Inferencia

Estadística

(Teoría y problemas)

I. Espejo Miranda
F. Fernández Palacín
M. A. López Sánchez
M. Muñoz Márquez
A. M. Rodríguez Chía
A. Sánchez Navas
C. Valero Franco

Edita: Servicio de Publicaciones de la Universidad de Cádiz
c/ Doctor Marañón, 3. 11002 Cádiz (España)
www.uca.es/publicaciones

ISBN: 978-84-9828-131-6

Inferencia Estadística (Revisión: Marzo 2007)
I. Espejo Miranda, F. Fernández Palacín, M. A. López Sánchez, M. Muñoz Márquez, A. M. Rodríguez Chía, A. Sánchez Navas, C. Valero Franco

`http://www.uca.es/teloydisren`

Capítulo 2

Estimación puntual

1. Introducción

En numerosas ocasiones, al realizar un estudio estadístico se conoce la estructura de la población que se pretende estudiar, con la salvedad de los parámetros que la caracterizan. Por ejemplo, la utilización de un aparato de medida objetivo garantiza, en general, que las mediciones obtenidas tendrán una distribución Normal, de la que se desconocerán sus parámetros: media y desviación típica. El objetivo que se persigue con las técnicas de estimación es el determinar de la forma más precisa dichos parámetros, de modo que la distribución quede completamente especificada.

En este capítulo y en los dos siguientes se abordará la Inferencia Estadística desde un punto de vista paramétrico, es decir, se parte del conocimiento (salvo parámetros) de la distribución de probabilidad que rige la población bajo estudio. De esta forma, se considera una población cuya función de distribución es $F_{\underline{\theta}}(x)$, donde $\underline{\theta} \in \mathbb{R}^k$ es un vector de parámetros desconocidos. En esta situación, el problema es cuantificar lo más exactamente posible el valor de $\underline{\theta}$ a partir de una muestra de tamaño n. La rama de la Estadística que se dedica a estudiar este tipo de problemas se llama *Teoría de la Estimación*, existiendo dos enfoques diferentes para llevar a cabo dicho estudio: la *estimación puntual* y la *estimación por intervalos*. En la primera, se estiman los parámetros a través de valores numéricos, mientras que en la segunda, queda garantizada su pertenencia a una región con un margen de seguridad prefijado.

Este capítulo se centra en la *estimación puntual*, si bien, la mayoría de los conceptos son generales y se utilizan también en la *estimación por intervalos* y en el *contraste de hipótesis*.

A efectos de notación se hará referencia a las características de la muestra con letras latinas, mientras que las de la población se designarán, en general, con la correspondiente letra griega. Así, por ejemplo, la varianza muestral será S^2, mientras que la poblacional se identificará por σ^2; con la media muestral seguirá utilizándose la notación usual, $\overline{X}$, mientras que la poblacional se denotará por μ. El objetivo que se perseguirá a lo largo del capítulo es el de obtener valores lo más precisos posibles de los parámetros desconocidos del modelo probabilístico.

2. Estadístico, Estimador y Estimación

Un *estadístico* $T(\underline{X})$, es una función de las variables muestrales que no depende de parámetros desconocidos. Se trata pues de una variable aleatoria, la cual tiene una distribución que se denomina distribución en el muestreo. El estadístico puede considerarse como un resumen o una compresión de la información suministrada por la muestra y, obviamente, va a ser más manejable que ésta. Nótese que puede ocurrir que en ese resumen se pierda alguna posible información que pudiera contener $\underline{X}$ acerca de los parámetros desconocidos. Por ello, el objetivo perseguido es que el estadístico $T(\underline{X})$ sea tal que el resumen que lleve a cabo se produzca sin pérdida de información relevante sobre los parámetros.

Dentro del conjunto de estadísticos destacan los *estimadores*, que son aquellos estadísticos que se construyen con la intención de estimar un parámetro de la población y que, consecuentemente, debe reunir condiciones que lo hagan deseable en algún sentido. Más adelante se darán criterios de bondad de un estimador.

Una *estimación* es el valor numérico que toma el estimador para una muestra concreta.

Ejemplo 2.1 *Sea X una variable aleatoria que sigue una distribución Normal de media desconocida, μ, y varianza σ^2. La función $T(\underline{X}) = \overline{X}$, es decir, la media muestral, es un estadístico y estimador de la media μ de la población. Si se toma la muestra $x_1 = 2'5$, $x_2 = 2$, $x_3 = 3'4$, $x_4 = 1'5$, $x_5 = 4$, el valor numérico $\overline{x} = 2'68$ es una estimación de μ.*

La necesidad de definir los estadísticos se debe a que, aunque con la muestra se ha reducido bastante la dimensión del problema, el excesivo tamaño de ésta obliga a comprimir aún más la información para obtener respuestas a las preguntas que puedan hacerse y, de esa forma, completar el proceso inferencial. El objetivo que se persigue al definir los estimadores es el de resumir la información muestral, en aras, de obtener valores próximos a los verdaderos valores de los parámetros desconocidos de la distribución de la población.

3. La función de verosimilitud

Sea X una variable aleatoria continua cuya distribución viene dada por una función de densidad $f_{\underline{\theta}}$, donde $\underline{\theta} \in \mathbb{R}^k$ es un vector de parámetros desconocidos. Para una muestra $\underline{x}$ extraída de dicha población, se define la *función de verosimilitud* como:

$$L(\underline{x}, \underline{\theta}) = f_{\underline{\theta}}(\underline{x}),$$

que en el caso de una muestra aleatoria simple toma la forma

$$L(\underline{x}, \underline{\theta}) = f_{\underline{\theta}}(\underline{x}) = \prod_{i=1}^{n} f_{\underline{\theta}}(x_i).$$

Si la variable aleatoria es discreta la función de verosimilitud se define de forma análoga, cambiando la función de densidad por la de probabilidad.

Se ha de hacer notar que la verosimilitud varía en los parámetros mientras que la muestra permanece constante. La importancia de dicha función queda perfectamente ilustrada en el caso de que la población bajo estudio sea discreta, ya que en tal caso la función de verosimilitud expresa la probabilidad de obtener una muestra en función del vector de parámetros $\underline{\theta}$.

4. Suficiencia

Anteriormente se ha comentado que los estadísticos realmente suponen una compresión o resumen de la información suministrada por la muestra, por ello, sería ideal que el estadístico contuviera toda la información relevante que posee la muestra respecto al parámetro que se está estimando. Si ocurre esto, se dice que el *estadístico* es *suficiente* para dicho parámetro. Formalmente, ello supone que la distribución conjunta de la muestra condicionada al estadístico, es independiente del parámetro.

