open green road

Guía Matemática

Variable aleatoria continua y distribución normal

tutor: Ismael Saldaña Caro

puntaje nacional.cl

1. Variable aleatoria continua y función de densidad de probabilidad

Previamente se definió la variable aleatoria continua (v.a.c.) como aquella cuyos elementos del recorrido (recuerde que es una función, y como tal posee dominio y recorrido) son no numerables, es decir, no se les puede asignar un orden. Por ejemplo, las estaturas de los estudiantes de un colegio, el tiempo de frecuencia de los trenes del Metro de Santiago, el peso (masa) de los integrantes de una familia, las temperaturas registradas durante un mes, etcétera. Todos estos ejemplos tienen en común que la exactitud de la medición depende de los instrumentos con que se cuente, de manera que siempre se puede llegar a una medición más precisa. Por ejemplo, la estatura de dos estudiantes es la misma e igual a 1,73 m. Si se cuenta con un instrumento de mayor precisión se puede afirmar que uno mide 1,734 m, mientras que el otro mide 1,731 m. Así, podemos ir aumentando la exactitud de las mediciones de tal forma que el orden relativo entre ellas varíe.

En resumen, una variable aleatoria continua es aquella que puede tomar cualquier valor numérico en un intervalo o conjunto de intervalos.

Al igual que para la variable aleatoria discreta, la función de probabilidad de la variable aleatoria continua describe cómo se distribuyen las probabilidades para cada uno de los valores que ésta puede adoptar. Dicha función recibe el nombre de **función de densidad de probabilidad** (o simplemente función de densidad), $f(x)$.

Dado que los elementos del dominio de $f(x)$ son valores continuos, su gráfica se representa mediante una curva en el plano cartesiano, donde el área bajo la curva entre dos puntos a y b representa la probabilidad de que la variable aleatoria tome un valor comprendido en dicho intervalo, tal como muestra la figura 1.

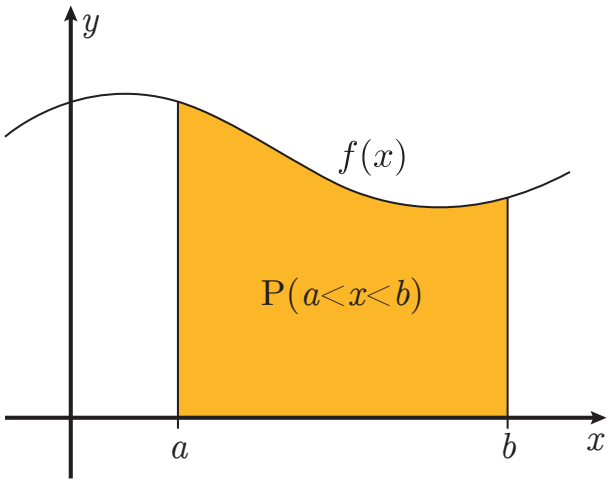


Figura 1: Gráfica de la función de densidad.

Recuerde que, al igual que para una variable aleatoria discreta, el dominio de la variable aleatoria continua corresponde al espacio muestral del experimento y su recorrido corresponde a los valores que ésta puede tomar (dependiendo de cómo se defina). Por otra parte, el dominio de la función de densidad de probabilidad corresponde al recorrido de la variable aleatoria, mientras que su recorrido corresponde a la probabilidad asociada a cada uno de los elementos del dominio, es decir, a los reales comprendidos entre 0 y 1.

Como consecuencia de lo anterior, para que f sea función de densidad de probabilidad se deben cumplir dos aspectos importantes:

1. $f(x)$ debe ser positivo o cero para todos y cada uno de los elementos del dominio de f. Esto porque por la propia definición, la probabilidad de un suceso toma valores entre 0 y 1, nunca negativos.
2. El área bajo la curva de f en su dominio debe ser 1. Esto porque, por definición, la suma de las probabilidades asociadas a cada uno de los elementos del dominio de f debe ser la unidad (o el 100 %).

Si la gráfica de f **no** cumple con alguno de los puntos mencionados, entonces **no** se trata de una función de densidad de probabilidad.

✎ ***Ejemplo***

La figura 2 muestra la función de densidad de probabilidad de una variable aleatoria continua X. Verifique que efectivamente es función de densidad.

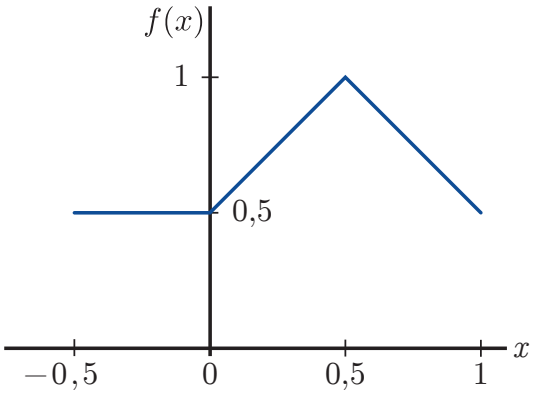


Figura 2: Función de densidad de X.

La gráfica de la figura 2 cumple con las dos condiciones de una función de densidad: f es positiva en todo su dominio y el área bajo la curva en su dominio es la unidad. Para verificar esta última afirmación considere la figura 3, en la cual se ha destacado el área de un triángulo y de un rectángulo formados a partir de f.

