle el estimador de θ por el método de los momentos y por el de máxima verosimilitud:

i) bernoulli $B(\theta)$	v) beta $B(\theta, 1)$
ii) poisson $P(\theta)$	vi) normal $N(\theta, 1)$
iii) geometrica $G(\theta)$	vii) normal $N(0, \theta)$
iv) exponencial $E(\theta)$	viii) uniforme $U(0, \theta)$

2. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de cada una de las siguientes poblaciones. Halle el estimador máximo verosímil de θ:

   i) $f(x;\theta) = \frac{1-\theta}{2}\, I_{\{0\}}(x) + \frac{1}{2}\, I_{\{1\}}(x) + \frac{\theta}{2}\, I_{\{2\}}(x) \quad 0 < \theta < 1$

   ii) $f(x;\theta) = \theta(1/x)^{\theta+1} I_{(1,\infty)}(x)$

   iii) $f(x;\theta) = e^{-(x-\theta)}\, I_{(\theta,\infty)}(x)$

   iv) Uniforme $U(\theta - 1/2, \theta + 1/2)$

   v) Uniforme $U(-\theta, \theta) \quad \theta > 0$

3. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de cada una de las siguientes poblaciones. Halle el estimador máximo verosímil para los parámetros respectivos:

   i) Normal $N(\mu, \sigma^2)$

   ii) Uniforme $U(\theta_1, \theta_2)$

   iii) Uniforme $U(\mu - \sqrt{3}\sigma, \mu + \sqrt{3}\sigma) \quad \sigma > 0$

   iv) $f(x; a, b) = \frac{1}{b} e^{-(x-a)/b} \quad I_{(a,\infty)}(x) \quad b > 0$

4. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con función de densidad:

$$f(x;p) = \frac{1}{1-q^2}\binom{2}{x} p^x q^{2-x}\, I_{\{1,2\}}(x) \;\text{ con } p > 0 \text{ y } p + q = 1$$

   Demuestre que $\hat{q} = \frac{2-\bar{x}}{\bar{x}}$ y $\hat{p} = \frac{2(\bar{x}-1)}{\bar{x}}$.

5. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con media μ y varianza σ^2. Sean $a_1, a_2, \ldots a_n$ números reales tales que $\sum a_i = 1$. Demuestre que:

    i) $T = \sum a_i X_i$ es un estimador insesgado para μ.

    ii) $Var(T)$ es mínima cuando $T = \bar{X}$.

6. Sean T_1 y T_2 estimadores independientes e insesgados para θ. La varianza de T_1 es el doble de la de T_2. Halle los valores de las constantes k_1 y k_2 tales que $T = k_1 T_1 + k_2 T_2$ sea un estimador insesgado para θ, y de mínima varianza.

7. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población bernoulli $B(\theta)$. Halle el estimador máximo verosímil de $\sigma^2(\theta) = \theta(1-\theta)$. ¿Es insesgado?

8. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con media μ y varianza σ^2. Demuestre que:

$$T = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (X_{i+1} - X_i)^2$$

   es un estimador insesgado para σ^2.

9. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con distribución uniforme $U(-\theta, \theta)$. Halle un estimador insesgado para θ basado en Y_n.

10. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población normal $N(\mu, \sigma^2)$ con μ conocida.

    i) Halle un estimador insesgado para σ basado en s. Halle su varianza.

    ii) Demuestre que:

$$T = \sqrt{\frac{\pi}{2}} \frac{\sum |X_i - \mu|}{n}$$

    es un estimador insesgado para σ. Halle su varianza.

11. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con distribución uniforme $U(\theta - 1/2, \theta + 1/2)$. Demuestre que $T = \frac{Y_1 + Y_n}{2}$ es un estimador insesgado para θ.

12. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con media μ y varianza σ^2. Demuestre que $\bar{x}$ y s^2 son estimadores insesgados y consistentes para μ y σ^2 respectivamente.

13. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población exponencial $E(\theta)$. Halle un estimador insesgado para:

    i) θ

    ii) $1/\theta$

    ¿Es consistente?

14. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población beta $B(\theta, 1)$. Halle un estimador insesgado para:

    i) θ

    ii) $1/\theta$

    ¿Es consistente?

15. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población uniforme $U(0, \theta)$. Halle un estimador insesgado para:

    i) θ

    ii) $-\theta$

    ¿Es consistente?

16. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con función de densidad:

    $$f(x; \theta) = \theta(1 + x)^{-(\theta+1)} \; I_{(0,\infty)}(x)$$

    Halle un estimador insesgado para:

    i) θ

    ii) $1/\theta$

    ¿Es consistente?

17. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con función de densidad:
$$f(x;\theta) = e^{-(x-\theta)}\, I_{(\theta,\infty)}(x)$$
Halle un estimador insesgado para θ. ¿Es consistente?

18. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con función de densidad:
$$f(x;\theta) = \frac{\theta(log(1/\theta))^x}{x!}\, I_{\{0,1\ldots\}}(x)$$
Halle un estimador insesgado para $\tau(\theta) = -log(\theta)$. ¿Es consistente?

19. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con función de densidad:
$$f(x;\theta) = \frac{2x}{\theta^2}\, I_{(0,\theta)}(x) \qquad \theta > 0$$
Halle un estimador insesgado para θ. ¿Es consistente?

20. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con función de densidad:
$$f(x;\theta) = \frac{\theta}{x^2}\, I_{(\theta,\infty)}(x) \qquad \theta > 0$$
Halle un estimador insesgado para θ. ¿Es consistente?

1.3.3. Suficiencia

Una propiedad deseable en un estimador es que resuma toda la información contenida en la muestra acerca del parámetro a estimar. Calificaremos como suficientes a aquellos estadísticos que cumplan con esta propiedad, en razón de que resultan "suficientes"para nuestros propósitos de estimación. Un razonamiento intuitivo que permite intentar una primera definición de la suficiencia es el siguiente: Si la distribución condicional $f(x_1, x_2 ... x_n/S = s_o)$ no depende de θ, quiere decir que una vez observado el valor de S, el conocimiento probabilístico contenido en la densidad muestral no nos dice nada acerca del parámetro; mal podemos obtener información sobre θ a partir de una distribución que no depende de θ.

Definicion 1.12 (Estadístico suficiente)**.**
Se dice que un estadístico S es suficiente si la función de densidad condicional de la muestra dado que $S = s_o$:

$$f(x_1, x_2, \dots x_n/S = s_o)$$

no depende de θ, para todo s_o.

