PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ
FACULTAD DE CIENCIAS SOCIALES
ESPECIALIDAD DE ECONOMÍA

PRÁCTICA DIRIGIDA No. 4
CURSO: EST 241 Estadística Inferencial
PROFESOR: Arturo Calderón G.
HORARIO: 0621
FECHA: 01 de junio de 2019
SEMESTRE: 2019-1

Los problemas del 1 a 4 serán tratados durante la práctica. El resto es para el trabajo personal del alumno.

Problema 1

En un estudio sobre costos de transacción, se registró el tiempo X que tarda un expediente en ser tramitado ante una institución pública. Como 'modelo de datos' se asumió que $X \sim U(x; 0{,}2\theta)$ y se decidió tomar una m.a. $(X_1, X_2, \dots, X_n)$ de tamaño n para calcular la media muestral $\overline{X}$ y usarla como aproximación de θ.

a) Si la muestra es de tamaño $n = 2$, ¿Con qué probabilidad el promedio de la muestra será mayor que el 10% del valor de θ?

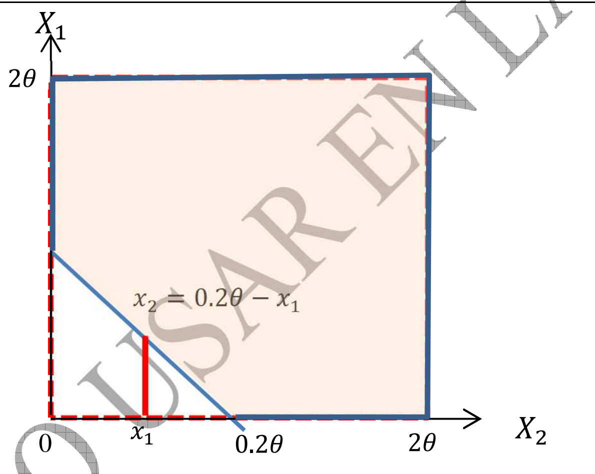
b) Como se usará $\overline{X}$ como aproximación de θ, el "error absoluto de aproximación o estimación" es $|\overline{X} - \theta|$ y lo usual es tratar de que este error no pase de una "margen de error" ε predeterminado. Por otra parte, el Teorema del Límite Central establece que cuando el tamaño de muestra n es grande $(n > 30)$, entonces $\overline{X} \sim N(\mu_X, \frac{\sigma_X^2}{n})$.
Aplique este teorema y para una m.a. de tamaño $n = 36$, calcule la probabilidad de que:
(1) El promedio de la muestra sea mayor que el 10% del valor de θ.
(2) El error absoluto de estimación pase del 25% del valor de θ.

c) En b), con 95% de probabilidad, si ε se fija como una fracción de θ: $\varepsilon = r\theta$, cuál es el valor de r?

d) Si se usa una muestra grande y se quiere que $|\overline{X} - \theta|$ no pase del 7% del valor de θcon 95% de probabilidad ¿Qué tamaño de muestra n necesitaría usar?

Solución:

a) (X_1, X_2) m.a. $\Rightarrow f_{X_1X_2}(x_1, x_2) = f_X(x_1) f_{X_2}(x_2) = \left(\frac{1}{2\theta}\right)^2 \quad 0 < x_1 \le 2\theta, 0 < x_2 \le 2\theta$



$$P(\overline{X} > 0.1\theta) = P(X_1 + X_2 > 0.2\theta) =$$

$$1 - P(X_1 + X_2 \le 0.2\theta) = 1 - \int_0^{0.2\theta} \int_0^{0.2\theta - x_1} \left(\frac{1}{2\theta}\right)^2 dx_2 dx_1 =$$

$$1 - \frac{1}{4\theta^2} \int_0^{0.2\theta} (0.2\theta - x_1) dx_1 = 1 - \frac{1}{4\theta^2} \left[0.2\theta x_1 - \frac{x_1^2}{2}\right]_0^{0.2\theta} =$$

$$1 - \frac{1}{4\theta^2}\left[(0.2\theta)^2 - \frac{(0.2\theta)^2}{2}\right] = 1 - 0.005 = 0.995$$

b) $n = 36 > 30 \Rightarrow \overline{X} \sim N(\mu_X, \frac{\sigma_X^2}{n})$ y de $X \sim U(x; 0{,}2\theta) \Rightarrow \mu_X = \frac{2\theta}{2}, \sigma_X^2 = \frac{(2\theta)^2}{12} = \frac{4\theta^2}{12} = \frac{\theta^2}{3}$ entonces

$$\overline{X} \sim N(\mu_X, \frac{\sigma_X^2}{n}) \Rightarrow \overline{X} \sim N(\theta, \frac{\theta^2}{3n}) \Rightarrow Z = \frac{\overline{X} - \theta}{\frac{\theta}{\sqrt{3n}}} \sim N(0,1)$$

