Inferencia estadística - Estimación puntual

La estadística provee técnicas que permiten obtener conclusiones generales a partir de un conjunto limitado – pero representativo – de datos. Cuando inferimos no tenemos garantía de que la conclusión que obtenemos sea exactamente correcta. Sin embargo, la estadística permite cuantificar el error asociado a la estimación.

La mayoría de las distribuciones de probabilidad dependen de cierto número de parámetros. Por ejemplo: $P(\lambda), N(\mu,\sigma^2), Bi(n,p),$ etc. Salvo que estos parámetros se conozcan, deben estimarse a partir de los datos.

El objetivo de la *estimación puntual* es usar una muestra para obtener números que, en algún sentido, sean los que mejor representan a los verdaderos valores de los parámetros

de interés.

Supongamos que se selecciona una muestra de tamaño *n* de una población. Antes de obtener la muestra no sabemos cuál será el valor de cada observación. Así, la primera observación puede ser considerada una v.a. X_1, la segunda una v.a. X_2, etc. Por lo tanto, antes de obtener la muestra denotaremos $X_1, X_2, \ldots, X_n$ a las observaciones y, una vez obtenida la muestra los valores observados los denotaremos $x_1, x_2, \ldots, x_n$.

Del mismo modo, antes de obtener una muestra, cualquier función de ella será una v.a., por ejemplo: $\overline{X}, \widetilde{X}, S^2, \max(X_1,\ldots,X_n),$ etc. Una vez obtenida la muestra los valores calculados serán denotados $\overline{x}, \widetilde{x}, s^2, \max(x_1,\ldots,x_n),$ etc.

<u>Definición</u>: Un estimador puntual de un parámetro θ es un valor que puede ser considerado representativo de θ y se indicará $\hat{\theta}$. Se obtiene a partir de alguna función de la muestra.

<u>Ejemplo</u>: Con el fin de estudiar si un dado es o no equilibrado, se arroja el dado 100 veces en forma independiente, obteniéndose 21 ases. ¿Qué valor podría utilizarse, en base a esa información, como estimación de la probabilidad de as? Parece razonable utilizar la frecuencia relativa de ases.

En este caso, si llamamos p a la probabilidad que queremos estimar, $\hat{p} = \dfrac{21}{100} = 0.21$

Métodos de estimación puntual

¿Cómo obtener estimadores para un problema dado? Estudiaremos dos métodos que proporcionan estimadores puntuales: el método de momentos y el método de máxima verosimilitud.

Método de momentos: La idea básica consiste en igualar ciertas características muestrales con las correspondientes características poblacionales. Recordemos la siguiente definición.

<u>Definición</u>: Sea X una v.a. con función de probabilidad puntual $p_X(x)$ en el caso discreto o función de densidad $f_X(x)$ en el caso continuo. Se denomina **momento de orden *k*** ($k \in N$) o momento poblacional de orden k a $E(X^k)$, es decir

$$E(X^k)=\begin{cases}\sum_x x^k\ p_X(x) & \text{en el caso discreto}\\ \int_{-\infty}^{\infty} x^k\ f_X(x)\,dx & \text{en el caso continuo}\end{cases}$$

si esas esperanzas existen.

Como ya hemos visto cuando estudiamos función generadora de momentos de una variable aleatoria, los momentos están relacionados con los parámetros de la distribución asociada.

Dada una muestra aleatoria $X_1, X_2, ..., X_n$, el **momento muestral de orden *k*** es

$$\frac{\sum_{i=1}^{n} X_i^k}{n}$$

<u>Definición</u>: Sea $X_1, X_2, ..., X_n$ una m.a. de una distribución con función de probabilidad puntual o función de densidad que depende de m parámetros $\theta_1, \theta_2, ..., \theta_m$. Los estimadores de momentos de $\theta_1, \theta_2, ..., \theta_m$ son los valores $\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_m$ que se obtienen igualando m momentos poblacionales con los correspondientes momentos muestrales. En general, se obtienen resolviendo el siguiente sistema de ecuaciones

$$\frac{\sum_{i=1}^{n} X_i^k}{n} = E\left(X^k\right) \qquad k = 1,2,...,m$$

<u>Ejemplos</u>: 1) Sea $X_1, X_2, ..., X_n$ una m.a. de una distribución exponencial de parámetro λ. Como hay un solo parámetro a estimar, basta plantear una ecuación basada en el primer momento.

$$\frac{\sum_{i=1}^{n} X_i}{n} = E\left(X\right) \quad\Rightarrow\quad \frac{\sum_{i=1}^{n} X_i}{n} = \frac{1}{\lambda} \quad\Rightarrow\quad \hat{\lambda} = \frac{n}{\sum_{i=1}^{n} X_i} \quad\Rightarrow\quad \hat{\lambda} = \frac{1}{\overline{X}}$$

2) Sea $X_1, X_2, ..., X_n$ una m.a. de una distribución $\Gamma(\alpha, \lambda)$. Como hay dos parámetros a estimar, planteamos un sistema de ecuaciones basadas en el primer y en el segundo momento.

