

# Tarea 3: EIVM e Intervalos de Confianza

Estadística Matemática

Mauricio Vazquez Moran

Aitana Acosta

Giancarlo De La Rosa

18 de Abril de 2024

## Estimadores insesgados de varianza mínima

A continuación, se muestra la tabla para cada pareja de  $n$  y  $\lambda$ :

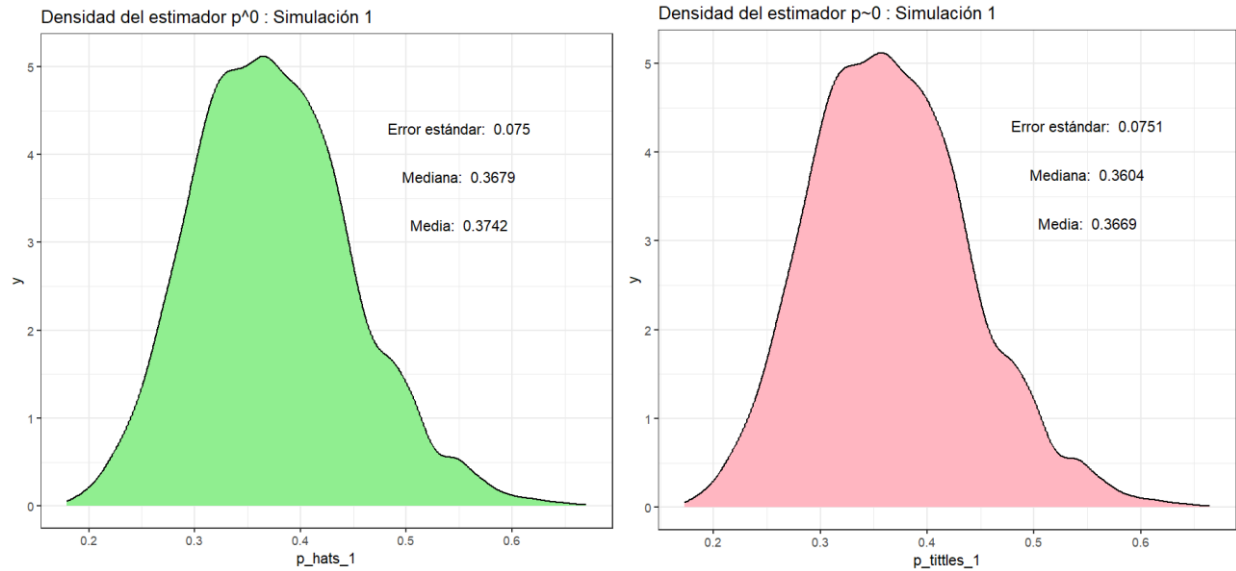
$\lambda$	$p_0$	$\theta$	$n$	$\hat{p}_0$	$ee(\hat{p}_0)$	$\tilde{p}_0$	$ee(\tilde{p}_0)$	$\sim p_0$	$ee(\sim p_0)$
1.0	0.3679	0.3770	25	0.375	0.0736	0.3678	0.0944	0.3677	0.0736
			50	0.3712	0.0526	0.3671	0.0683	0.3676	0.0526
			80	0.3696	0.0404	0.3679	0.0532	0.3673	0.0404
0.5	0.6065	0.6140	25	0.6131	0.0839	0.6095	0.0956	0.6071	0.0847
			50	0.6101	0.0597	0.6071	0.0681	0.607	0.06
			80	0.609	0.0481	0.6074	0.0545	0.6071	0.0482
2.0	0.1353	0.1422	25	0.1409	0.039	0.1346	0.068	0.1355	0.0383
			50	0.1373	0.0274	0.1338	0.0474	0.1346	0.0272
			80	0.1371	0.0212	0.1358	0.0375	0.1354	0.0211

El ejercicio ya nos demuestra la calidad de los estimadores. Con  $N=3,000$  simulaciones podemos apreciar con valores exactos y comparar con mayor certeza cuál es mejor para el presente caso.

Las conclusiones de sesgo y variabilidad que se plantean desde un inicio en el ejercicio se mantienen y  $\sim p_0$ , como un EIVM, es el mejor estimador para  $p_0$ . Sin embargo, las simulaciones nos presentan algunas curiosidades.