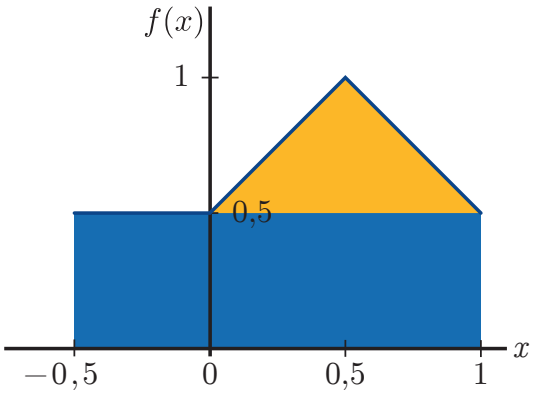


Figura 3: Áreas destacadas bajo la curva de la función de densidad de X.

Las áreas del rectángulo (a_r) y del triángulo (a_t) son respectivamente:

$$a_r = \text{base} \cdot \text{altura} = 1,5 \cdot 0,5 = 0,75$$

$$a_t = \frac{\text{base} \cdot \text{altura}}{2} = \frac{1 \cdot 0,5}{2} = 0,25$$

Luego, la suma de las áreas bajo la curva de f es $0,75 + 0,25 = 1$, con lo que la gráfica sí corresponde a una función de densidad. Aún más, de la misma gráfica se observa que el recorrido de X (dominio de f) corresponde a los reales comprendidos entre $-0,5$ y 1.

¿Cuál es la probabilidad de que X tome valores mayores o iguales que 0? ¿Cuál es la probabilidad de que X tome valores menores o iguales que 0? Dichas probabilidades corresponden al área bajo la curva descrita a la derecha e izquierda del eje de las ordenadas respectivamente, esto es, $\mathrm{P}(\mathrm{X} \geq 0) = 0{,}75$, mientras que $\mathrm{P}(\mathrm{X} \leq 0) = 0{,}25$. Verifíquelo.

Desafío 1



¿Cuál debe ser el valor de h para que la gráfica de la figura 4 sea función de densidad de probabilidad?

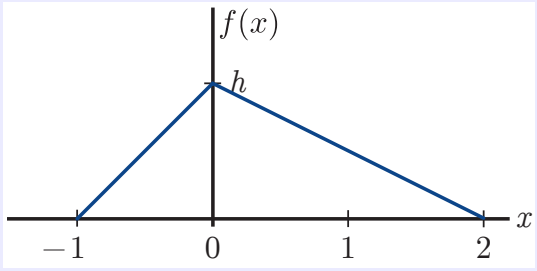


Figura 4: Función de densidad según h.

Respuesta

Otro aspecto importante de destacar de una variable aleatoria continua, es que la probabilidad de que ésta tome un valor en particular a es 0, es decir, $\mathrm{P}(\mathrm{X} = a) = 0$. La justificación es sencilla: suponga que se escoge un estudiante en particular de un colegio y se mide su estatura. Nos preguntamos cuál es la probabilidad de que dicho estudiante mida exactamente 1,71 m. Pues bien, como se explicó anteriormente, la exactitud de la medición dependerá del instrumento utilizado, de modo que un estudiante que eventualmente mide 1,71 m de altura, con otro instrumento de medición puede medir 1,712 m, con otro 1,7129 m, pero nunca llegaremos a un valor exacto.

Como consecuencia de lo anterior, la probabilidad de que una variable aleatoria continua tome valores dentro de un intervalo es la misma tanto si se considera uno, alguno o ambos extremos del intervalo, esto es, $\mathrm{P}(a < \mathrm{X} < b) = \mathrm{P}(a \leq \mathrm{X} \leq b) = \mathrm{P}(a \leq \mathrm{X} < b) = \mathrm{P}(a < \mathrm{X} \leq b)$.

Dado que la probabilidad de que una variable aleatoria continua tome un valor en particular es 0, se cumple que:

$$\begin{aligned} P(a < X < b) &= P(a \leq X \leq b) \\ &= P(a \leq X < b) \\ &= P(a < X \leq b) \end{aligned}$$

Desafío 2



¿Cuál(es) de las siguientes funciones puede(n) ser función de densidad de una variable aleatoria continua?

a) $f(x) = 0,5$, con $x \in [-1, 1]$

b) $f(x) = x - 1$, con $x \in [1, 3]$

c) $f(x) = -1$, con $x \in [-1, 0]$

d) $f(x) = |x|$ con $x \in [-1, 1]$

Respuesta

✍ *Ejercicios*

1. Determine la naturaleza de las siguientes variables: discretas o continuas.

   a) Número de vehículos que transita diariamente por las principales arterias de Santiago.

   b) Cantidad de líquido que consume una persona diariamente.

   c) La masa o el peso de una persona.

   d) Valor (en pesos) de una entrada al circo.

   e) Cantidad de masa utilizada en la elaboración de una marraqueta.

   f) Número de teclas que trae un teclado de computador.

   g) Tiempo de vida de una persona.

   h) Estatura de los habitantes de una región.

   i) Número de aplicaciones que se pueden descargar en un celular.

