
Estudios anteriores han demostrado que puede admitirse una distribución de Poisson para el número de hembras, en una determinada región, de la mosca tropical americana (*Dermatobia hominis*) la cual se caracteriza por poner sus huevos en un mosquito, pasando las larvas de la mosca a la piel de la persona cuya sangre ha chupado el mosquito.

Examinada la región en cuestión en 10 días elegidos al azar, se obtuvo el siguiente número de moscas hembra de la citada especie:

$$2, 1, 3, 3, 4, 2, 1, 2, 3, 2$$

Se pide:

a) **Estimar la ley de probabilidad que rige el mencionado número de moscas hembra de la región en estudio.**

b) **Determinar la estimación de máxima verosimilitud del número medio de moscas hembra de la región en cuestión.**

c) **Calcular el número mínimo de días que debe de muestrearse en la región en estudio, para que la diferencia entre el número medio de moscas hembra de la región y su estimación, se diferencien en menos de una, con probabilidad** $0'99$.

---

a) Del enunciado se desprende que para la variable aleatoria X, *número de moscas hembra de la región en estudio*, puede admitirse una distribución de Poisson $\mathcal{P}(\theta)$, lo cual significa que modelizamos, en cuanto a la forma, la ley de probabilidad que rige dicha variable aleatoria —la correspondiente a una distribución de Poisson–, dependiendo la determinación completa de dicha ley de probabilidad del valor que asignemos a θ. Estimando θ convenientemente, habremos estimado la ley de probabilidad de la variable en estudio.

Un método que nos proporciona buenos estimadores del parámetro o parámetros de la distribución es el de la *máxima verosimilitud*, independientemente de lo que representen aquellos —media, varianza u otra cosa—.

Para determinar el estimador de máxima verosimilitud de θ debemos formar, en primer lugar, la ecuación de verosimilitud. Como la función de masa de la distribución de Poisson $\mathcal{P}(\theta)$ es

$$p_\theta(x) = \frac{e^{-\theta}\,\theta^x}{x!} \qquad\qquad x = 0, 1, 2, ...$$

dicha función será

$$L(\theta) = \prod_{i=1}^{n} \frac{e^{-\theta}\,\theta^{x_i}}{x_i!} = \frac{e^{-n\theta}\,\theta^{\sum_1^n x_i}}{x_1! \cdot \ldots \cdot x_n!}$$

la cual tiene por logaritmo neperiano

$$\log L(\theta) = -n\theta + \sum_{i=1}^{n} x_i \log\theta - \log \prod_{i=1}^{n} x_i!.$$

Derivando esta función respecto a θ obtenemos la siguiente ecuación de verosimilitud:

$$\frac{\partial}{\partial\theta} \log L(\theta) = -n + \frac{\sum_{i=1}^{n} x_i}{\theta} = 0$$

la cual tiene como solución $\hat{\theta} = \overline{x}$; por tanto, el estimador de máxima verosimilitud de θ será la media muestral y, en consecuencia, estimaremos la ley de probabilidad que rige la variable en estudio por una $\mathcal{P}(\overline{x}) = \mathcal{P}(2'3)$.

(Obsérvese que, en la obtención del estimador de máxima verosimilitud, no se ha tenido en cuenta el que θ fuera la media de la distribución, es decir, la media de la población. En la obtención del estimador sólo hay que seguir los tres pasos antes indicados: obtención de la ecuación de verosimilitud, derivada (igualada a cero) de dicha función respecto al parámetro o parámetros que determinan totalmente la distribución modelo y, por último, resolución de la ecuación o sistema de ecuaciones determinado en el segundo paso.)

b) Como la media de la distribución de Poisson es el parámetro que la caracteriza (CB-sección 4.4.2), el cual hemos estimado en el apartado anterior, la estimación de máxima verosimilitud del número medio de moscas hembra de la región es $2'3$.

c) El problema aquí es el de determinar el menor tamaño muestral n tal que se cumpla la condición

$$P\left\{|\theta - \overline{x}| < 1\right\} = 0'99$$

o bien,

$$P\left\{|\overline{x} - \theta| < 1\right\} = 0'99$$

para lo que sería necesario conocer la distribución de la media muestral. Si el tamaño muestral resultante fuera grande, digamos mayor que 100,

podríamos utilizar la aproximación normal a la distribución de $\overline{x}$ dada en CB-sección 5.5, la cual es $\overline{x} \rightsquigarrow N(\theta, \sqrt{\overline{x}/n})$. En ese caso, tipificando en la última expresión sería

