Estadística II

Temas 2 y 3. Estimación puntual, suficiencia e información

Facultad de Ciencias de la Economía y de la Empresa (FCEE)

Curso 2025–2026 · 4.5 ECTS · 2º cuatrimestre

Francisco Rabadán Pérez · Raquel Ibar-Alonso · Ester Muñoz Céspedes

Departamento de Economía Aplicada I e Historia e Instituciones Económicas

Tema 3 · Objetivo

Al terminar este tema podrás:

- **Comprender qué es un estimador** y distinguir **parámetro, estadístico, estimación** y **error de estimación**.
- Medir la **calidad de un estimador** con el error cuadrático medio (**ECM**) y su descomposición: **Varianza + Sesgo²**.
- Identificar y aplicar **propiedades clave: insesgadez, consistencia y eficiencia (cota de Cramér–Rao).**
- **Introducir la suficiencia** como "**resumen sin pérdida**" y usar el **criterio de factorización de Fisher–Neyman** en ejemplos sencillos.

Cómo leer estas etiquetas

CORE | PLUS | ANEXO

CORE: ideas y procedimientos que sostienen el curso entero y que se practican de forma recurrente.

PLUS: material para profundizar o reforzar intuición; ayuda a consolidar.

ANEXO: material de referencia o consulta.

Universidad Rey Juan Carlos

Facultad de Ciencias de la Economía y de la Empresa

Tema 2 - 3 · Esquema

CORE

Qué vamos a construir paso a paso:

- **Bloque A. Estimación puntual**: parámetro, estimador, estimación y error de estimación
- **Bloque B. Calidad del estimador**: ECM= Var + Sesgo^2
- **Bloque C. Propiedades clave**: insesgadez, consistencia y eficiencia
- **Bloque D. Cota de Cramér–Rao (CCR)**: interpretación y uso básico
- **Bloque E. Suficiencia e información**: idea de estadístico suficiente + Fisher–Neyman (visión operativa)
- **Cierre**: resumen + tabla de fórmulas clave

Breviario

CORE

- $\boldsymbol{\theta}$:parámetro poblacional desconocido (valor fijo).
- $\widehat{\boldsymbol{\theta}}$: estimador de θ (variable aleatoria, depende de la muestra).
- $\boldsymbol{ECM}$: Error cuadrático medio
- $\boldsymbol{L(\theta|x)}$: función de verosimilitud (misma expresión que la conjunta, vista como función de θ con x fijo).
- $\mathbf{ln}\,\boldsymbol{L(\theta|x)}$: logaritmo neperiano de la función de verosimilitud (simplifica derivadas/productos).
- $\mathbf{I_n}(\boldsymbol{\theta})$: información de Fisher.
- **CCR**: Cota de Cramér-Rao.

Sección A. Conceptos básicos de estimación puntual

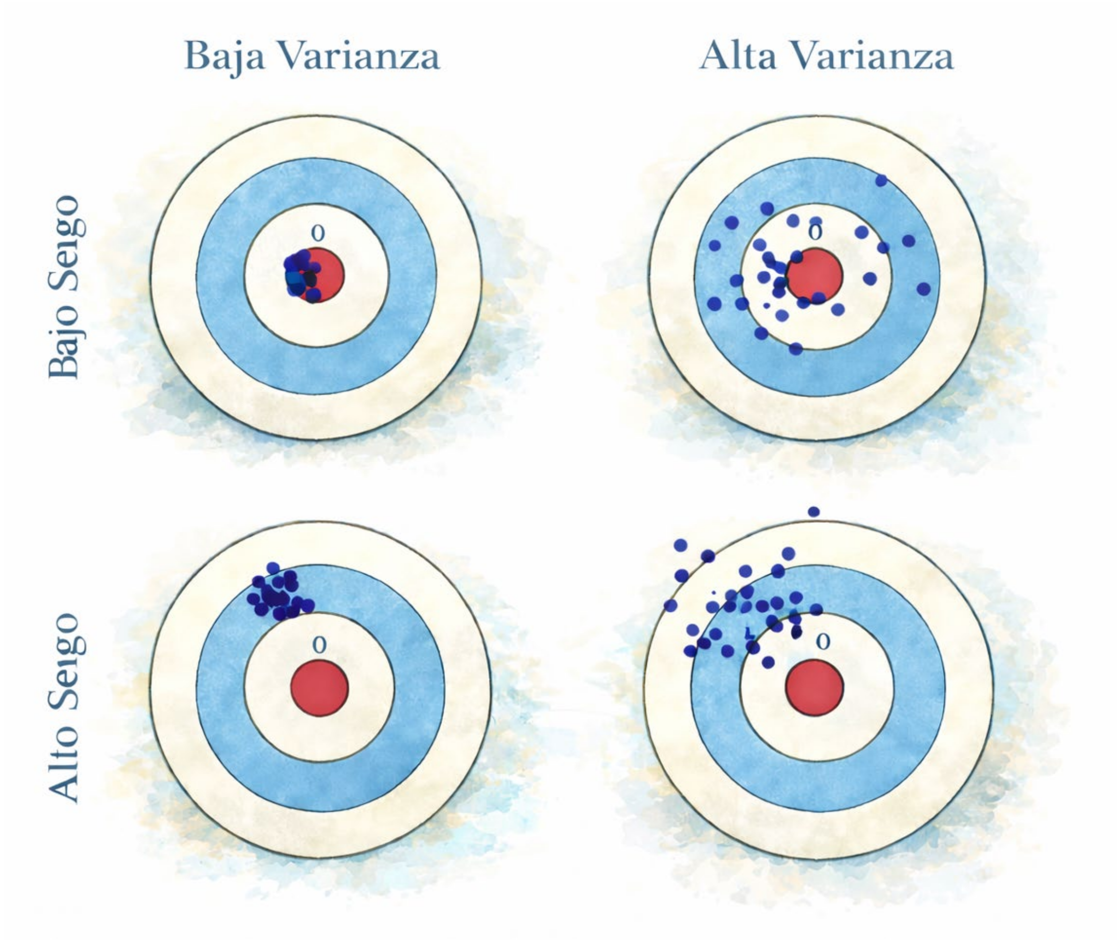
Parámetro, estimador y estimación

CORE

- **Parámetro (θ)**: valor fijo **desconocido** de la población (p. ej., **μ, p, σ^2**).
- **Estimador ($\hat{\theta}$)**: función de la muestra para aproximar θ (variable aleatoria).
- **Estimación**: el **valor numérico** que toma $\hat{\theta}$ cuando usamos los datos observados de una muestra concreta.
- Error de estimación: $\hat{\theta} - \theta$ (no se conoce porque θ es desconocido).

