Estadística Inferencial

Clifford Torres

Facultad de Ciencias (UNI)

June 3, 2019

Eficiencia relativa

Definición

Dados dos estimadores insesgados $\hat{\theta}_1$ y $\hat{\theta}_2$ de un parámetro θ, con varianzas $V(\hat{\theta}_1)$ y $V(\hat{\theta}_2)$, respectivamente, entonces la eficiencia de $\hat{\theta}_1$ con respecto a $\hat{\theta}_2$, denotada $eff(\hat{\theta}_1, \hat{\theta}_2)$, se define como la razón

$$eff(\hat{\theta}_1, \hat{\theta}_2) = \frac{V(\hat{\theta}_1)}{V(\hat{\theta}_2)}$$

Consistencia

Definición

Se dice que el estimador $\hat{\theta}_n$ es un estimador consistente de θ si, para cualquier número positivo ϵ,

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| \leq \epsilon) = 1$$

o bien, de forma equivalente,

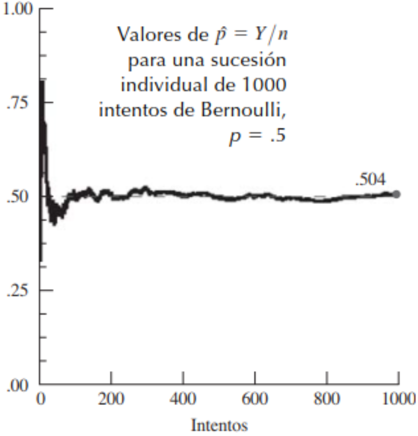
$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| > \epsilon) = 0.$$

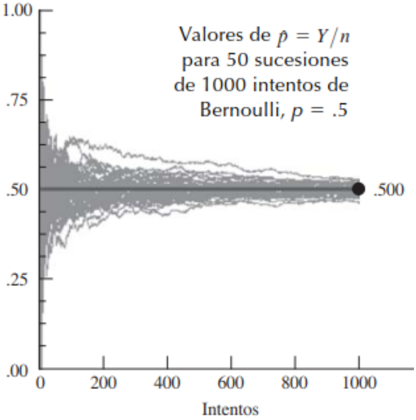
Teorema

Un estimador insesgado $\hat{\theta}_n$ para θ es un estimador consistente de θ si

$$\lim_{n\to\infty} V(\hat{\theta}_n) = 0.$$

Consistencia





Suficiencia

- Hasta aquí hemos seleccionado estimadores con base en la intuición (elegimos $\bar{Y}$ y S^2).
- Hemos visto que en ocasiones es más conveniente usar estimadores insesgados.
- Observe que hemos empleado la información en una muestra de tamaño n para calcular el valor de dos estadísticos que funcionan como estimadores para los parámetros de interés.
- En esta etapa los valores muestrales reales ya no son importantes; más bien, resumimos la información de la muestra que se relaciona con los parámetros de interés al usar los estadísticos $\bar{Y}$ y S^2.

Suficiencia

- Este proceso de resumir o reducir los datos a los dos estadísticos $\bar{Y}$ y S^2 conserva toda la información acerca de μ y σ^2 en el conjunto original de n observaciones muestrales? O bien, se ha perdido u ocultado alguna información acerca de estos parámetros en el proceso de reducir los datos?
- En esta sección presentamos métodos para hallar estadísticos que en cierto sentido resumen **toda** la información de una muestra acerca de un parámetro objetivo.
- Se dice que estos estadísticos tienen la propiedad de **suficiencia** o, dicho en una forma más sencilla, reciben el nombre de **estadísticos suficientes**.

Suficiencia

Definición

Sea $Y_1, Y_2, \cdots, Y_n$ una muestra aleatoria de una distribución de probabilidad con parámetro desconocido θ. Entonces se dice que el estadístico $U = g(Y_1, Y_2, \cdots, Y_n)$ es suficiente para θ si la distribución condicional de $Y_1, Y_2, \cdots, Y_n$, dada U, no depende de θ.

