Estadística II

Tema 4. Métodos de Estimación

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 4 · Objetivo

Al terminar este tema podrás:

- Entender el problema de estimación: a partir de una muestra, proponer un estimador del parámetro de interés.
- Construir estimadores usando tres métodos operativos:
  - Máxima verosimilitud (MV)
  - Método de los momentos (MM)
  - Mínimos cuadrados (MC)
- Aplicar cada método de forma operativa (plantear → resolver → interpretar).
- Interpretar el significado económico de un estimador (tasa, intensidad, media, volatilidad, ajuste).
- Conectar con el Tema 3: una vez obtenido un estimador, evaluar su calidad (insesgadez, consistencia, eficiencia / ECM).

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 4 · Esquema

CORE

- A. Máxima verosimilitud (MV)
- B. Propiedades prácticas del estimador MV
- C. Método de los momentos (MM)
- D. Mínimos cuadrados (MC)
- E. Resumen y ejercicios breves

CORE

Breviario

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

Tres métodos, tres ideas (visión global)

CORE

Método	Idea (en una línea)	¿Cuándo aparece en economía?
MV (Máxima verosimilitud)	Elegir el **estimador que hace más probable la muestra** observada	Estimar tasas, probabilidades, intensidades (impago, conversión, llegadas, demanda) cuando asumimos una distribución
MM (Método de los momentos)	**El estimador se obtiene al igualar momentos teóricos con momentos muestrales** (tantas ecuaciones como parámetros)	Cuando conocemos (o justificamos) momentos teóricos y queremos un estimador rápido y transparente
MC (Mínimos cuadrados)	Elegir estimadores que **minimizan la suma de errores al cuadrado**	Ajuste y regresión: relación entre variables (salario–educación, consumo–renta, demanda–precio), predicción

Sección A. Máxima verosimilitud (MV)

Máxima verosimilitud (MV): idea clave

CORE

- **Objetivo:** Elegir el valor del estimador $\hat{\theta}$ que hace más "verosímil" (más plausible) la muestra observada.

- **Intuición (muy operativa):**

1. Tenemos una **población con parámetro desconocido $\boldsymbol{\theta}$** (por ejemplo, $\pi, \lambda, \mu, \sigma$ …)
2. Observamos una **muestra aleatoria** simple de **tamaño n** $x_1, \dots, x_n$
3. Para distintos valores de θ, evaluamos hasta qué punto explican los datos.
4. El estimador **MV** es el $\hat{\theta}$ que mejor **aproxima $\boldsymbol{\theta}$ con la muestra**: el que maximiza la verosimilitud de la muestra observada.

- **Traducción económica**: Estimar el **parámetro que mejor justifica los datos observados**: proporción poblacional π (éxito/fracaso), intensidad media λ (conteos), media μ (nivel medio) o volatilidad σ (dispersión), según el modelo.

Verosimilitud $L(\theta)$: definición operativa

CORE

Situación

Tenemos una muestra aleatoria simple $x_1, \dots, x_n$ y una función de distribución que depende de un parámetro desconocido θ.

Idea clave

La **verosimilitud mide**, para cada valor de θ, **cómo de compatible es ese valor con los datos observados**.

Definición de la función de verosimilitud (para el caso de m.a.s.)

$$L(\theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

Interpretación práctica

- **Fijamos los datos** $x_1, \dots, x_n$
- Dejamos **variar $\boldsymbol{\theta}$**.
- El estimador de máxima verosimilitud $\widehat{\boldsymbol{\theta}}_{MV}$ **es el valor de $\boldsymbol{\theta}$ que maximiza $\boldsymbol{L(\theta)}$**

CORE

$L(\theta)$ no es una probabilidad "de θ"

Qué sí es

$L(\theta)$ es una **función del parámetro θ con los datos fijados**: nos permite comparar qué valores de θ encajan mejor con la muestra.

Qué NO es

No es correcto interpretar $L(\theta)$ como "la probabilidad de que θ sea este valor".

Cómo pensarlo (regla mental)

- Probabilidad/densidad: variable aleatoria x_i cambia, θ es fijo.
- Verosimilitud: los datos $x_1, \dots, x_n$ están fijados, θ es lo que cambia.

Objetivo

- Elegir $\hat{\boldsymbol{\theta}}_{MV}$ como el **valor de θ que hace máxima la verosimilitud** (derivar e igualar a 0).

Log-verosimilitud $\ell(\theta)$: por qué se usa

CORE

Problema

$L(\theta) = \prod_{i=1}^{n} f(x_i; \theta)$ es un producto: **derivar y maximizar puede ser incómodo**.