Usando que si X ~ $\Gamma(\alpha, \lambda)$, $E(X) = \dfrac{\alpha}{\lambda}$ y $V(X) = \dfrac{\alpha}{\lambda^2}$ y la relación: $V(X) = E(X^2) - (E(X))^2$,

$$\begin{cases} \dfrac{\sum_{i=1}^{n} X_i}{n} = E(X) \\ \dfrac{\sum_{i=1}^{n} X_i^2}{n} = E(X^2) \end{cases} \quad \Rightarrow \quad \begin{cases} \dfrac{\sum_{i=1}^{n} X_i}{n} = \dfrac{\alpha}{\lambda} \\ \dfrac{\sum_{i=1}^{n} X_i^2}{n} = \dfrac{\alpha}{\lambda^2} + \left(\dfrac{\alpha}{\lambda}\right)^2 \end{cases}$$

Reemplazando $\dfrac{\alpha}{\lambda} = \overline{X}$, en la segunda ecuación, se obtiene:

$$\frac{\sum_{i=1}^{n} X_i^2}{n} = \frac{\overline{X}}{\lambda} + \overline{X}^2$$

y, despejando λ:

$$\frac{\overline{X}}{\lambda} = \frac{\sum_{i=1}^{n} X_i^2}{n} - \overline{X}^2 \quad \Rightarrow \quad \hat{\lambda} = \frac{\overline{X}}{\dfrac{\sum_{i=1}^{n} X_i^2}{n} - \overline{X}^2}$$

Finalmente, reemplazando el estimador de λ en la primera ecuación, obtenemos el estimador de α:

$$\hat{\alpha} = \frac{\overline{X}^2}{\dfrac{\sum_{i=1}^{n} X_i^2}{n} - \overline{X}^2}$$

3) Sea $X_1, X_2, ..., X_n$ una m.a. de una distribución U(0,θ). Como hay un único parámetro a estimar, planteamos una ecuación basada en el primer momento.

$$\frac{\sum_{i=1}^{n} X_i}{n} = E(X) = \frac{\theta}{2} \qquad \Rightarrow \qquad \hat{\theta} = 2\overline{X}$$

4) Veamos por último un ejemplo que nos muestra que no siempre podemos utilizar los momentos en el orden natural. Sea $X_1, X_2, ..., X_n$ una m.a. de una distribución U(-θ,θ). Como hay un único parámetro a estimar, parece natural plantear una ecuación basada en el primer momento. Sin embargo, si lo hacemos,

$$\frac{\sum_{i=1}^{n} X_i}{n} = E(X) = 0$$

Observamos que el primer momento poblacional no depende de θ y por lo tanto no podemos despejar a partir de esta ecuación el estimador del parámetro. En este caso, es necesario plantear una ecuación basada en el segundo momento:

$$\frac{\sum_{i=1}^{n} X_i^2}{n} = E(X^2) = \frac{(2\theta)^2}{12} = \frac{\theta^2}{3} \qquad \Rightarrow \qquad \hat{\theta} = \sqrt{\frac{3\sum_{i=1}^{n} X_i^2}{n}}$$

Método de máxima verosimilitud: Este método fue introducido por Fisher en la década de 1920. Se basa en la idea de, dada una muestra, hallar los valores de los parámetros que hacen que la probabilidad de obtener dicha muestra sea máxima.

Ejemplo: Se realiza una encuesta de opinión a una m.a. de 20 personas. Se les formula una única pregunta que será respondida por Si o por NO. Sean $X_1, X_2, ..., X_{20}$ las v.a. correspondientes a la respuesta, tales que

$$X_i = \begin{cases} 1 & \text{si la persona } i \text{ responde SI} \\ 0 & \text{si la persona } i \text{ responde NO} \end{cases}$$

para *i* =1, 2, ..., 20 y sea $p = P(X_i = 1)$.

