

Estimación puntual y por intervalo

1.2 Propiedades deseables de los estimadores puntuales

Se conoce la familia de distribuciones a partir de la cual se obtiene la muestra, pero no puede identificarse el miembro de interés de esta, ya que no se conoce el valor del parámetro. Este último tiene que estimarse con base en los datos de la muestra.

El estimador de un parámetro θ debe tener una distribución de muestreo concentrada alrededor de θ y la varianza del estimador debe ser la menor posible. Para ampliar las propiedades anteriores, considérese la siguiente. Sea X_1, X_2, \dots, X_n una muestra aleatoria de tamaño n proveniente de una distribución con función de densidad $f(x, \theta)$, y sea $T = u(X_1, X_2, \dots, X_n)$ cualquier estadística. El problema es encontrar una función u que sea la que proporcione la mejor estimación de θ . Al buscar el mejor estimador de θ se hará uso de una cantidad muy importante que recibe el nombre de error cuadrático medio de un estimador.

Definición 1.1 Sea T cualquier estimador de un parámetro desconocido θ . Se define el error cuadrático medio de T como el valor esperado del cuadrado de la diferencia entre T y θ .

Para cualquier estadística T , se denotará el error cuadrático medio por $ECM(T)$; de esta forma

$$ECM(T) = E(T - \theta)^2.$$

Puede verse la razón del porque el error cuadrático medio es una cantidad importante para enjuiciar a los posibles estimadores de θ mediante el desarrollo de $ECM(T) = E(T - \theta)^2$; este es,

$$\begin{aligned} ECM(T) &= E(T^2 - 2\theta T + \theta^2) \\ &= E(T^2) - 2\theta E(T) + \theta^2 \\ &= \text{Var}(T) + E^2(T) - 2\theta E(T) + \theta^2 \\ &= \text{Var}(T) + [\theta - E(T)]^2. \end{aligned}$$

El error cuadrático medio de cualquier estimador es la suma de dos cantidades no negativas: una es la varianza del estimador y la otra es el cuadrado del sesgo del estimador. Estas dos cantidades se encuentran relacionadas en forma directa con las propiedades deseables de un estimador. La varianza de un estimador debe ser lo más pequeña posible mientras que la distribución de muestreo debe concentrarse del valor del parámetro. Por lo que debemos seleccionar el mejor estimador de θ ; es decir, la estadística que tenga el error cuadrático medio más pequeño posible entre todos los estimadores factibles de θ .

1.2.1 Estimadores insesgados

El término $\theta - E(T)$ recibe el nombre de sesgo del estimador.

Definición 1.2 Se dice que la estadística $T = u(X_1, X_2, \dots, X_n)$ es un estimador insesgado del parámetro θ , si $E(T) = \theta$ para todos los posibles valores de θ . De esta forma, para cualquier estimador insesgado de θ , la distribución de muestreo de T se encuentra centrada alrededor de θ y $ECM(T) = Var(T)$.

En la sección 7.4, demostramos que sin importar la distribución de la población de interés, $E(\bar{X}) = \mu$. Por lo tanto, la media muestral es un estimador insesgado de la media de la población μ para todo los valores de μ . De hecho, si X_1, X_2, \dots, X_n es una muestra aleatoria de la distribución de X con media μ , entonces cualquier X_i de la muestra es un estimador insesgado de μ , dado que $E(X_i) = \mu$ para todo $i = 1, 2, \dots, n$. Además si una estadística T es cualquier combinación lineal de las variables aleatorias de la muestra de manera tal que

$$T = a_1 X_1 + a_2 X_2 + \dots + a_n X_n$$

en donde $\sum_{i=1}^n a_i = 1$, entonces T es un estimador insesgado de μ dado que

$$\begin{aligned} E(T) &= \sum_{i=1}^n a_i E(X_i) \\ &= \sum_{i=1}^n a_i \mu \\ &= \mu. \end{aligned}$$

