Inferencia Estadística

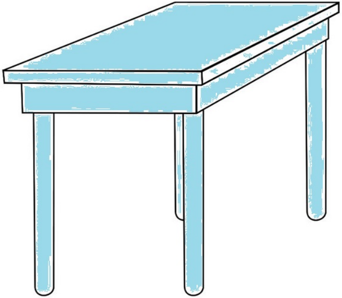
Clase 5 - Estimación Puntual

Paula Spano y Lucía Babino

Universidad Torcuato Di Tella

¿Solo estimaremos con promedios?

¿Cuánto mide la mesa?



La mesa es medida con un hilo por....

¿Quién mide la mesa? Juani

1. Juani tiene un ovillo de lana.
2. Apoya el extremo de la lana a un extremo de la mesa y la estira hasta un lugar aleatorio de la mesa pero sin pasarse.
3. Cortamos la lana hasta donde llegó y la medimos.
4. Repetimos los items 1-3 varias veces.

Objetivo: usar las medidas de las lanas para estimar la longitud de la mesa.

¿Cuánto mide la mesa?

Estas son las $n = 5$ primeras mediciones realizadas por Juani:

$$2.31 \quad 2.89 \quad 1.33 \quad 0.73 \quad 1.91$$

¿Cuál es el parámetro de interés?

Propongamos un modelo para estas mediciones.

¿Cómo estimamos la longitud de la mesa?

¿¿Propuestas??

¿Qué estimadores propusieron? ¿son consistentes?

> Conclusión: Veremos que no todo estimador consistente es un promedio.

¿Qué estimador elegimos?

Necesitamos de otro criterio de bondad.
Para definirlo tendremos que definir algunos conceptos.
Pero antes, realicemos algunos gráficos para ganar intuición.

Gráficos

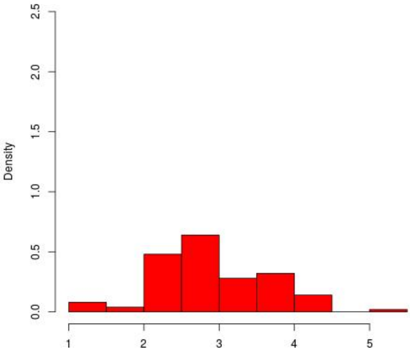
Muchos monos calculan $\widehat{\theta}_n = 2\overline{X}_n$ con $n = 5$

Estudiemos el comportamiento de $\widehat{\theta}_5 = 2\overline{X}_5$.

	Nombre	Mediciones					$\widehat{\theta}_5$
1	Juani	2.11	1.25	0.37	0.33	2.57	2.65
2	Chita	1.85	2.80	1.36	0.10	2.29	3.36
3	Perico	2.03	0.76	1.16	0.86	1.97	2.71
4	Bonita	0.29	2.06	1.90	2.76	2.79	3.92
5	Milo	0.06	1.27	2.00	1.07	0.17	1.83
6	Poli	1.72	1.93	2.73	2.27	2.80	4.58
7	Peperina	1.14	3.00	0.91	1.18	1.20	2.97
8	Tuco	2.91	1.34	2.49	2.43	1.48	4.26
9	Babá	1.27	0.54	0.45	1.04	0.34	1.46
10	Pepa	0.81	2.05	1.76	0.49	2.15	2.90
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.

¿Cómo podríamos visualizar estos datos?

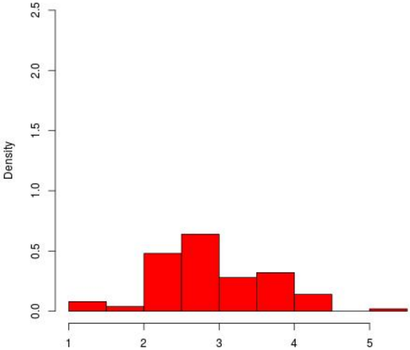
Histograma de $\widehat{\theta}_n = 2\overline{X}_n : n = 5$



	Nombre	$\widehat{\theta}_5$
1	Juani	2.65
2	Chita	3.36
3	Perico	2.71
4	Bonita	3.92
5	Milo	1.83
6	Poli	4.58
7	Peperina	2.97
8	Tuco	4.26
.	.	.
.	.	.