Ejemplo 1.11. *Sea $X_1, X_2, \dots X_n$ una muestra aleatoria de una población con distribución de poisson $P(\theta)$. Demuestre que $S = \sum X_i$ es un estadístico suficiente.*

En el ejemplo 1.1 vimos que la función de densidad muestral:

$$f_X(x_1, x_2, \dots x_n; \theta) = e^{-n\theta} \frac{\theta^{\sum x_i}}{\prod x_i!} \qquad x_i = 0, 1, \dots \quad \forall i = 1, 2 \dots n$$

Además, la variable suma $S = \sum X_i$ sigue una distribución de Poisson $P(n\theta)$, así que:

$$f_S(s; \theta) = e^{-n\theta} \frac{(n\theta)^s}{s!} \qquad s = 0, 1 \dots$$

En consecuencia, la densidad condicional:

$$f(x_1, x_2, \dots x_n/S = s_o) = \frac{f(x_1, x_2, \dots x_n; \theta)}{f_S(s_o, \theta)} = \frac{s_0!}{\prod x_i! \; n^{s_0}}$$

con $x_i = 0, 1 \dots \forall i$, tales que $\sum x_i = s_o$ y $s_o = 0, 1 \dots$, que no depende de θ, así que $S = \sum X_i$ es un estadístico suficiente.

Ejemplo 1.12. *Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con distribución de bernoulli $B(\theta)$. Demuestre que $S = \sum X_i$ es un estadístico suficiente.*

La función de densidad muestral:

$$\begin{aligned} f_X(x_1, x_2, \ldots x_n; \theta) &= \prod_{i=1}^{n} f(x_i; \theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} I_{\{0,1\}}(x_i) \\ &= \theta^{\sum x_i}(1-\theta)^{n-\sum x_i} \prod_{i=1}^{n} I_{\{0,1\}}(x_i) \end{aligned}$$

Además, la variable suma $S = \sum X_i$ sigue una distribución binomial $b(n, \theta)$, así que:

$$f_S(s; \theta) = \binom{n}{s} \theta^s (1-\theta)^{n-s} \qquad s = 0, 1 \ldots n$$

En consecuencia, la densidad condicional:

$$f(x_1, x_2, \ldots x_n / S = s_o) = \frac{f(x_1, x_2, \ldots x_n; \theta)}{f_S(s_o, \theta)} = \frac{1}{\binom{n}{s_0}}$$

con $x_i = 0, 1\ \forall i$, tales que $\sum x_i = s_o$ y $s_o = 0, 1 \ldots n$, que no depende de θ, así que $S = \sum X_i$ es un estadístico suficiente.

Ejemplo 1.13. *Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con distribución de exponencial $E(\theta)$. Demuestre que $S = \sum X_i$ es un estadístico suficiente.*

En el ejemplo 1.2 vimos que la función de densidad muestral:

$$f_X(x_1, x_2, \ldots x_n; \theta) = \theta^n e^{-\theta \sum x_i}$$

Además, la variable suma $S = \sum X_i$ sigue una distribución gamma $G(n, \theta)$, así que:

$$f_S(s; \theta) = \frac{\theta^n}{\Gamma(n)} s^{n-1} e^{-\theta s}\ I_{(0,\infty)}(s)$$

En consecuencia, la densidad condicional:

$$f(x_1, x_2, \ldots x_n / S = s_o) = \frac{f(x_1, x_2, \ldots x_n; \theta)}{f_S(s_o, \theta)} = \frac{\Gamma(n)}{s_0^{n-1}}$$

con $x_i > 0\ \forall i$, tales que $\sum x_i = s_o$ y $s_o > 0$, que no depende de θ, así que $S = \sum X_i$ es un estadístico suficiente.

Ejemplo 1.14. *Sea* $X_1, X_2, \dots X_n$ *una muestra aleatoria de una población con distribución de uniforme* $U(0, \theta)$*. Demuestre que* $S = Y_n = max\{X_i\}$ *es un estadístico suficiente.*

En el ejemplo 1.3 vimos que la función de densidad muestral:

$$f_X(x_1, x_2, \dots x_n; \theta) = \frac{1}{\theta^n} \prod_{i=1}^{n} I_{(0,\theta)}(x_i)$$

Además, la variable $S = Y_n$ tiene por función de densidad:

$$f_S(s; \theta) = \frac{ns^{n-1}}{\theta^n} \, I_{(0,\theta)}(s)$$

En consecuencia, la densidad condicional:

$$f(x_1, x_2, \dots x_n / S = s_o) = \frac{1}{ns_o^{n-1}}$$

con $x_i > 0 \ \forall i$, tales que $max\{x_i\} = s_o$ y $s_o > 0$, que no depende de θ, así que $S = Y_n$ es un estadístico suficiente.
A continuación presentaremos una definición alternativa de estadístico suficiente.

Definicion 1.13 (Estadístico suficiente)**.**
Se dice que un estadístico S *es suficiente si la función de densidad condicional de* T *dado que* $S = s_o$*:*

$$f(T/S = s_o)$$

no depende de θ*, para cualquier estadístico* T *y para todo* s_o*.*

Las definiciones 1.11 y 1.12 tienen alto valor intuitivo pero poco valor práctico. Requieren del cálculo de distribuciones condicionales, que no siempre son fáciles de obtener. Además no son definiciones “constructivas”en el sentido de que permiten verificar si un estadístico dado es suficiente, pero no permiten obtener estadísticos suficientes. Con este fin se presenta el siguiente teorema.

Teorema 1.8 (Criterio de factorización de Fisher)**.** *Sea* $X_1, X_2, \dots X_n$ *una muestra aleatoria de una población con función de densidad* $f(x; \theta)$*. Un estadístico* S *es suficiente si y solo si la densidad muestral puede factorizarse como:*

$$f(x_1, x_2, \dots x_n; \theta) = g(s; \theta) \, h(x_1, x_2, \dots x_n) \tag{1.13}$$

donde h es una función no negativa que no depende de θ, y g es no negativa y depende de las x_i sólo a través de s.

Ejemplo 1.15. *Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con distribución de poisson $P(\theta)$. Demuestre que $S = \sum X_i$ es un estadístico suficiente.*

La función de densidad muestral:

$$\begin{aligned} f_X(x_1, x_2, \ldots x_n; \theta) &= e^{-n\theta} \, \frac{\theta^{\sum x_i}}{\prod x_i!} \\ &= \big(e^{-n\theta}\theta^{\sum x_i}\big)\big(\frac{1}{\prod x_i!}\big) \\ &= g(s;\theta) \; h(x_1, x_2, \ldots x_n) \end{aligned}$$

así que $S = \sum X_i$ es un estadístico suficiente.

