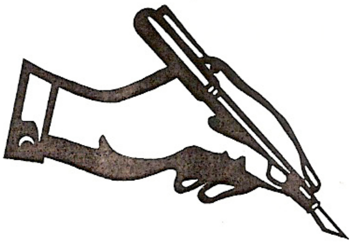
Estadística Teórica II



MUESTREO ALEATORIO SIMPLE

MUESTREO ALEATORIO SIMPLE. DISTRIBUCIÓN EN EL MUESTREO

1.- Con el objetivo de analizar el rendimiento académico de una promoción de licenciados universitarios se lleva a cabo un estudio en el que se emplea una m.a.s. de 3 licenciados. La variable que mide el rendimiento puede tomar tres valores según la calificación final obtenida:

1 – Aprobado
2 – Notable
3 – Sobresaliente

Por otra parte, en esa promoción hubo un total de 20 aprobados, 40 notables y 140 sobresalientes.

a) Hallar las distintas muestras que pueden extraerse y la probabilidad de obtención que tiene cada una de ellas.
b) Calcular le media de cada muestra, así como la distribución de probabilidad en el muestreo de la media.
c) Hacer lo mismo que en el apartado anterior con las varianzas.
d) Calcular la media y la varianza muestral y compararlas con la media y la varianza poblacionales.
e) Calcular la esperanza de la varianza muestral y compararla con la varianza poblacional.

a)

		1	**(1,1,1) , (1,1,2) , (1,1,3)**
	1	2	
		3	
		1	
1	2	2	(1,2,1) , (1,2,2) , **(1,2,3)**
		3	
		1	
	3	2	
		3	(1,3,1) , (1,3,2) , (1,3,3)
		1	(2,1,1) , (2,1,2) , (2,1,3)
	1	2	
		3	
		1	
2	2	2	**(2,2,1) , (2,2,2) , (2,2,3)**
		3	
		1	
	3	2	
		3	(2,3,1) , (2,3,2) , (2,3,3)
		1	(3,1,1) , (3,1,2) , (3,1,3)
	1	2	
		3	

2	1 2 3	(3,2,1) , (3,2,2) , (3,2,3)
3	1 2 3	**(3,3,1) , (3,3,2) , (3,3,3)**

Número muestras distintas	
(1,1,1)	1
(1,1,2) , (1,2,1) , (2,1,1)	3
(1,1,3) , (1,3,1) , (3,1,1)	3
(1,2,3) , (1,3,2) , (2,1,3) , (2,3,1) , (3,1,2) , (3,2,1)	6
(2,2,1) , (2,1,2) , (1,2,2)	3
(2,2,2)	1
(2,2,3) , (2,3,2) , (3,2,2)	3
(3,3,1) , (3,1,3) , (1,3,3)	3
(3,3,2) , (3,2,3) , (2,3,3)	3
(3,3,3)	1
	27

$$p(1) = \frac{20}{200} = 0{,}1$$

$$p(2) = \frac{40}{200} = 0{,}2$$

$$p(3) = \frac{140}{200} = 0{,}7$$

Muestras posibles	Número muestras	Probabilidad	$\overline{x}_i$
(1,1,1)	1	$1\,(0{,}1)^3 = 0{,}001$	1
(1,1,2)	3	$3.\,(0{,}1)^2\,(0{,}2) = 0{,}006$	4/3
(1,1,3)	3	$3.\,(0{,}1)^2\,(0{,}7) = 0{,}021$	5/3
(1,2,3)	6	$6.\,(0{,}1)\,(0{,}2)\,(0{,}7) = 0{,}084$	2
(2,2,1)	3	$3.\,(0{,}2)^2\,(0{,}1) = 0{,}012$	5/3
(2,2,2)	1	$1\,(0{,}2)^3 = 0{,}008$	2
(2,2,3)	3	$3.\,(0{,}2)^2\,(0{,}7) = 0{,}084$	7/3
(3,3,1)	3	$3.\,(0{,}7)^2\,(0{,}1) = 0{,}147$	7/3
(3,3,2)	3	$3.\,(0{,}7)^2\,(0{,}2) = 0{,}294$	8/3
(3,3,3)	1	$1\,(0{,}7)^3 = 0{,}343$	3

b) La distribución de probabilidad en el muestreo para la media:

$\overline{x}_i$	$P(\overline{x} = \overline{x}_i)$
1	$P(\overline{x} = 1) = 0{,}001$
4/3	$P(\overline{x} = 4/3) = 0{,}006$
5/3	$P(\overline{x} = 5/3) = 0{,}021 + 0{,}012 = 0{,}033$
2	$P(\overline{x} = 2) = 0{,}084 + 0{,}008 = 0{,}092$
7/3	$P(\overline{x} = 7/3) = 0{,}084 + 0{,}147 = 0{,}231$
8/3	$P(\overline{x} = 8/3) = 0{,}294$
1	$P(\overline{x} = 3) = 0{,}343$

c) La varianza de cada muestra y la distribución en el muestreo de la varianza muestral:

Muestras posibles	$\overline{x}_i = \sum x_i / n$	$\sum x_i^2 / n$	$\sigma^2_{\overline{x}_i} = \sum x_i^2 / n - \overline{x}^2$	Probabilidad
(1,1,1)	1	1	1 – 1 = 0	$1.\ (0,1)^3 = 0{,}001$
(1,1,2)	4/3	6/3	$6/3 - (4/3)^2 = 2/9$	$3.\ (0,1)^2\ (0,2) = 0{,}006$
(1,1,3)	5/3	11/3	$11/3 - (5/3)^2 = 8/9$	$3.\ (0,1)^2\ (0,7) = 0{,}021$
(1,2,3)	2	14/3	$14/3 - (2)^2 = 2/3$	6. (0,1) (0,2) (0,7) = 0,084
(2,2,1)	5/3	9/3	$9/3 - (5/3)^2 = 2/9$	$3.\ (0,2)^2\ (0,1) = 0{,}012$
(2,2,2)	2	12/3	$12/3 - (2)^2 = 0$	$1.\ (0,2)^3 = 0{,}008$
(2,2,3)	7/3	17/3	$17/3 - (7/3)^2 = 2/9$	$3.\ (0,2)^2\ (0,7) = 0{,}084$
(3,3,1)	7/3	19/3	$19/3 - (7/3)^2 = 8/9$	$3.\ (0,7)^2\ (0,1) = 0{,}147$
(3,3,2)	8/3	22/3	$22/3 - (8/3)^2 = 2/9$	$3.\ (0,7)^2\ (0,2) = 0{,}294$
(3,3,3)	3	27/3	$27/3 - (3)^2 = 0$	$1.\ (0,7)^3 = 0{,}343$

La distribución de probabilidad de la varianza muestral:

$\sigma^2_{\overline{x}_i}$	$P(\sigma^2_{\overline{x}} = \sigma^2_{\overline{x}_i})$
0	$P\left(\sigma^2_{\overline{x}} = 0\right) = 0{,}001 + 0{,}008 + 0{,}343 = 0{,}352$
2/9	$P\left(\sigma^2_{\overline{x}} = 2/9\right) = 0{,}006 + 0{,}012 + 0{,}084 + 0{,}294 = 0{,}396$
2/3	$P\left(\sigma^2_{\overline{x}} = 2/3\right) = 0{,}084$
8/9	$P\left(\sigma^2_{\overline{x}} = 8/9\right) = 0{,}021 + 0{,}147 = 0{,}168$

d) La media y la varianza de la media muestral y compararlas con la media y la varianza poblacionales:

$\overline{x}_i$	$P(\overline{x} = \overline{x}_i)$	$\overline{x}_i \,.\, P(\overline{x} = \overline{x}_i)$	$\overline{x}_i^2$	$\overline{x}_i^2 \,.\, P(\overline{x} = \overline{x}_i)$
1	0,001	0,001	1	0,001
4/3	0,006	4/3 . 0,006	16/9	16/9 . 0,006
5/3	0,033	5/3 . 0,033	25/9	25/9 . 0,033
2	0,092	2 . 0,092	4	4 . 0,092
7/3	0,231	7/3 . 0,231	49/9	49/9 . 0,231
8/3	0,294	8/3 . 0,294	64/9	64/9 . 0,294
3	0,343	3 . 0,343	9	9 . 0,343
		$\sum \overline{x}_i \,.\, P(\overline{x} = \overline{x}_i) = 2{,}6$		$\sum \overline{x}_i^2 \,.\, P(\overline{x} = \overline{x}_i) = 6{,}9067$

$$\left.\begin{aligned} E(\overline{x}) &= \sum \overline{x}_i \,.\, P(\overline{x} = \overline{x}_i) = 2{,}6 \\ E(\overline{x}^2) &= \sum \overline{x}_i^2 \,.\, P(\overline{x} = \overline{x}_i) = 6{,}9067 \\ V(\overline{x}) &= E(\overline{x}^2) - \left(E(\overline{x})\right)^2 = 6{,}9067 - (2{,}6)^2 = 0{,}1467 \end{aligned}\right\} \text{muestra}$$

$$\left.\begin{aligned} \mu &= \sum \overline{x}_i \,.\, P(\overline{x} = \overline{x}_i) = 1.0{,}1 + 2.0{,}2 + 3.0{,}7 = 2{,}6 \\ E(\overline{x}^2) &= \sum \overline{x}_i^2 \,.\, P(\overline{x} = \overline{x}_i) = 1^2.0{,}1 + 2^2.0{,}2 + 3^2.0{,}7 = 7{,}2 \\ \sigma^2 &= E(\overline{x}^2) - \mu^2 = 7{,}2 - (2{,}6)^2 = 0{,}44 \end{aligned}\right\} \text{población}$$

En consecuencia: $\begin{array}{l} \mu = E(\overline{x}) \\ \sigma^2 \neq V(\overline{x}) = \sigma^2_{\overline{x}} \end{array}$ Obsérvese que, $\sigma^2_{\overline{x}} = \dfrac{\sigma^2}{n} \mapsto 0{,}1467 = \dfrac{0{,}44}{3}$

e) Calcular la esperanza de la varianza muestral y compararla con la varianza poblacional.

$\sigma^2_{\overline{x}_i}$	$P(\sigma^2_{\overline{x}} = \sigma^2_{\overline{x}_i})$	$\sigma^2_{\overline{x}_i} . P(\sigma^2_{\overline{x}} = \sigma^2_{\overline{x}_i})$
0	0,352	0 . 0,352
2/9	0,396	2/9 . 0,352
2/3	0,084	2/3 . 0,352
8/9	0,168	8/9 . 0,352
		$\sum \sigma^2_{\overline{x}_i} . P(\sigma^2_{\overline{x}} = \sigma^2_{\overline{x}_i}) = 0{,}2933$

$E(\sigma^2_{\overline{x}}) = \sum \sigma^2_{\overline{x}_i} . P(\sigma^2_{\overline{x}} = \sigma^2_{\overline{x}_i}) = 0{,}2933$ (esperanza varianza muestral)

$\sigma^2 = E(\overline{x}^2) - \mu^2 = 7{,}2 - (2{,}6)^2 = 0{,}44$ (varianza poblacional)

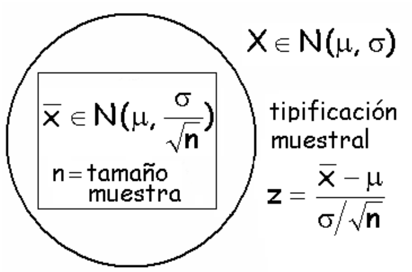
Se verifica la relación: $E(\sigma^2_{\overline{x}}) = 0{,}2933 = \dfrac{(3-1)\,(0{,}44)}{3} = \dfrac{(n-1)\,\sigma^2}{n}$

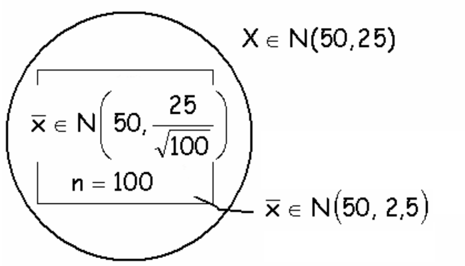
MUESTREO ALEATORIO SIMPLE. DISRIBUCIÓN DE LA MEDIA MUESTRAL CON VARIANZA CONOCIDA.