¿Qué nos dice el histograma? Por ejemplo, ¿cuál es el intervalo con mayor cantidad de datos?

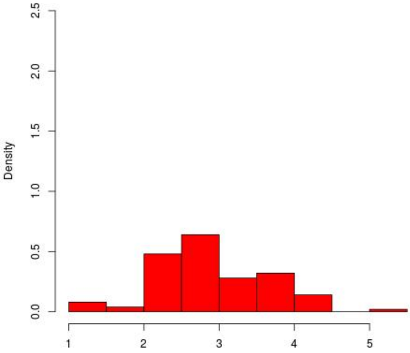
Histograma de $\widehat{\theta}_n = 2\overline{X}_n : n = 5$



	Nombre	$\widehat{\theta}_5$
1	Juani	2.65
2	Chita	3.36
3	Perico	2.71
4	Bonita	3.92
5	Milo	1.83
6	Poli	4.58
7	Peperina	2.97
8	Tuco	4.26
.	.	.
.	.	.

Un histograma es a la vez un estimador, ¿de quién?

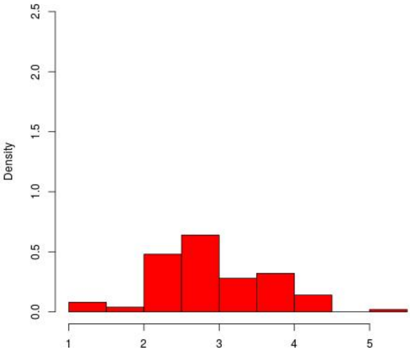
Histograma de $\widehat{\theta}_n = 2\overline{X}_n : n = 5$



	Nombre	$\widehat{\theta}_5$
1	Juani	2.65
2	Chita	3.36
3	Perico	2.71
4	Bonita	3.92
5	Milo	1.83
6	Poli	4.58
7	Peperina	2.97
8	Tuco	4.26
.	.	.
.	.	.

¿Qué forma parece tener la distribución de $\widehat{\theta}_n = 2\overline{X}_n$?

Histograma de $\widehat{\theta}_n = 2\overline{X}_n : n = 5$



	Nombre	$\widehat{\theta}_5$
1	Juani	2.65
2	Chita	3.36
3	Perico	2.71
4	Bonita	3.92
5	Milo	1.83
6	Poli	4.58
7	Peperina	2.97
8	Tuco	4.26
.	.	.
.	.	.

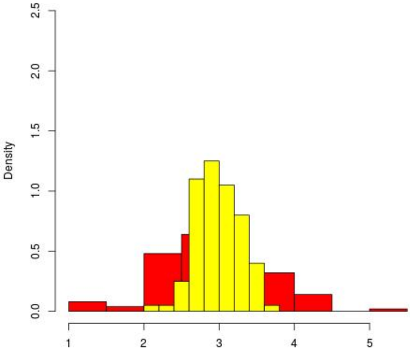
¿Qué pasa si ahora cada mono estima la mesa en base a 30 trozos de piolín?

Muchos monos con $n = 5$ y $n = 30$: $\widehat{\theta}_n = 2\overline{X}_n$

	Nombre	$\widehat{\theta}_5$	$\widehat{\theta}_{30}$
1	Juani	2.65	3.2
2	Chita	3.36	2.95
3	Perico	2.71	3.2
4	Bonita	3.92	3.23
5	Milo	1.83	2.93
6	Poli	4.58	2.9
7	Peperina	2.97	3.03
8	Tuco	4.26	2.79
9	Babá	1.46	3.41
10	Pepa	2.90	3.29
.	.	.	.
.	.	.	.
.	.	.	.

¿Cómo creen que será el histograma de $\widehat{\theta}_{30} = 2\overline{X}_{30}$ en relación al de $\widehat{\theta}_5 = 2\overline{X}_5$?

Histogramas de $\widehat{\theta}_n = 2\overline{X}_n : n = 5$ y 30



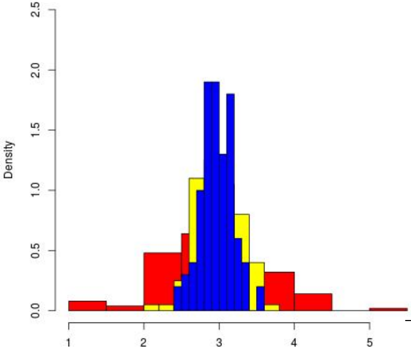
	Nombre	$\widehat{\theta}_5$	$\widehat{\theta}_{30}$
1	Juani	2.65	3.2
2	Chita	3.36	2.95
3	Perico	2.71	3.2
4	Bonita	3.92	3.23
5	Milo	1.83	2.93
6	Poli	4.58	2.9
7	Peperina	2.97	3.03
.	.	.	.
.	.	.	.