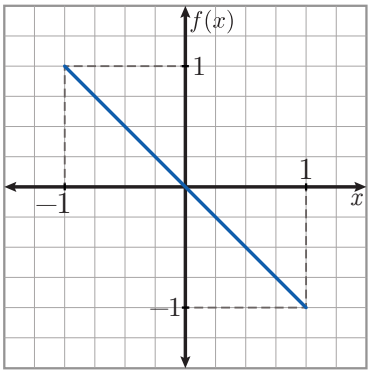
   j) Datos meteorológicos (temperatura y precipitaciones) registrados durante un periodo de tiempo.

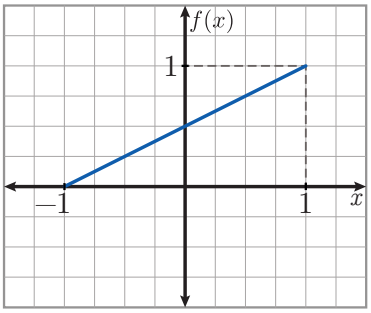
   k) Cantidad de goles marcados en un partido de fútbol.

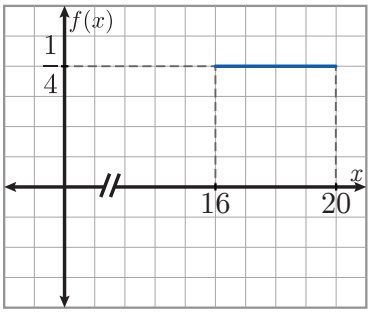
   l) Duración de la batería de un celular.

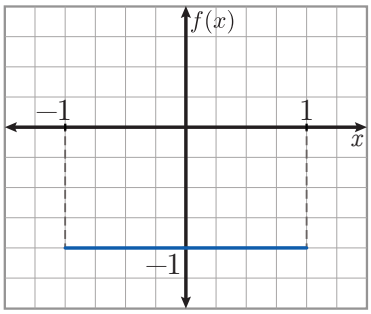
   m) Tiempo de reacción de una persona ante una emergencia.

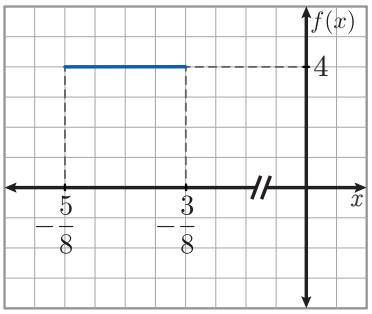
2. Determine cuál de las siguientes gráficas corresponde a una función de densidad de probabilidad.

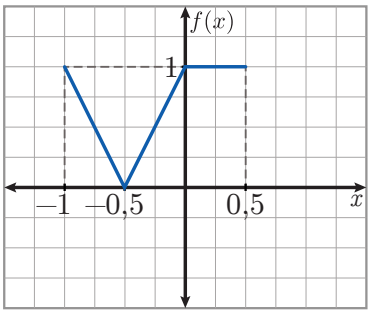
A)


B)


C)


D)


E)


F)


3. Sea $f = 0,2$, definida en el intervalo $[-2,3]$, función de densidad de la variable aleatoria continua X. Verifique que cumple con las condiciones de la función de densidad, determine el recorrido de X y calcule $P(X \leq 2)$, $P(X \geq -1)$, $P(-1 \leq X < 2)$ y $P(X = 0)$.

Respuestas:
Ejercicio 1: Variables aleatorias discretas: a), d), f), i) y k). Las restantes son variables aleatorias continuas.
Ejercicio 2: B, C, E y F son funciones de densidad.
Ejercicio 3: $P(X \leq 2) = P(X \geq -1) = 0,8$; $P(-1 \leq X < 2) = 0,6$; $P(X = 0) = 0$.

Supongamos ahora que se desea graficar la función de densidad de la variable aleatoria continua definida como la "estatura de los adultos chilenos". A falta de datos exactos, muy probablemente la función de densidad adopte la forma de la figura 5: la estatura del grueso de la población se encuentra próximo al valor promedio, mientras que una pequeña fracción de la población se ubica en los extremos de la distribución. En otras palabras, si eligiéramos una persona al azar y registrŕamos su estatura, difícilmente sería un valor próximo a 1,40 m o 2,05 m, siendo lo más común que su estatura bordee los 1,70 m.

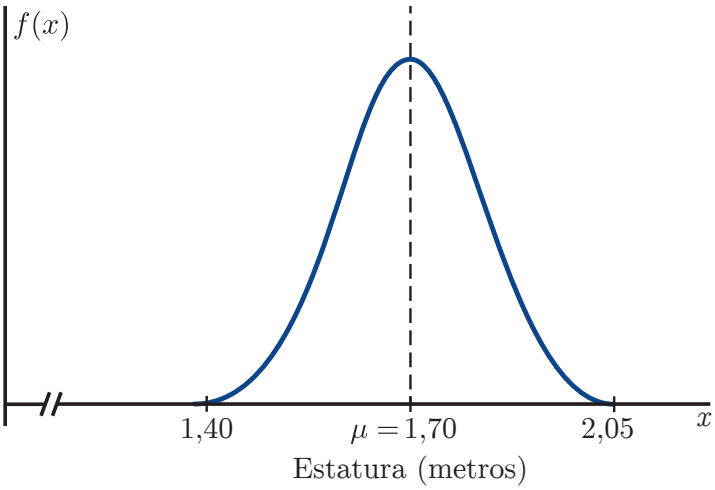


Figura 5: Función de densidad de la variable aleatoria definida como la "estatura de los adultos chilenos".