En la sección 7.5 se demostró que si la varianza muestral S^2 está dada por $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 / (n - 1)$. Entonces, cuando se muestrea una distribución normal, $E(S^2) = \sigma^2$. Lo que demostraremos que es insesgado sin importar cuál sea la distribución de la población de interés. Sea X_1, X_2, \dots, X_n una muestra aleatoria de alguna distribución con una función de densidad no específica. De esta manera, $E(X_i) = \mu$ y $Var(X_i) = \sigma^2$ para toda $i = 1, 2, \dots, n$. Entonces,

$$\begin{aligned} E(S^2) &= E \left[\sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n - 1} \right] \\ &= (n - 1)^{-1} E \left\{ \sum_{i=1}^n [(X_i - \mu) - (\bar{X} - \mu)]^2 \right\} \\ &= (n - 1)^{-1} E \left\{ \sum_{i=1}^n [(X_i - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu) + (\bar{X} - \mu)^2] \right\} \\ &= (n - 1)^{-1} \left[\sum_{i=1}^n E(X_i - \mu)^2 - 2n E(\bar{X} - \mu)^2 \right] \end{aligned}$$

pero por definición $E(X_i - \mu)^2 = Var(X_i) = \sigma^2$ y $E(\bar{X} - \mu)^2 = Var(\bar{X}) = \sigma^2 / n$. Por lo tanto,

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

En otras palabras, S^2 es un estimador insesgado de σ^2 sólo cuando el divisor es igual a $n - 1$. Es la razón por la que se divide por $n - 1$ en lugar de n . S no hará un estimador insesgado de σ .

1.2.2 Estimadores consistentes

Es razonable esperar que un buen estimador de un parámetro θ sea cada vez menor conforme crece el tamaño de la muestra. Se tendrá un mejor estimador de θ si se basa en 30 observaciones que si lo hace con sólo cinco.

Definición 1.3 Sea T el estimador de un parámetro θ , y sea T_1, T_2, \dots, T_n una secuencia de estimadores que representan a T con base en muestras de tamaño $1, 2, \dots, n$, respectivamente. Se dice que T es un estimador consistente (sencillo) para θ si

$$\lim_{n \rightarrow \infty} P(|T_n - \theta| \leq \epsilon) = 1$$

para todos los valores de θ y $\epsilon > 0$. El requisito constituye lo que se denomina convergencia en probabilidad. Es decir, si un estimador es consistente, converge en probabilidad al valor del parámetro que está intentando estimar conforme el tamaño de la muestra crece. Esto implica que la varianza de un estimador consistente T_n disminuye conforme n crece, y la media de T_n tiende hacia donde n crece. Para demostrar que \bar{X} es un estimador consistente de μ , primero se enuncia el teorema de desigualdad de Tchebysheff.

Teorema 1.1 Sea X una variable aleatoria con una función (densidad) de probabilidad $f(x)$ de manera tal que tanto $E(X) = \mu$ como $Var(X) = \sigma^2$ tienen un valor finito. Entonces

$$P(|X - \mu| \leq k\sigma) \geq 1 - \frac{1}{k^2}$$

o

$$P(|X - \mu| > k\sigma) \leq \frac{1}{k^2}$$

para cualquier constante $k \geq 1$. ■

Este teorema permite determinar los límites de las probabilidades de variables aleatorias discretas y continuas sin tener que especificar sus funciones (densidades) de probabilidad. Asegura que la probabilidad de que una variable aleatoria se aleje no más de k desviaciones estándar de la media, es mayor o igual a $1/k^2$ para algún valor de $k \geq 1$. Por ejemplo

$$P(|X - \mu| \leq 2\sigma) \geq 1 - \frac{1}{4}$$

y

$$P(|X - \mu| \leq 3\sigma) \geq 1 - \frac{1}{9}.$$

para cualquier variable aleatoria X con media μ y varianza σ^2 finitas.

Teorema 1.2 Sean X_1, X_2, \dots, X_n , n variables aleatorias IID, tales que $E(X_i) = \mu$ y $Var(X_i) = \sigma^2$ tienen un valor finito para $i = 1, 2, \dots, n$. Entonces, $\bar{X}_n = \sum_{i=1}^n X_i / n$ es un estimador consistente de μ .

