

“Introducción a la Estadística, Probabilidad e Inferencia”

Maestría en Estadística Aplicada
Facultad de Ciencias Económicas y Estadística
UNR

Introducción

La Teoría de Probabilidades permite modelar situaciones que exhiben aleatoriedad en términos de experimentos aleatorios involucrando espacios muestrales, eventos y distribuciones de probabilidad. Los axiomas de probabilidad permiten desarrollar un conjunto extenso de herramientas para calcular probabilidades para una amplia variedad de experimentos aleatorios y así estudiar el comportamiento de variables aleatorias definidas sobre ellos.

La aplicación de modelos probabilísticos a situaciones reales requiere llevar a cabo experimentos para recolectar datos con el fin de responder preguntas como:

- Cuáles son los valores de los parámetros (media, variancia, etc.) de una variable aleatoria de interés?
- Los datos observados, son consistentes con una distribución asumida?
- Los datos observados, son consistentes con un valor de un parámetro de una variable aleatoria?

Introducción

La Estadística se ocupa de la recolección y el análisis de datos y la elaboración de conclusiones a partir de los mismos. Los métodos estadísticos proveen los medios para responder preguntas como las anteriores.

Todo esto quiere decir que, dada una distribución de probabilidad, se pueden calcular probabilidades de distintos eventos. Por ejemplo, si $X \sim Bi(n = 100; p = 0,5)$, se puede calcular $P(40 \leq X \leq 60)$.

Pero en general los valores de los parámetros son desconocidos y el problema en el “opuesto”: sabiendo que $X \sim Bi(n = 100; p)$ donde el valor de p es desconocido, y habiendo observado $x = 32$, ¿qué puede decirse acerca de p ?

Se denomina inferencia estadística a cualquier procedimiento para conocer características de una población a partir de los valores observados en una muestra.

Introducción

En este contexto, denominaremos **parámetro** a cualquier resumen numérico de la población y **estadística** a cualquier resumen numérico calculado a partir de la muestra. En general los parámetros son desconocidos y el interés se centra en estimarlos.

El término **población** hace referencia al conjunto de todos los elementos o las unidades bajo estudio. Se colectan datos tomando **muestras** de la población. Las inferencias acerca de la población serán válidas si las unidades en la muestra son representativas de la población. Se requiere que las n observaciones constituyan una muestra aleatoria.

Sea X una variable aleatoria con una distribución de probabilidad que depende de un parámetro desconocido θ objeto de interés. Se selecciona una muestra aleatoria de n unidades de la población que se simboliza como $\mathbf{X} = (X_1, X_2, \dots, X_n)$, y no es más que un vector consistente de n variables aleatorias independientes con la misma distribución que X ,

Estimador y Estimación

Luego de seleccionar las n unidades se observan ciertos valores de las variables que se simbolizan con x_1, x_2, \dots, x_n . Los métodos estadísticos involucran realizar cálculos sobre los datos observados con el objetivo de concluir acerca de θ basado en $\mathbf{X} = (X_1, X_2, \dots, X_n)$

- Un **estimador** de θ es cualquier función $g(X_1, X_2, \dots, X_n)$ de la muestra que sea usada para aproximar el valor de θ .
- Una **estimación** de θ es el valor que toma $g(X_1, X_2, \dots, X_n)$ para los valores de una muestra particular x_1, x_2, \dots, x_n y habitualmente se escribe $\hat{\theta} = g(x_1, x_2, \dots, x_n)$.

Observación: Si bien es necesario hacer la distinción entre estimador y el valor que este toma, en esta unidad hablaremos del estimador pero lo simbolizaremos como $\hat{\theta}$ cuando deberíamos hacerlo como $g(X_1, X_2, \dots, X_n)$.

Distribución muestral de una estadística

Claramente, $\hat{\theta}$ en sí misma una variable aleatoria ya que varía de muestra en muestra (por ser función de X_1, X_2, \dots, X_n). Por lo tanto, las estimaciones e inferencias basadas en dicha estadística deben ser establecidas en términos de probabilidades.

