1. Un banco ha establecido un programa especial de entrenamiento para sus cajeros. Antes del entrenamiento el promedio de errores de transacción diarias en sus sucursales tenia una distribución normal con media $\mu = 0.02$ con una desviación standard $\sigma = 0.004$. Suponga que con el objetivo de establecer si el entrenamiento ha sido efectivo en la reducción de la media del promedio diario de errores de transacción, un mes despues de finalizado el entrenamiento se seleccionaron 16 sucursales al azar y se calculo para cada una de ellas el promedio del día de transacciones erroneas. Supongamos que el porcentaje de transacciones erroneas diarias despues del entrenamiento está normalmente distribuido y con la misma desviacion standard $\sigma = 0.004$ que antes del entrenamiento. Suponga que la media muestral del promedio de transacciones erroneas de las 16 sucursales fue de 0.019.

   (a) Establezca la hipótesis nula y alternativa adecuada para investigar si el entrenamiento fue efectivo.

   (b) ¿Considera que observar una media de 0.019 es evidencia convincente de que el entrenamiento ha sido efectivo? Para contestar esta pregunta plantee el test de hipótesis adecuado e interprete su resultado.

   (c) Si la verdadera media del porcentaje de transacciones diarias en todas las sucursales del banco después del entrenamiento es de 0.019 y se utiliza un test de nivel 0.01, con base en una muestra de tamaño 16, ¿cuál es la probabilidad de detectar este valor de la media poblacional?

   (d) ¿Qué tamaño muestral se requiere para satisfacer que el test tenga nivel igual a 0.01 y potencia para detectar a la media poblacional de 0.019 igual a 0.99?

2. Sean $X_1, \ldots, X_n$ una muestra aleatoria cuya distribución depende del parámetro $\theta > 0$ dada por la función de densidad

$$f_X(x|\theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}\ I_{[0,+\infty)}$$

   a) Halle el estimador de máxima verosimilitud $\widehat{\theta}_{MV}$ de θ.

3. Sea X la proporcion de tiempo que un estudiante elegido al azar en una cierta poblacion de estudiantes emplea estudiando matematicas. Supongamos que la funcion de densidad de X es de la forma

$$f(x;\theta) = \begin{cases} (\theta+1)\, x^{\theta} \text{ si } 0 \leq x \leq 1 \\ 0 \text{ de otro modo} \end{cases}$$

   donde $\theta > -1$ es desconocido.

   (a) Encontrar el estimador de máxima verosimilitud de θ

   (b) Sabiendo que $E(X) = \frac{\theta+1}{\theta+2}$ encontrar el estimador de momentos para θ.

   (c) Provea la formula para el estimador de máxima verosimilitud de $E(X)$

Solución ej 3

1. Las hipótesis son:

$$H_0: \quad \mu = 0.02 \qquad vs \qquad H_1: \quad \mu < 0.02$$

2.

$$P_{\mu=0.02}\left(\overline{X}_n \leq 0.019\right) = P\left(\frac{\overline{X}_n - 0.02}{\sqrt{\frac{0.004^2}{16}}} \leq \frac{0.019 - 0.02}{\sqrt{\frac{0.004^2}{16}}}\right) = P(Z \leq -1) \approx 0.1587$$

Si la regla de decisión fuese rechazar H_0 si se observase que la media muestral es 0.019 o menor, entonces se rechazaría la hipótesis nula cuando es verdadera en el 15.87% de los casos. Con este p-valor, no deberíamos considerar que hay suficiente evidencia a favor de que el entrenamiento haya sido efectivo.

Supongamos que la regla de decisión para el test original es rechazar H_0 si

$$\overline{X}_{16} \leq 0.02 - \underbrace{z_{0.99}}_{2.326} \cdot \frac{0.004}{\sqrt{16}} \approx 0.017674$$

(es decir, la media muestral es el cuantil 0.01 de la distribución de $\overline{X}_{16}$). Entonces la probabilidad de detectar que $\mu = 0.019$, es calcular la potencia del test:

$$\begin{aligned} P_{\mu=0.019}\left(\text{rechazar } H_0\right) &= P_{\mu=0.019}\left(\overline{X}_n \leq 0.017674\right) \\ &= P_{\mu=0.019}\left(\frac{\overline{X}_n - 0.019}{0.001} \leq \frac{0.017674 - 0.019}{0.001}\right) \\ &= P_{\mu=0.019}(Z \leq -1.326) = 0.0924 \end{aligned}$$

Para el último inciso buscamos encontrar n de manera que $\alpha = 0.01$ y la potencia sea 0.99.

