Estimadores II

Ejercicio 1

Si X es una v.a. uniforme discreta entre 1 y N:

$$E(X) = \frac{1}{N}\frac{N(N+1)}{2} = \frac{N+1}{2}$$

Por lo que el estimador de momentos de N es $\hat{N}_1 = 2\overline{X}_N - 1$. El estimador de máxima verosimilitud (EMV) es $\hat{N}_2 = \text{máx}\, X_i$.

En este caso, de una muestra de tamaño dos, estos quedan

$$\hat{N}_1 = X_1 + X_2 - 1, \quad \hat{N}_2 = \text{máx}\{X_1, X_2\}.$$

Me quedaría con la estimación de Beto ya que Beto usa como estimación el EMV de N, mientras que usando el método de los momentos la estimación sería $X_1 + X_2 - 1$, que no es lo que hace Ana.

Supongamos que Ana se ha equivocado y que simplemente se ha olvidado de restar 1 (su estimación en realidad es 20 y no 21).

Por un lado, vemos que $\mathrm{E}\left(\hat{N}_1\right) = 2(N+1)/2 - 1 = N$, por lo que $\hat{N}_1$ es insesgado. Este es un punto a favor de Ana. Además

$$\mathrm{ECM}(\hat{N}_1) = \mathrm{Var}\left(\hat{N}_1\right) = \mathrm{Var}\,(X_1) + \mathrm{Var}\,(X_2) - 1 = (N^2-1)/6.$$

Para hallar la distribución del máximo de X_1 y X_2, observar la tabla:

		X_2			
		1	2	$\cdots$	N
X_1	1	1	2	$\cdots$	N
	2	2	2	$\cdots$	N
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	N	N	N	$\cdots$	N

de donde vemos que la f.p.p. de $\hat{N}_2$ es

k	1	2	3	$\cdots$	N
$p(k)$	$1/N^2$	$3/N^2$	$5/N^2$	$\cdots$	$(2N-1)/N^2$

La esperanza es

$$\mathrm{E}\left(\hat{N}_2\right) = \frac{1}{N^2}\sum_{i=1}^{N}(2i-1)i = N\frac{(N+1)(4N-1)}{6N^2}.$$

Es decir, $\hat{N}_2$ es sesgado. Si N no es muy chico (sabemos que es mayor que 18), entonces $\mathrm{E}\left(\hat{N}_2\right) \approx 2N/3$, lo cual muestra que $\hat{N}_2$ tiene tendencia a estimar N por abajo.

Haciendo más cuentas se puede ver que

$$\mathrm{ECM}(\hat{N}_2) = \frac{(N-1)\left(N^2-N+1\right)}{6N}$$

Aquí surge una cuestión interesante. Si N es grande, el error cuadrático medio de $\hat{N}_2$ y $\hat{N}_1$ son equivalentes (ambos equivalentes a $N^2/6$). Luego, es mejor quedarse con la estimación de momentos (la de Ana si no se hubiera olvidado de restar 1).

Ejercicio 2

La esperanza es

$$\mathrm{E}\left(\overline{X}-\overline{Y}\right)=\mathrm{E}\left(\overline{X}\right)-\mathrm{E}\left(\overline{Y}\right)=\mu_A-\mu_B$$

por lo que el sesgo es cero.

El error cuadrático medio es igual a la varianza, que vale

$$\mathrm{Var}\left(\overline{X}-\overline{Y}\right)=\mathrm{Var}\left(\overline{X}\right)+\mathrm{Var}\left(\overline{Y}\right)=\frac{\sigma_A^2}{25}+\frac{\sigma_B^2}{30}=0.223$$

Ejercicio 3

1. $$L(\sigma)=\prod_{i=1}^{n}\frac{1}{2\sigma}e^{-\frac{|X_i|}{\sigma}}=\frac{1}{(2\sigma)^n}e^{-\frac{\sum_{i=1}^n|X_i|}{\sigma}}$$

   Aplicando logaritmo:

   $$h(\sigma)=-n\log(2\sigma)-\frac{\sum_{i=1}^n|X_i|}{\sigma}$$

   Derivando respecto a σ:

   $$h'(\sigma)=-n\frac{2}{2\sigma}+\frac{\sum_{i=1}^n|X_i|}{\sigma^2}$$

   De donde $\hat{\sigma}=\frac{\sum_{i=1}^n|X_i|}{n}$ es un punto crítico máximo.