(rojo: $n = 5$ y amarillo: $n = 30$)
Vemos que el histograma de $\widehat{\theta}_{30}$ está menos disperso que el de $\widehat{\theta}_5$
¿Por qué?

Muchos monos con $n = 5$, $n = 30$ y $n = 50$: $\widehat{\theta}_n = 2\overline{X}_n$

	Nombre	$\widehat{\theta}_5$	$\widehat{\theta}_{30}$	$\widehat{\theta}_{50}$
1	Juani	2.65	3.2	2.96
2	Chita	3.36	2.95	2.88
3	Perico	2.71	3.2	3.18
4	Bonita	3.92	3.23	3.18
5	Milo	1.83	2.93	2.81
6	Poli	4.58	2.9	2.59
7	Peperina	2.97	3.03	3.01
8	Tuco	4.26	2.79	3.1
9	Babá	1.46	3.41	3.01
10	Pepa	2.90	3.29	3.11
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

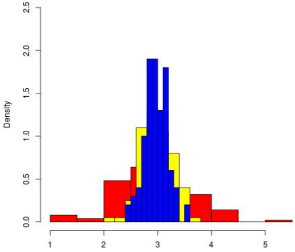
Histogramas de $\widehat{\theta}_n = 2\overline{X}_n : n = 5, 30$ y 50



	Nombre	$\widehat{\theta}_5$	$\widehat{\theta}_{30}$	$\widehat{\theta}_{50}$
1	Juani	2.65	3.2	2.96
2	Chita	3.36	2.95	2.88
3	Perico	2.71	3.2	3.18
4	Bonita	3.92	3.23	3.18
5	Milo	1.83	2.93	2.81
6	Poli	4.58	2.9	2.59
7	Peperina	2.97	3.03	3.01
.	.	.	.	.
.	.	.	.	.

(rojo: $n = 5$, amarillo: $n = 30$ y azul: $n = 50$)

Conclusiones



- La distribución de $\widehat{\theta}_n = 2\overline{X}_n$ se parece a una normal (para todos los n's) debido al TCL.
- A medida que n aumenta, la distribución de $\widehat{\theta}_n = 2\overline{X}_n$ tiene cada vez menor dispersión.

(Veamos qué pasa con $\widetilde{\theta}_n = \max\limits_{1 \leq i \leq n} \{X_i\}$)

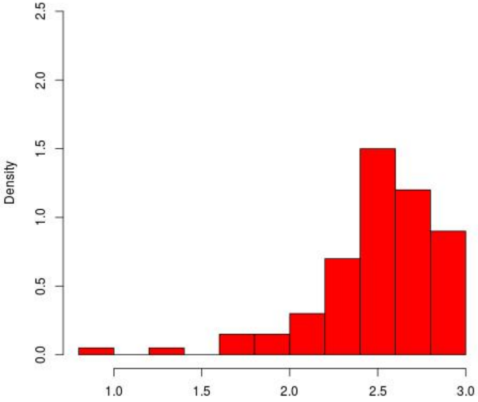
Muchos monos calculan $\widetilde{\theta}_n = \max_{1 \le i \le n}\{X_i\}$ con $n = 5$

Estudiemos el comportamiento de $\widetilde{\theta}_5 = \max\{X_1, \ldots, X_5\}$.