Solución práctica: Trabajamos con el **logaritmo**

$$\ell(\theta) = \sum_{i=1}^{n} \log f(x_i; \theta)$$

Ventajas (operativas)

- El producto se convierte en **suma** (más fácil de derivar).
- El máximo de $L(\theta)$ es el mismo que el máximo de $\ell(\theta)$ (el log es creciente).
- Evita números muy pequeños en productos largos.

Objetivo

- **Maximizar $\boldsymbol{\ell(\theta)}$** en lugar de $L(\theta)$.

MV: aplicar el método (plantear → resolver → interpretar)

CORE

Plantear

1. Identifica el parámetro θ (por ejemplo, π, λ, μ o σ).
2. Escribe la función de verosimilitud $L(\theta) = \prod_{i=1}^{n} f(x_i; \theta)$.
3. Pasa a log-verosimilitud $\ell(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(x_i; \theta)$

Resolver

4. Deriva respecto al parámetro e iguala a cero: $\ell'(\theta) = 0$
5. Comprueba que es un máximo (segunda derivada negativa o criterio equivalente).
6. Despeja θ y obtén $\hat{\theta}_{MV}$.

Interpretar

7) Explica qué estima $\hat{\theta}_{MV}$ en el contexto (proporción π, intensidad λ, nivel medio μ, dispersión σ).
8) Revisa que el resultado tenga sentido (rango del parámetro, unidades, orden de magnitud).

CORE

MV: Condición 1: ecuación del score

Idea: Para maximizar $\ell(\theta)$, buscamos los puntos donde la pendiente es cero.

Ecuación del score

$$S(\theta) = \ell'(\theta) = \frac{\partial \ell(\theta)}{\partial \theta}$$

Regla operativa

Candidatos a máximo: resolver

$$\ell'(\theta) = 0$$

Nota práctica

Si hay varios candidatos, comparamos $\ell(\theta)$ (o verificamos con la segunda derivada) y elegimos el que da mayor valor.

CORE

MV: Condición 2: comprobar que es un máximo

Idea: Resolver $\ell'(\theta) = 0$, da candidatos. Ahora debemos comprobar que el punto elegido maximiza la log-verosimilitud.

Criterio habitual (una variable)

Si $\hat{\theta}$ cumple $\ell'(\theta) = 0$ y además $\ell''(\theta) < 0$ entonces $\hat{\theta}$ es un máximo.

Alternativa práctica

Candidatos a máximo: resolver

$$\ell'(\theta) = 0$$

Nota práctica

Si no quieres usar $\ell''(\theta)$: compara el valor de $\ell(\theta)$ en los candidatos y quédate con el que dé mayor log-verosimilitud (si sólo hay un candidato, verificar que es máximo con $\ell''(\theta)$).

Ejemplo MV A1: Bernoulli/Binomial ⇒ $\hat{\pi}$

CORE

Contexto (éxito/fracaso): Observamos n resultados $x_i \in \{0,1\}$ (por ejemplo: impago sí/no, compra sí/no).

Modelo: $x_i \sim B(1, \pi)$, con $\pi \in (0,1)$

Verosimilitud.

$$L(\pi) = \prod_{i=1}^{n} \pi^{x_i}(1-\pi)^{1-x_i} =$$

$$= \pi^{\sum_{i=1}^{n} x_i}(1-\pi)^{n-\sum_{i=1}^{n} x_i}$$

Log-verosimilitud

$$\ell(\pi) = \sum_{i=1}^{n} x_i \log(\pi) + \left(n - \sum_{i=1}^{n} x_i\right) \log(1-\pi)$$

Score (resolver $\ell'(\theta) = 0$)

$$\ell'(\pi) = \frac{\sum_{i=1}^{n} x_i}{\pi} - \frac{n - \sum_{i=1}^{n} x_i}{1-\pi} = 0$$

Estimador MV

$$\hat{\pi}_{MV} = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x} = \mathrm{p}$$

Interpretación económica

$\hat{\pi}_{MV}$ es la proporción muestral de "éxitos": una estimación directa de la proporción poblacional.

