

TEMA 4

ESTIMACIÓN PUNTUAL

4.1 INTRODUCCIÓN

La definición de una distribución de probabilidad implica la necesidad de conocer tanto el modelo de probabilidad como los valores de los parámetros sin cuya concreción no es posible utilizar dicha distribución. En cuanto al primero podemos suponer su conocimiento previo, bien por el propio experimento aleatorio que estamos explicando como por el previo conocimiento que tengamos de la variable. Así pues, el investigador puede conocer el modelo de probabilidad pero requiere conocer los valores que podemos asignar a los parámetros. Esta situación se puede expresar formalmente de la siguiente manera: sea X una variable aleatoria poblacional cuya función de distribución $F(x)$ es conocida salvo un parámetro θ que es desconocido. El problema que se plantea es “aproximar” el valor de dicho parámetro mediante una muestra aleatoria simple (X_1, X_2, \dots, X_n) extraída de la población X .

La ESTIMACIÓN es la parte de la Inferencia Estadística que está dedicada a resolver el problema de asignar un valor al parámetro desconocido, es decir, el proceso por el que intentamos averiguar el valor de un parámetro poblacional a partir de un conjunto de datos observados que son la muestra aleatoria. Hay dos enfoques alternativos para abordar la estimación: el enfoque puntual y por intervalo. El primero busca dar un único valor al parámetro desconocido, consiguiendo una gran precisión pero poca fiabilidad. El segundo enfoque busca un intervalo de valores, es decir, un extremo inferior y otro superior, en el que esté incluido el parámetro desconocido con una probabilidad prefijada de antemano. De esta forma, el proceso tendrá mayor fiabilidad aunque no sea tan preciso.

Para ello extraemos una muestra de la población y se establece una función de los valores muestrales que denominamos **estimador**: $T=T(X_1, X_2, \dots, X_n)$. El valor que toma esta función con una muestra concreta es el valor que asignamos al parámetro desconocido y le denominaremos como **estimación puntual de θ** .

El párrafo anterior enfatiza la diferencia entre los conceptos de estimación puntual (resultado) y estimador (procedimiento de trabajo). Hay infinitas funciones de los valores muestrales, por lo tanto, puede haber infinitos estimadores, lo que implica que existen infinitas estimaciones puntuales. El problema que resuelve este tema es averiguar que estimaciones puntuales son más aceptables que otras. Para ello tendremos que ver que condiciones se tienen que cumplir para considerar una estimación “mejor” que otra. Dicho de otra forma, el objetivo de este tema se centra en el procedimiento de trabajo, es decir, en la selección del estimador adecuado. Esta selección incluye dos estudios: ¿qué métodos disponemos para construir estimadores? y ¿qué propiedades debemos exigirle para que la estimación sea “fiable”? Una vez elegido, se considerará dicho estimador y se evaluará con la muestra concreta extraída produciendo la correspondiente estimación puntual.

Ejemplo 4.1:

Supongamos que estamos estudiando el tiempo en minutos hasta que deja de funcionar un sistema informático. La variable la denominamos X y su distribución podría venir explicada mediante una variable aleatoria exponencial cuya media es λ . Para estimar dicho parámetro extraeremos una muestra aleatoria simple (X_1, \dots, X_n) y podemos utilizar varios estimadores:

$$T_1 = \bar{X}$$

$$T_2 = \max(X_1, \dots, X_n)$$

$$T_3 = \frac{\min(X_1, \dots, X_n) + \max(X_1, \dots, X_n)}{2}$$

Al fijar la muestra, por ejemplo, $n=5$ y los valores observados en 5 paradas al azar del sistema son: (5.6, 10.2, 3.4, 13.8, 9.7). Por lo tanto, tenemos tres estimaciones puntuales que son:

$$\hat{\lambda}_1 = 8.54$$

$$\hat{\lambda}_2 = 13.8$$

$$\hat{\lambda}_3 = 8.6$$

4.2 MÉTODOS DE CONSTRUCCIÓN DE ESTIMADORES

En este punto se plantean los criterios para seleccionar un estimador de los infinitos posibles como paso previo a la comprobación de las propiedades deseables que debe de tener. Existen diferentes métodos de estimación: método de los momentos, máxima verosimilitud, mínimos cuadrados, bayesiano, bootstrap,... Los dos métodos que se presentan se basan en el conocimiento de la distribución poblacional, uno de forma indirecta mediante las características poblacionales y el otro de forma directa mediante el uso de la función de cuantía o la función de densidad, dependiendo de si la variable poblacional es discreta o continua.