Observemos que las v.a. X_i son independientes y cada una de ellas tiene distribución Bi(1,*p*). Entonces, la función de probabilidad conjunta del vector $(X_1, X_2, ..., X_{20})$ es

$$p(x_1, x_2, ..., x_{20}) = p^{x_1}(1-p)^{1-x_1}\, p^{x_2}(1-p)^{1-x_2} \ldots p^{x_{20}}(1-p)^{1-x_{20}}$$

Si en la muestra obtenida se observan 7 NO's (0) y 13 SI's (1), sería

$$p(x_1, x_2, ..., x_{20}) = p^{13}(1-p)^7$$

La pregunta es: ¿qué valor de *p* hace que los valores muestrales obtenidos sean los más probables?

Es decir, buscamos el valor de p que hace máxima $p(x_1, x_2, ..., x_{20})$ o equivalentemente $\ln p(x_1, x_2, ..., x_{20})$ ya que ln es una función monótona creciente. Debemos maximizar la siguiente función de *p*

$$g(p) = \ln p(x_1, x_2, ..., x_{20}) = 13\ln(p) + 7\ln(1-p)$$

Para ello, como esta función es derivable respecto de *p*, buscamos los posibles puntos críticos, igualando a 0 la derivada primera.

$$0 = \frac{\partial g(p)}{\partial p} = \frac{13}{p} - \frac{7}{1-p} = \frac{13(1-p) - 7p}{p(1-p)} = \frac{13-20p}{p(1-p)} \quad \Leftrightarrow \quad 13\text{-}20p = 0 \quad \Leftrightarrow \quad \hat{p} = \frac{13}{20}$$

Este valor es en efecto el que maximiza *g*(*p*) pues

$$\left.\frac{\partial^2 g(p)}{\partial p^2}\right|_{p=13/20} = \left.-\frac{13}{p^2} - \frac{7}{(1-p)^2}\right|_{p=13/20} < 0$$

<u>Definición</u>: Sean $X_1, X_2, ..., X_n$ v.a. con función de probabilidad conjunta $p_{\vec{X}}(x_1, x_2, ..., x_n)$ o función de densidad conjunta $f_{\vec{X}}(x_1, x_2, ..., x_n)$ que depende de *k* parámetros $\theta_1, \theta_2, ..., \theta_m$. Cuando $(x_1, x_2, ..., x_n)$ son los valores observados y la función de probabilidad o de densidad conjunta se considera función de los parámetros $\theta_1, \theta_2, ..., \theta_m$, se denomina función de verosimilitud y se denota $L(\theta_1, \theta_2, ..., \theta_m)$.

Los estimadores de máxima verosimilitud (**EMV**) de $\theta_1, \theta_2, ..., \theta_m$ son los valores $\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_m$ que maximizan la función de verosimilitud, o sea los valores tales que

$$L(\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_m) \geq L(\tilde{\theta}_1, \tilde{\theta}_2, ..., \tilde{\theta}_m) \qquad \forall\, \tilde{\theta}_1, \tilde{\theta}_2, ..., \tilde{\theta}_m$$

La forma general de los EMV se obtiene reemplazando los valores observados x_i por las v.a. X_i.

<u>Ejemplos</u>: 1) Sea $X_1, X_2, ..., X_n$ una m.a. de una distribución exponencial de parámetro λ.

$$f(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} f_{X_i}(x_i) = \prod_{i=1}^{n} \lambda\, e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$$

por lo tanto, la función de verosimilitud es

$$L(\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$$

Observemos que no incluimos los indicadores porque, dado que el rango de la v.a. no depende del parámetro a estimar, podemos suponer que todas las observaciones son no negativas.

$$\ln L(\lambda) = n \ln(\lambda) - \lambda \sum_{i=1}^{n} x_i$$

$$\frac{\partial \ln L(\lambda)}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0 \qquad \Leftrightarrow \qquad \hat{\lambda} = \frac{n}{\sum_{i=1}^{n} X_i} = \frac{1}{\overline{X}}$$

Verificar que el punto crítico obtenido es en efecto un máximo.

Observemos que en este caso el EMV coincide con el de momentos.