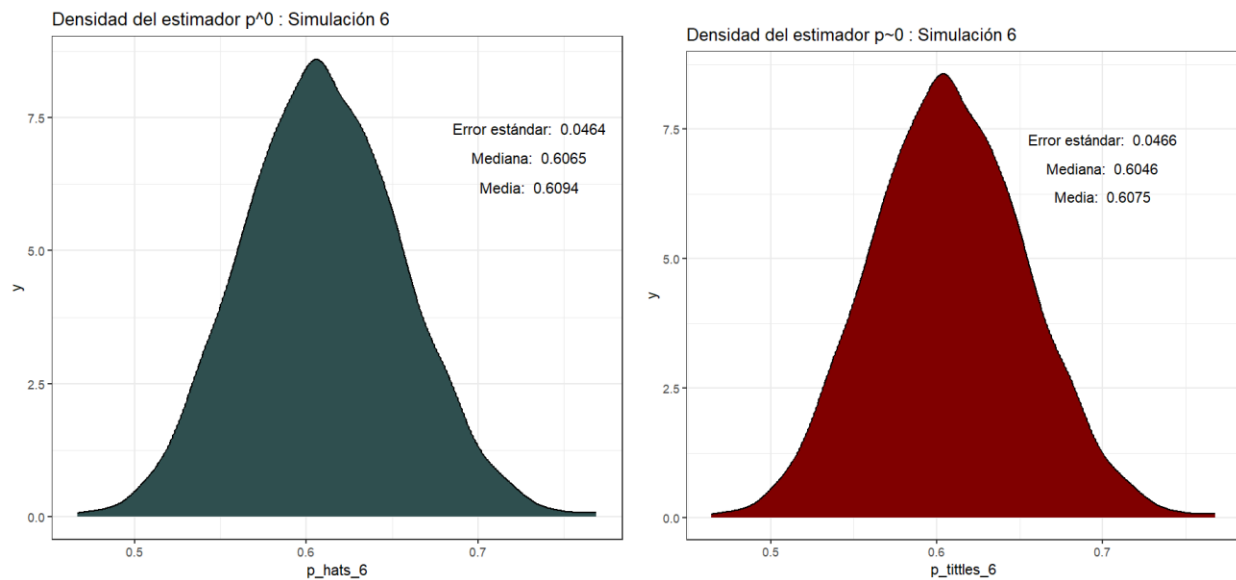
Por ejemplo,  $\hat{p}_0$  y  $\sim p_0$  se asemejan en variabilidad y en distribución son casi idénticas sin importar el valor de  $n$  ni de  $\lambda$ . Se volvió a correr la simulación y se mantuvo esta característica como se muestra en las Figura 1, 2 y 3. Esto nos lleva a hacer la siguiente conjetura:  **$\text{Var}(\sim p_0)$  tiende a  $\text{Var}(\hat{p}_0)$  cuando  $N$  es muy grande (o viceversa).** Realmente fallamos en calcular  $\text{Var}(\sim p_0)$  de forma analítica, pero creemos que este ejercicio muestra lo anterior.

Así mismo, encontramos que el error estándar de  $\tilde{p}_0$  es inversamente proporcional al tamaño de  $n$ . Esto es claro pues  $\tilde{p}_0$  está en función de una Bernoulli y entre más grande sea la población, más valores de  $Y_i=0$  habrá y la distribución se suaviza.

