Pauta Aux 2

P1 | $X_1, \ldots, X_n \sim \text{Geom}(p)$ $\quad p \in (0,1]$ desconocido

PMF: $f(k;p) = (1-p)^{k-1} p \quad \forall k \in \mathbb{N}_*$

a) → método de los momentos

Si $X \sim \text{Geom}(p) \Rightarrow \mathbb{E}[X] = \frac{1}{p}$ tenemos una forma de estimar p!

Utilizamos $\hat{\mathbb{E}}(x_1, \ldots, x_n) = \frac{1}{n}\sum^n x_i =: \bar{x} = \frac{1}{p}$ la esperanza empírica de la muestra $(x_1, \ldots, x_n)$, y por tanto

$$\boxed{\hat{p}_{MOM} = \frac{1}{\bar{x}}}$$

b) → Calcular estimador máxima verosimilitud

La verosimilitud es $\mathcal{L}(p|x) = \prod_k^n f(x_k;p) = \prod_k^n (1-p)^{x_k - 1} p$

$= (1-p)^{\sum^n x_k - n} p^n = (1-p)^{n\bar{x} - n} p^n$

$\Rightarrow \mathcal{L}(p|x) = (1-p)^{n\bar{x}-n} p^n$

La log-verosimilitud entonces es:

$$\ell(p|x) = \ln \mathcal{L}(p|x) = n(\bar{x}-1)\ln(1-p) + n \ln p$$

Derivémoslo c/r a p en el intervalo $(0,1)$

$$\frac{\partial}{\partial p}\ell(p|x) = n(\bar{x}-1)\frac{-1}{1-p} + n\frac{1}{p}$$

imponiendo $\frac{\partial}{\partial p}\ell(p|x) \overset{!}{=} 0$ se tiene que:

$$n\frac{1}{p} = n(\bar{x}-1)\frac{1}{1-p} \Rightarrow (1-p) = \bar{x}p - p \Rightarrow \boxed{\hat{p}_{MLE} = \frac{1}{\bar{x}}}$$

c) Aquí hubo una errata en el aux. para verlo, empecemos analizando lo siguiente:

$$\mathbb{E}[\bar{X}] = \mathbb{E}\left[\frac{1}{n}\sum^n X_i\right] = \frac{1}{n}\sum^n \mathbb{E}[X_i] = \frac{1}{n}\sum^n \frac{1}{p} = \frac{1}{p}$$

Notemos que la función $x \mapsto \frac{1}{x}$ es convexa. Por tanto, por la desig. de Jensen:

$$\mathbb{E}[\hat{p}_{MLE}] = \mathbb{E}\left[\frac{1}{\bar{X}}\right] > \frac{1}{\mathbb{E}[\bar{X}]} = p$$

/pues $\frac{\partial^2}{\partial x^2}\left[\frac{1}{x}\right] = \frac{1}{x^3} > 0 \; \forall x > 0$

y por tanto es sesgado.

P2| $X_1, \ldots, X_n \sim \text{Unif}[\theta, \theta+1]$ con $\theta \in \mathbb{R}$ desconocido

Recordemos que su función de distr. es

$$f(x; a, b) = \frac{1}{b-a} \mathbb{1}_{x \in [a,b]} \quad \forall x \in \mathbb{R}$$

Dada $(x_1 - x_n)$ muestra, la verosimilitud es:

$$\mathcal{L}(\theta | x) = \prod^n f(x_i; \theta, \theta+1) = \prod^n \mathbb{1}_{x_i \in [\theta, \theta+1]}$$

Recordando que si p y q son dos proposiciones, entonces $\mathbb{1}_p \cdot \mathbb{1}_q = \mathbb{1}_{p \wedge q}$

así:

$$\mathcal{L}(\theta | x) = \prod^n \mathbb{1}_{x_i \in [\theta, \theta+1]} = \mathbb{1}_{x_1 - x_n \in [\theta, \theta+1]}$$

Sin embargo, hay que pensar en la verosimilitud como fn de θ, y que los x_i están fijos, o "inyectados" en la verosimilitud.