La distribución muestral de una estadística $\hat{\theta}$ está dada por su distribución de probabilidad e indica que valores toman las estimaciones y con que probabilidad. Esta distribución muestral está caracterizada por sus momentos y permite calcular probabilidades que involucren a $\hat{\theta}$.

La distribución de $\hat{\theta}$ tiene asociada una esperanza y una variancia.

- La esperanza puede pensarse como el promedio de las estimaciones a través de todas las muestras posibles.
- La variancia mide la variabilidad de las estimaciones respecto a su esperanza. A mayor concentración de esta distribución alrededor de θ , mejor será el estimador.

Lo que veremos...

En esta clase definiremos propiedades deseables de los estimadores que servirán para compararlos. Consideraremos las siguientes cuestiones:

- Qué propiedades caracterizan a un buen estimador?
- Cómo se determina que un estimador es mejor que otro?
- Cómo se encuentra un buen estimador?

Veremos que la distribución muestral de una estadística permite determinar la precisión y la calidad de la misma.

Específicamente abordaremos las siguientes propiedades de los estimadores:

- Insesgamiento,
- Consistencia,
- Eficiencia,
- Suficiencia.

Insesgamiento

Sea X una variable aleatoria con una función (de densidad) de probabilidad que depende de un parámetro desconocido θ , $f(x, \theta)$. Idealmente, un buen estimador debería ser igual al parámetro, en promedio.

El estimador $\hat{\theta}$ es un estimador **insesgado** de θ si su valor esperado es igual al valor del parámetro, $E(\hat{\theta}) = \theta$.

Esta propiedad puede interpretarse en términos del promedio: si se obtuvieran repetidas muestras y para cada una de ellas se calculara la estimación, el promedio de estos valores sería igual al verdadero valor del parámetro que se está estimando.

Ejemplo: Sea $X \sim N(\mu, \sigma^2)$. Se extrae una muestra aleatoria de tamaño de n de esta población. Se considera a la *media muestral* para estimar a la media poblacional.

Insesgamiento

$$E(\bar{X}) = E\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{1}{n} E\left(\sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} n\mu = \mu$$

Por lo tanto, **la media muestral es un estimador insesgado de la media poblacional.**

Otros ejemplos de estimadores insesgados de la media poblacional son la *mediana* de las observaciones y cualquier observación, por ejemplo X_1 .

Ejemplo: Sea $X \sim N(\mu, \sigma^2)$. Se extrae una muestra aleatoria de tamaño de n de esta población. Se considera como estimador de la variancia poblacional a la estadística:

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

Insesgamiento

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

Esta estadística es un estimador sesgado de la variancia poblacional. Ahora bien, la **variancia muestral**,

$$S^2 = \frac{1}{n-1}\sum_{i=1}^n(X_i - \bar{X})^2$$

que no es más que el estimador anterior corregido por el factor $\frac{n}{n-1}$, es un estimador insesgado de la variancia poblacional σ^2 ya que:

$$E(S^2) = E\left(\frac{n}{n-1}\hat{\sigma}_M^2\right) = \frac{n}{n-1}E(\hat{\sigma}_M^2) = \sigma^2$$

Sesgo y Variancia de un Estimador

El **sesgo** de un estimador $\hat{\theta}$ para θ se define como:

