Bioestadística | **Universidad de la República**
Curso 2022 | **Facultad de Ciencias**

Práctico 4: Función de Distribución y Variables Aleatorias Continuas

Función de distribución de una variable aleatoria

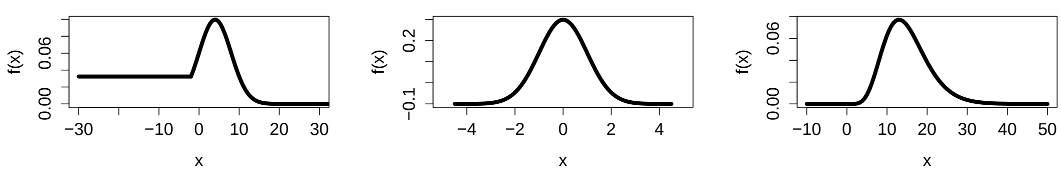
1. Sea X una variable aleatoria discreta tal que $Rec(X) = \{0, 1, 2, 4\}$ y la función de probabilidad puntual es $P(0) = \frac{1}{4}$, $P(1) = \frac{1}{4}$, $P(2) = \frac{1}{8}$ y $P(4) = \frac{3}{8}$. Establecer y graficar la función de distribución de la variable aleatoria X.

2. Sea F_X la función de distribución de una variable aleatoria X y los conjuntos $A = [-3, 1]$ y $B = (-2, 2)$.

$$F_X(x) = \begin{cases} 0 & \text{si} \quad x < -3 \\ 1/4 & \text{si} \quad -3 \leq x < 1 \\ 3/4 & \text{si} \quad 1 \leq x < 2 \\ 1 & \text{si} \quad x \geq 2 \end{cases}$$

   Realizar una gráfica de F_X. Establecer la función de probabilidad puntual de X y calcular $P(A)$, $P(B)$ y $P(B|A)$.

3. Indicar cuáles de las siguientes figuras podrían representar una función de densidad. Justificar.



Distribución uniforme

4. En pruebas de medición de distancia de frenado de automóviles, los vehículos que viajan a determinada velocidad tienden a recorrer distancias de frenado que están distribuidas uniformemente entre dos puntos a y b. Calcular la probabilidad de que uno de estos automóviles:
   (a) se detenga más cerca de a que de b.
   (b) se detenga de tal modo que la distancia a a sea por lo menos 3 veces mayor que la distancia a b.

5. Suponga que la concentración de cierto contaminante se encuentra distribuida uniformemente en el intervalo de 4 a 20 *ppm* (partes por millón). Si se considera tóxica una concentración de 15 *ppm* o más, ¿cuál es la probabilidad de que al tomar una muestra la concentración sea tóxica?

Distribución normal

Decimos que una variable aleatoria X tiene distribución normal de parámetros μ y $\sigma > 0$ ($X \sim N(\mu, \sigma)$) si su densidad es:

$$f_X(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}.$$

Cuando $\mu = 0$ y $\sigma = 1$, se dice que tiene distribución normal estándar y a su función de distribución se la escribe como $\Phi(x)$. Notar que su función de distribución es:

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} \mathrm{d}t.$$

El problema es que la función $e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}$ no tiene primitiva elemental (no existe una función $g(t)$ tal que $g'(t) = e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}$), por lo cual no podemos calcular $F_X(x)$ utilizando las herramientas que aprendimos en los cursos anteriores para calcular integrales. Existen algunos algoritmos computacionales que permiten aproximar el valor de $F_X(x)$ para cada valor de x.

Por suerte para nosotros, **R** *ya trae incorporado alguno de esos algoritmos, lo que nos permite calcular facilmente la función de distribución de una normal a partir de la función:*

$$\textbf{pnorm}(x, \text{ mean}= \mu, \text{ sd}= \sigma)$$

Es decir, en **R**:

- $F_X(x) = \textbf{pnorm}(x, \text{ mean}= \mu, \text{ sd}= \sigma)$
- *Cálculo de cuantiles:* si queremos averiguar cuál es el valor de x para el cual $F_X(x) = p$ para algún valor de $p \in (0, 1)$ dado, lo calculamos con la función $\textbf{qnorm}(p, \text{ mean}= \mu, \text{ sd}= \sigma)$

6. Sea Z una variable aleatoria de distribución normal estándar ($Z \sim \mathcal{N}(0, 1)$).
   (a) Calcular usando **R** el valor de $P\{Z \geq 1\}$
   (b) Deducir el valor de $P\{0 \leq Z \leq 1\}$ usando que la normal estándar es simétrica
   (c) Usando la primera parte y la simetría, deducir $P\{Z \geq -1\}$.
   (d) Usando las partes anteriores deducir $P\{-1 \leq Z \leq 1\}$;
   (e) Hallar el valor C que cumple:

   i) $P\{Z \geq C\} = 0.25$; ii) $P\{Z \leq C\} = 0.0287$; iii) $P\{-C \leq Z \leq C\} = 0.95$.