Ejemplo 1.16. *Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con distribución de bernoulli $B(\theta)$. Demuestre que $S = \sum X_i$ es un estadístico suficiente.*

La función de densidad muestral:

$$\begin{aligned} f_X(x_1, x_2, \ldots x_n; \theta) &= \theta^{\sum x_i}(1-\theta)^{n-\sum x_i} \\ &= \big(\theta^{\sum x_i}(1-\theta)^{n-\sum x_i}\big)\big(1\big) \\ &= g(s;\theta) \; h(x_1, x_2, \ldots x_n) \end{aligned}$$

así que $S = \sum X_i$ es un estadístico suficiente.

Ejemplo 1.17. *Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con distribución exponencial $E(\theta)$. Halle un estadístico suficiente.*

La función de densidad muestral:

$$\begin{aligned} f_X(x_1, x_2, \ldots x_n; \theta) &= \theta^n e^{\theta \sum x_i} \\ &= \big(\theta^n e^{\theta \sum x_i}\big)\big(1\big) \\ &= g(s;\theta) \; h(x_1, x_2, \ldots x_n) \end{aligned}$$

así que $S = \sum X_i$ es un estadístico suficiente.

Ejemplo 1.18. *Sea $X_1, X_2, \dots X_n$ una muestra aleatoria de una población con distribución uniforme $U(0, \theta)$. Halle un estadístico suficiente.*

La función de densidad muestral:

$$\begin{aligned} f_X(x_1, x_2, \dots x_n; \theta) &= \big(\frac{1}{\theta}\big)^n \prod_{i=1}^{n} I_{(0,\theta)}(x_i) \\ &= \big(\frac{1}{\theta^n}\, I_{(0,\theta)}(y_n)\big)\big(I_{(0,y_n)}(y_1)\big) \\ &= g(s;\theta)\ h(x_1, x_2, \dots x_n) \end{aligned}$$

así que $S = Y_n$ es un estadístico suficiente.

Teorema 1.9. *Si S es un estadístico suficiente, entonces cualquier estadístico, función uno a uno de S,$T = t(S)$, también es suficiente.*

Demostración.
Si $T = t(S) \Rightarrow S = t^{-1}(T)$
Por otro lado:
Si S es un estadístico suficiente $\Rightarrow$

$$\begin{aligned} f_X(x_1, x_2, \dots x_n; \theta) &= g(s;\theta)\ h(x_1, x_2, \dots x_n) \\ &= g(t^{-1}(t);\theta)\ h(x_1, x_2, \dots x_n) \\ &= g^*(t;\theta)\ h(x_1, x_2, \dots x_n) \end{aligned}$$

y por tanto T es suficiente. □

Teorema 1.10. *Si existe el estimador máximo verosímil de θ y es único, entonces es suficiente.*

Demostración.
La función de verosimilitud:

$$L(\theta) = \prod_{i=1}^{n} f(X_i;\theta) = g(s;\theta)\ h(x_1, x_2, \dots x_n)$$

siendo s un estadístico suficiente.
Como h no depende de θ, $L(\theta)$ alcanza su máximo en el mismo valor que $g(s;\theta)$. Como esta función depende de las x_i sólo a través de S, el valor que la maximiza dependerá de S:

$$\hat{\theta} = \hat{\theta}(S)$$

Esto quiere decir que $\hat{\theta}$ es función de un estadístico suficiente, y por el teorema 1.9, también es suficiente. □

1.3.4. Completitud

Definicion 1.14 (Estadístico completo)**.**
Se dice que un estadístico T es completo (o que su función de densidad es completa) si la única función de T con esperanza igual a cero es la función nula:

$$E[Z(t)] = 0 \Rightarrow Z(t) = 0 \quad \forall t$$

Veremos más adelante que esta propiedad junto con la suficiencia nos permitirá encontrar estimadores insesgados de varianza mínima.

Ejemplo 1.19. *Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con distribución uniforme $U(0, \theta)$. Demuestre que $T = Y_n$ es un estadístico completo.*

La función de densidad de T:

$$f_T(t) = \frac{nt^{n-1}}{\theta^n} \qquad 0 < t < \theta$$

La esperanza de $Z(T)$:

$$\begin{aligned} E[Z(T)] &= \int Z(t) f_T(t) dt = \int_0^\theta Z(t) \, \frac{nt^{n-1}}{\theta^n} \, dt \\ &= \frac{n}{\theta^n} \int_0^\theta Z(t) \, t^{n-1} \, dt \end{aligned}$$

Esta esperanza la igualamos a cero y vemos si ello implica que Z tiene que ser la función nula:

$$\begin{aligned} E[Z(T)] \quad &= 0 \quad \Rightarrow \quad \frac{n}{\theta^n} \int_0^\theta Z(t) \, t^{n-1} \, dt = 0 \quad \forall \theta \\ &\qquad\quad \Rightarrow \quad \int_0^\theta Z(t) \, t^{n-1} \, dy = 0 \quad \forall \theta \end{aligned}$$

derivando ahora con respecto a θ (asumiendo Z continua):

$$Z(\theta) \theta^{n-1} = 0 \quad \forall \theta$$

de donde:

$$Z(\theta) = 0 \quad \forall \theta$$

y por tanto $T = Y_n$ es un estadístico completo.

De la definición 1.14 se desprende el siguiente teorema:

Teorema 1.11. *Si T es un estadístico completo, entonces cualquier función uno a uno de T, $V = v(T)$, también es completo.*

Ejemplo 1.20. *Sea $X_1, X_2, \dots X_n$ una muestra aleatoria de una población con distribución de bernoulli $B(\theta)$. Demuestre que la media muestral es un estadístico completo.*

La función de densidad de $T = \sum X_i$:

$$f_T(t) = \binom{n}{t}\theta^t(1-\theta)^{n-t} \quad t = 0, 1 \dots n \quad 0 < \theta < 1$$

La esperanza de $Z(T)$:

$$\begin{aligned} E[Z(T)] &= \sum Z(t) f_T(t) = \sum_{t=0}^{n} Z(t)\binom{n}{t}\theta^t(1-\theta)^{n-t} \\ &= (1-\theta)^n \sum_{t=0}^{n} Z(t)\binom{n}{t}\left(\frac{\theta}{1-\theta}\right)^t \\ &= (1-\theta)^n \sum a_t k^t \end{aligned}$$

Si esta esperanza es cero $\forall\theta$, entonces:

$$\sum a_t k^t \quad = 0 \quad \forall k \quad \Rightarrow \quad a_t = 0 \quad \forall t$$

así que:

$$Z(t)\binom{n}{t} \quad = 0 \quad \forall t \quad \Rightarrow \quad Z(t) = 0 \quad \forall t$$

y por tanto T es un estadístico completo. Como la media muestral es una función uno a uno de T, también es un estadístico completo.