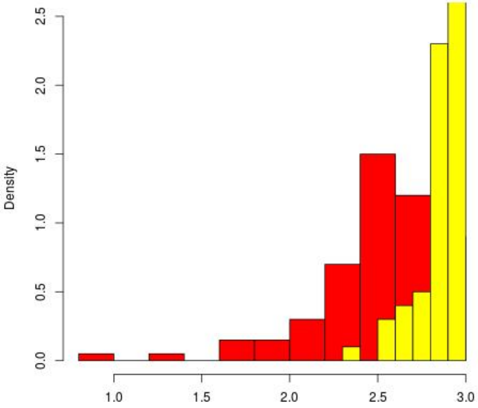
	Nombre	Mediciones					$\widetilde{\theta}_5$
1	Juani	2.11	1.25	0.37	0.33	2.57	2.57
2	Chita	1.85	2.80	1.36	0.10	2.29	2.29
3	Perico	2.03	0.76	1.16	0.86	1.97	2.03
4	Bonita	0.29	2.06	1.90	2.76	2.79	2.79
5	Milo	0.06	1.27	2.00	1.07	0.17	2.00
6	Poli	1.72	1.93	2.73	2.27	2.80	2.80
7	Peperina	1.14	3.00	0.91	1.18	1.20	3.00
8	Tuco	2.91	1.34	2.49	2.43	1.48	2.91
9	Babá	1.27	0.54	0.45	1.04	0.34	1.27
10	Pepa	0.81	2.05	1.76	0.49	2.15	2.15
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.

¿Cómo podríamos visualizar estos datos?

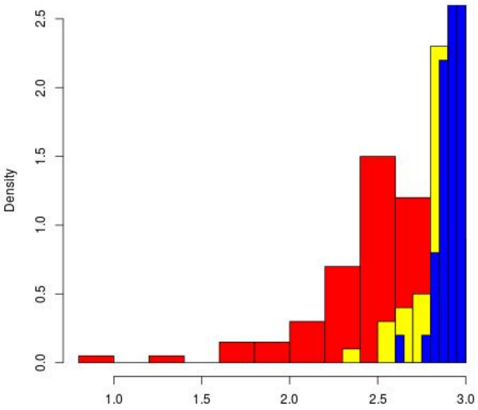
Histograma de $\widetilde{\theta}_n = \max_{1 \le i \le n}\{X_i\} : n = 5$



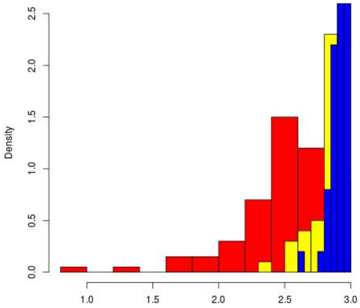
Histogramas de $\widetilde{\theta}_n = \max_{1 \leq i \leq n}\{X_i\} : n = 5$ y 30



Histogramas de $\widetilde{\theta}_n = \max_{1 \leq i \leq n} \{X_i\} : n = 5, 30$ y 50



Histogramas de $\widetilde{\theta}_n = \max_{1\leq i\leq n}\{X_i\} : n = 5, 30$ y 50

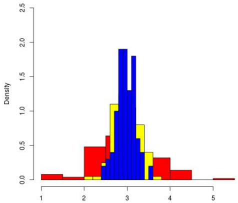


¿Qué ocurre a medida de n aumenta con la distribución de $\widetilde{\theta}_n = \max_{1\leq i\leq n}\{X_i\}$?

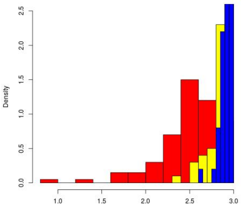
- ¿está cada vez menos dispersa?
- ¿se parece cada vez más a una normal?

Histogramas de $\widehat{\theta}_n$ y $\widetilde{\theta}_n$

Histogramas de $\widehat{\theta}_n$



Histogramas de $\widetilde{\theta}_n$



¿Similitudes y diferencias?

¿Cuál elegimos?

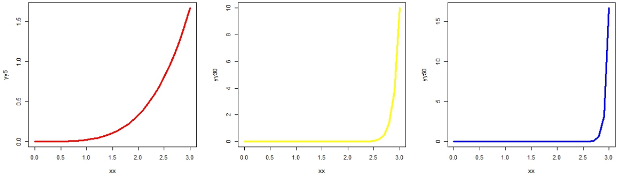
Ambos son consistentes? necesitamos otro criterio de bondad para decidir cuál elegir. Para definirlo, debemos plantear los conceptos de **sesgo** y **varianza** de un estimador.

Algunas definiciones

Un estimador $\widehat{\theta}_n = \widehat{\theta}_n(X_1, \dots, X_n)$ es una v. a. $\Rightarrow$ tiene:

- Distribución $\rightarrow$ distribución muestral.