2.- Se sabe que el peso de los jóvenes entre 14 y 18 años sigue una distribución normal con media 50 kg y desviación típica 25 kg. Para llevar a cabo un estudio del control de peso se seleccionan aleatoriamente 100 jóvenes cuyas edades se encuentran comprendidas en el intervalo señalado. Si el peso medio muestral está entre 45 y 70 kg se considera que están dentro de los límites normales. ¿Cuál es la probabilidad de que el peso esté fuera de control?

Solución:

v. a. X = "peso entre 14 y 18 años"





$$
\begin{aligned}
P[(\bar{x} < 45) \cup (\bar{x} > 70)] &= P(\bar{x} < 45) + P(\bar{x} > 70) = \\
&= P\left(\frac{\bar{x} - 50}{2{,}5} < \frac{45 - 50}{2{,}5}\right) + P\left(\frac{\bar{x} - 50}{2{,}5} > \frac{70 - 50}{2{,}5}\right) = \\
&= P(z < -2) + P(z > 8) = P(z > 2) + P(z > 8) = 0{,}0228 + 0 = \\
&= 0{,}0228
\end{aligned}
$$

MUESTREO ALEATORIO SIMPLE. DISRIBUCIÓN DE LA MEDIA MUESTRAL CON VARIANZA CONOCIDA Y CON VARIANZA DESCONOCIDA.

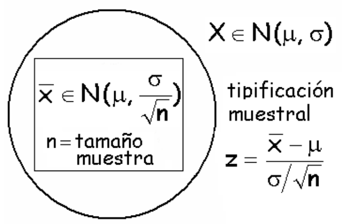
3.- Los barcos que hacen visitas guiadas por el Sena disponen de 60 asientos por barco y una capacidad máxima de 4.200 kg por viaje. Los dueños de la empresa de barcos saben por experiencia que los pesos de los turistas tienen una media de 71 kg y una dispersión, medida a través de la desviación típica, de 10 kg.

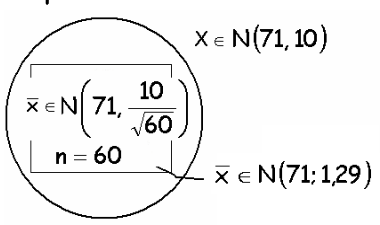
a) ¿Cuál es la probabilidad de que un grupo de 60 turistas, escogidos aleatoriamente en uno de los viajes, tenga un peso medio superior al total de la carga límite permitida?

b) ¿Cuál sería el resultado si la varianza poblacional fuera desconocida?. (Suponga que la desviación típica muestral es de 5 kg).

Solución:

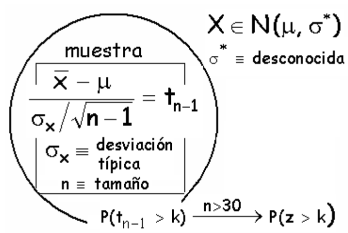
a) Peso = 4200/60 = 70 kg   v. a. X = "peso medio turistas"

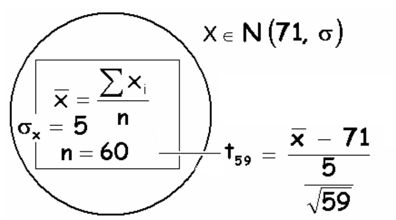




$$P(\bar{x} > 70) = P\left(\frac{\bar{x} - 71}{1,29} > \frac{70 - 71}{1,29}\right) = P(z > -0,77) = P(z < 0,77) =$$
$$= 1 - P(z > 0,77) = 1 - 0,2206 = 0,7794$$

b) En el muestreo de una población normal con varianza desconocida, y desviación típica muestral σ_x, la variable: $\dfrac{\bar{x} - \mu}{\dfrac{\sigma_x}{\sqrt{n-1}}} = t_{n-1}$





$$P(\bar{x} > 70) = P\left(\frac{\bar{x} - 71}{5/\sqrt{59}} > \frac{70 - 71}{5/\sqrt{59}}\right) = P(t_{59} > -1,53) \xrightarrow{n>30} P(z > -1,53) =$$
$$= P(z < 1,53) = 1 - P(z > 1,53) = 1 - 0,063 = 0,937$$

- **Interpolando:** $P(t_{59} > -1,53) = P(t_{59} < 1,53) = 1 - P(t_{59} > 1,53) = 1 - 0,069 = 0,931$

$$P(t_{59} > -1,53) = P(t_{59} < 1,53) = 1 - P(t_{59} > 1,53) = 1 - 0,069 = 0,931$$

$P\left(t_{60} > 1,53\right) = x$

	Abscisas	Áreas	Abscisas	Áreas
	1,296 – 1,671	0,1 - 0,05	0,37	0,05
	1,53 - 1,671	x – 0,05	0,14	x – 0,05

$$x = 0,05 + \frac{0,14 \,.\, 0,05}{0,37} = 0,069$$

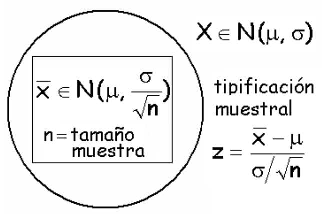
4.- La empresa Grano Sol vende galletas ecológicas en paquetes de 60 unidades. Los dueños saben que el peso de cada galleta es una variable aleatoria que tienen una media de 71 gr. y una dispersión, medida a través de la desviación típica, de 10 gr.

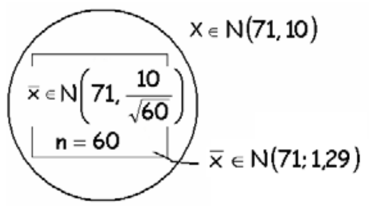
a) ¿Cuál es la probabilidad de que en un paquete de 60 galletas escogidas aleatoriamente, el peso medio de las galletas sea superior a 70 gramos?

b) ¿Cuál sería el resultado si la varianza poblacional fuera desconocida? (Suponga que la desviación típica muestral es de 5 kg, y una cuasidesviación típica de 5,04).

Solución:

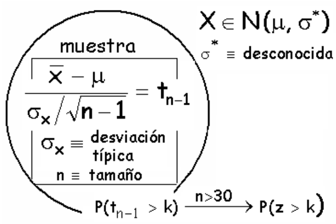
a) v. a. X = "peso de las galletas"

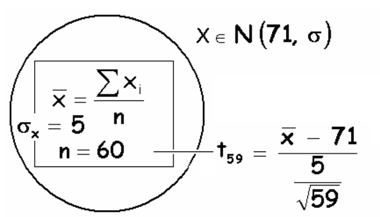




$$P(\bar{x} > 70) = P\left(\frac{\bar{x}-71}{1{,}29} > \frac{70-71}{1{,}29}\right) = P(z > -0{,}77) = P(z < 0{,}77) =$$
$$= 1 - P(z > 0{,}77) = 1 - 0{,}2206 = 0{,}7794$$

b) En el muestreo de una población normal con varianza desconocida, y desviación típica muestral σ_x, la variable: $\dfrac{\bar{x}-\mu}{\dfrac{\sigma_x}{\sqrt{n-1}}} = t_{n-1}$





$$P(\bar{x} > 70) = P\left(\frac{\bar{x}-71}{5/\sqrt{59}} > \frac{70-71}{5/\sqrt{59}}\right) = P(t_{59} > -1{,}53) \xrightarrow{n>30} P(z > -1{,}53) =$$
$$= P(z < 1{,}53) = 1 - P(z > 1{,}53) = 1 - 0{,}063 = 0{,}937$$

- **Interpolando:** $P(t_{59} > -1{,}53) = P(t_{59} < 1{,}53) = 1 - P(t_{59} > 1{,}53) = 1 - 0{,}069 = 0{,}931$

$P(t_{60} > 1{,}53) = x$

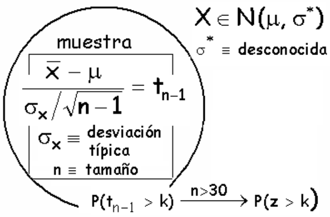
Abscisas	Áreas	Abscisas	Áreas
1,296 - 1,671	0,1 - 0,05	0,37	0,05
1,53 - 1,671	x - 0,05	0,14	x - 0,05

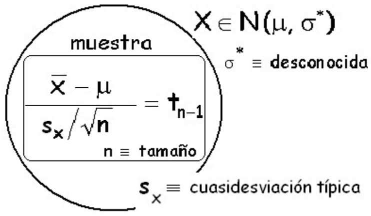
$$x = 0{,}05 + \frac{0{,}14 \cdot 0{,}05}{0{,}37} = 0{,}069$$

Adviértase que, si la varianza poblacional σ^2 es desconocida, la media muestral $\overline{x}$ sigue una t-Student con (n-1) grados de libertad, entonces:

$$\frac{\overline{x}-\mu}{\frac{\sigma_x}{\sqrt{n-1}}} = \frac{\overline{x}-\mu}{\frac{s_x}{\sqrt{n}}} \approx t_{n-1} \quad \text{y es una cantidad pivotal para } \mu$$

recordemos que $n.\sigma^2 = (n-1).s^2$





$$P(\overline{x} > 70) = P\left(\frac{\overline{x}-71}{5,04/\sqrt{60}} > \frac{70-71}{5,04/\sqrt{60}}\right) = P(t_{59} > -1,5369) = P(t_{59} < 1,5369) =$$
$$= 1 - P(t_{59} > 1,5369) = 1 - x = 1 - 0,06789 = 0,09321$$

$P(t_{60} > 1,5369) = x$

Abscisas	Áreas	Abscisas	Áreas
1,2961 – 1,6711	0,1 - 0,05	0,375	0,05
1,5369 - 1,6711	x - 0,05	0,1342	x - 0,05

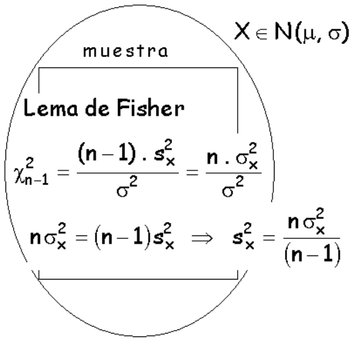
$$x = 0,05 + \frac{0,1342 \, . \, 0,05}{0,375} = 0,06789$$

MUESTREO ALEATORIO SIMPLE. DISRIBUCIÓN DE LA VARIANZA MUESTRAL.

5.- Se sabe por los datos censales que la variabilidad de la altura de alumnos de una clase medida a través de la varianza es de 15,3. No obstante, para estudiar la variabilidad en el muestreo de la varianza muestral se decide tomar una m.a.s. de 15 alumnos. ¿Cuál es la probabilidad de que la varianza muestral sea mayor que 15? Nota: Suponer que la estatura es una variable aleatoria normalmente distribuida.

Solución:

Para el análisis de la varianza muestral se utiliza el estadístico χ^2_{n-1} de Pearson con (n - 1) grados de libertad.



$\sigma^2 \equiv$ varianza poblacional

$\sigma_x^2 \equiv$ varianza muestral

$s_x^2 \equiv$ cuasivarianza muestral

v.a. X ="estatura": $X \in N\left(\mu;\sqrt{15,3}\right) \equiv N(\mu;\, 3,91)$

$$\chi^2_{n-1} \approx \frac{n\,.\,\sigma_x^2}{\sigma^2} \xrightarrow{n=15} \chi^2_{14} \approx \frac{15\,.\,\sigma_x^2}{15,3}$$

$$P\left(\sigma_x^2 > 15\right) = P\left(\frac{15\,.\,\sigma_x^2}{15,3} > \frac{15\,.\,15}{15,3}\right) = P\left(\chi^2_{14} > 14,7\right) = 0,4835$$

$$P\left(\chi^2_{14} > 14,7\right) = x$$

Abscisas	Áreas	Abscisas	Áreas
7,790 – 21,064	0,90 - 0,10	13,274	0,80
14,7 - 21,064	x – 0,10	6,364	x – 0,10

$$x = 0,10 + \frac{6,364\,.\,0,80}{13,274} = 0,4835$$

De otra parte, $\chi^2_{n-1} \approx \frac{(n-1)\,.\,s_x^2}{\sigma^2} \xrightarrow{n=15} \chi^2_{14} \approx \frac{14\,.\,s_x^2}{15,3}$

$$P\left(s_x^2 > 15\right) = P\left(\frac{14\,.\,s_x^2}{15,3} > \frac{14\,.\,15}{15,3}\right) = P\left(\chi^2_{14} > 13,725\right) = 0,5423$$

MUESTREO ALEATORIO SIMPLE. DISRIBUCIÓN DE LA DIFERENCIA DE MEDIAS MUESTRALES CON VARIANZAS CONOCIDAS.

