Antonio Toledo Galinso
Isaac Alejandro Pimentel
Felipe Castro

- **Fecha máxima de entrega:** Martes 15 de noviembre @ 19:00h.

Estimadores insesgados de varianza mínima

Considere $X_1, \ldots, X_n$ una muestra aleatoria de una distribución Poisson de media λ. Encuentre el estimador insesgado de mínima varianza de $p_0 = P(N = 0) = e^{-\lambda}$.

Con este fin note que

$$\hat{\lambda} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

es el estimador por el método de momentos y de máxima verosimilitud del parámetro λ. Claramente, el estimador es insesgado y se puede mostrar que es de varianza mínima pues, por ejemplo, su varianza, $\text{var}(\bar{X}) = \lambda/n$ alcanza la **cota de inferior de Cramér-Rao**. Sin embargo, el estimador

$$\hat{p}_0 = e^{-\hat{\lambda}}$$

es un estimador sesgado de p_0. De hecho, puesto que $n\hat{\lambda} \sim \text{Po}(n\lambda)$, se puede emplear la función generadora de momentos y sus propiedades para mostrar que

$$\theta = \mathbb{E}[\hat{p}_0] = \exp\left\{n\lambda(e^{-1/n} - 1)\right\}$$

Por otro lado, si define $Y_i = \mathbb{1}_{\{0\}}(X_i)$, para $i = 1, \ldots, n$, entonces las Y_i son *v.a.i.i.d.* distribuidas $\text{Ber}(p_0)$ y

$$\check{p}_0 = \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

es un estimador insesgado de p_0 con varianza $e^{-\lambda}(1 - e^{-\lambda})/n$.

Ahora bien, note que $T = \sum_{i=1}^{n} X_i$ es un estadístico suficiente para λ y $Y_1 = \mathbb{1}_{\{0\}}(X_1)$ es un estimador insesgado de $p_0 = e^{-\lambda}$. Se sigue entonces del **Teorema de Rao-Blackwell** que

$$\tilde{p}_0 = \mathbb{E}[Y_1|T] = \left(\frac{n-1}{n}\right)^T$$

es un estimador insesgado de menor varianza (¿cuál?) que $\check{p}_0$.

Finalmente, puesto que la distribución Poisson es miembro de la **familia exponencial de un parámetro**, el estadístico T es un estadístico **completo suficiente minimal**. Se concluye por el **Teorema de Lehmann-Scheffé** que $\tilde{p}_0$ es entonces un estimador insesgado de varianza mínima.

Para ilustrar lo anterior realice un ejercicio de simulación para completar la siguiente tabla:

lambda	p0	theta	n	p^0	ee(p^0)	pˇ0	ee(pˇ0)	p˜0	ee(p˜0)
1	0-3679	0.377	25	0.339596	0.074274	0.32	0.096446	0.332142	0.073576
			50	0.307279	0.052279	0.36	0.068197	0.303626	0.052026
			80	0.472367	0.041257	0.45	0.053915	0.470139	0.04113
0.5	0.6065	0.614	25	0.697676	0.085327	0.68	0.097704	0.692534	0.085776
			50	0.458406	0.060498	0.48	0.069087	0.454796	0.060653
			80	0.541994	0.047874	0.5	0.054618	0.539906	0.04795
2	0.1353	0.1422	25	0.090718	0.039796	0.2	0.068416	0.086352	0.038279
			50	0.120032	0.027606	0.1	0.048378	0.11748	0.027067
			80	0.147711	0.021665	0.175	0.038246	0.145941	0.021398

donde λ es el parámetro de la distribución Poisson; p_0 y θ definidos arriba; el promedio de los $\hat{p}_0$, $\check{p}_0$ y $\tilde{p}_0$; $\text{ee}(\cdot)$ los correspondientes errores estándar.

Simule $N = 3000$ ensayos para cada pareja λ y n. Concluya comentando el ejercicio.

