

Guía de Problemas para Evaluación 3

Se recomienda ver el último problema de esta guía, que corresponde a un intervalo en el caso exponencial y que es tratado de manera distinta a los vistos en clases.

Problema 1

Sea una m.a.s. $\{x_1, x_2, \dots, x_n\}$ de una distribución $N(\mu, \sigma^2)$ con media μ desconocida y $\sigma^2 = 9$.

- Dé el intervalo de confianza de largo mínimo para un nivel de confianza 0.95
- ¿Qué tamaño mínimo de muestra hay que tomar para tener un intervalo de largo igual o a lo más de 2/5?
- Suponga que X_i tiene una distribución $N(\mu, \nu^2)$, con ν^2 desconocido. Deduzca el intervalo de confianza para μ con $\alpha = 0.05$ tome $n=21$. Determine el largo del intervalo anterior.

Solución:

a) $X_i \rightarrow N(\mu, \sigma^2)$, entonces $\frac{(\bar{X} - \mu)\sqrt{n}}{\sigma} \rightarrow N(0,1)$

queremos que $P(a \leq \mu \leq b) = 1 - \alpha = 0.95$

$$\Rightarrow P\left(\frac{(\bar{x} - b)\sqrt{n}}{\sigma} \leq \frac{(\bar{x} - \mu)\sqrt{n}}{\sigma} \leq \frac{(\bar{x} - a)\sqrt{n}}{\sigma}\right) = 0.95$$

$$\Rightarrow P(z_1 \leq z \leq z_2) = 0.95$$

el intervalo de largo mínimo mas corto es el simétrico, entonces $z_1 = -z_2$

$$\Rightarrow P(z \geq z_2) = \frac{\alpha}{2} = 0.025. \quad \text{De la tabla } N(0,1), \quad z_2 = 1.96$$

$$\begin{aligned} \Rightarrow \frac{(\bar{x} - a)\sqrt{n}}{\sigma} &= 1.96 & \Rightarrow a &= \bar{x} - \frac{1.96\sigma}{\sqrt{n}} \\ & & \Rightarrow b &= \bar{x} + \frac{1.96\sigma}{\sqrt{n}} \end{aligned}$$

$$\Rightarrow I_1 = \left[\bar{x} - \frac{1.96\sigma}{\sqrt{n}}, \bar{x} + \frac{1.96\sigma}{\sqrt{n}} \right], \text{ pero } \sigma = \sqrt{9} = 3$$

$$\Rightarrow I_1 = \left[\bar{x} - \frac{5.88}{\sqrt{n}}, \bar{x} + \frac{5.88}{\sqrt{n}} \right]$$

b) El largo del intervalo es $L = b - a = \bar{x} + \frac{5.88}{\sqrt{n}} - \bar{x} + \frac{5.88}{\sqrt{n}} = 2 * \frac{5.88}{\sqrt{n}}$

Para que $L \leq \frac{2}{5}$, $\Rightarrow 2 * \frac{5.88}{\sqrt{n}} \leq \frac{2}{5} \Rightarrow 29.4 \leq \sqrt{n} \Rightarrow 864.36 \leq n$

c) Si $X_i \rightarrow N(\mu, \sigma^2)$ y quiero calcular el intervalo de la media, utilizo t-student

$\Rightarrow P\left(\frac{(\bar{x} - b)\sqrt{(n-1)}}{S_n} \leq t_{n-1} \leq \frac{(\bar{x} - a)\sqrt{(n-1)}}{S_n}\right) = 0.95$, suponiendo intervalo simétrico

$\Rightarrow P(t_{n-1} \geq t_2) = \frac{\alpha}{2} = 0.025 \Rightarrow t_2 = 2.086$

$\Rightarrow \frac{(\bar{x} - a)\sqrt{n-1}}{S_n} = 2.086 \quad \Rightarrow a = \bar{x} - \frac{2.086 S_n}{\sqrt{n-1}}$ y $b = \bar{x} + \frac{2.086 S_n}{\sqrt{n-1}}$

$\Rightarrow L = b - a = \frac{2 * 2.086 * S_n}{\sqrt{n-1}}$

Problema 2

Una empresa desea estimar el promedio de tiempo que necesita una secretaria para llegar a su trabajo. Se toma una m.a.s. de 26 secretarias y se encuentra un promedio muestral de 40 minutos. Suponiendo que el tiempo de trayecto proviene de una $N(\mu, \sigma^2)$, $\sigma^2=12$, dar un intervalo de confianza para la media μ .