La caracterización de la suficiencia de un estadístico se hace a partir del *criterio de factorización de Fisher–Neyman*, que dice que dada una m.a.s., $\underline{X}$, se tiene que un estadístico, $T(\underline{X})$, es suficiente para $\underline{\theta}$ si la función de verosimilitud admite la siguiente descomposición:

$$L(\underline{x}, \underline{\theta}) = g(T(\underline{x}), \underline{\theta}) h(\underline{x}),$$

donde g es una función no negativa, tanto del estadístico como del vector de parámetros, y h es una función no negativa exclusiva de los valores muestrales.

Ejemplo 2.2 *De una población distribuida según una Bernouilli de parámetro p se extrae una m.a.s. de tamaño n. Se trata de encontrar un estimador suficiente para el parámetro p. Para ello se considera la función de verosimilitud*

$$\begin{aligned} L(\underline{x}, p) &= P_p[(X_1, \ldots, X_n) = (x_1, \ldots, x_n)] \\ &= P_p[X_1 = x_1] \cdots P_p[X_n = x_n] \\ &= p^{x_1}(1-p)^{1-x_1} \cdots p^{x_n}(1-p)^{1-x_n} \\ &= p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}. \end{aligned}$$

Por el criterio de factorización, tomando

$$t = T(\underline{x}) = \textstyle\sum_{i=1}^n x_i, \qquad h(\underline{x}) = 1$$

y

$$g(t,p) \quad = \quad p^t(1-p)^{n-t},$$

se obtiene que $\sum_{i=1}^{n} X_i$ *es un estimador suficiente para* p.

5. Propiedades de los estimadores

Puesto que para un mismo parámetro pueden existir varios estimadores, a continuación se analizan una serie de propiedades que serían deseables para un estimador y que permiten elegir, entre dos de ellos, el mejor.

5.1. Estimador insesgado

Una propiedad deseable para un estimador es que su valor medio sea igual al parámetro que se quiere estimar. Dicha propiedad se llama *insesgadez*. Formalmente, un estimador $T(\underline{X})$ es insesgado o centrado para un parámetro θ, cuando $\mathrm{E}[T(\underline{X})] = \theta$.

Ejemplo 2.3 *La media muestral es un estimador insesgado para la media poblacional* μ, *cualquiera que sea la distribución de la población, ya que* $\mathrm{E}[\overline{X}] = \mu$.

Si se verifica que $\mathrm{E}[T(\underline{X})] = \theta + b(\theta)$ el estimador será sesgado o descentrado, siendo $b(\theta)$ su sesgo, excentricidad o error sistemático. Es interesante que un estimador sea insesgado porque tomará valores que estarán alrededor del valor del parámetro θ.

Ejemplo 2.4 *Si se considera la varianza muestral como estimador de la varianza poblacional, puede comprobarse que se trata de un estimador sesgado, ya que* $\mathrm{E}[S^2] = \frac{n-1}{n}\sigma^2$, *siendo su sesgo* $-\frac{1}{n}\sigma^2$. *Para demostrarlo, hay que tener en cuenta que la varianza muestral puede escribirse de la forma:*

$$\begin{aligned} S^2 &= \frac{\displaystyle\sum_{i=1}^{n}(X_i - \overline{X} + \mu - \mu)^2}{n} \\ &= \frac{1}{n}\sum_{i=1}^{n}\big((X_i - \mu) - (\overline{X} - \mu)\big)^2 . \end{aligned}$$

Desarrollando el cuadrado se obtiene

$$S^2 = \frac{1}{n}\left(\sum_{i=1}^{n}(X_i - \mu)^2 - n(\overline{X} - \mu)^2\right).$$

Calculando la esperanza de la varianza muestral a partir de la expresión anterior se tiene que

$$\mathrm{E}[S^2] = \frac{1}{n}\left(\sum_{i=1}^{n}\mathrm{E}[(X_i-\mu)^2]-n\,\mathrm{E}[(\overline{X}-\mu)^2]\right).$$

Teniendo en cuenta que la muestra es aleatoria simple y que la media muestral verifica que $\mathrm{E}[\overline{X}] = \mu$ *y que* $\mathrm{V}[\overline{X}] = \frac{\sigma^2}{n}$, *se tiene que la primera de las esperanzas que aparecen en el segundo miembro es, para todo* i, $\mathrm{E}[(X_i - \mu)^2] = \sigma^2$ *y la segunda,* $\mathrm{E}[(\overline{X} - \mu)^2] = \frac{\sigma^2}{n}$, *con lo que se llega a*

$$\mathrm{E}[S^2] = \frac{1}{n}\left(n\sigma^2 - n\frac{\sigma^2}{n}\right) = \frac{n-1}{n}\sigma^2.$$

Ahora bien, si se considera la cuasivarianza *muestral como estimador de la varianza poblacional en vez de considerar la varianza muestral, se llega a que éste último es insesgado. Para ello, basta tener en cuenta que la cuasivarianza se puede expresar en función de la varianza como* $S_c^2 = \frac{n}{n-1}S^2$, *entonces su esperanza viene dada por:*

$$\begin{aligned} \mathrm{E}[S_c^2] &= \mathrm{E}\left[\frac{n}{n-1}S^2\right] \\ &= \frac{n}{n-1}\,\frac{n-1}{n}\,\sigma^2 = \sigma^2. \end{aligned}$$

Cuando el sesgo $b(\theta)$ es tal que $\lim_{n\to\infty} b(\theta) = 0$, se dice que el estimador es *asintóticamente insesgado.*

Ejemplo 2.5 *Anteriormente se estudió que la varianza muestral era un estimador sesgado de la varianza poblacional, siendo su sesgo $b(\sigma) = -\frac{1}{n}\sigma^2$. Se observa que cuando $n \to \infty$ el sesgo $b(\sigma) \to 0$. Con lo cual, se tiene que la varianza muestral es un estimador asintóticamente insesgado del parámetro σ.*

5.2. Estimador eficiente

Puesto que lo que se intenta es obtener el valor del parámetro a través de un estimador, que es a su vez una variable aleatoria, una propiedad que también sería deseable es que la varianza de dicho estimador fuese lo más pequeña posible, dicha propiedad se denomina *eficiencia.* Se dice que un estimador T_1 es más eficiente que otro T_2, cuando ocurre que $\mathrm{Var}(\mathrm{T}_1)<\mathrm{Var}(\mathrm{T}_2)$. Un estimador es eficiente, en términos absolutos, cuando alcanza la llamada *Cota de Frechet–Cramer–Rao*, que para muestras aleatorias simples viene dada a través de la expresión

$$\mathrm{V}(T) \geq \frac{\left[\frac{\partial\,\mathrm{E}[T(\underline{X})]}{\partial\theta}\right]^2}{n\,\mathrm{E}\left[\left(\frac{\partial \log f_\theta(X)}{\partial\theta}\right)^2\right]},$$

donde el denominador de la expresión anterior se conoce como *cantidad de información de Fisher*, $I(\theta)$.

