

# 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  $\theta$  debe tener una distribución de muestreo concentrada alrededor de  $\theta$  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  $\theta$ . Al buscar el mejor estimador de  $\theta$  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  $\theta$ . Se define el error cuadrático medio de  $T$  como el valor esperado del cuadrado de la diferencia entre  $T$  y  $\theta$ .

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  $\theta$  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  $\theta$ ; es decir, la estadística que tenga el error cuadrático medio más pequeño posible entre todos los estimadores factibles de  $\theta$ .

### 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  $\theta$ , si  $E(T) = \theta$  para todos los posibles valores de  $\theta$ . De esta forma, para cualquier estimador insesgado de  $\theta$ , la distribución de muestreo de  $T$  se encuentra centrada alrededor de  $\theta$  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  $\mu$  para todo los valores de  $\mu$ . De hecho, si  $X_1, X_2, \dots, X_n$  es una muestra aleatoria de la distribución de  $X$  con media  $\mu$ , entonces cualquier  $X_i$  de la muestra es un estimador insesgado de  $\mu$ , 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  $\mu$  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  $\sigma^2$  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  $\sigma$ .

### 1.2.2 Estimadores consistentes

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

**Definición 1.3** Sea  $T$  el estimador de un parámetro  $\theta$ , 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  $\theta$  si

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

para todos los valores de  $\theta$  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  $\mu$ , 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  $\mu$  y varianza  $\sigma^2$  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  $\mu$ .

**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  $\epsilon$  es un número real positivo. Entonces,

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

Dado que  $\sigma^2$  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  $\mu$ . ■

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  $\theta$ . 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  $\theta$  tal que  $E(T) = \theta$  y  $Var(T)$  es menor que la varianza de cualquier otro estimador insesgado de  $\theta$  para todos los posibles valores de  $\theta$ . Se dice entonces que  $T$  es un estimador insesgado de varianza mínima de  $\theta$ .

La varianza de un estimador insesgado es la cantidad más importante para decidir que tan bueno es el estimador para estimar un parámetro  $\theta$ . Por ejemplo, sean  $T_1$  y  $T_2$  cualesquiera dos estimadores insesgados de  $\theta$ . Se dice que  $T_1$  es un estimador más eficiente de  $\theta$  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  $\theta$ , 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  $\theta$ . 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  $\theta$  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  $\theta$ .

#### 1.2.4 Estadísticas suficientes

Una estadística suficiente para un parámetro  $\theta$  es aquella que utiliza toda la información contenida en la muestra aleatoria con respecto a  $\theta$ . La utilidad de una estadística suficiente recae en el hecho de que si un estimador insesgado de un parámetro  $\theta$  es una función de una estadística suficiente, entonces tendrá la varianza más pequeña de entre todos los estimadores insesgados de  $\theta$  que no se encuentren basados en una estadística suficiente. De hecho, si existe el estimador eficiente de  $\theta$ , 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  $\theta$  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  $\theta$ . ■

### 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  $\theta$ . Si  $t = u(x_1, x_2, \dots, x_n)$  es el valor de  $\theta$  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  $\theta$ , 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  $\mu$  y  $\sigma^2$ .

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  $\mu$  como de  $\sigma^2$ , los estimadores MV de  $\mu$  y  $\sigma^2$  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  $\mu$  y con respecto a  $\sigma^2$  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  $\mu$ , sustituyendo este valor en la segunda y resolviendo para  $\sigma^2$ , 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  $\mu$  y  $\sigma^2$ .

Notemos que se ha introducido la acostumbrada notación de sombrero,  $\hat{\cdot}$ , para denotar un estimador MV. También notemos que el estimador MV de  $\sigma^2$  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  $\theta$ . Si  $g(\theta)$  es una función univaluada de  $\theta$ . Por ejemplo, dado que, cuando se muestrea una distribución normal, el estimador MV de  $\sigma^2$  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  $\sigma$  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  $\theta$ .

