Godinez Bravo Diego
Inferencia Estadística
Tarea 7

CIMAT

1. Las hojas de una planta se examinan buscando insectos. El número de insectos en una hoja sigue una distribución de Poisson con media λ, con la excepción de que muchas de las hojas no tienen insectos pues son inadecuadas para que se alimenten de ellas y esto no es simplemente el resultado de la variación aleatoria de la ley de Poisson.

a) Encuentre la probabilidad condicional de que una hoja contenga i insectos, dado que contiene al menos uno.

Solución

Dado que el número de insectos sigue una distribución de Poisson con parámetro λ

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad x = 0, 1, 2...$$

Y dada la definición de probabilidad condicional

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

entonces

$$P(X = i|X \geq 1) = \frac{P(X = i \cap X \geq 1)}{P(X \geq 1)}$$

Sabemos que $P(X \geq 1) = 1 - P(X = 0)$, y considerando que $X = i$ implica que $X \geq 1$; entonces

$$P(X = i|X \geq 1) = \frac{P(X = i)}{1 - P(X = 0)}$$

Considerando que $X \sim Poisson(\lambda)$

$$P(X = i|X \geq 1) = \frac{\frac{\lambda^i e^{-\lambda}}{i!}}{1 - \frac{\lambda^0 e^{-\lambda}}{0!}}$$

$$= \frac{\frac{\lambda^i e^{-\lambda}}{i!}}{1 - e^{-\lambda}}$$

De esta manera tenemos que la probabilidad condicional esta dada de la siguiente manera

$$P(X = i|X \geq 1) = \frac{\lambda^i e^{-\lambda}}{(1 - e^{-\lambda})i!}$$

b) Supongamos que se observan x_i hojas conteniendo i insectos ($i = 1, 2, 3, \ldots$), con $\sum x_i = n$. Muestre que el estimador de máxima verosimilitud de μ satisface la ecuación

$$\hat{\mu} = \bar{x}(1 - e^{-\hat{\mu}})$$

donde $\bar{x} = \sum i \frac{x_i}{n}$.

Solución

Dada la función de probabilidad condicional encontrada

$$P(X = i | X \geq 1) = \frac{\lambda^i e^{-\lambda}}{(1 - e^{-\lambda}) i!}$$

Encontramos la verosimilitud $L(\lambda)$ como sigue

$$L(\lambda) = \frac{\lambda^i e^{-\lambda}}{(1 - e^{-\lambda}) i!} * \frac{\lambda^i e^{-\lambda}}{(1 - e^{-\lambda}) i!} * \ldots * \frac{\lambda^i e^{-\lambda}}{(1 - e^{-\lambda}) i!} = \frac{\lambda^{\sum_{k=1}^n i_k} e^{-n\lambda}}{(1 - e^{-\lambda})^n \Pi_{k=1}^n i_k!}$$

Obtenemos el logaritmo de la verosimilitud

$$log\ L(\lambda) = log\Big(\lambda^{\sum_{k=1}^n i_k} e^{-n\lambda}\Big) - log\Big((1 - e^{-\lambda})^n \Pi_{k=1}^n i_k!\Big)$$

$$= log\Big(e^{-n\lambda} \lambda^{\sum_{k=1}^n i_k}\Big) - log\Big((1 - e^{-\lambda})^n \Pi_{k=1}^n i_k!\Big)$$

$$= -n\lambda + log\Big(\lambda^{\sum_{k=1}^n i_k}\Big) - log\Big((1 - e^{-\lambda})^n \Pi_{k=1}^n i_k!\Big)$$

$$= -n\lambda + \sum_{k=1}^n i_k\ log\Big(\lambda\Big) - nlog(1 - e^{-\lambda}) - log\Big(\Pi_{k=1}^n i_k!\Big)$$