Error de estimación

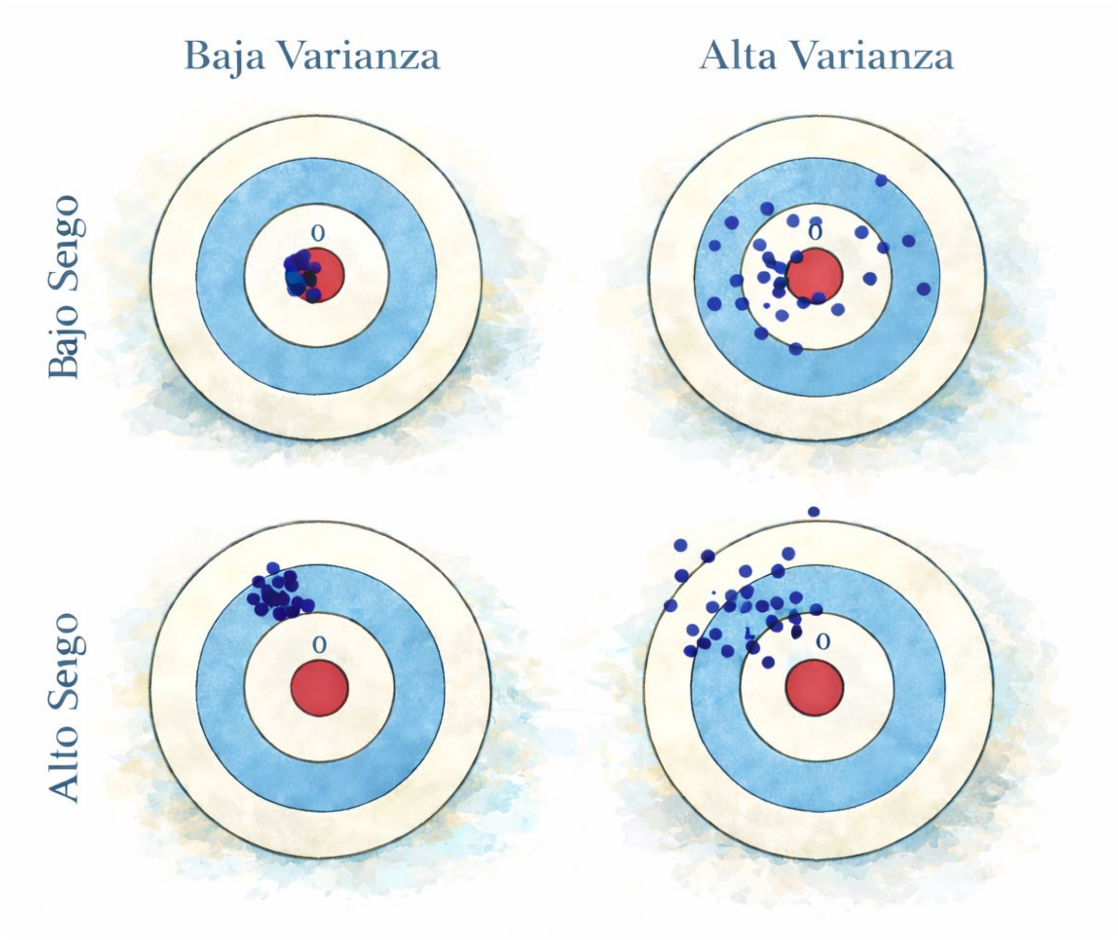
CORE



- **Error de estimación**: $\hat{\theta} - \theta$
- Mide la **diferencia** entre lo que **estimamos y el valor real** del parámetro.
- **No es observable** directamente porque θ **es desconocido**.
- Se analiza el **comportamiento de $\hat{\theta}$**: sesgo, varianza y ECM.

CORE

¿Qué es un "buen" estimador?



- Queremos que $\hat{\theta}$ esté **cerca de** θ
- "Cerca" tiene dos componentes:
  - **Sesgo (bias)**: desplazamiento **sistemático** respecto a θ
  - **Varianza**: **dispersión** de $\hat{\theta}$ entre muestras.
- Un **buen estimador** tiene **bajo sesgo** y **baja varianza.**

CORE

Sesgo del estimador

- **Sesgo de $\hat{\theta}$**:

$$Sesgo(\hat{\theta}) = E(\hat{\theta}) - \theta$$

- **Interpretación:** si repetimos el muestreo muchas veces, la diferencia entre el promedio de $\hat{\theta}$ y el valor verdadero θ.

- $\hat{\theta}$ **es Insesgado si y solo si:**

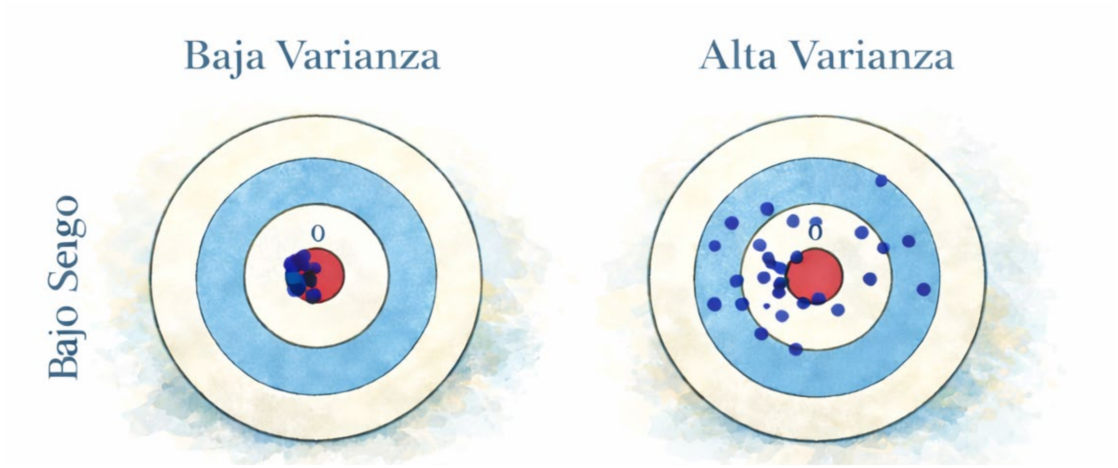
$$E(\hat{\theta}) = \theta$$

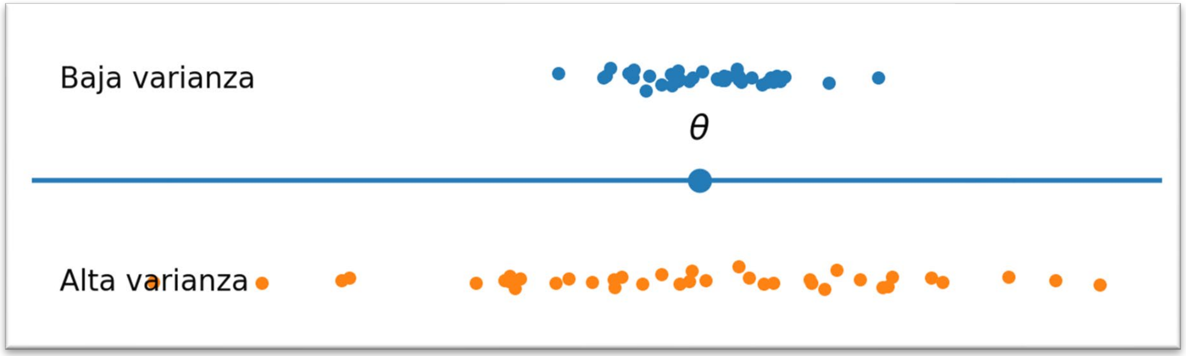
$\text{Sesgo} = E(\hat{\theta}) - \theta$

- **Ejemplos**:
- Si $E(\hat{\theta}) = \theta + 0{,}5 \rightarrow sesgo = 0{,}5$ ($\hat{\theta}$ sobreestima θ)
- Si $E(\hat{\theta}) = \theta - 0{,}5 \rightarrow sesgo = -0{,}5$ ($\hat{\theta}$ infraestima θ)

CORE

Varianza del estimador





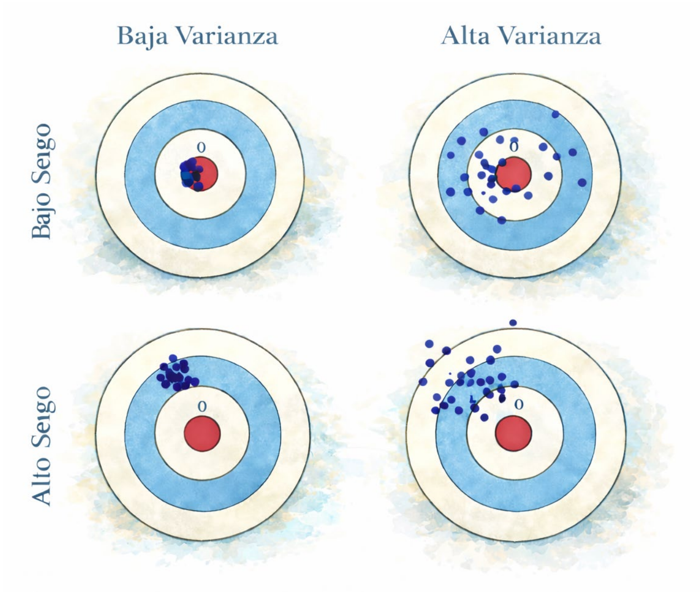
- **Varianza de $\hat{\theta}$**:

$$V(\hat{\theta})$$

- **Interpretación:** mide la dispersión de $\hat{\theta}$ cuando repetimos el muestreo.
- **Baja varianza** ⇒ estimaciones **estables** (poco cambian entre muestras).
- En general, al aumentar n, la varianza tiende a disminuir.

