PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ
FACULTAD DE CIENCIAS SOCIALES
ESTADÍSTICA INFERENCIAL
SEMESTRE 2019-I

SOLUCIONARIO PC4

1. **Problema 1 (6 puntos)**

Para el ingreso semanal (en dólares) de los microempresarios dedicados al reciclaje, un economista asume que éstos tienen una distribución Exponencial $Exp(x;\beta = 1/\theta)$ y desea "estimar" (aproximar) el ingreso medio μ_X usando la estadística $\overline{X}$ de una muestra "grande"de ingresos de n de este tipo de microempresarios.

a) **(3 puntos)** ¿La estimación que se obtenga, en valor esperado "tiende a coincidir" con μ_X? Si se desea que con 90 % de probabilidad o confiabilidad , el error de estimación o aproximación $\left|\overline{X} - \mu_X\right|$ no pase del 10 % del valor real de μ_X ¿Qué tamaño de muestra n se debiera usar?

Sabemos que

$$X \sim Exp(x;\beta = 1/\theta)$$

Luego

$$E\left(X\right) = \mu_X = \frac{1}{\beta} \rightarrow \mu_X = \theta, V\left(X\right) = \frac{1}{\beta^2} \rightarrow V\left(X\right) = \theta^2$$

$$E\left(\overline{X}\right) = E\left(\frac{\sum X}{n}\right) = \frac{\sum E\left(X\right)}{n} = \frac{n\theta}{n} = \theta = \mu_X$$

La estimación de $\overline{X}$ sí tiende a coincidir con θ, luego dado que $n > 30$

$$\overline{X} \sim N\left(\mu_X = \theta, \frac{V\left(X\right)}{n}\right)$$

Por lo tanto:

$$\overline{X} \sim N\left(\theta, \frac{\theta^2}{n}\right)$$

Nos piden:

$$P\left(\left|\overline{X} - \mu_X\right| \leq 0{,}1\mu_X\right) = 0{,}90$$

$$P\left(\left|\overline{X} - \theta\right| \leq 0{,}1\theta\right) = 0{,}90$$

Dado que $\overline{X}$ se distribuye como una normal con media θ, estandarizamos a una normal estandar para ello deberíamos restarle su media (θ) y dividirlo con respecto a su desviación estandar $\sqrt{\theta^2/n}$.

$$P\left(\frac{\left|\overline{X} - \theta\right|}{\frac{\theta}{\sqrt{n}}} \leq \frac{0{,}1\theta}{\frac{\theta}{\sqrt{n}}}\right) = 0{,}90$$

$$P\left(\left|Z\right| \leq 0{,}1\sqrt{n}\right) = 0{,}90$$

$$P\left(Z \leq 0{,}1\sqrt{n}\right) = 0{,}95$$

Luego

$$0{,}1\sqrt{n} = 1{,}645$$

$$n = 270{,}6025$$

$$n \approx 271$$

b) **(3 puntos)**Un colega del economista critica la "*especificación del modelo de datos*" rechazando el supuesto de distribución exponencial y sugiere que en verdad la distribución de ingresos es una uniforme $U(x; \alpha = 0, \beta = 2\theta)$ y concluyó que la confiabilidad, o probabilidad, de lograr el objetivo de estimación planteado en a) es menor de lo propuesto. Bajo el supuesto que la crítica fuera fundada, sería cierta la conclusión?

Sabemos que

$$X \sim U(x; \alpha = 0, \beta = 2\theta)$$

Luego

$$E(X) = \mu_X = \frac{2\theta + 0}{2} = \theta \to \mu_X = \theta, V(X) = \frac{(2\theta - 0)^2}{12} = \frac{4\theta^2}{12} \to V(X) = \frac{\theta^2}{3}$$

$$\overline{X} \sim N\left(\mu_X, \frac{V(X)}{n}\right), \text{ Asumiendo } n = 271 > 30$$

Por lo tanto:

$$\overline{X} \sim N\left(\theta, \frac{\theta^2}{3n}\right)$$

Nos piden:

$$P\left(\left|\overline{X} - \mu_X\right| \leq 0{,}1\mu_X\right)$$

$$P\left(\left|\overline{X} - \theta\right| \leq 0{,}1\theta\right)$$

$$P\left(\frac{\left|\overline{X} - \theta\right|}{\frac{\theta}{\sqrt{3n}}} \leq \frac{0{,}1\theta}{\frac{\theta}{\sqrt{3n}}}\right)$$

$$P\left(|Z| \leq 0{,}1\sqrt{3n}\right)$$

$$P(|Z| \leq 2{,}85)$$

$$2P(Z \leq 2{,}85) - 1 = 2(0{,}9978) - 1 = 0{,}9956$$

Falso: Las probabilidades de que el margen de error absoluto caiga dentro de los límites establecidos (10 % de la media) es mayor con una distribución uniforme.