Considere $X_1, \ldots, X_n$ una muestra aleatoria de una distribución Poisson de media λ. Encuentre el estimador insesgado de mínima varianza de $p_0 = P(N = 0) = e^{-\lambda}$.

Con este fin note que

$$\hat{\lambda} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

EMM

$E[X] = \lambda$, $E[X] \equiv m_1 \iff \boxed{\tilde{\lambda}_{EMM} = \bar{X}}$

$| \; f(x) = \frac{\lambda^x}{x!}\, e^{-\lambda} I_{\{0,1,\ldots\}}(x)$

$$L(\lambda; \underline{x}) = \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} e^{-n\lambda} \, \mathbb{I}_{\{0,1,\ldots\}}(x)$$

$$\ell(\lambda; \underline{x}) = \sum_{i=1}^{n} x_i \log \lambda - \log \prod_{i=1}^{n} x_i! - n\lambda \;, \quad \frac{d\ell}{d\lambda} = \frac{\sum_{i=1}^{n} x_i}{\lambda} - n = 0 \iff \boxed{\hat{\lambda}_{EMV} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{X}}$$

para maximizar

Son insesgados

$$E[\bar{X}] = \lambda \Rightarrow E[\bar{x} - \lambda] = 0$$

$$\frac{d\,\ell(\lambda, x)}{d\lambda} = \frac{x}{\lambda} - 1$$

$$\frac{d^2\ell(\lambda, x)}{d\lambda^2} = \frac{-x}{\lambda^2}$$

Alcanza CICR

$\tau(\lambda)' = 1$

$$CICR = \frac{1}{nI(\lambda)} \;, \quad I(\lambda) = -E\left[\frac{d^2\ell}{d\lambda^2}\right] = -E\left[\frac{-x}{\lambda^2}\right] = \frac{1}{\lambda^2}E[x] = \frac{1}{\lambda}$$

$$= \frac{1}{\frac{n}{\lambda}} = \frac{\lambda}{n} = var(\bar{x}) \;, \quad \hat{\lambda} = \tilde{\lambda} \text{ es UMVUE}$$

es el estimador por el método de momentos y de máxima verosimilitud del parámetro λ. Claramente, el estimador es insesgado y se puede mostrar que es de varianza mínima pues, por ejemplo, su varianza, $\text{var}(\bar{X}) = \lambda/n$ alcanza la **cota de inferior de Cramér-Rao**. Sin embargo, el estimador

$$\hat{p}_0 = e^{-\hat{\lambda}}$$

es un estimador sesgado de p_0. De hecho, puesto que $n\hat{\lambda} \sim \text{Po}(n\lambda)$, se puede emplear la función generadora de momentos y sus propiedades para mostrar que

$X = n\hat{\lambda} \iff -\hat{\lambda} = -\frac{X}{n}$

$$\theta = \mathbb{E}[\hat{p}_0] = \exp\left\{n\lambda(e^{-1/n} - 1)\right\}$$

$$m_X(t) = E[e^{tX}] = e^{\lambda(e^t - 1)} \Rightarrow E\left[e^{-t\frac{X}{n}}\right] = e^{\lambda(e^{-t/n} - 1)} = m_{-\frac{X}{n}}(t)$$

$$\Rightarrow m_{-\hat{\lambda}}(t) = E\left[e^{-t\frac{1}{n}\sum_{i=1}^{n} x_i}\right] \overset{ind}{=} E\left[e^{-t\frac{X}{n}}\right]^n = e^{n\lambda(e^{-t/n} - 1)} \quad \text{si } t = 1 \Rightarrow m_{-\hat{\lambda}}(1) = E[e^{-\hat{\lambda}}] = e^{n\lambda(e^{-1/n} - 1)} = E[\hat{p}_0]$$

$$\Rightarrow E\left[e^{-t\frac{2}{n}\sum_{i=1}^{n} x_i}\right] = e^{n\lambda(e^{-2t/n} - 1)} \;, \quad \text{si } t = 1 \Rightarrow E[\hat{p}_0^2] = e^{n\lambda(e^{-2/n} - 1)}$$

$$\hat{p}_0^2 = e^{-2\lambda} = e^{-\frac{2}{n}\sum x_i}$$

Para $var(\hat{p}_0)$

$$\Rightarrow \boxed{var(\hat{p}_0)} = e^{n\lambda(e^{-2/n} - 1)} - \left[e^{n\lambda(e^{-1/n} - 1)}\right]^2 = e^{n\lambda}\left[e^{(e^{-2/n} - 1)} - e^{2(e^{-1/n} - 1)}\right]$$