CORE

Ejemplo MV A1: Bernoulli/Binomial $\Rightarrow \hat{\pi}$

1. Si en una muestra Bernoulli de tamaño $n = 50$ observas $\sum_{i=1}^{n} x_i = 12$, ¿Cuál es $\hat{\pi}_{MV}$?

a) 0,12

b) 0,24

c) 0,50

d) 0,76

2. En el ejemplo Bernoulli, ¿qué cantidad necesitas conocer para calcular $\hat{\pi}_{MV}$?

a) $\sum_{i=1}^{n} x_i$

b) $\sum_{i=1}^{n} (\mathrm{x_i} - \bar{\mathrm{x}})^2$

c) $\min \mathrm{x_i}$

d) $\mathrm{x_1}$

3. Verdadero / Falso: maximizar $L(\pi)$ y maximizar $\ell(\pi) = \log L(\pi)$ da el mismo $\hat{\pi}$.

a) Verdadero

b) Falso

4. Si el resultado te da $\hat{\pi} = 1,08$, ¿qué concluyes?

a) Es correcto si n es grande.

b) Hay un error de planteamiento o de cálculo (porque $\pi \in [0,1]$

c) Debo dividir por $n - 1$

d) Falta comprobar la segunda derivada

CORE

Ejemplo MV A1: Bernoulli/Binomial ⇒ $\hat{\pi}$

1. **Si en una muestra Bernoulli de tamaño $n = 50$ observas $\sum_{i=1}^{n} x_i = 12$, ¿Cuál es $\hat{\pi}_{MV}$?**
   a) 0,12
   b) 0,24 ✅
   c) 0,50
   d) 0,76
2. **En el ejemplo Bernoulli, ¿qué cantidad necesitas conocer para calcular $\hat{\pi}_{MV}$?**
   a) $\sum_{i=1}^{n} x_i$ ✅
   b) $\sum_{i=1}^{n} (\mathrm{x_i} - \bar{\mathrm{x}})^2$
   c) $\min \mathrm{x_i}$
   d) $\mathrm{x_1}$
3. **Verdadero / Falso: maximizar $L(\pi)$ y maximizar $\ell(\pi) = \log L(\pi)$ da el mismo $\hat{\pi}$.**
   a) Verdadero ✅
   b) Falso
4. **Si el resultado te da $\hat{\pi} = 1,08$, ¿qué concluyes?**
   a) Es correcto si n es grande.
   b) Hay un error de planteamiento o de cálculo (porque $\pi \in [0,1]$ ✅
   c) Debo dividir por $n - 1$
   d) Falta comprobar la segunda derivada

Ejemplo MV A2: Poisson $\Rightarrow \hat{\lambda} = \bar{x}$

CORE

Contexto (conteos): Observamos n conteos $x_1, \dots, x_n$ (nº de clientes por hora, nº de incidencias al día).

Modelo: $x_i \sim Poisson(\lambda)$, con $\lambda > 0$.

Verosimilitud.

$$L(\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$$

Log-verosimilitud

$$\ell(\lambda) = \sum_{i=1}^{n} [-\lambda + x_i \log(\lambda) - \log(x_i!)]$$

Score (resolver $\ell'(\theta) = 0$)

$$\ell'(\lambda) = \sum_{i=1}^{n} \left[-1 + \frac{x_i}{\lambda}\right] = -n + \frac{\sum_{i=1}^{n} x_i}{\lambda} = 0$$

Estimador MV

$$\hat{\lambda}_{MV} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

Interpretación económica

$\hat{\lambda}$ es la intensidad media estimada: el número medio de sucesos por unidad de tiempo/espacio.

Ejemplo MV A3: Normal N $(\mu, \sigma) \Rightarrow \hat{\mu} = \bar{x}$

CORE

Contexto (v.a. continua): Observamos $x_1, \dots, x_n$ (por ejemplo: gasto medio, inflación mensual, rentabilidad diaria).

Modelo: $x_i \sim N(\mu, \sigma)$, con σ^2 conocida.

Verosimilitud.

$$L(\mu) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right]$$

Log-verosimilitud

$$\ell(\mu) = cte - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

Conclusión operativa

Maximizar $\ell(\mu)$ equivale a minimizar $\sum(x_i-\mu)^2$

Estimador MV

$$\hat{\mu}_{MV} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

Interpretación económica

$\hat{\mu}$ estima el nivel medio (tendencia central) de la variable en la población

CORE

Ejemplo MV A3: Normal N $(\mu, \sigma) \Rightarrow$ estimación de σ^2 y relación con S^2 y S_1^2

Contexto (v.a. Continua): Tenemos $x_1, \dots, x_n$ y asumimos que $x_i \sim N(\mu, \sigma)$

El parámetro de dispersión es σ^2(varianza poblacional).

Resultado MV (idea clave)

El estimador MV de σ^2es la media de los cuadrados de las desviaciones respecto a $\bar{x}$:

$$\widehat{\sigma^2}_{MV} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = S^2$$

Comparación (importante)

- $\widehat{\sigma^2}_{MV} = S^2$ (divide por n)
- S_1^2 (divide por n−1, ajuste por grados de libertad).