(1) $P(\overline{X} > 0.1\theta) = P\left(\dfrac{\overline{X}-\theta}{\frac{\theta}{\sqrt{3n}}} > \dfrac{0.1\theta-\theta}{\frac{\theta}{\sqrt{3n}}}\right) = P(Z > -0.9\sqrt{3n}) = P(Z > -9.35) = 1$

(2) $P(|\overline{X}-\theta| > 0.25\theta) = P(|Z| > 2.6) = 1 - P(-2.6 \le Z \le 2.6) = 1 - (0.9953 - 0.0047) = 0.0094$

c) $n = 36$ y se pide hallar r tal que $0.95 = P(|\bar{X}-\theta| < r\theta) = P\left(|Z| < \frac{r\theta}{\frac{\theta}{\sqrt{3n}}}\right) = P(|Z| < r\sqrt{3n}) =$
$P(|Z| < 10.4r) \Leftrightarrow P(-10.4r < Z < 10.4r) = 0.95 \Rightarrow P(Z < 10.4r) = 0.975 \Rightarrow 10.4r = 1.96 \Rightarrow r =$ 0.188 $\Rightarrow$ El margen de error $\varepsilon = r\theta$ es de 18.8% del valor real de θ.

d) Se quiere que $|\overline{X}-\theta|$ no pase del 7% del valor de θ con 95% de probabilidad, entonces planteamos:

$$0.95 = P(|\overline{X}-\theta| \le 0.07\theta) = P\left(\frac{|\overline{X}-\theta|}{\frac{\theta}{\sqrt{3n}}} \le \frac{0.07\theta}{\frac{\theta}{\sqrt{3n}}}\right) = P(|Z| \le 0.07\sqrt{3n}) = P(|Z| \le 0.07\sqrt{3n}) =$$

$$P(-0.07\sqrt{3n} \le Z \le 0.07\sqrt{3n}) \Rightarrow P(Z \le 0.07\sqrt{3n}) = 0.975 \Rightarrow 0.07\sqrt{3n} = 1.96 \Rightarrow n = 261.3 \cong 262$$

Problema 2

Sea $X \sim Exp(\beta = 1/\theta)$. Para una muestra aleatoria de tamaño $n = 2$, (X_1, X_2), se tiene las estadísticas: $\hat{\theta} = \frac{X_1+X_2}{2}$ y $\tilde{\theta} = \sqrt{X_1X_2}$. Halle el valor esperado de estas estadísticas y sus varianzas.

Nota: Recuerde que $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ y $\Gamma(p) = (p-1)\Gamma(p-1) \quad si\ p > 1$

Solución:

$$E(\hat{\theta}) = E\left(\frac{X_1+X_2}{2}\right) = \frac{1}{2}E(X_1+X_2) = \frac{1}{2}(E(X_1)+E(X_2)) = \frac{1}{2}(\theta+\theta) = \theta.$$

$$E(\tilde{\theta}) = E(\sqrt{X_1X_2}) = E(\sqrt{X_1}\sqrt{X_2}) = E(X_1^{1/2}X_2^{1/2}) = E(X_1^{1/2})E(X_2^{1/2}) = E(X^{1/2})E(X^{1/2}) = (E(X^{1/2}))^2$$

y

$$E(X^{1/2}) = \int_0^\infty x^{1/2} f_X(x)dx = \int_0^\infty x^{1/2}\frac{1}{\theta}e^{-\frac{x}{\theta}}dx \underset{x=\theta u}{=} \frac{1}{\theta}\int_0^\infty \theta^{1/2}u^{1/2}e^{-u}\theta du = \theta^{1/2}\underbrace{\int_0^\infty u^{1/2}e^{-u}du}_{\Gamma(\frac{3}{2})} = \Gamma(\frac{3}{2})\theta^{1/2} = (\frac{1}{2})\underbrace{\Gamma(\frac{1}{2})}_{\sqrt{\pi}}\theta^{1/2} =$$

$$= (\frac{1}{2})\sqrt{\pi\theta} \Rightarrow E(\tilde{\theta}) = \left(\left(\frac{1}{2}\right)\sqrt{\pi\theta}\right)^2 = \frac{\pi}{4}\theta$$

En cuanto a las varianzas: para abreviar, como $\hat{\theta} = \frac{X_1+X_2}{2}$ y la muestra es de tamaño $n = 2$, es claro que $\hat{\theta} = \bar{X}$, y de $\bar{X}$, en general sabemos que $E(\bar{X}) = \mu_X$ y $V(\bar{X}) = \frac{\sigma_X^2}{n}$. En este caso $X \sim Exp\left(x; \beta = \frac{1}{\theta}\right) \Rightarrow \mu_X = \theta;\ \sigma_X^2 = \theta^2$ y por tanto:

$V(\hat{\theta}) = V(\hat{\theta}) = V(\bar{X}) = \frac{\sigma_X^2}{n} = \frac{\theta^2}{n} = \frac{\theta^2}{2}$.