Demostración.- Se quiere demostrar que

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \leq \epsilon) = 1.$$

Dado que \bar{X}_n es una variable aleatoria tal que $E(\bar{X}_n) = \mu$ y $Var(\bar{X}_n) = \sigma^2/n$, se deduce del teorema de Tchebysheff que

$$P(|\bar{X}_n - \mu| > k\sigma/\sqrt{n}) \leq 1/k^2.$$

Sea k una constante positiva igual a $\epsilon\sqrt{n}/\sigma$, en donde ϵ es un número real positivo. Entonces,

$$P(|\bar{X}_n - \mu| > \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}.$$

Dado que σ^2 tiene un valor finito, tomando el límite de esta expresión conforme n tiende al infinito se tiene

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| > \epsilon) = 0.$$

Por lo tanto, se concluye que

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \leq \epsilon) = 1.$$

y \bar{X}_n es un estimador consistente de μ . ■

El teorema 8.2 también se conoce como la ley de los grandes números.

1.2.3 Estimadores insesgados de varianza mínima

Para un parámetro que posee un error cuadrático medio mínimo es difícil determinar un estimador para todo los posibles valores del parámetro. Pero podremos analizar uno que tenga un error cuadrático medio mínimo. Si se tiene una varianza mínima para todos los valores posibles de θ . Este estimador recibe el nombre de estimador insesgado de varianza mínima uniforme (VMU)

Definición 1.4 Sea X_1, X_2, \dots, X_n una muestra aleatoria de una distribución cuya función (densidad) de probabilidad es $f(x; \theta)$. Sea la estadística $T = u(X_1, X_2, \dots, X_n)$ un estimador de θ tal que $E(T) = \theta$ y $Var(T)$ es menor que la varianza de cualquier otro estimador insesgado de θ para todos los posibles valores de θ . Se dice entonces que T es un estimador insesgado de varianza mínima de θ .

La varianza de un estimador insesgado es la cantidad más importante para decidir que tan bueno es el estimador para estimar un parámetro θ . Por ejemplo, sean T_1 y T_2 cualesquiera dos estimadores insesgados de θ . Se dice que T_1 es un estimador más eficiente de θ que T_2 si $Var(T_1) \leq Var(T_2)$. Es muy común utilizar el cociente $Var(T_1)/Var(T_2)$ para determinar la eficiencia relativa de T_2 respecto de T_1 . Si los estimadores son sesgados, se emplea sus errores cuadráticos medios para determinar las eficiencias relativas.

Teorema 1.3 **Cota inferior de Cramér-Rao.** Sea X_1, X_2, \dots, X_n una muestra aleatoria de una distribución con una función (densidad) de probabilidad $f(x; \theta)$. Si T es un estimador insesgado de θ , entonces la varianza de T debe satisfacer la siguiente desigualdad

$$Var(T) \geq \frac{1}{n E \left[\left(\frac{\partial \ln f(X; \theta)}{\partial \theta} \right)^2 \right]}.$$

El teorema 8.3 establece un límite inferior para la varianza de un estimador de θ . Para un estimador insesgado cuya varianza se apeg a la cota inferior de Cramér-Rao, se tiene la siguiente definición. ■

Definición 1.5 Si T es cualquier estimador insesgado del parámetro θ tal que

$$\text{Var}(T) \geq \frac{1}{n E \left[\left(\frac{\partial \ln f(X; \theta)}{\partial \theta} \right)^2 \right]},$$

entonces se dice que T es un estimador eficiente de θ .

1.2.4 Estadísticas suficientes

Una estadística suficiente para un parámetro θ es aquella que utiliza toda la información contenida en la muestra aleatoria con respecto a θ . La utilidad de una estadística suficiente recae en el hecho de que si un estimador insesgado de un parámetro θ es una función de una estadística suficiente, entonces tendrá la varianza más pequeña de entre todos los estimadores insesgados de θ que no se encuentren basados en una estadística suficiente. De hecho, si existe el estimador eficiente de θ , se encontrará que este es una estadística suficiente. Un criterio para terminar una estadística suficiente está dado por el siguiente teorema, el cual se conoce como teorema de factorización de Neyman.