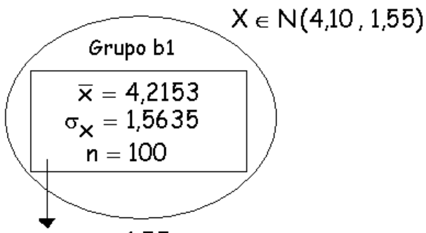
6.- Se desea analizar las diferencias de las calificaciones entre dos grupos de alumnos. Unos proceden del Grupo b1 y otros del Grupo b2. Para estudiar la distribución en el muestro de la diferencia de medias se toman m.a.s. independientes de ambas poblaciones obteniéndose la siguiente tabla:

	Grupo b1	Grupo b2
Tamaño de la población	200	150
Tamaño de la muestra	100	75
Media de la población	4,10	5,18
Media de la muestra	4,2153	5,3247
Desviación típica de la población	1,55	1,95
Desviación típica de la muestra	1,5635	1,8238

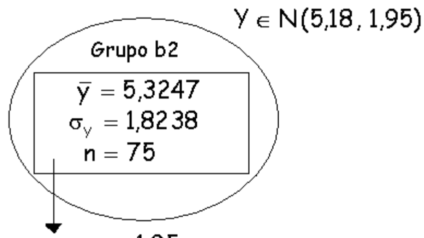
¿Cuál es la probabilidad de que la diferencia de medias muestrales sea mayor que uno?

Solución:

- v.a. X ="calificación del Grupo b1" $X \in N(4{,}10\,,\,1{,}55)$
- v.a. Y ="calificación del Grupo b2" $Y \in N(5{,}18\,,\,1{,}95)$



$\bar{x} \in N(4{,}1,\ \frac{1{,}55}{\sqrt{100}}) \equiv N(4{,}1\,,\,0{,}155)$



$\bar{y} \in N(5{,}18,\ \frac{1{,}95}{\sqrt{75}}) \equiv N(5{,}18,\ 0{,}225)$

Siendo X e Y independientes, la nueva variable $(X \pm Y)$ sigue también una distribución normal $N(\bar{x} \pm \bar{y}\ ;\ \sqrt{\sigma_x^2 + \sigma_y^2}\ )$

Con lo cual, $(\bar{x} - \bar{y}) \in N\left[(4{,}10 - 5{,}18);\ \sqrt{0{,}155^2 + 0{,}225^2}\right] \equiv N(-1.08\ ;0{,}2732)$

$$
\begin{aligned}
P(|\bar{x} - \bar{y}| > 1) &= P[(\bar{x} - \bar{y}) > 1] + P[(\bar{x} - \bar{y}) < -1] = \\
&= P\left[z > \frac{1 + 1{,}08}{0{,}2732}\right] + P\left[z < \frac{-1 + 1{,}08}{0{,}2732}\right] = P(z > 7{,}61) + P[z < 0{,}2928] = \\
&= 0 + [1 - P(z > 0{,}2928)] = 1 - 0{,}3859 = 0{,}6141
\end{aligned}
$$

o también,

$$P(|\bar{x} - \bar{y}| > 1) = 1 - P(|\bar{x} - \bar{y}| < 1) = 1 - P(-1 < (\bar{x} - \bar{y}) < 1) =$$

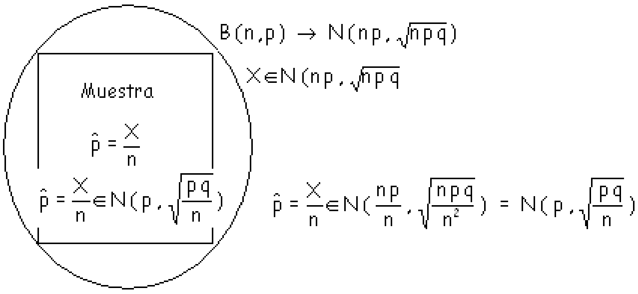
$$= 1 - P\left(\frac{-1+1{,}08}{0{,}2732} < \frac{(\overline{x}\ -\ \overline{y})+1.08}{0{,}2731} < \frac{1+1{,}08}{0{,}2732}\right) = 1 - P\left(0{,}2928 < z < 7{,}61\right) = 0{,}6141$$

MUESTREO ALEATORIO SIMPLE. DISRIBUCIÓN DE LA PROPORCIÓN MUESTRAL

7.- Un concesionario vende dos tipos de vehículos, unos de gama alta y otros de gama media. Los coches de gama alta suponen el 30% del total de los coches vendidos. ¿Cuál es la probabilidad de que entre los 100 últimos vehículos vendidos más del 35% sean de gama alta?

Solución:

La variable poblacional X = 'venta de coches gama alta' es una variable binomial B(100; 0,3), que sigue aproximadamente una distribución normal tal que $X \in N(np;\sqrt{npq})$



$$\hat{p} = \frac{X}{n} \xrightarrow{\text{Teorema Central Límite}} \hat{p} \in N\left(p\,,\sqrt{\frac{pq}{n}}\right)$$

$$\text{para } n = 100, \quad \hat{p} \in N\left(0{,}3\,;\sqrt{\frac{0{,}3.\ 0{,}7}{100}}\right) \equiv N(0{,}3;\ 0{,}0458)$$

$$P(\hat{p} > 0{,}35) = P\left(\frac{\hat{p}-0{,}3}{0{,}0458} > \frac{0{,}35-0{,}3}{0{,}0458}\right) = P(z > 1{,}0917) = 0{,}1375$$

Interpolando:

$$P(z > 1{,}0917) = x$$

Abscisas	Áreas	Abscisas	Áreas
0,1379 – 0,1357	1,09 – 1,1	0,0022	0,01
x - 0,1357	1,0917 – 1,1	x - 0,1357	0,0083

$$x = 0{,}1357 + \frac{0{,}0022\ .\ 0{,}0083}{0{,}01} = 0{,}1375$$

MUESTREO ALEATORIO SIMPLE. DISRIBUCIÓN DE LA DIFERENCIA DE PROPORCIONES MUESTRALES.

8.- Se sabe que los sábados por la noche un 70% de los conductores superan la tasa de alcoholemia permitida por la ley. Sin embargo esta cifra se reduce a un 40% los domingos por la noche. Durante un fin de semana, se quiere realizar un control de alcoholemia y comparar los resultados de los dos días. Se decide elegir al azar 40 vehículos de los que circulan el sábado por la noche y 35 del domingo. Calcular la probabilidad de que la proporción muestral de conductores que superan la tasa de alcoholemia permitida por la ley haya descendido más de un 10% del sábado al domingo.

Solución:

Sean las variables poblacionales:

X = "tasa de alcoholemia sábado", con $p_x = 0{,}7$

Y = "tasa de alcoholemia domingo", con $p_y = 0{,}4$

$$\hat{p}_x = \frac{X}{n_x} \xrightarrow{\text{Teorema Central Límite}} \hat{p}_x \in N\left(p_x\,,\ \sqrt{\frac{p_x\, q_x}{n_x}}\right)$$

$$\hat{p}_y = \frac{Y}{n_y} \xrightarrow{\text{Teorema Central Límite}} \hat{p}_y \in N\left(p_y\,,\ \sqrt{\frac{p_y\, q_y}{n_y}}\right)$$

$$\left[\hat{p}_x \pm \hat{p}_y\right] \in N\left(\left[p_x \pm p_y\right]\,;\ \sqrt{\left(\frac{p_x\, q_x}{n_x}\right) + \left(\frac{p_y\, q_y}{n_y}\right)}\right)$$

siendo las muestras: $n_x = 40\ ,\ \ n_y = 35$

$$\left[\hat{p}_x - \hat{p}_y\right] \in N\left(\left[0{,}7 - 0{,}4\right]\,;\ \sqrt{\left(\frac{0{,}7\,.\,0{,}3}{40}\right) + \left(\frac{0{,}4\,.\,0{,}6}{35}\right)}\right) \equiv N(0{,}3\ ;\ 0{,}11)$$

$$P(\hat{p}_x - \hat{p}_y > 0{,}1) = P\left(z > \frac{0{,}1 - 0{,}3}{0{,}11}\right) = P\left(z > -1{,}82\right) = 1 - P\left(z > 1{,}82\right) = 1 - 0{,}0344 = 0{,}9656$$

9.- Según los resultados de un estudio exhaustivo de la población un 80% de las mujeres entrevistadas afirman utilizar algún producto cosmético todos los días, mientras que en el caso de los hombres este porcentaje en la actualidad asciende 55%. Una pequeña firma de cosmética se plantea sacar al mercado una crema hidratante de uso específico para hombres, pero antes de crear esa nueva línea de negocio, decide realizar su propia encuesta sobre una pequeña muestra aleatoria: selecciona a 50 mujeres y a 60 hombres y les pregunta sobre sus hábitos cosméticos. Calcule la probabilidad de que la diferencia entre la proporción de mujeres que utiliza cosméticos respecto a la proporción de hombres que los utiliza sea inferior al 20%.

<u>Solución:</u>

Sean las variables poblacionales:

X = "mujeres utilizan algún producto cosmético", con $p_x = 0{,}8$
Y = "hombres utilizan algún producto cosmético", con $p_y = 0{,}55$

$$\hat{p}_x = \frac{X}{n_x} \xrightarrow{\text{Teorema Central Límite}} \hat{p}_x \in N\left(p_x\,,\ \sqrt{\frac{p_x\,q_x}{n_x}}\right)$$

$$\hat{p}_y = \frac{Y}{n_y} \xrightarrow{\text{Teorema Central Límite}} \hat{p}_y \in N\left(p_y\,,\ \sqrt{\frac{p_y\,q_y}{n_y}}\right)$$

$$\left[\hat{p}_x \pm \hat{p}_y\right] \in N\left(\left[p_x \pm p_y\right]\,;\ \sqrt{\left(\frac{p_x\,q_x}{n_x}\right)+\left(\frac{p_y\,q_y}{n_y}\right)}\right)$$ siendo las muestras: $n_x = 50\,,\ n_y = 60$

$$\left[\hat{p}_x - \hat{p}_y\right] \approx N\left(\left[0{,}8 - 0{,}55\right]\,;\ \sqrt{\left(\frac{0{,}8\,.\,0{,}2}{50}\right)+\left(\frac{0{,}55\,.\,0{,}45}{60}\right)}\right) \equiv N(0{,}25\,;\,0{,}0856)$$

$$P\left[(\hat{p}_x - \hat{p}_y) < 0{,}20\right] = P\left(z < \frac{0{,}2 - 0{,}25}{0{,}0856}\right) = P(z < -\,0{,}58) = P(z > 0{,}58) = 0{,}2810$$

Como la $P\left[(\hat{p}_x - \hat{p}_y) > 0{,}20\right] = 1 - 0{,}2810 = 0{,}719$ es bastante probable, se aconsejaría sacar el producto del mercado.

CÁLCULO DE PROPIEDADES BÁSICAS DE LOS ESTIMADORES (INSESGADEZ y EFICIENCIA)

10.- La variable aleatoria poblacional "renta de las familias" del municipio de Madrid se distribuye siguiendo un modelo $N(\mu, \sigma^2)$. Se extraen muestras aleatorias simples de tamaño 4. Como estimadores del parámetro μ, se proponen los siguientes:

$$\hat{\mu}_1 = \frac{x_1 + 2x_2 + 3x_3}{6}$$
$$\hat{\mu}_2 = \frac{x_3 - 4x_2}{-3}$$
$$\hat{\mu}_3 = \overline{x}$$

Se pide:

a) Comprobar si los estimadores son insesgados
b) ¿Cuál es el más eficiente?
c) Si tuviera que escoger entre ellos, ¿cuál escogería?. Razone su respuesta a partir del Error Cuadrático Medio.