Solución:

Se tiene entonces:

$N=36$

$X \rightarrow N(\mu, \sigma^2)$

Empíricamente: $\bar{X} = 40$

Sea el estadístico Z tal que $Z = \left(\frac{\bar{X} - \mu}{\sigma}\right) \cdot \sqrt{n} = \left(\frac{40 - \mu}{2\sqrt{3}}\right) \cdot \sqrt{36} = (40 - \mu) \cdot \sqrt{3}$

Luego:

$P(a < \mu < b) = 1 - \alpha$

$$P(-b < \mu < -a) = P[(40 - b) \cdot \sqrt{3} < (40 - \mu) \cdot \sqrt{3} < (40 - a) \cdot \sqrt{3}] = 1 - \alpha$$

$$P(z_1 < Z < z_2) = 1 - \alpha$$

$$P(Z > z_2) = \alpha/2$$

Para un nivel de confianza del 95%, $\alpha=0.05$.

$$\Rightarrow P(Z > z_2) = 0.025$$

$$\Rightarrow z_2 = 1.96 = (40 - a) \cdot \sqrt{3}$$

$$\Rightarrow a = 40 - \frac{1.96}{\sqrt{3}}$$

De igual forma se calcula para b, y se tiene:

$$\Rightarrow b = 40 + \frac{1.96}{\sqrt{3}}$$

Finalmente, el intervalo de confianza para la media será: $I = \left[40 - \frac{1.96}{\sqrt{3}}, 40 + \frac{1.96}{\sqrt{3}} \right]$.

Problema 3

Se dispone de 10 muestras de sangre, tomadas en las mismas condiciones a una misma persona. Se obtiene para cada una, las dosis de Colesterol (en grs.): 245, 248, 250, 247, 249, 247, 247, 246, 246, 248.

Cada medida puede considerarse como una realización particular de la variable X “tasa de colesterol”. Se supondrá $X \rightarrow N(\mu, \sigma^2)$.

a) Dé el nivel de confianza para μ al 95%, suponiendo $\sigma^2 = 1.5$

b) Dé el intervalo de confianza para la media cuando se desconoce la varianza, al mismo nivel de confianza.

Solución:

a) El estadístico a formar es: $Z = \left(\frac{\bar{X} - \mu}{\sigma} \right) \cdot \sqrt{n} \rightarrow N(0,1)$

$$\text{Pero } \sigma^2 = 1.5, n=10, \bar{X} = 247.5 = \frac{\sum X_i}{n}$$

$$P(a < \mu < b) = 1 - \alpha$$

$$P(-b < \mu < -a) = P\left[\left(\frac{247.5 - b}{\sqrt{1.5}}\right) \cdot \sqrt{10} < \left(\frac{247.5 - b}{\sqrt{1.5}} \mu\right) \cdot \sqrt{10} < \left(\frac{247.5 - a}{\sqrt{1.5}}\right) \cdot \sqrt{10}\right] = 1 - \alpha$$

$$\Rightarrow P(Z > z_2) = 0.025$$

Por tablas se obtiene que:

$$z_2 = 1.96 = \left(\frac{247.5 - a}{\sqrt{1.5}}\right) \cdot \sqrt{10}$$

$$247.5 - 1.96 \cdot \sqrt{\frac{3}{20}} = a$$

$$247.5 + 1.96 \cdot \sqrt{\frac{3}{20}} = b$$

Luego :

$$a \approx 246.7409$$

$$b \approx 248.2591$$

Entonces, el intervalo de confianza para la media es: [246.7409, 248.2591]

b) Al igual que en ocasiones anteriores, como la varianza no se conoce, formamos la t-student.