1.3.5. Eficiencia

Definicion 1.15 (Estimador eficiente)**.**
Se dice que un estimador T es eficiente con respecto a otro estimador W (ambos del mismo parámetro), si se cumple que $Var(T) \leq Var(W)$.

Definicion 1.16 (Eficiencia relativa)**.**
Se define como eficiencia relativa de un estimador T con respecto a otro estimador W (ambos del mismo parámetro), al cociente:

$$EFF(T/W) = \frac{Var(T)}{Var(W)}$$

Al comparar esta eficiencia con 1 se determina cuál de los dos estimadores tiene menor varianza.

Ejemplo 1.21. *Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con distribución uniforme $U(\theta - \frac{1}{2}, \theta + \frac{1}{2})$, y sea el estadístico $T = \frac{Y_1+Y_n}{2}$. Demuestre que:*

$$EFF(T/\bar{X}) = \frac{6n}{(n+1)(n+2)}$$

Por una parte:

$$Var(\bar{X}) = \frac{Var(X)}{n} = \frac{1}{12n}$$

Se demuestra además que:

$$Var(T) = \frac{1}{2(n+1)(n+2)}$$

Por lo tanto:

$$EFF(T/\bar{X}) = \frac{Var(T)}{Var(\bar{X})} = \frac{6n}{(n+1)(n+2)}$$

que es una función decreciente en n, y menor que 1 para $n \geq 2$. En consecuencia, en el caso de la distribución uniforme, el centro-recorrido T es más eficiente que la media muetral.

1.4. La clase exponencial de distribuciones

En este apartado vamos a presentar una familia de densidades que cumpl un papel muy importante en la inferencia estadística. Esta familia garantiza la existencia de estimadores con buenas propiedades.

Definicion 1.17 (Clase exponencial)**.** *Se dice que una familia de densidades $\{f(x;\theta)\}$ pertenece a la clase exponencial, si puede expresarse como:*

$$f(x;\theta) = a(\theta)b(x)e^{c(\theta)d(x)} \quad \forall x,\ \forall\theta \tag{1.14}$$

donde a, b, c y d son funciones convenientemente elegidas.

Ejemplo 1.22. *La distribución de poisson $P(\theta)$ pertenece a la clase exponencial.*

En efecto, como:

$$\begin{aligned} f(x;\theta) &= e^{-\theta}\ \frac{\theta^x}{x!}\ I_{\{0,1\dots\}}(x) \\ &= e^{-\theta}\ \frac{1}{x!}\ I_{\{0,1\dots\}}(x)\ \theta^x \\ &= (e^{-\theta})(\frac{1}{x!}\ I_{\{0,1\dots\}}(x))e^{(log\theta)(x)} \\ &= a(\theta)b(x)e^{c(\theta)d(x)} \end{aligned}$$

entonces la distribución de poisson pertenece a la clase exponencial.

Ejemplo 1.23. *La distribución exponencial $E(\theta)$ pertenece a la clase exponencial.*

En efecto, como:

$$\begin{aligned} f(x;\theta) &= \theta e^{-\theta x}\ I_{(0,\infty)}(x) \\ &= (\theta)(I_{(0,\infty)}(x))e^{(-\theta)(x)} \\ &= a(\theta)b(x)e^{c(\theta)d(x)} \end{aligned}$$

entonces la distribución exponencial pertenece a la clase exponencial.

El siguiente teorema nos muestra la forma general de la media y la varianza de la variable aleatoria $d(X)$:

Teorema 1.12. *Si $f(x;\theta)$ pertenece a la clase exponencial y $p(\theta) = \log a(\theta)$, entonces:*

$$\begin{aligned} i)\ \frac{d}{d\theta}\ f(x;\theta) &= f(x;\theta)(c'(\theta)d(x) + p'(\theta)) \\ ii)\ E[d(X)] &= -\frac{p'(\theta)}{c'(\theta)} \\ iii)\ \frac{d}{d\theta}E[d(X)] &= c'(\theta)Var[d(X)] \\ iv)\ Var[d(X)] &= \frac{c''\theta)p'(\theta) - c'(\theta)p''(\theta)}{(c'(\theta))^3} \end{aligned}$$

Demostración.

i) La función de densidad puede escribirse en la forma:

$$f(x;\theta) = a(\theta)b(x)e^{c(\theta)d(x)} = b(x)e^{c(\theta)d(x)+p(\theta)}$$

de donde:

$$\frac{d}{d\theta}\ f(x;\theta) = f(x;\theta)\ (c'(\theta)d(x) + p'(\theta))$$

ii) Por ser f una función de densidad:

$$\int f(x;\theta) = 1$$

si derivamos respecto de θ:

$$\int f(x;\theta)\ (c'(\theta)d(x) + p'(\theta))dx = 0$$

de donde:

$$\begin{aligned} 0 &= c'(\theta)\int f(x;\theta)d(x)dx + p'(\theta)\int f(x;\theta)dx \\ &= c'(\theta)E[d(X)] + p'(\theta) \end{aligned}$$

así que:

$$E[d(X)] = -\frac{p'(\theta)}{c'(\theta)}$$

iii) Por definición:

$$E[d(X)] = \int d(x)\ f(x;\theta)dx$$

derivando respecto de θ:

$$\begin{aligned} \frac{d}{d\theta}E[d(X)] &= \int d(x)\ f(x;\theta)(c'(\theta)d(x) + p'(\theta))dx \\ &= c'(\theta)\int (d(x))^2 f(x;\theta)dx + p'(\theta)\int d(x)\ f(x;\theta)dx \\ &= c'(\theta)E[(d(X))^2] + p'(\theta)E[d(X)] \\ &= c'(\theta)\{E[(d(X))^2] + \frac{p'(\theta)}{c'(\theta)}E[d(X)]\} \\ &= c'(\theta)\{E[(d(X))^2] - (E[d(X)])^2\} \\ &= c'(\theta)Var[d(X)] \end{aligned}$$

iv) Se desprende directamente de ii y iii.