CORE

Ejercicios rápidos



- **Si $\mathbf{E}(\hat{\boldsymbol{\theta}}) = \boldsymbol{\theta}$, entonces el estimador $\hat{\boldsymbol{\theta}}$ es:**

A) consistente B) insesgado C) eficiente D) suficiente

- **“Alta varianza” significa que $\hat{\boldsymbol{\theta}}$ cambia mucho entre muestras**

A) Verdadero B) Falso

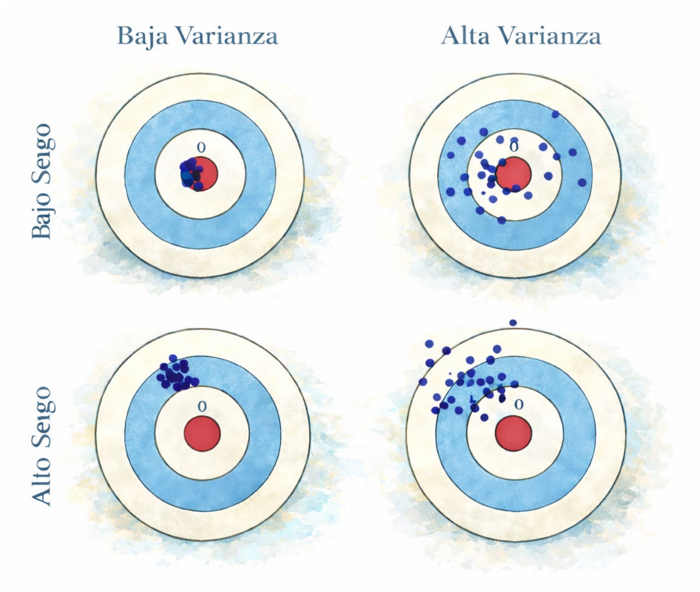
- **Si $\mathbf{E}(\hat{\boldsymbol{\theta}}) = \boldsymbol{\theta} + 0{,}5$, el sesgo es:**

A) −0,5 B) 0 C) 0,5 D) depende de n

- ¿Qué representa mejor el sesgo?
  - A) dispersión de $\hat{\theta}$
  - B) desplazamiento medio de $\hat{\theta}$ respecto a θ
  - C) el error $\hat{\theta} - \theta$ observado en una muestra
  - D) la varianza de la población

CORE

Ejercicios rápidos



- **Si $\mathrm{E}(\hat{\theta}) = \theta$, entonces el estimador es:**

A) consistente B) insesgado ✅ C) eficiente D) suficiente

- **"Alta varianza" significa que $\hat{\theta}$ cambia mucho entre muestras**

A) Verdadero ✅ B) Falso

- **Si $\mathrm{E}(\hat{\theta}) = \theta + 0{,}5$, el sesgo es:**

A) −0,5 B) 0 C) 0,5 ✅ D) depende de n

- ¿Qué representa mejor el sesgo?
  - A) dispersión de $\hat{\theta}$
  - B) desplazamiento medio respecto a θ ✅
  - C) el error $\hat{\theta} - \theta$ observado en una muestra
  - D) la varianza de la población

CORE

Error cuadrático medio (ECM)

- **En una muestra** no conocemos θ, pero sí podemos **evaluar la calidad media de $\hat{\theta}$** si repitiésemos el muestreo.
- Una **medida estándar** es el error cuadrático medio (ECM), o en inglés “*Mean Squared Error*” (MSE)
- Mide el error medio de estimación (al cuadrado) y se puede descomponer en varianza y sesgo².

- **Definición**

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

ECM = Varianza + Sesgo²

CORE

- El ECM combina dos fuentes de error:
  - **Varianza**: inestabilidad de $\hat{\theta}$ entre muestras.
  - **Sesgo²**: desplazamiento sistemático de $E[\hat{\theta}]$ respecto a θ.
- El mismo ECM puede resultar de combinaciones distintas (**trade-off**).

- **Definición**

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

- **Identidad**

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

$$ECM(\hat{\theta}) = V(\hat{\theta}) + \left[Sesgo(\hat{\theta})\right]^2$$

1. Sumo y resto $E(\hat{\theta})$

¿Por qué?

$$\hat{\theta}-\theta = \left(\hat{\theta}-E(\hat{\theta})\right) + \left(E(\hat{\theta})-\theta\right)$$

2. Elevo al cuadrado y tomo esperanza:

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

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

3. $E\left[\left(\hat{\theta}-E(\hat{\theta})\right)\ \left(E(\hat{\theta})-\theta\right)\right] =$

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

$$= [E(\hat{\theta})]^2 - E(\hat{\theta})\theta - [E(\hat{\theta})]^2 + E(\hat{\theta})\theta = 0$$

$$\left(E(\hat{\theta})-\theta\right)\cdot E\left[\left(\hat{\theta}-E(\hat{\theta})\right)\right] = 0$$

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

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

$$\boldsymbol{ECM(\hat{\theta}) = V(\hat{\theta}) + \left(Sesgo(\hat{\theta})\right)^2}$$

Ejercicio corto A3 · ECM = Varianza + Sesgo²

CORE

Enunciado

Sea un estimador $\hat{\theta}$ tal que:

- $\mathrm{E}\left[\hat{\theta}\right] = \theta + 0{,}2$
- $\mathrm{V}\left[\hat{\theta}\right] = 1/\mathrm{n}$

a) Calcula el sesgo y el sesgo².

b) Calcula el ECM como función de n.

c) ¿Para qué valores de n se cumple $ECM < 0{,}1$?

Respuesta

Dado $E\left[\hat{\theta}\right] = \theta + 0{,}2$ y $V\left[\hat{\theta}\right] = 1/n$

a) $\mathrm{Sesgo}(\hat{\theta}) = \mathrm{E}(\hat{\theta}) - \theta = 0{,}2\,;\ \left(\mathrm{Sesgo}(\hat{\theta})\right)^2 = 0{,}04$

b) $\mathrm{ECM}(\hat{\theta}) = \mathrm{V}(\hat{\theta}) + \left(\mathrm{Sesgo}(\hat{\theta})\right)^2 = \frac{1}{\mathrm{n}} + 0{,}04$

c) $\frac{1}{\mathrm{n}} + 0{,}04 < 0{,}1 \rightarrow \frac{1}{\mathrm{n}} < 0{,}06 \rightarrow \mathrm{n} > \frac{1}{0{,}06} = 16{,}66\ldots$

Por tanto, $\mathrm{n} \geq 17$

CORE

Propiedades deseables de un estimador

Un estimador $\hat{\theta}$ se suele evaluar por:

- **Insesgadez:** $\mathrm{E}(\hat{\theta}) = \theta$
- **consistencia:** $\hat{\theta} \xrightarrow{p} \theta$ cuando $n \to \infty$
- **Eficiencia**: varianza pequeña (entre insesgados)
- **Suficiencia**: $\mathrm{T}(X)$ recoge toda la información sobre θ
- **Invarianza** : si cambias la escala del parámetro, la estimación se obtiene aplicando el mismo cambio al estimador $\left(g(\theta) \to g(\hat{\theta})\right)$
- **Robustez**: poca sensibilidad a pequeños cambios (valores atípicos / supuestos / *outliers*)

Sección B. Insesgadez y consistencia

CORE

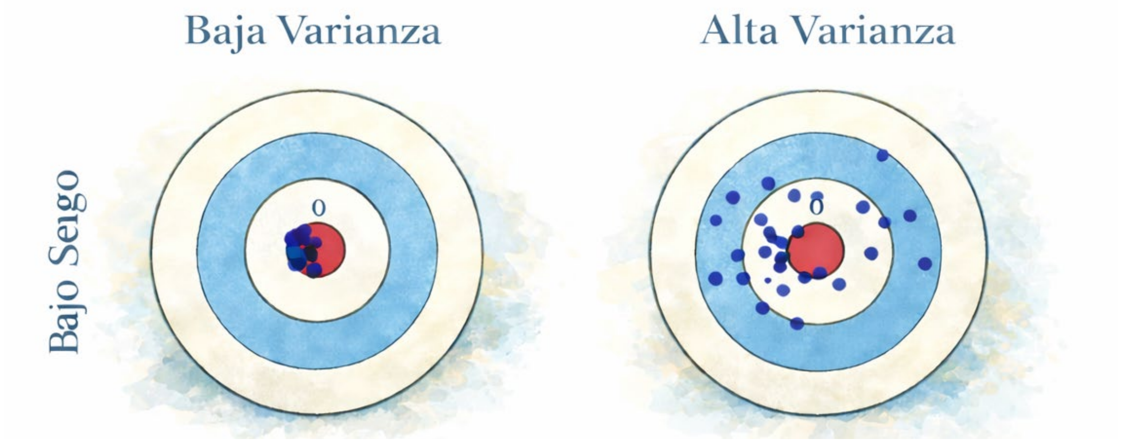
Insesgadez ≠ “siempre cerca”

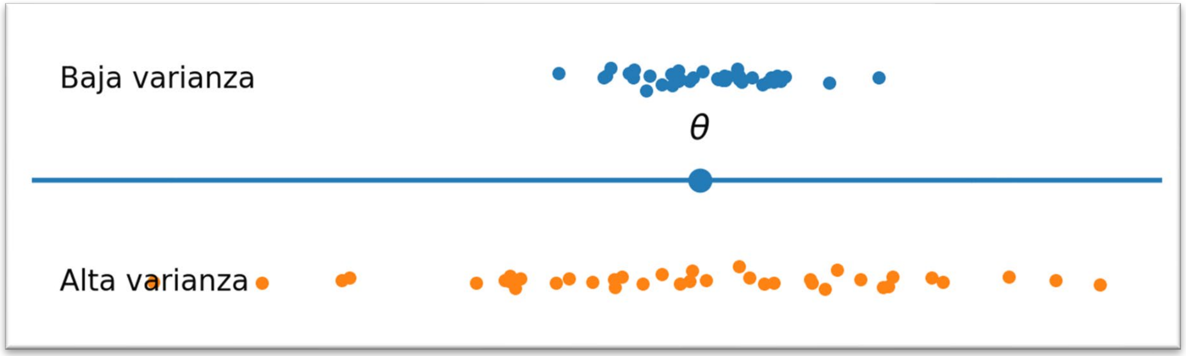
- **Insesgado significa:**

$$\mathrm{E}(\hat{\theta}) = \theta$$

- Pero una estimación concreta puede estar lejos si hay alta varianza.