Observación 2.1 *Para poder aplicar esta cota es necesario que se cumplan ciertas condiciones de regularidad de $f_\theta(x)$. Son las conocidas condiciones de regularidad de Fisher–Wolfowitz:*

1. *El campo de variación de la población de la cual se extrajo la muestra es independiente del parámetro θ, y por tanto, la muestra también lo es.*
2. *Existen, al menos, las dos primeras derivadas respecto al parámetro θ de la función $L(\underline{X}, \theta)$.*

3. *La derivación e integración, así como la suma en el caso discreto, son operaciones intercambiables.*

Cuando un estimador es más eficiente que otro pero a su vez tiene más sesgo, en general, se decide por aquel que tenga menor error cuadrático medio (ECM). El *error cuadrático medio* de un estimador se define como:

$$ECM(T) = \mathrm{E}\left[(T-\theta)^2\right] = \mathrm{V}[T] + (\theta - \mathrm{E}[T])^2 = \mathrm{V}[T] + b(\theta)^2,$$

es decir, la varianza del estimador más el cuadrado del sesgo.

Ejemplo 2.6 *Se quiere estimar el parámetro λ de una Poisson mediante la media de una muestra de tamaño n. ¿Es la media un estimador eficiente?*

La varianza de la media muestral es $\mathrm{V}[\overline{X}] = \frac{\lambda}{n}$ y la esperanza $\mathrm{E}[\overline{X}] = \lambda$. Calculando la Cota de Frechet–Cramer–Rao:

$$CFCR = \frac{\left(\dfrac{\partial\,\mathrm{E}[\overline{X}]}{\partial\lambda}\right)^2}{n\,\mathrm{E}\left[\left(\dfrac{\partial \log P_\lambda(X)}{\partial\lambda}\right)^2\right]}.$$

Se tiene que

$$\log P_\lambda(X) = -\lambda + x\log\lambda - \log(x!)$$

y su derivada respecto a λ

$$\frac{\partial \log P_\lambda(X)}{\partial\lambda} = -1 + \frac{x}{\lambda} = \frac{x-\lambda}{\lambda},$$

luego el denominador queda

$$\begin{aligned} n\,\mathrm{E}\left[\left(\frac{\partial \log P_\lambda(X)}{\partial\lambda}\right)^2\right] &= n\frac{\mathrm{E}\left[(X-\lambda)^2\right]}{\lambda^2} \\ &= n\frac{\mathrm{V}[X]}{\lambda^2} \\ &= n\frac{\lambda}{\lambda^2} = \frac{n}{\lambda}, \end{aligned}$$

y la Cota de Frechet–Cramer–Rao

$$CFCR = \frac{1}{n\,\mathrm{E}\left[\left(\frac{\partial \log P_\lambda(X)}{\partial \lambda}\right)^2\right]} = \frac{\lambda}{n}.$$

Como la varianza del estimador es igual a $\frac{\lambda}{n}$, se tiene que éste es eficiente.

5.3. Estimador consistente

Cuando un estimador no es insesgado se le exige que al menos sea *consistente*. Existen diversas definiciones de consistencia, en función de la convergencia que se utilice. Aquí se entenderá que un estimador es consistente cuando:

1. $\lim\limits_{n\to\infty} E(T_n) = \theta$

2. $\lim\limits_{n\to\infty} \mathrm{V}(T_n) = 0$

conociéndose dicha consistencia como *consistencia en probabilidad.*

Ejemplo 2.7 *La media muestral es un ejemplo de estimador consistente de la media poblacional μ: $\mathrm{E}[\overline{X}] = \mu$ y por tanto $\lim\limits_{n\to\infty} \mathrm{E}[\overline{X}] = \mu$ y $\mathrm{V}[\overline{X}] = \frac{\sigma^2}{n}$, con lo que se tiene que $\lim\limits_{n\to\infty} \mathrm{V}[\overline{X}] = 0$.*

6. Métodos de obtención de estimadores

A continuación, se estudian dos métodos que van a permitir obtener estimadores con unas cotas de bondad razonablemente buenas en relación con las propiedades que se acaban de describir. El primero de ellos, llamado *método de los momentos*, se basa en la correspondencia entre las características de la población y las de la muestra. El segundo, denominado de *máxima verosimilitud*, se apoya en la función de verosimilitud definida anteriormente.

6.1. Método de los momentos

Sea X una variable aleatoria tal que existen los r primeros momentos poblacionales con respecto al origen y cuya distribución depende de una serie de parámetros $\theta_1, \ldots, \theta_k$ desconocidos. En el caso de que el parámetro i-ésimo se pueda expresar en función de los r primeros momentos poblacionales con respecto al origen, es decir, $\theta_i = g_i(\alpha_1, \ldots, \alpha_r)$, para una muestra $(X_1, \ldots, X_n)$ el estimador obtenido a través del método de los momentos para dicho parámetro viene dado por $\hat{\theta}_i(\underline{X}) = g_i(a_1, \ldots, a_r)$, donde

$$\alpha_s = E[X_i^s] \qquad a_s = \frac{\sum_{i=1}^{n} X_i^s}{n}.$$

Propiedad 2.1 *Los estimadores obtenidos por el método de los momentos son consistentes, aunque, en general, no son insesgados ni tienen varianza mínima.*

Ejemplo 2.8 *Se quiere estimar la media y la varianza de una $N(\mu, \sigma)$ por el método de los momentos. Se sabe que $\mu = \mathrm{E}[X] = \alpha_1$, luego un estimador para μ por el método de los momentos resulta de sustituir α_1 por a_1, así*

$$\hat{\mu} = a_1 = \frac{\sum_{i=1}^{n} X_i}{n} = \overline{X}.$$

En cuanto a la varianza, se sabe que

$$\alpha_2 = \mathrm{E}[X^2] = \sigma^2 + \mu^2 \quad \Rightarrow \quad \sigma^2 = \alpha_2 - \mu^2,$$

luego sustituyendo α_2 por a_2 y μ por su estimador, se tiene que un estimador de σ^2 por el método de los momentos viene dado por:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} X_i^2}{n} - \overline{X}^2 = S^2.$$

6.2. Método de máxima verosimilitud

Este método ofrece generalmente mejores resultados que el anterior. En la misma situación se construye la función de verosimilitud, $L(\underline{x}, \underline{\theta})$, viniendo dados los estimadores máximos verosímiles por aquellas expresiones del vector de parámetros, $\underline{\theta}$, con $\underline{\theta} \in \mathbb{R}^k$, que hacen máxima dicha función. Por ser positiva y coincidir sus máximos con los de su logaritmo, el método se reduce a buscar la expresión del vector de parámetros, $\underline{\theta}$, que haga máxima la función $\log L(\underline{x}, \underline{\theta})$.