2. **Problema 2 (8 puntos)**

El tiempo X que pasa un consumidor cualquiera visitando la página web de una empresa que para hacer un pedido, es una v.a. con distribución Gamma $X \sim \Gamma(x; \alpha = 2, \beta)$ y se registra una m.a. de los tiempos $(X_1, X_2, ..., X_n)$ relativos a n consumidores con el objetivo de aproximar el valor del parámetro β usando la estadística $W = c\overline{X}$:

a) **(2 puntos)**Halle el valor de c de modo que "en promedio" W coincida con β, esto es, de modo que $E(W) = \beta$.

$$E(X) = \alpha\beta = 2\beta, V(X) = \alpha\beta^2 = 2\beta^2$$

$$W = c\overline{X} = c\frac{\sum X_i}{n}$$

$$E(W) = c\frac{\sum E(X_i)}{n} = c\frac{\sum E(X_i)}{n} = c\frac{2n\beta}{n} = 2c\beta$$

Luego

$$E(W) = \beta$$

$$2c\beta = \beta$$

$$c = 1/2$$

b) **(2 puntos)** En a) se decide usar un tamaño de muestra n grande. ¿Qué tamaño de muestra garantizaría que con 95 % de probabilidad, W diferirá de β (en valor absoluto) en menos de un 10 % de β, esto es que $P(|W-\beta| \leq 0{,}1\beta) = 0{,}95$?

Sabemos que para un tamaño de muestra n grande, $\overline{X}$ por el teorema del límite central, se distribuye con una distribución normal

$$\overline{X} \sim N\left(\mu_X, \frac{V(X)}{n}\right)$$

$$\overline{X} \sim N\left(2\beta, \frac{2\beta^2}{n}\right)$$

Luego,

$$P(|W-\beta| \leq 0{,}1\beta) = 0{,}95$$

$$P\left(\left|\frac{\overline{X}}{2} - \beta\right| \leq 0{,}1\beta\right) = 0{,}95$$

$$P\left(\left|\overline{X} - 2\beta\right| \leq 0{,}2\beta\right) = 0{,}95$$

$$P\left(\frac{\left|\overline{X} - 2\beta\right|}{\sqrt{\frac{2\beta^2}{n}}} \leq \frac{0{,}2\beta}{\sqrt{\frac{2\beta^2}{n}}}\right) = 0{,}95$$

$$P\left(|Z| \leq \frac{0{,}2\beta}{\sqrt{\frac{2}{n}}\beta}\right) = 0{,}95$$

$$P\left(|Z| \leq \frac{0{,}1 \times 2 \times \sqrt{n}}{\sqrt{2}}\right) = 0{,}95$$

$$P\left(|Z| \leq 0{,}1\sqrt{2n}\right) = 0{,}95$$

$$P\left(Z \leq 0{,}1\sqrt{2n}\right) = 0{,}975$$

$$0{,}1\sqrt{2n} = 1{,}96$$

$$n = 192{,}08$$

$$n \approx 193$$

c) **(2 puntos)** En b), si sólo hay dinero para tomar una muestra de $n = 36$ casos ¿Cuál sería ahora la probabilidad de lograr el objetivo $|W-\beta| \leq 0{,}1\beta$?

$$W \sim N\left(\beta, \frac{\beta^2}{2(36)}\right)$$

Luego

$$P(|W-\beta| \leq 0{,}1\beta)$$

$$P\left(\frac{|W-\beta|}{\frac{\beta}{\sqrt{2(36)}}} \leq \frac{0{,}1\beta}{\frac{\beta}{\sqrt{2(36)}}}\right)$$

$$P\left(|Z| \leq 0{,}1\sqrt{2(36)}\right)$$

$$2P(Z \leq 0{,}85) - 1$$

$$2(0{,}8023) - 1$$

$$P(|W-\beta| \leq 0{,}1\beta) = 0{,}6046$$

d) **(2 puntos)** En b), si sólo hay dinero para tomar una muestra de $n = 36$ casos ¿Cuál es la máxima diferencia $|W - \beta|$ (como % de β) que usted puede garantizar con 95 % de probabilidad?