Tratemos de trabar esa expresión:

$$x_1 - x_n \in [\theta, \theta+1] \iff \forall k = 1:n \quad \theta \leq x_k \wedge \theta + 1 \geq x_k$$

$$\iff \theta \leq \min_k x_k \wedge \theta + 1 \geq \max_k x_k$$

$$\iff \theta \in [\max_k x_k - 1, \min_k x_k]$$

Entonces:

$$\mathcal{L}(\theta | x) = \mathbb{1}_{x_1 - x_n \in [\theta, \theta+1]} = \mathbb{1}_{\theta \in [\max_k x_k - 1, \min_k x_k]}$$

y por tanto

$$\arg\max_\theta \mathcal{L}(\theta | x) = [\max_k x_k - 1, \min_k x_k]$$

P3]

a) N pob. total

K se capturan y marcan

n se capturan dsps de un mes

$P_N(k)$ prob. de que dsps de un mes k de los n estén marcados

$$P_N(k) = \frac{\underbrace{\binom{K}{k}}_{\text{formas de sacar k marcados entre K animales marcados.}}\binom{N-K}{n-k}}{\binom{N}{n}}$$

$\binom{N-K}{n-k}$ } formas de sacar n-k no marcados de entre N-K

$\binom{N}{n}$ } formas de sacar n animales de entre N animales.

Esto es: $P_N \sim \text{Hiper Geom}(N, \tilde{K}, \tilde{n})$ (con $\tilde{K}$ y $\tilde{n}$ "fijos", pero N es el parámetro de los modelos)

b) $$\frac{P_N(k)}{P_{N-1}(k)} = \frac{\cancel{\binom{K}{k}}\binom{N-K}{n-k}}{\binom{N}{n}} \frac{\binom{N-1}{n}}{\cancel{\binom{K}{k}}\binom{N-1-K}{n-k}} = \frac{\binom{N-1}{n}}{\binom{N}{n}} \frac{\binom{N-K}{n-k}}{\binom{N-1-K}{n-k}}$$

$$= \frac{\frac{(N-1)!}{\cancel{n!}(N-1-n)!}}{\frac{N!}{\cancel{n!}(N-n)!}} \cdot \frac{\frac{(N-K)!}{\cancel{(n-k)!}(N-K-n+k)!}}{\frac{(N-K-1)!}{\cancel{(n-k)!}(N-K-n+k-1)!}}$$

$$= \frac{1}{N}\cdot(N-n)\cdot(N-K)\cdot\frac{1}{N-K-n+k} = \frac{(N-K)(N-n)}{N(N-K-n+k)}$$ //

$P_N(k) \geq P_{N-1}(k) \Leftrightarrow (N-K)(N-n) \geq N(N-K-n+k)$

$\Leftrightarrow \cancel{N^2} \cancel{-KN} \cancel{-nN} + nK \geq \cancel{N^2} \cancel{-NK} \cancel{-nN} + Nk$

$\Leftrightarrow nK \geq Nk \Leftrightarrow N \leq \frac{nK}{k}$

$\therefore P_{\bullet}(k)$ creciente con $N \in \{0, \_, \lfloor \frac{Kn}{k} \rfloor\}$

decreciente en $N \in \{\lfloor \frac{Kn}{k} \rfloor + 1, \_\_ \}$

$\Rightarrow \hat{N}_{MLE} = \lfloor \frac{Kn}{k} \rfloor$

$K = 10$, $n = 5$, $k = 1$ } $\hat{N}_{MLE} = 50$

La interpretación es que, digamos, originalmente habían $N = 100$ animales, pero se tomó una muestra de $K = 10$ delfines, entonces en la segunda vez se hará un re-escalamiento de K, por medio del segundo vistazo. Este factor de reescalamiento de K es $\frac{n}{k}$, que se

c) Aquí: $K = 10$, $n = 5$, $k = 1$

$$\hat{N}_{MLE} = K \cdot \frac{n}{k} = 10 \cdot \frac{5}{1} = 50.$$

P4) $X_1 — X_n \sim \text{Gamma}(k, \lambda)$ con, $k>0$ y $\lambda>0$ desconocidos

su pdf: $f(x; k, \lambda) = \frac{1}{\Gamma(k)\lambda^k} x^{k-1} e^{-\frac{x}{\lambda}} \mathbb{1}_{x>0}$

Se puede calcular que para $X \sim \text{Gamma}(k, \lambda)$, su función generadora de momentos es

$$M_X(t) = \mathbb{E}[e^{tX}] = (1-\lambda t)^{-k} \quad \text{para } t < \frac{1}{\lambda}$$

Derivando:

$$\frac{d}{dt} M_X(t) = (1-\lambda t)^{-k-1}(-k)\cdot(-\lambda) = \lambda k (1-\lambda t)^{-k-1}$$

$$\frac{d^2}{dt^2} M_X(t) = \lambda^2 k (1-\lambda t)^{-k-2} (-k-1)(-\lambda) = \lambda^2 k (k+1)(1-\lambda t)^{-k-2}$$