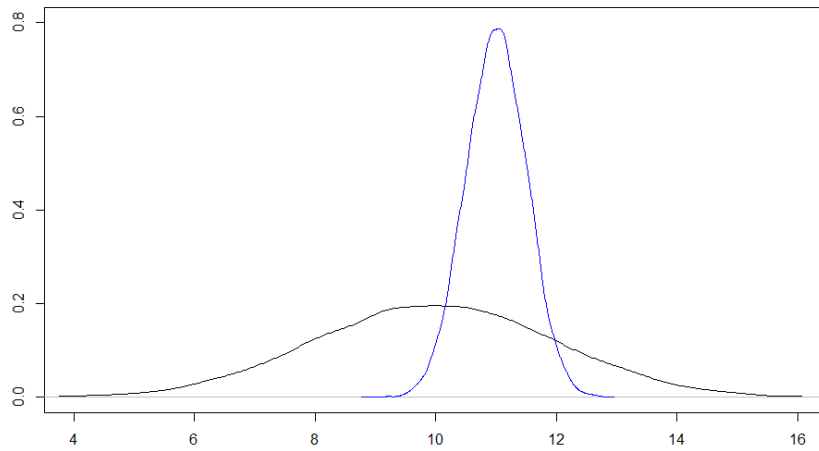
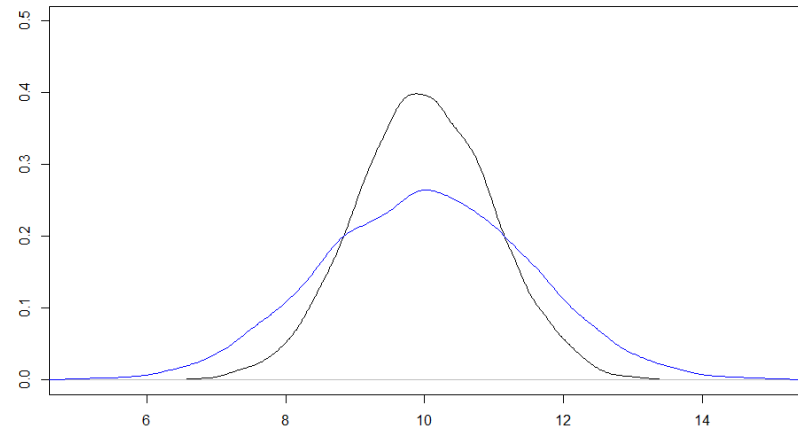
$$B = E(\hat{\theta}) - \theta.$$

En el caso de estimadores insesgados, lógicamente el sesgo es nulo. Un sesgo grande implica que en promedio, el valor de $\hat{\theta}$ caerá lejos de θ y esto no es deseable. Muchas veces la elección de estimadores se restringe al conjunto de estimadores insesgados. Pero en algunos casos, se acepta un sesgo pequeño. Además, para algunos parámetros pueden no existir estimadores insesgados.

La variancia de un estimador mide la variabilidad de las estimaciones respecto a su valor esperado. Por lo tanto, sería ideal utilizar estimadores insesgados con variancia pequeña porque así las estimaciones tenderían a aproximarse a su valor esperado, lo cual, por ser de estimadores insesgados, significa aproximarse al verdadero valor del parámetro.

Sesgo y Variancia de un Estimador

Supongamos que interesa estimar la media de una población cuyo valor es 10. Entre dos estimadores insesgados preferimos el de menor variancia (curva negra).



Aunque es deseable que un estimador sea insesgado, puede haber ocasiones en las cuales podríamos preferir estimadores sesgados (curva azul).

Error Cuadrático Medio

Una medida que considera tanto la variancia como el sesgo de un estimador es el **Error Cuadrático Medio** definido como:

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

El ECM es el valor esperado de la distancia entre el estimador y el parámetro a ser estimado. Los estimadores con ECM pequeño conducirán a estimaciones que en promedio estarán más cerca del parámetro de interés y por lo tanto serán preferidos. Esta medida puede descomponerse de la siguiente manera:

$$\begin{aligned} ECM &= E \left[(\hat{\theta} - \theta)^2 \right] = E \left[(\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta)^2 \right] \\ &= E \left[(\hat{\theta} - E(\hat{\theta}))^2 \right] + [E(\hat{\theta}) - \theta]^2 = V(\hat{\theta}) + B^2 \end{aligned}$$

El ECM resulta ser igual a la variancia del estimador más el sesgo al cuadrado.

Error Cuadrático Medio

Un buen estimador tendrá ECM pequeño porque esto implica que los valores que toma caerán cercanos a θ , es decir tendrá sesgo y variancia pequeños. Si el estimador es insesgado del parámetro, el ECM es simplemente la variancia del estimador.

Entre dos estimadores será preferible aquel con menor variancia. La comparación de estimadores insesgados con estimadores sesgados puede ser engañosa. Es posible encontrar un estimador sesgado con menor ECM que uno insesgado. En algunas situaciones, puede ser preferible el estimador sesgado.

