Tarea 5 - Análisis de datos
Giovanni Gamaliel López Padilla

Problema 3

La vida útil de un cierto tablet sigue una distribución normal con promedio 3 años y σ^2=0.9

a) **Calcula la probabilidad que funcionará más de cuatro años.**

Se tiene que $\mu = 3$ y $\sigma^2 = 0.9$. Entonces, al tratarse de una distribución normal, la probabilidad que la tablet funcione más de cuatro años es:

$$\begin{aligned} P(X > 4) &= 1 - P(x \leq 4) \\ &= 1 - \int_{-\infty}^{4} \mathcal{N}(3, 0.9)dx \\ &= 1 - 0.8540797 \\ P(X > 4) &= 0.1459203 \end{aligned}$$

lo anterior fue calculado con la función pnorm de R.

b) **Se quiere determinar la duración de la garantía. ¿Para cuántos meses máximos se debe dar la garantía para que la probabilidad que el tablet se descomponga antes del fin de la garantía sea no mayor que 0.25? La respuesta debe ser en términos de meses enteros.**

Se tiene que $\mu = 3$ y $\sigma^2 = 0.9$. Entonces, al tratarse de una distribución normal, lo que se quiere encontrar es el valor de x tal que:

$$\begin{aligned} P(X) &< 0.25 \\ \int_{-\infty}^{x} \mathcal{N}(3, 0.9)dx &= 0.25 \\ x &= 2.360123 \end{aligned}$$

Esto fue calculado con la función qnorm de R. Entonces se tiene que en 2.360123 años sería el límite para la gararantía. Levando esta cantidad al el entero menor en meses, el máximo de meses es 28 meses.

Problema 4

Se redondean 50 números reales al número entero más cercano. Suponiendo que los errores de redondeo tienen una distribución uniforme sobre (-0.5, 0.5), calcula

una aproximación para la probabilidad que la suma con los valores redondeados tiene un error mayor que 3 a la suma con los valores exactos.

Sea $X_1, X_2, \ldots, X_n$ variables aleatorias independientes y $X_i \sim \mathcal{U}(-0.5, 0.5)$, entonces se tiene que $\mu = 0$ y $\sigma^2 = \frac{1}{12}$. Por ser μ y σ finitos, se obtiene que:

$$\lim_{n \to 0} P\left(\frac{\frac{S_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq y\right) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} exp\left(-\frac{x^2}{2}\right) dx \tag{1}$$

donde $S_n = X_1 + X_2 + \cdots + X_n$. De la ecuación 1, se obtiene que para n suficientemente grande:

$$S_n \sim \mathcal{N}(n\mu, n\sigma^2)$$

Por lo que, se puede obtener una aproximación para cualquier n. Entonces, tomando los datos se obtiene que:

$$\begin{aligned} P(S_n > 3) &= 1 - P(S_n \leq 3) \\ P(S_n > 3) &\approx 1 - 0.9291777 \\ P(S_n > 3) &\approx 0.07082235 \end{aligned}$$

El valor de $P(S_n \leq 3)$ fue calculado usando la función pnorm de R.

Problema 5

El tiempo de ejecucion de algoritmo A es $\mathcal{N}(2, 1)$ y de algoritmo B es $\mathcal{N}(4, 2)$. Si los tiempos son independientes entre sí, calcula la probabilidad que en una corrida B es más rapido que A.

La probabilidad que el algoritmo B obtenga un tiempo menor que A es definida como $P(B < A)$. Sea C el conjunto de valores que toma la variable aleatoria X tal que:

$$C = \{x \in X : \mathcal{N}(x, 4, 2) < \mathcal{N}(x, 2, 1)\}$$

Encontrando el conjunto de C, se tiene lo siguiente:

$$\begin{aligned} \frac{1}{\sigma_1\sqrt{2\pi}} exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right) &= \frac{1}{\sigma_2\sqrt{2\pi}} exp\left(-\frac{(x-\mu_2)^2}{2\sigma_2^2}\right) \\ -ln\left(\sigma_1\sqrt{2\pi}\right) - \frac{(x-\mu_1)^2}{2\sigma_1^2} &= -ln\left(\sigma_2\sqrt{2\pi}\right) - \frac{(x-\mu_2)^2}{2\sigma_2^2} \\ \frac{(x-\mu_2)^2}{2\sigma_2^2} - \frac{(x-\mu_1)^2}{2\sigma_1^2} &= ln\left(\sigma_1\sqrt{2\pi}\right) - ln\left(\sigma_2\sqrt{2\pi}\right) \\ \sigma_1^2(x-\mu_2)^2 - \sigma_2^2(x-\mu_1)^2 &= 2\sigma_1^2\sigma_2^2 ln\left(\frac{\sigma_1}{\sigma_2}\right) \end{aligned}$$

$$(\sigma_1^2 - \sigma_2^2)x^2 + 2(\mu_1\sigma_2^2 - \mu_2\sigma_1^2)x + \sigma_1^2\mu_2^2 - \sigma_2^2\mu_1^2 - 2\sigma_1^2\sigma_2^2 ln\left(\frac{\sigma_1}{\sigma_2}\right) = 0 \tag{2}$$