Generando los momentos:

$$\mathbb{E}[X] = \frac{d}{dt} M_X(t) \Big|_{t=0} = \lambda k$$

$$\mathbb{E}[X^2] = \frac{d^2}{dt^2} M_X(t) \Big|_{t=0} = \lambda^2 k (k+1)$$

Ahora tenemos un sistema de Ecs. para λ y k. usando la esperanza empírica:

$$\frac{1}{n} \sum^n x_i =: \overline{x} = \lambda k$$

$$\frac{1}{n} \sum^n x_i^2 =: \overline{x^2} = \lambda^2 k (k+1)$$

Notemos que

$$\overline{x^2} = \lambda^2 k (k+1) = \underbrace{\lambda^2 k^2}_{\overline{x}^2} + \underbrace{\lambda^2 k}_{\lambda \overline{x}} = \overline{x}^2 + \lambda \overline{x} \Rightarrow \hat{\lambda}_{MOM} = \frac{\overline{x^2} - \overline{x}^2}{\overline{x}}$$

Además, $\hat{\sigma}^2 = \frac{1}{n} \sum^n (x_i - \overline{x})^2 = \frac{1}{n} \sum^n x_i^2 - \left(\frac{1}{n} \sum x_i\right)^2 = \overline{x^2} - \overline{x}^2$

$$\Rightarrow \hat{\lambda}_{MOM} = \frac{\widehat{\sigma^2}}{\overline{x}}$$

Y teníamos que $\overline{x} = \lambda k$, $\Rightarrow \hat{k}_{MOM} = \frac{\overline{x}}{\hat{\lambda}} = \frac{\overline{x}}{\frac{\widehat{\sigma^2}}{\overline{x}}} = \frac{\overline{x}^2}{\widehat{\sigma^2}}$

$$\therefore \hat{\lambda}_{MOM} = \frac{\widehat{\sigma^2}}{\overline{x}} \qquad \hat{k}_{MOM} = \frac{\overline{x}^2}{\widehat{\sigma^2}}$$

b) Asumiendo que tenemos k fijo, tenemos que la familia de modelos paramétricos es:

$$\mathcal{M}_\Theta = \{ \text{Gamma}(\tilde{k}, \lambda) \mid \lambda \in \mathbb{R}_+ = \Theta \}$$

y la verosimilitud en este caso es:

$$\mathcal{L}(\lambda \mid x) = \prod_i^n f(x_i; \tilde{k}, \lambda) \quad / \text{denotaré } \tilde{k} \text{ por } k \text{ por simplicidad (flojera)}$$

$$= \prod_i^n \frac{1}{\Gamma(k)\lambda^k} x_i^{k-1} e^{-x_i/\lambda} \underbrace{\mathbb{1}_{x_i>0}}_{\to 1} \quad \text{pues } \forall i \; x_i > 0 \text{ y no entra } \lambda$$

$$= \frac{1}{\Gamma(k)^n} \lambda^{-nk} \left(\prod^n x_i\right)^{k-1} e^{-\frac{\sum x_i}{\lambda}}$$

$$= \frac{\left(\prod^n x_i\right)^{k-1}}{\Gamma(k)^n} \lambda^{-nk} e^{-\frac{n\bar{x}}{\lambda}}$$

y la log-verosimilitud es:

$$\ell(\lambda \mid x) = \underbrace{(k-1)\sum^n \log x_i - n \log \Gamma(k)}_{\text{cte } c/r \text{ a } \lambda} - nk \log \lambda - \frac{n\bar{x}}{\lambda}$$

Derivamos!

$$\frac{\partial}{\partial \lambda} \ell(\lambda \mid x) = -nk \frac{1}{\lambda} + \frac{n\bar{x}}{\lambda^2}$$

Imponemos $= a\ 0$

$$-\frac{\cancel{n}k}{\cancel{\lambda}} + \frac{\cancel{n}\bar{x}}{\lambda^{\cancel{2}}} \overset{!}{=} 0 \iff -k + \frac{\bar{x}}{\lambda} = 0 \iff \lambda = \frac{\bar{x}}{k}$$

$\hookrightarrow \lambda > 0$

y por tanto

$$\boxed{\hat{\lambda}_{MLE} = \frac{\bar{x}}{\tilde{k}}}$$

P5) $X_1, \ldots, X_n \sim \text{Unif}[a, b]$ $a<b$ desconocidos

Se sabe que los primeros dos momentos de una uniforme son:

$$\mathbb{E}[U] = \frac{a+b}{2}$$
$$\mathbb{E}[U^2] = \frac{a^2+ab+b^2}{3}$$

(basta integrar las funciones, puede ser un ejercicio para desempolvar las habilidades de integración)