Ejemplo: Sea X una variable aleatoria con distribución exponencial con media λ y sean los siguientes estimadores para $\theta = 1/\lambda$: $\hat{\theta}_1 = \frac{1}{n} \sum_{i=1}^n X_i$ y $\hat{\theta}_2 = n * \min(X_1, X_2, \dots, X_n)$

Error Cuadrático Medio

$\hat{\theta}_1$ es la media muestral (estimador insesgado del parámetro) y por lo tanto:

$$ECM_{\hat{\theta}_1} = E \left[\left(\hat{\theta}_1 - \frac{1}{\lambda} \right)^2 \right] = Var(\hat{\theta}_1) = \frac{n}{n^2 \lambda^2} = \frac{1}{n \lambda^2}$$

$\hat{\theta}_2$ es una variable aleatoria exponencial con media $\frac{1}{n\lambda}$:

$$E[\hat{\theta}_2] = E[n * \min(X_1, X_2, \dots, X_n)] = \frac{n}{n\lambda} = \frac{1}{\lambda}$$

Por lo tanto:

$$\begin{aligned} ECM_{\hat{\theta}_2} &= E \left[\left(\hat{\theta}_2 - \frac{1}{\lambda} \right)^2 \right] = Var(\hat{\theta}_2) = n^2 Var(\min(X_1, X_2, \dots, X_n)) = \frac{n^2}{n^2 \lambda^2} \\ &= \frac{1}{\lambda^2} \end{aligned}$$

En conclusión, la media muestral es preferible ya que tiene menor ECM.

Consistencia

Otro criterio para evaluar estimadores considera su comportamiento a medida que el tamaño de la muestra crece y es llamado propiedad de consistencia. Sea X una variable aleatoria con una distribución de probabilidad que depende de un parámetro desconocido θ , $f(x, \theta)$.

El estimador $\hat{\theta}$ para θ es un estimador **consistente** si

$$P[|\hat{\theta} - \theta| > \varepsilon] \xrightarrow{n \rightarrow \infty} 0 \quad \forall \varepsilon > 0$$

Es decir, si converge en probabilidad al valor del parámetro. Esta es una propiedad asintótica que establece que un estimador es consistente si al aumentar el tamaño de muestra la distribución de valores $\hat{\theta}$ se concentra cada vez más alrededor de θ .

Los estimadores consistentes, ya sean insesgados o sesgados, tienden al valor correcto del parámetro θ a medida que el tamaño de la muestra crece.

Consistencia

Es relativamente fácil verificar si un estimador es insesgado o no. También es fácil comparar las variancias de dos estimadores insesgados. Sin embargo, verificar la convergencia aplicando la definición anterior no es tan sencillo. El siguiente resultado puede ser útil.

Teorema: Sea $\hat{\theta}$ un estimador de θ basado en una muestra de tamaño n . Si

- $E(\hat{\theta}) \xrightarrow{n \rightarrow \infty} \theta$
- $\text{Var}(\hat{\theta}) \xrightarrow{n \rightarrow \infty} 0$

entonces $\hat{\theta}$ es un estimador consistente de θ .

El teorema anterior es una condición suficiente pero no necesaria. Es decir, que se verifiquen estas condiciones asegura que el estimador es consistente pero haber algún estimador para el cual estas condiciones no se verifiquen y aún así sea consistente.

Consistencia

Ejemplo: En el caso de la media muestral de una muestra aleatoria de tamaño n de una población $N(\mu, \sigma^2)$:

$$E(\bar{X}) = \mu$$
$$V(\bar{X}) = \frac{\sigma^2}{n}$$

Luego, la media muestral es un estimador consistente de la media poblacional.

En cambio, si utilizamos como estimador a la primera observación, este no resulta consistente pues:

$$E(X_1) = \mu$$
$$V(X_1) = \sigma^2$$

En general, la media muestral es un estimador consistente de la media poblacional para cualquier distribución.

Función de Densidad Conjunta de la Muestra

Sea X una variable aleatoria. Escribiremos $f(x, \theta)$ tanto para la función de probabilidad, $p(x) = P(X = x)$, si X es discreta, como para la función de densidad de probabilidad $f(x)$, si X es continua. La función $f(x, \theta)$ depende del parámetro desconocido θ en el cual estamos interesados.