2. $$E(\hat{\sigma})=E\left(\frac{\sum_{i=1}^n|X_i|}{n}\right)=\frac{\sum_{i=1}^n E(|X_i|)}{n}=E(|X_i|)=\int_{-\infty}^{infty}|x|\frac{e^{-\frac{|x|}{\sigma}}}{2\sigma}dx=\int_{-0}^{\infty}x\frac{e^{-\frac{x}{\sigma}}}{\sigma}dx=\sigma$$

   Otra forma de calcular $E(|X_i|)$ es observando que $|X_i|$ es una v.a. exponencial de parámetro $\frac{1}{\sigma}$.

   $$\begin{aligned}Var(\hat{\sigma})=E(\hat{\sigma}^2)-E(\hat{\sigma})^2&=E\left(\left(\frac{\sum_{i=1}^n|X_i|}{n}\right)^2\right)-\sigma^2\\&=\frac{\sum_{i=1}^n E(|X_i|^2)+\sum_{i\neq j}2E(|X_i||X_j|)}{n^2}-\sigma^2\\&=\frac{nE(|X_i|^2)}{n^2}+\frac{\frac{n(n+1)}{2}2E(|X_i|)E(|X_j|)}{n^2}-\sigma^2\\&=\frac{E(|X_i|^2)}{n}+E(|X_i|)^2+\frac{1}{n}E(|X_i|)^2-\sigma^2=\frac{E(|X_i|^2)}{n}+\frac{\sigma^2}{n}\end{aligned}$$

   Ahora:

   $$E(|X_i|^2)=\int_{-\infty}^{\infty}|x|^2\frac{e^{-\frac{|x|}{\sigma}}}{2\sigma}dx=\int_{-0}^{\infty}x^2\frac{e^{-\frac{x}{\sigma}}}{\sigma}dx=2\sigma^2$$

Entonces:

$$Var(\hat{\sigma}) = \frac{2\sigma^2}{n} - \frac{\sigma^2}{n} = \frac{\sigma^2}{n}$$

Otra forma de calcular esta varianza es:

$$Var(\hat{\sigma}) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} |X_i|\right) = \frac{n}{n^2} Var(|X_i|) = \frac{\sigma^2}{n}$$

donde hemos usado que las variables X_i son independientes, igualmente distribuidas, y que $Var(|X_i|) = \sigma^2$ por ser $|X_i|$ exponencial.

3. $ECM(\hat{\sigma}) = Var(\hat{\sigma})$ pues su sesgo es 0, y como $Var(\hat{\sigma}) = \frac{\sigma^2}{n} \underset{n}{\rightarrow} 0$ es consistente.
4. Sea $\tilde{\sigma}$ la estimación pedida, sustituyendo en la expresión hallada en (1), entonces $\tilde{\sigma} = 3,403$.

Ejercicio 4

1. $Rec(X) = \mathbb{N}^+$.

$$P(X = k) = P([T] + 1 = k) = P(k - 1 \leq T < k) = \int_{k-1}^{k} \lambda e^{-\lambda t} dt = -e^{-\lambda t}\Big|_{t=k-1}^{t=k} =$$

$$= -e^{-\lambda k} + e^{-\lambda(k-1)} = e^{-\lambda(k-1)}(1 - e^{-\lambda}) = (1-p)^{k-1} p \Rightarrow X \sim Geo(p)$$

donde $p = 1 - e^{-\lambda}$.