Solución:

a) Un estimador $\hat{\theta}$ es insesgado (o centrado) cuando se verifica $E(\hat{\theta}) = \theta$

$$E(\hat{\mu}_1) = E\left[\frac{x_1 + 2x_2 + 3x_3}{6}\right] = \frac{1}{6} E\left[x_1 + 2x_2 + 3x_3\right] =$$
$$= \frac{1}{6}\left[E(x_1) + 2E(x_2) + 3E(x_3)\right] = \frac{1}{6}\left[6\mu\right] = \mu$$

$$E(\hat{\mu}_2) = E\left[\frac{x_1 - 4x_2}{-3}\right] = -\frac{1}{3} E\left[x_1 - 4x_2\right] = -\frac{1}{3}\left[E(x_1) - 4E(x_2)\right] =$$
$$= -\frac{1}{3}\left[-3\mu\right] = \mu$$

$$E(\hat{\mu}_3) = E\left[\frac{x_1 + x_2 + x_3 + x_4}{4}\right] = \frac{1}{4} E\left[x_1 + x_2 + x_3 + x_4\right] =$$
$$= \frac{1}{4}\left[E(x_1) + E(x_2) + E(x_3) + E(x_4)\right] = \frac{1}{4}\left[4\mu\right] = \mu$$

Los tres estimadores son insesgados o centrados.

b) El estimador más EFICIENTE es el que tenga menor varianza.

$$V\left[\hat{\mu}_1\right] = V\left[\frac{x_1 + 2x_2 + 3x_3}{6}\right] = \frac{1}{36} V\left[x_1 + 2x_2 + 3x_3\right] =$$
$$= \frac{1}{36}\left[V(x_1) + 4V(x_2) + 9V(x_3)\right] = \frac{1}{36}\left[14\sigma^2\right] = \frac{14}{36}\sigma^2 = 0{,}39\sigma^2$$

$$V\left[\hat{\mu}_2\right] = V\left[\frac{x_1 - 4x_2}{-3}\right] = \frac{1}{9}\,V\left[x_1 - 4x_2\right] = \frac{1}{9}\left[V(x_1) + 16\,V(x_2)\right] =$$

$$= \frac{1}{9}\left[17\sigma^2\right] = \frac{17}{9}\sigma^2 = 1{,}89\,\sigma^2$$

$$V\left[\hat{\mu}_3\right] = V\left[\frac{x_1 + x_2 + x_3 + x_4}{4}\right] = \frac{1}{16}\,V\left[x_1 + x_2 + x_3 + x_4\right] =$$

$$= \frac{1}{16}\left[V(x_1) + V(x_2) + V(x_3) + V(x_4)\right] = \frac{1}{16}\left[4\sigma^2\right] = \frac{4}{16}\sigma^2 = 0{,}25\,\sigma^2$$

El estimador $\hat{\mu}_3$ es el más eficiente.

c) Escogería el estimador que presentase menor Error Cuadrático Medio (ECM)

$$ECM(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = V(\hat{\theta}) + \left[\underbrace{E(\hat{\theta}) - \theta}_{\text{sesgo}}\right]^2 \qquad \text{sesgo } b(\hat{\theta}) = \left[E(\hat{\theta}) - \theta\right]$$

$$\text{Si } \underbrace{E(\hat{\theta}) = \theta}_{\text{insesgado}} \Rightarrow ECM(\hat{\theta}) = V(\hat{\theta})$$

Como los tres estimadores son insesgados (centrados), me decido por el que menor varianza presenta, puesto que coincidirá con el que menor ECM tiene, es decir, escojo el estimador $\hat{\mu}_3$

Adviértase que si el estimador $\hat{\theta}$ es insesgado: $ECM(\hat{\theta}) = V(\hat{\theta})$

ESTIMADORES SESGADOS: CÁLCULO SESGO Y ESTIMACIÓN PUNTUAL

11.- La variable aleatoria X representa los gastos mensuales de una empresa, cuya función de densidad es $f(\theta, x) = \theta x^{\theta-1}$ con $\theta > 0$ y $0 < x < 1$. Se realiza una m.a.s. de tamaño 3, y se proponen tres estimadores:

$$\hat{\theta}_1 = \bar{x}$$

$$\hat{\theta}_2 = \frac{x_1^2 + 2x_2^2 + 3x_3^2}{6}$$

$$\hat{\theta}_3 = \frac{x_3 - 2x_1 + 4x_2}{6}$$

a) Calcule los sesgos
b) Si la muestra que se obtiene es (0,7 ; 0,1 ; 0,3), calcule las estimaciones puntuales.
c) ¿Cuáles son las funciones estimadas para las estimaciones anteriores?

Solución:

Un estimador $\hat{\theta}$ es insesgado (centrado) cuando $E(\hat{\theta}) = \theta$.
Un estimador $\hat{\theta}$ es sesgado cuando $E(\hat{\theta}) = \theta + \underbrace{b(\hat{\theta})}_{\text{sesgo}} \Rightarrow b(\hat{\theta}) = E(\hat{\theta}) - \theta$

X = "gastos mensuales de la empresa"
$f(\theta, x) = \theta x^{\theta-1}$ con $\theta > 0$ y $0 < x < 1$ m.a.s. con n = 3

- Sesgo del estimador $\hat{\theta}_1 = \bar{x}$

$$\hat{\theta}_1 = \bar{x} \Rightarrow E(\hat{\theta}_1) = E\left[\frac{x_1 + x_2 + x_3}{3}\right] = \frac{1}{3} E\left[x_1 + x_2 + x_3\right] = \frac{1}{3}(3\mu) = \mu \text{ (media poblacional)}$$

donde $\mu = \int_{-\infty}^{\infty} x\, f(x, \theta)\, dx = \int_0^1 x\, f(x, \theta)\, dx = \int_0^1 x\, \theta x^{\theta-1}\, dx = \int_0^1 \theta\, x^{\theta}\, dx = \left[\frac{\theta\, x^{\theta+1}}{\theta+1}\right]_0^1 = \frac{\theta}{\theta+1}$

El sesgo: $b(\hat{\theta}_1) = E(\hat{\theta}_1) - \theta = \dfrac{\theta}{\theta+1} - \theta = -\dfrac{\theta^2}{\theta+1}$

- Sesgo del estimador $\hat{\theta}_2 = \dfrac{x_1^2 + 2x_2^2 + 3x_3^2}{6}$

$$E(\hat{\theta}_2) = E\left[\frac{x_1^2 + 2x_2^2 + 3x_3^2}{6}\right] = \frac{1}{6}\left[\underbrace{E(x_1^2)}_{\alpha_2} + 2\underbrace{E(x_2^2)}_{\alpha_2} + 3\underbrace{E(x_3^2)}_{\alpha_2}\right] = \frac{1}{6}(6\alpha_2) = \alpha_2 \quad (*)$$

donde α_2 es el momento de orden 2 respecto al origen.

$$\alpha_2 = E(x^2) = \int_{-\infty}^{\infty} x^2 \, f(x,\theta)\, dx = \int_0^1 x^2 \, f(x,\theta)\, dx = \int_0^1 x^2 \, \theta \, x^{\theta-1}\, dx = \int_0^1 \theta \, x^{\theta+1}\, dx =$$

$$= \left[\frac{\theta \, x^{\theta+2}}{\theta+2}\right]_0^1 = \frac{\theta}{\theta+2}$$

entonces,

$$E(\hat{\theta}_2) = E\left[\frac{x_1^2 + 2x_2^2 + 3x_3^2}{6}\right] = \frac{1}{6}\left[\underbrace{E(x_1^2)}_{\alpha_2} + 2\underbrace{E(x_2^2)}_{\alpha_2} + 3\underbrace{E(x_3^2)}_{\alpha_2}\right] = \alpha_2 = \frac{\theta}{\theta+2}$$

El sesgo: $b(\hat{\theta}_2) = E(\hat{\theta}_2) - \theta = \dfrac{\theta}{\theta+2} - \theta = -\dfrac{\theta^2+\theta}{\theta+2}$

- Sesgo del estimador $\hat{\theta}_3 = \dfrac{x_3 - 2x_1 + 4x_2}{6}$

$$E(\hat{\theta}_3) = E\left[\frac{x_3 - 2x_1 + 4x_2}{6}\right] = \frac{1}{6}\, E\left[x_3 - 2x_1 + 4x_2\right] = \frac{1}{6}\,(3\mu) = \frac{1}{2}\,\mu$$

$$\mu = \int_{-\infty}^{\infty} x \, f(x,\theta)\, dx = \int_0^1 x \, f(x,\theta)\, dx = \int_0^1 x \, \theta x^{\theta-1}\, dx = \int_0^1 \theta \, x^{\theta}\, dx = \left[\frac{\theta \, x^{\theta+1}}{\theta+1}\right]_0^1 = \frac{\theta}{\theta+1}$$

El sesgo: $b(\hat{\theta}_3) = E(\hat{\theta}_3) - \theta = \dfrac{1}{2}\left(\dfrac{\theta}{\theta+1}\right) - \theta = -\dfrac{2\theta^2+\theta}{2(\theta+1)}$

b) Si la muestra que se obtiene es (0,7 ; 0,1 ; 0,3), calcule las estimaciones puntuales.

$\hat{\theta}_1 = \dfrac{0,7 + 0,1 + 0,3}{3} = 0,367$

$\hat{\theta}_2 = \dfrac{0,7^2 + 2.0,1^2 + 3.0,3^2}{6} = 0,13$

$\hat{\theta}_3 = \dfrac{0,3 - 2.0,7 + 4.0,1}{6} = -\,0,117 \;\mapsto$ no puede ser, puesto que $\hat{\theta} > 0$

c) ¿Cuáles son las funciones estimadas para las estimaciones anteriores?

$\hat{\theta}_1 \Rightarrow f(0,367,\, x) = 0,367\; x^{0,367-1} = 0,367\; x^{-0,633}$

$\hat{\theta}_2 \Rightarrow f(0,13,\, x) = 0,13\; x^{0,13-1} = 0,367\; x^{-0,87}$

CÁLCULO EFICIENCIA RELATIVA Y ERROR CUÁDRATICO MEDIO

12.- Sea una población con media μ de la que se extraen m.a.s. de tamaño n. Considere los siguientes estimadores de la media:

$$\hat{\mu}_1 = \overline{x} \qquad\qquad \hat{\mu}_2 = \frac{1}{n+1}\sum_{i=1}^{n} x_i$$

a) Estudie la insesgadez, la eficiencia relativa y la consistencia de ambos estimadores.
b) Elija uno de los dos en término del error cuadrático medio.

Solución:

a) Insesgadez

Un estimador $\hat{\theta}$ es insesgado (o centrado) cuando se verifica $E(\hat{\theta}) = \theta$

Un estimador $\hat{\theta}$ es sesgado cuando $E(\hat{\theta}) = \theta + \underbrace{b(\hat{\theta})}_{\text{sesgo}} \Rightarrow \underbrace{b(\hat{\theta})}_{\text{sesgo}} = E(\hat{\theta}) - \theta$

Un estimador $\hat{\theta}$ es *asintóticamente insesgado* si su posible sesgo tiende a cero al aumentar el tamaño muestral que se calcula: $\lim\limits_{n\to\infty} b(\tilde{\theta}) = 0$

$$E(\hat{\mu}_1) = E(\overline{x}) = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n} E\left(\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n}(n\mu) = \mu$$

$$b(\hat{\mu}_1) = E(\hat{\mu}_1) - \mu = \mu - \mu = 0$$

$$E(\hat{\mu}_2) = E\left(\frac{1}{n+1}\sum_{i=1}^{n} x_i\right) = \frac{1}{n+1} E\left(\sum_{i=1}^{n} x_i\right) = \frac{1}{n+1}\sum_{i=1}^{n} E(x_i) = \frac{1}{n+1}(n\mu) = \frac{n\mu}{n+1}$$

$$b(\hat{\mu}_2) = E(\hat{\mu}_2) - \mu = \frac{n\mu}{n+1} - \mu = \frac{n\mu - n\mu - \mu}{n+1} = \overbrace{\underbrace{-\frac{\mu}{n+1}}_{\text{sesgado asintoticamente}}}^{\to 0 \text{ cuando 'n' aumenta}}$$

- Eficiencia

Sean $\hat{\theta}_1$ y $\hat{\theta}_2$ dos estimadores insesgados de un parámetro desconocido θ.