En el caso de $\tilde{\theta} = \sqrt{X_1X_2}$, usaremos la identidad $V(\tilde{\theta}) = E(\tilde{\theta}^2) - \left(E(\tilde{\theta})\right)^2$. Como $E(\tilde{\theta}) = \frac{\pi}{4}\theta$ ya es conocido, sólo falta $E(\tilde{\theta}^2)$ y $E(\tilde{\theta}^2) = E\left(\left(\sqrt{X_1X_2}\right)^2\right) = E(X_1X_2) \overset{\times indep}{\triangleq} E(X_1)E(X_2) = \theta^2$. El resto es rutina.

Problema 3

Si el capital inicial (en <u>cientos</u> de dólares) de una microempresa es una v.a. continua X con distribución normal $N(\mu, \sigma^2)$ y se toma una m.a. de tamaño n

a) Halle la función generatriz de momentos $M_{\bar{X}}(t)$ y verifique que corresponde a una distribución $N\left(\mu, \frac{\sigma^2}{n}\right)$ $\forall n$.

b) Si $\sigma^2 = 1$, $n = 9$, calcule la probabilidad de que la media $\overline{X}$ de la muestra difiera de la media poblacional μ en menos de 20 dólares

c) Si $\sigma^2 = 1$ ¿Qué tamaño de muestra garantiza que con 95% de probabilidad la diferencia $|\overline{X} - \mu|$ no pasará de 20 dólares?

d) Si $\sigma^2 = 1$ y μ sólo puede ser $\mu = 1$ o $\mu = 5$, si $n = 3$ y toda la muestra dio valores entre 2 y 6 cientos de dólares ¿Cuál valor de μ escogería? Use probabilidades para decidir.

Solución:

a) $M_{\bar{X}}(t) = E\left(e^{t\bar{X}}\right) = E\left(e^{t\frac{1}{n}\sum_{j=1}^{n} X_j}\right) = E\left(e^{\frac{t}{n}X_1} \times e^{\frac{t}{n}X_2} \times \ldots \times e^{\frac{t}{n}X_n}\right) \overset{(1)}{\cong} E\left(e^{\frac{t}{n}X_1}\right) \times E\left(e^{\frac{t}{n}X_2}\right) \times \ldots \times E\left(e^{\frac{t}{n}X_n}\right) \overset{(2)}{\cong}$ $E\left(e^{\frac{t}{n}X}\right) \times E\left(e^{\frac{t}{n}X}\right) \times \ldots \times E\left(e^{\frac{t}{n}X}\right) = M_X\left(\frac{t}{n}\right) \times M_X\left(\frac{t}{n}\right) \times \ldots \times M_X\left(\frac{t}{n}\right) = \left(M_X\left(\frac{t}{n}\right)\right)^n \overset{(3)}{\cong} \left(e^{\frac{t}{n}\mu + \frac{t^2}{2n^2}\sigma^2}\right)^n =$ $e^{t\mu + \frac{t^2}{2}\left(\frac{\sigma^2}{n}\right)}$ $\forall n \Rightarrow M_{\bar{X}}(t) = e^{t\mu + \frac{t^2}{2}\left(\frac{\sigma^2}{n}\right)}$ $\forall n$, que corresponde a la F.G.M. de una distribución $N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow$ $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \forall n$.

(1) $\{X_1, X_2, \ldots, X_n\}$ son independientes; (2) $\{X_1, X_2, \ldots, X_n\}$ tienen todas la misma distribución que X;

(3) $X \sim N(\mu, \sigma^2) \Leftrightarrow M_X(l) = e^{l\mu + \frac{l^2}{2}\sigma^2} \forall l$

b) Por a), como la muestra se toma de una distribución normal, se cumple $\bar{X} \sim N\left(\mu_X, \frac{\sigma_X^2}{n}\right)$ $\forall n$, o sea no es necesario tener tamaño de muestra grande: $\bar{X} \sim N\left(\mu, \frac{1}{9}\right)$ y se pregunta por $P(|\bar{X} - \mu| < 0.2)$, que se obtiene estandarizando y usando la tabla Z: $P(|\bar{X} - \mu| < 0.2) = P\left(\frac{|\bar{X}-\mu|}{\frac{1}{\sqrt{9}}} < \frac{0.2}{\frac{1}{\sqrt{9}}}\right) = P(|Z| < 0.6) = 0.7257 - 0.2743 = 0.4515$; que es una probabilidad muy baja (confiabilidad de 45.15%)