De la misma forma, se espera que la gráfica de la función de densidad de la variable aleatoria continua definida como el "puntaje promedio obtenido en la PSU" adopte la forma de la figura 5, es decir, que la mayor parte de los estudiantes obtengan puntajes cercanos al promedio, mientras que un pequeño grupo destaque por sobre el resto, ya sea por obtener puntajes muy altos o muy bajos.

La misma gráfica aplica para diversos contextos, como por ejemplo las temperaturas registradas durante el transcurso del día, la masa (peso) de un grupo de personas, las mediciones científicas, las horas de sueño de un grupo de personas, etcétera.

La curva que describe la función de densidad de los ejemplos anteriores es conocida como "campana de Gauss". Muchos fenómenos naturales, sociales y psicológicos pueden ser modelados por esta curva, de modo que resulta útil estudiarla en lo que se conoce como distribución normal, distribución de Gauss o distribución gaussiana.

2. Distribución normal

La función de densidad de una variable aleatoria continua con distribución normal queda definida por la expresión:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{1}$$

Donde μ y σ corresponden a la media aritmética y desviación estándar de los datos respectivamente.

Evidentemente, la expresión anterior resulta visualmente compleja y aterradora, sin embargo, usted no debe aprenderla de memoria, sino que debe comprender las implicancias que de ella se desprenden, tal como se verá más adelante.

Una variable aleatoria continua X con distribución normal de media μ y desviación estándar σ se denota como $X \sim N(\mu, \sigma)$ y la gráfica de su función de densidad es una curva con forma de campana, tal como se señaló previamente.

Las características a tener presente en la gráfica de una distribución normal son:

1. El área bajo la curva es igual a 1, correspondiente al 100 %.
2. Sus extremos se prolongan de modo indefinido en ambas direcciones del eje de las abscisas, y teóricamente nunca lo intersectan.
3. Es simétrica con respecto a la media aritmética, μ, es decir, el área bajo la curva a la izquierda de μ es la misma que el área a su derecha. Cuando $\mu = 0$, el eje de simetría de la distribución es el eje de las ordenadas, tal como muestra la figura 6.

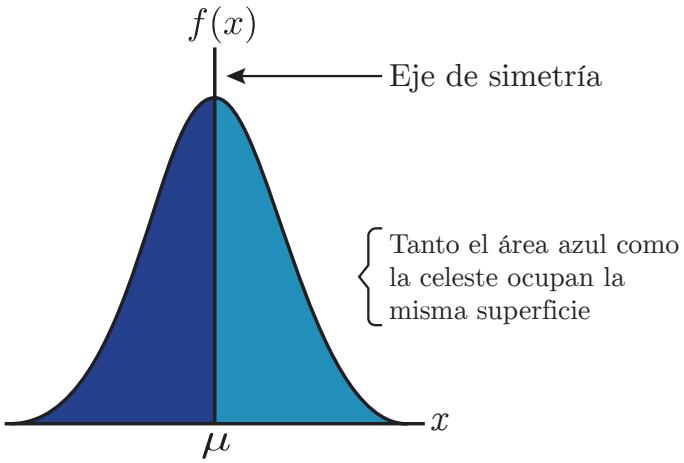


Figura 6: Eje de simetría y área bajo la curva de la gráfica de una variable aleatoria continua con distribución normal.

4. La media aritmética de la distribución de datos determina la ubicación de la curva: si μ es negativo, el punto máximo de la gráfica se ubicará en el sector negativo del eje de las abscisas. Al contrario, si μ es positivo, el punto máximo de la gráfica se ubicará en el sector positivo. Cuando μ es 0, el punto máximo de la gráfica se encuentra en el eje ordenado. Advierta que cualquiera sea el caso, el valor de $f(\mu)$ es siempre un máximo positivo. La figura 7 ilustra las dos ideas previas.

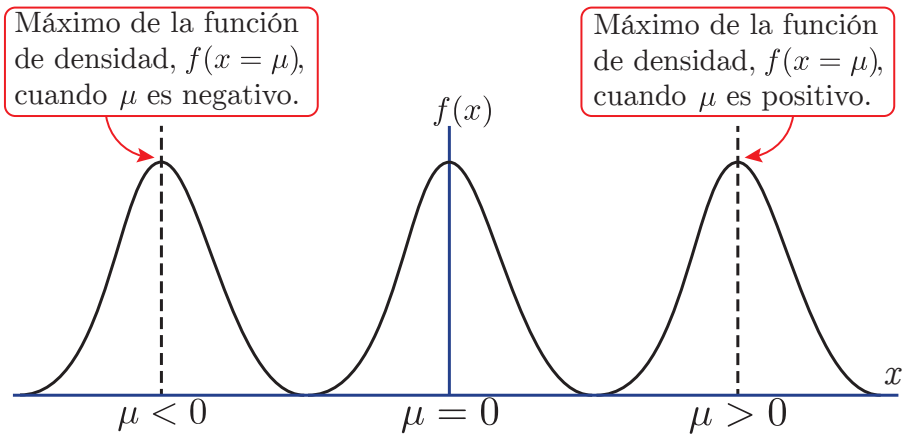


Figura 7: Ubicación de la función de densidad según el valor de la media aritmética para una distribución normal.