### 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

En algunas situaciones de muestreo, en forma especial en las pruebas de duración, el procedimiento de prueba puede terminar antes de proporcionar una muestra aleatoria completa. Una prueba típica de duración consiste en artículos iguales (tales como componentes eléctricos o mecánicos) seleccionados en forma aleatoria de un proceso y operando en un medio cuidadosamente controlado hasta que el artículo falla. Si la prueba de duración se termina sólo cuando todas las unidades de la muestra han fallado, se dice que la muestra aleatoria de tiempos está completa. Pero por varias razones la prueba termina ya sea después de un lapso de tiempo predeterminado  $x_0$  o después de que falla un determinado número de unidades  $m \leq n$ . Las dos condiciones producen muestras censuradas. Si  $X_0$  es un lapso de fijo de tiempo, el número de unidades que fallan de las  $n$ , desde el comienzo de la prueba hasta el tiempo  $x_0$  es una variable aleatoria; esta constituye una muestra censurada de tipo *I*. Si  $m$  es fijo y el tiempo de terminación  $X_m$  es la variable aleatoria, se dice que la muestra es de tipo *II*.

Supongamos que los tiempos de duración de las unidades son variables aleatorias  $X_1, X_2, \dots, X_n$  independientes exponencialmente distribuidas, con una función

$$f(x; \theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}.$$

El interés recae en encontrar el estimador de máxima verosimilitud del parámetro  $\theta$ . La función de verosimilitud para un muestreo censurado del tipo *II* es la probabilidad conjunta de que fallen  $m$  unidades en los tiempos  $x_1, x_2, \dots, x_m$  en ese orden, y sobrevivan  $n - m$  unidades con un tiempo de supervivencia igual a  $x_m$ . La parte de la función de verosimilitud que corresponde a las  $m$  unidades que han fallado en los tiempos  $x_1, x_2, \dots, x_m$  es  $f(x_1; \theta) \cdot f(x_2; \theta) \cdot \dots \cdot f(x_m; \theta)$ . Pero esta sólo es una de las posibles formas en que pueden fallar  $m$  unidades de un total de  $n$ . El número total de formas es  $n! / (n - m)!$ . La probabilidad de que  $n - m$  unidades sobrevivan un tiempo  $x_m$ , está dada por la función de confiabilidad a tiempo  $x_m$ ; de esta forma, para la distribución exponencial,

$$P(X > x_m) = e^{-\frac{x_m}{\theta}}.$$

Por lo tanto, la función de verosimilitud Es

$$\begin{aligned} L(x_1, x_2, \dots, x_m; \theta) &= \frac{n!}{(n - m)!} \left[ \frac{1}{\theta} e^{-\frac{x_1}{\theta}} \cdot \dots \cdot \frac{1}{\theta} e^{-\frac{x_1}{\theta}} e^{-\frac{x_1}{\theta}} \cdot \dots \cdot e^{-\frac{x_1}{\theta}} \right] \\ &= \frac{n!}{(n - m)!} \left[ \frac{1}{\theta^m} e^{-\frac{1}{\theta} \sum_{i=1}^m x_i} \cdot e^{-\frac{n-m}{\theta} x_m} \right] \\ &= \frac{n!}{(n - m)!} \left( \frac{1}{\theta} e^{-\frac{1}{\theta} T_m} \right). \end{aligned}$$

Donde

$$T_m = \sum_{i=1}^m x_i + (n - m)x_m.$$

Tomando el logaritmo natural de  $L$ , se tiene

$$\ln L(x_1, x_2, \dots, x_m; \theta) = \ln(n!) - \ln[(n - m)!] - m \ln \theta - \frac{1}{\theta} T_m.$$

Entonces



$$\frac{d [\ln L(x_1, x_2, \dots, x_m; \theta)]}{d\theta} = -\frac{m}{\theta} + \frac{1}{\theta^2} T_M$$

e igualando la derivada a cero, el estimador de máxima verosimilitud de  $\theta$  es

$$\hat{\theta} = \frac{\sum_{i=1}^m x_i + (n-m)x_m}{m}.$$

#### 1.4 Estimación por intervalo

Una media muestral, por ejemplo  $\bar{x} = 200$  unidades, es un estimador puntual de un parámetro desconocido. Este estimador, ¿implica que la demanda media desconocida no sea mayor de 250 ni menor de 150?. En este punto es difícil saberlo, ya que no se tiene ninguna indicación del posible error en el estimador puntual. El error en el estimado puntual se mide en términos de la variación muestral del correspondiente estimador.