□

Teorema 1.13. *Sea $X_1, X_2, \dots X_n$ una muestra aleatoria de una población con función de densidad perteneciente a la clase exponencial, entonces el estadístico $S = \sum d(X_i)$ es suficiente y completo.*

Demostración. (solamente la suficiencia)
La función de densidad muestral:

$$\begin{aligned} f_X(x_1, x_2, \dots x_n; \theta) &= \prod_{i=1}^{n} f(x_i; \theta) = \prod_{i=1}^{n} [a(\theta) b(x_i) e^{c(\theta) d(x_i)}] \\ &= [a(\theta)]^n [\prod b(x_i)] \ e^{c(\theta) \sum d(x_i)} \\ &= [a(\theta)]^n e^{c(\theta) \sum d(x_i)}] \quad [\prod b(x_i)] \\ &= g(s; \theta) \ h(x_1, x_2 \dots x_n) \end{aligned}$$

donde $S = \sum d(x_i)$, así que S es un estadístico suficiente. □

Teorema 1.14. *Sea $X_1, X_2, \dots X_n$ una muestra aleatoria de una población con función de densidad perteneciente a la clase exponencial, y sea $\tau(\theta) = -p'(\theta)/c'(\theta)$. Entonces:*

i) El estimador máximo verosímil de $\tau(\theta)$ es $T = \sum d(X_i)/n$.

ii) T es insesgado, consistente, suficiente y completo.

Demostración.

i) La función de verosimilitud:

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} f(x_i; \theta) = \prod_{i=1}^{n} b(x_i) e^{c(\theta) d(x_i) + p(\theta)} \\ &= [\prod_{i=1}^{n} b(x_i)] e^{np(\theta) + c(\theta) \sum d(x_i)} \end{aligned}$$

tomando logaritmo:

$$l(\theta) = log[\prod_{i=1}^{n} b(x_i)] + np(\theta) + c(\theta)\sum d(x_i)$$

derivando respecto de θ (asumiento la diferenciabilidad de $p(\theta)$ y $c(\theta)$:

$$\frac{d}{d\theta}l(\theta) = np'(\theta) + c'(\theta)\sum d(x_i)$$

igualando a cero se obtiene:

$$\hat{\tau}(\theta) = \frac{\sum d(x_i)}{n} = T$$

ii) Por el teorema 1.12:

$$E(T) = E\big[\frac{\sum d(x_i)}{n}\big] = \frac{1}{n}E\big[\sum d(x_i)\big] = E[d(X)] = \tau(\theta)$$

y además:

$$Var(T) = Var\big[\frac{\sum d(x_i)}{n}\big] = \frac{1}{n^2}\sum Var\big[d(x_i)\big] = \frac{1}{n}Var[d(X)]$$

así que $lim_{n\to\infty}Var[d(X)] = 0$, y por lo tanto T es insesgado y consistente.
Por el teorema 1.13, $S = \sum d(X_i)$ es suficiente y completo, y en consecuencia T = S/n también es suficiente y completo.

□

1.5. Estimadores insesgados de mínima varianza

De acuerdo con las propiedades estudiadas, mantendremos el criterio de que el mejor estimador T de un determinado parámetro $\tau(\theta)$ es aquél que sea insesgado, consistente, suficiente, completo y eficiente. Nuestro procedimiento de búsqueda del mejor estimador consistirá en hallar un estimador insesgado y eficiente (de mínima varianza), y luego evaluar las propiedades restantes.

Definicion 1.18 (Estimador insesgado de mínima varianza)**.** *Sea* $X_1, X_2, \ldots X_n$ *una muestra aleatoria de* $f(x;\theta)$. *Se dice que* T^* *es un estimador de mínima varianza (eimv) para* $\tau(\theta)$ *si:*

i) $E(T^*) = \tau(\theta)$.

ii) $Var(T^*) \leq Var(T) \quad \forall T$ *insesgado para* $\tau(\theta)$.

Si este estimador resulta consistente, suficiente y completo, lo consideraremos como el "mejor"estimador de $\tau(\theta)$.

1.5.1. Cota de Cramer-Rao

Una manera de obtener estimadores insesgados de mínima varianza consiste en establecer una cota inferior para la varianza de los estimadores insesgados, y concluir que si un determinado estimador insesgado tiene una varianza igual a esta cota, entonces es el estimador de mínima varianza. Esta cota inferior se conoce como cota de Cramer-Rao.

Antes de enunciar y demostrar el teorema de Cramer-Rao, recordaremos la conocida desigualdad de Cauchy-Schwarz:

Proposicion 1.1 (Desigualdad de Cauchy-Schwarz)**.** *Sean X e Y variables aleatorias cuyas medias y varianzas existen, entonces:*

$$Cov^2(X, Y) \leq Var(X)Var(Y)$$

cumpliéndose la igualdad si y solo si Y-E(Y) = k (X - E(X)). (De este resultado proviene la definición de coeficiente de correlación).

Teorema 1.15 (Cota de Cramer-Rao)**.** *Sea* $X_1, X_2, \ldots X_n$ *una muestra aleatoria de* $f(x;\theta)$, *la cual satisface las siguientes condiciones de regularidad:*

i) Existe la variable aleatoria:

$$\begin{aligned} V &= \frac{\partial}{\partial\theta} log \prod f(x_i;\theta) = \frac{1}{\prod f(x_i;\theta)} \frac{\partial}{\partial\theta} \prod f(x_i;\theta) \\ &= \sum \frac{\partial}{\partial\theta} log f(x_i;\theta) \end{aligned}$$

su esperanza y su varianza.

ii) Se cumple que:

$$\frac{\partial}{\partial\theta}\int\prod f(x_i;\theta)\,dx_1dx_2\dots dx_n =$$
$$\int\frac{\partial}{\partial\theta}\prod f(x_i;\theta)\,dx_1dx_2\dots dx_n$$

iii) Se cumple que:

$$\frac{\partial}{\partial\theta}\int t(x_1,x_2\dots x_n)\prod f(x_i;\theta)\,dx_1dx_2\dots dx_n =$$
$$\int t(x_1,x_2\dots x_n)\frac{\partial}{\partial\theta}\prod f(x_i;\theta)\,dx_1dx_2\dots dx_n$$

Entonces, para todo estimador insesgado T de $\tau(\theta)$ se tiene que:

$$Var(T)\geq\frac{(\tau(\theta))^2}{n\,E[(\frac{\partial}{\partial\theta}log\,f(x;\theta))^2]} = CCR$$

cumpliéndose la igualdad si y solo si existe una función $k(\theta,n)$ tal que:

$$\sum\frac{\partial}{\partial\theta}log f(x_i;\theta) = k(\theta,n)\,(T-\tau(\theta))$$