5. La altura de la curva (su punto máximo) depende del valor de la desviación estándar: a menor valor de la desviación estándar, mayor es su punto máximo. Al contrario, a mayor desviación estándar, menor es su punto máximo. La razón matemática de lo anterior involucra a la ecuación (1). Conceptualmente, una distribución cuya desviación estándar es pequeña nos indica que los datos se encuentran próximos a su media aritmética, aumentando así la probabilidad (área bajo la curva) en torno a μ, lo que se traduce en que la gráfica aumenta su punto máximo. Al contrario, una distribución cuya desviación estándar es grande nos indica que los datos son muy dispersos, de modo que la probabilidad no se concentra en torno a μ, sino que se reparte de forma más equitativa, lo que a su vez se traduce en que la gráfica disminuye su punto máximo.

   Ahora, advierta que al aumentar la "altura" de la gráfica de una distribución normal, su "ancho" debe reducirse a fin de que el área bajo la curva siempre sea 1. Al disminuir la "altura" de la gráfica de una distribución normal, su "ancho" debe aumentar a fin de que el área bajo la curva siempre sea 1. La figura 8 ilustra esta idea comparando ambos casos con una variable aleatoria continua que distribuye $\mathrm{N}(0,1)$.

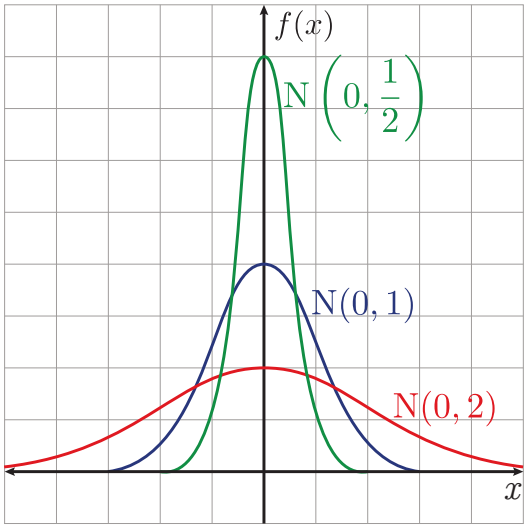


   Figura 8: Gráfica de la función de densidad para una variable aleatoria con distribución normal y según su desviación estándar.

6. La media aritmética es el valor en torno al cual es más probable que la variable aleatoria tome valores, disminuyendo la probabilidad a medida que nos alejamos de ella. En particular, se cumple que:

   • El 68,26 % de los valores de la variable aleatoria continua con distribución normal se ubica dentro de más o menos una desviación estándar de su media.

   • El 95,45 % de los valores de la variable aleatoria continua con distribución normal se ubica dentro de más o menos dos desviaciones estándar de su media.

   • El 99,73 % de los valores de la variable aleatoria continua con distribución normal se ubica dentro de más o menos tres desviaciones estándar de su media.

La figura 9 ilustra lo anterior.

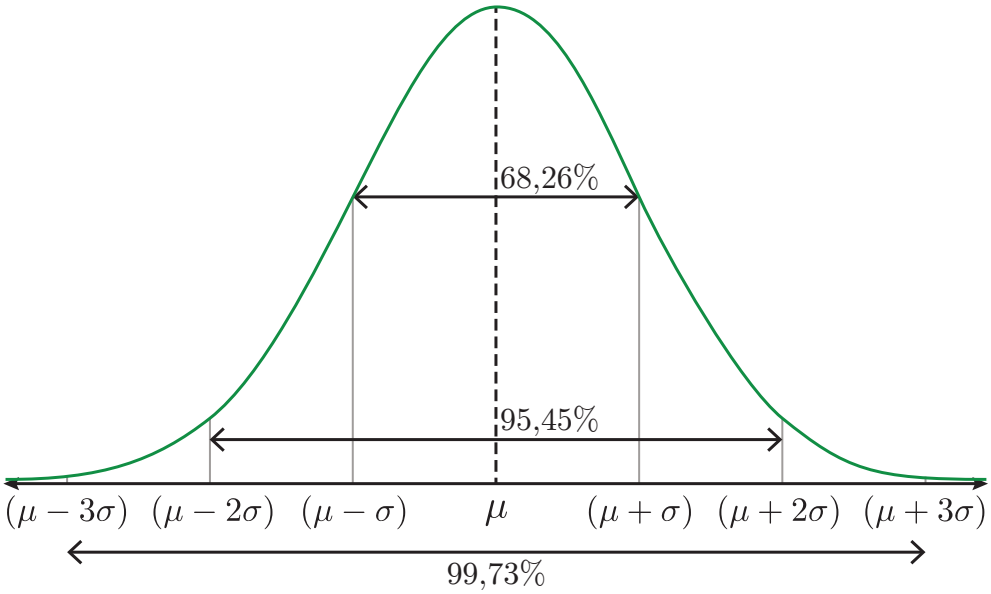
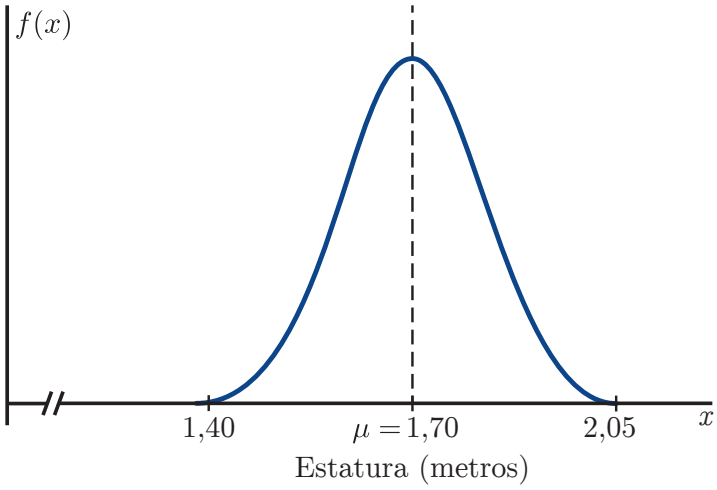