Por ejemplo, supóngase que la desviación estándar de la media muestral  $\bar{X}$  es 60 unidades. De acuerdo con el teorema central del límite puede argumentarse que  $\bar{X} \rightarrow N(\mu, 60)$ , conforme  $n \rightarrow \infty$ . De esta forma, la probabilidad de que  $\bar{X}$  se encuentre dentro de dos desviaciones estándar alrededor de  $\mu$ , es de, aproximadamente, 0.95. En otras palabras, para  $n$  grande,

$$P(|\bar{X} - \mu| < 120) = 0.95,$$

o

$$P(-120 < \bar{X} - \mu < 120) = 0.95.$$

Restando  $X$  y multiplicando por  $-1$  en el interior de los paréntesis, se tiene

$$P(\bar{X} - 120 < \mu < \bar{X} + 120) = 0.95.$$

Si se sustituye el estimado para  $\bar{x} = 200$ , se tiene

$$P(80 < \mu < 320) = 0.95,$$

lo que sugiere que es enteramente posible que la demanda sea tan grande como 250 unidades o tan pequeña como 150 unidades, siempre que d.e. ( $\bar{X}$ )=60. Por otro lado, supóngase que la desviación estándar de  $\bar{X}$  es igual a 10. Entonces, la expresión correspondiente a dos desviaciones estándar, es

$$P(\bar{X} - 20 < \mu < \bar{X} + 20) = 0.95,$$

y para  $\bar{X} = 200$ ,

$$P(180 < \mu < 220) = 0.95.$$

En este caso es poco probable que  $\mu$  sea tan grande como 250 o tan pequeño como 150. En ambos casos la clave está en la desviación estándar del estimador puntual. En esencia, para la estimación del intervalo se considera, tanto el estimador puntual del parámetro  $\theta$ , como su distribución de muestreo.

El intervalo  $[\bar{X} - 120, \bar{X} + 120]$  es un intervalo aleatorio, y la probabilidad de que este intervalo contenga el valor esperado de  $\mu$  es de 0.95. En otras palabras, si se obtuviesen muestras del mismo tamaño en forma repetida de una población, y cada vez que estas se seleccionan, se calculan los valores específicos para el intervalo aleatorio  $[\bar{X} - 120, \bar{X} + 120]$ ; entonces debe esperarse que un 95% de estos intervalos contengan el valor de la media desconocida  $\mu$ . La probabilidad de 0.95 para el intervalo aleatorio sugiere que la confianza en que el intervalo (80, 320) contenga el valor de la media desconocida  $\mu$  es alta. Así, cuando se escribe

$$P(80 < \mu < 320) = 0.95,$$

no se está formulando ninguna proposición probabilística en el sentido clásico, sino más bien se expresa un grado de confianza. El intervalo  $(80, 320)$  recibe el nombre de intervalo de confianza del 95% para  $\mu$ .

En términos generales, la construcción de un intervalo de confianza para un parámetro desconocido  $\theta$ , consiste en encontrar una estadística suficiente  $T$  y relacionarla con otra variable aleatoria  $X' = f(T; \theta)$ , donde  $X$  involucra a  $\theta$  pero la distribución de  $X$  no contiene a  $\theta$ , así como tampoco a ningún otro parámetro desconocido. Entonces se seleccionan dos valores  $x_1$  y  $x_2$  tales que

$$P(x_1 < X < x_2) = 1 - \alpha,$$

donde  $1 - \alpha$  recibe el nombre de **coeficiente de confianza**. Mediante una manipulación algebraica de las dos expresiones, se puede modificar el contenido entre paréntesis y expresarlo como

$$P(h_1(T) < \theta < h_2(T)) = 1 - \alpha,$$

donde  $h_1(T)$  y  $h_2(T)$  son funciones de la estadística  $T$  y de esta forma, variables aleatorias. Al seguirse el desarrollo podemos encontrar intervalos de confianza unilaterales, de la forma