Interpretación económica

σ^2 mide la volatilidad poblacional; S^2 y S_1^2 son medidas muestrales de esa volatilidad con distinta corrección.

CORE

Ejercicios rápidos (Sección A · MV)

1. **Poisson: si en $n = 20$ días observas $\sum x_i = 64$ incidencias, ¿ $\hat{\lambda}_{MV}$ *es* … ?**
   a) 1,6
   b) 3,2
   c) 6,4
   d) 64
2. **Normal $N(\mu, \sigma)$ con σ conocida: el estimador MV de μ es…**
   a) $\sum \mathrm{x_i}$
   b) $\bar{x}$
   c) s^2
   d) s_1^2

3. **Si calculas S^2 , ¿con qué coincide en el modelo normal?**
   a) $\widehat{\sigma^2}_{\mathrm{MV}}$
   b) $\hat{\sigma}_{\mathrm{MV}}$
   c) $\hat{\mu}_{\mathrm{MV}}$
   d) No coincide con ningún MV
4. **Si para una muestra con $\mathrm{n} = 10$ obtienes $s_1^2 = 4$, entonces S^2 vale…**
   a) 3,6
   b) 4
   c) 4,4
   d) 40

Ejercicios rápidos (Sección A · MV)

CORE

1. **Poisson: si en $n = 20$ días observas $\sum x_i = 64$ incidencias, ¿ $\hat{\lambda}_{MV}$ *es* ...?**
   a) 1,6
   b) 3,2 ✅
   c) 6,4
   d) 64
2. **Normal $N(\mu, \sigma)$ con σ conocida: el estimador MV de μ es...**
   a) $\sum x_i$
   b) $\bar{x}$ ✅
   c) s^2
   d) s_1^2
3. **Si calculas S^2 , ¿con qué coincide en el modelo normal?**
   a) $\widehat{\sigma^2}_{MV}$ ✅
   b) $\hat{\sigma}_{MV}$
   c) $\hat{\mu}_{MV}$
   d) No coincide con ningún MV
4. **Si para una muestra con $n = 10$ obtienes $S_1^2 = 4$, entonces S^2 vale...**
   a) 3,6 ✅
   b) 4
   c) 4,4
   d) 40

Sección B. MV: propiedades prácticas del estimador

CORE

Propiedades del estimador máximo verosímil

Un estimador máximo verosímil ($\hat{\theta}_{MV}$), cuando se cumplen las condiciones regulares habituales (identificabilidad, diferenciabilidad del log-likelihood, información de Fisher finita, etc.), presenta las siguientes propiedades fundamentales:

1. **Consistencia:** $\hat{\theta}_{MV} \xrightarrow{p} \theta$
2. **Eficiencia asintótica:** $V(\hat{\theta}_{MV})$ alcanza asintóticamente la CCR
3. **Normalidad asintótica:**

   Bajo condiciones regulares, $\hat{\theta}_{MV} \xrightarrow{d} \mathrm{N}\left(\theta, \frac{1}{\sqrt{nI(\theta)}}\right)$
4. **Invarianza:** $\hat{\theta}_{MV}$ es invariante

$$\hat{\theta}_{MV} = g\big(T(X)\big)$$

5. En general, $\hat{\theta}_{MV}$ **no es insesgado** en muestras finitas.
6. $\hat{\theta}_{MV}$ suelen escribirse como **función del estadístico suficiente** (pero no tiene por qué cumplirse siempre)

Invarianza del estimador MV

CORE

Enunciado (regla útil)

Si $\hat{\theta}_{MV}$ es el estimador de máxima verosimilitud de θ, entonces para cualquier transformación $g(\cdot)$:

$$\widehat{g(\theta)}_{MV} = g\left(\hat{\theta}_{MV}\right)$$

Interpretación práctica:

No hace falta rehacer toda la verosimilitud si el parámetro que te interesa es una función de otro.

Micro-ejemplo Normal $N(\mu,\sigma)$

Si ya sabemos que $\widehat{\sigma^2}_{MV} = S^2$, entonces: $\hat{\sigma}_{MV} = \sqrt{\widehat{\sigma^2}_{MV}} = \sqrt{S^2} = S$

Estadísticos resumen y suficiencia (idea útil)

CORE

Idea clave

En muchos modelos, para estimar un parámetro no hace falta "toda" la muestra: basta con un resumen $T(X)$.

Definición (suficiencia, en una frase):

Un estadístico $T(X)$ es suficiente para θ si, una vez conocido $T(X)$, el resto de los datos no aporta información adicional sobre θ.