Figura 9: Probabilidad de $f(x)$ en torno a la media aritmética y según intervalo de σ.

✎ ***Ejemplo***

Considere que la desviación estándar de las estaturas del ejemplo referido a la figura 5 es $\sigma = 0,02$ m, de modo que podemos afirmar que la variable aleatoria ("estatura de los adultos chilenos") distribuye $\mathrm{N}(1,7;\ 0,02)$. ¿Qué se puede afirmar al respecto? (Tenga presente que los datos son ejemplificadores y no corresponden a la realidad).



El problema consiste en una distribución normal de media 1,7 m y desviación estándar 0,02 m. A partir de esta información de inmediato se puede afirmar que:

• La estatura promedio de los adultos chilenos es 1,7 m, con una desviación estándar de 2 cm.

• El máximo de la gráfica de la distribución normal se ubica en el primer cuadrante del plano cartesiano (lado positivo del eje de las abscisas).

• El 50 % de los chilenos mide menos de 1,7 m. La mitad restante mide más de 1,7 m (¿qué pasa con exactamente 1,7 m?). Esto a razón de la simetría de la gráfica con respecto a su media.

• El 68,26 % de los adultos chilenos mide entre 1,68 m y 1,72 m, correspondiente al intervalo $[\mu - \sigma,\ \mu + \sigma]$.

• El 95,45 % de los adultos chilenos mide entre 1,66 m y 1,74 m, correspondiente al intervalo $[\mu - 2\sigma,\ \mu + 2\sigma]$.

• El 99,73 % de los adultos chilenos mide entre 1,64 m y 1,76 m, correspondiente al intervalo $[\mu - 3\sigma,\ \mu + 3\sigma]$.

Desafío 3



Dada una variable aleatoria continua con distribución normal, determine la probabilidad de que ésta tome valores mayores que μ–3σ, μ–2σ, μ–σ, $\mu + \sigma$, $\mu + 2\sigma$, $\mu + 3\sigma$, menores que cada uno de los valores anteriores y entre μ y cada uno de los valores anteriores.

Respuesta

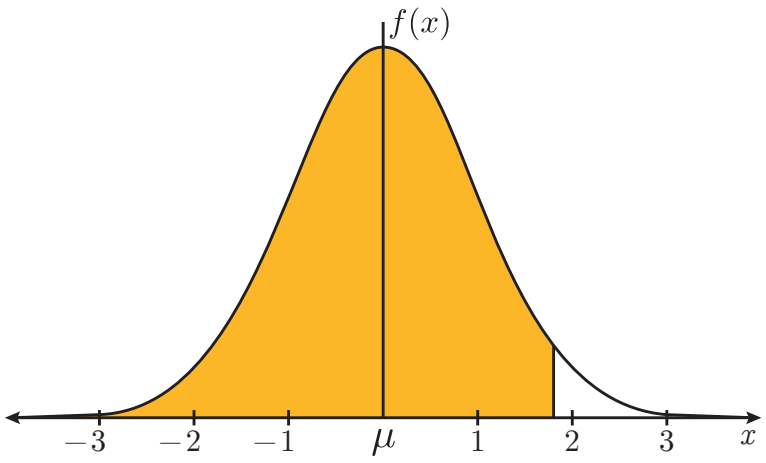
3. Distribución normal estándar

Un caso particular de variable aleatoria continua con distribución normal es cuando sus parámetros son $\mu = 0$ y $\sigma = 1$. Dicha distribución se conoce como **distribución normal estándar** y se denota como $X \sim N(0, 1)$.

Su utilidad radica en que, como indica su nombre, ha sido estandarizada respecto de una variable aleatoria de diferentes parámetros, y a partir de ella se han elaborado tablas que permiten determinar de manera rápida las probabilidades de que un valor de la variable se encuentre dentro de un intervalo particular. En efecto, el cuadro 1 en el anexo al final de este documento permite calcular la probabilidad de que la variable aleatoria X con distribución normal estándar sea menor que un valor z dado, esto es, $P(X \leq z)$. Los siguientes ejemplos ilustran su utilización.