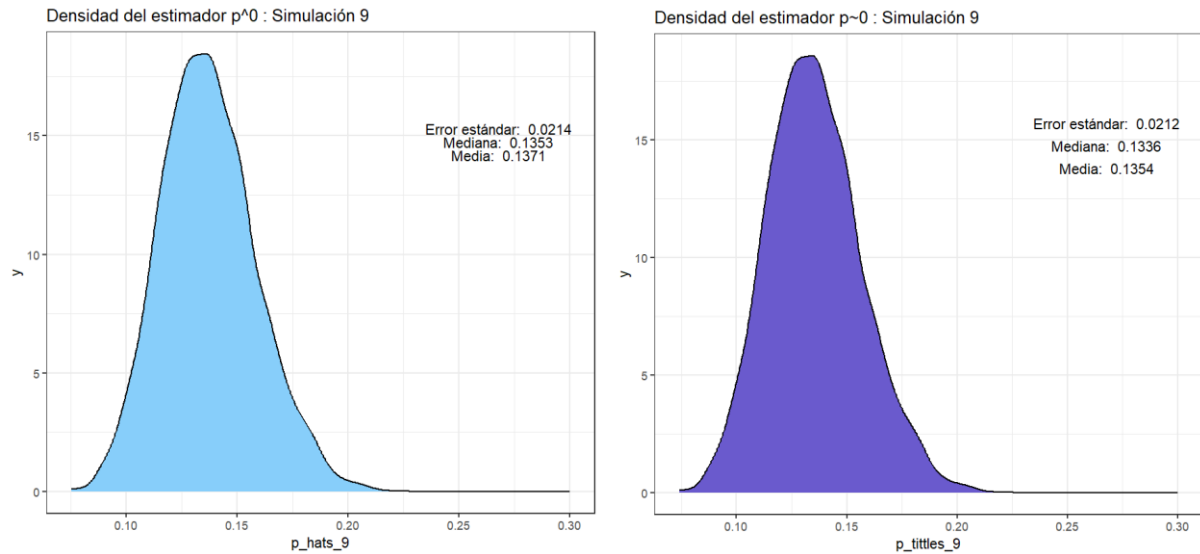


**Figura 1:** Se muestran las gráficas de distribución para  $\hat{p}_0$  y  $\tilde{p}_0$  con  $n=25$  y  $\lambda=1.0$ .

El error estándar de ambos es prácticamente el mismo, al igual que su distribución.



**Figura 2:** Mismo caso con  $n=50$  y  $\lambda=0.5$



**Figura 3:** Aquí tenemos esto con  $n=80$  y  $\lambda=2.0$ . En este caso notamos un poco de diferencia en tanto amplitud de la curva de la distribución.

### Amplitud de los intervalos de probabilidad

$X \sim N(0, \sigma)$		Colas del mismo peso			Amplitud mínima		
$\sigma$	$1 - \alpha$	$q_i$	$q_s$	$\omega$	$q_i$	$q_s$	$\omega$
1.0	0.90	-1.644854	1.644854	3.29707	-1.644854	1.644854	3.29707
	0.95	-1.959964	1.959964	3.919928	-1.959964	1.959964	3.919928
	0.99	-2.575829	2.575829	5.151659	-2.575829	2.575829	5.151659
2.0	0.90	-3.289707	3.289707	6.579415	-3.289707	3.289707	6.579415
	0.95	-3.919928	3.919928	7.839856	-3.919928	3.919928	7.839856
	0.99	-5.151659	5.151659	10.30332	-5.151659	5.151659	10.30332
5.0	0.90	-8.224268	8.224268	16.44854	-8.224268	8.224268	16.44854
	0.95	-9.79982	9.79982	19.59964	-9.79982	9.79982	19.59964
	0.99	-12.87915	12.87915	25.75829	-12.87915	12.87915	25.75829

$Y \sim \chi_n^2$		Colas del mismo peso			Amplitud mínima		
$n$	$1 - \alpha$	$q_i$	$q_s$	$\omega$	$q_i$	$q_s$	$\omega$
5	0.90	1.145476	11.0705	9.925021	0.475794	9.43322	8.95742
	0.95	0.8312116	12.8325	12.00129	0.291955	11.18723	10.89527
	0.99	0.4117419	16.7496	16.33786	0.100251	15.12616	15.0259
10	0.90	3.940299	18.30704	14.36674	2.984672	16.67871	13.69404
	0.95	3.246973	20.48318	17.2362	2.41612	18.86264	16.44652
	0.99	2.155856	25.18818	23.03232	1.500815	23.53574	22.03493
20	0.90	10.85081	31.41043	20.55962	9.749783	29.84016	20.09037
	0.95	9.590777	34.16961	24.57883	8.496118	32.52211	24.02599
	0.99	7.433844	39.99685	32.563	6.558438	38.40366	31.84522

Para detalles del código ejecutado en las partes 1 y 2, consultar:

[https://github.com/MauricioVazquezM/Mathematical\\_Statistics\\_Spring2024/blob/main/ASSIGNMENT%203/Assignment3.R](https://github.com/MauricioVazquezM/Mathematical_Statistics_Spring2024/blob/main/ASSIGNMENT%203/Assignment3.R)

## Comparación de medias y varianzas de poblaciones normales

1. Construya un intervalo del 90 % de confianza para el tiempo medio de procesamiento  $\mu_1$  del algoritmo 1 si sabe que su desviación estándar es de 1 segundo.