7. Sea Y una variable con distribución normal de parámetros μ y σ ($Y \sim \mathcal{N}(\mu, \sigma)$).
   (a) Si sabemos que $\mu = 50$ y $\sigma = 10$. Calcular (usando **R**) las siguientes probabilidades:
   a) $P\{Y \leq 65\}$; b) $P\{Y \leq 25\}$; c) $P\{42 \leq Y \leq 62\}$; d) $P\{38 \leq Y \leq 47\}$.
   (b) Si sabemos que $\sigma = 5$ y $P\{Y \geq 78.5\} = 0.6681$. Calcular el parámetro μ.
   Sugerencia: utilizar que $F_Y(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$ siendo Φ la función de distribución de una normal estándar.
   (c) Si sabemos que $\mu = 19$ y $P\{Y \leq 22.5\} = 0.63683$. Calcular el parámetro σ.

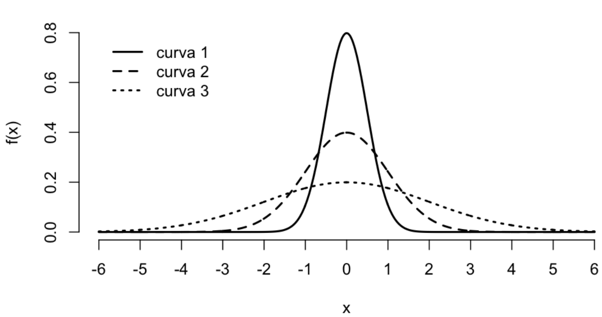
8. (a) Sea $Z \sim N(0, 1)$. Verificar que:

$$P(Z \in [-1, 1]) \simeq 0.68, \quad P(Z \in [-2, 2]) \simeq 0.95 \quad \text{y} \quad P(Z \in [-3, 3]) \simeq 0.997.$$

   (b) Sea $X \sim \mathcal{N}(\mu, \sigma)$. Hallar las siguientes probabilidades:

$$P(X \in [\mu - \sigma, \mu + \sigma]), \quad P(X \in [\mu - 2\sigma, \mu + 2\sigma]) \quad \text{y} \quad P(X \in [\mu - 3\sigma, \mu + 3\sigma]).$$

   (c) La siguiente figura muestra la densidad de tres variables aleatorias normales centradas. Indicar aproximadamente cuánto vale el desvío estándar σ en cada caso (*regla del tres-σ*).



Distribución exponencial

9. Se ha comprobado que el tiempo de vida de cierto tipo de marcapasos sigue una distribución exponencial con media de 16 años. [1]

   (a) Calcular la probabilidad de que a una persona a la que se le ha implantado este marcapasos se le deba reimplantar otro antes de que transcurran 20 años.

   (b) Si el marcapasos lleva funcionando correctamente 5 años en un paciente, calcular la probabilidad de que haya que cambiarlo antes de que transcurran 25 años.

   (c) Interpretar ambos resultados.

10. El tiempo de reparación de un instrumento tiene una distribución exponencial, con media de 22 minutos.[2]

    (a) Hallar la probabilidad de que el tiempo de reparación sea menor a diez minutos.

    (b) El costo de reparación es de 2000 pesos por cada media hora o fracción. Calcular la probabilidad de que una reparación cueste 4000 pesos.

    (c) Calcular el tiempo de reparación (T_r) que se debe asignar a cada reparación para que la probabilidad de que $X > T_r$ sea igual a 0.1.

Otras variables aleatorias continuas

11. Decimos que X es una variable aleatoria de *Laplace* de parámetros μ y $\sigma > 0$ ($X \sim L(\mu, \sigma)$) si:

$$f_X(x) = \frac{1}{2\sigma} e^{-\frac{|x-\mu|}{\sigma}} \qquad \forall x \in \mathbb{R}.$$

(esta es una versión bilateral de la v.a. exponencial). Sea $X \sim L(1, 2)$.

   (a) Calcular $P\{X > 1\}$. Sugerencia: observar que f_X es simétrica respecto de 1.

   (b) Calcular $P\{0 < X < 5\}$

12. Decimos que X es una variable aleatoria *Gumbel* de parámetros μ y $\beta > 0$ ($X \sim Gumbel(\mu, \beta)$) si su función de distribución es:

$$F_X(x) = e^{-e^{-\frac{x-\mu}{\beta}}} \qquad \forall x \in \mathbb{R}.$$

   (a) Hallar la función de densidad de esta variable aleatoria.

   (b) Supongamos que la altura máxima (en metros) que puede alcanzar el Río Uruguay a la altura de la ciudad de Salto respecto del cero local en un año puede modelarse mediante una variable aleatoria Gumbel de parámetros $\mu = 6$ y $\beta = 1$.

      i. Calcular la probabilidad de que el río tape la marca del cero local.

      ii. Calcular la probabilidad de que el agua supere la marca de 6 metros respecto del cero local.

[1] la media es el inverso del parámetro λ, es decir, hay que tomar $\lambda = 1/16$
[2] es decir, $\lambda = 1/22$