Propiedad 2.2 *Los estimadores máximo–verosímiles (M.V.) son asintóticamente insesgados, asintóticamente Normales y de existir un estimador eficiente éste es el máximo–verosímil.*

Ejemplo 2.9 *Sea X una variable aleatoria que sigue una $B(p)$. Para encontrar el estimador máximo–verosímil para p, se construye en primer lugar la función de verosimilitud:*

$$L(\underline{x}, p) = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}.$$

Si $T(\underline{x})$ es tal que

$$\max_p L(\underline{x}, p) = L(\underline{x}, T(\underline{x})),$$

entonces también se verifica que

$$\max_p \log L(\underline{x}, p) = \log L(\underline{x}, T(\underline{x})).$$

De esta forma, se tiene que

$$\log L(\underline{x}, p) = \sum_{i=1}^n x_i \log p + (n - \sum_{i=1}^n x_i) \log(1-p),$$

luego

$$\frac{\partial \log L(\underline{x}, p)}{\partial p} = \frac{\sum_{i=1}^n x_i}{p} - \frac{n - \sum_{i=1}^n x_i}{1-p} = 0.$$

De donde se obtiene que $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

A continuación, habría que comprobar que para ese valor de p *la función* $\log L(\underline{x}, p)$ *alcanza un máximo. Se puede probar que* $\frac{\partial^2 \log L(\underline{x},p)}{\partial p^2} \leq 0$ *y por tanto,* $T(\underline{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i$ *es el estimador máximo–verosímil de* p.

7. Estimación de parámetros en poblaciones Normales

Puesto que la mayoría de problemas que se abordan en la Inferencia Estadística asumen la hipótesis de normalidad de la población bajo estudio, a partir de ahora, se le va a dar un trato diferenciado, particularizando cualquier estudio para esta distribución. Sea pues X una variable aleatoria con distribución $\mathrm{N}(\mu,\sigma)$. Los estimadores máximo–verosímiles de los parámetros μ y σ son, respectivamente, la media y la desviación típica muestral, con las propiedades que les confiere el método utilizado. Se trata ahora de estudiar las distribuciones de ambos estimadores.

No obstante, antes de continuar con esta cuestión, se hará un inciso para estudiar una serie de distribuciones que se derivan de la Normal.

7.1. Distribuciones derivadas de la Normal

Las distribuciones derivadas de la Normal tienen gran importancia en la Inferencia Estadística, ya que serán las distribuciones de una amplia familia de estimadores. Todas ellas se obtienen como combinación y/o promedios de variables Normales, están tabuladas y se caracterizan sólo por el número de Normales tipificadas que entran en su composición; a dicho número se le llama grado(s) de libertad, justificándose este nombre por los motivos que se desarrollarán en los próximos temas.

7.1.1. Distribución Chi–cuadrado

Sean $\mathrm{Z}_1, \mathrm{Z}_2, \ldots, \mathrm{Z}_n$, n variables $N(0,1)$ independientes, la variable

χ^2_n definida como

$$\chi^2_n = Z_1^2 + Z_2^2 + \cdots + Z_n^2$$

sigue una distribución Chi–cuadrado con n grados de libertad. Dicha variable puede interpretarse como el cuadrado de la distancia euclídea desde el origen de coordenadas al punto $(\mathrm{Z}_1, \mathrm{Z}_2, \ldots, \mathrm{Z}_n)$. La variable se caracteriza únicamente por el número de Normales tipificadas que entran en su composición y la independencia de éstas hace fácil el cálculo de los momentos. Así

$$\mathrm{E}[\chi^2_n] = n \qquad y \qquad \mathrm{V}[\chi^2_n] = 2n.$$

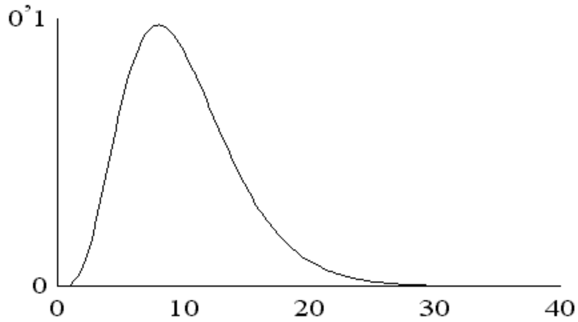

Figura 2.1: Distribución Chi–cuadrado

La función de densidad de la distribución Chi–cuadrado es asimétrica, siendo sólo distinta de cero para valores positivos de la variable. Tiene una asíntota cuando los valores tienden a infinito y para $n > 2$ tiene forma campaniforme.

Propiedades 2.3

1. *La distribución Chi–cuadrado es un caso particular de una distribución Gamma, en concreto, es una* $\Gamma(\frac{1}{2}, \frac{n}{2})$. *Recuérdese que* $X \sim \Gamma(a;p)$ *si tiene como función de densidad*

$$f(x) = \begin{cases} \dfrac{a^p}{\Gamma(p)} e^{-ax} x^{p-1} & si\ x > 0 \\ 0 & si\ x \leq 0 \end{cases}$$

2. *La función característica de un* χ_n^2 *es* $\varphi(t) = (1-2it)^{-\frac{n}{2}}$.

3. *La suma de dos Chi–cuadrado independientes con* n_1 *y* n_2 *grados de libertad es una nueva variable Chi–cuadrado con* $n_1 + n_2$ *grados de libertad.*

4. *Cuando* n *es mayor que 100 se verifica la siguiente aproximación:*

$$\sqrt{2\chi_n^2} \quad \cong \quad N(\sqrt{2n-1}, 1).$$

Ejemplo 2.10 *La velocidad (cm/seg) de un objeto de masa 1 Kg., viene dada por una variable aleatoria* V *que sigue una* $N(0, 25)$. *Si* $K = \frac{mV^2}{2}$, *donde* m *es la masa del objeto, es la variable aleatoria que representa la energía cinética de dicho objeto, se pide calcular la probabilidad de que la energía cinética sea menor que 200.*

Puesto que $m = 1$, *se tiene que*

$$\begin{aligned} P(K < 200) &= P\left(\tfrac{mV^2}{2} < 200\right) \\ &= P\left(\tfrac{V^2}{625} < \tfrac{200\cdot 2}{625}\right) \\ &= P\left(\tfrac{V^2}{625} < 1'28\right) \\ &= P(\chi_1^2 < 1'28) = 0'725. \end{aligned}$$

7.1.2. Distribución t de Student

Sean Z y χ_n^2 dos variables aleatorias independientes que siguen una distribución $N(0,1)$ y una Chi–cuadrado con n grados de libertad, respectivamente. La variable aleatoria

$$t_n = \frac{Z}{\sqrt{\frac{\chi_n^2}{n}}},$$

sigue una distribución t de Student con n grados de libertad. La distribución es simétrica respecto a cero, con una varianza mayor que la $N(0,1)$ y tiende a ésta a medida que n lo hace hacia infinito (se puede considerar que sus probabilidades coinciden a partir de un n superior a 120).