- Sabemos que el intervalo de una normal %confianza  $100(1-\alpha)\%$  para  $\mu$  con  $\sigma^2$  conocida es  $\left[ \bar{x} \pm z_{(1-\alpha/2)} \cdot \frac{\sigma}{\sqrt{n}} \right]$ .

→ Intervalo para  $\mu_1$  si  $\sigma=1$  con 90%, i.e  $\alpha=.1$

$$\bar{x} \approx 9.93, n=25, z_{(.95)} \approx 1.65 \quad \left( 9.93 - z_{(.95)} \cdot \frac{1}{\sqrt{5}}, 9.93 + z_{(.95)} \cdot \frac{1}{\sqrt{5}} \right)$$

∴ IC para  $\mu_1$  con 90% confianza : (9.6, 10.26)

2. Construya el correspondiente intervalo para el tiempo medio  $\mu_2$  del algoritmo 2, con un nivel de 0.95 de confianza reconociendo que no conoce su desviación estándar.

- Sabemos que el intervalo %confianza  $100(1-\alpha)\%$  de una normal para  $\mu$  con  $\sigma^2$  desconocida es  $\left[ \bar{x} \pm t_{(n-1)}^{(1-\alpha/2)} \cdot \frac{\sqrt{s^2}}{\sqrt{n}} \right]$

estimamos  $\sigma^2$  usando  $S^2$  (var muestral):

$$s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2 = \frac{1}{24} \sum (x_i - 10.43)^2 \approx 2.43$$

→ Intervalo para  $\mu_1$  con 95% de confianza si:  $s^2 \approx 2.43$ ,  $\bar{x} = 10.43$ ,  $t_{(24)}^{(.975)} = 2.064$

$$\left( 10.43 - 2.064 \left( \sqrt{2.43} \right)^{1/5}, 10.43 + 2.064 \left( \sqrt{2.43} \right)^{1/5} \right)$$

∴ IC para  $\mu_1$  %95% de confianza  $\approx (9.3865, 11.0735)$

4. Responda el inciso anterior pero con base a un intervalo del 95 % de confianza para el cociente de varianzas  $\theta = \sigma_1^2 / \sigma_2^2$ .

→ Sabemos que el intervalo c/confianza  $100(1-\alpha)\%$  de  $\theta = \frac{\sigma_1^2}{\sigma_2^2}$  con  $\bar{x}_1, \bar{x}_2$  muestras normales con  $\mu_1, \mu_2$  desconocidas es:

$$\left[ \frac{s_1^2}{s_2^2} F_{(n_1-1, n_2-1)}^{(\alpha/2)}, \frac{s_1^2}{s_2^2} F_{(n_1-1, n_2-1)}^{(1-\alpha/2)} \right]$$

∴ el IC si  $s_1^2 = 1.025$ ,  $s_2^2 = 2.43$ ,  $F_{(24, 24)}^{(0.025)} = .44067$ ,  $F_{(24, 24)}^{(0.975)} = 2.269$  : (.1859, .957)

3. ¿Podría suponer que la variabilidad del tiempo de procesamiento es la misma para los dos algoritmos? Para esto, construya un intervalo de confianza del 95 % para la desviación estándar  $\sigma_i$  de cada uno de los algoritmos y concluya.

• Sabemos que el intervalo c/confianza  $100(1-\alpha)\%$  de una normal para  $\sigma$  con  $\mu$  desconocida es  $\left(0, \frac{(n-1)s^2}{\chi^2_{(n-1)}(\alpha)}\right)$ .

→ Intervalo para  $\sigma_1$  con 95% de confianza si  $s^2 = 1.025$ ,  $\chi^2_{(24)}^{(0.05)} = 36.415$  : (0, .6755)

→ Intervalo para  $\sigma_2$  con 95% de confianza si  $s^2 = 2.43$ ,  $\chi^2_{(24)}^{(0.05)} = 36.415$  : (0, 1.6015)

★ Notamos que la variabilidad del algoritmo 2 es mayor que la del algoritmo 1.