Luego:

$$\Rightarrow T = \frac{(\bar{X} - \mu) \cdot \sqrt{n-1}}{S_n} \rightarrow t_{n-1}$$

Pero, no conocemos S_n . (El resto de los valores se entregan en el enunciado.). Por definición:

$$S_n^2 = \frac{1}{n} \cdot \sum (X_i - \bar{X})^2 = 2.05 \Rightarrow S_n \approx 1.4318$$

$$\Rightarrow T = \frac{(247.5 - \mu) \cdot \sqrt{9}}{\sqrt{2.05}}$$

Entonces:

$$P(a < \mu < b) = 1 - \alpha$$

$$P(-b < \mu < -a) = 1 - \alpha$$

$$P\left[\frac{(247.5-b)\cdot\sqrt{9}}{\sqrt{2.05}} < \frac{(247.5-\mu)\cdot\sqrt{9}}{\sqrt{2.05}} < \frac{(247.5-a)\cdot\sqrt{9}}{\sqrt{2.05}}\right] = 0.95$$

$$P(t_1 < T < t_2) = 0.95$$

$$P(T > t_2) = 0.025$$

Como:

$$T \rightarrow t_9 \Rightarrow t_2 = 2.262 \text{ , obtenido de las tablas}$$

Igualando términos:

$$t_2 = 2.262 = \left(\frac{247.5-a}{\sqrt{2.05}}\right) \cdot \sqrt{9}$$

$$247.5 - \frac{2.262}{3} \cdot \sqrt{2.05} = a$$

$$247.5 + \frac{2.262}{3} \cdot \sqrt{2.05} = b$$

Luego :

$$a \approx 246.4204$$

$$b \approx 248.5796$$

Entonces, el intervalo de confianza para la media es: [246.4204, 248.5796]

Problema 4

Sea X una v.a. de distribución exponencial de parámetro β : $f(x|\beta) = \beta e^{-\beta x}$ si $x > 0$.

a) Muestre que la v.a. $Y = \frac{X}{a}$, en que a es una constante positiva, sigue una distribución exponencial de parámetro $a\beta$.

b) Sean los valores muestrales x_1, x_2, \dots, x_n de X . Dé el estimador de Máxima Verosimilitud de β .

c) Sea x_1, x_2, \dots, x_n una muestra aleatoria simple proveniente de la distribución $N(1, 4)$. Calcule la probabilidad de que el promedio \bar{x} pertenezca al intervalo $\left[1 - 1.96 \frac{2}{\sqrt{n}}, 1 + 1.96 \frac{2}{\sqrt{n}}\right]$.

d) Supongamos que se extraen 100 muestras independientes del mismo tamaño. Deduzca el número esperado de veces que la media muestral \bar{x} pertenezca al intervalo $\left[1 - 1.96 \frac{2}{\sqrt{n}}, 1 + 1.96 \frac{2}{\sqrt{n}}\right]$.

Solución:

$$\text{a) } Y = \frac{X}{a} \Rightarrow F_Y(y) = P(Y \leq y) = P(X \leq ay) = F_X(ay) \Rightarrow$$

$$f_Y(y) = af_X(ay) = a\beta e^{-a\beta y} \Rightarrow Y \sim \exp(a\beta)$$

$$\text{b) } L = \beta^n e^{-\beta \sum x_i} \Rightarrow \log(L) = n \log(\beta) - \beta \sum x_i$$

$$\frac{\partial \log(L)}{\partial \beta} = 0 \Rightarrow \frac{n}{\hat{\beta}} = \sum x_i \Rightarrow \hat{\beta} = \frac{1}{\bar{x}}.$$

$$\text{c) } X_i \sim N(1, 4). \Rightarrow \bar{x} \sim N(1, \frac{4}{n}). \text{ Si } Z \sim N(0, 1). \Rightarrow P(|Z| \leq 1.96) = 0.95$$

$$\Rightarrow P(\bar{x} \in [1 - 1.96 \frac{2}{\sqrt{n}}, 1 + 1.96 \frac{2}{\sqrt{n}}]) = 0.95.$$

d) Si se repiten 100 muestras, se esperan que 95 de las muestras tengan su media muestral que cae en el intervalo.