4.2.1 Método de los momentos

Es un método sencillo e intuitivo propuesto por Pearson. Se basa en la idea de que las propiedades poblacionales y muestrales deben ser similares, es decir, las características poblacionales y muestrales coincidirán si la muestra es representativa.

El resumen de la información contenida en la muestra (momentos muestrales) tendrá que aproximarse a la información contenida en la población (momentos poblacionales) cuando el tamaño muestral tiende a infinito y, por lo tanto, la muestra se asemeja más a la población. El método consiste en plantear un sistema de ecuaciones, tantas como parámetros deseamos estimar, igualando los momentos muestrales con los momentos poblacionales. Éstos están expresados a través de los parámetros que son las incógnitas que debemos despejar. Resuelto el sistema de ecuaciones, cada parámetro viene expresado mediante una función de los momentos muestrales y son éstas las que reciben el nombre de **estimadores por el método de los momentos**.

Ejemplo 4.2:

Sea X la variable de interés y conocemos que su distribución es uniforme continua entre a y b . Tomamos una muestra y nuestro objetivo es encontrar los estimadores por el método de los momentos de los dos parámetros desconocidos: a y b . Tengo que plantear dos ecuaciones (porque tengo dos incógnitas) basadas en los dos momentos poblacionales, igualando éstos a los momentos muestrales.

$$\text{Media poblacional} \quad \mu = E[X] = \frac{a+b}{2} = \bar{X} \quad \text{Media muestral}$$

$$\text{Varianza poblacional} \quad \sigma^2 = V[X] = \frac{(b-a)^2}{12} = S^2 \quad \text{Varianza muestral}$$

En la primera ecuación despejo: $a = 2\bar{X} - b$ y sustituyo en la segunda ecuación:

$$(2b - 2\bar{X})^2 = 12S^2 \Rightarrow 2b - 2\bar{X} = 2S\sqrt{3} \Rightarrow b = \bar{X} + S\sqrt{3}$$

Volviendo a la primera ecuación obtenemos el estimador de a:

$$a = \bar{X} - S\sqrt{3}$$

Los estimadores de a y b por el método de los momentos de una distribución uniforme son:

$$\hat{a} = \bar{X} - S\sqrt{3}$$

$$\hat{b} = \bar{X} + S\sqrt{3}$$

4.2.2 Método de máxima verosimilitud

En primer lugar, vamos a construir y a definir la **función de verosimilitud** que es la base de este método. La función de verosimilitud es la función de probabilidad o función de densidad de la muestra aleatoria extraída, dependiendo de si la variable poblacional es discreta o continua. Si la muestra es aleatoria simple, entonces los datos de la muestra son independientes e idénticamente distribuidos según la variable poblacional X, cuya función de cuantía es $P\{X=x\}$ o cuya función de densidad es $f(x)$. La función de verosimilitud para una m.a.s. (X_1, X_2, \dots, X_n) vendrá dada por:

$$L(\theta) = \begin{cases} \prod_{i=1}^n P\{X = x_i; \theta\} & \text{si X es discreta} \\ \prod_{i=1}^n f(x_i; \theta) & \text{si X es continua} \end{cases}$$

Puesto que los datos de la muestra serán conocidos a lo largo del proceso inferencial, la única incógnita de esta función será el parámetro desconocido θ . Por lo tanto, se interpreta como una función de dicho parámetro, indicando la creencia o lo verosímil que es un valor del parámetro a la vista de los datos observados en la muestra.