c) $0.95 = P(|\bar{X} - \mu| < 0.2) = P\left(\frac{|\bar{X}-\mu|}{\frac{1}{\sqrt{n}}} < \frac{0.2}{\frac{1}{\sqrt{n}}}\right) = P\left(|Z| < 0.2\sqrt{n}\right) \Rightarrow P\left(Z < 0.2\sqrt{n}\right) = 0.975 \Rightarrow 0.2\sqrt{n} =$ $1.96 \Rightarrow n \cong 91$

d) Se sabe que ocurrió $2 \le X_i \le 6 \ \forall i$, y como $P(2 \le X_i \le 6 \ \forall i) = \prod_{i=1}^{3} P(2 \le X_i \le 6) = \left(P(2 \le X \le 6)\right)^3$ $\left(F_X(6) - F_X(2)\right)^3 = \left(F_Z(6-\mu) - F_Z(2-\mu)\right)^3$. Reemplazando valores posibles de μ, $\mu = 1$ y $\mu = 5$ se obtienen las dos probabilidades: 0.0040 y 0.5927 respectivamente: Lo más probable es que $\mu = 5$, escogeríamos $\mu = 5$. (Recordar que $X \sim N(\mu, \sigma) \Rightarrow Z = \frac{(X-\mu)}{\sigma} \sim N(0,1)$ y en este caso como $\sigma = 1$, $Z = (X - \mu)$).

Problema 4

En un modelo sobre determinación de precios de comerciantes minoristas se propone que el comerciante agrega una cantidad X de soles al precio mínimo sugerido por el fabricante, con X v.a.c. con función de densidad:

$f_X(x) = \begin{cases} \theta x & 0 < x \le 1 \\ 1 - \frac{\theta}{2} & 1 < x \le 2 \end{cases}$ donde $\theta > 0$ es parámetro de valor desconocido por estimar o aproximar.

Dada $(X_1, X_2, \ldots, X_n)$ m.a. se propone como aproximación de θ a $\hat{\theta} = c\left(\frac{3}{2} - \bar{X}\right)$. Halle c de modo que $\hat{\theta}$ "tienda a coincidir" con θ, esto es, de modo que $E(\hat{\theta}) = \theta$.

Solución:

Queremos hallar la constante c tal que la estadística $\hat{\theta}$ definida por $\hat{\theta} = c\left(\frac{3}{2} - \bar{X}\right)$ tenga como valor esperado o de "tendencia en el equilibrio" al parámetro θ, esto es $E[\hat{\theta}] = \theta$. Veamos:

$E[\hat{\theta}] = E\left[c\left(\frac{3}{2} - \bar{X}\right)\right] = cE\left[\left(\frac{3}{2} - \bar{X}\right)\right] = c\left(\frac{3}{2} - E(\bar{X})\right)$ y por teoría del curso sabemos que $E(\bar{X}) = \mu_X$ y del capítulo 1 del curso, sabemos que

$$\mu_X = \int_0^2 x f_X(x)dx = \int_0^1 x\theta x\, dx + \int_1^2 x\left(1 - \frac{\theta}{2}\right)dx = \left[\theta\frac{x^3}{3}\right]_0^1 + \left(1 - \frac{\theta}{2}\right)\left[\frac{x^2}{2}\right]_1^2 = \frac{\theta}{3} + \left(1 - \frac{\theta}{2}\right)\frac{3}{2} = \frac{3}{2} - \frac{5}{12}\theta$$

$\Rightarrow c\left(\frac{3}{2} - E(\bar{X})\right) = c\left(\frac{3}{2} - \frac{3}{2} + \frac{5}{12}\theta\right) = c\frac{5}{12}\theta$

Reemplazando en $E[\hat{\theta}] = \theta$ tenemos la condición $c\frac{5}{12}\theta = \theta \Rightarrow c\frac{5}{12} = 1 \Rightarrow c = \frac{12}{5}$ es el valor pedido.

Problema 5

a) Si $X_1, X_2, \ldots, X_n$ es m.a. tomada de una distribución uniforme $U(0,\theta)$ y sea la estadística $W = \sum_{j=1}^{n} \alpha_j X_j$. Halle los valores de las constantes α_1,α_2,…, α_n tales que $E(W) = \theta$ y $V(W)$ sea mínima. (Aplique multiplicadores de Lagrange)

b) Si $(X_1, X_2, \ldots, X_n)$ es m.a. tomada de una distribución distribución exponencial $Exp(x, \beta = 1/\theta)$ y se define la variable aleatoria $Y = Mín\{X_j\}$, halle $P(Y > y)$ y la distribución acumulativa de Y $F_Y(y) = P(Y \leq y)$ y luego, derivando, halle la función de densidad de Y. Si se usara Y para aproximar o estimar el valor de θ, ¿Esperaría que su estimación coincidiera con el parámetro θ?