Problema 5

Una fábrica trabaja con dos máquinas, una de tipo A y la otra de tipo B. El costo semanal X de reparación para las máquinas de tipo A sigue una distribución normal $N(\mu, \sigma^2)$. El costo semanal Y de reparación para las máquinas de tipo B sigue una distribución normal $N(\mu, 2\sigma^2)$.

Se tienen dos muestras independientes: una muestra aleatoria simple x_1, x_2, \dots, x_n de costos para las máquinas de tipo A y una muestra aleatoria simple y_1, y_2, \dots, y_m de costos para las máquinas de tipo B y se consideran los estimadores de μ de la forma

$$\hat{\mu} = \sum_{i=1}^n a_i X_i + \sum_{j=1}^m b_j Y_j.$$

a) Determine una condición sobre los a_i y b_j para que el estimador $\hat{\mu}$ de μ sea insesgado.

b) Determine los a_i y b_j para que el estimador $\hat{\mu}$ de μ sea insesgado y de varianza mínima. Deduzca la expresión de $\hat{\mu}$.

c) El estimador $\hat{\mu}$ de b) es de mínima varianza entre los estimadores insesgados de la forma $\hat{\mu} = \sum_{i=1}^n a_i X_i + \sum_{j=1}^m b_j Y_j$. Usando la desigualdad de Cramer-Rao verifique si $\hat{\mu}$ es de mínima varianza entre todos los estimadores insesgados de μ .

d) Determine los a_i y b_j para que el estimador $\mu^* = \sum_{i=1}^n a_i X_i + \sum_{j=1}^m b_j Y_j$ minimice

$$E\{(\mu^* - \mu)^2\}.$$

e) Deduzca la varianza de μ^* . ¿Cuál de $\hat{\mu}$ o μ^* tiene la varianza más pequeña?

Solución:

$$a) \quad E(\hat{\mu}) = \sum_{i=1}^n a_i E(X_i) + \sum_{j=1}^m b_j E(Y_j) = \mu \left(\sum_{i=1}^n a_i + \sum_{j=1}^m b_j \right) = \mu \Rightarrow \boxed{\sum_{i=1}^n a_i + \sum_{j=1}^m b_j = 1}$$

$$b) \quad V(\hat{\mu}) = \sum_{i=1}^n a_i^2 \text{Var}(X_i) + \sum_{j=1}^m b_j^2 \text{Var}(Y_j) = \sigma^2 \left(\sum_{i=1}^n a_i^2 + 2 \sum_{j=1}^m b_j^2 \right). \text{ Hay que minimizar } \text{Var}(\hat{\mu})$$

$$\text{sueto a } \sum_{i=1}^n a_i + \sum_{j=1}^m b_j = 1 \Rightarrow Q = \text{Var}(\hat{\mu}) - \lambda \left(\sum_{i=1}^n a_i + \sum_{j=1}^m b_j - 1 \right)$$

$$\left. \begin{aligned} \frac{\partial Q}{\partial a_i} = 0 &\Rightarrow 2\sigma^2 a_i - \lambda = 0 \Rightarrow a_i = \frac{\lambda}{2\sigma^2} \quad (a_i \text{ todos iguales}) \\ \frac{\partial Q}{\partial b_j} = 0 &\Rightarrow 4\sigma^2 b_j - \lambda = 0 \Rightarrow b_j = \frac{\lambda}{4\sigma^2} \quad (b_j \text{ todos iguales}) \end{aligned} \right\} \Rightarrow a_i = 2b_j$$

$$\text{De } \sum_{i=1}^n a_i + \sum_{j=1}^m b_j = 1 \text{ se deduce que } \boxed{a_i = \frac{2}{2n+m}, b_j = \frac{1}{2n+m}} \text{ y}$$

$$\boxed{\hat{\mu} = \left(\frac{1}{2n+m} \right) \left(2 \sum_i X_i + \sum_j Y_j \right)}$$

Es decir que es una media ponderada de todas las

observaciones, en donde los X_i tiene un peso doble del peso de los Y_j .