- **Idea clave: insesgadez habla de promedio, no de precisión.**





Insesgadez: dos matices útiles

CORE

- **Insesgado:**

$$\mathrm{E}(\hat{\theta}) = \theta$$

- En ese caso, el error cuadrático medio **mide solo la variabilidad entre muestras**.

$$\hat{\theta}\ insesgado \rightarrow \mathrm{ECM}(\hat{\theta}) = \mathrm{V}(\hat{\theta})$$

- **Insesgadez asintótica:**

$$\mathrm{E}(\hat{\theta}) - \theta \rightarrow 0 \quad (n \rightarrow \infty)$$

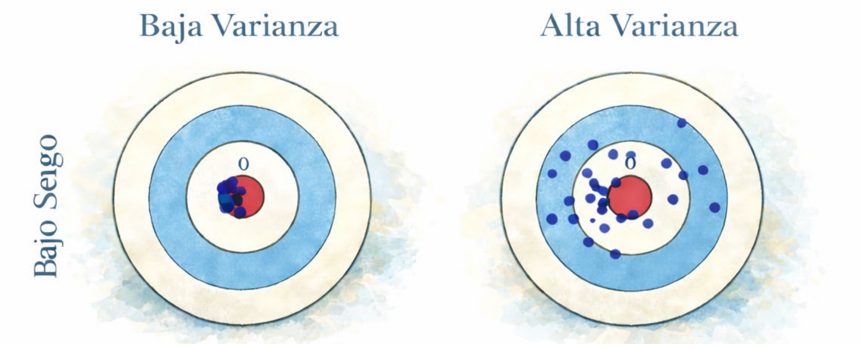
- **Ejemplo clásico (varianza muestral):**

$$\mathrm{E}(S^2) = \frac{n-1}{n}\sigma^2 \Rightarrow$$

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

$$Sesgo(S^2) = -\frac{\sigma^2}{n} \xrightarrow[n\to\infty]{} 0$$

Conclusión: S^2 es **sesgado**, pero **asintóticamente insesgado**.



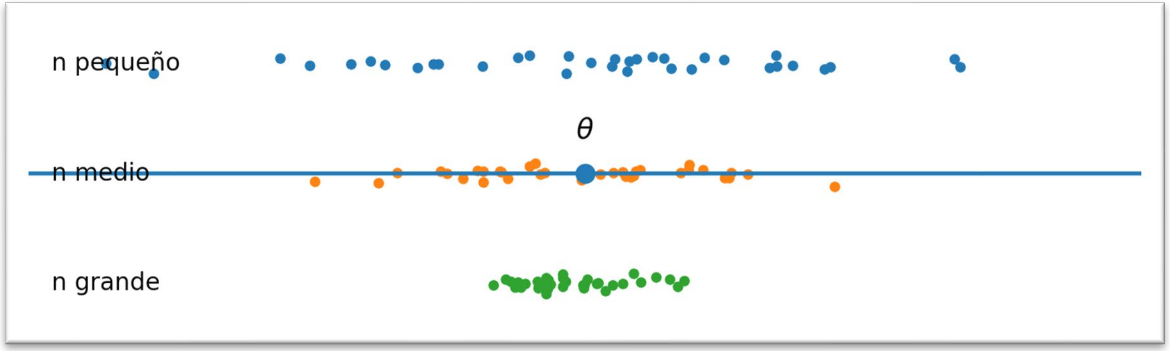
El estadístico insesgado correspondiente: S_1^2

CORE

Consistencia (idea y definición)

Intuición

- Un estimador $\hat{\theta}$ es consistente si, cuando aumenta el tamaño muestral n, $\hat{\theta}$ se aproxima a θ.
- **Idea clave:** con muestras grandes, el error $\hat{\theta} - \theta$ tiende a hacerse pequeño.



Definición (convergencia en probabilidad):

$$\hat{\theta} \xrightarrow{p} \theta \quad (n \to \infty)$$

- Equivalente operacional: para todo $\varepsilon > 0$,

$$P\left(\left|\hat{\theta} - \theta\right| > \varepsilon\right) \to 0$$

- **Decisión**: para cualquier $\varepsilon > 0$, miras qué ocurre con $P\left(\left|\hat{\theta} - \theta\right| > \varepsilon\right)$ cuando $n \to \infty$; Si tiende a 0 para todo ε, entonces $\hat{\theta}$ es consistente.

- **Lectura:** la probabilidad de "fallar por más de ε" se hace casi cero cuando n crece.

¿Cómo comprobar consistencia? (regla práctica)

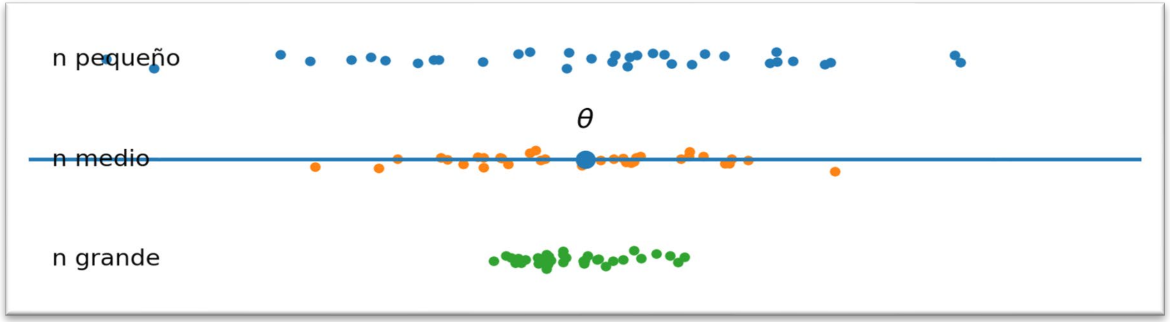
CORE

Regla útil (suficiente): **si se cumple que, cuando $n \to \infty$,**

- $\mathrm{E}(\hat{\theta}) \to \theta$; $\left(Sesgo(\hat{\theta}) \to 0\right)$
- $\mathrm{V}(\hat{\theta}) \to 0$

Entonces $\hat{\theta}$ es consistente.

- **Idea clave:** la distribución de $\hat{\theta}$ se concentra alrededor de θ.



Ejemplo de 1 minuto

Sea $\mathrm{E}(\hat{\theta}) = \theta + 1/n$; $\mathrm{V}(\hat{\theta}) = 4/n$

Si $n \to \infty$,

- $\mathrm{E}(\hat{\theta}) \to \theta$
- $\mathrm{V}(\hat{\theta}) \to 0$

Y, por tanto, $\hat{\theta}$ **es consistente**

- **Idea informal**: Si la dispersión se hace pequeña $(Var \to 0)$ y el centro se acerca a θ $\left(Sesgo(\hat{\theta}) \to 0\right)$, entonces $\hat{\theta}$ termina "pegándose" a θ.

Ejercicios rápidos B1 · Consistencia

CORE

1. Si $\hat{\theta} \xrightarrow{p} \theta$, entonces:

a) $\mathrm{E}(\hat{\theta}) = 0$ para todo n.

b) $\mathrm{P}(|\hat{\theta} - \theta| > \varepsilon) \to 0$ para todo $\varepsilon > 0$.

c) $\mathrm{V}(\hat{\theta})$ es constante

d) $\hat{\theta} = 0$ siempre

2. Si $\mathbf{E}[\hat{\theta}] = \theta$ + 1/n y $\mathbf{V}[\hat{\theta}] = \mathbf{1/n}$, entonces $\hat{\theta}$ es:

a) sesgado y no consistente

b) insesgado y consistente

c) sesgado y consistente

d) insesgado y no consistente

3. ¿Cuál de estas condiciones es suficiente para consistencia?

a) $\mathrm{E}(\hat{\theta}) = 0$

b) $\mathrm{V}(\hat{\theta}) \to 0$

c) $\mathrm{E}(\hat{\theta}) \to \theta$ y $\mathrm{V}(\hat{\theta}) \to 0$

d) $\mathrm{V}(\hat{\theta})$ mínima.