La distribución t de Student compara a una $N(0,1)$ con un promedio de n variables $N(0,1)$. Sus momentos principales son:

$$\mathrm{E}[t_n] = 0 \qquad \mathrm{V}[t_n] = \frac{n}{n-2} \quad (n > 2).$$

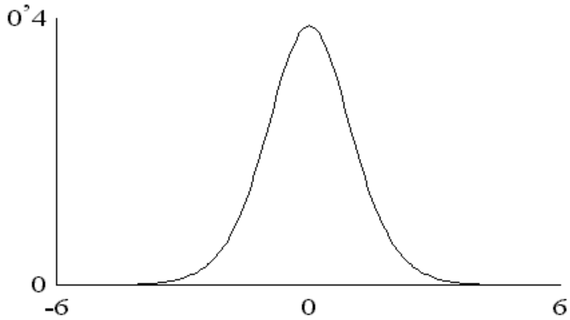

Figura 2.2: Distribución t de Student

Ejemplo 2.11 *Sea V una variable aleatoria que sigue una t_{20}. Se quiere hallar a tal que $P(|V| > a) = 0'01$. Para ello basta tener en cuenta que la distribución t de Student es simétrica, con lo cual*

$$\begin{aligned} P(|V| > a) &= P(V > a) + P(V < -a) \\ &= 2P(V > a) = 0'01. \end{aligned}$$

Así pues, el a requerido es el que verifica

$$P(V > a) = 0'005,$$

de donde se obtiene, buscando en las tablas, que $a = 2'845$.

7.1.3. Distribución $\mathcal{F}$ de Snedecor–Fisher

La distribución $\mathcal{F}$ se define como el cociente entre dos variables independientes Chi–cuadrado divididas por sus grados de libertad, es decir

$$\mathcal{F}_{n,m} = \frac{\frac{\chi^2_n}{n}}{\frac{\chi^2_m}{m}}.$$

La distribución está caracterizada por los grados de libertad n y m, siendo su forma esencialmente la misma de la Chi–cuadrado. Sus características más importantes son:

$$\mathrm{E}[\mathcal{F}_{n,m}] = \frac{m}{m-2} \qquad \mathrm{V}[\mathcal{F}_{n,m}] = \frac{2m^2(m+n-2)}{n(m-2)^2(m-4)}.$$

Propiedades 2.4

1. *De la definición se deduce que si* $X \sim \mathcal{F}_{n,m} \Rightarrow \frac{1}{X} \sim \mathcal{F}_{m,n}$.
2. *La distribución t de Student al cuadrado es un caso particular de la $\mathcal{F}$. Esto es, si*

$$t_n = \frac{Z}{\sqrt{\frac{\chi_n^2}{n}}} \quad \Rightarrow \quad t_n^2 = \frac{Z^2}{\frac{\chi_n^2}{n}}$$

siendo $t_n^2 \sim \mathcal{F}_{1,n}$

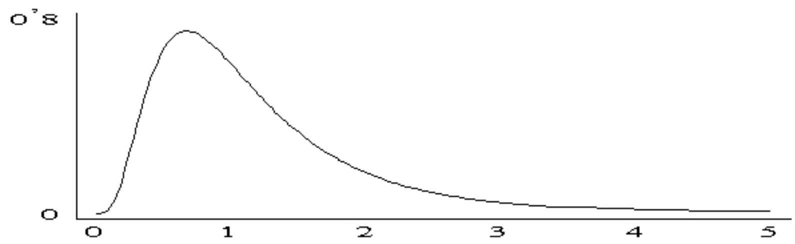

Figura 2.3: Distribución $\mathcal{F}$ de Fisher–Snedecor

Ejemplo 2.12 *Hallar el valor b tal que $P(F < b) = 0'01$, sabiendo que la variable aleatoria F sigue una distribución $\mathcal{F}_{7,20}$.*

Como $\mathcal{F}_{7,20} = \frac{1}{\mathcal{F}_{20,7}}$, *se tiene entonces que*

$$P(\mathcal{F}_{7,20} < b) = 0'01 \quad \Leftrightarrow \quad P\left(\frac{1}{\mathcal{F}_{7,20}} > \frac{1}{b}\right) = 0'01,$$

luego

$$P(\mathcal{F}_{20,7} > \frac{1}{b}) = 0'01 \quad \Rightarrow \quad \frac{1}{b} = 6'15.$$

De donde $b = 0'162$.

7.2. Distribución de la media muestral

Como ya se ha visto en ejemplos anteriores, la media muestral tiene esperanza μ y varianza σ^2/n; además por ser combinación lineal de variables Normales es a su vez Normal, es decir:

$$\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

Lo anterior también sería, aproximadamente, cierto para una variable X no Normal siempre que n sea suficientemente grande, como garantiza el Teorema Central del Límite.

7.3. Distribución de la varianza muestral

La relación que existe entre la media y la varianza muestral viene dada por el *teorema de Fisher–Cochran*:

Teorema 2.1 *Las variables aleatorias $\overline{X}$ y S^2 son independientes y el estadístico $\frac{nS^2}{\sigma^2}$ tiene distribución Chi–cuadrado con n-1 grados de libertad.*

Se obviará la demostración de lo anterior, que de forma equivalente y en función de la cuasivarianza muestral, puede expresarse como:

$$\frac{(n-1)S_c^2}{\sigma^2} \sim \chi^2_{n-1}. \tag{2.1}$$

De los momentos de una Chi–cuadrado se puede deducir:

$$\mathrm{E}[S^2] = \frac{n-1}{n}\sigma^2 \qquad \mathrm{V}[S^2] = \frac{2(n-1)}{n^2}\sigma^4.$$

Esto indica, como ya se estudió, que S^2 no es un estimador insesgado de σ^2, por lo que en la mayoría de los casos se toma como estimador de la varianza poblacional la cuasivarianza muestral, S_c^2, también denominada *varianza muestral corregida*. La razón de no elegir siempre la cuasivarianza es que su ECM es mayor que el de la varianza.