Teorema 1.4 Sea X_1, X_2, \dots, X_n una muestra aleatoria de una distribución con una función de densidad de probabilidad $f(x; \theta)$. Se dice que la estadística $T = u(X_1, X_2, \dots, X_n)$ es una estadística suficiente para θ si y sólo si la función de verosimilitud puede factorizarse de la siguiente manera:

$$L(x_1, x_2, \dots, x_n; \theta) = h(t; \theta) g(x_1, x_2, \dots, x_n)$$

para cualquier valor $t = u(x_1, x_2, \dots, x_n)$ de T y en donde $g(x_1, x_2, \dots, x_n)$ no contiene al parámetro θ . ■

1.3 Métodos de estimación puntual

Se estudiará cómo obtener estimadores que, de manera general, tenga buenas propiedades. Específicamente se considerarán los métodos de máxima verosimilitud y el de momentos.

1.3.1 Método por máxima verosimilitud

El método de estimación por máxima verosimilitud, selecciona como estimador a aquél valor del parámetro que tiene la propiedad de maximizar el valor de la probabilidad de la muestra aleatoria observada. En otras palabras, el método de máxima verosimilitud consiste en encontrar el valor del parámetro que maximiza la función de verosimilitud.

Definición 1.6 Sea X_1, X_2, \dots, X_n una muestra aleatoria de una distribución con función (densidad) de probabilidad $f(x; \theta)$, y sea $L(x_1, x_2, \dots, x_n; \theta)$ la verosimilitud de la muestra como función de θ . Si $t = u(x_1, x_2, \dots, x_n)$ es el valor de θ para el cual el valor de la función de verosimilitud es máxima, entonces $T = u(X_1, X_2, \dots, X_n)$ es el estimador de máxima verosimilitud de θ , y t es el estimador de máxima verosimilitud.

Tiene la propiedad deseable de proporcionar estimadores que son funciones de estadísticas suficientes, siempre y cuando el estimador MV sea único. Además, el método MV proporciona el estimador eficiente, si es que existe. Sin, embargo los estimadores MV son generalmente sesgados. El procedimiento para obtener este tipo de estimadores es relativamente directo. Debido a la naturaleza de la función

de verosimilitud se escoge, por lo común, maximizar el logaritmo natural de $L(\theta)$. Esto es, en muchas ocasiones más fácil obtener el estimador MV maximizando $\ln L(\theta)$ que $L(\theta)$.

Ejemplo 1.6 En un experimento binomial se observan $X = x$ éxitos en n ensayos. Obtener el estimador de máxima verosimilitud del parámetro binomial p .

Respuesta.- En este caso la función de verosimilitud es idéntica a la probabilidad de que $X = x$; de esta forma

$$L(x; p) = \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x}, \quad 0 \leq p \leq 1.$$

Entonces,

$$\ln L(x; p) = \ln(n!) - \ln[(n-x)!] - \ln(x!) + (n-x) \ln(1-p).$$

Para encontrar el valor de p , para el cual $\ln L(x; p)$ tiene un valor máximo, se toma la primera derivada con respecto a p y se iguala a cero:

$$\frac{\partial [\ln L(x; p)]}{\partial p} = \frac{x}{p} - \frac{n-x}{1-p} = 0.$$

Después de resolver para p , se obtiene el estimador MV de p el cual recibe el nombre de proporción muestral X/n , y el estimador MV es $p = x/n$. Para confirmar que este valor maximiza a $\ln L(x; p)$, se toma la segunda derivada con respecto a p y se evalúa en x/n :

$$\frac{\partial^2 [\ln L(x; p)]}{\partial p^2} = -\frac{np(1-p) + (x-np)(1-2p)}{[p(1-p)]^2}$$

y

$$\left. \frac{\partial^2 [\ln L(x; p)]}{\partial p^2} \right|_{x/n} = -\frac{x}{(x/n)^2 [1 - (x/n)]} \quad x/n < 1,$$

lo que confirma el resultado, dado que la segunda derivada es negativa. Para un ejemplo específico, si se observan $x = 5$ con base en 25 ensayos independientes, el estimador MC de p es 0.2. ■