Sea X_1, X_2, \dots, X_n una muestra aleatoria de X . Luego la función de densidad conjunta de X_1, X_2, \dots, X_n está dada por:

$$f(x_1, x_2, \dots, x_n, \theta) = f(x_1, \theta)f(x_2, \theta) \dots f(x_n, \theta) = \prod_{i=1}^n f(x_i, \theta)$$

$f(x_1, x_2, \dots, x_n, \theta)$ representa a $P[X_1 = x_1, X_2 = x_2, \dots, X_n = x_n]$, la función de probabilidad conjunta, si X es discreta, o a $f(x_1, x_2, \dots, x_n)$, la función de densidad conjunta de X_1, X_2, \dots, X_n , si X es continua.

Eficiencia

En general es deseable encontrar un estimador insesgado con la menor variancia posible ya que dicho estimador producirá las estimaciones más precisas, es decir, estimaciones cercanas al verdadero valor del parámetro de interés.

Un estimador insesgado es **eficiente** si tiene la mínima variancia posible.

De esta definición surgen la pregunta: es posible determinar cuál es la variancia mínima de los estimadores para un parámetro?

Una respuesta a esta pregunta está dada por la desigualdad de Cramér-Rao que establece una cota inferior para la variancia de los estimadores insesgados. Pero además permite asegurar que un estimador insesgado cuya variancia alcanza ese límite, entonces posee la mínima variancia y por lo tanto el mínimo error cuadrático medio.

Eficiencia

La cota en cuestión está dada por el recíproco de la medida denominada *información de Fisher* que se define como:

$$I_n(\theta) = E \left[\left\{ \frac{\partial}{\partial \theta} f(\theta, \mathbf{x}) \right\}^2 \right] = E \left[\left\{ \frac{\partial}{\partial \theta} \ln f(\theta, x_1, x_2, \dots, x_n) \right\}^2 \right]$$

Se puede probar que, para todo estimador insesgado $\hat{\theta}$ de θ , se verifica que:

$$\text{Var}(\hat{\theta}) \geq \frac{1}{I_n(\theta)}$$

$(I_n(\theta))^{-1}$ es la cota de Cramer-Rao.

Este resultado muestra que la variancia de cualquier estimador insesgado no puede ser menor que esta cota que depende de la Información de Fisher. Si encontramos un estimador insesgado cuya variancia alcanza dicho valor mínimo, será el mejor estimador (ningún estimador se comportará mejor que ese), es decir será un estimador eficiente.

Eficiencia

Observaciones:

- Una forma equivalente para la Información de Fisher más útil para cálculos es:

$$I_n(\theta) = -E \left[\frac{\partial^2}{\partial \theta^2} \ln f(\theta, \mathbf{x}) \right]$$

- En la expresión que define a la Información de Fisher, el término entre llaves es llamada función score, la cual se define como la derivada parcial del logaritmo de la función de densidad conjunta de la muestra con respecto al parámetro θ . $I_n(\theta)$ coincide con la variancia de la función score. Esta función mide la tasa a la cual cambia la función log-verosimilitud a medida que θ cambia. El análisis de esta función permite concluir que valores grandes de $I_n(\theta)$ permitirán obtener estimadores con mejor performance en términos de menores variancias que valores pequeños de $I_n(\theta)$.

Eficiencia

- La medida de Información de Fisher está definida para toda distribución de probabilidad $f(x, \theta)$ y estimador $\hat{\theta}$ tales que:
 1. para todo x tal que $f(x, \theta) > 0$, $\frac{\partial}{\partial \theta} \ln f(x, \theta)$ es finita;
 2. las operaciones de integración con respecto a x y derivación con respecto a θ pueden ser intercambiadas:

$$\frac{\partial}{\partial \theta} \int \hat{\theta} f(x, \theta) dx = \int \hat{\theta} \frac{\partial}{\partial \theta} f(x, \theta) dx$$

siempre que el miembro derecho de la ecuación sea finito.