Demostración.
La esperanza de V:

$$\begin{aligned}
E(V) &= \int v\prod f(x_i;\theta)\,dx_1dx_2\dots dx_n\\
&= \int(\frac{1}{\prod f(x_i;\theta)}\frac{\partial}{\partial\theta}\prod f(x_i;\theta))\prod f(x_i;\theta)\,dx_1dx_2\dots dx_n\\
&= \int\frac{\partial}{\partial\theta}\prod f(x_i;\theta)\,dx_1dx_2\dots dx_n\\
&= \frac{\partial}{\partial\theta}\int\prod f(x_i;\theta)\,dx_1dx_2\dots dx_n\\
&= 0
\end{aligned}$$

Su varianza:

$$
\begin{aligned}
Var(V) &= E[V^2] \\
&= E[\big(\sum \frac{\partial}{\partial\theta} log f(x_i;\theta)\big)^2] \\
&= E[\sum(\frac{\partial}{\partial\theta} log f(x_i;\theta))^2] + E[\sum\sum(\frac{\partial}{\partial\theta} log f(x_i;\theta))(\frac{\partial}{\partial\theta} log f(x_j;\theta))] \\
&= nE[(\frac{\partial}{\partial\theta} log f(x;\theta))^2] + \sum\sum Cov(\frac{\partial}{\partial\theta} log f(x_i;\theta), \frac{\partial}{\partial\theta} log f(x_j;\theta)) \\
&= nE[(\frac{\partial}{\partial\theta} log f(x;\theta))^2]
\end{aligned}
$$

Además la covarianza:

$$
\begin{aligned}
Cov(T,V) &= E[TV] \\
&= \int t(x_1 \dots x_n)(\frac{1}{\prod f(x_i;\theta)}\, \frac{\partial}{\partial\theta} \prod f(x_i;\theta)) \prod f(x_i;\theta)\, dx_1 \dots dx_n \\
&= \int t(x_1 \dots x_n)\, \frac{\partial}{\partial\theta} \prod f(x_i;\theta)\, dx_1 \dots dx_n \\
&= \frac{\partial}{\partial\theta} \int t(x_1 \dots x_n) \prod f(x_i;\theta)\, dx_1 \dots dx_n \\
&= \frac{\partial}{\partial\theta}\, [\tau(\theta)] = [\tau'(\theta)]
\end{aligned}
$$

De acuerdo con la proposición 1.1:

$$
\begin{aligned}
&Cov^2(T,V) \leq Var(T)Var(V) \quad \Rightarrow \\
&(\tau'(\theta))^2 \leq Var(T)\, nE[(\frac{\partial}{\partial\theta} log f(x;\theta))^2]
\end{aligned}
$$

de donde:

$$
Var(T) \geq \frac{(\tau'(\theta))^2}{nE[(\frac{\partial}{\partial\theta} log f(x;\theta))^2]} = CCR
$$

Además, se cumple la igualdad si y solo si existe una función $k(\theta,n)$ tal que:

$$
V - E(V) = k(\theta,n)(T - E(T))
$$

o sea si:

$$
\sum \frac{\partial}{\partial\theta} log f(x_i;\theta) = k(\theta,n)(T - \tau(\theta)) \tag{1.15}
$$

□

Corolario 1.15.1. *Si se cumple (1.16) entonces Var(T) = CCR y por tanto T es el eimv para $\tau(\theta)$.*

El siguiente teorema afirma que las condiciones de regularidad del teorema 1.15 se cumplen únicamente en el caso de que $f(x;\theta)$ pertenezca a la familia exponencial.

Teorema 1.16. *Una función de densidad $f(x;\theta)$ satisface las condiciones de regularidad si y solo si pertenece a la clase exponencial.*

Esto quiere decir que la búsqueda de estimadores insesgados de mínima varianza por medio de la cota de Cramer-Rao está restringida al caso de poblacioines con función de densidad perteneciente a la familia exponencial.

Corolario 1.16.1. *Si $f(x;\theta)$ pertenece a la clase exponencial, entonces $T = \sum d(X_i)/n$ es el eimv de $\tau(\theta)$).*

En efecto, en la demostración del teorema 1.14 se encontró que:

$$\frac{\partial}{\partial\theta} log\, f(x_i;\theta) = p'(\theta) + c'(\theta)d(x_i)$$

luego:

$$\begin{aligned}\sum \frac{\partial}{\partial\theta} log\, f(x_i;\theta) &= np'(\theta) + c'(\theta)\sum d(x_i)\\ &= nc'(\theta)\Big[\frac{\sum d(x_i)}{n} - (-\frac{p'(\theta)}{c'(\theta)}\Big]\end{aligned}$$

así que T es el eimv de $\tau(\theta)$.

1.5.2. Suficiencia y completitud

En este apartado veremos como combinar las propiedades de suficiencia y completitud para establecer un procedimiento de búsqueda de estimadores de mínima varianza.

Teorema 1.17 (Rao - Blackwell)**.** *Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de $f(x;\theta)$. Sea S un estadístico suficiente y T un estimador insesgado de $\tau(\theta)$. Entonces[2]:*

[2] Los apartados ii y iii son válidos aún cuando S no sea suficiente.

i) $T' = E(T/S)$ *es un estadístico*

ii) $E(T') = \tau(\theta)$

iii) $Var(T') \leq Var(T)$

Demostración.

i) Si S es un estadístico suficiente $\Rightarrow$ f(t/s) no depende de $\theta \Rightarrow$
$T' = E(T/S)$ no depende de $\theta \Rightarrow T'$ es un estadístico.

ii) Si T es un estimador insesgado de $\tau(\theta)$:
$E(T') = E(E(T/S)) = E(T) = \tau(\theta)$

iii) Como:

$$\begin{aligned} Var(T) &= E(Var(T/S)) + Var(E(T/S)) \\ &= E(Var(T/S)) + Var(T') \end{aligned}$$

entonces:
$Var(T) - Var(T') = E(Var(T/S)) \geq 0$
de donde:
$Var(T') \leq Var(T)$

$\square$

Este teorema afirma que si se tiene un estimador insesgado T y un estadístico suficiente S, estos pueden combinarse para hallar un estimador insesgado T' mejor que T ya que su varianza es menor o igual. El siguiente teorema plantea que si S además de ser suficiente también es completo, entonces al combinarlo con T se halla no solo un estimador insesgado mejor que T, sino el estimador insesgado de mínima varianza.