2) Sea $X_1, X_2, ..., X_n$ una m.a. de una distribución $N(\mu,\sigma^2)$.

$$f(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} f_{X_i}(x_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sigma^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

Por lo tanto la función de verosimilitud es

$$L(\mu,\sigma) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sigma^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

y maximizarla equivale a maximizar su logaritmo

$$\ln L(\mu,\sigma) = -n \ln\left(\sqrt{2\pi}\right) - n \ln(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

$$\begin{cases} \dfrac{\partial \ln L(\mu,\sigma)}{\partial \mu} = \dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n}(x_i - \mu) = 0 \\ \dfrac{\partial \ln L(\mu,\sigma)}{\partial \sigma} = -\dfrac{n}{\sigma} + \dfrac{1}{\sigma^3}\displaystyle\sum_{i=1}^{n}(x_i - \mu)^2 = 0 \end{cases} \Leftrightarrow \begin{cases} \displaystyle\sum_{i=1}^{n}(x_i - \mu) = 0 \\ -n\sigma^2 + \displaystyle\sum_{i=1}^{n}(x_i - \mu)^2 = 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} \mu = \dfrac{\sum_{i=1}^{n} x_i}{n} \\ \sigma = \sqrt{\dfrac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}} \end{cases}$$

y, reemplazando el valor estimado de μ en la segunda ecuación, se obtienen los EMV de los parámetros

$$\hat{\mu} = \overline{X} \qquad \hat{\sigma} = \sqrt{\dfrac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}}$$

3) Sea $X_1, X_2, ..., X_n$ una m.a. de una distribución U(0,θ).

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

y la función de verosimilitud es

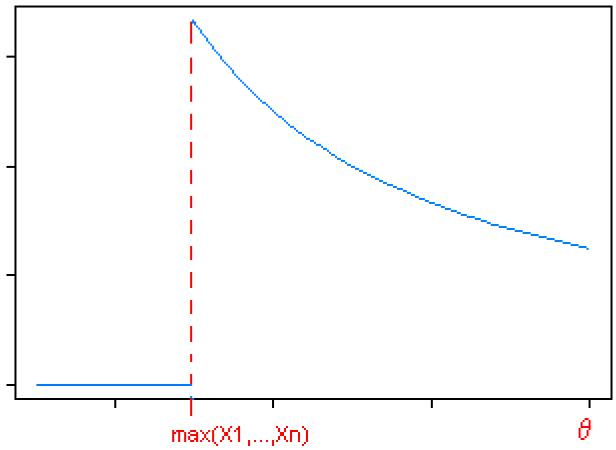
$$L(\theta) = \frac{1}{\theta^n} \prod_{i=1}^{n} I_{(0,\theta)}(x_i)$$

Observemos que, en este caso, no es posible tomar logaritmo ni derivar porque el parámetro (argumento de la función de verosimilitud) determina el soporte de la densidad. Analicemos cómo es esta función a fin hallar su máximo

$$L(\theta) = \begin{cases} \dfrac{1}{\theta^n} & \text{si } 0 < x_i < \theta \ \ \forall i \\ 0 & \text{en caso contrario} \end{cases} = \begin{cases} \dfrac{1}{\theta^n} & \text{si } \max\limits_{1 \le i \le n}(x_i) < \theta \\ 0 & \text{en caso contrario} \end{cases}$$

$$= \begin{cases} \dfrac{1}{\theta^n} & \text{si } \theta > \max\limits_{1\le i\le n}(x_i) \\ 0 & \text{si } \theta \le \max\limits_{1\le i\le n}(x_i) \end{cases}$$

Grafiquemos $L(\theta)$ como función de θ.



Como se puede observar, el máximo de la función de verosimilitud se alcanza en $\theta = \max\limits_{1\le 2\le n}(x_i)$ y por lo tanto el EMV del parámetro es

$$\hat{\theta} = \max_{1\le i\le n}(X_i)$$

Propiedad de Invarianza de los EMV: Sea $\hat{\theta}$ el EMV de θ y sea h una función inyectiva con dominio en el rango de valores posibles de θ, entonces el EMV de $h(\theta)$ es $h(\hat{\theta})$. Por ejemplo, en el caso de una m.a. de una distribución N(μ, σ^2) hemos visto que el EMV de σ es

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

entonces el EMV de σ^2 es

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

pues la función h(x)=x^2 es inyectiva si su dominio se restringe a los reales positivos, es decir si h :$\Re^{\ge 0} \to \Re$.

En general, sean $\hat{\theta}_1,...,\hat{\theta}_m$ los EMV de $\theta_1,...,\theta_m$ y sea una función $h:\Re^k \to \Re$, ¿bajo qué condiciones el EMV de $h(\theta_1,...,\theta_m)$ es $h(\hat{\theta}_1,...,\hat{\theta}_m)$? Esta propiedad, denominada propiedad de invarianza de los EMV, se cumple si la función h puede ser completada a una función inyectiva.