Por otro lado, si define $Y_i = \mathbb{1}_{\{0\}}(X_i)$, para $i = 1, \ldots, n$, entonces las Y_i son *v.a.i.i.d.* distribuidas $\mathrm{Ber}(p_0)$ y

$$\check{p}_0 = \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

es un estimador insesgado de p_0 con varianza $e^{-\lambda}(1 - e^{-\lambda})/n$.

$E[\bar{Y}] = P_0$, es insesgado

$$\mathrm{var}[\bar{Y}] = \frac{\sigma_Y^2}{n} = \frac{P_0(1-P_0)}{n} = \frac{e^{-\lambda}(1-e^{-\lambda})}{n}$$

Ahora bien, note que $T = \sum_{i=1}^{n} X_i$ es un estadístico suficiente para λ y $Y_1 = \mathbb{1}_{\{0\}}(X_1)$ es un estimador insesgado de $p_0 = e^{-\lambda}$. Se sigue entonces del **Teorema de Rao-Blackwell** que

$$\tilde{p}_0 = \mathbb{E}[Y_1|T] = \left(\frac{n-1}{n}\right)^T$$

es un estimador insesgado de menor varianza (¿cuál?) que $\check{p}_0$.

T es est. suf.

$$f(\underline{x}, \lambda) = \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} e^{-n\lambda}\, \mathbb{I}_{\{0,1,\ldots\}}(x) = \underbrace{\lambda^{\sum_{i=1}^{n} x_i} \cdot e^{-n\lambda}}_{g(S(\underline{x}), \lambda)} \cdot \underbrace{\frac{1}{\prod_{i=1}^{n} x_i!}\, \mathbb{I}_{\{0,1,\ldots\}}(x)}_{h(\underline{x})}$$

Por Teo de Factorización NF

$T = \sum_{i=1}^{n} x_i$ es estadístico suficiente

$Y_1 \sim Be(P_0) \Rightarrow E[Y_1] = P_0 \therefore$ es estimador insesgado para $P_0 = e^{-\lambda}$

Tarea 2

P.D. $\tilde{p}_0 = \left(\frac{n-1}{n}\right)^T$

Veamos que

ya que $\sum_{i=1}^{n} x_i \sim Poi(n\lambda)$, $\sum_{i=2}^{n} x_i \sim Poi(n\lambda)$

$$P[Y_1 = 1 \mid T = s] = P\left(x_1 = 0 \mid \sum_{i=1}^{n} x_i = s\right) \overset{\text{ind}}{=} \frac{P(x_1 = 0)\, P\left(\sum_{i=2}^{n} x_i = s\right)}{P\left(\sum_{i=1}^{n} x_i = s\right)} = \frac{(\cancel{e^{-\lambda}})(\cancel{e^{-(n-1)\lambda}}\{(n-1)\lambda\}^s / s!)}{\cancel{e^{-n\lambda}} (n\lambda)^s / \cancel{s!}}$$

$x_1 = 0$

$$= \frac{[(n-1)\lambda]^s}{[n\lambda]^s} = \left[\frac{n-1}{n}\right]^s$$

$$E[Y_1 \mid T] = 1 \cdot \left[\frac{n-1}{n}\right]^T + 0 = \left[\frac{n-1}{n}\right]^T$$