Solución:

a) $(X_1, X_2, \ldots, X_n)$ es m.a. de una distribución uniforme $U(0,\theta)$ y $W = \sum_{i=1}^{n} \alpha_j X_j$, donde hay que hallar $\{\alpha_j\}$ tales que $\theta = E(W) = E\left(\sum_{i=1}^{n} \alpha_j X_j\right) = \sum_{i=1}^{n} \alpha_j E(X_j) = \sum_{i=1}^{n} \alpha_j \frac{\theta}{2} \Leftrightarrow \sum_{i=1}^{n} \alpha_j = 2$ pues $X_j \sim U(0,\theta) \forall j$ y por tanto (ver formulario) $E(X_j) = \frac{\theta}{2}$. Es decir los $\{\alpha_j\}$ deben satisfacer la condición $\sum_{i=1}^{n} \alpha_j = 2$.

Pero además los $\{\alpha_j\}$ deben minimizar $V(W) = V\left(\sum_{i=1}^{n} \alpha_j X_j\right) = \sum_{i=1}^{n} \alpha_j^2 V(X_j) = \sum_{i=1}^{n} \alpha_j^2 \frac{\theta^2}{12} = \frac{\theta^2}{12} \sum_{i=1}^{n} \alpha_j^2$, porque habiendo independencia, la varianza de la suma es la suma de varianzas y $V(X_j) = \frac{\theta^2}{12}$.

En consecuencia los coeficientes $\{\alpha_j\}$ minimizan $\sum_{i=1}^{n} \alpha_j^2$ cumpliendo además la restricción $\sum_{i=1}^{n} \alpha_j = 2$ (omitimos $\frac{\theta^2}{12}$ porque es positiva y constante, así que minimizar $\frac{\theta^2}{12} \sum_{i=1}^{n} \alpha_j^2$ equivale a minimizar $\sum_{i=1}^{n} \alpha_j^2$).

Como se trata de una minimización con restricciones, podemos aplicar Multiplicadores de Lagrange:

$L(\alpha_1, \alpha_2, \ldots, \alpha_n, \lambda) = \sum_{i=1}^{n} \alpha_j^2 + \lambda\left(2 - \sum_{i=1}^{n} \alpha_j\right)$. Resolviendo se obtiene $\alpha_j = \frac{2}{n}$.

b) $P(Y > y) = P(Mín\{X_i\} > y) = P(\cap_{i=1}^{n} (X_i > y)) = \prod_{i=1}^{n} P(X_i > y) = \prod_{i=1}^{n} P(X > y) = \left(P(X > y)\right)^n =$

$$\left(\int_y^{\infty} \frac{1}{\theta} e^{-\frac{x}{\theta}} dx\right)^n = \left(\left[-e^{-\frac{x}{\theta}}\right]_y^{\infty}\right)^n = e^{-\frac{n}{\theta}y} \Rightarrow F_Y(y) = 1 - e^{-\frac{n}{\theta}y} \Rightarrow f_Y(y) = \left(\frac{n}{\theta}\right) e^{-\left(\frac{n}{\theta}\right)y} \Rightarrow Y \sim Exp\left(y; \left(\frac{n}{\theta}\right)\right)$$

$$\Rightarrow E(Y) = \frac{\theta}{n}$$

Si usáramos Y como aproximación de θ, sistemáticamente estaríamos obteniendo valores estimados más bajos que el valor verdadero de θ, estaríamos subestimando el parámetro.

Problema 6

a) Si (X_1, X_2, X_3) es muestra aleatoria de una distribución $U(0,1)$, halle $P([Mín\{X_j\} > 0.25] \cap [Máx\{X_j\} < 0.75])$

b) Si $(X_1, X_2, \ldots, X_n)$ es muestra aleatoria de una distribución $U(0,1)$, halle la media y la varianza de la estadística $P = \prod_{j=1}^{n} X_j$.

Solución:

a) $P\left(\left(Mín\{X_j\} > 0.25\right) \cap \left(Máx\{X_j\} < 0.75\right)\right) = P\left(0.25 < X_j < 0.75\ \forall j\right) =$

$P\left((0.25 < X_1 < 0.75) \cap (0.25 < X_2 < 0.75) \cap (0.25 < X_3 < 0.75)\right) =$

$P(0.25 < X_1 < 0.75) \times P(0.25 < X_2 < 0.75) \times P(0.25 < X_3 < 0.75) =$

$\left(P(0.25 < X < 0.75)\right)^3 = 0.5^3 = 0.125$ pues $P(0.25 < X < 0.75) = \int_{0.25}^{0.75} 1 dx = 0.5$

b) $E(P) = E\left(\prod_{j=1}^{n} X_j\right) = E(X_1) \times E(X_2) \times \ldots \times E(X_n) = \left(E(X)\right)^n = \left(\frac{1}{2}\right)^n$ (debido a la independencia entre componentes de toda muestra aleatoria).