Ejemplos (lo que ya hemos visto)

- Bernoulli/**Binomial**: $T(X) = \sum x_i$ es suficiente para π.
- **Poisson**: $T(X) = \sum x_i$ es suficiente para λ.
- **Normal** $N(\mu, \sigma)$ con σ^2 conocida $T(X) = \bar{x}$ es suficiente para μ.

Lectura práctica

Por eso los estimadores MV suelen escribirse como función de esos resúmenes:

$$\hat{\theta}_{MV} = g\big(T(X)\big)$$

CORE

Estadístico suficiente ⇒ MV depende solo de $T(X)$

Criterio de factorización (idea clave)

Si $T(X)$ es suficiente para θ, la verosimilitud puede escribirse como

$$L(\theta|x) = h(x)g(T(x), \theta)$$

donde $h(x)$ no depende de θ.

Consecuencia directa para MV:

Al maximizar respecto a θ, el factor $h(x)$ no influye, así que:

$$\hat{\theta}_{MV} = \arg max_{\theta}\, L(\theta|x) = \arg max_{\theta}\, g(T(x), \theta)$$

Por tanto, el estimador MV depende de la muestra solo a través del estadístico suficiente:

$$\hat{\theta}_{MV} = g\big(T(x)\big)$$

Lectura práctica

Una vez calculas $T(x)$, toda la información relevante del modelo para estimar θ está en ese resumen.

Aplicación de suficiencia: ¿qué $T(X)$ basta en cada modelo?

CORE

Idea

Si $T(X)$ es suficiente para θ, el estimador MV se puede escribir como función de $T(X)$:

$$\hat{\theta}_{MV} = g(T(X))$$

Modelos típicos (resumen suficiente):

- Bernoulli/**Binomial**: $T(X) = \sum x_i$; $\hat{\pi}_{MV} = \frac{T}{n} = p$
- **Poisson**: $T(X) = \sum x_i$; $\hat{\lambda}_{MV} = \bar{x}$
- **Normal** $N(\mu, \sigma)$ con σ^2 conocida $T(X) = \sum x_i$; $\hat{\mu}_{MV} = \bar{x}$

Lectura práctica

En estos modelos, para estimar el parámetro basta con la media.

CORE

Propiedades asintóticas del estimador MV (idea)

1. Idea

Cuando $n \to \infty$, el estimador se acerca al valor real del parámetro:

$$\hat{\theta}_{\mathrm{MV}} \xrightarrow[\mathrm{p}]{} \theta$$

2. Eficiencia asintótica (idea)

Con n grande, MV suele alcanzar la menor varianza posible (**cota de Cramér–Rao)** bajo condiciones regulares.

3. Normalidad asintótica.

Para n grande, $\hat{\theta}_{MV}$ se aproxima a una Normal centrada en θ, con varianza que cae como $1/n$:

$$\hat{\theta}_{\mathrm{MV}} \xrightarrow{d} \mathrm{N}\left(\theta, \frac{1}{\sqrt{nI(\theta)}}\right)$$

[$\mathrm{I}(\theta) =$ información de Fisher del modelo]

Ejercicios rápidos (Sección B · Propiedades MV)

CORE

1. **Invarianza: si $\hat{\lambda}_{MV} = 3,2$ para una Poisson, el estimador MV de 2λ es …**
   a) 1,6
   b) 3,2
   c) 6,4
   d) 2/3,2
2. **Consistencia (interpretación): $\hat{\theta}_n \xrightarrow[p]{} \theta$ significa que cuando n crece…**
   a) $\hat{\theta}_n$ se vuelve exactamente igual a θ para cualquier muestra
   b) $\hat{\theta}_n$ tiende a estar cada vez más cerca de θ con alta probabilidad
   c) $\hat{\theta}_n$ siempre es insesgado
   d) la varianza muestral divide por n−1
3. **Normalidad asintótica (lectura): para n grande, el error $\hat{\theta} - \theta$ suele…**
   a) ser aproximadamente uniforme
   b) ser aproximadamente normal
   c) Crecer con n
   d) desaparecer exactamente
4. **Suficiencia (aplicación): en Bernoulli, para calcular $\hat{\pi}_{MV}$ basta conocer**
   a) $\sum x_i$
   b) $\sum(x_i - \bar{x})^2$
   c) $\min x_i$
   d) x_1

Ejercicios rápidos (Sección B · Propiedades MV)