c) Calculemos la cota inferior de la desigualdad de Cramer-Rao. La función de verosimilitud de las $(n+m)$ valores muestrales es:

$$\begin{aligned} L &= \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp\left(-\frac{1}{2\sigma^2} \sum_i (X_i - \mu)^2\right) \left(\frac{1}{\sqrt{4\pi\sigma^2}} \right)^m \exp\left(-\frac{1}{4\sigma^2} \sum_j (Y_j - \mu)^2\right) \\ \log(L) &= -\left(\frac{n}{2}\right) \log(2\pi\sigma^2) - \left(\frac{m}{2}\right) \log(4\pi\sigma^2) - \frac{1}{\sigma^2} \left\{ \frac{1}{2} \sum_i (X_i - \mu)^2 + \frac{1}{4} \sum_j (Y_j - \mu)^2 \right\} \\ \frac{\partial \log(L)}{\partial \mu} &= \frac{1}{\sigma^2} \left\{ \sum_i (X_i - \mu) + \frac{1}{2} \sum_j (Y_j - \mu) \right\} \\ \frac{\partial^2 \log(L)}{\partial \mu^2} &= -\frac{1}{\sigma^2} \left\{ n + \frac{m}{2} \right\} = -\frac{1}{\sigma^2} \left\{ \frac{2n+m}{2} \right\} \Rightarrow \boxed{\text{Var}(\hat{\mu}) \geq \frac{2\sigma^2}{2n+m}} \end{aligned}$$

Por otro lado:

$Var(\hat{\mu}) = \sigma^2 \left(\sum_i a_i^2 + 2 \sum_j b_j^2 \right) = \frac{2\sigma^2}{2n+m} \Rightarrow \hat{\mu}$ es un estimador insesgado de mínima varianza para μ .

d) Se tiene $\mu^* = \sum_{i=1}^n a_i X_i + \sum_{j=1}^m b_j Y_j$ y $E\{(\mu^* - \mu)^2\} = Var(\mu^*) + \{E(\mu^* - \mu)\}^2$.

$$Var(\mu^*) = \sigma^2 \left(\sum_i a_i^2 + 2 \sum_j b_j^2 \right) \text{ y } E(\mu^* - \mu) = \mu \left(\sum_i a_i + \sum_j b_j - 1 \right)$$

$$E\{(\mu^* - \mu)^2\} = \sigma^2 \left(\sum_i a_i^2 + 2 \sum_j b_j^2 \right) + \mu^2 \left(\sum_i a_i + \sum_j b_j - 1 \right)^2$$

Minimizando:

$$\frac{\partial E\{(\mu^* - \mu)^2\}}{\partial a_i} = 2\sigma^2 a_i + 2\mu^2 \left(\sum_i a_i + \sum_j b_j - 1 \right) = 0 \Rightarrow a_i = \frac{\mu^2}{\sigma^2} \left(1 - \sum_i a_i - \sum_j b_j \right)$$

$$\frac{\partial E\{(\mu^* - \mu)^2\}}{\partial b_j} = 2\sigma^2 b_j + 2\mu^2 \left(\sum_i a_i + \sum_j b_j - 1 \right) = 0 \Rightarrow b_j = \frac{\mu^2}{2\sigma^2} \left(1 - \sum_i a_i - \sum_j b_j \right)$$

Luego todos los b_j son iguales y todos los a_i son iguales a $2b_j$. Sea $b_j = b$ y $a_i = 2b$.

$$\text{Entonces } \sum_i a_i = \frac{n\mu^2}{\sigma^2} \left(1 - \sum_i a_i - \sum_j b_j \right) \text{ o sea } 2nb = \frac{n\mu^2}{\sigma^2} (1 - 2nb - mb)$$

$$\text{Finalmente: } b = \frac{\mu^2}{2\sigma^2 + (2n+m)\mu^2} \text{ y } \mu^* = \sum_{i=1}^n a_i X_i + \sum_{j=1}^m b_j Y_j = 2b \sum X_i + b \sum Y_j$$