Ejemplo 4.3:

Sea X ="número de clientes que entran en una hora" una variable con distribución Poisson de media λ . Se toma una muestra aleatoria de 5 horas seleccionadas al azar y se anotan los clientes que han entrado en esas cinco horas: (15, 9, 12, 14, 8). La función de verosimilitud para esta muestra sería:

$$L(\lambda) = \prod_{i=1}^5 P\{X = X_i\} = e^{-5\lambda} \frac{\lambda^{15}}{15!} \frac{\lambda^9}{9!} \frac{\lambda^{12}}{12!} \frac{\lambda^{14}}{14!} \frac{\lambda^8}{8!} = e^{-5\lambda} \frac{\lambda^{58}}{15!9!12!14!8!}$$

Ahora podemos medir lo verosímil que pueden parecer diferentes valores del parámetro (el valor 11,6 corresponde a la media muestral observada):

λ	8	10	11,6	12	14
$L(\lambda)$	$1,2732 \cdot 10^{-7}$	$2,4140 \cdot 10^{-6}$	$4,4356 \cdot 10^{-6}$	$4,2886 \cdot 10^{-6}$	$1,4869 \cdot 10^{-6}$
$L(11,6)/L(\lambda)$	34,84	1,84	1	1,03	2,98

Es más verosímil creer que el parámetro es 11,6 que 8 porque la probabilidad de observar esta muestra si la media es 11,6 es 34,84 veces mayor que si la media fuese 8. La probabilidad de observar la muestra si la media es 11,6 es 1,84 veces mayor que si el parámetro fuese 10, por lo tanto, es más verosímil el valor 11,6 que la media 10. De la misma forma, la probabilidad de observar la muestra si la media fuese 11,6 es 2,98 veces mayor que si la media fuese 14, por lo tanto, es más verosímil el valor 11,6 que el 14.

Dada la función de verosimilitud como una medida sobre la creencia o plausibilidad de un determinado valor del parámetro tomando como base de información la muestra aleatoria observada, el método de estimación se basa en pensar que entre varios posibles sucesos ocurrirá el que tenga mayor probabilidad. Esto implica que entre los posibles valores del parámetro será mejor estimación, aquella que tenga mayor verosimilitud o que sea más plausible a la vista de los datos. El método de máxima verosimilitud se basa en buscar el estimador cuya verosimilitud es mayor que la de cualquier otro, es decir, el estimador máximo verosímil (MLE) de θ será $\hat{\theta}$ si cumple que:

$$L(\hat{\theta}) = \max_{\theta} L(\theta)$$

Puesto que la función logaritmo neperiano es monótona creciente y con el objetivo de simplificar los cálculos del máximo, la expresión anterior tiene la misma solución que la siguiente:

$$\ln L(\hat{\theta}) = \max_{\theta} \ln L(\theta)$$

siendo $\hat{\theta}$ el estimador máximo verosímil (MLE) de θ .

Para buscar el máximo, basta con derivar respecto del parámetro e igualar a cero. Al despejar el parámetro quedará una función de los valores muestrales y ésta es el estimador máximo verosímil, siempre que la segunda derivada cumpla la condición de máximo, es decir, que sea negativa.

Ejemplo 4.3:

Sea X una variable dicotómica que indica la preferencia o no por un artículo de consumo (es decir, 1=está dispuesto a comprarlo y 0=no lo compraría) en una región geográfica. Su distribución es Bernoulli y tiene un parámetro desconocido p =proporción de clientes potenciales en la región. La cuestión es construir el estimador máximo verosímil de p . Para ello, se selecciona una muestra aleatoria de personas y se les consulta su preferencia o no por el producto, recogiendo los datos en (X_1, X_2, \dots, X_n) .