4. En términos intuitivos, consistencia significa que al crecer n :

a) El sesgo aumenta

b) Las estimaciones se dispersan más

c) El estimador se acerca a θ con “alta probabilidad”

d) El estimador es suficiente

Ejercicios rápidos B1 · Consistencia

CORE

1. Si $\hat{\theta} \xrightarrow{p} \theta$, entonces:

a) $E(\hat{\theta}) = 0$ para todo n.

b) $P(|\hat{\theta} - \theta| > \varepsilon) \to 0$ para todo $\varepsilon > 0$. ✅

c) $V(\hat{\theta})$ es constante

d) $\hat{\theta} = 0$ siempre

2. Si $E[\hat{\theta}] = \theta + 1/n$ y $V[\hat{\theta}] = 1/n$, entonces $\hat{\theta}$ es:

a) sesgado y no consistente

b) insesgado y consistente

c) sesgado y consistente ✅

d) insesgado y no consistente

3. ¿Cuál de estas condiciones es suficiente para consistencia?

a) $E(\hat{\theta}) = 0$

b) $V(\hat{\theta}) \to 0$

c) $E(\hat{\theta}) \to \theta$ y $V(\hat{\theta}) \to 0$ ✅

d) $V(\hat{\theta})$ mínima.

4. En términos intuitivos, consistencia significa que al crecer n :

a) El sesgo aumenta

b) Las estimaciones se dispersan más

c) El estimador se acerca a θ con “alta probabilidad” ✅

d) El estimador es suficiente

Respuestas

Insesgadez y consistencia: no son lo mismo

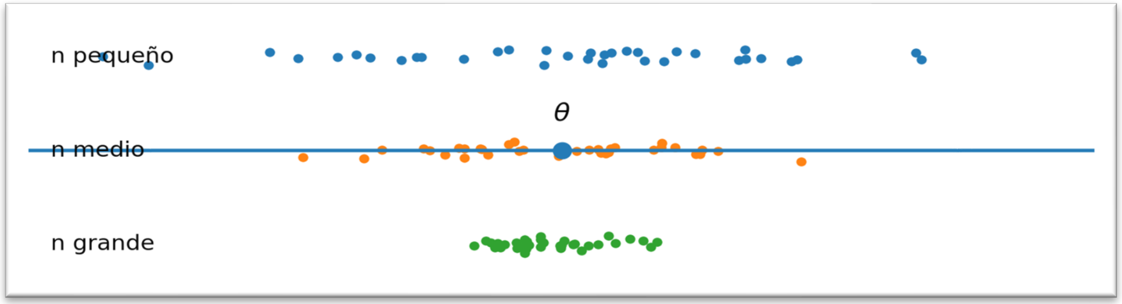
CORE

Insesgadez (para cada n)

- $\mathrm{E}(\hat{\theta}) = \theta$
- Habla del **promedio** de $\hat{\theta}$ en muestreos repetidos.
- No garantiza baja varianza.

Consistencia (cuando $n \to \infty$)

- $\hat{\theta} \xrightarrow{p} \theta$
- Habla del **comportamiento con muestras grandes**.
- Puede haber **sesgo para** n **finito**, pero $sesgo \to 0$ cuando $n \to \infty$.



Mini-test

¿Puede un estimador ser consistente y sesgado (para n finito)? Sí ✅

Ejercicio corto B2

CORE

Sea $\hat{\theta}_n$ tal que:

$$E(\hat{\theta}_n) = \theta + \frac{2}{n}, \qquad V(\hat{\theta}_n) = \frac{1}{n}$$

a) ¿Es insesgado para n finito?

b) ¿Es consistente? Justifica con condiciones suficientes (la regla práctica).

c) Calcula el ECM y su límite cuando $n \to \infty$.

Solución

a) No.

$$Sesgo(\hat{\theta}_n) = E(\hat{\theta}_n) - \theta = \frac{2}{n} \neq 0$$

Solución

b) Sí, es consistente, porque:

$$E(\hat{\theta}_n) \to \theta \;\; y \; V(\hat{\theta}_n) \to 0 \;\; (n \to \infty)$$

c)

$$ECM(\hat{\theta}_n) = V(\hat{\theta}_n) + \left(Sesgo(\hat{\theta}_n)\right)^2 = \frac{1}{n} + \left(\frac{2}{n}\right)^2 = \frac{1}{n} + \frac{4}{n^2}$$

$$\lim_{n\to\infty} ECM(\hat{\theta}_n) = \lim_{n\to\infty} \left(\frac{1}{n} + \frac{4}{n^2}\right) = 0$$

Nota: al cumplirse b) y c) $\Rightarrow \hat{\theta}_n \underset{p}{\to} \theta$

Idea clave

Consistencia no exige insesgadez para n finito: basta con que el sesgo $\to 0$ y la varianza $\to 0$ cuando n aumenta.

Sección C. Eficiencia y cota de Cramér–Rao

Eficiencia del estimador $\hat{\theta}$

CORE

- **El estimador eficiente** es el de mínima varianza**:** Entre los estimadores insesgados de un mismo parámetro poblacional, el eficiente es aquel que tiene mínima varianza.
  Para saber si un estimador es el eficiente se puede comparar con la **Cota de Cramér–Rao (CCR),** lo vemos más adelante.

- **Eficiencia relativa:** Un estimador $\hat{\boldsymbol{\theta}}_{\mathbf{1}}$ es **más eficiente** que otro $\hat{\theta}_2$ si tiene **menor varianza**:

$$V\left(\widehat{\theta}_1\right) < V\left(\widehat{\theta}_2\right)$$

  La eficiencia relativa solo tiene sentido comparar estimadores insesgados. Si los estimadores comparados son sesgados, se prefiere el de menor Error Cuadrático Medio (ECM).

Eficiencia del estimador (idea operativa)

CORE

- **En sentido amplio**, diremos que $\hat{\theta}_1$ es **más eficiente** que otro $\hat{\theta}_2$ si tiene **menor ECM**:

$$\boldsymbol{ECM(\hat{\theta}) = V(\hat{\theta}) + \left(Sesgo(\hat{\theta})\right)^2}$$

  - **Si** ambos son **insesgados**, entonces comparar eficiencia equivale a **comparar Varianzas**, porque

$$ECM(\hat{\theta}) = V(\hat{\theta})$$

- **Interpretación**: entre estimadores, preferimos el que produce **estimaciones más estables** (menos dispersas).

mini-ejemplo

Dos estimadores insesgados de θ:

- $\mathrm{V}(\hat{\theta}_1) = 0{,}20$
- $\mathrm{V}(\hat{\theta}_2) = 0{,}05$

→ $\hat{\theta}_2$ **es más eficiente**

Eficiencia vs consistencia

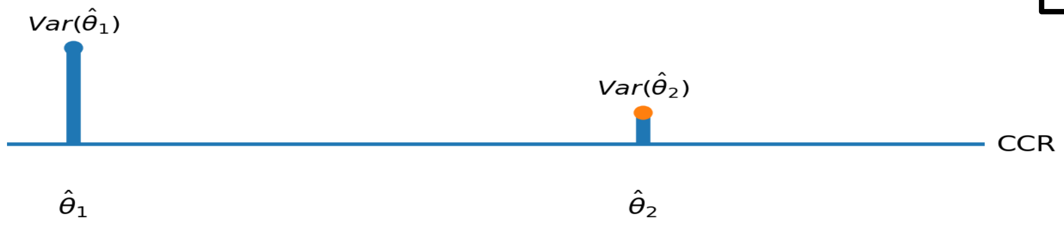
- **Consistencia**: con n grande $\hat{\theta} \to \theta$.
- **Eficiencia**: para un n dado, entre insesgados, preferimos el que tiene **menor varianza** (más precisión).

Cota de Cramér–Rao (CCR)

CORE

Intuición:

- La CCR proporciona un **límite inferior para la varianza** de cualquier estimador de θ **conocida la distribución poblacional** (derivable).
- **Interpretación**: existe una precisión "máxima" (mínima varianza) que no se puede superar dadas la distribución y el tamaño muestral (n).
- **Para todo estimador insesgado,** $V(\hat{\theta}) \geq CCR\ (\theta)$. Si alcanza la igualdad, es eficiente en sentido CCR.



- **Información de Fisher:** $I_n(\theta)$ de una m.a.s(n)

$$I_1(\theta) = E\left[\left(\frac{\partial(\ln f(X;\theta))}{\partial\theta}\right)^2\right]; \quad I_n(\theta) = nI_1(\theta)$$

La CCR se define:

- Si $\hat{\theta}$ es insesgado: $\mathrm{CCR}(\theta) = \frac{1}{\mathrm{I_n}(\theta)}$
- Si $\hat{\theta}$ es sesgado: $\mathrm{CCR}(\theta) = \frac{(1+\mathrm{b}'(\theta))^2}{\mathrm{I_n}(\theta)}$

Donde

- $\mathrm{b}(\theta) = \mathrm{E}(\hat{\theta}) - \theta$ y
- $\mathrm{b}'(\theta)$ es la derivada primera del sesgo.