Derivando e igualando a cero para obtener los puntos críticos

$$\frac{d}{d\lambda} log\ L(\lambda) = -n + \sum_{k=1}^n i_k\ \frac{1}{\lambda} - n\ \frac{e^{-\lambda}}{(1 - e^{-\lambda})}$$

$$= -n + \frac{\sum_{k=1}^n i_k}{\lambda} - \frac{ne^{-\lambda}}{(1 - e^{-\lambda})}$$

$$= \frac{\sum_{k=1}^n i_k}{\lambda} - \frac{ne^{\lambda}}{-(1 - e^{-\lambda})} - n$$

$$= \frac{\sum_{k=1}^n i_k}{\lambda} - \frac{ne^{-\lambda} + n(1 - e^{-\lambda})}{(1 - e^{-\lambda})}$$

$$= \frac{\sum_{k=1}^n i_k}{\lambda} - \frac{ne^{-\lambda} + n - ne^{-\lambda}}{(1 - e^{-\lambda})}$$

$$= \frac{\sum_{k=1}^n i_k}{\lambda} - \frac{n}{(1 - e^{-\lambda})}$$

derivada $\frac{d}{d\lambda} log\ L(\lambda)$.

$$\frac{\sum_{k=1}^n i_k}{\lambda} - \frac{n}{(1 - e^{-\lambda})} = 0$$

$$\frac{\sum_{k=1}^{n} i_k}{\lambda} = \frac{n}{(1-e^{-\lambda})}$$

$$\lambda = \frac{\sum_{k=1}^{n} i_k(1-e^{-\lambda})}{n}$$

$$\lambda = \frac{\sum_{k=1}^{n} i_k}{n}\ (1-e^{-\lambda})$$

sea $\bar{k} = \frac{\sum_{k=1}^{n} i_k}{n}$

$$\lambda = \bar{k}\ (1-e^{-\lambda})$$

Por lo tanto, estimador de máxima verosimilitud de μ satisface la ecuación propuesta.

c) Determine $\hat{\mu}$ numéricamente para el caso $\bar{x} = 3.2$. Utilice R.

Resultados

Implementando una función en R se estimó el valor númerico de λ

$$\lambda = 3.0695$$

2. Los siguientes son los tiempos (en horas) entre fallas sucesivas del sistema de aire acondicionado en un avión:

97	51	11	4	141	18	142	68	77
80	1	16	106	206	82	54	31	216
46	111	39	63	18	191	18	163	24

a) Suponiendo que estas son observaciones independientes de una distribución exponencial con media θ, encuentre $\hat{\theta}_{MLE}$.

Solución

Dado que los tiempos siguen una distribución exponencial $X \sim Exp(\theta)$

$$f(x) = \frac{1}{\theta}e^{-\frac{x}{\theta}} \quad x > 0$$

Encontramos la verosimilitud $L(\theta)$

$$L(\theta) = \frac{1}{\theta}e^{-\frac{x_1}{\theta}} * \frac{1}{\theta}e^{-\frac{x_2}{\theta}} * \cdots * \frac{1}{\theta}e^{-\frac{x_n}{\theta}} = \frac{1}{\theta^n}e^{-\frac{1}{\theta}\sum_{i=1}^n x_i}$$

Obtenemos el logaritmo

$$log\ L(\theta) = log\Big(\frac{1}{\theta^n}e^{-\frac{1}{\theta}\sum_{i=1}^n x_i}\Big)$$

$$= log\Big(\frac{1}{\theta^n}\Big) + log\Big(e^{-\frac{1}{\theta}\sum_{i=1}^n x_i}\Big)$$

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

$$= -nlog(\theta) - \frac{1}{\theta}\sum_{i=1}^n x_i$$

Derivando y obteniendo los puntos críticos

$$\frac{d}{d\theta}log\ L(\theta) = -\frac{n}{\theta} - \sum_{i=1}^n x_i \left(-\frac{1}{\theta^2}\right)$$

$$= -\frac{n}{\theta} + \frac{\sum_{i=1}^n x_i}{\theta^2}$$

derivada $\frac{d}{d\theta}log\ L(\theta)$.

$$-\frac{n}{\theta} + \frac{\sum_{i=1}^n x_i}{\theta^2} = 0$$

$$\frac{-n\theta^2 + \theta\ \sum_{i=1}^n x_i}{\theta^3} = 0$$

$$\frac{-n\theta + \sum_{i=1}^n x_i}{\theta^2} = 0$$

cuando $\theta = \frac{\sum_{i=1}^n x_i}{n}$.