$$\mu^* = b(2 \sum X_i + \sum Y_j) = \frac{\mu^2}{2\sigma^2 + (2n+m)\mu^2} (2 \sum X_i + \sum Y_j)$$

e) De aquí se deduce la varianza de μ^* :

$$Var(\mu^*) = \left(\frac{\mu^2}{2\sigma^2 + (2n+m)\mu^2} \right)^2 (4n\sigma^2 + m\sigma^2) = \frac{2(2n+m)\sigma^2}{(2\sigma^2 / \mu^2 + (2n+m))^2}$$

$$\text{Como } Var(\hat{\mu}) = \frac{2\sigma^2}{2n+m} = \frac{2(2n+m)\sigma^2}{(2n+m)^2}, \text{ se deduce que } \boxed{Var(\hat{\mu}) > Var(\mu^*)}$$

Problema 6

- a) Sea x_1, x_2, \dots, x_n una muestra aleatoria simple proveniente de la distribución $N(1,4)$. Calcule la probabilidad de que el promedio \bar{x} pertenezca al intervalo $[1 - 1.65 \frac{2}{\sqrt{n}}, 1 + 1.65 \frac{2}{\sqrt{n}}]$.
- b) Dé el largo del intervalo obtenido en 1.1 para una muestra aleatoria de tamaño 100. Para este mismo intervalo dé el tamaño de muestra requerido para tener un nivel de confianza de 95%.

Solución:

a) $X_i \sim N(1,4) \Rightarrow \bar{x} \sim N(1, \frac{4}{n})$. Si $Z = \frac{\bar{x} - 1}{2/\sqrt{n}} \sim N(0,1)$. $\Rightarrow P(|Z| \leq 1.65) = 0.90$
 $\Rightarrow P(\bar{x} \in [1 - 1.65 \frac{2}{\sqrt{n}}, 1 + 1.65 \frac{2}{\sqrt{n}}]) = 0.90$

b) Tenemos un intervalo a 90% de confianza y $n=100$,

$$P(\bar{x} \in [1 - 1.65 \frac{2}{\sqrt{n}}, 1 + 1.65 \frac{2}{\sqrt{n}}]) = P(\bar{x} \in [1 - 1.65 \frac{2}{10}, 1 + 1.65 \frac{2}{10}]) = [0.67, 1.33].$$

Para el tamaño de muestra n : $P(\bar{x} \in [0.67, 1.33]) = 0.95$

Ahora bien $\sqrt{n}(\bar{x} - 1) / 2 \sim N(0,1)$, luego $P(\sqrt{n}(\bar{x} - 1) / 2 \in [-1.96, 1.96]) = 0.95$. o sea

$$P(\bar{x} \in [1 - 1.96 \frac{2}{\sqrt{n}}, 1 + 1.96 \frac{2}{\sqrt{n}}]) = 0.95 \Rightarrow [1 - 1.96 \frac{2}{\sqrt{n}}, 1 + 1.96 \frac{2}{\sqrt{n}}] = [0.67, 1.33]$$

Finalmente: $1 + 1.96 \frac{2}{\sqrt{n}} = 1.33$ y $n > 141$.

Problema 7

Una fábrica de cecinas encarga a dos empresas de estudios de mercados *BOL* y *MIOCHE* que estimen el consumo promedio mensual μ de jamón por persona en la población chilena. La empresa *BOL* obtiene los consumos mensuales de jamón por persona sobre una muestra aleatoria simple $\{X_1, X_2, \dots, X_n\}$ de tamaño n de media muestral

$\bar{X} = \frac{1}{n} \sum_i X_i$ y varianza muestral $S^2 = \frac{1}{n} \sum_i (X_i - \bar{X})^2$ y la empresa *MIOCHE* una muestra

aleatoria simple $\{Y_1, Y_2, \dots, Y_m\}$ de tamaño m de media muestral $\bar{Y} = \frac{1}{m} \sum_i Y_i$ y varianza

muestral $T^2 = \frac{1}{m} \sum_i (Y_i - \bar{Y})^2$. Como las estimaciones de μ , \bar{X} y \bar{Y} , dadas respectivamente por *BOL* y por *MIOCHE*, no son iguales, la fábrica decide combinar las dos estimaciones y propone usar el estimador $\hat{\mu}_1 = \frac{1}{2} \bar{X} + \frac{1}{2} \bar{Y}$.