$V(P) = E(P^2) - \left(E(P)\right)^2 = E(P^2) - \left(\left(\frac{1}{2}\right)^n\right)^2 = E(P^2) - \left(\frac{1}{2}\right)^{2n}$ y como

$E(P^2) = E[(X_1 \times X_2 \times \ldots \times X_n)^2] = E[(X_1^2 \times X_2^2 \times \ldots \times X_n^2)] = E(X_1^2)E(X_2^2) \ldots . E(X_n^2) = \left(E(X^2)\right)^n$ y

$E(X^2) = \sigma_X^2 + \mu_X^2 = \frac{1}{12} + \frac{1}{4} = \frac{1}{3}$, entonces $E(P^2) = \left(\frac{1}{3}\right)^n$ y $V(P) = \left(\frac{1}{3}\right)^n - \left(\frac{1}{2}\right)^{2n} = \left(\frac{1}{3}\right)^n - \left(\frac{1}{4}\right)^n$

Problema 7

Sea $(X_1, X_2, \ldots, X_k)$ v.a. de componentes independientes:

a) Si $X_j \sim b(n = 1, p)$ $\quad j = 1,2, \ldots, k$; halle la función generatriz de momentos de la suma $T = \sum_{j=1}^{k} X_j$ y use este resultado verificar que la distribución de T es una Binomial

b) Si $X_j \sim N(0,1) \quad j = 1,2,\ldots,k$; halle la función generatriz de momentos de la suma $T = \sum_{j=1}^{k} X_j$ y use este resultado verificar que la distribución de T es Normal

Solución:
Tanto a) como b) se basan el mismo hecho: las $\{X_i\}$ son independientes y con idéntica distribución (son *i.i.d*), por tanto el esperado de productos es producto de valores esperados y todas tienen la misma función generatriz de momentos. Asimilado esto:

$M_T(t) = E(e^{tT}) = E\left(e^{t\sum_{i=1}^{k} X_i}\right) = E\left(e^{\sum_{i=1}^{k} tX_i}\right) = E(e^{tX_1} \times e^{tX_2} \times \ldots \times e^{tX_k}) =$
$= E(e^{tX_1}) \times E(e^{tX_2}) \times \ldots \times E(e^{tX_k}) = M_{X_1}(t) \times M_{X_2}(t) \times \ldots \times M_{X_k}(t) = \left(M_X(t)\right)^k$ y lo que va cambiando con X es la respectiva función generatriz $M_X(t)$, así:

En a) $X \sim B(n = 1, p) \Leftrightarrow M_X(t) = (pe^t + q)^1 \Rightarrow M_T(t) = (pe^t + q)^k$ que corresponde a una distribución binomial $B(n = k, p)$
y en el caso de b) $M_T(t) = (e^{\frac{t^2}{2}})^k = e^{\frac{t^2}{2}k}$ que corresponde a una $N(0, k)$

Problema 8
Asuma que el gasto mensual X en cabinas de Internet, es una v.a. con distribución $\Gamma(x; \alpha = 1, \beta)$ y se piensa tomar una m.a. **grande** de **n** usuarios de cabinas para aproximar el valor de β, usando para ello la media muestral $\overline{X}$. Aplique el "Teorema del Límite Central para $\overline{X}$" y resuelva las siguientes preguntas:

a) ¿Qué tamaño de muestra garantizaría que con 95% de probabilidad, la media $\overline{X}$ de la muestra diferirá de β en menos de un 20% de β?.
b) En a), si sólo hay dinero para tomar una muestra de $n = 40$ usuarios, calcule la probabilidad de lograr lo deseado. ¿Cuál es la máxima diferencia $|\overline{X} - \mu|$ (como % de β) que se puede garantizar con 95% de probabilidad?

Solución:

El contexto general aplicable es que *n* grande ($n > 30$) implica que $\overline{X} \sim N(\mu_X, \frac{\sigma_X^2}{n})$ y lo que va cambiando de problema en problema es X y μ_X con σ_X^2. En este problema $X \sim \Gamma(x; \alpha = 1, \beta) \Rightarrow \mu_X = \beta, \sigma_X^2 = \beta^2$ y por tanto $\overline{X} \sim N(\mu_X = \beta, \frac{\sigma_X^2}{n} = \frac{\beta^2}{n})$ o equivalentemente $Z = \frac{\overline{X} - \beta}{\beta / \sqrt{n}} \sim N(0,1)$.

a) "$\overline{X}$ de la muestra diferirá de β menos de un 20% de β" equivale a $|\overline{X} - \beta| < 0.2\beta$ y se tiene