$$P\left\{|Z| < \frac{\sqrt{n}}{\sqrt{\overline{x}}}\right\} = 0'99$$

con $Z \rightsquigarrow N(0,1)$. La tabla 3 de la distribución normal nos dice que a la derecha de $2'575$ hay un área de probabilidad $0'005$, por lo que dentro del intervalo $(-2'575, 2'575)$ habrá un área de probabilidad $1 - 2 \times 0'005 = 0'99$. Es decir, que es

$$P\left\{|Z| < 2'575\right\} = 0'99.$$

Por tanto, debe ser

$$\frac{\sqrt{n}}{\sqrt{\overline{x}}} = 2'575$$

o bien,

$$\sqrt{n} = 2'575 \cdot \sqrt{\overline{x}} = 2'575 \cdot \sqrt{2'3}$$

es decir,

$$n = 2'575^2 \cdot 2'3 = 15'25.$$

A medida que n aumenta, disminuye la varianza de la distribución normal de $\overline{x}$ antes utilizada, por lo que ésta estaría cada vez más concentrada entorno a su media, θ, y sería cada vez mayor la probabilidad del suceso puesto como condición al inicio del apartado. En suma, el *menor* tamaño muestral requerido es $n = 16$.

Observemos, sin embargo, que hemos utilizado una aproximación normal, la cual es buena cuando el tamaño muestral es grande; como no ha resultado tan grande como lo inicialmente supuesto, vamos a calcular directamente la probabilidad en cuestión:

$$P\left\{|\overline{x} - \theta| < 1\right\}$$

utilizando el que $n \cdot \overline{x} = 16 \cdot \overline{x} \rightsquigarrow \mathcal{P}(16\,\theta)$ como se indica en la sección de CB antes mencionada. Será

$$P\left\{|\overline{x}-\theta|<1\right\}=P\left\{\theta-1<\overline{x}<\theta+1\right\}=P\left\{16(\theta-1)<\sum_{i=1}^{16}X_i<16(\theta+1)\right\}.$$

Si es $\theta=2'3$, esta probabilidad será

$$P\left\{20'8<\sum_{i=1}^{16}X_i<52'8\right\}$$

con $\sum_{i=1}^{16}X_i\rightsquigarrow\mathcal{P}(36'8)$. Por tanto,

$$\begin{aligned}
P\{20'8<\sum_{i=1}^{16}X_i<52'8\} &= P\{\sum_{i=1}^{16}X_i=21\}+P\{\sum_{i=1}^{16}X_i=22\}+...+P\{\sum_{i=1}^{16}X_i=52\}\\
&= \frac{e^{-36'8}\,36'8^{21}}{21!}+\frac{e^{-36'8}\,36'8^{22}}{22!}+...+\frac{e^{-36'8}\,36'8^{52}}{52!}\\
&= F(52)-F(20)\\
&= 0'992986-0'001840\\
&= 0'991146>0'99
\end{aligned}$$

como queríamos y en donde F representa la función de distribución de la $\mathcal{P}(36'8)$.

Como últimos cálculos a realizar que confirman el que $n=16$ es el menor tamaño muestral requerido, se tiene que, si fuera $n=15$, sería $n\cdot\overline{x}=15\cdot\overline{x}=\sum_{i=1}^{15}X_i\rightsquigarrow\mathcal{P}(15\,\theta)=\mathcal{P}(15\cdot 2'3)=\mathcal{P}(34'5)$ y

$$\begin{aligned}
P\left\{|\overline{x}-\theta|<1\right\} &= = P\left\{15(2'3-1)<\sum_{i=1}^{15}X_i<15(2'3+1)\right\}\\
&= P\left\{19'5<\sum_{i=1}^{15}X_i<49'5\right\}=F(49)-F(19)\\
&= 0'992295-0'002957=0'989338<0'99
\end{aligned}$$

con $F\rightsquigarrow\mathcal{P}(34'5)$ no llegando, por tanto, a la probabilidad exigida, y si fuera $n=17$, sería $\sum_{i=1}^{17}X_i\rightsquigarrow\mathcal{P}(17\cdot 2'3)=\mathcal{P}(39'1)$ y

$$
\begin{aligned}
P\left\{|\overline{x}-\theta|<1\right\} &= P\left\{22'1<\sum_{i=1}^{17}X_i<56'1\right\} \\
&= F(56)-F(22) \\
&= 0'995758-0'002135=0'993623>0'99
\end{aligned}
$$

con $F \rightsquigarrow \mathcal{P}(39'1)$, mayor que la probabilidad solicitada aunque no siendo $n=17$ el menor tamaño muestral que lo cumple.

Por tanto, se concluye que es razonable utilizar la aproximación normal para la media de una muestra procedente de la distribución de Poisson, al menos en una primera etapa. Si el tamaño muestral resultante no es muy grande conviene realizar unos cálculos, similares a los anteriores, que confirmen el resultado obtenido.