La función de verosimilitud es:

$$L(\lambda) = \prod_{i=1}^n P\{X = X_i\} = p^{\sum X_i} (1-p)^{n-\sum X_i}$$

Se toma el logaritmo neperiano:

$$\ln L(\lambda) = \ln \left(p^{\sum X_i} (1-p)^{n-\sum X_i} \right) = \sum X_i \ln p + (n - \sum X_i) \ln(1-p)$$

Se calcula la primera derivada respecto del parámetro p a estimar:

$$\frac{\partial \ln L(\lambda)}{\partial p} = \frac{1}{p} \sum X_i - \frac{1}{1-p} (n - \sum X_i)$$

Se iguala a cero y se despeja el parámetro p :

$$\frac{1}{p} \sum X_i = \frac{1}{1-p} (n - \sum X_i) \Rightarrow \sum X_i = np \Rightarrow p = \frac{1}{n} \sum X_i$$

El estimador máximo verosímil de p es la media muestral, es decir, la proporción muestral:

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

4.3 CRITERIO DE SELECCIÓN DE ESTIMADORES

Al ser desconocido el parámetro nunca sabremos con exactitud hasta qué punto la estimación puntual está cerca o lejos del valor del parámetro, es decir, la aplicación de la estimación puntual en vez del verdadero valor del parámetro conduce a cometer un error más o menos importante. Un criterio de selección del estimador a utilizar se debe basar en el estudio de este error que estará estrechamente ligado a la idea de precisión en la estimación.

Por otro lado, hay que recordar que una muestra aleatoria es una variable aleatoria lo que implica que el estimador también es una variable aleatoria. Esto supone que el error en la estimación, entendiéndolo como la diferencia entre el estimador y el valor del parámetro, es una variable aleatoria cuya distribución deberíamos tener en cuenta.

El criterio de selección más simple sería el ERROR CUADRÁTICO MEDIO, es decir, el promedio de las distancias cuadráticas entre las posibles estimaciones puntuales y el verdadero valor de θ . Así pues, dado un estimador $\hat{\theta}$ de θ , se define el error cuadrático medio como:

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

Esta medida de bondad del estimador se puede expresar como:

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

El problema que aparece al utilizar este criterio es que en algunas ocasiones la elección del estimador depende del posible verdadero valor del parámetro y éste es desconocido. Así pues, la aplicabilidad de este criterio disminuye considerablemente lo que implica que deberemos analizar su descomposición y formular las propiedades oportunas según ésta.

Ejemplo 4.4:

Sea una variable aleatoria X =rentabilidad mensual, en tanto por cien, de un activo que cotiza en Bolsa y suponemos que sigue una distribución normal con desviación típica $\sigma=1$. Queremos estimar su media μ (rentabilidad mensual media) y tenemos dos posibles estimadores:

$$T_1 = \bar{X}$$

$$T_2 = \frac{n\bar{X}}{n+1}$$

Para decidir qué estimador es mejor compararemos sus errores cuadráticos medios y estudiaremos cuál es menor.

$$ECM(T_1) = V[T_1] + (E[T_1] - \mu)^2 = V[T_1] = \frac{\sigma^2}{n} = \frac{1}{n}$$

$$ECM(T_2) = V[T_2] + (E[T_2] - \mu)^2 = \frac{n\sigma^2}{(n+1)^2} + \frac{\mu^2}{(n+1)^2} = \frac{n + \mu^2}{(n+1)^2}$$