  ¿Cómo visualizamos la dist. de los estimadores en el ej.? ¿La podemos conocer?



Algunas definiciones

Un estimador $\widehat{\theta}_n = \widehat{\theta}_n(X_1, \ldots, X_n)$ es una v. a. $\Rightarrow$ tiene:

- Distribución $\rightarrow$ distribución muestral.

  ¿Cómo visualizamos la dist. de los estimadores en el ej.? ¿La podemos conocer?

- Esperanza: $\mathbb{E}(\widehat{\theta}_n)$
- Varianza: $\mathbb{V}(\widehat{\theta}_n)$
  - Desvío estándar: $\sqrt{\mathbb{V}(\widehat{\theta}_n)} = \mathsf{SE}(\widehat{\theta}_n) \rightarrow$ Error estándar.

Los siguientes criterios de bondad que vamos a definir están estrechamente ligados a la esperanza y al error estándar de un estimador (varianza).

Esperanza y varianza de estimadores

Definiciones

- Definimos al Sesgo (bias) de un estimador $\widehat{\theta}_n$ como:

$$\mathbb{B}(\widehat{\theta}_n) = \mathbb{E}\left(\widehat{\theta}_n\right) - \theta$$

- Un estimador $\widehat{\theta}_n$ se dice insesgado si

$$\mathbb{B}(\widehat{\theta}_n) = 0 \quad \text{es decir si} \quad \mathbb{E}\left(\widehat{\theta}_n\right) = \theta$$

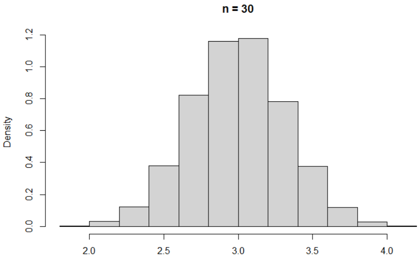
- Un estimador $\widehat{\theta}_n$ se dice que es asintóticamente insesgado si

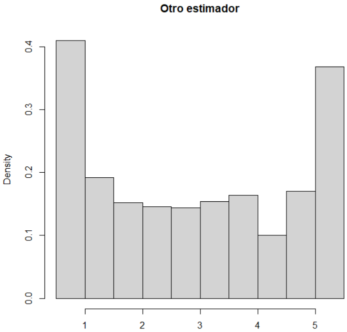
$$\mathbb{B}(\widehat{\theta}_n) \underset{n\to\infty}{\longrightarrow} 0 \quad \text{es decir si} \quad \mathbb{E}\left(\widehat{\theta}_n\right) \underset{n\to\infty}{\longrightarrow} \theta$$

Otras características de bondad de un estimador es que resulte insesgado o bien, asintóticamente insesgado.

Estimadores insesgados

La insesgadez es una característica buena de un estimador, pero, ¿será suficiente?





- ¿parece insesgado para 3?
- ¿es "bueno"?
- ¿qué lo diferencia del primero?

Conclusión

1. Que un estimador sea insesgado no garantiza que sea bueno.

2. Para que un estimador sea "bueno" necesitamos que tenga:

   - poco sesgo: $\mathbb{B}(\widehat{\theta}_n) = \mathbb{E}\left(\widehat{\theta}_n\right) - \theta \cong 0$

     (el promedio de las estimaciones está cerca de θ)

   - poca varianza: $\mathbb{V}(\widehat{\theta}_n) \cong 0$ o, equivalentemente, **poco error estándar**.

     (las estimaciones están cerca entre sí, tienen poca dispersión)

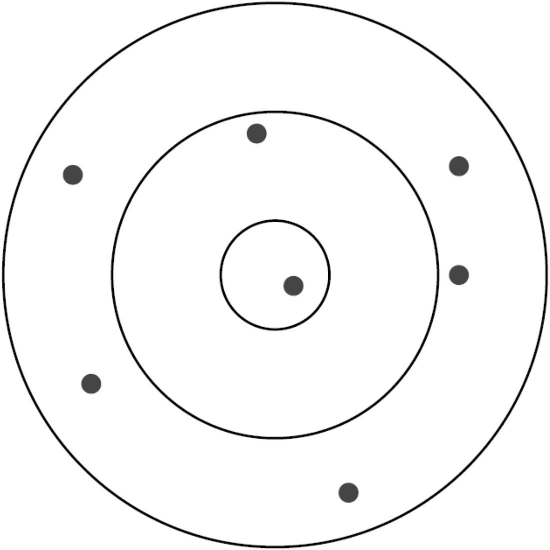
Dilema: si tenemos un estimador con poco sesgo y mucha varianza y otro con mucho sesgo y poca varianza, ¿cuál elegimos?