$$W \sim N\left(\beta, \frac{\beta^2}{2\,(36)}\right)$$

Luego

$$P\left(|W - \beta| \leq r\beta\right) = 0{,}95$$

$$P\left(\frac{|W - \beta|}{\frac{\beta}{\sqrt{2(36)}}} \leq \frac{0{,}1\beta}{\frac{\beta}{\sqrt{2(36)}}}\right)$$

$$P\left(|Z| \leq r\sqrt{2\,(36)}\right)$$

$$P\left(Z \leq 6r\sqrt{2}\right) = 0{,}975$$

$$6r\sqrt{2} = 1{,}96$$

$$r = 23{,}09\,\%$$

3. **Problema 3 (6 puntos)**

Para aproximar la tasa de subempleo θ en un sector de la economía, se genera un a variable indicadora de "sub-empleo por ingresos" X que se mide sobre cada miembro de la PEA del sector, mediante: $X = 1$ si la persona está subempleada; $X = 0$ si la persona no está subempleada. Considerando que se toma al azar a un miembro de la PEA del sector, se plantea que $P\,(X = 1) = \theta = Tasa\ de\ Subempleo$. En este contexto, si se toma al azar una muestra de n miembros de la PEA y se registra X en cada uno, obteniendo una m.a.$(X_1, X_2, ..., X_n)$ de la población de X.

a) **(2 puntos)** Halle o tabule la función de probabilidad de X y luego μ_X y σ_X^2

$$P(x) = \theta^X\,(1 - \theta)^{1-X}\,, X = 0, 1$$

Dado que la forma de distribución que hemos planteado corresponde a una binomial, recordamos que

$$E\,(X) = np, Var\,(X) = npq$$

Dado que en este caso $n = 1, p = \theta$ y $q = 1 - \theta$, entonces:

$$E\,(X) = \theta, Var\,(X) = \theta\,(1 - \theta)$$

b) **(2 puntos)** Al tomar una pequeña m.a. piloto de tamaño $n = 5$ se obtuvo $X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 0, X_5 = 1$ y le plantean que θ sólo puede ser $\theta = 0{,}2$ o $\theta = 0{,}6$ ¿Cuál valor escogería? Justifique.

Halloamos la probabilida de la muestra:

$$P\,(1, 1, 0, 0, 1) = P\,(X = 1)^3\,P\,(X = 0)^2 = \theta^3\,(1 - \theta)^2$$

Luego evaluamos la probabilidad para $\theta = 0{,}2$ y $\theta = 0{,}6$

$$P\,(1, 1, 0, 0, 1 \mid \theta = 0{,}2) = 0{,}00512$$

$$P\,(1, 1, 0, 0, 1 \mid \theta = 0{,}6) = 0{,}03456$$

Dado que la probabilidad de obtener los resultados conjuntos $X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 0, X_5 = 1$ es mayor cuando $\theta = 0{,}6$, entonces escogería este valor.

c) **(2 puntos)** Se desea aproximar θ usando la estadística $T = \sum_{j=1}^{n} c_j X_j$ que es una combinación lineal de los valores de la muestra aleatoria. Halle lo coeficientes $\{c_j\}$ de modo que $E(T) = \theta$ y $V(T)$ sea mínima.

Dada la distribución de X ,entonces

$$T = \sum_{j=1}^{n} c_j X_j$$

$$E(T) = \sum_{j=1}^{n} c_j E(X_j) = \sum_{j=1}^{n} c_j (\theta) = \theta$$

$$\sum_{j=1}^{n} c_j = 1$$

$$V(T) = \sum_{j=1}^{n} c_j^2 V(X_j) = \sum_{j=1}^{n} c_j^2 \theta (1-\theta) = \theta (1-\theta) \sum_{j=1}^{n} c_j^2$$

Luego dado que θ podríamos considerarla como un factor fijo (tasa de subempleo), el ejercicio pide que $V(T)$ sea mínima s.a $E(T) = \theta$

$$\min V(T)$$

$$s.a. E(T) = \theta$$

$$\min \theta (1-\theta) \sum_{j=1}^{n} c_j^2$$

$$s.a. \sum_{j=1}^{n} c_j = 1$$

Dado que θ y $(1-\theta)$ son positivos, el problema se puede reducir a:

$$\min \sum_{j=1}^{n} c_j^2$$

$$s.a. \sum_{j=1}^{n} c_j = 1$$

$$L = \sum_{j=1}^{n} c_j^2 - \lambda \left(\sum_{j=1}^{n} c_j - 1 \right)$$

$$\frac{\partial L}{\partial c_j} = 2c_j - \lambda = 0$$

$$c_j = \frac{\lambda}{2}$$

$$\sum_{j=1}^{n} c_j = 1$$

$$\sum_{j=1}^{n} \frac{\lambda}{2} = 1$$

$$\frac{n\lambda}{2} = 1$$

$$\lambda = \frac{2}{n}$$

$$c_j = \frac{1}{n}$$

Luego

$$c_1 = c_2 = ... = c_n = \frac{1}{n}$$

Por lo tanto el estimador sería $\overline{X}$.