Definición

Sean $y_1, y_2, \cdots, y_n$ observaciones muestrales tomadas de variables aleatorias correspondientes $Y_1, Y_2, \cdots, Y_n$ cuya distribución depende de un parámetro θ.
Entonces, si $Y_1, Y_2, \cdots, Y_n$ son variables aleatorias discretas, la verosimilitud de la muestra, $L(y_1, y_2, \cdots, y_n|\theta)$, se define como la probabilidad conjunta de $y_1, y_2, \cdots, y_n$.
Si $Y_1, Y_2, \cdots, Y_n$ son variables aleatorias continuas, la verosimilitud $L(y_1, y_2, \cdots, y_n|\theta)$ se define como la densidad conjunta evaluada en $y_1, y_2, \cdots, y_n$.

Suficiencia

Teorema

Sea U un estadístico basado en la muestra aleatoria $Y_1, Y_2, \cdots, Y_n$. Entonces U es un estadístico suficiente para la estimación de un parámetro θ si y sólo si la verosimilitud $L(y_1, y_2, \cdots, y_n|\theta)$ se puede factorizar en dos funciones no negativas,

$$L(y_1, y_2, \cdots, y_n|\theta) = g(u, \theta) \times h(y_1, y_2, \cdots, y_n)$$

donde $g(u, \theta)$ es una función sólo de u y θ y $h(y_1, y_2, \cdots, y_n)$ no es una función de θ.

Suficiencia

El Teorema se puede usar para demostrar que hay muchos posibles estadísticos suficientes para cualquier parámetro poblacional.

1. De acuerdo con la Definición o el criterio de factorización, la muestra aleatoria por sí misma es un estadístico suficiente.
2. si $Y_1, Y_2, \cdots, Y_n$ denotan una muestra aleatoria de una distribución con una función de densidad con parámetro θ, entonces el conjunto de estadísticos de orden $Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(n)}$, que es una función de $Y_1, Y_2, \cdots, Y_n$, es suficiente para θ.

Conceptos previos: Teorema de RaoBlackwell

Definición

Si Y_1 y Y_2 son dos variables aleatorias cualesquiera, el valor esperados condicional de $g(Y_1)$, dado que $Y_2 = y_2$, se define que es

$$E[g(Y_1)|Y_2 = y_2] = \int_{-\infty}^{\infty} g(y_1) f(y_1|y_2) dy$$

si Y_1 y Y_2 son continuas conjuntamente y

$$E[g(Y_1)|Y_2 = y_2] = \sum_{y_1} g(y_1) p(y_1|y_2) dy$$

si Y_1 y Y_2 son discretas conjuntamente.

Conceptos previos: Teorema de RaoBlackwell

Teorema

Si Y_1 *y* Y_2 *son dos variables aleatorias, entonces*

$$E[Y_1] = E[E[Y_1|Y_2]],$$

donde en el lado derecho de la ecuación el valor esperado interior es con respecto a la distribución condicional de Y_1 *dada* Y_2 *y el valor esperado exterior es con respecto a la distribución de* Y_2*.*

Teorema

Si Y_1 *y* Y_2 *son dos variables aleatorias, entonces*

$$V[Y_1] = E[V[Y_1|Y_2]] + V[E[Y_1|Y_2]].$$

Teorema de RaoBlackwell

- Los estadísticos suficientes desempeñan un importante papel para determinar buenos estimadores para parámetros.
- Si $\hat{\theta}$ es un estimador insesgado para θ y si U es un estadístico suficiente para θ, entonces hay una función de U que también es un estimador insesgado para θ y tiene una varianza no mayor que $\hat{\theta}$.
- Si buscamos estimadores insesgados con varianzas pequeñas, podemos restringir nuestra búsqueda a estimadores que sean funciones de estadísticos suficientes