✎ ***Ejemplo***

1. Dada una variable aleatoria continua que distribuye $\mathrm{N}(0,1)$, calcule la probabilidad de que tome un valor menor que 1,87.



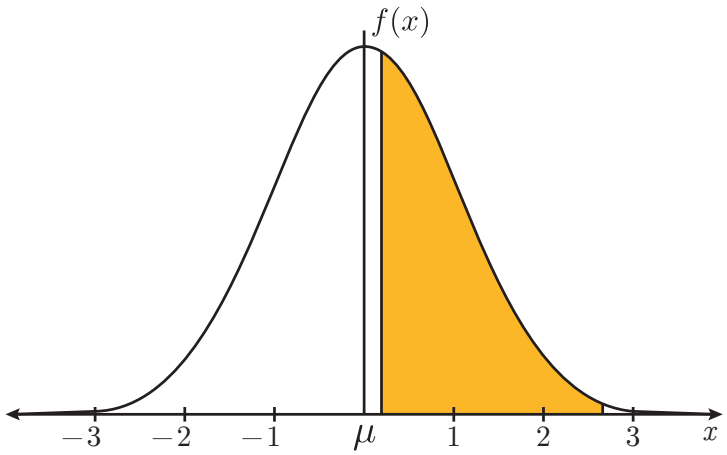
El cuadro 1 en el anexo permite calcular la probabilidad de que la variable aleatoria continua con distribución normal estándar sea menor que un valor z dado. En este caso en particular $z = 1,87$, de modo que debemos buscar dicho valor en la tabla y ver su probabilidad asociada.

La primera columna (vertical) de la tabla indica el valor que adopta la variable aleatoria continua cuando solo se considera su parte entera y su primer decimal. Por otro lado, la primera fila (horizontal) indica el valor del segundo decimal de la variable aleatoria. Así, la probabilidad asociada a un valor z dado corresponde al número que se encuentra en la intersección de la fila y columna que lo componen.

Con lo anterior, la probabilidad asociada a $z = 1,87$ se ubica en la intersección de la fila cuyo valor es 1,8 y de la columna cuyo valor es 0,07. La probabilidad buscada es 0,96926. Así, la probabilidad de que la variable aleatoria continua tome valores menores que 1,87 es aproximadamente un 97 %.

x	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
	⋮				⋮					⋮
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,9608	0,96164	0,96246	0,96327
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97128	0,97193	0,97257	0,9732	0,97381	0,97441	0,975	0,97558	0,97615	0,9767

2. Dada una variable aleatoria continua que distribuye N$(0, 1)$, calcule la probabilidad de que tome un valor entre 0,16 y 2,93.



Nuevamente, para resolver el problema utilizamos el cuadro 1 en el anexo. No obstante, ahora no estamos interesados en obtener la probabilidad de que la variable aleatoria sea menor a un z dado, sino que ésta se encuentre entre dos valores, a saber, entre 0,16 y 2,93. Para solucionar este problema advierta que la probabilidad buscada es equivalente a la diferencia entre P(X $<$ 2,93) y P(X $<$ 0,16), esto es:

$$\mathrm{P}(0,16 < \mathrm{X} < 2,93) = \mathrm{P}(\mathrm{X} < 2,93) - \mathrm{P}(\mathrm{X} < 0,16)$$

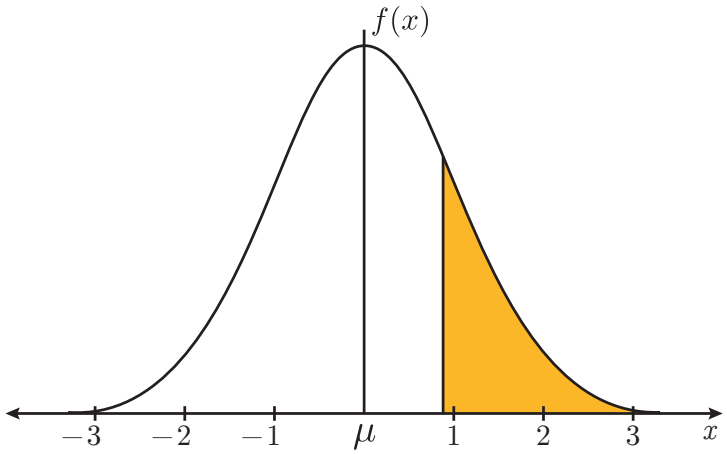
Reemplazando en la expresión anterior con los valores proporcionados por el cuadro 1 en el anexo:

$$\mathrm{P}(0,16 < \mathrm{X} < 2,93) = 0,97882 - 0,56356$$

$$\mathrm{P}(0,16 < \mathrm{X} < 2,93) = 0,41526$$

Así, la probabilidad de que la variable aleatoria tome un valor entre 0,16 y 2,93 es 0,41526, que a su vez equivale aproximadamente al 41,5 %.

3. Dada una variable aleatoria continua que distribuye N$(0, 1)$, calcule la probabilidad de que tome un valor mayor que 0,94.