Propiedades de los estimadores y criterios de selección

Observemos que, dada una muestra $X_i, X_2,..., X_n$, un estimador puntual del parámetro θ, obtenido en base a ella, es una v.a. $\hat{\theta}$. La diferencia

$$\hat{\theta} - \theta$$

es el error de estimación y una estimación será más precisa cuanto menor sea este error. Este error es también una v.a. dado que depende de la muestra obtenida. Para algunas muestra será positivo, para otras negativo. Una propiedad deseable es que la esperanza del error sea 0, es decir que "en promedio" el error obtenido al estimar a partir de diferentes muestras sea cero.

Definición: Un estimador puntual $\hat{\theta}$ del parámetro θ es **insesgado** si

$$E(\hat{\theta}) = \theta \qquad \forall\, \theta$$

Si $\hat{\theta}$ no es insesgado, se denomina sesgo de $\hat{\theta}$ a $b(\hat{\theta}) = E(\hat{\theta}) - \theta$.

Por lo tanto, un estimador es insesgado si su distribución tiene como valor esperado al parámetro que se desea estimar.

Definición: Un estimador puntual $\hat{\theta}$ del parámetro θ basado en una muestra $X_1,..., X_n$,es **asintóticamente insesgado** si

$$E(\hat{\theta}) \xrightarrow[n\to\infty]{} \theta \qquad \forall\, \theta$$

Ejemplos: 1) Sea *X:* número de éxitos en *n* repeticiones de un experimento binomial con probabilidad de éxito igual a *p*. Entonces *X* ~ Bi(*n*,*p*) y hemos visto que el EMV de *p* es $\hat{p} = X / n$, o sea la frecuencia relativa de éxitos. Verifiquemos que este estimador es insesgado.

$$E(\hat{p}) = E\left(\frac{X}{n}\right) = \frac{E(X)}{n} = \frac{np}{n} = p \qquad \forall\, p$$

y, por lo tanto, es insesgado.

2) Sea $X_1, X_2,..., X_n$ una m.a. de una distribución N(μ,σ^2). Los EMV de μ y σ^2 son

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

Como $E(\hat{\mu}) = \mu \quad \forall\, \mu$, este estimador es insesgado.

Verifiquemos que el estimador de la varianza no lo es.

$$E(\hat{\sigma}^2) = E\left(\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n}\right) = \frac{1}{n}E\left(\sum_{i=1}^{n}\left(X_i^2 - 2X_i\overline{X} + \overline{X}^2\right)\right)$$

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

$$= \frac{1}{n}E\left(\sum_{i=1}^{n}X_i^2 - n\overline{X}^2\right) = \frac{1}{n}E\left(\sum_{i=1}^{n}X_i^2\right) - E(\overline{X}^2) = \frac{n}{n}E(X_1^2) - E(\overline{X}^2)$$

$$= \left[V(X_1) + \left(E(X_1)\right)^2\right] - \left[V(\overline{X}) + \left(E(\overline{X})\right)^2\right] = \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 = \frac{n-1}{n}\sigma^2$$

Por lo tanto el EMV de la varianza no es insesgado, pero es asintóticamente insesgado ya que su esperanza tiende a σ^2 cuando el tamaño de la muestra tiende a infinito.

<u>Ejercicio</u>: Verificar que la varianza muestral $S^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1}$ es un estimador insesgado de la varianza poblacional cualquiera sea la distribución.

3) Sea $X_1, X_2, ..., X_n$ una m.a. de una distribución U(0,θ). El estimador de momentos de θ es $\hat{\theta} = 2\overline{X}$ y el EMV es $\hat{\theta} = \max_{1 \le i \le n}(X_i)$.

El estimador de momentos es insesgado. En efecto,

$$E(\hat{\theta}) = 2E(\overline{X}) = 2\frac{\theta}{2} = \theta \qquad \forall\ \theta$$

Verificaremos que el EMV no lo es. Para ello, necesitamos obtener la densidad de la v.a. $U = \max_{1 \le i \le n}(X_i)$.

Recordemos que, si $X_1, X_2, ..., X_n$ es una m.a. de una distribución U(0,θ), entonces

$$F_U(u) = \left(F_X(u)\right)^n = \begin{cases} 0 & \text{si } u \le 0 \\ \left(\dfrac{u}{\theta}\right)^n & \text{si } 0 < u < \theta \\ 1 & \text{si } u \ge \theta \end{cases}$$

entonces

$$f_U(u) = n\left(\frac{u}{\theta}\right)^{n-1} \frac{1}{\theta} I_{(0,\theta)}(u) .$$

Calculemos la esperanza del EMV.

$$E\left(\max(X_i)\right) = E(U) = \int_0^\theta u\, n\left(\frac{u}{\theta}\right)^{n-1} \frac{1}{\theta} du = \frac{n}{\theta^n}\int_0^\theta u^n du = \frac{n}{\theta^n} \left.\frac{u^{n+1}}{n+1}\right|_0^\theta = \frac{n}{n+1}\theta$$

Entonces, el EMV no es insesgado pero es asintóticamente insesgado.