Teorema

Sea $\hat{\theta}$ un estimador insesgado para θ tal que $V(\hat{\theta}) < \infty$. Si U es un estadístico suficiente para θ, definamos $\hat{\theta}^* = E[\hat{\theta}|U]$. Entonces, para toda θ,

$$E[\hat{\theta}^*] = \theta \qquad y \qquad V[\hat{\theta}^*] \leq V[\hat{\theta}].$$

Teorema de RaoBlackwell

- El Teorema implica que un estimador insesgado para θ con una varianza pequeña es, o puede hacerse que sea, una función de un estadístico suficiente.
- Si tenemos un estimador insesgado para θ, podríamos mejorarlo con el uso del resultado del Teorema.
- Inicialmente puede parecer que si se aplica una vez el teorema de RaoBlackwell para obtener un mejor estimador insesgado y luego se aplica nuevamente al nuevo estimador resultante se obtiene un estimador insesgado aún mejor (NO).
- Esto es, si usamos el mismo estadístico suficiente en aplicaciones sucesivas del teorema de RaoBlackwell, no ganamos nada después de la primera aplicación.
- La única forma en que aplicaciones sucesivas pueden llevar a mejores estimadores insesgados es si usamos un estadístico suficiente distinto cada vez que se reaplica el teorema.

Teorema de RaoBlackwell

- Debido a que numerosos estadísticos son suficientes para un parámetro u asociado con una distribución... qué estadístico suficiente debemos usar cuando apliquemos este teorema?
- Para las distribuciones que estudiamos, el criterio de factorización de manera típica identifica un estadístico U que mejor resume la información de los datos acerca del parámetro θ. Tales estadísticos reciben el nombre de **estadísticos suficientes mínimos**

Estimación insesgada de varianza mínima

- En los casos que consideramos, estos estadísticos poseen otra propiedad (completabilidad) que garantiza que, si aplicamos el Teorema usando U, no sólo obtenemos un estimador con una varianza más pequeña sino que también obtenemos en realidad un estimador insesgado para θ con varianza mínima.
- Este estimador recibe el nombre de estimador insesgado de varianza mínima (MVUE, por sus siglas en inglés).
- Por tanto,
  1. Si empezamos con estimador insesgado para un parámetro θ,
  2. el estadístico suficiente obtenido por medio del criterio de factorización,
  3. la aplicación del teorema de RaoBlackwell en general lleva a un MVUE para el parámetro.

Método de momentos

- Estudiaremos uno de los métodos más antiguos para obtener estimadores puntuales: el método de momentos.
- Es un procedimiento muy sencillo para hallar un estimador para uno o más parámetros poblacionales.
- El k-ésimo momento de una variable aleatoria Y, tomado alrededor del origen, es
$$\mu_k = E[Y^k]$$
donde el correspondiente k-ésimo momento muestral es el promedio
$$m_k = \frac{1}{n}\sum_{i=1}^{n} Y_i^k$$
- El método de momentos está basado en la idea de que los momentos muestrales deben dar buenas estimaciones de los momentos poblacionales correspondientes.

Método de momentos

Es decir, m_k debe ser un buen estimador de μ_k , para $k = 1, 2, \cdots$. Entonces, debido a que los momentos poblacionales $\mu_1, \mu_2, \cdots, \mu_k$ son funciones de los parámetros poblacionales, podemos igualar los correspondientes momentos poblacionales y muestrales y despejar los estimadores deseados. En consecuencia, el método de momentos se puede expresar como sigue:

Método de momentos. Escoja como estimaciones los valores de los parámetros que son soluciones de las ecuaciones $m_k = \mu_k$, para $k = 1, 2, \cdots, t$, donde t es el número de parámetros por estimar.