Por otro lado, puesto que $\frac{\overline{X}-\mu}{\sigma}\sqrt{n} \sim N(0,1)$ y $(n-1)\frac{S_c^2}{\sigma^2} \sim \chi^2_{n-1}$ y como además estos estadísticos son independientes, se tiene que

$$\frac{\frac{\overline{X}-\mu}{\sigma}\sqrt{n}}{\sqrt{\frac{(n-1)S_c^2}{(n-1)\sigma^2}}} = \frac{\overline{X}-\mu}{S_c}\sqrt{n} \sim t_{n-1}.$$

7.4. Distribución de la diferencia de medias muestrales

Se consideran dos poblaciones independientes representadas por las variables X e Y, con distribuciones respectivas $N(\mu_1, \sigma_1)$ y $N(\mu_2, \sigma_2)$, de las cuales se dispone de sendas muestras, $\underline{X}$ y $\underline{Y}$, de tamaños n_1 y n_2 respectivamente. Es fácil ver que su diferencia de medias muestrales se distribuye como:

$$\overline{X} - \overline{Y} \sim N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right).$$

7.5. Distribución del cociente de varianzas muestrales

Dadas las dos variables anteriores X e Y independientes y según la ecuación (2.1), se tiene que:

$$\frac{\frac{S_{c_1}^2}{\sigma_1^2}}{\frac{S_{c_2}^2}{\sigma_2}} \sim \mathcal{F}_{n_1-1, n_2-1}.$$

Ejemplo 2.13 *En una clase de ciencias, se toma una m.a.s. de 51 alumnos que se presentaron al examen de matemáticas y otra, independiente de la anterior, de 19 alumnos presentados al examen de física. Se sabe que las notas de los alumnos tanto en matemáticas como en física siguen una Normal con la misma* *dispersión. Se pretende averiguar cuál es la probabilidad de que la varianza observada en la primera muestra sea al menos el doble de la segunda.*

Sea S_M^2 la varianza muestral de las notas correspondientes a matemáticas y S_F^2 la varianza muestral de las notas de física. Puesto que se trata de muestras independientes y teniendo en cuenta que

$$51\frac{S_M^2}{\sigma^2} \sim \chi_{50}^2 \quad y \quad 19\frac{S_F^2}{\sigma^2} \sim \chi_{18}^2$$

se tiene que

$$\frac{51 \cdot 18 \cdot S_M^2}{50 \cdot 19 \cdot S_F^2} \quad \sim \quad \mathcal{F}_{50,18}.$$

Así pues,

$$\begin{aligned} P\left(\frac{S_M^2}{S_F^2} \geq 2\right) &= P\left(\frac{51 \cdot 18 \cdot S_M^2}{50 \cdot 19 \cdot S_F^2} \geq 1'93\right) \\ &= P(\mathcal{F}_{50,18} \geq 1'93) \\ &= 0'0632. \end{aligned}$$

7.6. Distribución de la proporción muestral

Utilizando una particularización del Teorema Central del Límite, se sabe que de forma asintótica, para una población Bernouilli, B(p), se tiene que la distribución de la proporción muestral $\hat{p} = \overline{X}$ puede aproximar por una Normal, tal que

$$\hat{p} \quad \cong \quad N\left(p, \sqrt{\frac{pq}{n}}\right).$$

Si se tienen dos poblaciones Bernouilli, entonces:

$$\hat{p}_1 - \hat{p}_2 \quad \cong \quad N\left(p_1 - p_2, \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}\right).$$

8. Ejercicios

8.1. Ejercicios resueltos

2.1 Sea $\underline{X}$ una m.a.s. de tamaño n extraída de una población Normal de media μ y varianza σ^2.

a) Halle los estimadores de máxima verosimilitud para los parámetros μ y σ^2.

b) ¿Es $\overline{X}$ un estimador eficiente de μ? ¿Y suficiente?

c) Encuentre un estimador insesgado de $\mu^2 + \sigma^2$.

d) Si $\mu = 0$, encuentre un estimador suficiente de σ^2.

Solución:

a) A partir de la m.a.s. de tamaño n, se construye la función de verosimilitud:

$$L = L(\underline{x}; \mu, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{ -\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{2\sigma^2} \right\}.$$

Para encontrar los estimadores de máxima verosimilitud de μ y σ^2, hay que encontrar los máximos de la función de verosimilitud, L, o equivalentemente, los máximos de la función $\log L$:

$$\log L = -\frac{n}{2}\log(\sigma^2) - \frac{n}{2}\log 2\pi - \frac{\displaystyle\sum_{i=1}^{n}(x_i - \mu)^2}{2\sigma^2}.$$

Para ello, habrá que resolver el siguiente sistema:

$$\begin{aligned}
\frac{\partial \log L}{\partial \mu} &= \frac{\displaystyle\sum_{i=1}^{n}(x_i - \mu)}{\sigma^2} = 0 \\
\frac{\partial \log L}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{\displaystyle\sum_{i=1}^{n}(x_i - \mu)^2}{2\sigma^4} = 0
\end{aligned}$$

cuya solución proporciona los estimadores máximo–verosímiles de μ y σ^2:

$$\hat{\mu} = \overline{X}$$

$$\hat{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n} = S^2.$$

b) Supuesta conocida σ^2 y teniendo en cuenta que $T(\underline{X}) = \overline{X}$ es insesgado para μ, la cota de Frechet–Cramer–Rao viene dada por:

$$CFCR = \frac{1}{n\,\mathrm{E}\left[\left(\frac{\partial \log f_\mu(X)}{\partial \mu}\right)^2\right]}.$$

Haciendo cálculos,

$$\log f_\mu(x) = -\log\sigma - \frac{1}{2}\log 2\pi - \frac{(x-\mu)^2}{2\sigma^2},$$

$$\frac{\partial \log f_\mu(x)}{\partial \mu} = \frac{x-\mu}{\sigma^2} \Rightarrow \left(\frac{\partial \log f_\mu(x)}{\partial \mu}\right)^2 = \frac{(x-\mu)^2}{\sigma^4},$$

$$\mathrm{E}\left[\left(\frac{\partial \log f_\mu(X)}{\partial \mu}\right)^2\right] = \frac{1}{\sigma^2}.$$

Por tanto, la cota de Frechet–Cramer–Rao vale

$$CFCR = \frac{1}{n\frac{1}{\sigma^2}} = \frac{\sigma^2}{n},$$

y puesto que $\mathrm{V}[T(\underline{X})] = \mathrm{V}[\overline{X}] = \frac{\sigma^2}{n}$, se deduce que $T(\underline{X}) = \overline{X}$ es eficiente para μ.

Estos cálculos no hubiesen sido necesarios si se hubiera recurrido a las propiedades de los estimadores máximo–verosímiles. Como $\overline{X}$ es máximo-verosímil para μ, de existir un estimador eficiente sería él.