Teorema 1.18 (Lehmann-Scheffe)**.** *Sea* $X_1, X_2, \ldots X_n$ *una muestra aleatoria de* $f(x;\theta)$*. Sea* S *un estadístico suficiente y completo, y* $T^* = t^*(S)$ *un estimador insesgado de* $\tau(\theta)$*, función de* S*. Entonces* T^* *es un eimv para* $\tau(\theta)$*.*

Demostración.
Si S es suficiente y T es insesgado, entonces por el teorema anterior tenemos que $T' = E(T/S)$ es insesgado para $\tau(\theta)$ y además $Var(T') \leq Var(T)$.
Repitiendo el mismo argumento podemos encontrar otro estimador insesgado $T'' = E(T'/S)$ tal que $Var(T'') \leq Var(T')$.
Podemos esperar entonces que al aplicar sucesivamente este razonamiento un cierto número de veces encontraríamos el estimador insesgado de mínima varianza. Sin embargo, si consideramos que:

$$T' = E(T/S) = g(S) \text{ y } E(T') = \tau(\theta)$$
$$T'' = E(T'/S) = h(S) \text{ y } E(T'') = \tau(\theta)$$

y denotamos:

$$Z(S) = g(S) - h(S)$$

entonces:

$$E[Z(S)] = E[g(S) - h(S)] = 0$$

y como S es completo, esto implica que $Z(S) = 0$ y por tanto $T = T'$.
Es decir, al aplicar una sola vez el teorema de Rao-Blackwell se obtiene el estimador insesgado de mínima varianza, y en las sucesivas repeticiones se obtiene el mismo estimador. □

Es importante anotar que en general no hace falta hallar un estimador insesgado T y luego calcular la esperanza condicional $T^* = E(T/S)$, sino que bastará con obtener un estadístico $T^* = t^*(S)$, función de S, que sea insesgado.

El siguiente teorema es una consecuencia directa de los teoremas 1.14 y 1.18:

Teorema 1.19. *Si $f(x;\theta)$ pertenece a la clase exponencial entonces $T = \sum d(X_i)/n$ es el estimador insesgado de mínima varianza de $\tau(\theta) = -p'(\theta)/c'(\theta)$.*

Ejemplo 1.24. *Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con distribución exponencial $E(\theta)$. Halle el eimv para θ.*

Sabemos ya que $f(x;\theta)$ pertenece a la clase exponencial, que la suma $S = \sum X_i$ es un estadístico suficiente y completo, que la media muestral es el eimv para $1/\theta$ y que su varianza alcanza la cota de Cramer-Rao.

Para hallar el eimv de θ tenemos que encontrar un estimador insesgado para θ que sea función de S.

Como sabemos que $E(S/n) = 1/\theta$, probamos entonces con n/S:

$$\begin{aligned} E(n/S) &= \int_{-\infty}^{\infty} (n/s)\ f_S(s) ds = \int_0^{\infty} \frac{n}{s}\ \big(\frac{\theta^n}{\Gamma(n)}\ s^{n-1}\ e^{-\theta s}\big) ds \\ &= \frac{n\theta^n}{\Gamma(n)} \int_0^{\infty} s^{n-1}\ e^{-\theta s} ds = \frac{n}{n-1}\ \theta \end{aligned}$$

Esto quiere decir que:

$$T^* = \frac{n-1}{n}\ \frac{n}{S} = \frac{n-1}{S}$$

es el estimador insesgado de mínima varianza para θ, ya que es un estimador insesgado, y función de otro que es suficiente y completo.

Ejemplo 1.25. *Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con distribución uniforme $U(0, \theta)$. Halle el eimv para θ.*

Sabemos ya que Y_n es un estadístico suficiente y completo, y que su esperanza es igual a $\frac{n}{n+1}\ \theta$. Entonces $T^* = \frac{n+1}{n}\ Y_n$ es el eimv para θ, ya que es un estimador insesgado, y función de otro que es suficiente y completo.

1.6. Propiedad asintótica de los estimadores máximo verosímiles

Teorema 1.20. *Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de $f(x;\theta)$, la cual satisface las condiciones de regularidad, y sea $T = \sum d(X_i)/n$ el esimador máximo verosímil de $\tau(\theta) = -p'(\theta)/c'(\theta)$. Entonces la sucesión de variables aleatorias $\{Z_n\}$ definidas por:*

$$Z_n = \frac{T_n - \tau(\theta)}{\sigma_n(\theta)}$$

converge en distribución a la variable $Z \sim N(0, 1)$, siendo:

$$\sigma_n^2(\theta) = \frac{(\tau(\theta))^2}{n\, E[(\frac{\partial}{\partial\theta} log\, f(x;\theta))^2]} = CCR$$

Ejercicios 1.2

1. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de cada una de las siguientes poblaciones. Halle estadísticos suficientes utilizando el criterio de factorización:

i) bernoulli $B(\theta)$	v) beta $B(\theta,1)$
ii) poisson $P(\theta)$	vi) normal $N(\theta,1)$
iii) geometrica $G(\theta)$	vii) normal $N(0,\theta)$
iv) exponencial $E(\theta)$	viii) uniforme $U(0,\theta)$

2. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de cada una de las siguientes poblaciones. Halle estadísticos suficiente utilizando el criterio de factorización:

$$\text{i) } f(x;\theta) = \theta\ (1+x)^{-(\theta+1)}\ I_{(0,\infty)}(x) \qquad \theta > 0$$
$$\text{ii) } f(x;\theta) = e^{-(x-\theta)}\ I_{(\theta,\infty)}(x)$$
$$\text{iii) } f(x;\theta) = \frac{2x}{\theta^2}\ I_{(0,\theta)}(x) \qquad \theta > 0$$
$$\text{iv) } f(x;\theta) = \frac{\theta}{x^2}\ I_{(\theta,\infty)}(x) \qquad \theta > 0$$
$$\text{v) } f(x;\theta) = \frac{\theta(log(1/\theta))^x}{x!}\ I_{\{0,1\ldots\}}(x) \qquad \theta > 0$$
$$\text{vi) } f(x;\theta) = \frac{1}{x\sqrt{2\pi}} e^{-(logx-\theta)^2/2}\ I_{(0,\infty)}(x)$$

3. Sea $f(x;\theta)$ una función de densidad perteneneciente a la clase exponencial. Demuestre que:

$$\text{E[d(x)]} = -\,\frac{p'(\theta)}{c'(\theta)} \quad \text{y} \quad \text{Var[d(x)]} \;=\; \frac{c''\theta)p'(\theta) - c'(\theta)p''(\theta)}{(c'(\theta))^3}$$

4. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una función de densidad perteneneciente a la clase exponencial, y sea $\tau(\theta) = -p'(\theta)/c'(\theta)$. Demuestre que:

   i) El estimador máximo verosímil de $\tau(\theta)$ es $T = \sum d(X_i)/n$

   ii) T es insesgado, consistente y suficiente.