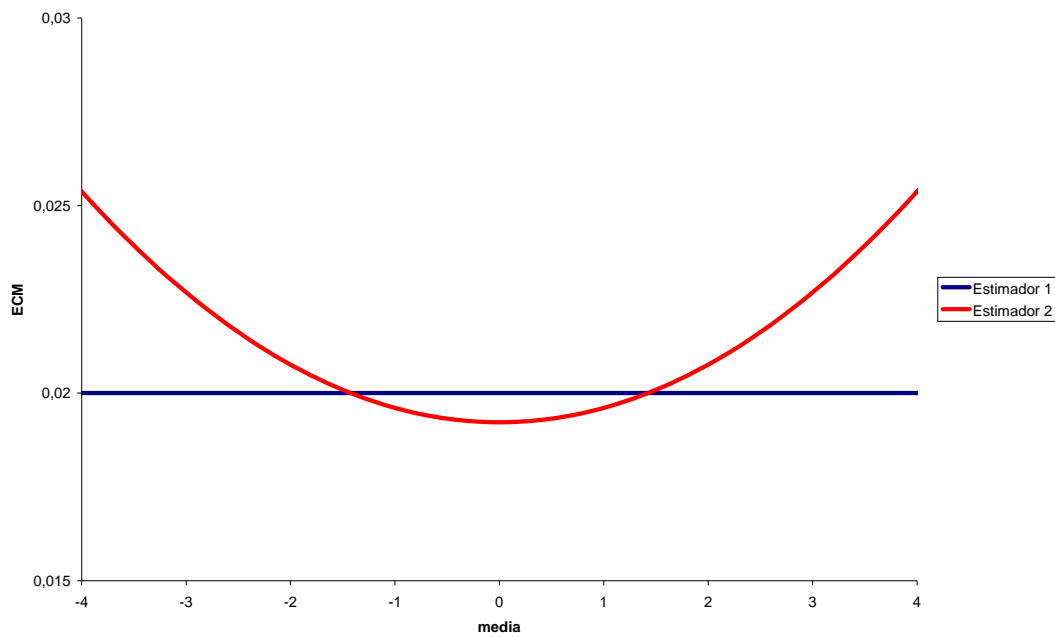
Veamos cuando es mejor el segundo estimador, es decir, cuando su error cuadrático es menor:

$$ECM(T_2) < ECM(T_1) \Rightarrow \frac{n + \mu^2}{(n+1)^2} < \frac{1}{n} \Rightarrow \mu^2 < \frac{2n+1}{n} \Rightarrow -\sqrt{\frac{2n+1}{n}} < \mu < \sqrt{\frac{2n+1}{n}}$$

Por el contrario será mejor el primer estimador:

$$ECM(T_2) > ECM(T_1) \Rightarrow \frac{n + \mu^2}{(n+1)^2} > \frac{1}{n} \Rightarrow \mu^2 > \frac{2n+1}{n} \Rightarrow \begin{cases} \mu < -\sqrt{\frac{2n+1}{n}} \\ \mu > +\sqrt{\frac{2n+1}{n}} \end{cases}$$

Este resultado indica que la elección del estimador dependerá del valor del parámetro μ (recordar que es desconocido). Por ejemplo, para una muestra de tamaño $n=50$ representamos en el gráfico siguiente los errores cuadráticos medios de ambos estimadores para posibles valores de la media entre -4% y 4%. Observamos que el segundo estimador (línea roja) es mejor cuando el posible valor de la media está entre -1,42% y 1,42% mientras que el primer estimador (línea azul) será mejor si la media toma valores mayores que 1,42% o menores que -1,42%. Por lo tanto, este criterio no es aplicable porque depende de la cantidad desconocida que es el objetivo de nuestro estudio: el parámetro. Como dicho parámetro es desconocido no podemos determinar que estimador es mejor y no podemos elegir entre ambos.



Volviendo al criterio de mínimo error cuadrático medio, vamos a analizar su descomposición:

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

El error cuadrático medio depende por un lado de la varianza (variabilidad del estimador) y, por otro, de la discrepancia entre el valor medio del estimador y el verdadero valor del parámetro. Son dos cantidades positivas, así que, para minimizar la suma de ambas tendremos que exigir que ambas cantidades sean mínimas, lo que implicará definir las propiedades que deben tener un buen estimador.

4.4 PROPIEDADES DE LOS ESTIMADORES

Dada la descomposición del error cuadrático medio, el segundo término se minimiza y, además, toma el valor cero, cuando la media del estimador coincide con el parámetro. Esto conduce a la propiedad de **insesgadez** de un estimador respecto del parámetro de interés e implica que el error cuadrático medio coincide con la varianza del estimador.