Cuando hay más de un estimador insesgado para un mismo parámetro, ¿cómo decidimos cuál conviene usar? Por ejemplo, sea $X_1, X_2, ..., X_n$ una m.a. de una distribución N(μ,σ^2). Es inmediato verificar que los siguientes son todos estimadores insesgados de μ:

$$\hat{\mu}_1 = \overline{X}$$
$$\hat{\mu}_2 = \frac{X_1 + X_2}{2}$$
$$\hat{\mu}_3 = X_1$$

Sin embargo, las varianzas de estos estimadores son

$$V(\hat{\mu}_1) = \frac{\sigma^2}{n}$$
$$V(\hat{\mu}_2) = \frac{\sigma^2}{2}$$
$$V(\hat{\mu}_3) = \sigma^2$$

y parece natural elegir el estimador más preciso, es decir el de menor varianza.

Principio de estimación insesgada de mínima varianza: Entre todos los estimadores insesgados de θ, elegir el de menor varianza. El estimador resultante se denomina **IMVU** (insesgado de mínima varianza uniformemente). Existe una metodología que permite hallar estimadores IMVU en muchas situaciones.

Teorema: Sea $X_1, X_2, ..., X_n$ una m.a. de una distribución N(μ,σ^2). Entonces $\overline{X}$ es estimador IMVU de μ.

A partir de este resultado deducimos que, si se tiene evidencia de que la m.a. proviene de una distribución Normal, parece conveniente usar $\overline{X}$ como estimador de μ. Sin embargo, si los datos no son Normales este estimador podría llegar a ser una pésima elección.

Ejemplo: Sean las siguientes distribuciones simétricas alrededor del parámetro μ

a) N(μ,σ^2) : $f(x) = \dfrac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

b) Cauchy de parámetro μ : $f(x) = \dfrac{1}{\pi\left(1 + (x-\mu)^2\right)}$

c) U(μ -1, μ +1) : $f(x) = \dfrac{1}{2} I_{(\mu-1,\mu+1)}(x)$

La distribución de Cauchy tiene forma de campana como la distribución Normal, pero tiene colas más pesadas que ésta. La distribución Uniforme no tiene colas, por lo tanto podríamos decir que tiene colas más livianas que la Normal.

Consideremos los siguientes estimadores de μ:

$$\hat{\mu}_1 = \overline{X} \qquad \hat{\mu}_2 = \widetilde{X} \qquad \hat{\mu}_3 = \frac{\max(X_i) + \min(X_i)}{2}$$

En el caso a), $\hat{\mu}_1$ es IMVU y por lo tanto, es la elección correcta.

En el caso b), $\hat{\mu}_1$ y $\hat{\mu}_3$ son malos porque ambos son muy sensibles a la presencia de observaciones atípicas y la distribución Cauchy produce una importante proporción de ellas. Por lo tanto la mejor elección entre estos tres estimadores sería $\hat{\mu}_2$. También podríamos utilizar una media podada.

En el caso c) el mejor estimador es $\hat{\mu}_3$ porque la distribución no tiene colas.

Error standard de un estimador: Al informar el resultado de una estimación puntual es necesario brindar información sobre la precisión de la estimación.

Definición: El error standard de un estimador $\hat{\theta}$ es su desviación standard, es decir

$$\sigma_{\hat{\theta}} = \sqrt{V(\hat{\theta})}$$

Si el error standard depende de parámetros desconocidos, éstos se reemplazan por un estimador y se obtiene el error standard estimado.

Ejemplo: Sea $X_1, X_2, ..., X_n$ una m.a. de una distribución N(μ,σ^2). Entonces $\overline{X}$ es el EMV de μ y su error standard es

$$\sigma_{\overline{X}} = \sqrt{V(\overline{X})} = \sqrt{\frac{\sigma^2}{n}}$$

Como depende del parámetro σ , podemos reemplazarlo por la varianza muestral y obtenemos el error standard estimado

$$\hat{\sigma}_{\overline{X}} = \sqrt{\frac{S_X^2}{n}} = \sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n(n-1)}}$$

Definición: Sea $\hat{\theta}$ un estimador de θ, su error cuadrático medio es

$$ECM(\hat{\theta}) = E\left[\left(\hat{\theta} - \theta\right)^2\right]$$

Si el estimador $\hat{\theta}$ es insesgado el error cuadrático medio es igual a la varianza del estimador.