2. $\hat{\mu} = \overline{X}_n$ es insesgado por ser $E(\overline{X}_n) = \mu$, es consistente por la LFGN, y es asintóticamente normal por el TCL.
3. Si $X \sim Geo(p) \Rightarrow E(X) = \frac{1}{p} \Rightarrow \mu = \frac{1}{p} = \frac{1}{1-e^{-\lambda}}$

$$\Rightarrow 1 - e^{-\lambda} = \frac{1}{\lambda} \Rightarrow e^{-\lambda} = 1 - \frac{1}{\lambda} \Rightarrow -\lambda = \log\left(1 - \frac{1}{\mu}\right) = \log\left(\frac{\mu - 1}{\mu}\right)$$

$$\Rightarrow \lambda = \log\left(\frac{\mu}{\mu - 1}\right)$$

Por tanto, un estimador consistente de λ es:

$$\hat{\lambda} = \log\left(\frac{\overline{X}_n}{\overline{X}_n - 1}\right)$$

ya que $\log\left(\frac{x}{x-1}\right)$ es continua si $x > 1$.

4. $\overline{X}_n = 54,7$, entonces $\hat{\lambda} = 0,01845$.

Ejercicio 5

1. La esperanza es $\int_\theta^\infty xe^{-(x-\theta)}dx = \theta + 1$, por lo que el estimador de momentos es $\hat{\theta} = \overline{X}_n - 1$.
2. El sesgo de $\hat{\theta}$ es igual a
$$\mathrm{E}\left(\hat{\theta}\right) - \theta = (\theta+1) - 1 - \theta = 0,$$
por lo que $\hat{\theta}$ es insesgado.

   La varianza de $p(x;\theta)$ es
$$\int_\theta^\infty (x-(\theta+1))^2 e^{-(x-\theta)}dx = 1$$
   El error cuadrático medio de $\hat{\theta}$ es igual a la varianza y vale entonces
$$\mathrm{ECM}(\hat{\theta}) = \frac{1}{n}.$$
3. La función de verosimilitud es
$$L(\theta) = \begin{cases} e^{-\sum_{i=1}^n X_i} e^{n\theta} & \text{si } \min\{X_i\} \geq \theta \\ 0 & \text{si no} \end{cases}$$
   Es una función creciente hasta $\min\{X_i\}$, y luego vale cero, por lo que su máximo se da en $\hat{\theta}_2 = \max\{X_i\}$.
4. La fda de $\hat{\theta}_2$ es
$$\begin{aligned} F(t) &= 1 - P\left(\hat{\theta}_2 > t\right) \\ &= 1 - P(X_1 > t, \ldots, X_n > t) \\ &= 1 - \prod_{i=1}^n P(X_i > t) \\ &= \begin{cases} 0 & \text{si } t < \theta \\ 1 - e^{-n(t-\theta)} & \text{si } t \geq \theta \end{cases} \end{aligned}$$
   de donde $n\hat{\theta}_2$ tiene densidad $p(x;n\theta)$.

   Luego el sesgo es
$$\mathrm{E}\left(\hat{\theta}_2\right) - \theta = \frac{n\theta+1}{n} - \theta = \frac{1}{n}.$$
   La varianza es
$$\mathrm{Var}\left(\hat{\theta}_2\right) = \frac{1}{n^2}$$
   por lo que el error cuadrático medio es $2/n^2$.
5. Es preferible el estimador $\hat{\theta}_2$, ya que aunque sea segado, el sesgo tiende a cero cuando $n \to \infty$, y el error cuadrático medio tiende a cero más rápido que el de $\hat{\theta}_1$ ($1/n^2$ contra $1/n$).

Ejercicio 6

1. La probabilidad de que $X = 0$ es $e^{-\mu}$. Por la LGN, la frecuencia relativa de veces que se observa cero tiende a la probabilidad, de donde $Y/n \xrightarrow{d} e^{-\mu}$. Es decir, Y/n es un estimador consistente de $e^{-\mu}$. Como la función logaritmo es continua, $\hat{\mu}$ es un estimador de $-\ln(e^{-\mu}) = \mu$.
2. Podemos aproximar
$$\frac{Y}{n} = e^{-\hat{\mu}} \approx 1 - \hat{\mu}.$$
Tomando varianza a ambos lados deducimos que $\mathrm{Var}(\hat{\mu}) \approx p_0(1-p_0)/n$.