En cuanto a la suficiencia de $T(\underline{X}) = \overline{X}$, a partir de la función de máxima verosimilitud

$$L = L(\underline{x};\mu,\sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right\},$$

desarrollando la expresión de la exponencial,

$$-\frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^2} = -\frac{\sum_{i=1}^n x_i^2 + n\mu(\mu - 2\overline{x})}{2\sigma^2}$$

se llega a que, si se define

$$\begin{array}{rcl} g(T(\underline{x}),\mu) & = & \exp\left\{-\frac{n\mu(\mu-2\overline{x})}{2\sigma^2}\right\} \\ h(\underline{x}) & = & (2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{\sum_{i=1}^n x_i^2}{2\sigma^2}\right\}, \end{array}$$

por el criterio de factorización, el estimador $T(\underline{X}) = \overline{X}$ es suficiente para μ.

c) Puesto que

$$\mathrm{E}[\overline{X}] = \mu;\ \mathrm{V}[\overline{X}] = \frac{\sigma^2}{n},$$

se tiene que

$$\mathrm{E}[\overline{X}^2] = \mathrm{V}[\overline{X}] + \mathrm{E}[\overline{X}]^2 = \frac{\sigma^2}{n} + \mu^2.$$

Como, además,

$$\mathrm{E}[S^2] = \frac{n-1}{n}\sigma^2,$$

el estimador que se busca es $T(\underline{X}) = \overline{X}^2 + S^2$, ya que

$$\mathrm{E}[T(\underline{X})] = \mathrm{E}[\overline{X}^2] + \mathrm{E}[S^2] = \sigma^2 + \mu^2.$$

d) En este caso,

$$f_{\sigma^2}(x) = (2\pi\sigma^2)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2\sigma^2}x^2\right\}$$

y la función de verosimilitud es

$$L = L(\underline{x},\sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^k x_i^2\right\}.$$

Por el criterio de factorización, tomando

$$g(T(\underline{x}),\sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}T(\underline{x})\right\},$$

con $T(\underline{x}) = \sum_{i=1}^k x_i^2$, y $h(\underline{x}) = 1$, se tiene que $T(\underline{X}) = \sum_{i=1}^k X_i^2$ es estimador suficiente para σ^2.

8.2. Ejercicios propuestos

2.1. Dadas W, X, Y y Z cuatro variables aleatorias independientes e idénticamente distribuidas según una $N(0, 5)$.

a) Si $S = 2W + 3X - Y + Z + 30$, obtenga $P(S \leq 42)$.

b) Si $T = W^2 + X^2 + Y^2 + Z^2$, obtenga a verificando que $P(T \leq a) = 0'025$.

c) Si $U = \sqrt{\frac{W^2+X^2+Y^2+Z^2}{4}}$, obtenga $P(U \leq 6'973)$.

2.2. Sean X e Y dos variables aleatorias que siguen una t_{36} y una χ^2_{62} respectivamente.

a) Halle x tal que $P(|X| > x) = 0'05$.

b) Obtenga y tal que $P(|Y| > y) = 0'05$.

2.3. Se sabe que la anchura de las piezas fabricadas por una cierta máquina, medida en centímetros, se distribuye según una Normal de media 10 y desviación típica $0'25$. Si se toma una m.a.s. de 25 piezas, calcule:

a) $P(9'68 \leq \overline{X} \leq 10'1)$.

b) $P(S^2 \leq 0'19)$.

2.4. Se quiere estudiar la altura de los alumnos de tercero de ESO y se estimó, en experiencias anteriores, que dicha característica se distribuye según una Normal de media 167 cm. y varianza $10'24$ cm^2. Si se toma una m.a.s. de 10 alumnos,

a) Calcule la probabilidad de que la media muestral de las alturas de los 10 alumnos no sea inferior a 165 cm.

b) Halle la probabilidad de que la cuasivarianza muestral de las alturas de los 10 alumnos sea superior a $15'90$ cm^2.

2.5. Se extrae (X_1, X_2, X_3, X_4) m.a.s. de una población X distribuida según una Exp$(\frac{1}{\theta})$. Dados los estadísticos

$$Y_1(\underline{X}) = \frac{1}{6}(X_1 + X_2) + \frac{1}{3}(X_3 + X_4)$$

$$Y_2(\underline{X}) = \frac{X_1 + 2X_2 + 3X_3 + 4X_4}{5}$$

$$Y_3(\underline{X}) = \frac{X_1 + X_2 + X_3 + X_4}{4},$$

estudie cuáles son insesgados para θ.

2.6. Dada una población distribuida normalmente con media desconocida y varianza igual a 25, se extrae una m.a.s. de tamaño 3 y se consideran los estimadores de la media

$$\begin{aligned} Y(\underline{X}) &= 0'65X_1 + 0'25X_2 + 0'1X_3 \\ Z(\underline{X}) &= 2X_3 - X_1 \\ T(\underline{X}) &= \frac{X_1 + X_2 + X_3}{3}. \end{aligned}$$

Estudie cuál de los tres estimadores es el mejor desde el punto de vista del sesgo y la eficiencia.

2.7. Sea (X_1, X_2, X_3) una m.a.s. procedente de una población que se distribuye normalmente. Sean

$$T_1(\underline{X}) = \frac{X_1 + 2X_2 + 3X_3}{6} \qquad \text{y} \qquad T_2(\underline{X}) = \frac{X_1 - 4X_2}{-3}$$

dos estimadores de μ.

a) Demuestre que ambos son insesgados.

b) Pruebe que $T_1(\underline{X})$ es más eficiente que $T_2(\underline{X})$.

2.8. Sea X una variable aleatoria distribuida según una $N(\mu, \sigma)$. Calcule un estimador insesgado de $\mu^2 + 6\mu$.

2.9. De una población $N(\mu, 2)$ se extrae una m.a.s. $\underline{Y}$ de tamaño $n = 4$. Para el siguiente estimador de la media

$$T(\underline{Y}) = 0'2Y_1 + 0'4Y_2 + cY_3 + dY_4,$$

calcule c y d para que $T(\underline{Y})$ sea insesgado y eficiente.

2.10. Sea $(X_1, \ldots, X_n)$ una m.a.s. extraída de una población que sigue una $B(p)$. Considérense los estimadores:

$$T_1(\underline{X}) = \overline{X} \quad \text{y} \quad T_2(\underline{X}) = \frac{\sum_{i=1}^n X_i^2}{n}.$$

a) Demuestre que ambos son insesgados.
b) Estudie cuál es más eficiente.
c) ¿Son consistentes?

2.11. Sea $(X_1, \ldots, X_n)$ una m.a.s. extraída de una población que sigue una $P(\lambda)$.

a) Pruebe que $T(\underline{X}) = \sum_{i=1}^n X_i$ es suficiente para λ.

b) Estudie la consistencia del estadístico $U = \frac{T^2 - T}{n^2}$ para el parámetro λ^2.