Método de momentos

- Para resumir, el método de momentos permite generar estimadores de parámetros desco- nocidos al igualar los correspondientes momentos muestrales y poblacionales.
- El método es fácil de emplear y proporciona estimadores consistentes, pero los estimadores obtenidos por este método en ocasiones no son funciones de estadísticos suficientes.
- En consecuencia, es frecuente que los estimadores del método de momentos no sean eficientes y en muchos casos sean sesgados.
- Las virtudes básicas de este método son su facilidad de aplicación y que a veces proporciona estimadores con propiedades razonables.

Método de máxima verosimilitud

- En la clase anterior, se presento un método para obtener un estimador insesgado de varianza mínima (MVUE) por un parámetro objetivo: usando el criterio de factorización junto con el teorema de Rao-Blackwell.
- Tal método requiere que encontremos alguna función de un estadístico suficiente mínimo que es un estimador insesgado para el parámetro objetivo. Aun cuando tenemos un método para hallar un estadístico suficiente, la determinación de la función del estadístico suficiente mínimo que proporciona un estimador insesgado puede ser en gran medida una cuestión de azar.
- Asimismo, el método de momentos es intuitivo y fácil de aplicar, pero por lo general no lleva a los mejores estimadores.

Método de máxima verosimilitud

El método de máxima verosimilitud que con frecuencia proporciona estimadores insesgados de varianza mínima (MVUE).

Motivación: Suponga que tenemos una caja que contiene tres pelotas. Sabemos que cada una de las pelotas puede ser roja o blanca, pero no sabemos el número total de cualquiera de los colores. No obstante, podemos muestrear aleatoriamente dos de las pelotas sin restitución. Si nuestra muestra aleatoria contiene dos pelotas rojas.

¿Cuál sería una buena estimación del número total de pelotas rojas en la caja?

Método de máxima verosimilitud

Este ejemplo ilustra un método para hallar un estimador que puede aplicarse a cualquier situación. La técnica, llamada método de máxima verosimilitud, selecciona como estimaciones los valores de los parámetros que maximizan la verosimilitud (la función de probabilidad conjunta o función de densidad conjunta) de la muestra observada.

Método de máxima verosimilitud. Suponga que la función de verosimilitud depende de k parámetros

$$\theta_1, \theta_2, \cdots, \theta_k.$$

Escoja como estimaciones los valores de los parámetros que maximicen la verosimilitud

$$L(y_1, y_2, \cdots, y_n | \theta_1, \theta_2, \cdots, \theta_k).$$

Método de máxima verosimilitud

- Para destacar el hecho de que la función de verosimilitud es una función de los parámetros $\theta_1, \theta_2, \cdots, \theta_k$, a veces expresamos la función de verosimilitud como $L(\theta_1, \theta_2, \cdots, \theta_k)$.
- Es común referirnos a estimadores de máxima verosimilitud como a los MLE, por sus siglas en inglés.

Método de máxima verosimilitud

- Hemos visto que los estadísticos suficientes que mejor resumen los datos tienen propiedades deseables y con frecuencia se pueden usar para determinar un estimador insesgado de varianza mínima (MVUE) para parámetros de interés.
- Si U es cualquier estadístico suficiente para la estimación de un parámetro θ, incluyendo el estadístico suficiente obtenido del uso óptimo del criterio de factorización, el MLE es siempre alguna función de U.
- Esto es, el **MLE depende de las observaciones muestrales sólo mediante el valor de un estadístico suficiente**.

Método de máxima verosimilitud

- En efecto, sólo necesitamos observar que si U es un estadístico suficiente para θ, el criterio de factorización implica que la verosimilitud puede ser factorizada como

$$L(\theta) = L(y_1, y_2, \cdots, y_n|\theta) = g(u, \theta)h(y_1, y_2, \cdots, y_n).$$

  donde $g(u, \theta)$ es una función de sólo u y θ y $h(y_1, y_2, \cdots, y_n)$ no depende de θ.
- Por tanto, se deduce que

$$\ln[L(\theta)] = \ln[g(u, \theta)] + \ln[h(y_1, y_2, \cdots, y_n)].$$

- Observe que $\ln[h(y_1, y_2, \cdots, y_n)]$ no depende de θ y por tanto maximizar $\ln[L(\theta)]$ con respecto a θ es equivalente a maximizar $\ln[g(u, \theta)]$ con respecto a θ.
- Como $\ln[g(u, \theta)]$ depende de los datos sólo mediante el valor del estadístico suficiente U, el MLE para θ es siempre alguna función de U.