Por lo tanto

$$\hat{\theta} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Considerando los datos proporcionados

$$\hat{\theta}_{MLE} = \frac{\sum_{i=1}^{27} x_i}{27}$$

$$\hat{\theta}_{MLE} = \frac{97 + 51 + \cdots + 163 + 24}{27} = \frac{2074}{27}$$

$$\hat{\theta}_{MLE} = 76.8148$$

b) Haga una tabla de frecuencias para estos datos usando las clases $(0, 50], (50, 100], (100, 200]$ y $(200, \infty]$. Calcule el estimador de máxima verosimilitud de las frecuencias esperadas para estas clases bajo el supuesto en el inciso anterior. ¿La distribución exponencial parece ser un modelo adecuado para los datos?

Resultados

Sabemos que los tiempos siguen una distribución exponencial $X \sim Exp(\theta)$

$$f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}} \quad x > 0$$

Por lo tanto, calculamos las frecuencias esperadas utilizando la función de distribución $f(x)$ considerando $\hat{\theta} = 76.8148$, y evaluando en los intervalos propuestos

$$freq_i = n \int_{-\infty}^{\infty} f(x) dx$$

En este caso

$$freq_i = 27 \int_{-\infty}^{\infty} \frac{1}{76.8148} e^{-\frac{x}{76.8148}} dx$$

Se calculan las frecuencias con ayuda de recursos computacionales evaluando la integral en cada uno de los intervalos propuestos.

Intervalo $(0, 50]$

$$freq = 27 \int_{0}^{50} \frac{1}{76.8148} e^{-\frac{x}{76.8148}} dx$$

$$freq = 12.9176$$

Intervalo $(50, 100]$

$$freq = 27 \int_{50}^{100} \frac{1}{76.8148} e^{-\frac{x}{76.8148}} dx$$

$$freq = 6.7374$$

Intervalo $(100, 200]$

$$freq = 27 \int_{100}^{200} \frac{1}{76.8148} e^{-\frac{x}{76.8148}} dx$$

$$freq = 5.3468$$

Intervalo $(200, \infty]$

$$freq = 27 \int_{200}^{\infty} \frac{1}{76.8148} e^{-\frac{x}{76.8148}} dx$$

$$freq = 1.9980$$

Intervalos	Frecuencia Esperada	Frecuencia Observada
$(0, 50]$	12.9176	11
$(50, 100]$	6.7374	8
$(100, 200]$	5.3468	6
$(200, \infty]$	1.9980	2

Podemos afirmar que el modelo propuesto es preciso al momento de describir cómo se distribuyen los tiempos entre fallas sucesivas del sistema. Esto principalmente porque las frecuencias obtenidas se aproximan en gran medida a las frecuencias observadas.

3. Se hicieron 27 mediciones de los rendimientos de dos procesos industriales, con los siguientes resultados:

$$\begin{aligned} &\text{Proceso 1:} \quad n_1 = 11,\ \bar{y}_1 = 6.23,\ s_1^2 = 3.79 \\ &\text{Proceso 2:} \quad n_2 = 16,\ \bar{y}_2 = 12.74,\ s_2^2 = 4.17 \end{aligned}$$

Suponiendo que los rendimientos se distribuyen normalmente con la misma varianza, encuentre intervalos de confianza para las medias μ_1 y μ_2 y para la diferencia de medias $\mu_1 - \mu_2$.

