Inferencia: Estimación por intervalos

Grado en Ingeniería Biomédica
Clase 2

Jorge Calero Sanz

Departamento de Teoría de la Señal y Comunicaciones
Universidad Rey Juan Carlos

20 de febrero de 2023

Propiedades de los estimadores

Propiedades de los estimadores

Propiedades de los estimadores

Propiedades de los estimadores

- En un problema de estimación, siempre partimos de la pregunta: *¿Qué parámetro queremos estimar?*

Propiedades de los estimadores

Propiedades de los estimadores

- En un problema de estimación, siempre partimos de la pregunta: *¿Qué parámetro queremos estimar?*
- Si el parámetro a estimar es θ, un estimador es un estadístico $\hat{\theta} = T(X_1, X_2, ..., X_n)$.

Propiedades de los estimadores

Propiedades de los estimadores

- En un problema de estimación, siempre partimos de la pregunta: *¿Qué parámetro queremos estimar?*
- Si el parámetro a estimar es θ, un estimador es un estadístico $\hat{\theta} = T(X_1, X_2, ..., X_n)$.
- Un estadístico es una función que depende de la muestra. Para una muestra dada, el estadístico devuelve un valor concreto.

Propiedades de los estimadores

Propiedades de los estimadores

- Para estimar un parámetro, puede haber más de un estimador.

Propiedades de los estimadores

Propiedades de los estimadores

- Para estimar un parámetro, puede haber más de un estimador.
- Por ejemplo, para la varianza σ^2, hemos visto que teníamos la varianza muestral S^2, y la cuasivarianza muestral S_*^2.

Propiedades de los estimadores

Propiedades de los estimadores

- Para estimar un parámetro, puede haber más de un estimador.
- Por ejemplo, para la varianza σ^2, hemos visto que teníamos la varianza muestral S^2, y la cuasivarianza muestral S_*^2.
- ¿Por qué elegir un estimador y no otro?.

Propiedades de los estimadores

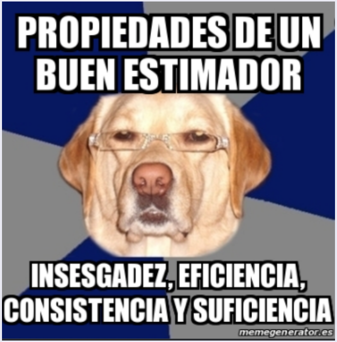
Propiedades de los estimadores

- Para estimar un parámetro, puede haber más de un estimador.
- Por ejemplo, para la varianza σ^2, hemos visto que teníamos la varianza muestral S^2, y la cuasivarianza muestral S_*^2.
- ¿Por qué elegir un estimador y no otro?.

Propiedades de los estimadores

Propiedades de los estimadores

Existen varias propiedades deseables que pueden tener los estimadores. Veamos las más importantes:



Propiedades de los estimadores

Sesgo

- Un estimador se dice que es **insesgado** si

Propiedades de los estimadores

Sesgo

- Un estimador se dice que es **insesgado** si $E[\hat{\theta}] = \theta$.
- Ejemplos: La media muestral $\bar{X}$ para la media poblacional, la cuasivarianza muestral S_*^2 para la varianza poblacional.
- Llamaremos sesgo a $b(\hat{\theta}) = E[\hat{\theta}] - \theta$.

Propiedades de los estimadores

Error cuadrático medio (ECM)

- Como hemos visto, los estimadores puntuales son variables aleatorias.
- Si son insesgadas, su esperanza es el parámetro que queremos encontrar.

Propiedades de los estimadores

Error cuadrático medio (ECM)

- Como hemos visto, los estimadores puntuales son variables aleatorias.
- Si son insesgadas, su esperanza es el parámetro que queremos encontrar.
- Sin embargo, para una muestra dada, lo normal es que cometamos un error.