donde la ecuación 2 es una ecuación cuadratica. Tomando en cuenta los parametros $\mu_1 = 2$, $\sigma_1 = 1$, $\mu_2 = 4$ y $\sigma_2 = 2$, las soluciones son $x = \{-3.063706, 3.063706\}$. Como las distribuciones representan tiempos, entonces las solución negativa será despreciada. Por lo que la el conjunto C puede ser escrito como:

$$C = \{x \in X : 0 < x < 3.063706\}$$

Entonces

$$\begin{aligned} P(x \in C) &= P(0 < x < 3.063706) \\ P(x \in C) &= P(x < 3.063706) - P(x < 0) \end{aligned} \tag{3}$$

La probabilidad señalada en la ecuación 3 es la señalada en la figura 1.

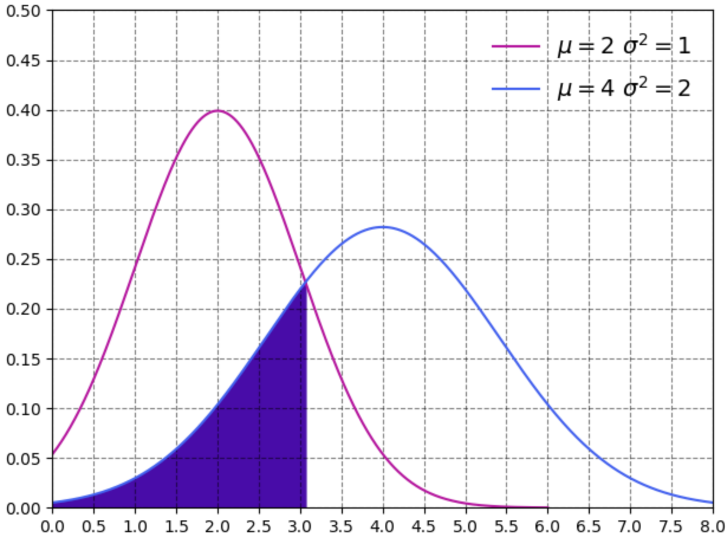


Figura 1: Área a integrar en la ecuación 2.

Por lo tanto

$$\begin{aligned} P(x \in C) &= P(x < 3.063706) - P(x < 0) \\ &= 0.2539664 - 0.002338867 \\ P(x \in C) &= 0.2516276 \end{aligned}$$

Problema 6

Alguine tiene una reserva de 100 focos. Se sabe que el tiempo de vida de un foco sigue una distribución exponencial con promedio 5 horas. Se prende un foco a la vez; si se descompone, se cambia el foco inmediatamente hasta agotar la reserva. Si los tiempos de vida son v.a. independientes entre sí calcula una aproximación para la probabilidad que aun hay un foco prendido despues de 525 horas.

Sea $X_1, X_2, \ldots, X_n$ variables aleatorias independientes y $X_i \sim Exp(\beta = 5)$, entonces se tiene que $\mu = 5$ y $\sigma^2 = 25$. Por ser μ y σ finitos, se puede usar el teorema del límite central (ecuación 1). Entonces

$$S_n \sim \mathcal{N}(n\mu, n\sigma^2)$$

Por lo tanto:

$$\begin{aligned}
P(S_n > 525) &= 1 - P(S_n \leq 525) \\
P(S_n > 525) &\approx 1 - 0.6914625 \\
P(S_n > 525) &\approx 0.3085375
\end{aligned}$$

El valor de $P(S_n \leq 525)$ fue calculado usando la función pnorm de R.

Problema 7

Se construyen dos vectores de dimensión 100, X y Y con $X_i, Y_i \sim \mathcal{U}\left(-\frac{1}{10}, \frac{1}{10}\right)$ y todos independientes entre sí. Calcula una aproximación para la probabilidad de que $\langle X, Y \rangle$ es mayor que 0.005.

Sea Z una variable aleatoria tal que: Z=XY, entonces calculando EZ y $Var(Z)$, se obtiene lo siguiente:

Para EZ:

$$EZ = EXY$$

como X y Y son variables aleatorias independientes, entonces:

$$EZ = EXEY$$

donde $EX = EY = 0$, esto es porque $X, Y \sim \mathcal{U}\left(-\frac{1}{10}, \frac{1}{10}\right)$, por lo tanto $EZ = 0$.