2.12. Sea $(X_1, \ldots, X_n)$ una m.a.s de una $U(0, \theta)$, $\theta > 0$.

a) Sea $M_n(\underline{X}) = \text{máx}\{X_1, \ldots, X_n\}$. Pruebe que $M_n(\underline{X})$ es consistente para θ. ¿Es insesgado?

b) Si $Y_n(\underline{X}) = 2\overline{X}$, estudie la consistencia para θ.

c) Demuestre que el estadístico $Z_n(\underline{X}) = \frac{n+1}{n} M_n(\underline{X})$ es insesgado y más eficiente que $Y_n(\underline{X})$.

2.13. De una población distribuida según una $B(m, p)$, se extrae una m.a.s. de tamaño n. Estudie la insesgadez del estadístico $T(\underline{X}) = \frac{\overline{X}}{m}$ respecto al parámetro p y demuestre su eficiencia.

2.14. Considérese una m.a.s. de tamaño n extraída de una población Normal de media μ y varianza σ^2.

a) Encuentre un estimador suficiente de σ^2 cuando $\mu = 0$.

b) Busque un estimador suficiente de μ. ¿Es ese estimador eficiente?

c) Demuestre que $T(\underline{X}) = S^2$ no es un estimador eficiente de σ^2.

2.15. De una población con función de densidad

$$f_\theta(x) = \frac{1}{\theta}\, e^{-\frac{x}{\theta}}, \qquad x \geq 0$$

se extrae una m.a.s. de tamaño n. Si se estima el parámetro θ a través de la media muestral:

a) Demuestre que es consistente.
b) Estudie su eficiencia.

2.16. Estudie la eficiencia del estimador $T(\underline{X}) = \overline{X}$ del parámetro b de la función de densidad $\Gamma(b; a)$, para una m.a.s. de tamaño n.

2.17. Dada una m.a.s. de tamaño n extraída de una población $N(\mu, \sigma)$, se quiere estimar la media μ mediante

$$T(\underline{X}) = k \sum_{j=1}^{n} jX_j.$$

a) Obtenga k para que $T(\underline{X})$ sea insesgado.
b) Estudie si $T(\underline{X})$ es eficiente.
c) ¿Es consistente?

2.18. Para estimar la media de una población normalmente distribuida, se extrae una m.a.s. de tamaño n y se consideran los estadísticos

$$T_1(\underline{X}) = \overline{X} \qquad \text{y} \qquad T_2(\underline{X}) = \frac{\sum_{i=1}^{n} iX_i}{\sum_{i=1}^{n} i}.$$

Determine cual es el mejor estimador desde el punto de vista de la insesgadez y la eficiencia.

2.19. Sea $\underline{X}$ una m.a.s. extraída de una población $\text{Exp}\,(\frac{1}{\theta})$. Pruebe que el estadístico $Y_n(\underline{X}) = n\,\text{mín}\{X_i\},\ i = 1, \ldots, n$ es insesgado, pero no consistente para θ.

2.20. De una población distribuida según una Exponencial de función de densidad

$$f_\alpha(x) = \alpha e^{-x\alpha} \quad x > 0,$$

se extrae una m.a.s. de tamaño n.

a) Demuestre que $T(\underline{X}) = \sum_{i=1}^{n} X_i$ es suficiente para α.
b) Pruebe que el estimador $U = \frac{n-1}{T}$ es consistente para α.

2.21. Encuentre un estimador por el método de máxima verosimilitud para el parámetro λ de una Poisson.

2.22. Dada una m.a.s. extraída de una población que sigue una $\text{Exp}(\frac{1}{\theta})$, encuentre un estimador máximo-verosímil para el parámetro θ.

2.23. Determine un estimador, por el método de los momentos, de los parámetros en los siguientes casos:

a) $P(\lambda)$.
b) $\text{Exp}(\theta)$.
c) $\Gamma(a; b)$.

2.24. Sea la variable aleatoria X que sigue la distribución de Pascal:

$$f(x) = p(1-p)^x, \quad \text{donde } x = 0, 1, 2, \ldots \text{ y } 0 < p < 1.$$

Busque un estimador de p por el método de los momentos.

2.25. Obtenga un estimador, por el método de los momentos, para el parámetro a de la distribución que tiene por función de densidad

$$f_a(x) = \frac{2(a-x)}{a^2} \quad 0 < x < a.$$

2.26. Una fábrica produce botones cuyo diámetro varía aleatoriamente entre dos valores a y b. Supuesto que el diámetro se ajusta a una variable aleatoria distribuida uniformemente, estime, a partir de la muestra

$$10'20 \quad 10'22 \quad 10'10 \quad 10'14,$$

los parámetros a y b por el método de los momentos.

2.27. La función de densidad de una variable aleatoria es:

$$f_\theta(x) = (\theta + 1)x^\theta \quad 0 < x < 1.$$

Halle el estimador de θ utilizando:

a) El método de los momentos.
b) El método de máxima verosimilitud.

2.28. Una determinada empresa quiere planificar su producción. Ellos calculan que el producto que ofrecen puede gustar entre el 40 % y el 50 % de los habitantes de su ciudad, pero tras tomar una muestra aleatoria de 10 individuos observan que sólo tres muestran interés por el producto. Teniendo ésto en cuenta, ¿cuál de las dos proporciones contempladas deberán tomar en consideración con base en el criterio de máxima verosimilitud?

2.29. Sea X una variable aleatoria con función de probabilidad

$$P_\theta(X = x) = \theta(1 - \theta)^{x-1} \quad x = 1, 2, \ldots \quad 0 \leq \theta \leq 1.$$

Encuentre un estimador del parámetro θ por el método de máxima verosimilitud.

2.30. Obtenga un estimador por el método de los momentos para el parámetro a en una distribución Pareto, cuya función de densidad viene dada por

$$f_a(x) = \frac{a x_0^a}{x^{a+1}} \quad x > x_0.$$

2.31. Sea X una variable una variable aleatoria que tiene por función de densidad

$$f_\theta(x) = 2\theta^{-2}(1-x) \quad 0 < x < 1.$$

Obtenga un estimador de máxima verosimilitud para el parámetro θ.

2.32. Sea $(X_1, \ldots, X_n)$ una m.a.s. extraída de una población con distribución $U(a; b)$. Obtenga estimadores de a y b por el método de máxima verosimilitud.

2.33. Para una distribución que tiene función de densidad

$$f_\theta(x) = e^{-(x-\theta)}, \quad x \geq \theta,$$

calcule el estimador de máxima verosimilitud para el parámetro θ.