$$P[g_1(T) < \theta] = 1 - \alpha$$

o

$$P[\theta < g_2(T)] = 1 - \alpha.$$

#### 1.4.1 Intervalos de confianza para $\mu$ cuando se muestrea una distribución normal con varianza conocida

El interés recae en la construcción de un intervalo de confianza de un  $100(1 - \alpha)\%$  sobre  $\mu$  y donde  $\alpha$  es un número pequeño, tal que  $0 < \alpha < 1$ . La construcción de un intervalo de confianza se hace con base en el mejor estimador de  $\mu$ , explícitamente de la media muestral  $\bar{X}$ .

Sea

$$P(-120 < \bar{X} - \mu < 120) = 0.95.$$

Sumando  $\mu$ , se tiene

$$P(\mu - 120 < \bar{X} < \mu + 120) = 0.95.$$

De esta forma, los límites  $\mu - 120$  y  $\mu + 120$  son funciones de los posibles valores de  $\mu$ . Por lo tanto, y en general, se puede escribir

$$P[g_1(\mu) < \bar{X} < g_2(\mu)] = 1 - \alpha$$

de manera tal que

$$\int_{-\infty}^{g_1(\mu)} f(\bar{x}; \mu) d\bar{x} = \frac{\alpha}{2}$$

y

$$\int_{g_2(\mu)}^{\infty} f(\bar{x}; \mu) d\bar{x} = \frac{\alpha}{2},$$

donde  $f(\bar{x}; \mu)$  es la función de densidad de la distribución de muestreo de  $\bar{X}$ , y  $g_1(\mu)$  y  $g_2(\mu)$  son funciones de  $\mu$  las cuales no contiene a ningún otro parámetro desconocido.

Determinemos  $g_1(\mu)$  y  $g_2(\mu)$ . Dado que  $\bar{X} \sim N(\mu, \sigma/\sqrt{n})$ , la normal estándar  $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ , y

$$P[g_1(\mu) < \bar{X} < g_2(\mu)] = P\left[\frac{g_1(\mu) - \mu}{\sigma/\sqrt{n}} < Z < \frac{g_2(\mu) - \mu}{\sigma/\sqrt{n}}\right] = 1 - \alpha.$$

Pero ya que  $P(z_{\alpha/2} < Z < z_{1-\alpha/2}) = 1 - \alpha$ , en donde los valores cuantiles  $z_{\alpha/2}$  y  $z_{1-\alpha/2}$  son tales que  $P(Z < z_{\alpha/2}) = \alpha/2$  y  $P(Z < z_{1-\alpha/2}) = 1 - \alpha/2$ , respectivamente. Se sigue que

$$\frac{g_1(\mu) - \mu}{\sigma/\sqrt{n}} = z_{\alpha/2}$$

y

$$\frac{g_2(\mu) - \mu}{\sigma/\sqrt{n}} = z_{1-\alpha/2}.$$

Dando solución en términos de  $g_1(\mu)$  y  $g_2(\mu)$ , respectivamente, se obtienen

$$g_1(\mu) = \mu - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

y

$$g_2(\mu) = \mu + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}.$$

(Dado que para la normal estándar  $z_{\alpha/2} = -z_{1-\alpha/2}$  puede sustituirse  $-z_{1-\alpha/2}$  para  $z_{\alpha/2}$  en la ecuación de arriba). Por lo tanto,

$$P\left(\mu - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} < \bar{X} < \mu + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

Al manipular las desigualdades que se encuentran dentro de los paréntesis se tiene

$$P\left(\bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Si se reemplaza la variable aleatoria  $X$  por el estimador  $\bar{x}$  calculando a partir de los datos de una muestra de tamaño  $n$ , un intervalo de confianza del  $100(1 - \alpha)\%$  para  $\mu$ , es

$$\bar{x} \pm z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}},$$

en donde  $\bar{x} - z_{1-\alpha/2}(\sigma/\sqrt{n})$  y  $\bar{x} + z_{1-\alpha/2}(\sigma/\sqrt{n})$  reciben el nombre de límites de confianza inferiores y superiores, respectivamente, para  $\mu$ .