a) Calcule la esperanza y la varianza del estimador en función de la varianza σ^2 en la población. Deduzca las propiedades del estimador.

b) Sea $\hat{\mu}_a = a\bar{X} + (1-a)\bar{Y}$ en donde $0 \leq a \leq 1$. Encuentre el valor de 'a' que minimiza la varianza de $\hat{\mu}_a$. Concluye.

c) Bajo el supuesto de normalidad del consumo obtenga el estimador $\hat{\mu}_2$ de máxima verosimilitud. Calcule su esperanza y varianza. Concluye sobre sus propiedades. ¿Cual de los dos estimadores $\hat{\mu}_1$ y $\hat{\mu}_2$ les parece mejor?

d) Dé las distribuciones de los estadísticos $\frac{nS^2}{\sigma^2}$, $\frac{mT^2}{\sigma^2}$ y $\frac{(n+m-2)\sigma^2}{\sigma^2}$ y sus varianzas

donde $\sigma^2 = \frac{1}{n+m-2}(nS^2 + mT^2)$.

Solución:

a) $E(\hat{\mu}_1) = \frac{1}{2}E(\bar{X}) + \frac{1}{2}E(\bar{Y}) = \mu$.

Dado que las muestras son independientes

$$Var(\hat{\mu}_1) = \frac{1}{4}Var(\bar{X}) + \frac{1}{4}Var(\bar{Y}) = \frac{1}{4}\left(\frac{1}{n} + \frac{1}{m}\right)\sigma^2.$$

$\hat{\mu}_1 = \frac{1}{2}\bar{X} + \frac{1}{2}\bar{Y}$ es insesgado, de varianza que converge a 0 cuando n y/o m tienden a infinito. Es consistente.

b) Sea $\hat{\mu}_a = a\bar{X} + (1-a)\bar{Y}$ en donde $0 \leq a \leq 1$.

$$Var(\hat{\mu}_a) = a^2Var(\bar{X}) + (1-a)^2Var(\bar{Y}) = \left[\frac{a^2}{n} + \frac{(1-a)^2}{m}\right]\sigma^2.$$

Se obtiene un mínimo para $a = \frac{n}{n+m}$.

Se puede concluir que si $n \neq m$ el estimador insesgado $\hat{\mu}_1 = \frac{1}{2}\bar{X} + \frac{1}{2}\bar{Y}$ no es de mínima varianza.

c) El estimador $\hat{\mu}_2$ de máxima verosimilitud es igual a $\hat{\mu}_2 = \frac{n\bar{X} + m\bar{Y}}{n+m}$ que es la solución

óptima de 2.2. $E(\hat{\mu}_2) = \mu$. $Var(\hat{\mu}_2) = \frac{\sigma^2}{n+m}$. Se prefiere $\hat{\mu}_2$ ya que es insesgado y tiene menos varianza que $\hat{\mu}_1$.

$$d) \frac{nS^2}{\sigma^2} \sim \chi_{n-1}^2, \frac{mT^2}{\sigma^2} \sim \chi_{m-1}^2 \text{ y } \frac{(n+m-2)\sigma^2}{\sigma^2} \sim \chi_{n+m-2}^2.$$

Problema 8

La fábrica LUMINA S.A. produce ampolletas de uso residencial. El jefe de la sección de producción lo ha contratado a Ud. para que lo ayude a estimar la tasa de vida útil de su producto estrella, la ampolleta DURALUM. De acuerdo a sus conocimientos, Usted sabe que la vida útil de una ampolleta X (en años) se puede modelar según la distribución exponencial siguiente: $f(x | \lambda) = \lambda e^{-\lambda x}$ con $x > 0$

El jefe ha tomado previamente una muestra de 30 ampolletas y encuentra que la duración promedio de ellas es de 5 años. Se define la variable $y = \sum_{i=1}^{30} x_i$.