Propiedades de los estimadores

Error cuadrático medio (ECM)

- Como hemos visto, los estimadores puntuales son variables aleatorias.
- Si son insesgadas, su esperanza es el parámetro que queremos encontrar.
- Sin embargo, para una muestra dada, lo normal es que cometamos un error. Para cuantificar cuánto se equivoca el estimador $\hat{\theta}$ nos referimos a su **error cuadrático medio (ECM)** y se define como

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

Propiedades de los estimadores

Error cuadrático medio (ECM)

- Como hemos visto, los estimadores puntuales son variables aleatorias.
- Si son insesgadas, su esperanza es el parámetro que queremos encontrar.
- Sin embargo, para una muestra dada, lo normal es que cometamos un error. Para cuantificar cuánto se equivoca el estimador $\hat{\theta}$ nos referimos a su **error cuadrático medio (ECM)** y se define como

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

- Si el estimador es insesgado ($b(\hat{\theta}) = 0$), entonces $ECM[\hat{\theta}] = Var(\hat{\theta})$

Propiedades de los estimadores

Eficiencia

- Si tenemos dos estimadores, diremos que uno es más **eficiente** que otro si la varianza del primero es menor que la varianza del segundo.

Propiedades de los estimadores

Eficiencia

- Si tenemos dos estimadores, diremos que uno es más **eficiente** que otro si la varianza del primero es menor que la varianza del segundo.
- Esto nos permite comparar dos estimadores que tengamos, pero diremos que un estimador es **eficiente** cuando posea mínima varianza.

Propiedades de los estimadores

Eficiencia

- Si tenemos dos estimadores, diremos que uno es más **eficiente** que otro si la varianza del primero es menor que la varianza del segundo.
- Esto nos permite comparar dos estimadores que tengamos, pero diremos que un estimador es **eficiente** cuando posea mínima varianza.
- ¿Cómo sabemos si nuestro estimador es el de mínima varianza?

Propiedades de los estimadores

Cota de Cramer-Rao

- Existe una cota inferior para la varianza de un estimador llamada **Cota de Cramer-Rao**,

Propiedades de los estimadores

Cota de Cramer-Rao

- Existe una cota inferior para la varianza de un estimador llamada **Cota de Cramer-Rao**, que es:

$$Var(\hat{\theta}) \geq \frac{(1 + b(\hat{\theta}))^2}{n\mathcal{I}(\theta)}$$

- donde $\mathcal{I}(\theta)$ es la **información de Fisher**. Intuitivamente, es una medida de cuánta información tenemos de la distribución de la que procede la muestra.

Propiedades de los estimadores

Cota de Cramer-Rao

- Si $f(x, \theta)$ es la función de densidad de la población, y suponiendo ciertas condiciones de regularidad, la información de Fisher es:

Propiedades de los estimadores

Cota de Cramer-Rao

- Si $f(x,\theta)$ es la función de densidad de la población, y suponiendo ciertas condiciones de regularidad, la información de Fisher es:

$$\mathcal{I}(\theta) = E\left[\left(\frac{\partial logf(x;\theta)}{\partial\theta}\right)^2\right]$$

- Y, con un poco de cuentas, equivalentemente:

$$\mathcal{I}(\theta) = -E\left[\frac{\partial^2 logf(x;\theta)}{\partial\theta^2}\right]$$

Propiedades de los estimadores

Cota de Cramer-Rao

- La cota de Cramer-Rao la podemos expresar del siguiente modo:

$$Var(\hat{\theta}) \geq \frac{(1+b(\hat{\theta}))^2}{-nE\left[\left(\frac{\partial logf(x,\theta)}{\partial\theta}\right)^2\right]} = \frac{(1+b(\hat{\theta}))^2}{-nE\left[\frac{\partial^2 logf(x,\theta)}{\partial\theta^2}\right]}$$

- Si la varianza de un estimador coincide con la cota de Cramer-Rao, entonces el estimador es **eficiente**.