Bajo condiciones regulares (supuestos técnicos para que exista $I(\theta)$ y sea válida la derivación).

CCR en una muestra i.i.d. (tamaño n)

CORE

$$\boldsymbol{Si\ V(\hat{\theta}) = CCR(\theta)}\ \textbf{entonces}\ \boldsymbol{\hat{\theta}}\ \textbf{es el eficiente}$$

$$\Longrightarrow$$

$$\not\Longleftarrow$$

Ojo: Puede existir un estimador $\hat{\theta}$ que sea el eficiente y que no cumpla la igualdad: $V(\hat{\theta}) = CCR(\theta)$

Explicación intuitiva:

- Para una misma población:

$$Si \quad \mathrm{n} \uparrow \quad \Rightarrow \quad I_n(\theta) \uparrow \quad \Rightarrow \quad CCR(\theta) \downarrow$$

- **consecuencia**: la varianza mínima posible puede ser menor para una muestra mayor.

Idea: más datos ⇒ más información ⇒ mayor precisión potencial.

$I_1(\theta)$ es información esperada por observación; el *score* varía con los datos, pero su esperanza (información) es la misma para cada x_i.

Cálculo de $I_n(\pi)$ (Bernoulli)

PLUS

idea operativa:

$$I_n(\theta) = n \cdot E\left[\left(\frac{\partial(\ln f(x;\theta))}{\partial\theta}\right)^2\right]$$

Pasos:

1. Escribe $f(x;\theta)$ (una observación)
2. Calcula $\ln f(x;\theta)$
3. Deriva: $\frac{\partial(\ln f(x;\theta))}{\partial\theta}$
4. Eleva al cuadrado y toma esperanza.
5. Multiplica por n.

Si $x \sim B(1,\pi)$; estimador p:

1. $f(x;\pi) = \pi^x(1-\pi)^{1-x}, x \in \{0,1\}$
2. $\ln f(x;\pi) = x\ln\pi + (1-x)\ln(1-\pi)$
3. $\frac{\partial(\ln f(X;\pi))}{\partial\pi} = \frac{x}{\pi} - \frac{1-x}{1-\pi} = \frac{x-\pi}{\pi(1-\pi)}$
4. $E\left[\left(\frac{\partial(\ln f(x;\pi))}{\partial\pi}\right)^2\right] = \frac{E[(x-\pi)^2]}{\pi^2(1-\pi)^2} = \frac{V(x)}{\pi^2(1-\pi)^2} = \frac{1}{\pi(1-\pi)}$
5. $I_n(\pi) = n \cdot \frac{1}{\pi(1-\pi)} = \frac{n}{\pi(1-\pi)}$

- **Nota**: Si $x \sim B(1,\pi)$; entonces $V(x) = \pi(1-\pi)$;

CCR en $B(1,\pi)$: varianza mínima para estimadores insesgados

CORE

Planteamiento:

Si $x_1, \dots, x_n \sim B(1,\pi)$ v.a.i.i.d. y $\hat{\pi}$ es un estimador insesgado de π, entonces:

- Ya hemos calculado:

$$I_n(\pi) = \frac{n}{\pi(1-\pi)}$$

- Calculamos la CCR (π) para estimador insesgado:

$$CCR(\pi) = \frac{1}{I_n(\pi)} = \frac{\pi(1-\pi)}{n}$$

- Aplicando la CCR:

$$V(\hat{\pi}) \geq CCR(\pi)$$

$$V(\hat{\pi}) = \mathrm{V(p)} = \frac{\pi\,(1-\pi)}{n} = CCR(\pi)$$

Por lo tanto, $\hat{\boldsymbol{\pi}} = \boldsymbol{p}$ **es el estimador eficiente de** $\boldsymbol{\pi}$

Interpretación (2 líneas):

- Ningún estimador insesgado de π puede tener varianza menor que $\frac{\pi(1-\pi)}{n}$
- A mayor n, menor cota → mayor precisión potencial

Ejercicio corto C2 · CCR en $B(1,\pi)$

CORE

Planteamiento:

Si $x_1, \dots, x_n \sim B(1,\pi)$ v.a.i.i.d. y define el estimador:

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

1. Calcula $E(p)$. ¿Es p insesgado para π?
2. Calcula $V(p)$
3. Compara $V(p)$ con la CCR:

$$CCR(\pi) = \frac{\pi(1-\pi)}{n}$$

¿p alcanza la cota?

1. $E(p) = \pi \;\rightarrow p \; es \; insesgado.$
2. Como $V(X_i) = \sigma^2 = \pi(1-\pi) \; y \; son$ $independientes$

$$V(p) = \frac{\sigma^2}{n} = \frac{\pi(1-\pi)}{n}$$

3. $V(p) = CCR(\pi) \rightarrow p$ es el estimador eficiente (en sentido CCR) para π en $B(1,\pi)$

- **Nota**: Este resultado explica por qué **la proporción muestral es el estimador ‘natural’ de π.**

Ejercicios rápidos C1 · Cramér–Rao

CORE

1. La CCR proporciona:

a) un valor exacto de $V(\hat{\theta})$
b) un límite inferior para $V(\hat{\theta})$
c) un límite superior para el sesgo
d) un test de hipótesis

2) Si un estimador insesgado alcanza la CCR, entonces:

a) es consistente
b) es suficiente
c) es el eficiente (en sentido CCR)
d) es robusto

3) En general, al aumentar n en una muestra i.i.d., suele ocurrir que:

a) $I_n(\theta)$ disminuye
b) la CCR aumenta
c) $I_n(\theta)$ aumenta y la CCR disminuye
d) la CCR no cambia

4) La CCR se aplica a:

a) cualquier estimador, sesgado o no
b) solo a estimadores robustos
c) estimadores insesgados y bajo condiciones regulares
d) solo a distribuciones normales

Ejercicios rápidos C1 · Cramér–Rao

CORE

1. La CCR proporciona:

a) un valor exacto de $V(\hat{\theta})$

b) un límite inferior para $V(\hat{\theta})$ ✅

c) un límite superior para el sesgo

d) un test de hipótesis

2) Si un estimador insesgado alcanza la CCR, entonces:

a) es consistente

b) es suficiente

c) es el eficiente (en sentido CCR) ✅

d) es robusto

3) En general, al aumentar n en una muestra i.i.d., suele ocurrir que:

a) $I_n(\theta)$ disminuye

b) la CCR aumenta

c) $I_n(\theta)$ aumenta y la CCR disminuye ✅

d) la CCR no cambia

4) La CCR se aplica a:

a) cualquier estimador, sesgado o no

b) solo a estimadores robustos

c) estimadores insesgados y bajo condiciones regulares ✅

d) solo a distribuciones normales

Sección D.
Suficiencia (y criterio de Fisher–Neyman)

Suficiencia: “resumen sin pérdida” (idea)

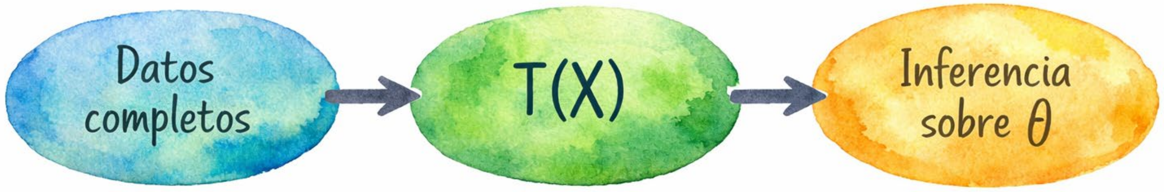
CORE

Definición intuitiva, concisa:

- Un estadístico $T(X)$ es **suficiente** para θ si **contiene toda la información que aporta la muestra** sobre θ.
- **Intuición**: una vez conoces $T(X)$, el resto de los datos **no aportan nada más** para inferir θ.
- **Objetivo**: **reducir datos** sin perder información sobre el parámetro.