CORE

1. **Invarianza: si $\hat{\lambda}_{MV} = 3,2$ para una Poisson, el estimador MV de 2λ es …**
   a) 1,6
   b) 3,2
   c) 6,4 ✅
   d) 2/3,2
2. **Consistencia (interpretación): $\hat{\theta}_n \underset{p}{\rightarrow} \theta$ significa que cuando n crece…**
   a) $\hat{\theta}_n$ se vuelve exactamente igual a θ para cualquier muestra
   b) $\hat{\theta}_n$ tiende a estar cada vez más cerca de θ con alta probabilidad ✅
   c) $\hat{\theta}_n$ siempre es insesgado
   d) la varianza muestral divide por n−1
3. **Normalidad asintótica (lectura): para n grande, el error $\hat{\theta} - \theta$ suele…**
   a) ser aproximadamente uniforme
   b) ser aproximadamente normal ✅
   c) Crecer con n
   d) desaparecer exactamente
4. **Suficiencia (aplicación): en Bernoulli, para calcular $\hat{\pi}_{MV}$ basta conocer**
   a) $\sum x_i$ ✅
   b) $\sum(x_i - \bar{x})^2$
   c) $\min x_i$
   d) x_1

Sección C. Método de los momentos (MM)

Método de los momentos (MM): definición operativa

CORE

Momento teórico

El momento poblacional $(\mathrm{E(x)} = \mu, \mathrm{E}\left(\mathrm{x}^2\right) = \alpha_2, \mathrm{V(x)} = \sigma^2, \dots)$ se expresa en función del parámetro θ.

Momento muestral

Se calcula en función de las variables muestrales: $\bar{x} = \frac{1}{n}\sum x_i\,; a_2 = \frac{1}{n}\sum x_i^2\,, \dots$

Regla del método (MM)

Si el modelo tiene k parámetros, elegimos k momentos y resolvemos el sistema formado por k ecuaciones:

$$\text{momento teórico}_j = \text{momento muestral}_j \qquad para\ j = 1, \dots, k$$

Interpretación económica

MM produce estimadores “transparentes”: se obtienen igualando indicadores del modelo con indicadores observados (medias, varianzas, etc.).

MM: aplicar el método (plantear → resolver → interpretar)

CORE

Plantear

1. Identifica el/los parámetros desconocidos (1 parámetro o varios).
2. Elige tantos momentos como parámetros (por ejemplo, $\mathrm{E(x)}, \mathrm{E}\left(\mathrm{x}^2\right), \mathrm{V(x)}, \ldots$)
3. Escribe los momentos teóricos en función de los parámetros.

Momento muestral

4. Calcula los momentos muestrales correspondientes $\bar{x}, a_2, S^2, \ldots$
5. Igualas: momento teórico = momento muestral.
6. Resuelves el sistema y obtienes $\hat{\theta}_{MM}$.

Interpretar

7. Explica qué está estimando el parámetro en tu contexto (media, dispersión, intensidad…).
8. Comprueba que el resultado tiene sentido (rango, unidades, orden de magnitud).

CORE

Propiedades del estimador por el método de los momentos

Un estimador por el método de los momentos ($\hat{\theta}_{MM}$), en general, no es un estimador óptimo. Su comportamiento respecto a las propiedades:

1. **Consistencia:** $\hat{\theta}_{MM} \xrightarrow{p} \theta$
2. **Eficiencia:** $\hat{\theta}_{MM}$ no tiene por qué ser eficiente ni alcanzar la CCR
3. **Invarianza:** $\hat{\theta}_{MV}$ es invariante.

$$\hat{\theta}_{MV} = g\big(T(X)\big)$$

4. En general, $\hat{\theta}_{MM}$ **no es insesgado** en muestras finitas
5. $\hat{\theta}_{\mathrm{MM}}$ no es, en general función del estadístico suficiente.

Ejemplo MM C1: Poisson $\Rightarrow \hat{\lambda}_{MM} = \bar{x}$

CORE

Contexto (conteos)

Observamos n conteos $x_1, \dots, x_n$ (nº de clientes por hora, nº de incidencias al día).

Modelo: $x_i \sim Poisson(\lambda)$, con $\lambda > 0$.

Momento teórico (media del modelo)

$$E(x) = \lambda$$

Momento muestral

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

Estimador por momentos

$$\hat{\lambda}_{MM} = \bar{x} \quad ; \ \mathrm{T(x)} = \bar{x}$$

Interpretación económica: $\hat{\lambda}$ estima la intensidad media: el número medio de sucesos por unidad de tiempo/espacio.

Ejemplo MM C2: Uniforme $U(0, \theta) \Rightarrow \hat{\theta}_{MM} = 2\bar{x}$

CORE

Contexto

Observamos $x_1, \dots, x_n$ con valores entre 0 y un “tope” θ (por ejemplo: un límite máximo técnico o normativo).

Modelo:

$$x_i \sim U(0, \theta)$$

con $\theta > 0$.