Ejemplo: Sea $X \sim N(\mu, \sigma^2)$ siendo σ^2 conocida. Se extrae una muestra aleatoria de tamaño de n de esta población: X_1, X_2, \dots, X_n . La distribución conjunta de la muestra es:

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

Eficiencia

Aplicando logaritmo: $\ln f(\mathbf{x}, \mu) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$

Luego: $\frac{\partial l(\mathbf{x}, \mu)}{\partial \mu} = \frac{n}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = \frac{n}{\sigma^2} (\bar{x} - \mu)$

Por lo tanto,

$$I_n(\theta) = E \left[\left(\frac{\partial \ln f(\mathbf{x}, \mu)}{\partial \mu} \right)^2 \right] = E \left[\frac{n^2}{\sigma^4} (\bar{x} - \mu)^2 \right] = \frac{n^2}{\sigma^4} E[(\bar{x} - \mu)^2]$$

Dado que $E[(\bar{x} - \mu)^2] = \text{Var}(\bar{X}) = \frac{\sigma^2}{n}$, la información de Fisher resulta

$I_n(\theta) = \frac{n}{\sigma^2}$. Finalmente, la cota de Cramer-Rao es $(I_n(\theta))^{-1} = \frac{\sigma^2}{n}$.

Dado que $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$, la media muestral es un estimador eficiente de la media poblacional μ .

Eficiencia

Observaciones:

- La *eficiencia* de un estimador insesgado se define como el cociente entre su variancia y la mínima variancia posible (Cota de Cramer-Rao).
- La *eficiencia relativa* de un estimador $\hat{\theta}_1$ con respecto a otro estimador $\hat{\theta}_2$ ambos insesgados para θ , se define como el cociente:

$$E_{\hat{\theta}_1/\hat{\theta}_2} = \frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1)}$$

Si $E_{\hat{\theta}_1/\hat{\theta}_2} > 1$ entonces $V(\hat{\theta}_2) > V(\hat{\theta}_1)$ y por lo tanto $\hat{\theta}_1$ es mejor que $\hat{\theta}_2$.

En el estudio de las propiedades de los estimadores es común evaluar la *eficiencia relativa asintótica*.

Eficiencia

- Sea $g(\theta)$ una función continua y derivable con derivada no nula. Una cota inferior para estimadores insesgados de dicha función es $\left(\frac{\partial g(\theta)}{\partial \theta}\right)^2 \frac{1}{I_\theta}$
- En el caso de k parámetros de interés, $\theta_1, \theta_2, \dots, \theta_k$, sean $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k$, k estimadores insesgados de dichos parámetros. La información de Fisher es una matriz de orden $k \times k$ cuyos elementos son

$$I_n(\boldsymbol{\theta}) = E \left[\frac{\partial \ln f(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_i} \frac{\partial \ln f(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_j} \right]$$

Sea I^{ij} el elemento (i, j) de la inversa de la matriz de información. Luego resulta que

$$V(\hat{\theta}_i) \geq I^{ii}$$

Suficiencia

Dado que el proceso de estimación puntual consiste en resumir los n valores observados en un único valor $\hat{\theta}$, es importante determinar si en este proceso se pierde información. Por ejemplo, si $\hat{\theta} = X_1$, resulta evidente que no se utiliza toda la información de la muestra.

En muchos casos un estimador puede proporcionar toda la información que contiene la muestra acerca del parámetro θ . Entonces preferimos trabajar con dicho estimador en lugar de hacerlo con los n valores observados por la sencilla razón de que es más fácil manejar una sola variable aleatoria que n . Debemos determinar qué se entiende al decir que un estimador contiene toda la información acerca de un parámetro incluida en los n valores muestrales.

Sea X una variable aleatoria con una distribución de probabilidad que depende de un parámetro desconocido θ , $f(x, \theta)$.

Suficiencia

Un estimador $\hat{\theta}$ es suficiente para θ si la distribución condicional de X_1, X_2, \dots, X_n dado $\hat{\theta}$ es independiente θ .

De esta definición surge el siguiente resultado: si la densidad conjunta de la muestra puede descomponerse en el producto de dos factores de la siguiente forma:

$$f(x_1, x_2, \dots, x_n, \theta) = h(\hat{\theta}, \theta)k(x_1, x_2, \dots, x_n)$$

donde $k(x_1, x_2, \dots, x_n)$ no depende del parámetro θ , entonces $\hat{\theta}$ es un estadístico suficiente para θ .