Propiedades de los estimadores

Cota de Cramer-Rao

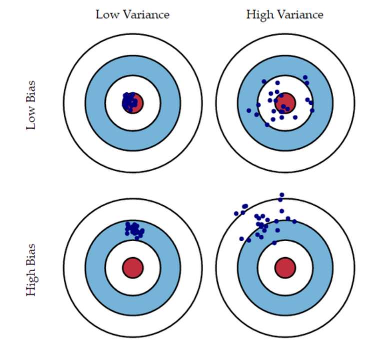
- La cota de Cramer-Rao la podemos expresar del siguiente modo:

$$Var(\hat{\theta}) \geq \frac{(1+b(\hat{\theta}))^2}{-nE\left[\left(\frac{\partial logf(x,\theta)}{\partial\theta}\right)^2\right]} = \frac{(1+b(\hat{\theta}))^2}{-nE\left[\frac{\partial^2 logf(x,\theta)}{\partial\theta^2}\right]}$$

- Si la varianza de un estimador coincide con la cota de Cramer-Rao, entonces el estimador es **eficiente**.
- Si además es insesgado (el numerador de la cota nos quedaría 1), decimos que es de **eficiencia absoluta**.

Propiedades de los estimadores

Sesgo vs varianza



Propiedades de los estimadores

Consistencia

- El sesgo y el ECM son propiedades que nos hablan de cómo es el estimador, independientemente del tamaño muestral.
- Como hemos visto para la media muestral, a medida que la muestra tiene mayor tamaño n,

Propiedades de los estimadores

Consistencia

- El sesgo y el ECM son propiedades que nos hablan de cómo es el estimador, independientemente del tamaño muestral.
- Como hemos visto para la media muestral, a medida que la muestra tiene mayor tamaño n, el estimador tiene pinta de normal cada vez más puntiaguada, o dicho bien, tiene menor varianza.
- Esta propiedad se conoce como **consistencia (a veces por robustez)** y de forma rigurosa se expresa como

$$\lim_{n} P(\mid \hat{\theta}_n - \theta \mid > \varepsilon) = 0$$

o equivalentemente:

$$\lim_{n} P(\mid \hat{\theta}_n - \theta \mid < \varepsilon) = 1$$

Propiedades de los estimadores

Consistencia

- Para probar esta propiedad, es útil conocer la **Desigualdad de Chebyshev**, que establece que para una variable aleatoria X con media μ o varianza σ^2, se tiene:

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

- Otra forma equivalente de enunciarla es que para $k > 0$:

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

Propiedades de los estimadores

Propiedades asintóticas

- La idea de ver que pasa si la muestra aumenta se puede extender a las definiciones de insesgado y eficiente.
- Un estimador es **asintóticamente insesgado** si $\lim_{n\to\infty} E[\hat{\theta}_n] = \theta$.
- Un estimador es **asintóticamente eficiente** si su varianza siempre es menor que la del resto de estimadores cuando la muestra crece $(n \to \infty)$.

Propiedades de los estimadores

Propiedades asintóticas

Encontrar estimadores

Encontrar estimadores

- Hasta ahora hemos visto principalmente que para una normal $N(\mu, \sigma^2)$, el estimador para μ es $\overline{X} = \frac{1}{n}\sum X_i$, y para la varianza σ^2 tenemos la cuasivarianza muestral $S_*^2 = \frac{1}{n-1}\sum(X_i - \overline{X})^2$.
- ¿Qué estimadores hay para otros parámetros? Por ejemplo, para la prevalencia p de una binomial, o el parámetro θ de una variable cuya función de densidad sea

$$f(x) = \frac{\theta - x}{\theta^2}, \text{ si } 0 \leq x \leq \theta$$

Encontrar estimadores

Método de los momentos

- El método de los momentos consiste en dar como estimadores de los momentos poblacionales los momentos muestrales.
- Es decir, estableceremos ecuaciones donde igualamos:
  - Momentos poblacionales $\mu_X^k = E[X^k]$.
  - Momentos muestrales $m_X^k = \frac{1}{n} \sum X^k$.
- Veámoslo con una variable aleatoria definida por la función de densidad:

$$f(x;\theta) = \frac{\theta - x}{\theta^2}, \text{ si } 0 \leq x \leq \theta$$

Encontrar estimadores

Método de los momentos

- Primero calculamos el primer momento poblacional:

$$\mu_X^1 = E[X] := \int x \cdot f(x)dx =$$

Encontrar estimadores

Método de los momentos

- Primero calculamos el primer momento poblacional:

$$\mu_X^1 = E[X] := \int x \cdot f(x)dx = \int_0^{\theta} x \cdot \frac{\theta - x}{\theta^2} dx$$