Proposición: $ECM(\hat{\theta}) = V(\hat{\theta}) + \left[b(\hat{\theta})\right]^2$, siendo $b(\hat{\theta}) = E(\hat{\theta}) - \theta$ el sesgo del estimador.

Dem:

$$ECM(\hat{\theta}) = E\left[\left(\hat{\theta} - \theta\right)^2\right] = E\left[\left(\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta\right)^2\right]$$

$$= E\left[\left(\hat{\theta} - E(\hat{\theta})\right)^2 + \left(E(\hat{\theta}) - \theta\right)^2 + 2\left(\hat{\theta} - E(\hat{\theta})\right)\left(E(\hat{\theta}) - \theta\right)\right]$$

$$= E\left[\left(\hat{\theta} - E(\hat{\theta})\right)^2\right] + E\left[\left(E(\hat{\theta}) - \theta\right)^2\right] + 2E\left[\left(\hat{\theta} - E(\hat{\theta})\right)\left(E(\hat{\theta}) - \theta\right)\right]$$

Usando que la esperanza de una v.a. es una constante y la esperanza de una constante es igual a ésta, se obtiene

$$ECM(\hat{\theta}) = \underbrace{E\left[\left(\hat{\theta} - E(\hat{\theta})\right)^2\right]}_{V(\hat{\theta})} + \underbrace{\left(E(\hat{\theta}) - \theta\right)^2}_{\left(b(\hat{\theta})\right)^2} + 2\left(E(\hat{\theta}) - \theta\right)\underbrace{\left(E(\hat{\theta}) - E(\hat{\theta})\right)}_{0}$$

y, por lo tanto, $ECM(\hat{\theta}) = V(\hat{\theta}) + \left[b(\hat{\theta})\right]^2$, como queríamos demostrar.

<u>Principio de estimación de menor error cuadrático medio</u>: Dados dos o más estimadores del parámetro θ, elegir el de menor ECM.

Este principio se reduce, en el caso de estimadores insesgados, al de mínima varianza entre los insesgados mencionado más arriba, ya que el error cuadrático medio se reduce a la varianza cuando un estimador es insesgado. Sin embargo, nos permite además seleccionar, por ejemplo, entre un estimador insesgado y otro que no lo es, en base a la varianza y al sesgo. Si el estimador sesgado tiene una varianza mucho menor que el insesgado, podría ser preferible su uso.

<u>Definición</u>: Sea $X_1, X_2, ..., X_n$ una m.a de una distribución que depende de un parámetro θ y sea $\hat{\theta}_n$ un estimador puntual de θ basado en esa muestra. Diremos que $\{\hat{\theta}_n\}$ es una sucesión **consistente** (o más brevemente que $\hat{\theta}_n$ es un estimador consistente de θ) si

$$\hat{\theta}_n \xrightarrow{p} \theta$$

es decir, si $\forall\, \varepsilon > 0,\ P\left(\left|\hat{\theta}_n - \theta\right| > \varepsilon\right) \xrightarrow[n\to\infty]{} 0$.

<u>Ejemplo</u>: Sea $X_1, X_2, ..., X_n$ una m.a de una distribución con $E(X_i) = \mu$ y $V(X_i) = \sigma^2 < \infty$, entonces $\overline{X}$ es un estimador consistente de μ. En efecto, aplicando la desigualdad de Chebyshev,

$$P\left(\left|\overline{X} - \mu\right| > \varepsilon\right) \leq \frac{V(\overline{X})}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 0 \qquad \forall\, \varepsilon > 0$$

<u>Ejercicio</u>: Verificar que, en este ejemplo, $\hat{\mu} = \dfrac{X_1 + X_2}{2}$ no es consistente de μ.

<u>Proposición</u>: Sea $X_1, X_2, ..., X_n$ una m.a de una distribución que depende de un parámetro θ y sea $\hat{\theta}_n$ un estimador de θ basado en la muestra de tamaño *n*. Si

a) $E(\hat{\theta}_n) \xrightarrow[n\to\infty]{} \theta$ (o sea, si el estimador es asintóticamente insesgado)

b) $V(\hat{\theta}_n) \xrightarrow[n\to\infty]{} 0$

entonces, $\hat{\theta}_n$ es consistente de θ.