Decimos que $\hat{\theta}_1$ es *más eficiente* que $\hat{\theta}_2$ si se verifica que $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$

La *eficiencia relativa* se mide por el ratio: $\dfrac{Var(\hat{\theta}_1)}{Var(\hat{\theta}_2)}$

$$V(\hat{\mu}_1) = V(\overline{x}) = V\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(x_i) = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}$$

$$V(\hat{\mu}_2) = V\left(\frac{1}{n+1}\sum_{i=1}^{n} x_i\right) = \frac{1}{(n+1)^2}\sum_{i=1}^{n} V(x_i) = \frac{1}{(n+1)^2}(n\sigma^2) = \frac{n}{(n+1)^2}\sigma^2$$

$$\text{eficiencia relativa} \equiv \frac{Var(\hat{\mu}_1)}{Var(\hat{\mu}_2)} = \frac{\sigma^2/n}{n\sigma^2/(n+1)^2} = \frac{(n+1)^2}{n^2} > 1 \mapsto Var(\hat{\mu}_1) > Var(\hat{\mu}_2)$$

El estimador $\hat{\mu}_2$ tiene menor varianza, por lo que es ***más eficiente*** que $\hat{\mu}_1$

- Consistencia

Un estimador $\hat{\theta}$ *consistente* es un estimador asintóticamente insesgado cuya varianza tiende a cero al aumentar el tamaño muestral.

El estimador $\hat{\theta}$ es *consistente* cuando $\begin{cases} \lim\limits_{n\to\infty} E(\hat{\theta}) = \theta \\ \lim\limits_{n\to\infty} V(\hat{\theta}) = 0 \end{cases}$

$$\hat{\mu}_1 \equiv \begin{cases} \lim\limits_{n\to\infty} E(\hat{\mu}_1) = \lim\limits_{n\to\infty} E(\overline{x}) = \mu \\ \lim\limits_{n\to\infty} V(\hat{\mu}_1) = \lim\limits_{n\to\infty} \dfrac{\sigma^2}{n} = 0 \end{cases} \qquad \textit{es consistente}$$

$$\hat{\mu}_2 \equiv \begin{cases} \lim\limits_{n\to\infty} E(\hat{\mu}_2) = \lim\limits_{n\to\infty} \left(\mu - \dfrac{1}{n+1}\,\mu\right) = \mu \\ \lim\limits_{n\to\infty} V(\hat{\mu}_2) = \lim\limits_{n\to\infty} \left[\dfrac{n}{(n+1)^2}\,\sigma^2\right] = 0 \end{cases} \qquad \textit{es consistente}$$

c) Elegir uno de los dos en término del error cuadrático medio.

El Error Cuadrático Medio (ECM) de un estimador $\hat{\theta}$ viene definido:

$$ECM(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = V(\hat{\theta}) + \Big[\underbrace{E(\hat{\theta}) - \theta}_{\text{sesgo}}\Big]^2 \qquad \text{sesgo } b(\hat{\theta}) = E(\hat{\theta}) - \theta$$

Si $\underbrace{E(\hat{\theta}) = \theta}_{\text{insesgado}} \Rightarrow ECM(\hat{\theta}) = V(\hat{\theta})$

$$ECM(\hat{\mu}_1) = V(\hat{\mu}_1) + \left[b(\hat{\mu}_1)\right]^2 = \frac{\sigma^2}{n} + 0 = \frac{\sigma^2}{n}$$

$$ECM(\hat{\mu}_2) = V(\hat{\mu}_2) + \left[b(\hat{\mu}_2)\right]^2 = \frac{n}{(n+1)^2}\,\sigma^2 + \left(\frac{1}{n+1}\,\mu\right)^2 = \frac{n\sigma^2 + \mu^2}{(n+1)^2}$$

El estimador $\hat{\mu}_1$ será el que presenta menor ECM cuando $\underline{ECM(\hat{\mu}_1) \le ECM(\hat{\mu}_2)}$

En esta línea,

$$\frac{\sigma^2}{n} \le \frac{n\sigma^2 + \mu^2}{(n+1)^2} = \frac{n\sigma^2}{(n+1)^2} + \frac{\mu^2}{(n+1)^2} \Rightarrow \frac{\sigma^2}{n} - \frac{n\sigma^2}{(n+1)^2} \le \frac{\mu^2}{(n+1)^2} \Rightarrow$$

$$\frac{(n+1)^2\sigma^2 - n^2\sigma^2}{n(n+1)^2} \leq \frac{\mu^2}{(n+1)^2} \Rightarrow \frac{(n+1)^2 - n^2}{n}\sigma^2 \leq \mu^2 \Rightarrow$$

$$\Rightarrow \frac{2n+1}{n}\sigma^2 \leq \mu^2 \Rightarrow \frac{2n+1}{n} \leq \frac{\mu^2}{\sigma^2}$$

$$\begin{cases} \text{Si } \dfrac{2n+1}{n} \leq \dfrac{\mu^2}{\sigma^2} & \mapsto \hat{\mu}_1 \text{ se elige antes que } \hat{\mu}_2 \\ \text{Si } \dfrac{2n+1}{n} \geq \dfrac{\mu^2}{\sigma^2} & \mapsto \hat{\mu}_2 \text{ se elige antes que } \hat{\mu}_1 \end{cases}$$

CÁLCULO INSESGADEZ E EFICIENCIA

13.- El peso en kilos de los jamones vendidos por una empresa sigue una distribución normal con varianza 4 y peso medio desconocido. Se conoce que el peso medio de los jamones vendidos es superior a 5 kg, y se toman m.a.s. de tamaño 4 para estimar θ. ¿Cuál de los dos estimadores sería el mejor respondiendo a la insesgadez y eficiencia?

$$\hat{\theta}_1 = \frac{X_1 + X_2 + X_3}{4} \qquad\qquad \hat{\theta}_2 = \frac{X_1 + X_2}{2}$$

Solución:

- Un estimador es insesgado (centrado) si $E(\hat{\theta}) = \theta$
  Un estimador es sesgado si $E(\hat{\theta}) = \theta + b(\hat{\theta}) \quad \mapsto \quad \underbrace{b(\hat{\theta})}_{\text{sesgo}} = E(\hat{\theta}) - \theta$

La v.a X_i ='peso en kg de los jamones' sigue una distribución normal de varianza 4

Para estudiar la insesgadez de los estimadores hallamos sus esperanzas:

- $E(\hat{\theta}_1) = E\left[\frac{X_1 + X_2 + X_3}{4}\right] = \frac{1}{4}\left[E(X_1) + E(X_2) + E(X_2)\right] = \frac{3}{4}\theta$

El sesgo del estimador $\hat{\theta}_1$ será: $b(\hat{\theta}_1) = E(\hat{\theta}_1) - \theta = \frac{3}{4}\theta - \theta = -\frac{1}{4}\theta$

- $E(\hat{\theta}_2) = E\left[\frac{X_1 + X_2}{2}\right] = \frac{1}{2}\left[E(X_1) + E(X_2)\right] = \frac{2}{2}\theta = \theta$

El estimador $\hat{\theta}_2$ es insesgado, $b(\hat{\theta}_2) = 0$

Atendiendo al sesgo se elige $\hat{\theta}_2$

- Para analizar la eficiencia relativa de los dos estimadores se calculan las respectivas varianzas

$$V(\hat{\theta}_1) = V\left[\frac{X_1 + X_2 + X_3}{4}\right] = \frac{1}{16}\left[\underbrace{V(X_1 + X_2 + X_3)}_{\substack{\text{las observaciones}\\ \text{son independientes}}}\right] = \frac{1}{16}\left[V(X_1) + V(X_2) + V(X_2)\right] =$$

$$\overset{V(X_i)=4}{=} \frac{1}{16}\,12 = \frac{12}{16} = \frac{3}{4}$$

$$V(\hat{\theta}_2) = V\left[\frac{X_1 + X_2}{2}\right] = \frac{1}{4}\left[\underbrace{V(X_1 + X_2)}_{\substack{\text{las observaciones}\\ \text{son independientes}}}\right] = \frac{1}{4}\left[V(X_1) + V(X_2)\right] \overset{V(X_i)=4}{=} \frac{1}{4}\,8 = 2$$

Respecto a la varianza se elige el estimador $\hat{\theta}_1$ por ser el de menor varianza.

Tenemos propiedades contrapuestas, de modo que el estimador insesgado $\hat{\theta}_2$ es el de mayor varianza. Elegiremos el estimador en base al error cuadrático medio (ECM):

$$\text{ECM} = \text{Varianza} + (\text{sesgo})^2 \equiv \begin{cases} \text{ECM}(\hat{\theta}_1) \;=\; \dfrac{3}{4} + \left(-\dfrac{\theta}{4}\right)^2 = \dfrac{\theta^2 + 12}{16} \\ \\ \text{ECM}(\hat{\theta}_2) \;=\; 2 + 0 \;=\; 2 \end{cases}$$

Se analiza cuando es mayor el ECM del primer estimador $\hat{\theta}_1$: $\text{ECM}(\hat{\theta}_1) > \text{ECM}(\hat{\theta}_2)$

$$\frac{\theta^2 + 12}{16} > 2 \quad \Rightarrow \quad \theta^2 > 20 \quad \Rightarrow \quad |\theta| > \sqrt{20} \approx 4{,}47$$

Si θ es en valor absoluto mayor que 4,47, el error cuadrático medio de $\hat{\theta}_1$ es mayor, con lo que se elige el estimador $\hat{\theta}_2$.

Como sabemos que el peso medio de los jamones es superior a 5 kg, no queda duda que el estimador a elegir (con menor error cuadrático medio) es $\hat{\theta}_2$.

14.- La distribución del peso de las manzanas de una determinada cosecha sigue una distribución normal, cuyo peso medio es desconocido y cuya desviación típica es 7 gramos. Se pide:

a) Analizar cuál de los estimadores $\hat{\mu}_1$, $\hat{\mu}_2$ del peso medio es mejor respecto del sesgo y de la eficiencia, para una muestra aleatoria simple de tamaño cinco.

b) Si $\hat{\mu}_1 = \dfrac{\sum_{i=1}^{5} X_i}{5}$ y $\hat{\mu}_2 = X_1 + 2X_2 + 3X_3 - 4X_4 - X_5$, obtener los pesos medios estimados a partir de la siguiente muestra (125, 135, 130, 137, 142).

Solución.-

a) El peso de las manzanas sigue una distribución $N(\mu, 7)$

Calculamos las esperanzas de los estimadores para analizar el sesgo de los estimadores

$$E(\hat{\mu}_1) = E\left[\sum_{i=1}^{5} X_i \Big/ 5\right] = \frac{1}{5} E\left[\sum_{i=1}^{5} X_i\right] = \frac{1}{5}\sum_{i=1}^{5} E\left[X_i\right] \overset{E(X_i)=\mu}{=} \frac{1}{5}(5\mu) = \mu$$

$$E(\hat{\mu}_2) = E(X_1 + 2X_2 + 3X_3 - 4X_4 - X_5) = E(X_1) + 2E(X_2) + 3E(X_3) - 4E(X_4) - E(X_5) =$$
$$= \mu + 2\mu + 3\mu - 4\mu - \mu = \mu$$

Los estimadores $\hat{\mu}_1$, $\hat{\mu}_2$ son insesgados (centrados).

b) Para analizar la eficiencia de los estimadores calculamos sus varianzas:

$$V(\hat{\mu}_1) = V\left[\sum_{i=1}^{5} X_i \Big/ 5\right] = \frac{1}{25} V\left[\sum_{i=1}^{5} X_i\right] = \frac{1}{25}\sum_{i=1}^{5} V\left[X_i\right] \overset{V(X_i)=7^2}{=} \frac{1}{25}(5.\,49) = \frac{49}{5}$$

$$V(\hat{\mu}_2) = V(X_1 + 2X_2 + 3X_3 - 4X_4 - X_5) = V(X_1) + 4V(X_2) + 9V(X_3) + 16V(X_4) + V(X_5) =$$
$$= (49) + 4\,(49) + 9\,(49) + 16\,(49) + (49) = 31\,(49) = 1519$$

Como los dos estimadores son insesgados y $V(\hat{\mu}_1) < V(\hat{\mu}_2)$ se elige como mejor el estimador $\hat{\mu}_1$, que es el peso medio de la muestra de las cinco manzanas.