Vemos que mientras más grande es el tamaño de la muestra, más pequeño es el ancho del intervalo; o para un coeficiente de confianza  $1 - \alpha$  más grande, mayor es el ancho de intervalo. Es decir, un tamaño grande de la muestra disminuirá la varianza del estimador, y un coeficiente de confianza grande incrementa el valor cuantil dando como resultado un intervalo más amplio.

#### 1.4.2 Intervalos de confianza para $\mu$ cuando se muestrea una distribución normal con varianza desconocida

Recordemos que cuando se muestrea una  $N(\mu, \sigma)$ , donde tanto  $\mu$  como  $\sigma^2$  son desconocidos, la variables aleatoria

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

tiene distribución  $t$  de Student con  $n - 1$  grados de libertad. Por lo tanto, es posible determinar el valor cuantil  $t_{1-\alpha/2, n-1}$  de  $T$ , para el cual

$$P(-t_{1-\alpha/2, n-1} < T < t_{1-\alpha/2, n-1}) = 1 - \alpha,$$

en donde el valor cuantil es tal que  $P(T < t_{1-\alpha/2, n-1}) = \alpha/2$  y  $P(T < t_{1-\alpha/2, n-1}) = 1 - \alpha/2$ . Al sustituir para  $T$ , se tiene

$$P\left(-t_{1-\alpha/2, n-1} < \frac{\bar{X} - \mu}{S/\sqrt{n}} < t_{1-\alpha/2, n-1}\right) = 1 - \alpha.$$

y

$$P\left(\bar{X} - t_{1-\alpha/2, n-1} \frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{1-\alpha/2, n-1} \frac{S}{\sqrt{n}}\right) = 1 - \alpha.$$

Por lo tanto, el intervalo  $\bar{X} \pm t_{1-\alpha/2, n-1} S/\sqrt{n}$  es un intervalo aleatorio y la probabilidad de que este contenga el valor verdadero de  $\mu$ , es  $1 - \alpha$ . De esta forma, dados ñps datps de una muestra aleatoria de tamaño  $n$  a partir de las cuales se calculan los estimados  $\bar{x}$  y  $s^2$ . Un intervalo de confianza del  $100(1 - \alpha)\%$  para  $\mu$  es

$$\bar{x} \pm t_{1-\alpha/2, n-1} \frac{s}{\sqrt{n}}.$$

### 1.4.3 Intervalo de confianza para la diferencia de medias cuando se muestrean dos distribuciones normales independientes

Sean  $X_1, X_2, \dots, X_{n_X}$  e  $Y_1, Y_2, \dots, Y_{n_Y}$  dos muestras aleatorias de dos distribuciones normales independientes, con medias  $\mu_X$  y  $\mu_Y$  y varianzas  $\sigma_X^2$  y  $\sigma_Y^2$ , respectivamente. Se desea construir un intervalo de confianza para la diferencia  $\mu_X - \mu_Y$ . Supóngase que se conocen los valores de las varianzas. Entonces la variable aleatoria

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}}$$

es  $N(0, 1)$ . De esta forma es posible encontrar el valor cuantil  $z_{1-\alpha/2}$  tal que

$$P(-z_{1-\alpha/2} < Z < z_{1-\alpha/2}) = 1 - \alpha.$$

De donde se tiene

$$P\left(\bar{X} - \bar{Y} - z_{1-\alpha/2} \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}} = 1 - \alpha < \mu_X - \mu_Y < \bar{X} - \bar{Y} + z_{1-\alpha/2} \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}\right) = 1 - \alpha,$$

que es un intervalo aleatorio que no contiene parámetros desconocidos. Un intervalo de confianza del  $100(1 - \alpha)\%$  para  $\mu_X - \mu_Y$  es

$$\bar{x} - \bar{y} \pm z_{1-\alpha/2} \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}},$$

en donde el valor cuantil  $z_{1-\alpha/2}$  es tal que  $P(Z < z_{1-\alpha/2}) = 1 - \alpha/2$ .