Ejercicio 7

1. La función de verosimilitud es
$$L(\theta) = \left(\frac{2+\theta}{4}\right)^{1997} \cdot \left(\frac{1-\theta}{4}\right)^{906} \cdot \left(\frac{1-\theta}{4}\right)^{904} \cdot \left(\frac{\theta}{4}\right)^{32}$$
Tomando logaritmo
$$\ell(\theta) = 1997\ln(2+\theta) + 1810\ln(1-\theta) + 32\ln(\theta) - 3839\ln 4$$
Derivando
$$\ell'(\theta) = \frac{1997}{2+\theta} - \frac{1810}{1-\theta} + \frac{32}{\theta} = 0$$
que es equivalente a
$$1997(1-\theta)\theta - 1810(2+\theta)\theta + 32(2+\theta)(1-\theta) = 0$$
Reordenando
$$-3839\theta^2 - 1655\theta + 64 = 0$$
Cuya solución positiva es
$$\theta = \frac{1655 \pm \sqrt{1655^2 + 4.3839 \cdot 64}}{2 \cdot 3839} = 0.0357$$
2. Como $X_A \sim \mathrm{Bin}(n, p_A)$, tenemos
$$E\left(\hat{\theta}_1\right) = E\left(\frac{4X_A}{n} - 2\right) = \frac{4}{n}E(X_A) - 2 = \frac{4}{n}np_A - 2 = \theta$$
Del mismo modo, como $X_D \sim \mathrm{Bin}(n, p_D)$, tenemos
$$E\left(\hat{\theta}_2\right) = E\left(\frac{4X_D}{n}\right) = \frac{4}{n}E(X_D) = \frac{4}{n}np_D = \theta$$
Para las varianzas tenemos
$$\mathrm{Var}\left(\hat{\theta}_1\right) = \frac{16}{n^2}\mathrm{Var}(X_A) = \frac{16p_A(1-p_A)}{n} = \frac{4-\theta^2}{n}$$

y

$$\operatorname{Var}\left(\hat{\theta}_2\right)=\frac{16}{n^2}\operatorname{Var}(X_D)=\frac{16p_D(1-p_D)}{n}=\frac{\theta(4-\theta)}{n}$$

Por último

$$\text{Eficiencia }=\frac{\operatorname{Var}\left(\hat{\theta}_1\right)}{\operatorname{Var}\left(\hat{\theta}_2\right)}=\frac{4-\theta^2}{\theta(4-\theta)}$$

Notar que si θ es chico $\hat{\theta}_2$ es más eficiente $\hat{\theta}_1$. Ya que estimamos en nuestro caso θ chico, nos conviene usar $\hat{\theta}_2$ como estimador.

Ejercicio 8

1. Por la fórmula de la probabilidad total

$$\begin{aligned}P(MM)&=P(MM|\text{Idénticos})P(\text{Idénticos})+P(MM|\text{No idénticos})P(\text{Idénticos})\\&=\frac{1}{2}\alpha+\frac{1}{4}(1-\alpha)\\&=\frac{2\alpha+1-\alpha}{4}\\&=\frac{1+\alpha}{4}\end{aligned}$$

Análogamente se calcula $P(HH)$.

Por otro lado

$$\begin{aligned}P(HM)&=1-P(MM)-P(HH)\\&=1-2P(MM)\\&=1-2\left(\frac{1+\alpha}{4}\right)\\&=\frac{1-\alpha}{2}\end{aligned}$$

2.

$$\begin{aligned}\ell(\alpha)&=\text{ Constante }+(n_1+n_2)\log(1+\alpha)+n_3\log(1-\alpha)\\\ell'(\alpha)&=\frac{n_1+n_2}{1+\alpha}-\frac{n_3}{1-\alpha}\\\hat{\alpha}&=\frac{n_1+n_2-n_3}{n_1+n_2+n_3}\end{aligned}$$

Observar que $n_1+n_2+n_3=n$.