15.- Supongamos que la distribución de ingresos de una cierta población es una variable aleatoria con media μ desconocida y varianza σ^2 también desconocida. Si queremos estimar el ingreso medio de la población mediante una m.a.s. de tamaño n, respecto de la insesgadez y de la eficiencia. ¿Cuál de los dos estimadores elegiríamos?

$$\hat{\mu}_1 = \frac{\sum_{i=1}^{n} X_i}{n-1} \qquad \hat{\mu}_2 = \frac{\sum_{i=1}^{n} X_i}{n}$$

Solución:

- Un estimador es insesgado (centrado) si $E(\hat{\theta}) = \theta$
  Un estimador es sesgado si $E(\hat{\theta}) = \theta + b(\hat{\theta}) \quad \mapsto \quad \underbrace{b(\hat{\theta})}_{\text{sesgo}} = E(\hat{\theta}) - \theta$

La v.a X_i ='ingresos de cierta población'' sigue una distribución normal $N(\mu, \sigma)$

Para analizar el sesgo de los estimadores, hallamos la esperanza:

$$E(\hat{\mu}_1) = E\left(\sum_{i=1}^{n} X_i \Big/ n-1\right) = \frac{1}{n-1} E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n-1} \sum_{i=1}^{n} E(X_i) = \frac{1}{n-1} (n\mu) = \frac{n}{n-1} \mu$$

El sesgo del estimador $\hat{\mu}_1$ será: $b(\hat{\mu}_1) = E(\hat{\mu}_1) - \mu = \dfrac{n}{n-1}\mu - \mu = \dfrac{1}{n-1}\mu$

$$E(\hat{\mu}_2) = E\left(\sum_{i=1}^{n} X_i \Big/ n\right) = \frac{1}{n} E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \frac{1}{n} (n\mu) = \mu$$

El estimador $\hat{\mu}_2$, que es la media muestral, es insesgado (centrado).

- La eficiencia de los estimadores se analiza a través de su varianza:

$$V(\hat{\mu}_1) = V\left(\sum_{i=1}^{n} X_i \Big/ n-1\right) = \frac{1}{(n-1)^2} V\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{(n-1)^2} \sum_{i=1}^{n} V(X_i) = \frac{1}{(n-1)^2} (n\sigma^2) = \frac{n\sigma^2}{(n-1)^2}$$

$$V(\hat{\mu}_2) = V\left(\sum_{i=1}^{n} X_i \Big/ n\right) = \frac{1}{n^2} V\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n} \sum_{i=1}^{n} V(X_i) = \frac{1}{n^2} (n\sigma^2) = \frac{\sigma^2}{n}$$

El estimador más eficiente será el de menor varianza. Comparando las varianzas de los estimadores:

$$V(\hat{\mu}_2) = \frac{\sigma^2}{n} < \frac{n\sigma^2}{(n-1)^2} = V(\hat{\mu}_1) \quad \text{puesto que } (n-1)^2 < n^2$$

El estimador $\hat{\mu}_2$, que es la media muestral, es el mejor tanto al sesgo como a la eficiencia.

COMPRENSIÓN DE LA VEROSIMILITUD

CÁLCULO DE LOS ESTIMADORES MÁXIMO VERSOSÍMILES. PROPIEDADES

16.- Una urna contiene bolas blancas y negras. Sea p la probabilidad de extraer una bola blanca cuando se realiza una extracción al azar. Asociado a este experimento aleatorio tenemos la variable aleatoria X que puede tomar los valores:

X = 1 si la bola extraída es blanca
X = 0 si la bola extraída es negra

La distribución de probabilidad será una B(1; p): $P(X=x) \ = \ p^x\,(1-p)^{1-x}$

Se selecciona una muestra aleatoria con reemplazamiento de tamaño 3 (x_1, x_2, x_3), siendo x_i la variable aleatoria a la extracción i-ésima, y suponemos que ha resultado la siguiente relación (B, N, B). Como el parámetro p es desconocido pretendemos saber, entre los valores, $p = 0{,}65$ y $p = 0{,}73$ qué valor hace más probable la aparición de dicha extracción.

Solución.-

Si la muestra (B, N, B) es independiente, siendo $\begin{cases} P(B) = p \\ P(N) = 1-p \end{cases}$

$$P(B, N, B) \ = \ P(B\cap N\cap B) \ = \ P(B).P(N).P(B) \ = \ p.(1-p).p \ = \ p^2.\,(1-p)$$

$$\text{entonces} \begin{cases} p = 0{,}65: \quad P(B,N,B) \ = \ 0{,}65^2.0{,}35 \ = 0{,}1479 \\ \\ p = 0{,}73: \quad P(B,N,B) \ = \ 0{,}73^2.0{,}27 \ = 0{,}1439 \end{cases}$$

Resulta más probable (p = 0,65), siendo más verosímil.

FUNCIÓN DE VEROSIMILITUD DE LA MUESTRA.- Sea $(X_1, \cdots, X_n)$ una muestra aleatoria de una población X con función de masa (o función de densidad f_θ) donde $\theta = (\theta_1, \ \cdots \ , \theta_n)$. El estimador de máxima verosimilitud de θ es el formado por los valores $(\hat\theta_1, \ \cdots \ , \hat\theta_n)$ que maximizan lo que llamaremos *función de verosimilitud* de la muestra $(x_1, \cdots, x_n)$ obtenida:

$$L(\theta) \ = \ L(x_1, \ \cdots, x_n;\theta) \ = \ \begin{cases} P(x_1,\theta)\cdots P(x_n,\theta) \ \ \text{caso discreto} \\ f_\theta(x_1)\cdots f_\theta(x_n) \ \ \text{caso continuo} \end{cases}$$

Si consideramos la m.a.s. (x_1, x_2, x_3), siendo las variables aleatorias x_i independientes, tomando los valores 0, 1, con distribución B(1, p), la distribución de probabilidad asociada será:

$$\left.\begin{array}{l} P(x_1, p) = P(X = x_1) = p^{x_1}(1-p)^{1-x_1} \\ P(x_2, p) = P(X = x_2) = p^{x_2}(1-p)^{1-x_2} \\ P(x_3, p) = P(X = x_3) = p^{x_3}(1-p)^{1-x_3} \end{array}\right\} \; x_i = 1, 0 \;\; \text{sea bola blanca o negra}$$

La función de verosimilitud será:

$$L(p) = \prod_{i=1}^{3} P(x_i, p) = p^{x_1}(1-p)^{1-x_1} \,.\, p^{x_2}(1-p)^{1-x_2} \,.\, p^{x_3}(1-p)^{1-x_3} =$$

$$= p^{x_1+x_2+x_3}(1-p)^{3-(x_1+x_2+x_3)}$$

En la muestra (B, N, B) el valor que toma la función de verosimilitud será:

$$L(p) = p^{1+0+1}(1-p)^{3-(1+0+1)} = p^2 .\, (1-p)$$

17.- Un atleta olímpico de salto de altura se enfrenta a un listón de 2,3 metros. Su entrenador desea estudiar el comportamiento del saltador. Sabe que el número de saltos fallidos por hora es una variable aleatoria distribuida como una Poisson de parámetro λ.

a) Calcular el estimador máximo verosímil del parámetro λ.

b) Analizar sus propiedades.

Solución.-

a)

FUNCIÓN DE VEROSIMILITUD DE LA MUESTRA (EMV).- Sea $(x_1, \cdots, x_n)$ una muestra aleatoria de una población X con función de masa P_θ (o función de densidad f_θ) donde $\theta = (\theta_1, \cdots, \theta_n)$. El estimador de máxima verosimilitud de θ es el formado por los valores $(\hat{\theta}_1, \cdots, \hat{\theta}_n)$ que maximizan lo que llamaremos *función de verosimilitud* de la muestra $(x_1, \cdots, x_n)$ obtenida:

$$L(\theta) = L(X;\theta) = L(x_1, \cdots, x_n; \theta) = \begin{cases} P(x_1, \theta) \cdots P(x_n, \theta) & \text{caso discreto} \\ f_\theta(x_1) \cdots f_\theta(x_n) & \text{caso continuo} \end{cases}$$

En muchas ocasiones, la forma más cómoda de encontrar el estimador de máxima verosimilitud es considerar $[\ln L(\theta)]$ en vez de $L(\theta)$, ya que es más fácil de manejar y presenta los mismos máximos y mínimos, y despejamos $\theta = (\theta_1, \cdots, \theta_n)$ de la ecuación:

$$\frac{\vartheta \ln L(\theta)}{\vartheta \theta} = 0$$

Sea la v.a. X = 'número de saltos fallidos por hora'

En la distribución de Poisson: $P(X = x) = \dfrac{\lambda^x}{x!} e^{-\lambda}$ $\quad \begin{cases} E(X) = \lambda \\ V(X) = \lambda \end{cases}$

En una muestra aleatoria simple de tamaño n, la función de verosimilitud $L(X, \lambda)$:

$$L(\lambda) = L(X, \lambda) = \prod_{i=1}^{n} P(x_i, \lambda) = \frac{\lambda^{x_1}}{x_1!} e^{-\lambda} \cdots \frac{\lambda^{x_n}}{x_n!} e^{-\lambda} = \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} e^{-n\lambda}$$

$$L(X, \lambda) = \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} e^{-n\lambda} \Rightarrow \ln L(X, \lambda) = \ln\left[\frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} e^{-n\lambda}\right] = \ln(\lambda^{\sum_{i=1}^{n} x_i}) - \ln(\prod_{i=1}^{n} x_i!) + \ln(e^{-n\lambda}) =$$

$$= \sum_{i=1}^{n} x_i \operatorname{Ln}\lambda - \sum_{i=1}^{n} \operatorname{Ln}(x_i!) - n\lambda$$

$$\ln L(X,\lambda) = \sum_{i=1}^{n} x_i \, \text{Ln}\lambda - \sum_{i=1}^{n} \text{Ln}(x_i!) - n\lambda$$

$$\frac{\vartheta \ln L(X,\lambda)}{\vartheta\lambda} = \sum_{i=1}^{n} x_i \frac{1}{\lambda} - n = 0 \quad \Rightarrow \quad \hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$$

Lo que nos dice que el Estimador de Máxima Verosimilitud (EMV) del parámetro λ vendría dado por la media muestral: $\text{EMV}(\lambda) = \bar{x}$

b) Analizar las propiedades

- <u>Insesgadez</u>

El estimador sería insesgado (centrado) si $E(\hat{\lambda}) = \lambda$

$$E(\hat{\lambda}) = E\left[\frac{\sum_{i=1}^{n} x_i}{n}\right] = \frac{1}{n} E(\sum_{i=1}^{n} x_i) = \frac{1}{n} \sum_{i=1}^{n} E(x_i) = \frac{1}{n}(n\lambda) = \lambda$$

- <u>Eficiencia</u>

Para que un estimador sea *<u>eficiente</u>* tiene que ser *centrado* y de *varianza mínima*.
La varianza mínima se analiza en virtud de la <u>acotación de Cramer-Rao:</u>

$$V(\hat{\lambda}) \geq \frac{1}{nE\left[\frac{\vartheta \ln f(x,\lambda)}{\vartheta\lambda}\right]^2} \qquad \text{acotación de Cramer - Rao}$$

Ahora bien, $f(x,\lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$

$$\ln f(x,\lambda) = \ln\left[\frac{\lambda^x}{x!} e^{-\lambda}\right] = x\ln\lambda - \ln(x!) - \lambda$$

$$\frac{\vartheta \ln f(x,\lambda)}{\vartheta\lambda} = \frac{x}{\lambda} - 1 = \frac{x-\lambda}{\lambda}$$

$$E\left[\frac{\vartheta \ln f(x,\lambda)}{\vartheta\lambda}\right]^2 = E\left[\frac{x-\lambda}{\lambda}\right]^2 = \frac{1}{\lambda^2} E(x-\lambda)^2 = \frac{1}{\lambda^2} E(x-\bar{x})^2 = \frac{1}{\lambda^2} V(x) = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$$

En consecuencia, $V(\hat{\lambda}) \geq \frac{1}{n\frac{1}{\lambda}} = \frac{\lambda}{n}$

El resultado nos dice que el menor valor de la varianza del estimador sería λ/n.
$\hat{\lambda} = \bar{x}$ (calculado por el EMV). Sabemos $V(\bar{x}) = \frac{\lambda}{n}$, lo que muestra que el estimador empleado es *eficiente*.

$$V(\overline{x}) = V\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(x_i) = \frac{1}{n^2}(n\lambda) = \frac{\lambda}{n}$$

- Consistencia

Un estimador $\hat{\lambda}$ *consistente* es un estimador asintóticamente insesgado cuya varianza tiende a cero al aumentar el tamaño muestral.