Solución

Dado que se trata de una población normal con media μ_i y varianza σ^2 desconocida; podemos calcular los intervalos de confianza como

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

De esta manera, para el **primer proceso** industrial tenemos

$$\bar{x} = 6.23 \quad s^2 = 3.79 \quad n - 1 \text{ grados de libertad}$$

consultando el valor t en tablas t-Student con 10 grados de libertad y $\alpha = 0.05$

$$6.23 \pm 2.2281 * \frac{\sqrt{3.79}}{\sqrt{11}}$$

$$6.23 \pm 2.2281 * \frac{1.9467}{3.3166}$$

$$6.23 \pm 2.2281 * 0.5869$$

$$6.23 \pm 1.3078$$

Intervalo de confianza para μ_1

$$IC = (4.9221, 7.5378)$$

Por otro lado, para el **segundo proceso** industrial tenemos

$$\bar{x} = 12.74 \quad s^2 = 4.17 \quad n - 1 \text{ grados de libertad}$$

consultando el valor t en tablas t-Student con 15 grados de libertad y $\alpha = 0.05$

$$12.74 \pm 2.1315 * \frac{\sqrt{4.17}}{\sqrt{16}}$$

$$12.74 \pm 2.1315 * \frac{2.0420}{4}$$

$$12.74 \pm 2.1315 * 0.5105$$

$$12.74 \pm 1.0881$$

Intervalo de confianza para μ_2

$$IC = (11.6519, 13.8281)$$

Considerando que ambos rendimientos siguen la misma distribución con varianza desconocida tal que $\sigma_1^2 = \sigma_2^2 = \sigma^2$; podemos estimar el intervalo de confianza para ambas poblaciones de la siguiente manera

$$\bar{x} - \bar{y} \pm t_{\alpha/2}\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}$$

Dado que σ^2 es desconocida, debemos de estimarla considerando ambas fuentes de información

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Sustituyendo los valores de ambos procesos

$$S_p^2 = \frac{(11 - 1)(3.79) + (16 - 1)(4.17)}{11 + 16 - 2}$$

$$= \frac{(10)(3.79) + (15)(4.17)}{25}$$

$$= \frac{(37.9 + 62.55)}{25}$$

$$S_p^2 = \frac{104.45}{25} = 4.018$$

Considerando un valor de t con $n_1 + n_2 - 2$ grados de libertad y valor $\alpha = 0.05$

$$\bar{x} - \bar{y} \pm t_{\alpha/2}\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}$$

$$6.23 - 12.74 \pm (2.0595)(\sqrt{4.018})\sqrt{\frac{1}{11} + \frac{1}{16}}$$

$$-6.51 \pm (2.0595)(2.0044)(\sqrt{0.1534})$$

$$-6.51 \pm 1.6168$$

Intervalo de confianza para $\mu_1 - \mu_2$

$$IC = (-8.1268, -4.8932)$$

Ejercicios Notas.

Ejercicio pg. 95

Otro estimador para σ^2 en poblaciones normales es
$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Tenemos que,

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n}(n-1)S^2 = \frac{n-1}{n}S^2,$$

lo cual implica que

$$E(S_n^2) = E\left(\frac{n-1}{n}S^2\right) = \frac{n-1}{n}E(S^2) = \frac{n-1}{n}\sigma^2 \neq \sigma^2.$$

Por lo tanto, S_n^2 no es un estimador insesgado para σ^2. En este caso se dice que S_n^2 subestima en promedio el valor de σ^2; se dice que hay un sesgo.

Calcula el sesgo.

Solución

Si un estimador $\hat{\theta}$ no es insesgado para θ, se dice que es sesgado y se define el sesgo de $\hat{\theta}$ como

$$Sesgo(\hat{\theta}) = E(\hat{\theta}) - \theta$$

Cuando es sesgo es positivo se dice que el estimador está sesgado a la derecha y si es negativo se dice que está sesgado a la izquierda.