Si las varianzas  $\sigma_X^2$  y  $\sigma_Y^2$  se desconocen pero son iguales, entonces la variable aleatoria

$$T = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{S_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}}$$

tiene una distribución t de Student con  $k = n_X + n_Y - 2$  grados de libertad. Al seguir el procedimiento anterior, se tiene que un intervalo de confianza del  $100(1 - \alpha)\%$  para  $\mu_X - \mu_Y$ , es

$$\bar{x} - \bar{y} \pm t_{1-\alpha/2, n_X + n_Y - 2} s_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}},$$

en donde el estimado combinando de la varianza común es

$$s_p^2 = \frac{(n_X - 1)s_X^2 + (n_Y - 1)s_Y^2}{n_X + n_Y - 2}.$$

#### 1.4.4 Intervalo de confianza para $\sigma^2$ cuando se muestrean dos distribución normal con media desconocida

Recordemos que la distribución de muestreo de  $(n-1)S^2/\sigma^2$  es chi-cuadrada con  $n-1$  grados de libertad. Entonces, es posible determinar los valores cuantiles  $\chi^2_{\alpha/2, n-1}$  y  $\chi^2_{1-\alpha/2, n-1}$ , tales que

$$P \left[ \chi^2_{\alpha/2, n-1} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{1-\alpha/2, n-1} \right] = 1 - \alpha.$$

De donde

$$P \left[ \frac{1}{\chi^2_{\alpha/2, n-1}} > \frac{\sigma^2}{(n-1)S^2} > \frac{1}{\chi^2_{1-\alpha/2, n-1}} \right] = 1 - \alpha$$

Entonces el intervalo

$$\left[ \frac{(n-1)S^2}{\chi^2_{1-\alpha/2, n-1}}, \frac{(n-1)S^2}{\chi^2_{\alpha/2, n-1}} \right]$$

es un intervalo aleatorio el cual contiene a  $\sigma^2$  y a parámetros conocidos con una probabilidad de  $1 - \alpha$ . De esta forma, con base en los datos de una muestra aleatoria de tamaño  $n$ , se calcula el estimado  $s^2$  y un intervalo de confianza del  $100(1 - \alpha)\%$  para  $\sigma^2$ , es de  $(n-1)s^2/\chi^2_{1-\alpha/2, n-1}$  a  $(n-1)s^2/\chi^2_{\alpha/2, n-1}$ . Es interesante notar que la variable aleatoria pivotal es  $(n-1)S^2/\sigma^2$  ya que su función de densidad, dada por

$$f(y; n-1) = \begin{cases} \frac{1}{\Gamma[(n-1)/2] 2^{(n-1)/2}} y^{[(n-1)/2]-1} e^{-y/2} & y > 0, \\ 0 & \text{para cualquier otro valor.} \end{cases}$$

no contiene ningún parámetro desconocido.

#### 1.4.5 Intervalos de confianza para el cociente de dos varianzas cuando se muestrean dos distribuciones normales independientes

Supongamos que se tienen muestras aleatorias provenientes de dos distribuciones normales con medias y varianzas desconocidas. Sean  $n_X$  y  $n_Y$ , el tamaño de las muestras y  $S_X^2$  y  $S_Y^2$  las varianzas muestrales. El interés se centra en construir un intervalo de confianza para el cociente  $\sigma_Y^2/\sigma_X^2$  de las dos variables poblacionales. Recordemos que la variable aleatoria  $(S_X^2/\sigma_X^2)(S_Y^2/\sigma_Y^2)$  tiene una distribución  $F$  con  $n_X - 1$  y  $n_Y - 1$  grados de libertad. Entonces puede escribirse

$$P \left( a < \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} < b \right) = 1 - \alpha.$$

en donde  $a$  y  $b$  son los valores cuantiles inferiores y superiores de una distribución  $F$  tal que

$$a = \frac{1}{f_{1-\alpha/2, n_Y-1, n_X-1}} \quad \text{y} \quad b = f_{1-\alpha/2, n_X-1, n_Y-1}.$$