Dem: La demostración es inmediata, a partir de la desigualdad de Chebyshev, si el estimador es insesgado. No daremos la demostración en el caso general.

<u>Ejemplos</u>: 1) Sea $X_1, X_2, ..., X_n$ una m.a de una distribución con $E(X_i) = \mu$ y $V(X_i) = \sigma^2 < \infty$, entonces $\overline{X}$ es un estimador consistente de μ. En efecto, $E(\overline{X}) = \mu$ y $V(\overline{X}) = \dfrac{\sigma^2}{n}$. Por lo tanto, se satisfacen las dos condiciones de la Proposición.

2) Sea $X_1, X_2, ..., X_n$ una m.a. de una distribución U(0,θ). Hemos demostrado antes que el EMV de θ, $\hat{\theta} = \max\limits_{1 \le i \le n}(X_i)$ es asintóticamente insesgado pues $E(\hat{\theta}) = \dfrac{n}{n+1}\theta$. Para probar que es consistente, verificaremos que su varianza tiende a cero cuando el tamaño de la muestra tiende a infinito. Pero

$$V(\hat{\theta}) = E(\hat{\theta}^2) - \left[E(\hat{\theta})\right]^2 = E(\hat{\theta}^2) - \left[\frac{n}{n+1}\theta\right]^2$$

entonces, debemos calcular la esperanza del cuadrado de la v.a. $U = \max\limits_{1 \le i \le n}(X_i)$, recordando que su densidad está dada por

$$f_U(u) = n\left(\frac{u}{\theta}\right)^{n-1} \frac{1}{\theta} I_{(0,\theta)}(u)$$

$$E(U^2) = \int_0^\theta u^2 n\left(\frac{u}{\theta}\right)^{n-1} \frac{1}{\theta}\, du = \frac{n}{\theta^n}\int_0^\theta u^{n+1} du = \frac{n}{\theta^n} \left.\frac{u^{n+2}}{n+2}\right|_0^\theta = \frac{n}{n+2}\theta^2$$

Entonces,

$$V(\hat{\theta}) = \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\right)^2\theta^2 = \left(\frac{n}{n+2} - \frac{n^2}{(n+1)^2}\right)\theta^2 = \frac{n}{(n+2)(n+1)^2}\theta^2 \xrightarrow[n\to\infty]{} 0$$

Por lo tanto, el EMV es consistente.

3) Este último ejemplo ilustra como demostrar la consistencia de un estimador a partir de la Ley de los Grandes Números y de las propiedades de la convergencia en probabilidad.

En primer lugar recordemos que si $X_1, X_2, ..., X_n, ...$ e $Y_1, Y_2, ..., Y_n, ...$ son sucesiones de v.a. tales que $X_n \xrightarrow{p} a$ e $Y_n \xrightarrow{p} b$, entonces:

a) $X_n \pm Y_n \xrightarrow{p} a \pm b$

b) $X_n Y_n \xrightarrow{p} ab$

c) $\dfrac{X_n}{Y_n} \xrightarrow{p} \dfrac{a}{b}$ si $b \neq 0$

d) $g(X_n) \xrightarrow{p} g(a)$ si g es una función continua en a.

e) si c_n es una sucesión numérica tal que $c_n \longrightarrow c$, entonces $c_n X_n \xrightarrow{p} ca$ ¿??

Sea $X_1, X_2, ..., X_n$ una m.a de una distribución con $E(X_i) = \mu$ y $V(X_i) = \sigma^2 < \infty$, demostraremos que la varianza muestral S_X^2 es un estimador consistente de la varianza poblacional.

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

Por la Ley de los Grandes Números $\overline{X} \xrightarrow{p} \mu$, entonces por la propiedad d) $\overline{X}^2 \xrightarrow{p} \mu^2$.

Por otra parte, aplicando nuevamente la Ley de los Grandes Números

$$\frac{\sum_{i=1}^{n} X_i^2}{n} \xrightarrow{p} E(X^2) = V(X) + \left[E(X)\right]^2 = \sigma^2 + \mu^2$$

Como además $\dfrac{n}{n-1} \to 1$, se obtiene

$$S_X^2 = \frac{n}{n-1}\left(\frac{\sum_{i=1}^{n} X_i^2}{n} - \overline{X}^2\right) \xrightarrow{p} \sigma^2 + \mu^2 - \mu^2 = \sigma^2$$

y por lo tanto la varianza muestral es un estimador consistente de σ^2.