Momento teórico (media del modelo)

$$E(x) = \frac{\theta}{2}$$

Momento muestral

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

Igualación MM

$$\frac{\theta}{2} = \bar{x}$$

Estimador por momentos

$$\hat{\theta}_{MM} = 2\bar{x} \;\; ; \;\; \mathrm{T(x)} = \bar{x}$$

Interpretación económica: $\hat{\theta}$ estima el “tope” θ a partir del nivel medio observado.

MM vs MV: cuándo coinciden (mensaje práctico)

CORE

Idea

A veces, MM y MV llevan al mismo estimador; otras veces no.

Ejemplos del tema

- Poisson:

  $MM\colon \hat{\lambda} = \bar{x}$

  $MV\colon \hat{\lambda} = \bar{x}$

→ coinciden

- Bernoulli/Binomial:

  $MM\colon \hat{\pi} = \bar{x}$

  $MV\colon \hat{\pi} = \bar{x}$

→ coinciden

Ejemplos del tema

- Uniforme $U(0, \theta)$:

  $MM\colon \hat{\theta} = 2\bar{x}$

  $MV\colon \hat{\theta} = \max\{x_1, \dots, x_n\}$

→ No coinciden

Lectura práctica

- MM suele ser más “directo” (igualar indicadores).
- MV suele estar más ligado al ajuste global del modelo y respeta mejor las restricciones del soporte (por eso en Uniforme el MV cae en el máximo observado).

Ejercicios rápidos (Sección C · MM vs MV)

CORE

1. **Poisson: si $\overline{x} = 4,5$, el estimador por momentos para λ es …**
   a) 0,22
   b) 2,25
   c) 4,5
   d) 9
2. **Uniforme $U(0, \theta)$; si $\overline{x} = 6$, el estimador MM para θ es…**
   a) 3
   b) 6
   c) 12
   d) $\max x_i$
3. **Bernoulli/Binomial: si $\sum x_i = 18$ y $n = 60$, el estimador por momentos de π es …**
   a) 0,18
   b) 0,3
   c) 0,6
   d) 18
4. **Verdadero / Falso: en Uniforme $U(0, \theta)$, MM y MV siempre coinciden.**
   a) Verdadero
   b) Falso

CORE

Ejercicios rápidos (Sección C · MM vs MV)

1. **Poisson: si $\bar{x} = 4,5$, el estimador por momentos para λ es …**
   a) 0,22
   b) 2,25
   c) 4,5 ✅
   d) 9
2. **Uniforme $U(0, \theta)$; si $\bar{x} = 6$, el estimador MM para θ es…**
   a) 3
   b) 6
   c) 12 ✅
   d) $\max x_i$
3. **Bernoulli/Binomial: si $\sum x_i = 18$ y $n = 60$, el estimador por momentos de π es …**
   a) 0,18
   b) 0,3 ✅
   c) 0,6
   d) 18
4. **Verdadero / Falso: en Uniforme $U(0, \theta)$, MM y MV siempre coinciden.**
   a) Verdadero
   b) Falso ✅

Sección D. Mínimos cuadrados (MC)

Mínimos cuadrados (MC): definición operativa

CORE

Situación

Tenemos una población con un parámetro θ desconocido y se extrae una m.a.s.(n).

Error (residuo)

$$e_i(\theta) = x_i - \theta$$

Criterio de mínimos cuadrados

Elegimos θ para minimizar la suma de errores al cuadrado:

$$f(\theta) = \sum_{i=1}^{n} e_i(\theta)^2 = \sum_{i=1}^{n} (x_i - \theta)^2 \Rightarrow \ \hat{\theta}_{MC} = \arg\min_{\theta} f(\theta)$$

Interpretación económica

Buscamos el parámetro que produce el mejor ajuste “promedio” (penalizando mucho los errores grandes).

Ejemplo MC D1: ajustar una constante $\Rightarrow \hat{\theta} = \bar{x}$

CORE

Problema

Queremos aproximar todos los datos por una constante θ (un “nivel medio”):

Error (residuo)

$$e_i(a) = x_i - \theta$$

Función objetivo (mínimos cuadrados)

$$f(\theta) = \sum_{i=1}^{n} (x_i - \theta)^2$$

Resultado (ver Estadística 1, media = centro de gravedad)

$$\hat{\theta} = \bar{x}$$

Interpretación económica

Si usamos un único número para resumir el nivel de la variable (gasto medio, precio medio, etc.), el mejor ajuste en mínimos cuadrados es la media muestral.