La proposición de probabilidad se puede expresar como

$$P \left( a < \frac{S_X^2}{S_Y^2} \cdot \frac{\sigma_Y^2}{\sigma_X^2} < b \right) = 1 - \alpha$$

o

$$P \left( \frac{aS_Y^2}{S_X^2} < \frac{\sigma_Y^2}{\sigma_X^2} < \frac{bS_Y^2}{S_X^2} \right) = 1 - \alpha$$

De esta manera, un intervalo de confianza del  $100(1 - \alpha)\%$  para  $\sigma_Y^2/\sigma_X^2$  está dado por

$$\left( \frac{as_Y^2}{s_X^2}, \frac{bs_Y^2}{s_Y^2} \right).$$

#### 1.4.6 Intervalos de confianza para el parámetro de proporción $p$ cuando se muestrea una distribución binomial

El número de unidades defectuosas es una variable aleatoria binomial, si se supone una probabilidad constante e independiente. En una muestra aleatoria de tamaño  $n$  el parámetro  $p$  que representa la proporción de artículos defectuosos es desconocido. Se desea determinar un intervalo de confianza para  $p$ . Se demostró que el estimador de máxima verosimilitud de  $p$ , denotado por  $\hat{P}$ , es

$$\hat{P} = \frac{X}{n}.$$

en donde  $X$  es binomial con parámetros  $n$  y  $p$ . Nótese que  $\hat{P}$  es un estimador insesgado de  $p$ , ya que

$$E(\hat{P}) = \frac{1}{n}E(X) = \frac{np}{n} = p.$$

La varianza de  $\hat{P}$  se puede obtener de la siguiente forma:

$$\begin{aligned} \text{Var}(\hat{P}) &= \text{Var}\left(\frac{X}{n}\right) \\ &= \frac{1}{n^2} [np(1-p)] \\ &= \frac{p(1-p)}{n}. \end{aligned}$$

Recuérdese que para  $n$  grande, la variable aleatoria  $(X - np)/\sqrt{np(1-p)}$  es aproximadamente  $N(0, 1)$ . Entonces puede demostrarse que la distribución De

$$\frac{\hat{P} - p}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}}$$

también tiende a  $N(0, 1)$  para  $n$  grande. De esta forma, la probabilidad de intervalo aleatorio

$$\left[ \hat{P} - z_{1-\alpha/2} \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}, \hat{P} + z_{1-\alpha/2} \sqrt{\frac{\hat{P}(1-\hat{P})}{n}} \right]$$

es en forma aproximada,  $1 - \alpha$  para  $n$  grande. De acuerdo con lo anterior, un intervalo de confianza aproximado del  $100(1 - \alpha)\%$  para el parámetro de proporción  $p$ , es

$$\left[ \hat{p} - z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right],$$

en donde el estimador de máxima verosimilitud  $\hat{p} = x/n$  se obtiene de la muestra aleatoria de tamaño  $n$ .

Con respecto al muestreo de una distribución binomial, un problema que surge, en forma frecuente, es el de estimar el tamaño de la muestra necesario de manera tal que con una confiabilidad de  $100(1 - \alpha)\%$  aproximadamente, el estimado del parámetro de proporción se encuentre a no más de  $\epsilon$  unidades de  $p$ . Dado el estimador de máxima verosimilitud  $X/n$ , puede expresarse

$$\left[ \hat{p} - z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$

como

$$P\left(\left|\frac{X}{n} - p\right| < \epsilon\right) = 1 - \alpha.$$

en donde

$$\epsilon = z_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}}$$

Al resolver para  $n$ , se obtiene

$$n = \frac{z_{1-\alpha/2}^2 p(1-p)}{\epsilon^2}.$$

Utilizaremos el valor máximo de  $p$ , es decir  $p = 1/2$ . Esto será válido siempre que el tamaño de la muestra sea suficientemente grande.

Para ilustrar el uso de la distribución  $t$  de Student sigue siendo válido para inferencias con respecto a las medias, aún a pesar de que se muestree una distribución que no es normal.

## 1.5 Estimación bayesiana