Ejemplo 1.7 Sea X_1, X_2, \dots, X_n una muestra aleatoria de una distribución normal con una función de densidad de probabilidad

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Determinar los estimadores de μ y σ^2 .

Respuesta.- Para este problema se procederá de la misma forma que en el caso de un solo parámetro. Dado que la función de verosimilitud depende tanto de μ como de σ^2 , los estimadores MV de μ y σ^2 son los valores para los cuales la función de verosimilitud tiene un valor máximo. De acuerdo con lo anterior

$$\begin{aligned} L(x_1, x_2, \dots, x_n; \mu, \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x_1-\mu)^2}{2\sigma^2}} \dots e^{-\frac{(x_n-\mu)^2}{2\sigma^2}} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2}, \end{aligned}$$

y

$$\ln [L(x_1, x_2, \dots, x_n; \mu, \sigma^2)] = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2.$$

Después de obtener las primeras derivadas parciales con respecto a μ y con respecto a σ^2 e igualándolas a cero, se tiene

$$\frac{\partial [\ln L(\mu, \sigma^2)]}{\partial \mu} = -\frac{2}{2\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0$$

y

$$\frac{\partial [\ln L(\mu, \sigma^2)]}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 = 0.$$

Resolviendo la primera ecuación para μ , sustituyendo este valor en la segunda y resolviendo para σ^2 , se tiene

$$\hat{\mu} = \sum_{i=1}^n \frac{x_i}{n} = \bar{x}$$

y

$$\hat{\sigma}^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n}.$$

■

Si tomamos la segunda derivada notaremos que estos valores maximizan la función de verosimilitud, ellos son los estimadores MV de μ y σ^2 .

Notemos que se ha introducido la acostumbrada notación de sombrero, $\hat{\cdot}$, para denotar un estimador MV. También notemos que el estimador MV de σ^2 es sesgado, confirmándose de esta manera un comentario anterior en el sentido en el que los estimadores MV no necesariamente son insesgados.

El método de máxima verosimilitud posee otra propiedad deseable conocida como propiedad de invarianza. Sea $\hat{\theta} = u(X_1, X_2, \dots, X_n)$ el estimador de máxima verosimilitud de θ . Si $g(\theta)$ es una función univaluada de θ . Por ejemplo, dado que, cuando se muestrea una distribución normal, el estimador MV de σ^2 es

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2,$$

por la propiedad de invarianza, el estimador MV de la desviación estándar σ es

$$\hat{\sigma} = \left[\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^{\frac{1}{2}}.$$

Como ejemplo adicional de la propiedad de invarianza, el estimador MV de la función de confiabilidad Weibull es

$$\hat{R}(t) = e^{-\frac{t^\alpha}{\hat{\theta}}}$$

donde $\hat{\theta}$ es el estimador MV del parámetro de escalar θ .

1.3.2 Método de los momentos

Definición 1.7 Sea X_1, X_2, \dots, X_n una muestra aleatoria de una distribución con función (densidad) de probabilidad $f(x; \theta)$. El r -ésimo momento alrededor de cero se define como

$$M'_r = \frac{1}{n} \sum_{i=1}^n X_i^r.$$

El método de los momentos proporciona una alternativa razonable cuando no se pueden determinar los estimadores de máxima verosimilitud. Recuérdese que los parámetros son, en general, funciones de los momentos teóricos. En esencia, el método se implementa igualando tantos momentos muestrales con los correspondientes momentos teóricos tantas veces como sea necesario para determinar un estimador de momentos para un parámetro desconocido.

1.3.3 Estimación por máxima verosimilitud para muestras censuradas