$0.95 = P(|\overline{X} - \beta| < 0.2\beta) = P(\frac{|\overline{X} - \beta|}{\beta / \sqrt{n}} < \frac{0.2\beta}{\beta / \sqrt{n}}) = P(|Z| < 0.2\sqrt{n}) = P(-0.2\sqrt{n} < Z < 0.2\sqrt{n})$. No podemos entrar a la tabla Z con 0.95 de probabilidad porque esta tabla no tiene áreas centrales sino acumuladas, pero dada la simetría de la distribución de Z (graficar las áreas ayudará a entender la probabilidad 0.975)

$0.95 = P(-0.2\sqrt{n} < Z < 0.2\sqrt{n}) \Leftrightarrow P(Z < 0.2\sqrt{n}) = 0.975 \Rightarrow 0.2\sqrt{n} = 1.96 \Rightarrow n = 9.8^2 \cong 97$.

b) Se pregunta ahora, si $n = 40$ hallar r tal que $0.95 = P(|\overline{X} - \beta| < r\beta)$, que se resuelve con

$0.95 = P(|\overline{X} - \beta| < r\beta) = P(\frac{|\overline{X} - \beta|}{\beta / \sqrt{40}} < \frac{r\beta}{\beta / \sqrt{40}}) = P(|Z| < r\sqrt{40}) = P(-r\sqrt{40} < Z < r\sqrt{40}) \Rightarrow$

$0.975 = P(Z < r\sqrt{40}) \Rightarrow r\sqrt{40} = 1.96 \Rightarrow r = 1.96 / \sqrt{40} = 0.31 \, o \, 31\% \, de \ \beta$

Problema 9
Sea X v.a. continua con función de densidad $f_X(x) = \frac{2x}{\theta^2} \, 0 < x \leq \theta$ y sea $(X_1, X_2, \ldots, X_n)$ muestra aleatoria de esta distribución.

a) Halle la media y la varianza de la estadística $\overline{X}$, como funciones explícitas de θ
b) Se desea usar la media $\overline{X}$ de la muestra para aproximar μ, usando una muestra grande de $n = 36$ casos. ¿Con qué probabilidad ocurrirá que el error de aproximación $|\overline{X} - \mu|$ no pasará del 5% del valor real de μ?

c) Sea la estadística $Y = Máx\{X_1, X_2, \dots, X_n\}$. Halle la función de densidad $f_Y(y)$ de Y y el valor esperado de esta estadística
Sugerencia: Calcule primero la distribución acumulativa $F_Y(y)$ y luego, derivando, obtenga $f_Y(y)$.

Sugerencias:

a) $E(\bar{X}) = E(X) = \mu_X$ y $V(\bar{X}) = \frac{\sigma_X^2}{n}$. Sólo hay que hallar $E(X) = \int_0^\theta x\frac{2x}{\theta^2}dx = \frac{2}{\theta^2}\int_0^\theta x^2\,dx = \frac{2}{\theta^2}\left[\frac{x^3}{3}\right]_0^\theta = \frac{2}{3}\theta$,
$E(X^2) = \int_0^\theta x^2\frac{2x}{\theta^2}dx = \frac{2}{\theta^2}\int_0^\theta x^3\,dx = \frac{2}{\theta^2}\left[\frac{x^4}{4}\right]_0^\theta = \frac{\theta^2}{2} = \frac{1}{2}\theta^2$; $V(X) = \sigma_X^2 = E(X^2) - [E(X)]^2 = \frac{1}{2}\theta^2 - \frac{4}{9}\theta^2 = \frac{\theta^2}{18}$.

b) Similar al problema 8b): "el error de aproximación $|\bar{X} - \mu|$ no pasará del 5% del valor real de μ" equivale a $|\bar{X} - \mu| \leq 0.05\mu$ y de $n = 36 > 30$ sabemos que $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{36}\right) \Rightarrow Z_{\bar{X}} = \frac{(\bar{X}-\mu)}{\frac{\sigma}{6}} \sim N(0,1)$ y en este caso $\mu = \frac{2}{3}\theta$ y $\sigma = \sqrt{\frac{\theta^2}{18}} = \frac{\theta}{3\sqrt{2}}$ de modo que $|\bar{X} - \mu| \leq 0.05\mu \Leftrightarrow |\bar{X} - \mu| \leq 0.05\frac{2}{3}\theta$ y el resto es rutina.