- Proponga un intervalo de confianza al 95% para la tasa de vida útil λ (HINT: use el Teorema Central del Límite para la variable “y”, resolviendo la ecuación cuadrática asociada).
- Muestre que la variable $T = 2\lambda Y$ sigue una distribución Chi-cuadrado con “2n” grados de libertad (Hint: utilice la función generatriz de los momentos o la función característica).
- Encuentre L y U tales que $P(U < T < L) = 0,95$ y luego deduzca un intervalo de confianza al 95% para λ .
- Compare los dos intervalos de confianza. ¿Cuál intervalo preferiría Ud. Cuando el tamaño muestral es menor a 25 ? Explique.

Solución:

a) Un intervalo de confianza para λ utilizando $y = \sum_{i=1}^{30} x_i$ equivale a encontrar a y b tales

que $P(a \leq \lambda \leq b) = 1 - \alpha = 0.95$

Para utilizar el Teorema central del límite calculamos

$$E(y) = E(\sum x_i) = nE(x_i) = \frac{n}{\lambda} \quad \text{y la } Var(y) = \frac{n}{\lambda^2}$$

$$\text{Luego } z = \frac{y - E(y)}{\sqrt{Var(y)}} = \frac{y - \frac{n}{\lambda}}{\sqrt{\frac{n}{\lambda^2}}} \longrightarrow N(0,1)$$

$$\Rightarrow P(a \leq \lambda \leq b) = P\left(\frac{1}{b} \leq \frac{1}{\lambda} \leq \frac{1}{a}\right) = P\left(\frac{y - \frac{n}{b}}{\sqrt{\frac{n}{\lambda^2}}} \leq \frac{y - \frac{n}{\lambda}}{\sqrt{\frac{n}{\lambda^2}}} \leq \frac{y - \frac{n}{a}}{\sqrt{\frac{n}{\lambda^2}}}\right) = P(-z_1 \leq z \leq z_1) = 0.95$$

$$z_1 = 1,96 \Rightarrow P\left(\left|\frac{y - \frac{n}{\lambda}}{\sqrt{\frac{n}{\lambda^2}}}\right| \leq 1,96\right) \text{ Luego desarrollamos la ecuación cuadrática asociada.}$$

$$\left|\frac{y - \frac{n}{\lambda}}{\sqrt{\frac{n}{\lambda^2}}}\right| \leq 1,96 /^2 \Rightarrow \frac{(y^2 - 2\frac{ny}{\lambda} + \frac{n^2}{\lambda^2})}{\frac{n}{\lambda^2}} \leq 3,8416 \text{ Desarrollando la ecuación de segundo grado}$$

quedaría cómo: $y^2 \lambda^2 - 2ny\lambda + n^2 - 3,8416n \leq 0$ desarrollando para la igualdad para $n=30$ e $y=150$ años.

$$\lambda = \frac{n}{y} \pm \frac{z}{y} \sqrt{n} \Rightarrow P(0,12843 \leq \lambda \leq 0,27157) = 0,95$$

b) Utilizando la función generatriz de los momentos:

$$\varphi_Y(t) = E(e^Y) = [E(e^X)]^n = \left(\frac{\lambda}{\lambda - t}\right)^n.$$

$$E(e^{2\lambda Y}) = \varphi_Y(2\lambda t) = \left(\frac{\lambda}{\lambda - 2\lambda t}\right)^n = \left(\frac{1}{1 - 2t}\right)^n \text{ que es la función generatriz de los}$$

momentos de la χ^2_{2n} .

c) $P(L \leq T \leq U) = 0,95$ La distribución de T ya es conocida por lo que solo se necesita calcular los valores para L y U.

$T \longrightarrow \chi^2_{2n}$ tomando colas equiprobables $\alpha/2$.

$$P(T \leq L) = 0.025 \text{ y } P(T \geq U) = 0.025 \text{ Luego } L = 40,482 \text{ y } U = 83,298$$

$$\text{Luego } \frac{L}{2y} = 0,13494 \text{ y } \frac{U}{2y} = 0,27766 \Rightarrow P(0,13494 \leq \lambda \leq 0,27766) = 0,95$$

d) Por TCL.