El estimador $\hat{\lambda}$ es *consistente* cuando $\begin{cases} \lim\limits_{n\to\infty} E(\hat{\lambda}) = \lambda \\ \lim\limits_{n\to\infty} V(\hat{\lambda}) = 0 \end{cases}$

$$\lim_{n\to\infty} E(\hat{\lambda}) = \lim_{n\to\infty} \lambda = \lambda$$

$$\lim_{n\to\infty} V(\hat{\lambda}) = \lim_{n\to\infty} \frac{\lambda}{n} = 0$$

El estimador $\hat{\lambda}$ es consistente

18.- En una gran piscifactoría hay una proporción desconocida de peces de una especie A. Para obtener información sobre esta proporción, vamos a ir sacando peces al azar.

a) Si la proporción de peces de la especie A es p., ¿cuál es la probabilidad de que el primer pez de la especie A sea el décimo que extraemos?.
b) Tres personas realizan, independientemente unas de otras, el proceso de sacar peces al azar hasta encontrarse con el primero de tipo A:
   - La primera persona obtiene el primer pez tipo A en la décima extracción.
   - La segunda persona obtiene el primer pez tipo A en la decimoquinta extracción.
   - La tercera persona obtiene el primer pez tipo A en la decimoctava extracción.

Escribir la función de verosimilitud y obtener la estimación de máxima verosimilitud de la proporción p.

<u>Solución</u>.-

El objetivo fundamental del ejercicio es estimar, por máxima verosimilitud, el parámetro p = "proporción de peces de la especie A".

a) P(primer pez tipo A en la décima extracción) = $(1-p)^9\, p$

b) La función de verosimilitud L(p) = P(Resultados muestrales obtenidos)

L(p) = P(primer pez tipo A en la décima extracción y primer pez tipo A en la decimoquinta extracción y primer pez tipo A en la decimoctava extracción)

$$L(p) = \left((1-p)^9\, p\right)\left((1-p)^{14}\, p\right)\left((1-p)^{17}\, p\right) = (1-p)^{40}\, p^3$$

$$\log\left[L(p)\right] = \log\left((1-p)^{40}\, p^3\right) = \log(1-p)^{40} + \log p^3 = 40\log(1-p) + 3\log p$$

$$\frac{\log\left[L(p)\right]}{dp} = \frac{-40}{1-p} + \frac{3}{p} = 0 \quad \mapsto \quad \hat{p} = \frac{3}{43}$$

19.- Las personas de un país se clasifican según dos características: color de los ojos (claros u oscuros) y sexo (hombre o mujer). Las dos características son independientes.

a) Obtenemos una muestra al azar de la población con los siguientes resultados:
   - 200 mujeres con ojos claros
   - 150 hombres con ojos claros
   - 350 mujeres con ojos oscuros
   - 300 hombres con ojos oscuros

Obtener la estimación de máxima verosimilitud de p = P(hombres) y q = P(ojos oscuros)

b) Si tomamos 8 personas al azar de ese país, ¿cuál es la probabilidad de encontrar alguna mujer de ojos oscuros?. Y si la muestra que tomamos es de 200 personas, ¿cuál es la probabilidad de que haya más de 60 mujeres de ojos oscuros?

Solución.-

a) Las probabilidades de los cuatro posibles resultados muestrales son:

- P(mujer con ojos claros) = $(1-p)\,q$
- P(hombre con ojos claros) = $p\,q$
- P(mujer con ojos oscuros) = $(1-p)(1-q)$
- P(hombre con ojos oscuros) = $p\,(1-q)$

La función de verosimilitud L(p, q) = P(resultados muestrales obtenidos)

$$L(p,q) = \big((1-p)\,q\big)^{200}\,(p\,q)^{150}\,\big((1-p)(1-q)\big)^{350}\,\big(p\,(1-q)\big)^{300} = p^{450}\,(1-p)^{550}\,q^{350}\,(1-q)^{650}$$

$$\log L(p,q) = \log\left(p^{450}\,(1-p)^{550}\,q^{350}\,(1-q)^{650}\right) = 450\,\log p + 550\,\log(1-p) + 350\,\log q + 650\,\log(1-q)$$

$$\frac{\vartheta \log L(p,q)}{\vartheta p} = \frac{450}{p} - \frac{550}{1-p} = 0 \quad \mapsto \quad \hat{p} = 0{,}45$$

$$\frac{\vartheta \log L(p,q)}{\vartheta q} = \frac{350}{q} - \frac{650}{1-q} = 0 \quad \mapsto \quad \hat{q} = 0{,}35$$

b) Conocemos que P(mujer con ojos oscuros) = $(1-p)(1-q) = 0{,}24$

La variable aleatoria X = "número de mujeres con ojos oscuros, entre 8" sigue una distribución binomial $B(n=8\,;\,p=0{,}24)$

$$P(X \geq 1) = 1 - P(X=0) = 1 - \binom{8}{0}\,(0{,}24)^0\,(0{,}76)^8 = 0{,}89$$

La variable Y = "número de mujeres con ojos oscuros, entre 200" sigue una distribución binomial $B(n = 20\,;\, p = 0{,}24)$, que por ser el tamaño de la muestra grande (n = 200) y p no próximo a cero (p = 0,24) aproximamos por la distribución normal

$$B(n = 20\,;\, p = 0{,}24) \approx N(\mu = np = 48\,;\, \sigma = \sqrt{npq} = \sqrt{200\,(0{,}24)(0{,}76)} = 6{,}04)$$

$$P(Y > 60) = P\left(\frac{Y - 48}{6{,}04} > \frac{60 - 48}{6{,}04}\right) = P(z > 1{,}99) = 0{,}0233$$

20.- Calcular el estimador máximo verosímil del parámetro 'a' de las siguientes funciones:

a) $f(x;\, a) = a^2\, e^{-ax}$ siendo $x \geq 0$ en muestras aleatorias simples de tamaño n.

b) $f(x;\, a) = a\, e^{-ax}$ para $x \geq 0$, $a > 0$ en muestras aleatorias simples de tamaño 2.

Solución.-

a) $f(x;\, a) = a^2\, e^{-ax}$ donde $x \geq 0$ en m.a.s. de tamaño n

La función de verosimilitud

$$L = L(x_1, x_2, \cdots, x_n; a) = (a^2\, e^{-a x_1}).(a^2\, e^{-a x_2}) \cdots (a^2\, e^{-a x_n}) = a^{2n}\, e^{-a \sum\limits_{i=1}^{n} x_i}$$

aplicando logaritmos neperianos: $\log L = \log\left(a^{2n}\, e^{-a \sum\limits_{i=1}^{n} x_i}\right) = 2n \log a - a \sum\limits_{i=1}^{n} x_i$

derivando respecto de 'a' e igualando a cero:

$$\frac{d(\log L)}{da} = \frac{2n}{a} - \sum_{i=1}^{n} x_i = 0 \;\Rightarrow\; \hat{a} = \frac{2n}{\sum\limits_{i=1}^{n} x_i} = \frac{2}{\overline{x}} \qquad\qquad \hat{a} = \frac{2}{\overline{x}}$$

b) Sea $f(x;\, a) = a\, e^{-ax}$ para $x \geq 0$, $a > 0$ en m.a.s. de tamaño 2

La función de verosimilitud $L = L(x_1, x_2; a) = (a\, e^{-a x_1}).(a\, e^{-a x_2}) = a^2\, e^{-a(x_1 + x_2)}$

aplicando logaritmos neperianos: $\log L = \log\left(a^2\, e^{-a(x_1+x_2)}\right) = 2 \log a - a(x_1 + x_2)$

derivando respecto de 'a' e igualando a cero:

$$\frac{d(\log L)}{da} = \frac{2}{a} - (x_1 + x_2) = 0 \;\Rightarrow\; \hat{a} = \frac{2}{x_1 + x_2} = \frac{1}{\overline{x}}$$

21.- Sea la distribución $N(\mu;\sigma)$, con media μ conocida y varianza desconocida. Calcular la estimación máximo-verosimíl de la varianza en muestras aleatorias simples de tamaño n.

Solución.-

La función de verosimilitud es:

$$L(X;\mu,\sigma^2)=\left[\frac{1}{\sqrt{2\pi\sigma^2}}\,e^{-\frac{(x_1-\mu)^2}{2\sigma^2}}\right]\left[\frac{1}{\sqrt{2\pi\sigma^2}}\,e^{-\frac{(x_2-\mu)^2}{2\sigma^2}}\right]\cdots\left[\frac{1}{\sqrt{2\pi\sigma^2}}\,e^{-\frac{(x_n-\mu)^2}{2\sigma^2}}\right]=$$

$$=\frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2)^{\frac{n}{2}}}\,e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}$$

tomando logaritmos neperianos, se tiene:

$$\log\left[L(X;\mu,\sigma^2)\right]=\log\left[\frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2)^{\frac{n}{2}}}\,e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}\right]=-\frac{n}{2}\log(2\pi)-\frac{n}{2}\log(\sigma^2)-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}$$

y derivando respecto a σ^2 e igualando a cero:

$$\frac{d\log\left[L(X;\mu,\sigma^2)\right]}{d\sigma^2}=-\frac{n}{2\sigma^2}+\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^4}=0$$

como $\sigma^2>0$, el estimador máximo verosímil de σ^2 será: $\hat{\sigma}^2=\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}$

Conviene observar que el estimador no es la varianza muestral, dado que las desviaciones de los valores muestrales lo son con respecto a la media poblacional μ y no respecto a la media muestral $\overline{x}$.

22.- Sea la distribución $N(\mu;\sigma)$, con la media y varianza desconocidas. Calcular los estimadores máximo-verosímiles de μ y σ^2.

Solución.-

La función de verosimilitud es:

$$L(X;\mu,\sigma^2)=\left[\frac{1}{\sqrt{2\pi\sigma^2}}\,e^{-\frac{(x_1-\mu)^2}{2\sigma^2}}\right]\left[\frac{1}{\sqrt{2\pi\sigma^2}}\,e^{-\frac{(x_2-\mu)^2}{2\sigma^2}}\right]\cdots\left[\frac{1}{\sqrt{2\pi\sigma^2}}\,e^{-\frac{(x_n-\mu)^2}{2\sigma^2}}\right]=$$

$$=\frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2)^{\frac{n}{2}}}\,e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}$$

tomando logaritmos neperianos, se tiene:

$$\log\left[L(X;\mu,\sigma^2)\right]=\log\left[\frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2)^{\frac{n}{2}}}\,e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}\right]=-\frac{n}{2}\log(2\pi)-\frac{n}{2}\log(\sigma^2)-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}$$

y derivando respecto a μ y σ^2, e igualando a cero:

$$\frac{\vartheta\log\left[L(X;\mu,\sigma^2)\right]}{\vartheta\mu}=\frac{\sum_{i=1}^{n}(x_i-\mu)}{\sigma^2}=0$$

$$\frac{\vartheta\log\left[L(X;\mu,\sigma^2)\right]}{\vartheta\sigma^2}=-\frac{n}{2\sigma^2}+\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^4}=0$$

resolviendo el sistema resulta: $\hat{\mu}=\overline{x}$ y $\hat{\sigma}^2=\dfrac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n}=\sigma_x^2$

Los estimadores máximo-verosímiles de μ y σ^2 son la media y la varianza muestrales.

CÁLCULO DE ESTIMADOR POR EL MÉTODO DE LOS MOMENTOS

23.- Sea una población definida por:

$$\left.\begin{array}{l} P(\xi=-1)=\dfrac{1-\theta}{2} \\ P(\xi=0)=\dfrac{\theta+\lambda}{2} \\ P(\xi=1)=\dfrac{1-\lambda}{2} \end{array}\right\} \quad \begin{array}{l} 0<\theta<1 \\ 0<\lambda<1 \end{array}$$

Estimar los parámetros θ y λ por el método de los momentos, estudiando si son insesgados.