Ahora para un n genérico, tenemos que si la regla de decisión para el test original es rechazar H_0 si

$$\overline{x}_n \leq 0.02 - \underbrace{z_{0.99}}_{2.326} \cdot \frac{0.004}{\sqrt{n}}$$

(es decir, la media muestral observada es el cuantil 0.01 de la distribución de $\overline{X}_n$)

$$\begin{aligned} P_{\mu=0.019}\left(\text{Rechazar } H_0\right) &= P_{\mu=0.019}\left(\overline{X}_n \leq 0.02 - 2.326 \cdot \frac{0.004}{\sqrt{n}}\right) \\ &= P_{\mu=0.019}\left(\frac{\overline{X}_n - 0.019}{\frac{0.004}{\sqrt{n}}} \leq \frac{0.02 - 2.326 \cdot \frac{0.004}{\sqrt{n}} - 0.019}{\frac{0.004}{\sqrt{n}}}\right) = 0.99 \\ &= P\left(Z \leq \left(0.001 - \frac{0.009304}{\sqrt{n}}\right)\frac{\sqrt{n}}{0.004}\right) = 0.99 \\ &= P\left(Z \leq \frac{1}{4}\sqrt{n} - 2.326\right) = 0.99 \end{aligned}$$

Por lo tanto

$$\frac{1}{4}\sqrt{n} - 2.326 = z_{0.99} = 2.326$$

Despejando n, obtenemos que $n = 346.26$. Por lo tanto, se necesita una muestra de 347 datos para poder garantinzar que la potencia sea de 0.99 y el nivel de significación de 0.01.

Solución ej 2

a Calculamos la función de verosimilitud.

$$\mathcal{L}_n(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} e^{-\frac{X_i}{\theta}} = \left(\frac{1}{\theta}\right)^n e^{-\frac{1}{\theta}\sum_{i=1}^{n} X_i}$$

Tomando el logaritmo de la verosimilitud, se tiene que:

$\log(\mathcal{L}_n(\theta)) = n\log(\frac{1}{\theta}) - \frac{1}{\theta}\sum_{i=1}^{n} X_i$

Obtenemos la condición de primer orden respecto de θ, $-\frac{n}{\theta} + \frac{\sum_{i=1}^{n} X_i}{\theta^2} = 0$

De donde obtenemos el EMV (estimador de máxima verosimilitud): $\hat{\theta}_{MV} = \frac{\sum_{i=1}^{n} X_i}{n}$

Solución ej 3

(a) y (c) Calculamos el EMV de θ y luego por la propiedad invarianza obtenemos el EMV de $E(X)$. Para una muestra i.i.d. de tamaño n, tenemos la siguiente función de verosimilitud

$$\mathcal{L}_n(\theta) = \prod_{i=1}^{n}(\theta+1)X_i^{\theta} = (\theta+1)^n \prod_{i=1}^{n} X_i^{\theta}$$

Tomando el logaritmo de la función de verosimilitud

$\log \mathcal{L}_n(\theta) = n\log(\theta+1) + \theta \sum_{i=1}^{n} \log(X_i)$

Sacando la CPO del problema de maximización respecto de θ:

$\frac{n}{\hat{\theta}_{MV}+1} + \sum_{i=1}^{n}\log(X_i) = 0 \quad \frac{n}{\hat{\theta}_{MV}+1} = -\sum_{i=1}^{n}\log(X_i) \quad \hat{\theta}_{MV} = -\frac{1}{\frac{\sum_{i=1}^{n}\log(X_i)}{n}} - 1$

Luego por la propiedad invarianza, el estimador de máxima verosimilitud de $E(X)$ es

$$\frac{\hat{\theta}_{MV}+1}{\hat{\theta}_{MV}+2}$$