Propiedades del estimador por el método de mínimos cuadrados

CORE

Si : $x_i = \theta + u_i$ y $E[u_i] = 0$, entonces:

1. $\hat{\theta}_{MC}$ **es consistente**

2. $\mathbf{V}(\hat{\theta}_{MC}) = \frac{\boldsymbol{\sigma}^2}{\mathbf{n}}$

3. $\hat{\theta}_{MC}$ **es insesgado**.

4. $\hat{\theta}_{MC}$ es normal si la población es normal.

Ejercicios rápidos (Sección D · MC)

CORE

1. **En mínimos cuadrados, el parámetro se elige para…**
   a) Maximizar $L(\theta)$
   b) Minimizar $f(\theta) = \sum e_i(\theta)^2$
   c) Igualar $E(x) = \bar{x}$
   d) Maximizar $\ell(\theta)$
2. **Ajuste por constante: si los datos son 2,4,6,8, el valor $\hat{\theta}_{MC}$ es…**
   a) 4
   b) 5
   c) 6
   d) 20

3. **En la regresión simple por mínimos cuadrados elige a y b para minimizar……**
   a) $\sum(x_i - \bar{x})^2$
   b) $\sum(y_i - \bar{y})^2$
   c) $\sum[y_i - (a + bx)]^2$
   d) $\sum \log f(x_i; \theta)$

4. **Verdadero / Falso: el criterio de mínimos cuadrados penaliza más los errores grandes que los pequeños.**
   a) Verdadero
   b) Falso

Ejercicios rápidos (Sección D · MC)

CORE

1. **En mínimos cuadrados, el parámetro se elige para…**
   a) Maximizar $L(\theta)$
   b) Minimizar $f(\theta) = \sum e_i(\theta)^2$ ✅
   c) Igualar $E(x) = \bar{x}$
   d) Maximizar $\ell(\theta)$
2. **Ajuste por constante: si los datos son 2,4,6,8, el valor $\hat{\theta}_{MC}$ es…**
   a) 4
   b) 5 ✅
   c) 6
   d) 20
3. **En la regresión simple por mínimos cuadrados elige a y b para minimizar……**
   a) $\sum(x_i - \bar{x})^2$
   b) $\sum(y_i - \bar{y})^2$
   c) $\sum[y_i - (a + bx)]^2$ ✅
   d) $\sum \log f(x_i; \theta)$
4. **Verdadero / Falso: el criterio de mínimos cuadrados penaliza más los errores grandes que los pequeños.**
   a) Verdadero ✅
   b) Falso

Cierre de tema

Resumen: MV vs MM vs MC (cuándo usar cada uno)

CORE

MV (máxima verosimilitud)

- Cuando tienes un **modelo probabilístico** $f(x;\theta)$.
- Estima θ **maximizando** $\boldsymbol{L(\theta)}$ (o $\ell(\theta)$)
- Suele dar muy buen comportamiento con **muestras grandes**.

MM (método de los momentos)

- Cuando conoces **momentos teóricos** del modelo (media, varianza, …).
- Estima **igualando** momentos **teóricos** con momentos **muestrales**.
- Es **simple y transparente** (se apoya en indicadores como $\bar{x}, S^2$, …)

MC (mínimos cuadrados)

- Cuando defines un **modelo de ajuste** y **puedes medir errores** $e_i(\theta)$.
- Estima minimizando $f(\theta) = \sum e_i(\theta)^2$.
- Recupera resultados conocidos: ajustar una constante → $\bar{x}$; ajuste de rectas (Estadística I).

Tema 4 (E2) – Propiedades generales de MV / MM / MC

CORE

Método	Idea	Modelo prob.	Invarianza	Consistencia	Normalidad asint.	Insesgadez finita
MV	Max ℓ(θ)	✅	✅	✅	✅	⚠️
MM	Iguala momentos	⚠️	⚠️	✅	⚠️	⚠️
MC	Min Σe^2	✗	✗	✅	⚠️	⚠️

Estadístico	Modelo típico	MV	MM	MC
$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$	Normal (μ), Poisson (λ)	✅	✅	✅ (const.)
$p = \frac{1}{n}\sum_{i=1}^{n} x_i \; ; x_i = \{0, 1\}$	Bernoulli/ Binomial (π)	✅	✅	✅ (ajuste constante con 0/1→ $\hat{a} = \bar{x} = \hat{\pi}$)
$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$	Normal (σ^2)	✅ ($\hat{\sigma}^2$)	✅ (μ_2)	✅ (como ECM residual del ajuste constante)
$s_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$	Normal (σ^2)	✗	✗	✗

Leyenda: ✅ =SI, ⚠️ =DEPENDE, ☐ =NO

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.