Dada la propiedad de insesgadez, tendremos que minimizar la varianza del estimador lo que lleva a definir la propiedad de **eficiencia**, como el estimador cuya

varianza es la mínima entre todos los estimadores insesgados con el mismo tamaño muestral.

Otra idea razonable es que el estimador se aproxime más al verdadero valor del parámetro cuánto mayor sea el tamaño muestral, es decir, cuánta mayor información tengamos. Sería deseable que si el tamaño muestral tiende a infinito, entonces estaríamos observando toda la población, el error de estimación desapareciese. De aquí se deduce la propiedad de **consistencia**, como la conveniencia de que el estimador se aproxime cada vez más al valor del parámetro con una probabilidad muy alta cuando el tamaño muestral tiende a infinito.

4.4.1 Inssegadez

Un estimador $\hat{\theta}$ se dice **insesgado** si su esperanza matemática coincide con el parámetro que quiere estimar, es decir, cuando se verifica que:

$$E[\hat{\theta}] = \theta$$

La diferencia entre el valor esperado del estimador y el valor del parámetro recibe el nombre de sesgo y se denota por $b(\hat{\theta}) = E[\hat{\theta}] - \theta$. Un estimador $\hat{\theta}$ se dice **sesgado** si el sesgo es distinto de cero e implica que sistemáticamente el estimador está produciendo infravaloraciones o sobrevaloraciones del parámetro según sea el signo del sesgo.

Ejemplo 4.5:

Dada una variable aleatoria X , con media m y desviación típica s , demostramos los siguientes resultados en el tema anterior:

$$E[\bar{X}] = \mu$$

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

$$E[S_1^2] = \sigma^2$$

Por lo tanto, podemos decir que la media muestral es un estimador insesgado para la media poblacional. La cuasivarianza muestral es un estimar insesgado para la varianza poblacional. La varianza muestral es un estimador sesgado (infravalora) de la varianza poblacional.

4.4.2 Eficiencia

Un estimador $\hat{\theta}$ se dice **eficiente** cuando es insesgado y posee mínima varianza respecto a cualquier otro estimador insesgado.

Para poder asegurar que la varianza es mínima tendríamos que establecer una cota inferior para cualquier estimador, de tal forma que si encontramos un estimador cuya varianza coincide con dicha cota, llamada cota de Frechet-Cramer-Rao, entonces podemos asegurar que posee la mínima varianza posible.

La acotación de Frechet-Cramer-Rao viene dada por:

Bajo ciertas condiciones de regularidad, dado un estimador insesgado $\hat{\theta}$ se verifica que:

$$V[\hat{\theta}] \geq \frac{1}{E\left[\left(\frac{\partial \ln L(\theta)}{\partial \theta}\right)^2\right]}$$

donde $L(\theta)$ es la función de verosimilitud. Al denominador de la cota de Frechet-Cramer-Rao (cota FCR) se le denomina $I(\theta)$ cantidad de información de Fisher y, en el caso de una muestra aleatoria simple, se puede expresar como el producto del tamaño muestral y la cantidad de información suministrada por un solo elemento:

$$I(\theta) = nI_1(\theta) = n \cdot E\left[\left(\frac{\partial \ln f(x_i)}{\partial \theta}\right)^2\right] = n \cdot \left(-E\left[\frac{\partial^2 \ln f(x_i)}{\partial \theta^2}\right]\right)$$

Por lo tanto, la cota FCR para un parámetro θ se expresa como $I^{-1}(\theta)$.

Esta acotación permite otra definición alternativa de eficiencia:

Un estimador $\hat{\theta}$ se dice **eficiente** cuando es insesgado y su varianza coincide con la cota de Frechet-Cramer-Rao.

Ejemplo 4.6:

Sea X una variable aleatoria con distribución Bernoulli de parámetro p y vamos a utilizar el estimador de proporción muestral. Nos preguntamos si dicho estimador, que sabemos es insesgado para la proporción poblacional, es eficiente. Para ello calculamos la cota de Frechet-Cramer-Rao y la comparamos con la varianza del estimador.