$E[E[Y_1 \mid T]] = E[Y_1] = P_1 = e^{-\lambda}$, es insesgado

$$\mathrm{var}(E[Y_1 \mid T]) = \mathrm{var}(Y_1) - E(\mathrm{var}(Y_1 \mid T)) = (e^{-\lambda})(1 - e^{-\lambda}) - E(\mathrm{var}(Y_1 \mid T))$$

$$\operatorname{var}(Y_1|T) = E[Y_1^2|T] - E[Y_1|T]^2 = \left[\frac{n-1}{n}\right]^T - \left[\frac{n-1}{n}\right]^{2T}, \text{ ya que } E[Y_1^2|T] = 1^2\left[\frac{n-1}{n}\right]^T + 0$$

$$= \left[\frac{n-1}{n}\right]^T \left[1 - \left[\frac{n-1}{n}\right]^2\right]$$

$$P[Y_1 = x | T = s] = \frac{\frac{\cancel{e^{-\lambda}}\lambda^x}{x!} \cdot \cancel{e^{-(n-1)\lambda}}[(n-1)\lambda]^s / \cancel{s!}}{\cancel{e^{-n\lambda}}(n\lambda)^s / \cancel{s!}} = \frac{\lambda^x}{x!}\left(\frac{n-1}{n}\right)^s$$

$$E[\operatorname{var}(Y_1|T)] = \sum_{y=0}^{1} \left[\frac{n-1}{n}\right]^T \left[1 - \left[\frac{n-1}{n}\right]^2\right] \frac{\lambda^y}{y!}\left(\frac{n-1}{n}\right)^T$$

$$T_2 = \sum_{i=2}^{n} x_i$$

$$= \left[1 - \left[\frac{n-1}{n}\right]^2\right]\left[\left[\frac{n-1}{n}\right]^{2T} + \left[\frac{n-1}{n}\right]^{2T_2} \cdot \lambda\right]$$

Por último, para mostrar que es UMVUE y por lo tanto de menor que $\check{P}_o$ mostraré que $\sum_{i=1}^{n} x_i = T$ es completo

Basta con mostrar que $\underline{X}$ es de la fam. exponencial

$$f(\underline{x}, \lambda) = \frac{\lambda^{\sum_{i=1}^n x_i}}{\sum_{i=1}^n x_i!} e^{-n\lambda}\, \mathbb{I}_{\{0,1,\dots\}}(x) = e^{\left(\underbrace{\log\lambda \sum_{i=1}^n x_i}_{c(\theta)\,T(x)} - \underbrace{\log \sum_{i=1}^n x_i!}_{S(x)} \underbrace{- n\lambda}_{d(\theta)}\right)} \mathbb{1}_{\{0,1,\dots\}}(x)$$

$\therefore$ $\underline{X}$ pertence a la familia exponencial

Por esto $\sum_{i=1}^{n} x_i = T$ es un estadístico completo

Entonces por el Teo de Lehmann Scheffé

$\hat{P}_0$ es el UMVUE

La CICR sería

$$\tau'(\lambda) = -e^{-\lambda} \Rightarrow \tau'(\lambda)^2 = e^{-2\lambda} \qquad \frac{d\ell(\lambda,x)}{d\lambda} = \frac{x_i}{\lambda} - n \qquad \frac{d^2\ell}{d\lambda} = -\frac{x_i}{\lambda^2}$$

$$I(\lambda) = -E\left[\frac{d^2\ell}{d\lambda}\right] = -E\left[\frac{X}{\lambda^2}\right] = \frac{1}{\lambda}$$

$$\Rightarrow CICR = \frac{\tau'(\lambda)}{nI(\lambda)} = \frac{\lambda e^{-2\lambda}}{n} \qquad \operatorname{var}(\hat{p}_0) \geq CICR$$

Amplitud de los intervalos de probabilidad

Sea X una población que es modelada razonablemente por una distribución normal centrada en cero y con desviación estándar σ. Complete la siguiente tabla donde q_i y q_s denotan los límites inferior y superior del intervalo de probabilidad $1-\alpha$ y $w = q_s - q_i$ es la amplitud del intervalo, para los distintos valores de σ. Se construirán dos intervalos bajo criterios distintos: a) **colas del mismo peso** - cada una de las colas tiene $\alpha/2$ de probabilidad; b) **amplitud mínima** - los límites se escogen de manera que w sea lo más pequeño posible, manteniendo el nivel $1-\alpha$ de probabilidad.