5. Demuestre que las siguientes densidades pertenecen a la clase exponencial:

i) bernoulli $B(\theta)$	v) beta $B(\theta, 1)$
ii) poisson $P(\theta)$	vi) normal $N(\theta, 1)$
iii) geometrica $G(\theta)$	vii) normal $N(0, \theta)$
iv) exponencial $E(\theta)$	viii) gamma $G(r, \theta)$ (r conocido)

6. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con función de densidad:
$$f(x;\theta) = e^{-(x-\theta)} e^{-e^{-(x-\theta)}}$$
Halle el estimador máximo verosímil de $e^{-\theta}$. ¿Es insesgado?. ¿Es consistente?. ¿Es suficiente?. ¿Es completo?

7. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con función de densidad:
$$f(x;\theta) = 2\theta x e^{-\theta x^2} I_{(0,\infty)}(x) \qquad \theta > 0$$
Halle el estimador máximo verosímil de $1/\theta$. ¿Es insesgado?. ¿Es consistente?. ¿Es suficiente?. ¿Es completo?

8. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con función de densidad:
$$f(x;\theta) = \frac{1}{\theta}\, x^{\frac{1-\theta}{\theta}} I_{(0,1)}(x) \qquad \theta > 0$$
Halle el estimador máximo verosímil de θ. ¿Es insesgado?. ¿Es consistente?. ¿Es suficiente?. ¿Es completo?

9. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con función de densidad:
$$f(x;\theta) = \left(\frac{\theta}{2}\right)^{|x|} (1-\theta)^{1-|x|} I_{\{-1,0,1\}}(x) \quad 0 < \theta < 1$$
Halle el estimador máximo verosímil de θ. ¿Es insesgado?. ¿Es consistente?. ¿Es suficiente?. ¿Pertenece $f(x;\theta)$ a la clase exponencial?

10. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población uniforme $U(\theta - 1/2, \theta + 1/2)$ y sea el estimador $T = (Y_1 + Y_n)/2$. Demuestre que:

$$EFF(T/\bar{x}) = \frac{6n}{(n+1)(n+2)}$$

11. Sea $f(x;\theta)$ una función d densidad que cumple con las condiciones de regularidad. Demuestre que:

$$E\Big[\big(\frac{\partial}{\partial\theta}\, log\, f(x;\theta)\big)^2\Big] = -E\Big[\frac{\partial^2}{\partial\theta^2}\, log\, f(x;\theta)\Big]$$

12. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de $f(x;\theta)$. Sea S un estadístico suficiente y T un estimador insesgado para θ. Demuestre que:

    i) $T' = E(T/S)$ es un estimador insesgado para θ

    ii) $Var(T') \leq Var(T)$

13. Sea X_1, X_2 una muestra aleatoria de una población exponencial $E(1/\theta)$. Sea los estadísticos $S = X_1 + X_2$ y $T = X_2$. Demuestre que:

    i) S es suficiente y T es insesgado para θ

    ii) $T' = E(T/S)$ es insesgado para θ

    iii) $Var(T') \leq Var(T)$

14. Sea X_1, X_2 una muestra aleatoria de una población exponencial $E(1/\theta)$. Sea Y_1, Y_2 la muestra ordenada y $V = E(Y_2/Y_1)$. Demuestre que:

    i) $E(V) = E(Y_2)$

    ii) $Var(V) \leq Var(Y_2)$

    Comente.

15. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con función de densidad perteneciente a la Clase Exponencial. Demuestre que $T = \sum d(X_i)/n$ es un estimador insesgado de mínima varianza para $\tau(\theta) = -p'(\theta)/c'(\theta)$.

16. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población uniforme $U(0,\theta)$. ¿Es el estimador $T = \frac{n+1}{n} Y_n$ insesgado y de mínima varianza para θ?. Compare su varianza con la CCR.

17. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de cada una de las siguientes poblaciones. Halle el parámetro $\tau(\theta)$ para el cual existe un eimv cuya varianza alcanza la CCR.

i) bernoulli $B(\theta)$	v) beta $B(\theta, 1)$
ii) poisson $P(\theta)$	vi) normal $N(\theta, 1)$
iii) geometrica $G(\theta)$	vii) normal $N(0, \theta)$
iv) exponencial $E(\theta)$	viii) gamma $G(r, \theta)$ (r conocido)

18. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de cada una de las siguientes poblaciones. Halle el eimv para los parámetros que se indican.

i)	bernoulli $B(\theta)$	θ, $1-\theta$, $\theta(1-\theta)$, θ^2
ii)	poisson $P(\theta)$	θ
iii)	geometrica $G(\theta)$	$(1-\theta)/\theta$
iv)	exponencial $E(\theta)$	θ, $1/\theta$
v)	beta $B(\theta, 1)$	θ, $1/\theta$
vi)	normal $N(\theta, 1)$	θ,θ^2
vii)	normal $N(0, \theta)$	θ,$\sqrt{\theta}$
viii)	uniforme $U(0, \theta)$	θ
ix)	$f(x;\theta) = \theta(1+x)^{-(\theta+1)}\ I_{(0,\infty)}(x)$	θ, $1/\theta$
x)	$f(x;\theta) = e^{-(x-\theta)}\ I_{(\theta,\infty)}(x)$	θ
xi)	$f(x;\theta) = 2x/\theta^2\ I_{(0,\theta)}(x)$	θ
xii)	$f(x;\theta) = \theta/x^2\ I_{(\theta,\infty)}(x)$	θ
xiii)	$f(x;\theta) = 2\theta x e^{-\theta x^2}\ I_{(0,\infty)}(x)$	$1/\theta, \theta, 1/\sqrt{\theta}, \sqrt{\theta}$
xiv)	$f(x;\theta) = \frac{\theta}{1-\theta}\ x^{(2\theta-1)/(1-\theta)}\ I_{(0,1)}(x)$	$1/\theta$
xv)	$f(x;\theta) = e^{-(x-\theta)}e^{-e^{-(x-\theta)}}$	$e^{-\theta}$

19. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una población con función de densidad $f(x;\theta)$. Sea $\tau(\theta) = P(X \in A)$ y S un estadístico suficiente y completo. Demuestre que:

   i) $T = I_A(X_i)$ es un estimador insesgado para $\tau(\theta)$, para cualquier i.

   ii) $T^* = P(X_i \in A/S)$ es un eimv para $\tau(\theta)$.

20. Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de cada una de las siguientes poblaciones. Halle el eimv para los parámetros que se indican.

i)	bernoulli $B(\theta)$	θ
ii)	poisson $P(\theta)$	$e^{-\theta}, \theta e^{-\theta}, e^{-2\theta}$
iii)	geometrica $G(\theta)$	θ
iv)	exponencial $E(\theta)$	$e^{-k\theta} \quad (k > 0)$