Encontrar estimadores

Método de los momentos

- Primero calculamos el primer momento poblacional:

$$\mu_X^1 = E[X] := \int x \cdot f(x)dx = \int_0^\theta x \cdot \frac{\theta - x}{\theta^2} dx$$

$$= \frac{1}{\theta^2} \int_0^\theta x \cdot \theta - x^2 dx = \frac{1}{\theta^2} \cdot \left(\frac{x^2\theta}{2} - \frac{x^3}{3} \right]_{x=0}^{x=\theta} = \frac{\theta}{2} - \frac{\theta}{3} = \frac{\theta}{6}$$

Encontrar estimadores

Método de los momentos

- Ahora igualamos $\mu_X^1 = m_X^1$, es decir, $\frac{\theta}{6} = \overline{X} = \frac{1}{n} \sum X_i$.
- Despejando el parámetro θ, obtenemos el estimador:

$$\hat{\theta} = 6 \cdot \overline{X}$$

- Este método nos proporciona una manera de encontrar estimadores, pero no necesariamente son los buenos. Tendremos que estudiar luego sus propiedades.

Como encontrar estimadores

Método de máxima verosimilitud

- El método de máxima verosimilitud se basa en el supuesto de que la muestra que hemos sacado debería de ser la muestra más probable de salirnos (la más verosímil).

Como encontrar estimadores

Método de máxima verosimilitud

- El método de máxima verosimilitud se basa en el supuesto de que la muestra que hemos sacado debería de ser la muestra más probable de salirnos (la más verosímil).
- Sea la función de densidad de la población $f(X_i, \theta)$, y tomamos una muestra aleatoria $(X_1, X_2, ..., X_n)$, podemos calcular la función de densidad conjunta de la muestra, que llamaremos **función de verosimilitud**, y se expresa como:

Como encontrar estimadores

Método de máxima verosimilitud

- El método de máxima verosimilitud se basa en el supuesto de que la muestra que hemos sacado debería de ser la muestra más probable de salirnos (la más verosímil).
- Sea la función de densidad de la población $f(X_i, \theta)$, y tomamos una muestra aleatoria $(X_1, X_2, ..., X_n)$, podemos calcular la función de densidad conjunta de la muestra, que llamaremos **función de verosimilitud**, y se expresa como:

$$L(\theta|X_1, X_2, ..., X_n) := f(X_1|\theta) \cdot f(X_2|\theta) \cdot ... f(X_n|\theta) = \prod_{i=1}^{n} f(X_i|\theta)$$

Como encontrar estimadores

Método de máxima verosimilitud

- Además, vamos a tomar el logaritmo de esa función:

Como encontrar estimadores

Método de máxima verosimilitud

- Además, vamos a tomar el logaritmo de esa función:

$$l(\theta|X_1, X_2, ..., X_n) := \ln L(\theta|X_1, X_2, ..., X_n)$$

- El estimador de máxima verosimilitud es áquel que maximiza la función de verosimilitud, es decir:

Como encontrar estimadores

Método de máxima verosimilitud

- Además, vamos a tomar el logaritmo de esa función:

$$l(\theta|X_1, X_2, ..., X_n) := \ln L(\theta|X_1, X_2, ..., X_n)$$

- El estimador de máxima verosimilitud es áquel que maximiza la función de verosimilitud, es decir:

$$EMV(\theta) = max_\theta \; l(\theta|X_1, X_2, ..., X_n)$$

Como encontrar estimadores

Método de máxima verosimilitud

- Si la muestra que hemos obtenido es la más probable de salir, es que es el máximo de la función L sobre todos los valores de θ.