De manera similar si Y denota una población que es modelada por una distribución χ^2 con n grados de libertad, complete la parte correspondiente de la tabla para los distintos grados de libertad.

$X \sim \mathrm{N}(0,\sigma)$		colas del mismo peso			amplitud mínima		
σ	$1-\alpha$	q_i	q_s	w	q_i	q_s	w
	0.90	-1.645	1.645	3.290			
1.0	0.95	-1.960	1.960	3.920	Igual		
	0.99	-2.576	2.576	5.152			
	0.90	-3.290	3.290	6.579			
2.0	0.95	-3.920	3.920	7.840	Igual		
	0.99	-5.152	5.152	10.303			
	0.90	-8.224	8.224	16.449			
5.0	0.95	-9.800	9.800	19.600	Igual		
	0.99	-12.879	12.879	25.758			

$Y \sim \chi^2_n$		colas del mismo peso			amplitud mínima		
n	$1-\alpha$	q_i	q_s	w	q_i	q_s	w
	0.90	1.145	11.070	9.925	0.476	9.434	8.957
5	0.95	0.831	12.833	12.001	0.296	11.191	10.895
	0.99	0.412	16.750	16.338	0.101	15.127	15.026
	0.90	3.940	18.307	14.367	3.017	16.711	13.693
10	0.95	3.247	20.483	17.236	2.414	18.860	16.447
	0.99	2.156	25.188	23.032	1.498	23.533	22.035
	0.90	10.851	31.410	20.560	9.786	29.876	20.090
20	0.95	9.591	34.170	24.579	8.584	32.607	24.023
	0.99	7.434	39.997	32.563	6.544	38.340	31.845

Comparación de medias y varianzas de poblaciones normales

Para la solución de un problema mediante procedimientos numéricos se disponen de dos algoritmos que compiten en eficiencia. La siguiente tabla muestra los tiempos de procesamiento (en segundos) que se llevan ambos algoritmos en 25 distintas computadoras (diferentes CPU's, RAM, discos de almacenamiento y versiones del sistema operativos Linux). Responda los siguientes incisos, suponiendo que el tiempo de procesamiento entre las distintas computadoras es independiente y razonablemente modelado por la distribución normal.

	Algoritmo			Algoritmo	
Computadora	1	2	Computadora	1	2
1	10.53	11.24	14	11.12	11.85
2	12.22	11.53	15	9.84	10.58
3	9.01	8.82	16	11.47	11.23
4	11.31	11.20	17	10.89	8.99
5	9.90	10.00	18	10.53	12.50
6	8.17	9.07	19	9.54	8.02
7	9.92	9.21	20	11.03	13.87
8	9.61	9.88	21	10.43	13.18
9	9.68	9.00	22	10.93	12.31
10	8.86	9.84	23	11.70	12.61
11	10.85	11.40	24	10.58	12.19
12	7.97	8.41	25	9.96	10.50
13	8.77	9.70			

1. Construya un intervalo del 90 % de confianza para el tiempo medio de procesamiento μ_1 del algoritmo 1 si sabe que su desviación estándar es de 1 segundo.
2. Construya el correspondiente intervalo para el tiempo medio μ_2 del algoritmo 2, con un nivel de 0.95 de confianza reconociendo que no conoce su desviación estándar.
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 σ_i de cada uno de los algoritmos y concluya.
4. Responda el inciso anterior pero con base a un intervalo del 95 % de confianza para el cociente de varianzas $\theta = \sigma_1^2/\sigma_2^2$.

Bibliografía

Mood, Graybill and Boes (1974). *Introduction to the Theory of Statistics*. 3rd. Ed. McGraw-Hill.

1. Construya un intervalo del 90 % de confianza para el tiempo medio de procesamiento μ_1 del algoritmo 1 si sabe que su desviación estándar es de 1 segundo.