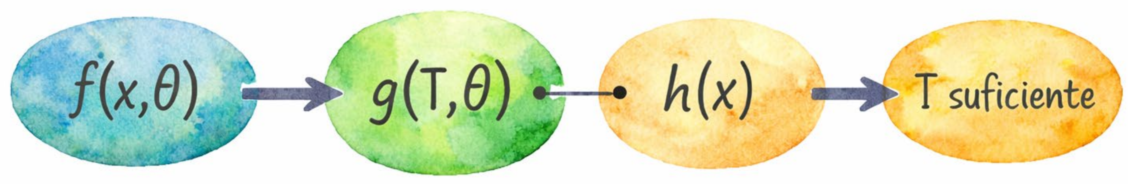
ejemplos rápidos

- Si $x_i \sim B(1, \pi)$, un resumen natural es $\sum x_i$ (número de éxitos).
- Si $x_i \sim Poisson(\lambda)$, un resumen natural es $\sum x_i$ (número total de eventos).
- **Nota**: En muchos modelos, basta con sumas/medias para estimar el parámetro.



Criterio de Fisher–Neyman

CORE



Idea intuitiva:

- **Para decidir si $T(X)$ es suficiente, miramos si la verosimilitud $L(\theta|x)$ se puede separar** en:
  - una parte que depende de los datos solo a través de $T(x)$ y de θ.
  - y otra parte que depende de los datos, pero no de θ.

Si se puede, entonces $T(X)$ es suficiente

Enunciado

$T(X)$ es suficiente para θ si la función conjunta de la muestra se puede escribir como

$$f(x_1, \dots, x_n; \theta) = g(T(x), \theta) \cdot h(x)$$

Donde $h(x)$ no depende de θ.

Cómo se usa

1. Escribe $f(x_1, \dots, x_n; \theta)$
2. Reordena para aislar $T(x)$.
3. Identifica $g(\cdot)$ y $h(\cdot)$.

Fisher–Neyman en la Normal: intuición con fórmulas

CORE

Modelo:

Sea un m.a.s. $x_i \sim N(\mu, \sigma)$ v.a.i.i.d. con σ^2 conocida

Función conjunta (verosimilitud)

$$f(x_1, \dots, x_n; \mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x_i - \mu)^2}{2\sigma^2}\right]$$

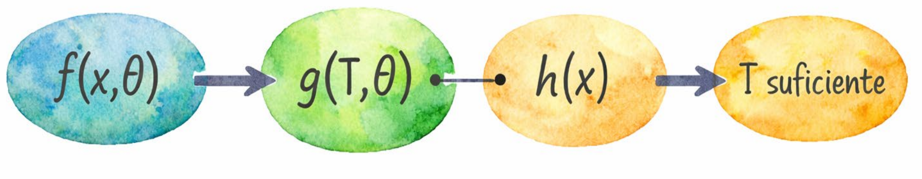
Reescribimos el exponente

$$\prod_{i=1}^{n} \exp\left[-\frac{(x_i - \mu)^2}{2\sigma^2}\right] = \exp\left[-\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{2\sigma^2}\right] = \exp\left[-\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - \mu)^2}{2\sigma^2}\right]$$

Factorización (Fisher–Neyman)

$$f(x_1, \dots, x_n; \mu) = \underbrace{exp\left(\frac{-n(\bar{x} - \mu)^2}{2\sigma^2}\right)}_{depende\ de\ \mu\ solo\ vía\ \bar{x}} \cdot \underbrace{\left[(2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum (x_i - \bar{x})^2\right)\right]}_{h(x_1, \dots, x_n)\ no\ depende\ de\ \mu}$$

Conclusión: $\bar{x}$ es estadístico suficiente para μ

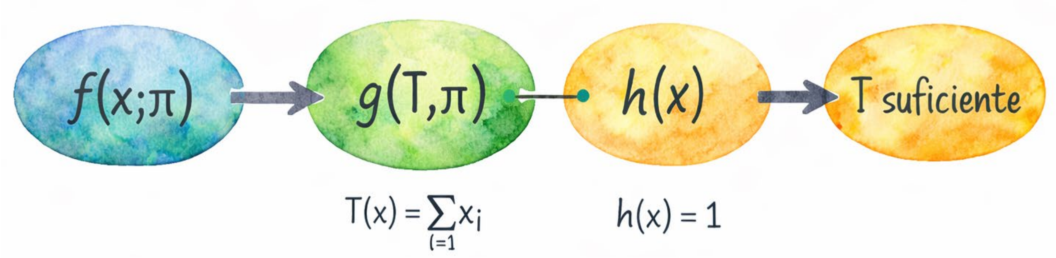


Comentario

1. El término $\sum_{i=1}^{n}(x_i - \bar{x})^2$ describe la forma interna de la muestra, pero **no cambia la verosimilitud relativa entre valores de μ.**
2. Por eso, una vez conocida $\bar{x}$, el resto de los datos es irrelevante para inferir μ.

Ejercicio corto · Suficiencia en Bernoulli $B(1,\pi)$

PLUS



Pregunta

Sea $x_1, \dots, x_n \sim B(1,\pi)$ v.a.i.i.d.

1. Escribe la función conjunta $f(x_1, \dots, x_n; \pi)$
2. Factoriza para mostrar que

$$T(x_1, \dots, x_n) = \sum_{i=1}^{n} x_i$$

Solución

1. Como $f(x;\pi) = \pi^x (1-\pi)^{1-x}$,

$$f(x_1, \dots, x_n; \pi) = \prod_{i=1}^{n} \pi^{x_i} (1-\pi)^{1-x_i} = \pi^{\sum x_i} (1-\pi)^{n-\sum x_i}$$

2. Identificamos:

$$g(T(x), \pi) = \pi^{\sum x_i} (1-\pi)^{n-\sum x_i} \ , \qquad h(x) = 1$$

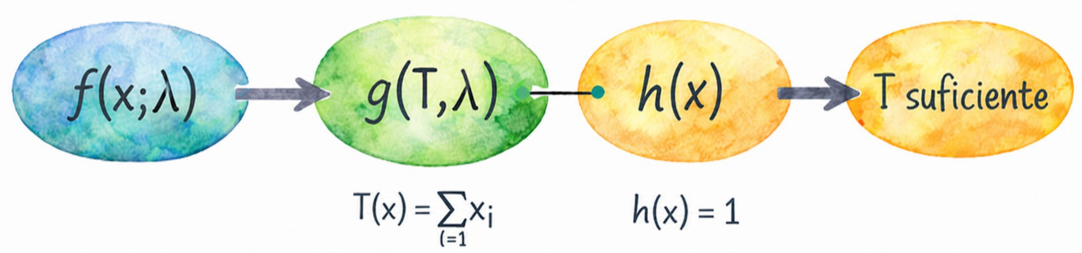
Como $h(x)$ no depende de π, por Fisher–Neyman:

$$T(X) = \sum_{i=1}^{n} x_i \ es\ suficiente\ para\ \pi.$$

Interpretación: basta saber “número de éxitos” para inferir π

Ejercicio corto · Suficiencia en Poisson(λ)

PLUS



Pregunta

Sea $x_1, \dots, x_n \sim Poisson(\lambda)$ con realización $x_1, \dots, x_n$

1. Escribe la función conjunta $f(x_1, \dots, x_n; \lambda)$
2. Factoriza para mostrar que

$$T(x_1, \dots, x_n) = \sum_{i=1}^{n} x_i$$

es suficiente para λ.

Solución

1. Como $f(x;\lambda) = e^{-\lambda}\frac{\lambda^x}{x!}$,

$$f(x_1, \dots, x_n; \lambda) = \prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}\prod_{i=1}^{n}\frac{1}{x_i!}$$

2. Identificamos:

$$f(x_1, \dots, x_n; \lambda) = g(T(x_1, \dots, x_n), \lambda) \cdot h(x_1, \dots, x_n)$$

$$g(T(x_1, \dots, x_n), \lambda) = e^{-n\lambda}\lambda^{T(x)}, \qquad h(x_1, \dots, x_n) = \prod_{i=1}^{n}\frac{1}{x_i!}$$

Como $h(x_1, \dots, x_n)$ no depende de λ, por Fisher–Neyman:

$$T(X) = \sum_{i=1}^{n} x_i \ es\ suficiente\ para\ \lambda.$$

Interpretación: basta saber “número de ocurrencias” para inferir λ.

Ejercicios rápidos · Suficiencia (test)

CORE

1. $T(X)$ es suficiente para θ si:

a) $T(X)$ es insesgado.

b) $T(X)$ tiene varianza mínima.

c) $f(X;\theta) = g(T(X),\theta)h(X)$, con $h(X)$ que no depende de θ.

d) $T(X)$ es robusto.