c) $F_Y(y) = P(Máx\{X_i\} \leq y) = P(\bigcap_{i=1}^n (X_i \leq y)) = \prod_{i=1}^n P(X_i \leq y) = (F_X(y))^n = \left(\int_0^y \frac{2x}{\theta^2}\right)^n = \left(\frac{1}{\theta^2}\int_0^y 2x\right)^n = \left(\frac{y^2}{\theta^2}\right)^n = \frac{y^{2n}}{\theta^{2n}} \Rightarrow F_Y(y) = \frac{y^{2n}}{\theta^{2n}}$ $\quad 0 < y \leq \theta$; derivando se tiene $f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{d}{dy}\frac{y^{2n}}{\theta^{2n}} = \frac{2ny^{2n-1}}{\theta^{2n}}$ $\quad 0 < y \leq \theta$, luego se procede a calcular $E(Y) = \int_0^\theta y\frac{2ny^{2n-1}}{\theta^{2n}}dy = \frac{2n}{\theta^{2n}}\int_0^\theta y\frac{y^{2n-1}}{\theta^{2n}}dy = \frac{2n}{\theta^{2n}}\left[\frac{y^{2n+1}}{2n+1}\right]_0^\theta = \left(\frac{2n}{2n+1}\right)\theta$.

Problema 10
Un economista asume que los salarios, en dólares, de los trabajadores que recién se integran a una corporación, siguen una distribución normal $N(\mu, \sigma^2 = 70^2)$. Para estudiar estos salarios, selecciona al azar una muestra de n de este tipo de trabajadores.

a) Si se quiere aproximar o estimar el salario medio μ usando la estadística $\bar{X}$ de modo que con 95% de probabilidad o confiabilidad, el error de estimación o aproximación $|\bar{X} - \mu|$ no pase de 20 dólares. ¿Qué tamaño de muestra debiera usar?
b) Si $n = 4$ y μ fuera 800 ¿qué tan probable es que el salario del que gane menos en la muestra supere los 700 dólares?
c) Un colega del economista critica la "especificación del modelo de datos" rechazando el supuesto de normalidad y afirma que la distribución de salarios es una distribución exponencial con una media de 800 dólares. Si fuera correcta la crítica, ¿con qué probabilidad, aproximadamente, una muestra de $n = 64$ trabajadores arrojará una media de salarios que supere los 900 dólares?

Soluciones:
a)
Dada la distribución normal de los salarios entonces

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \rightarrow \overline{X} \sim N\left(\mu, \frac{70^2}{n}\right)$$

Luego

$$P\left(|\overline{X} - \mu| \leq 20\right) = 95\,\%$$

Estandarizando

$$P\left(\frac{|\overline{X} - \mu|}{\sqrt{\frac{\sigma^2}{n}}} \leq \frac{20}{\frac{70}{\sqrt{n}}}\right) = 95\,\%$$

$$P\left(|Z| \leq \frac{2}{7}\sqrt{n}\right) = 95\,\% \rightarrow \underbrace{P\left(Z \leq \frac{2}{7}\sqrt{n}\right)}_{97.5\,\%} - \underbrace{P\left(Z \leq -\frac{2}{7}\sqrt{n}\right)}_{2.5\,\%} = 95\,\% \rightarrow \frac{2}{7}\sqrt{n} = 1{,}96$$

Por lo tanto

$$n = 47{,}0596 \rightarrow n \approx 48$$

b)

Dado que hay 4 muestras, y nos piden el salario del que gane menos tendríamos supere los 700 dólares

$$P\left(\min\left\{X_i\right\} > 700\right)\ \ \forall i = 1,2,3,4$$

$$\begin{aligned}
&= P\left(X_1 > 700\right) \cap P\left(X_2 > 700\right) \cap P\left(X_3 > 700\right) \cap P\left(X_4 > 700\right) \\
&= P\left(\bigcap_{i=1}^{4} P\left(X_i > 700\right)\right) = \prod_{i=1}^{4} P\left(X_i > 700\right) = \left[P\left(X > 700\right)\right]^4 \\
&= \left[1 - P\left(X \leq 700\right)\right]^4 = \left[1 - P\left(\frac{X-\mu}{\sigma} \leq \frac{700-800}{70}\right)\right]^4 \\
&= \left[1 - P\left(Z \leq -1{,}43\right)\right]^4 = \left[1 - 0{,}0764\right]^4 = 0{,}7276
\end{aligned}$$

c)

Dado que $n = 64 > 30$ ("n" grande), aplicamos el teorema del límite central, luego

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \rightarrow \mu = \frac{1}{\beta} = 800, \sigma^2 = \frac{1}{\beta^2} = 800^2 \rightarrow \overline{X} \sim N\left(800, \frac{800^2}{n}\right)$$

$$\begin{aligned}
P\left(\overline{X} > 900\right) &= 1 - P\left(\overline{X} < 900\right) = 1 - P\left(\frac{\overline{X}-\mu}{\sigma} < \frac{900-800}{800/\sqrt{n}}\right) \\
&= 1 - P\left(Z < \frac{1}{8}n\right) = 1 - P\left(Z < 1\right) = 0{,}1587
\end{aligned}$$

$$P\left(\overline{X} > 900\right) = 0{,}1587$$

ACG./SAMP.