Como encontrar estimadores

Método de máxima verosimilitud

- Si la muestra que hemos obtenido es la más probable de salir, es que es el máximo de la función L sobre todos los valores de θ.
- Para hallar el máximo de L, podemos derivar sobre θ, aunque en general, se suele maximizar la función $l = \ln(L)$, que es más fácil.

Como encontrar estimadores

Método de máxima verosimilitud

- Ejemplo: Sobre una muestra de 35 estudiantes de ingeniería biomédica, 12 les encanta Estadística. ¿ Cuál es la prevalencia de *que les guste Estadística* de la población de estudiantes de ingeniería biomédica?

Como encontrar estimadores

Método de máxima verosimilitud

- Ejemplo: Sobre una muestra de 35 estudiantes de ingeniería biomédica, 12 les encanta Estadística. ¿ Cuál es la prevalencia de *que les guste Estadística* de la población de estudiantes de ingeniería biomédica?
- Cada observación (alumno) es una Bernoulli con $X_i = 1$ si le gusta Estadística y lo hará con probabilidad p, que es el parámetro que quiero conocer, y $X_i = 0$ si no lo hace con $P(X_i = 0) = 1 - p$.

Como encontrar estimadores

Método de máxima verosimilitud

- Ejemplo: Sobre una muestra de 35 estudiantes de ingeniería biomédica, 12 les encanta Estadística. ¿ Cuál es la prevalencia de *que les guste Estadística* de la población de estudiantes de ingeniería biomédica?
- Cada observación (alumno) es una Bernoulli con $X_i = 1$ si le gusta Estadística y lo hará con probabilidad p, que es el parámetro que quiero conocer, y $X_i = 0$ si no lo hace con $P(X_i = 0) = 1 - p$.
- La función de verosimilitud para esta muestra será:

$$L(X_1, ..., X_{35}, p) = p^{12} \cdot (1 - p)^{23}$$

Como encontrar estimadores

Método de máxima verosimilitud

- ¿Qué valor de p hace que L alcance el máximo? Tomando logaritmos, tenemos:

$$l = \ln L = 12 \ln(p) + 23 \ln(1-p)$$

- Derivando respecto de p, e igualando a 0:

$$\frac{\partial log(L)}{\partial p} = \frac{12}{p} - \frac{23}{1-p} = 0$$

Como encontrar estimadores

Método de máxima verosimilitud

- ¿Qué valor de p hace que L alcance el máximo? Tomando logaritmos, tenemos:

$$l = \ln L = 12 \ln(p) + 23 \ln(1-p)$$

- Derivando respecto de p, e igualando a 0:

$$\frac{\partial log(L)}{\partial p} = \frac{12}{p} - \frac{23}{1-p} = 0$$

$$\frac{12}{p} = \frac{23}{1-p}$$

Como encontrar estimadores

Método de máxima verosimilitud

- ¿Qué valor de p hace que L alcance el máximo? Tomando logaritmos, tenemos:

$$l = \ln L = 12 \ln(p) + 23 \ln(1-p)$$

- Derivando respecto de p, e igualando a 0:

$$\frac{\partial log(L)}{\partial p} = \frac{12}{p} - \frac{23}{1-p} = 0$$

$$\frac{12}{p} = \frac{23}{1-p}$$

$$12(1-p) = 23p$$

$$12 = 35p$$

Como encontrar estimadores

Método de máxima verosimilitud

- Luego $\hat{p} = 12/35$ es el punto máximo.

Como encontrar estimadores

Método de máxima verosimilitud

- Luego $\hat{p} = 12/35$ es el punto máximo.
- Es decir, el estimador $\hat{p}$ de la proporción muestral, es de máxima verosimilitud para poblaciones binomiales.

Como encontrar estimadores

Método de máxima verosimilitud

El estimador de máxima verosimilitud que obtengamos tiene, al menos, las siguientes propiedades:

- Asintóticamente insesgado.
- Asintóticamente eficiente.