2. En un m.a.s(n), $\mathbf{x_i \sim B(1, \pi)}$, un estadístico suficiente para π es:

a) $T(X) = \max\{x_1, \ldots, x_n\}$

b) $T(X) = \sum_{i=1}^{n} |x_i - \bar{x}|$

c) $T(X) = \sum x_i$

d) $T(X) = \sum (x_i^2 - \bar{x})$

3. En $Poisson(\lambda)$, un estadístico suficiente para λ es

a) $\sum x_i$

b) $\max x_i$

c) $\min x_i$

d) $\sum x_i^2$

4. En el criterio de Fisher–Neyman, la parte $\mathbf{h(x)}$ debe:

a) depender de θ

b) depender solo de $T(X)$

c) No depender de θ

d) ser constante igual a 0

Ejercicios rápidos · Suficiencia (test)

CORE

1. $T(X)$ es suficiente para θ si:

a) $T(X)$ es insesgado.

b) $T(X)$ tiene varianza mínima.

c) $f(\mathrm{X};\theta)=g(T(X),\theta)h(X)$, con $h(\mathrm{X})$ que no depende de θ. ✅

d) $T(X)$ es robusto.

2. En un m.a.s(n), $\mathbf{x_i \sim B(1,\pi)}$, un estadístico suficiente para π es:

a) $\mathrm{T}(X)=\max\{x_1,\dots,x_n\}$

b) $T(X)=\sum_{i=1}^{n}|x_i-\bar{x}|$

c) $T(X)=\sum x_i$ ✅

d) $T(X)=\sum(\mathrm{x}_\mathrm{i}^2-\bar{x})$

3. En $Poisson(\lambda)$, un estadístico suficiente para λ es

a) $\sum x_\mathrm{i}$ ✅

b) $\max x_\mathrm{i}$

c) $\min \mathrm{x_i}$

d) $\sum \mathrm{x_i^2}$

4. En el criterio de Fisher–Neyman, la parte $\mathbf{h(x)}$ debe:

a) depender de θ

b) depender solo de $T(\mathrm{X})$

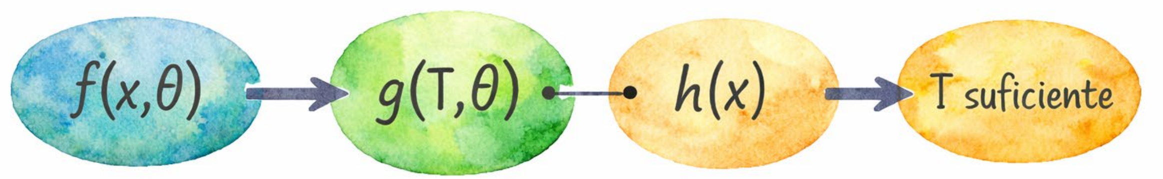
c) No depender de θ ✅

d) ser constante igual a 0

Respuestas

Suficiencia: por qué importa (cierre)

CORE



idea clave:

- **Suficiencia = resumen sin pérdida:** $T(X)$ retiene toda la información de la muestra sobre el parámetro.
- **Ventaja práctica:** reduce el problema de trabajar con n datos a trabajar con **un resumen $\boldsymbol{T(X)}$** (más simple).
- **Cómo se comprueba**: criterio de Fisher–Neyman:

$L(\theta|X) = g(T(X), \theta)h(X)$ con $h(X)$ que no depende de θ.

Ejemplos

- $x_i \sim B(1, \pi): T(X) = \sum x_i$(número de éxitos) es suficiente para π.
- $x_i \sim Poisson(\lambda): T(X) = \sum x_i$(total de eventos) es suficiente para λ.

Mini-frase final

Si $T(X)$ es suficiente, no perdemos información sobre el parámetro al reemplazar la muestra por $T(X)$.

Sección E. Invarianza y robustez

Invarianza y robustez (dos propiedades prácticas)

CORE

Invarianza

Idea: si tenemos un parámetro poblacional (θ) y construimos uno nuevo (η) transformándolo con una función g:

$$\eta = g(\theta)$$

Entonces, **si el estimador $\hat{\theta}$ es invariante**, el estimador del nuevo parámetro ($\hat{\eta}$) es la transformación g aplicada al estimador del parámetro inicial :

$$\hat{\eta} = g(\hat{\theta})$$

Ejemplo rápido:

Si $\eta = 2\mu$ y $\hat{\mu} = \bar{x}$, entonces $\hat{\eta} = 2\,\bar{x}$

Robustez

- **Idea:** un estimador es robusto si **cambia poco ante valores atípicos** o pequeñas desviaciones del modelo.
- **Ejemplo rápido:** la **media es sensible** a *outliers*; la **mediana suele ser más robusta.**
- **Mini-test:** ¿Qué es más robusto ante un outlier extremo? Mediana ✅

Cierre de tema

Tema 3 · Propiedades y fórmulas mínimas

CORE

Ideas clave

- Sesgo y Varianza explican el comportamiento de un estimador.
- ECM($\hat{\theta}$) = V ($\hat{\theta}$) + [Sesgo ($\hat{\theta}$)]²

Propiedades

- **Insesgadez** (finita): $E\left(\hat{\theta}\right) = \theta$
- **Insesgadez asintótica**: $E\left(\hat{\theta}\right) - \theta \to 0$
- **Consistencia**: $\hat{\theta}_n \underset{p}{\to} \theta$
- **Eficiencia**: la menor $V\left(\hat{\theta}\right)$
- **Eficiencia relativa**: $\min\limits_{i=1,\ldots m}(V\left(\hat{\theta}_i\right))$
- **Suficiencia** (idea): resumen $T(X)$ sin pérdida de información relevante.
- **Invarianza**: $g(\theta) \to g\left(\hat{\theta}\right)$
- **Robustez**: poca sensibilidad a *outliers*/supuestos.

Formulas mínimas

- $Sesgo\left(\hat{\theta}\right) = E\left(\hat{\theta}\right) - \theta$
- $ECM\left(\hat{\theta}\right) = V\left(\hat{\theta}\right) + \left(Sesgo\left(\hat{\theta}\right)\right)^2$
- $\hat{\theta}_n \underset{p}{\to} \theta$
- CCR
  - $V\left(\hat{\theta}\right) \geq \frac{1}{I_n(\theta)}$; $V\left(\hat{\theta}\right) \geq \frac{(1+b'(\theta))^2}{I_n(\theta)}$
  - $I_n(\theta) = nI_1(\theta)$
  - $I_1(\theta) = E\left[\left(\frac{\partial(\ln f(X;\theta))}{\partial\theta}\right)^2\right]$

Tema 3 · Propiedades de estadísticos básicos

CORE

Estadístico (parámetro)	Insesgadez	Consistencia	Eficiencia (CCR)	Suficiencia (idea)
$\bar{x}$ (para μ)	✅ $E(\bar{x}) = \mu$	✅ $V(\bar{x}) = \frac{\sigma^2}{n} \to 0 \Rightarrow \bar{x} \xrightarrow[p]{} \mu$;	✅ Normal, σ^2 conocida	✅ Normal (σ^2 conocida): $T = \sum X_i$
S^2 (para σ^2)	❌ $E(S^2) = \frac{n-1}{n}\sigma^2$ ✅ $E(S^2) \xrightarrow[n\to\infty]{} \sigma^2$	✅ (con v.a.i.i.d.) $S^2 \xrightarrow[p]{} \sigma^2$	❌ Normal, μ desconocida: no alcanza CCR	✅ Normal: $T = \left(\sum x_i, \sum x_i^2\right)$
S_1^2 (para σ^2)	✅ $\mathrm{E}(\mathrm{S}_1^2) = \sigma^2$	✅ (con i.i.d.) $S_1^2 \xrightarrow[p]{} \sigma^2$	❌ Normal, μ desconocida: no alcanza CCR	✅ Normal: $T = \left(\sum x_i, \sum x_i^2\right)$
$p = \bar{x}$ con $x_i \sim B(1, \pi)$ (para π)	✅ $\mathrm{E(p)} = \pi$	✅ $V(p) = \frac{\pi(1-\pi)}{n}$	✅ Bernoulli: $V(p) = \frac{1}{I_n(\pi)} = CCR(\pi)$	✅ Bernoulli: $T = \sum x_i$

Invarianza: estos estimadores son invariantes en el sentido habitual si $\eta = g(\theta)$, un estimador natural es $\hat{\eta} = g(\hat{\theta})$.

Robustez (idea): $\bar{x}, S^2, S_1^2$ y p son **sensibles a valores atípicos**; para robustez se usan alternativas (p. ej., mediana, MAD, etc.).

Nota: “Eficiencia (CCR) y suficiencia se entienden en un **modelo concreto** (p.ej., Bernoulli para π, Normal para μ, σ^2)”

Bibliografía

- **Ruiz-Maya, L., Martín-Pliego López, F. J. (3.ª ed.). Fundamentos de Inferencia Estadística. Thomson–Paraninfo.**
- Casella, G., & Berger, R. L. (2002). Statistical Inference (2nd ed.). Duxbury/Thomson Learning.
- DeGroot, M. H., & Schervish, M. J. (2012). Probability and Statistics (4th ed.). Pearson.