Para calcular la varianza, observar primero que

$$\begin{aligned}\hat{\alpha}&=\frac{n_1+n_2-n_3}{n}\\&=\frac{n_1+n_2-(n-n_1-n_2)}{n}\\&=\frac{2(n_1+n_2)-n}{n}\end{aligned}$$

Entonces

$$\begin{aligned}
\operatorname{Var}(\hat{\alpha}) &= \operatorname{Var}\left(\frac{2(n_1+n_2)-n}{n}\right) \\
&= \frac{4}{n^2}\operatorname{Var}(n_1+n_2) \\
&= \frac{4}{n^2}\left(\operatorname{Var}(n_1)+\operatorname{Var}(n_2)+2\operatorname{Cov}(n_1,n_2)\right)
\end{aligned}$$

Notar que n_1 y n_2 son binomiales con probabilidad de éxito $(1+\alpha)/4$. Luego

$$\operatorname{Var}(n_1)=\operatorname{Var}(n_2)=n\left(\frac{1+\alpha}{4}\right)\left(\frac{3-\alpha}{4}\right)$$

Definimos $Y_i=1$ si el i-ésimo par de gemelos es HH y 0 si no, y $X_i=1$ si el i-ésimo par de gemelos es MM y 0 si no. Claramente $n_1=\sum_{i=1}^n Y_i$ y $n_2=\sum_{i=1} X_i$, y también $Y_iX_i=0$ ya que un par de gemelos no puede ser HH y MM a la vez. Usando estas definiciones, tenemos

$$\begin{aligned}
\operatorname{Cov}(n_1,n_2) &= \operatorname{E}(n_1n_2)-\operatorname{E}(n_1)\operatorname{E}(n_2) \\
&= \operatorname{E}\left(\left(\sum_{i=1}^n Y_i\right)\left(\sum_{j=1}^n X_j\right)\right)-\frac{n(1+\alpha)}{4}\frac{n(1+\alpha)}{4} \\
&= \operatorname{E}\left(\sum_{i=1}^n Y_iX_i+\sum_{i\neq j} Y_iX_j\right)-n^2\left(\frac{1+\alpha}{4}\right)^2 \\
&= \operatorname{E}\left(\sum_{i=1}^n 0\right)+\operatorname{E}\left(\sum_{i\neq j} Y_iX_j\right)-n^2\left(\frac{1+\alpha}{4}\right)^2 \\
&= 0+\left(n^2-n\right)\operatorname{E}(Y_i)\operatorname{E}(X_j)-n^2\left(\frac{1+\alpha}{4}\right)^2 \\
&= \left(n^2-n\right)\left(\frac{1+\alpha}{4}\right)^2-n^2\left(\frac{1+\alpha}{4}\right)^2 \\
&= -n\left(\frac{1+\alpha}{4}\right)^2
\end{aligned}$$

Sustituyendo estos resultados de nuevo en la expresión para la varianza, tenemos

$$\begin{aligned}
\operatorname{Var}(\hat{\alpha}) &= \frac{4}{n^2}\left(n\left(\frac{1+\alpha}{4}\right)\left(\frac{3-\alpha}{4}\right)+n\left(\frac{1+\alpha}{4}\right)\left(\frac{3-\alpha}{4}\right)+2\left(-n\left(\frac{1+\alpha}{4}\right)^2\right)\right) \\
&= \frac{1}{n}\left[\frac{(1+\alpha)(3-\alpha)}{2}+\frac{(1+\alpha)^2}{2}\right] \\
&= \frac{(1+\alpha)4}{2n} \\
&= \frac{2(1+\alpha)}{n}
\end{aligned}$$