Entonces planteamos:

$$\frac{1}{n}\sum^n x_i =: \bar{x} = \frac{a+b}{2} \quad (1)$$
$$\frac{1}{n}\sum^n x_i^2 =: \overline{x^2} = \frac{a^2+ab+b^2}{3} \quad (2)$$

De (1) tenemos que:

$$a+b = 2\bar{x} \Rightarrow a = 2\bar{x} - b \quad (3)$$

Reemplazando (3) en (2):

$$3\overline{x^2} = (2\bar{x}-b)^2 + (2\bar{x}-b)b + b^2 = 4\bar{x}^2 - 4b\bar{x} + b^2 + 2\bar{x}b \cancel{- b^2} \cancel{+ b^2}$$
$$= b^2 - 2b\bar{x} + \bar{x}^2 + 3\bar{x}^2 = (b-\bar{x})^2 + 3\bar{x}^2$$

Así: $(b-\bar{x})^2 = 3(\overline{x^2} - \bar{x}^2) = 3\widehat{\sigma^2}$ (4)

(Recordar que $\widehat{\sigma^2} = \overline{x^2} - \bar{x}^2 = \frac{1}{n}\sum^n (x_i - \bar{x})^2$)

Reemplazando (3) en (4):

$$(a-\bar{x})^2 = 3\widehat{\sigma^2}$$

Así, tenemos que
$$b - \bar{x} = \pm\sqrt{3\widehat{\sigma^2}}$$
$$a - \bar{x} = \pm\sqrt{3\widehat{\sigma^2}}$$

Pero $a<b$, así que:

$$\hat{a}_{MOM} = \bar{x} - \sqrt{3\widehat{\sigma^2}}$$
$$\hat{b}_{MOM} = \bar{x} + \sqrt{3\widehat{\sigma^2}} \quad (5)$$
//

Tratemos de interpretarlo. Sabemos que $\widehat{\sigma^2}$ tratará de estimar la varianza, la cuál es:

$$\widehat{\sigma^2} \approx \frac{1}{12}(b-a)^2$$

Así que si lo reemplazamos en (5)

$$\hat{a} \approx \bar{x} - \sqrt{3\,\frac{(b-a)^2}{12}} = \bar{x} - \sqrt{\frac{(b-a)^2}{4}} = \bar{x} - \frac{b-a}{2}$$

$$\hat{b} \approx \bar{x} + \frac{b-a}{2}$$

Así que, tomando desde la mitad, el factor $\sqrt{3\widehat{\sigma^2}}$ te trata de decir que vayas la longitud del intervalo $(b-a)$ y le restes la mitad, mientras que para b sería sumarle la mitad.

b) Para la máxima verosimilitud, hacemos una estrategia como en la P2):

$$\mathcal{L}(a,b;x) = \prod^{n} \frac{1}{b-a} \mathbb{1}_{x_i \in [a,b]} = (b-a)^{-n} \mathbb{1}_{x_1 \ldots x_n \in [a,b]}$$

Vemos que: $x_1 \ldots x_n \in [a,b]$

$$\Leftrightarrow \forall i \;\; a \leq x_i \wedge b \geq x_i$$

$$\Leftrightarrow a \leq \min_k x_k \wedge b \geq \max_k x_k$$

Entonces

$$\mathcal{L}(a,b;x) = (b-a)^{-n} \cdot \mathbb{1}_{a \leq \min_k x_k} \cdot \mathbb{1}_{b \geq \max_k x_k}$$

Notemos que si fijamos un $\tilde{a} \leq \min_k x_k$, $\mathcal{L}(\tilde{a}, b; x)$ es decreciente en b, con un máximo en $\hat{b} = \max_k x_k$. Si ahora fijamos $\hat{b} = \max_k x_k$ $\mathcal{L}(a,b;x)$ es creciente con un tope en $\hat{a} = \min_k x_k$. Por tanto

$$\boxed{\hat{b}_{MLE} = \max_k x_k \qquad \hat{a}_{MLE} = \min_k x_k}$$

Estos estimadores tienen sentido, pues si tenemos muchas muestras, el $\max_k x_k$ se acerca el valor de b, al igual que $\min_k x_k$ lo hará al parámetro a.