$\uparrow$

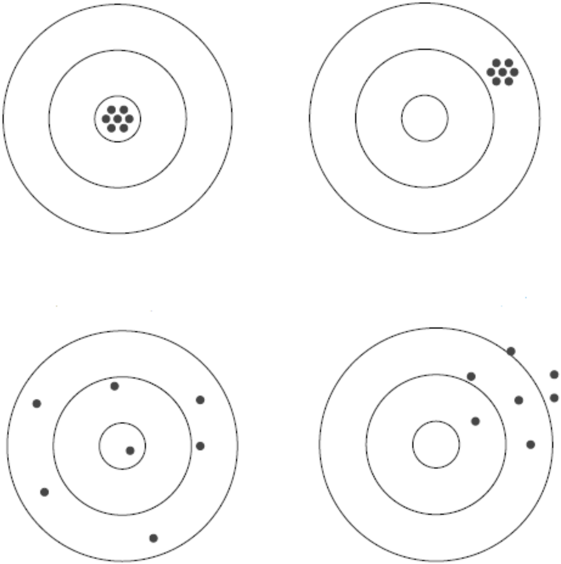
Trade off sesgo-varianza

Trade off Sesgo-varianza

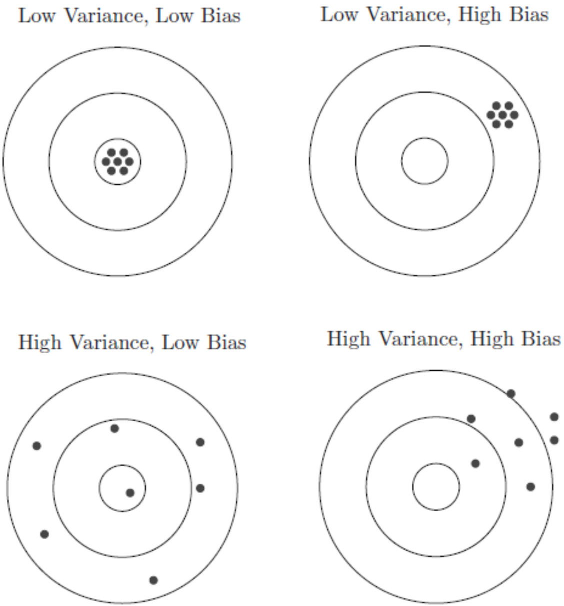
Estimación Puntual



¿Sesgo y Varianza en cada escenario?



Trade off Sesgo - Varianza: ¿cuál elegimos?



Error Cuadrático Medio

Error Cuadrático Medio

Definición

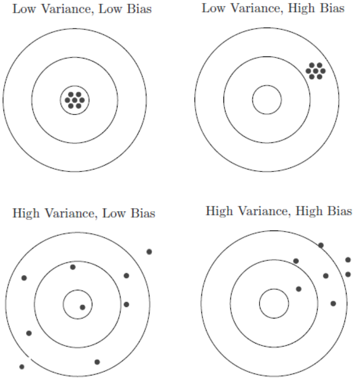
El error cuadrático medio de un estimador $\widehat{\theta}_n$ se define como

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

¿¿Cómo queremos que sea $\mathrm{ECM}(\widehat{\theta}_n)$??

¿¿Cómo se relaciona $\mathrm{ECM}(\widehat{\theta}_n)$ con el *trade off sesgo - varianza*??

ECM y Trade off Sesgo - Varianza



¿Qué estimador tiene menor $\mathrm{ECM}(\widehat{\theta}_n) = \mathbb{E}\left[\left(\widehat{\theta}_n - \theta\right)^2\right]$?

El ECM será chico cuando el sesgo y la varianza sean chicos.