Para $Var(Z)$, se tiene que:

$$\begin{aligned}
Var(Z) &= Var(XY) \\
&= E(XY)^2 - (EXY)^2 \\
&= EX^2Y^2 - (EXEY)^2 \\
&= EX^2EY^2
\end{aligned}$$

calculando EX^2, se obtiene lo siguiente:

$$
\begin{aligned}
EX^2 &= \int_{-\frac{1}{10}}^{\frac{1}{10}} x^2 \left(\frac{1}{\frac{1}{10} + \frac{1}{10}} \right) dx \\
&= \int_{-\frac{1}{10}}^{\frac{1}{10}} 5x^2 dx \\
EX^2 &= \frac{1}{300}
\end{aligned}
$$

Entonces, como X, Y tienen la misma distribución $EX^2 = EY^2 = \frac{1}{300}$, por ende:

$$Var(Z) = \frac{1}{9000}$$

Por lo tanto, para la variable aleatoria Z, se tiene que $\mu = 0$ y $\sigma^2 = \frac{1}{90000}$, al ser finitas se puede usar el teorema del límite central (ecuación 1). Entonces:

$$
\begin{aligned}
P(S_n > 0.005) &= 1 - P(S_n \leq 0.005) \\
P(S_n > 0.005) &\approx 1 - 0.5596177 \\
P(S_n > 0.005) &\approx 0.4403823
\end{aligned}
$$

El valor de $P(S_n \leq 0.005)$ fue calculado usando la función pnorm de R.

Problema 8

Usando el hecho que para cualquier a, b, c: máx(a, b) $\leq$ c es equivalente a $a \leq c$ y $b \leq c$, calcula (a) la función de distribución de máx(X, Y) si X y Y son v.a. independientes uniformes sobre [-1, 1] y (b) E(máx(X, Y)).

a) Calculando F(max(X,Y))

Sea C una variable aleatoria tal que:

$$C = \{C \in c : max(X, Y) \leq c\}$$

donde X,Y son variables aleatorias independientes. Entonces, se tiene que la distribución de C es:

$$
\begin{aligned}
F(C < c) &= P(max(X, Y) < c) \\
&= P(x < c, y < c)
\end{aligned}
$$

como x,y son independientes entonces:

$$\begin{aligned} F(C<c) &= P(x<c, y<c) \\ &= P(x<c)P(y<c) \end{aligned}$$

Calculando $P(x<c)$

$$P(x<c) = \int_{-1}^{c} \frac{1}{2} dx$$
$$P(x<c) = \frac{1}{2}(c+1)$$

Como $P(x<c) = P(y<c)$, ya que X y Y tienen la misma distribución, entonces:

$$\begin{aligned} F(C<c) &= P(x<c)P(y<c) \\ &= (P(x<c))^2 \\ &= \left(\frac{1}{2}(c+1)\right)^2 \\ F(C<c) &= \frac{1}{4}(c+1)^2 \end{aligned}$$

b) Calculando E(max(X,Y))
Se tiene que

$$F(C<c) = \frac{1}{4}(c+1)^2$$

entonces, calculando su densidas de distribución de tiene que:

$$f(c) = \frac{1}{2}(c+1)$$

Por lo que calculando la esperanza, se obtiene que:

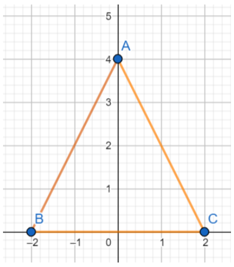
$$EC = \int_{-1}^{1} \frac{1}{2} c(c+1) dc \tag{4}$$
$$= \int_{-1}^{1} \frac{1}{2}(c^2+c) dc \tag{5}$$
$$= \frac{1}{6}c^3 + \frac{1}{4}c^2 \Big|_{-1}^{1} \tag{6}$$
$$EC = \frac{1}{3} \tag{7}$$

por lo tanto:

$$E(max(X,Y)) = \frac{1}{3}$$

Problema 9

Se elige al azar un punto (X, Y) adentro del siguiente triángulo. Calcula Cov(X, Y).



Se tiene que la densidad de probabilidad conjunta es:

$$f(x,y) = \begin{cases} \frac{1}{8} & -2 \leq x \leq 2, 0 \leq y \leq -2|x| + 4 \\ 0 & c.o.p \end{cases}$$

Se tiene que

$$Cov(X,Y) = EXY - EXEY$$

Calculando EXY

$$\begin{aligned} EXY &= \int_{-2}^{2} \int_{0}^{-2|x|+4} \frac{xy}{8} dydx \\ &= \int_{-2}^{2} \frac{x}{8} \left(\frac{1}{2}(-2|x|+4)^2 \right) \\ &= \int_{-2}^{0} \frac{x}{8} \left(\frac{1}{2}(2x+4)^2 \right) + \int_{0}^{2} \frac{x}{8} \left(\frac{1}{2}(-2x+4)^2 \right) \\ &= -\frac{1}{3} + \frac{1}{3} \\ EXY &= 0 \end{aligned}$$

Calculando EX

$$
\begin{aligned}
EX &= \int_{-2}^{2} \int_{0}^{-2|x|+4} \frac{x}{8} dy dx \\
&= \int_{-2}^{2} \frac{x}{8}(-2|x|+4)dx \\
&= \int_{-2}^{0} \frac{x}{8}(2x+4)dx + \int_{0}^{2} \frac{x}{8}(-2x+4)dx \\
&= -\frac{1}{3} + \frac{1}{3} \\
EX &= 0
\end{aligned}
$$

como $EX = 0$, ya no es necesaio calcular EY, ya que $0(EY) = 0$, por lo tanto:

$$Cov(X, Y) = 0$$