Ejemplo: Sea $X \sim N(\mu, 1)$. Se extrae una muestra aleatoria de tamaño de n de esta población: x_1, x_2, \dots, x_n .

Suficiencia

La distribución conjunta de la muestra es:

$$f(x_1, x_2, \dots, x_n, \mu) = \prod_{i=1}^n \frac{1}{(2\pi)^{\left(\frac{1}{2}\right)}} e^{-\frac{1}{2}(x_i - \mu)^2} = \frac{1}{(2\pi)^{\left(\frac{n}{2}\right)}} e^{-\frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2}$$

Pero:

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

Entonces:

$$f(x_1, x_2, \dots, x_n, \mu) = \frac{1}{(2\pi)^{\left(\frac{n}{2}\right)}} e^{-\frac{1}{2}(n\mu^2 - 2\mu \sum_{i=1}^n x_i)} e^{-\frac{1}{2} \sum_{i=1}^n x_i^2}$$

Suficiencia

Luego,

$$h(\hat{\theta}, \mu) = e^{-\frac{1}{2}(n\mu^2 - 2\mu \sum_{i=1}^n x_i)}$$

y

$$k(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi)^{\left(\frac{n}{2}\right)}} e^{-\frac{1}{2} \sum_{i=1}^n x_i^2}$$

Por lo tanto, $\sum_{i=1}^n x_i$ es una estadística suficiente para μ .

Suficiencia

Una propiedad importante dice que si X es una variable aleatoria cuya función de densidad es $f(x, \theta)$, X_1, X_2, \dots, X_n , una muestra aleatoria de dicha variable, $t = u(X_1, X_2, \dots, X_n)$ una estadística suficiente para θ y $s = v(t)$ una función de t con inversa única, entonces s es también una estadística suficiente de θ .

Luego, en el ejemplo anterior, dado que para $X \sim N(\mu, 1)$, la estadística $\sum_{i=1}^n x_i$ es suficiente para μ , podemos asegurar que $\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$ es también una estadística suficiente para μ .

A lo largo de esta unidad vimos que el estimador \bar{X} es insesgado, consistente, eficiente y suficiente para μ .

Familias Exponenciales

Se dice que una variable aleatoria X pertenece a la familia exponencial de k parámetros si su distribución puede ser escrita como:

$$f(x, \theta) = \exp \left(\sum_{j=1}^k A_j(\theta) B_j(x) + C(x) + D(\theta) \right)$$

Donde $A_j(\theta)$ $j = 1, \dots, k$ y $D(\theta)$ son funciones sólo de los parámetros θ y $B_j(x)$ $j = 1, \dots, k$ y $C(x)$ son funciones sólo de x .

Puede mostrarse que varias de las distribuciones que ya se han visto pertenecen a esta familia. Por ej: Binomial, Poisson, Exponencial, Gamma, Normal.

Un lema demuestra que existe un vector de k estadísticas suficientes para un vector de parámetros si y sólo si la distribución de X pertenece a la familia exponencial.

Familias Exponenciales

En ese caso, resulta:

$$\begin{aligned} f(\mathbf{x}, \boldsymbol{\theta}) &= \prod_{i=1}^n f(\mathbf{x}_i, \boldsymbol{\theta}) = \prod_{i=1}^n \exp \left(\sum_{j=1}^k A_j(\boldsymbol{\theta}) B_j(x_i) + C(x_i) + D(\boldsymbol{\theta}) \right) \\ &= \exp \left(\sum_{j=1}^k A_j(\boldsymbol{\theta}) \sum_{i=1}^n B_j(x_i) + \sum_{i=1}^n C(x_i) + nD(\boldsymbol{\theta}) \right) \end{aligned}$$

Y en esta expresión,

$$\mathbf{T} = \left(\sum_{i=1}^n B_1(x_i), \sum_{i=1}^n B_2(x_i), \dots, \sum_{i=1}^n B_k(x_i) \right)$$

es el vector de estadísticas suficientes para $\boldsymbol{\theta}$.