Método de máxima verosimilitud

- En consecuencia, si un MLE para un parámetro se puede hallar y luego ajustar para ser insesgado, el estimador resultante es con frecuencia un MVUE del parámetro en cuestión.
- Suponga que $\hat{\theta}$ es el MLE para un parámetro θ. Sea $t(\theta)$ una función de θ que posee una inversa única [es decir, si $\beta = t(\theta)$, entonces $\theta = t^{-1}(\beta)$. Demuestre que $t(\hat{\theta})$ es el MLE de $t(\theta)$.
- Usted demostrará que si $t(\theta)$ es una función biunívoca de θ y si $\hat{\theta}$ es el MLE para θ, entonces el MLE de $t(\theta)$ está dado por

$$\widehat{t(\theta)} = t(\hat{\theta})$$

Este resultado, a veces conocido como la propiedad de *invarianza de los MLE*, también se cumple para cualquier función de un parámetro de interés (no sólo funciones biunívocas)[1].

[1]Véanse más detalles en la obra de Casella y Berger (2002).

Algunas propiedades de los estimadores de máxima verosimilitud con muestras grandes

- Los estimadores de máxima verosimilitud también tienen propiedades interesantes cuando se trabaja con muestras grandes.
- *Suponga que $t(\theta)$ es una función derivable de θ.*
- *Afirmamos por la propiedad de invarianza* que si $\hat{\theta}$ es el MLE de θ, entonces el MLE de $t(\theta)$ está dado por $t(\hat{\theta})$.
- En algunas condiciones de regularidad que se cumplen para las distribuciones que consideraremos, *$t(\hat{\theta})$ es un estimador consistente para $t(\theta)$*.
- Además, para tamaños muestrales grandes,

$$Z = \frac{t(\hat{\theta}) - t(\theta)}{\sqrt{\dfrac{\left[\dfrac{\partial t(\theta)}{\partial \theta}\right]^2}{nE\left[-\dfrac{\partial \ln f(Y|\theta)}{\partial^2 \theta}\right]}}}$$

Algunas propiedades de los estimadores de máxima verosimilitud con muestras grandes

- Z tiene aproximadamente una distribución normal estándar.
- En esta expresión, la cantidad $f(Y|\theta)$ del denominador es la función de densidad correspondiente a la distribución continua de interés, evaluada en el valor aleatorio Y.
- Si deseamos un intervalo de confianza para $t(\theta)$, podemos usar a Z como la cantidad pivote. Siguiendo el proceso que usamos para obtener intervalos de confianza, tenemos el siguiente intervalo de confianza aproximado de muestra grande $100(1-\alpha)\%$ para $t(\theta)$:

Algunas propiedades de los estimadores de máxima verosimilitud con muestras grandes

$$
\begin{aligned}
[L, U] &\equiv t(\hat{\theta}) \pm z_{\alpha/2} \sqrt{\frac{\left[\dfrac{\partial t(\theta)}{\partial \theta}\right]^2}{nE\left[-\dfrac{\partial \ln f(Y|\theta)}{\partial^2 \theta}\right]}} \\
\text{haciendo} \quad & \theta = \hat{\theta} \\
&\approx t(\hat{\theta}) \pm z_{\alpha/2} \sqrt{\frac{\left[\dfrac{\partial t(\hat{\theta})}{\partial \hat{\theta}}\right]^2}{nE\left[-\dfrac{\partial \ln f(Y|\hat{\theta})}{\partial^2 \hat{\theta}}\right]}}.
\end{aligned}
$$