Error cuadrático medio

Teorema

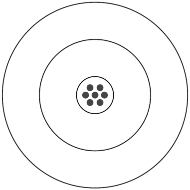
$$ECM\left(\widehat{\theta}_n\right) = \mathbb{B}(\widehat{\theta}_n)^2 + \mathbb{V}\left(\widehat{\theta}_n\right)$$

Corolario

Si $\widehat{\theta}_n$ es insesgado, entonces

$$\mathrm{ECM}\left(\widehat{\theta}_n\right) = V\left(\widehat{\theta}_n\right)$$

Gráficamente:



Por lo tanto, si un estimador es insesgado, tener un ECM chico, es equivalente a tener varianza chica.

Error cuadrático medio

Teorema

$$ECM\left(\widehat{\theta}_n\right) = \mathbb{B}(\widehat{\theta}_n)^2 + \mathbb{V}\left(\widehat{\theta}_n\right)$$

Demostración (optativa):

$$\mathrm{ECM}\left(\widehat{\theta}_n\right) = \mathbb{E}\left[\left(\widehat{\theta}_n - \theta\right)^2\right] = \mathbb{E}\left[\left(\widehat{\theta}_n - \mathbb{E}(\widehat{\theta}_n) + \mathbb{E}(\widehat{\theta}_n) - \theta\right)^2\right]$$

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

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

$$= \mathbb{V}(\widehat{\theta}_n) + \left(\mathbb{E}(\widehat{\theta}_n) - \theta\right)^2 + 2\left(\mathbb{E}(\widehat{\theta}_n) - \theta\right)^2 \mathbb{E}\left(\widehat{\theta}_n - \mathbb{E}(\widehat{\theta}_n)\right)$$

$$= \mathbb{V}(\widehat{\theta}_n) + \mathbb{B}(\widehat{\theta}_n)^2 + 2\left(\mathbb{E}(\widehat{\theta}_n) - \theta\right)^2 \left(\mathbb{E}(\widehat{\theta}_n) - \mathbb{E}(\widehat{\theta}_n)\right)$$

$$= \mathbb{V}(\widehat{\theta}_n) + \mathbb{B}(\widehat{\theta}_n)^2.$$

Error cuadrático medio

Teorema

$$ECM\left(\widehat{\theta}_n\right) = \mathbb{B}(\widehat{\theta}_n)^2 + \mathbb{V}\left(\widehat{\theta}_n\right)$$

Dilema: si tenemos un estimador con poco sesgo y mucha varianza y otro con mucho sesgo y poca varianza, ¿cuál elegimos?

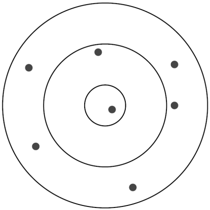
$\uparrow$

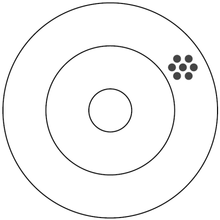
¡El de menor ECM!

Error cuadrático medio

Teorema

$$ECM\left(\widehat{\theta}_n\right) = \mathbb{B}(\widehat{\theta}_n)^2 + \mathbb{V}\left(\widehat{\theta}_n\right)$$





Característica de bondad de un estimador: Poseer error cuadrático medio bajo.

Criterios de bondad

Criterios de bondad de estimadores:

- **Consistencia:** $\widehat{\theta}_n \xrightarrow{p} \theta$
- **ECM** $\left(\widehat{\theta}_n\right) = \mathbb{E}\left[\left(\widehat{\theta}_n - \theta\right)^2\right]$ **bajo**

¿Cómo se relacionan estos criterios?

Teorema

$$\textit{Si } ECM\left(\widehat{\theta}_n\right) \underset{n\to\infty}{\longrightarrow} 0 \quad \Rightarrow \quad \widehat{\theta}_n \xrightarrow{p} \theta$$

Demostración en Wasserman (fuera del alcance del curso).

Volviendo al Ejemplo

Teníamos $X_1, \ldots, X_n \sim \mathcal{U}(0, \theta)$ i.i.d.

Estimadores de θ: $\widehat{\theta}_n = 2\overline{X}_n$ y $\widetilde{\theta}_n = \max\limits_{1 \leq i \leq n} \{X_i\}$.

Ambos son consistentes, ¿cuál elegimos?

¡El de menor ECM!

Ejercicios de la práctica que pueden hacer

Práctica 3: Ejercicios 3 al 5 y 7 al 13.