Sea el estimador $\hat{\theta} = S_n^2 = \frac{1}{n}(n-1)S^2 = \frac{n-1}{n}S^2$, entonces

$$Sesgo(\hat{\theta}) = E(S_n^2) - \theta$$

$$= \frac{n-1}{n}\sigma^2 - \sigma^2$$

$$= \frac{n\sigma^2 - \sigma^2}{n} - \sigma^2$$

$$= \frac{n\sigma^2 - \sigma^2 - n\sigma^2}{n}$$

Por lo tanto, el sesgo del estimador $\theta = \sigma^2$ es

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

Ejercicio pg. 103

Demostrar que la función de densidad $X_{(n)}$ y su varianza son

$$f_{X_{(n)}}(x) = n\frac{x^{n-1}}{\theta^n} \quad 0 < x < \theta$$

y

$$V(X_{(n)}) = \frac{n}{(n+1)^2(n+2)}\theta^2$$

respectivamente.

Solución

Dado que se consideró una distribución uniforme en el intervalo $(0, \theta)$, entonces

$$f_X(x) = \frac{1}{\theta} \quad 0 < x < \theta$$

y

$$F_X(x) = \frac{x}{\theta}$$

Por lo tanto, podemos calcular la función de densidad como sigue

$$f_{X_{(n)}}(x) = n[F_X(x)]^{n-1} f_X(x)$$

$$f_{X_{(n)}}(x) = n\Big[\frac{x}{\theta}\Big]^{n-1} \cdot \frac{1}{\theta}$$

$$= n\Big[\frac{x^{n-1}}{\theta^{n-1}}\Big] \cdot \frac{1}{\theta}$$

$$= n\Big[\frac{x^{n-1}}{\theta^{n-1+1}}\Big]$$

De esta manera comprobamos que la **función de densidad** $f_{X_{(n)}}(x)$ esta definida como

$$f_{X_{(n)}}(x) = n\frac{x^{n-1}}{\theta^n}$$

Por otro lado

$$V(X) = E[X^2] - E[X]^2$$

Calculamos cada uno de los elementos.

$E[X]$

$$E[X] = \int_{-\infty}^{\infty} x\, f(x) dx$$

$$= \int_0^{\theta} x\; n\; \frac{x^{n-1}}{\theta^n} dx$$

$$= \frac{n}{\theta^n} \int_0^{\theta} x^n\; dx$$

$$= \frac{n}{\theta^n} \Big[\frac{x^{n+1}}{n+1}\Big]_0^{\theta}$$

$$= \frac{n}{\theta^n}\Big(\frac{\theta^{n+1}}{n+1}\Big)$$

$$E[X] = \frac{n\theta}{n+1}$$

$E[X^2]$

$$E[X^2] = \int_0^\theta x^2 \; n \; \frac{x^{n-1}}{\theta^n} dx$$

$$= \int_0^\theta n \; \frac{x^{n+1}}{\theta^n} dx$$

$$= \frac{n}{\theta^n} \int_0^\theta x^{n+1} dx$$

$$= \frac{n}{\theta^n}\Big[\frac{x^{n+2}}{n+2}\Big]_0^\theta$$

$$= \frac{n}{\theta^n}\Big(\frac{\theta^{n+2}}{n+2}\Big)$$

$$E[X^2] = \frac{n\theta^2}{n+2}$$

Por lo tanto

$$V(X_{(n)}) = \frac{n\theta^2}{n+2} - \Big(\frac{n\theta}{n+1}\Big)^2$$

$$= \frac{n\theta^2(n+1)^2 - n^2\theta^2(n+2)}{(n+2)(n+1)^2}$$

$$= \frac{n\theta^2(n^2+2n+1) - n\theta^2(n^2+2n)}{(n+2)(n+1)^2}$$

$$= \frac{n\theta^2(n^2+2n+1-n^2-2n)}{(n+2)(n+1)^2}$$

$$= \frac{n\theta^2(1)}{(n+2)(n+1)^2}$$

Por lo tanto, comprobamos que la **varianza** $V(X_{(n)})$ esta definida de la siguiente manera

$$V(X_{(n)}) = \frac{n\theta^2}{(n+2)(n+1)^2}$$