IC de nivel $1-\alpha=0.90$ para la media conocida la varianza

$$\left(\bar{X}_n - z_{1-\alpha/2}\sigma/\sqrt{n},\ \bar{X}_n + z_{1-\alpha/2}\sigma/\sqrt{n}\right)$$

1- IC para alg. 1 con $\alpha=0.10$ y $\sigma=1$

IC = (9.8638, 10.5218)

2. Construya el correspondiente intervalo para el tiempo medio μ_2 del algoritmo 2, con un nivel de 0.95 de confianza reconociendo que no conoce su desviación estándar.

IC de nivel $1-\alpha=0.95$ para la media con varianza desconocida

$$\left(\bar{X}_n - t_{1-\alpha/2;n-1}S/\sqrt{n},\ \bar{X}_n + t_{1-\alpha/2;n-1}S/\sqrt{n}\right)$$

IC = (10.0281 , 11.3423)

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 σ_i de cada uno de los algoritmos y concluya.

Ic de nivel $1-\alpha=0.95$ para varianza

$$\left(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2;\ n-1}},\ \frac{(n-1)S^2}{\chi^2_{\alpha/2;\ n-1}}\right)$$

Alg. 1	Alg. 2
IC = (0.8481, 1.5111)	IC = (1.2431, 2.2147)

Podrían llegar a tener la misma variabilidad porque los intervalos se intersectan, aunque el rango que intersecta es menor al resto de ambos intervalos.

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

Sean $\boldsymbol{X}_{n_1} = (X_1, \ldots, X_{n_1})$ y $\boldsymbol{Y}_{n_2} = (Y_1, \ldots, Y_{N_2-1})$ muestras aleatorias de poblaciones $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$ y $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$ independientes, respectivamente. Por construir un intervalo de confianza para $\theta = \sigma_1^2/\sigma_2^2$.

Sean $S_1^2 = \sum_{i=1}^{n_1}(X_i - \bar{X})^2/(n_1 - 1)$ y $S_2^2 = \sum_{j=1}^{n_2}(Y_j - \bar{Y})^2/(n_2 - 1)$. Luego, $\dfrac{n_1 - 1}{\sigma_1^2} S_1^2 \sim \chi^2_{n_1-1}$ y $\dfrac{n_2 - 1}{\sigma_2^2} S_2^2 \sim \chi^2_{n_2-1}$ independientes por lo que

$$Q = \frac{\frac{n_1-1}{\sigma_1^2} S_1^2/(n_1 - 1)}{\frac{n_2-1}{\sigma_2^2} S_2^2/(n_2 - 1)} = \frac{1}{\theta} \frac{S_1^2}{S_2^2} \sim F_{\nu_1;\nu_2}$$

con $\nu_1 = n_1 - 1$ y $\nu_2 = n_2 - 1$. Se sigue entonces que para 1_1 y q_2 se tiene que

$$\gamma = \mathbb{P}\left(q_1 \le Q \le q_2\right) = \mathbb{P}\left(\frac{S_1^2/S_2^2}{q_2} \le \theta \le \frac{S_1^2/S_2^2}{q_1} \right) = 1 - \alpha$$

Finalmente,

$$\left(\frac{S_1^2/S_2^2}{F(1 - \alpha/2; n_1 - 1, n_2 - 1)}, \ \frac{S_1^2/S_2^2}{F(\alpha/2; n_1 - 1, n_2 - 1)} \right)$$

constituye un intervalos de $100(1 - \alpha)$ % de confianza para $\theta = \sigma_1^2/\sigma_2^2$.

I_c de nivel $1 - \alpha = 0.95$ para el cociente de varianzas

$IC = (0.2051, 1.0564)$

Si podrían ser iguales con este intervalo de 95% de confianza

ya que incluye al 1 y si $\frac{\sigma_1^2}{\sigma_2^2} = 1 \Rightarrow \sigma_1^2 = \sigma_2^2$