Solución.-

MÉTODO DE LOS MOMENTOS.- El procedimiento consiste en igualar momentos poblacionales respecto al origen (α_r) a los correspondientes momentos muestrales respecto al origen (a_r), formando así tantas ecuaciones como parámetros poblacionales se pretenden estimar:

$$\begin{cases} \alpha_1 = E(X)=\mu \;\Rightarrow\; \hat{\alpha}_1 = a_1 = \dfrac{\sum_{i=1}^{n} x_i}{n} = \bar{x} \\ \alpha_2 = E(X^2) \;\Rightarrow\; \hat{\alpha}_2 = a_2 = \dfrac{\sum_{i=1}^{n} x_i^2}{n} \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ \alpha_r = E(X^r) \;\Rightarrow\; \hat{\alpha}_r = a_r = \dfrac{\sum_{i=1}^{n} x_i^r}{n} \end{cases}$$

Puesto que hay que estimar dos parámetros hay que calcular los dos primeros momentos.

$$\overbrace{\begin{array}{l} \alpha_1 = \mu = E(\xi) = \sum_i x_i P(\xi=x_i) = (-1)\left(\dfrac{1-\theta}{2}\right) + (0)\left(\dfrac{\theta+\lambda}{2}\right) + (1)\left(\dfrac{1-\lambda}{2}\right) = \dfrac{\theta-\lambda}{2} \\ \alpha_2 = E(\xi^2) = \sum_i x_i^2\, P(\xi=x_i) = (-1)^2\left(\dfrac{1-\theta}{2}\right) + (0)^2\left(\dfrac{\theta+\lambda}{2}\right) + (1)^2\left(\dfrac{1-\lambda}{2}\right) = \dfrac{2-\theta-\lambda}{2} \end{array}}^{\text{momentos poblacionales}}$$

$$\overbrace{a_1 = \bar{x} = \dfrac{\sum_i x_i}{n} \qquad a_2 = \dfrac{\sum_i x_i^2}{n}}^{\text{momentos muestrales}}$$

$$\left.\begin{array}{l} \alpha_1 = a_1 \;\Rightarrow\; \dfrac{\theta - \lambda}{2} = \overline{x} \;\Rightarrow\; \theta - \lambda = 2\overline{x} \\[2ex] \alpha_2 = a_2 \;\Rightarrow\; \dfrac{2 - \theta - \lambda}{2} = a_2 \;\Rightarrow\; -\theta - \lambda = 2a_2 - 2 \end{array}\right\} \begin{array}{l} \theta - \lambda = 2\overline{x} \\ -\theta - \lambda = 2a_2 - 2 \end{array} \left\{\begin{array}{l} \hat{\lambda} = 1 - a_2 - \overline{x} \\ \hat{\theta} = 1 - a_2 + \overline{x} \end{array}\right.$$

- <u>Insesgadez</u>

Un estimador $\hat{\theta}$ es insesgado (o centrado) cuando se verifica $E(\hat{\theta}) = \theta$

$$E(\hat{\theta}) = E(1 - a_2 + \overline{x}) = 1 - E(a_2) + E(\overline{x}) = 1 - \alpha_2 + \mu = 1 - \left(\frac{2 - \theta - \lambda}{2}\right) + \left(\frac{\theta - \lambda}{2}\right) = \theta$$

$$E(\hat{\lambda}) = E(1 - a_2 - \overline{x}) = 1 - E(a_2) - E(\overline{x}) = 1 - \alpha_2 - \mu = 1 - \left(\frac{2 - \theta - \lambda}{2}\right) - \left(\frac{\theta - \lambda}{2}\right) = \lambda$$

Los estimadores θ y λ son insesgados.

CÁLCULO DE ESTADÍSTICOS. FUNCIÓN DE DENSIDAD

24.- Una muestra aleatoria $(X_1, \cdots, X_n)$ de la población tiene como función de densidad $f_\theta(x) = \begin{cases} \theta x^{\theta-1} & \text{si } x \in (0,1) \\ 0 & \text{en el resto} \end{cases} \quad \theta > 0$

a) Hallar un estadístico suficiente
b) Estimador de máxima verosimilitud de θ
c) Estimador de θ por el método de los momentos

Solución.-

a)

Un estimador $\hat\theta$ es suficiente cuando no da lugar a una pérdida de información. Es decir, cuando la información basada en $\hat\theta$ es tan buena como la que hiciera uso de toda la muestra.
Para identificar estadísticos suficientes se utiliza el teorema de factorización, que dice que dada una muestra aleatoria $(x_1, \cdots, x_n)$ de una población X con función de masa P_θ (o función de densidad f_θ) un estadístico $\hat\theta$ es suficiente para θ si y sólo sí:

$$\begin{cases} P_\theta(x_1, \cdots, x_n) = g\left[\hat\theta(x_1, \cdots, x_n), \theta\right] . h(x_1, \cdots, x_n) & \textit{caso discreto} \\ f_\theta(x_1, \cdots, x_n) = g\left[\hat\theta(x_1, \cdots, x_n), \theta\right] . h(x_1, \cdots, x_n) & \textit{caso continuo} \end{cases}$$

Para encontrar un estadístico suficiente $\hat\theta$ hay que factorizar la función de verosimilitud de la forma: $L(\theta) = g(\hat\theta, \theta) . h(x_1, \cdots, x_n)$

$$L(\theta) = f_\theta(x_1) f_\theta(x_2) \cdots f_\theta(x_n) = (\theta x_1^{\theta-1})(\theta x_2^{\theta-1}) \cdots (\theta x_n^{\theta-1}) = \theta^n (x_1 \cdots x_n)^{\theta-1}$$

Por tanto, $\hat\theta = x_1, \cdots, x_n$ es un estadístico suficiente.

b) $L(\theta) = \theta^n (x_1 \cdots x_n)^{\theta-1}$

$$\ln L(\theta) = \ln\left(\theta^n (x_1 \cdots x_n)^{\theta-1}\right) = \ln \theta^n + \ln \prod_{i=1}^n x_i^{\theta-1} = \ln \theta^n + \sum_{i=1}^n \ln(x_i^{\theta-1})$$

$$\ln L(\theta) = n \ln\theta + (\theta-1) \sum_{i=1}^n \ln(x_i) \Rightarrow \frac{\vartheta \ln L(\theta)}{\vartheta \theta} = \frac{n}{\theta} + \sum_{i=1}^n x_i = 0 \mapsto \hat\theta = -\frac{n}{\sum_{i=1}^n \ln(x_i)}$$

c) Se plantea la ecuación $E(X) = \overline{x}$

$$\overline{x} = E(X) = \int_0^1 x\, f_\theta(x)\, dx = \int_0^1 x\ \theta\, x^{\theta - 1}\, dx = \int_0^1 \theta\, x^{\theta}\, dx = \theta \left[\frac{x^{\theta + 1}}{\theta + 1} \right]_0^1 = \frac{\theta}{\theta + 1}$$

$$\overline{x}\ (\theta + 1) = \theta \quad \Rightarrow \quad \hat{\theta} = \frac{\overline{x}}{1 - \overline{x}}$$

25.- Una muestra aleatoria $(X_1, \cdots, X_n)$ de la población tiene como función de densidad $f_\theta(x) = \begin{cases} e^{-x+\theta} & \text{si } x > 0 \\ 0 & \text{en el resto} \end{cases}$

a) Hallar un estimador por el método de los momentos de θ
b) Estudiar si el estimador encontrado en el apartado anterior es insesgado para estimar el parámetro θ

Solución.-

a) Se plantea la ecuación: $E[X] = \overline{x}$

$$\overline{x} = E[X] = \int_\theta^\infty x\, f_\theta(x)\, dx = \overbrace{\int_\theta^\infty x\, e^{-x+\theta}\, dx}^{\text{int egración por partes}} = \theta + 1 \quad \Rightarrow \quad \hat{\theta} = \overline{x} - 1$$

b) Un estimador es insesgado o centrado cuando su valor probable coincide con el valor
del parámetro a estimar. Es decir, $E(\hat{\theta}) = \theta$

$$E(\hat{\theta}) = E(\overline{x} - 1) = E(\overline{x}) - 1 = (\theta + 1) - 1 = \theta$$

integración por partes
$$\begin{cases} \int \underbrace{x}_{u}\, \underbrace{e^{-x+\theta} dx}_{dv} = \underbrace{x}_{u}(\underbrace{-e^{-x+\theta}}_{v}) - \int \underbrace{-e^{-x+\theta}}_{v} \underbrace{dx}_{du} = -x e^{-x+\theta} - e^{-x+\theta} = \\ \quad = -(1+x)\, e^{-x+\theta} = -e^{-\theta}\left(\dfrac{1+x}{e^x}\right) \\ \int_\theta^\infty x\, e^{-x+\theta} dx = -e^{-\theta}\left(\dfrac{1+x}{e^x}\right)_\theta^\infty = 1 + \theta \end{cases}$$

26.- Una muestra aleatoria $(X_1, \cdots, X_n)$ de la población tiene como función de densidad $f_\theta(x) = \begin{cases} \theta^2 x\, e^{-\theta x} & \text{si } x > 0 \\ 0 & \text{en el resto} \end{cases}$

Hallar el estimador de máxima verosimilitud de θ

Solución.-

La función de verosimilitud $L(\theta)$:

$$L(\theta) = f_\theta(x_1)\, f_\theta(x_2) \cdots f_\theta(x_n) = (\theta^2 x_1\, e^{-\theta x_1})\,(\theta^2 x_2\, e^{-\theta x_2}) \cdots (\theta^2 x_n\, e^{-\theta x_n})$$

$$= \theta^{2n}\,(x_1 \cdots x_n)\; e^{-(\theta x_1 + \theta x_2 + \cdots + \theta x_n)} = \theta^{2n}\,(x_1 \cdots x_n)\; e^{-\theta \sum_{i=1}^{n} x_i}$$

$$L(\theta) = \theta^{2n}\,(x_1 \cdots x_n)\; e^{-\theta \sum_{i=1}^{n} x_i} \Rightarrow \ln L(\theta) = \ln\left[\theta^{2n}\,(x_1 \cdots x_n)\; e^{-\theta \sum_{i=1}^{n} x_i}\right]$$

$$\ln L(\theta) = (2n)\ln\theta + \ln\prod_{i=1}^{n} x_i - \theta \sum_{i=1}^{n} x_i \Rightarrow \ln L(\theta) = (2n)\ln\theta + \sum_{i=1}^{n}\ln x_i - \theta \sum_{i=1}^{n} x_i$$

$$\frac{\vartheta \ln L(\theta)}{\vartheta\theta} = \frac{2n}{\theta} - \sum_{i=1}^{n} x_i = 0 \Rightarrow \hat{\theta} = \frac{2n}{\sum_{i=1}^{n} x_i}$$

27.- El coseno X del ángulo con el que se emiten los electrones en un proceso radioactivo es una variable aleatoria con función de densidad

$$f_\theta(x) = \begin{cases} (1+\theta x)/2 & -1 \le x \le 1 \\ 0 & \text{en el resto} \end{cases} \qquad -1 \le \theta \le 1$$

Consideremos una muestra aleatoria $(X_1, \cdots, X_n)$ de esta variable aleatoria

a) Obtener el estimador θ por el método de los momentos

b) Calcular la varianza de este estimador y demostrar que es consistente

<u>Solución.-</u>

a) Se plantea la ecuación $E[X] = \overline{x}$

$$\overline{x} = E[X] = \int_{-1}^{1} x \, \frac{1+\theta x}{2}\, dx = \left[\frac{x^2}{2} + \frac{\theta x^3}{6}\right]_{-1}^{1} = \frac{\theta}{3} \quad \Rightarrow \quad \hat{\theta} = 3\overline{x}$$

b) $V(\hat{\theta}) = V(3\overline{x}) = 9V(\overline{x}) = 9\,\dfrac{V(X)}{n} = \dfrac{9}{n}\,V(X)$

$$V(X) = E(X^2) - [E(X)]^2 = \int_{-1}^{1} x^2\, \frac{1+\theta x}{2}\, dx - \left[\frac{\theta}{3}\right]^2 = \left[\frac{x^3}{6} + \frac{\theta x^4}{8}\right]_{-1}^{1} - \left[\frac{\theta}{3}\right]^2 = \frac{3-\theta^2}{9}$$

de donde, $V(\hat{\theta}) = \dfrac{9}{n}\,V(X) = \dfrac{9}{n}\left[\dfrac{3-\theta^2}{9}\right] = \dfrac{3-\theta^2}{n}$

Para probar que $\hat{\theta}$ es consistente para estimar θ es suficiente probar $\begin{cases} \lim\limits_{n\to\infty} E(\hat{\theta}) = \theta \\ \lim\limits_{n\to\infty} V(\hat{\theta}) = 0 \end{cases}$

$$\lim_{n\to\infty} E(\hat{\theta}) = \lim_{n\to\infty} E(3\overline{x}) = \lim_{n\to\infty} 3E(\overline{x}) = 3E(X) = 3\,\frac{\theta}{3} = \theta$$

$$\lim_{n\to\infty} V(\hat{\theta}) = \lim_{n\to\infty} V(3\overline{x}) = \lim_{n\to\infty} \frac{3-\theta^2}{n} = 0$$

Por tanto, queda probado que $\hat{\theta}$ es consistente para estimar θ