El cálculo de la cota de Frechet-Cramer-Rao, es decir, de la mínima varianza posible de un estimador insesgado de p es:

$$P\{X = x_i\} = p^{x_i} (1-p)^{1-x_i}$$
$$\ln P\{X = x_i\} = x_i \ln p + (1-x_i) \ln(1-p)$$

$$\frac{\partial nP\{X = x_i\}}{\partial p} = \frac{x_i}{p} - \frac{1 - x_i}{1 - p} = \frac{x_i - p}{p(1 - p)}$$

$$I(p) = nI_i(p) = nE\left[\frac{(x_i - p)^2}{p^2(1 - p)^2}\right] = \frac{n}{p(1 - p)}$$

$$V[p^*] \geq \frac{p(1 - p)}{n} = \text{cota FCR}(p)$$

La varianza del estimador proporción muestral es:

$$V[\hat{p}] = \frac{p(1 - p)}{n}$$

Por lo tanto, coincide con la cota de Frechet-Cramer-Rao y el estimador proporción muestral es eficiente para la proporción poblacional de una variable de tipo Bernoulli.

4.4.3 Consistencia

Un estimador $\hat{\theta}$ se dice **consistente** cuando el error de estimación está acotado superiormente con una probabilidad que tiende a 1 cuando el tamaño muestral crece a infinito. Se puede expresar de la siguiente forma:

$$\lim_{n \rightarrow \infty} P\{|\hat{\theta} - \theta| < \varepsilon\} = 1 \quad \forall \varepsilon > 0$$

Esta propiedad significa que el error que cometemos al usar el estimador es tan pequeño como sea necesario y para ello solo tenemos que utilizar el tamaño muestral conveniente, es decir, tomar el conjunto de información adecuado para controlar la posible diferencia entre el valor real y el valor estimado.

Se puede demostrar que si el estimador es insesgado y su varianza tiende a cero cuando el tamaño muestral crece a infinito entonces dicho estimador es consistente:

$$\left. \begin{array}{l} E[\hat{\theta}] = \theta \\ \lim_{n \rightarrow \infty} V[\hat{\theta}] = 0 \end{array} \right\} \Rightarrow \hat{\theta} \text{ es consistente}$$

Ejemplo 4.6:

El estimador media muestral sabemos que cumple:

$$E[\bar{X}] = \mu$$

$$V[\bar{X}] = \frac{\sigma^2}{n}$$

Por lo tanto, la media muestral es insesgado para la media poblacional y su varianza tiende a cero cuando el tamaño muestral crece (porque es una constante dividida por el tamaño muestral). En esas condiciones podemos afirmar que la media muestral es un estimador consistente para la media poblacional, es decir, que el error de estimación será tan pequeño como queramos y para ello solo necesitamos aumentar el tamaño muestral.

4.5 EVALUACIÓN DE LAS PROPIEDADES EN LOS ESTIMADORES

A continuación presentamos una evaluación de las propiedades de los estimadores conseguidos por los métodos de construcción vistos previamente.

Las propiedades que tienen los estimadores conseguidos por el método de los momentos son:

- Suelen ser estimadores consistentes.
- No son estimadores insesgados, porque la esperanza matemática no respeta todo tipo de transformaciones.
- No son estimadores eficientes, porque sólo utiliza algunas propiedades de la variable poblacional.

Las propiedades que poseen los estimadores de máxima verosimilitud son:

- Son estimadores consistentes.
- Si existe un estimador eficiente, éste será el estimador máximo verosímil.
- Cuando el tamaño muestral crece a infinito, el estimador máximo verosímil tiene una distribución aproximadamente normal cuya media es el verdadero valor del parámetro y cuya varianza coincide con la cota de Frechet-Cramer-Rao.
- La estimación máximo verosímil es invariante frente a transformaciones del parámetro: Si $\hat{\theta}$ es el MLE de θ entonces $g(\hat{\theta})$ será el MLE de $g(\theta)$, siendo $g(\cdot)$ una aplicación.