A partir del cuadro 1 del anexo se obtiene la probabilidad de que la variable aleatoria tome un valor menor a 0,94. Por otra parte, el problema hace referencia a la probabilidad del suceso complementario, es decir:

$$\mathrm{P}(\mathrm{X} > 0,94) = 1 - \mathrm{P}(\mathrm{X} < 0,94)$$

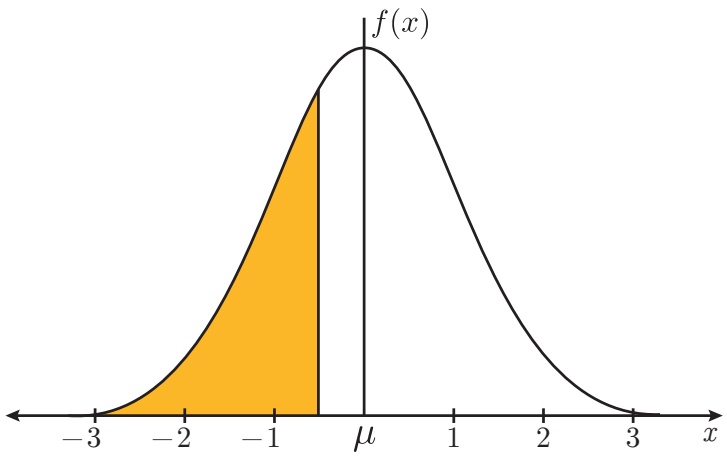
Luego, reemplazando valores en la expresión anterior de acuerdo con el cuadro 1 del anexo:

$$\mathrm{P}(\mathrm{X} > 0,94) = 1 - 0,82639$$

$$\mathrm{P}(\mathrm{X} > 0,94) = 0,17361$$

Con lo anterior, hay aproximadamente un 17 % de probabilidad de que la variable aleatoria tome valores mayores que 0,94.

4. Dada una variable aleatoria continua que distribuye $\mathrm{N}(0,1)$, calcule la probabilidad de que tome un valor menor que $-0,59$.



En este caso el cálculo no es directo, pues el cuadro 1 del anexo nos entrega las probabilidades asociadas para valores positivos de la variable aleatoria, no así negativos. Para dar con la solución vamos a ocupar el hecho de que la distribución normal estándar es simétrica con respecto del eje de las ordenadas, de modo que $\mathrm{P}(\mathrm{X} < -0,59)$ es equivalente a $\mathrm{P}(\mathrm{X} > 0,59)$. Este último caso es análogo al anterior, siendo 0,2776 la probabilidad buscada ($\approx 27,8\,\%$). Su desarrollo se deja propuesto.

Desafío 4



Dada una variable aleatoria continua que distribuye $\mathrm{N}(0,1)$, ¿cuál debe ser el valor de a tal que la probabilidad de tomar valores entre $-2,53$ y a sea del 89,031 %?

Respuesta

Previamente se afirmó que una distribución normal estándar es (como su nombre lo indica) una estandarización de una distribución normal de diferentes parámetros ($\mu \neq 0$ o $\sigma \neq 1$) a fin de simplificar cálculos y representaciones. Entonces, la pregunta que resta hacerse es: ¿cuál es el proceso que permite estandarizar una distribución normal?

Sea x uno de los valores que puede tomar una variable aleatoria continua X que distribuye N(μ, σ). Se define z como el valor que puede tomar la variable aleatoria continua Z de la siguiente manera:

$$z = \frac{x - \mu}{\sigma} \tag{2}$$

Luego, Z es una variable aleatoria continua con distribución normal estándar. Verifiquemos esta situación:

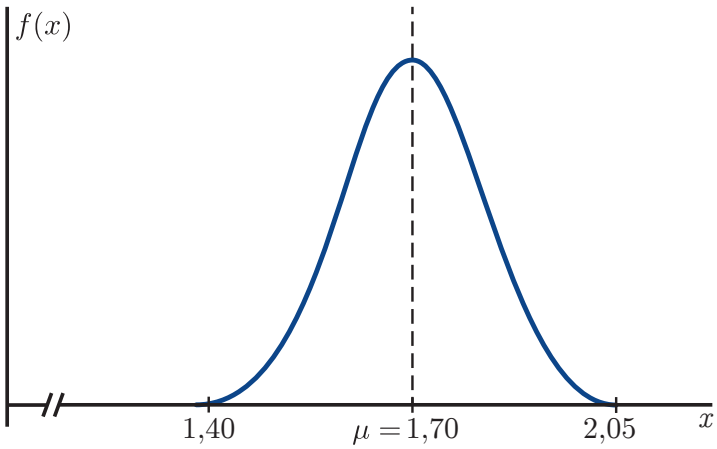
• Cuando X $\sim$ N(0, 1), implica que $z = \dfrac{x - 0}{1} = x$. Luego, tanto X como Z toman los mismos valores, y como consecuencia ambas variables distribuyen en forma normal estándar.

• Cuando X $\sim$ N(μ, σ), el valor de z cuando $x = \mu$ es $z = \dfrac{\mu - \mu}{\sigma} = 0$. Es decir, la media aritmética de X (que es distinta de 0) se traslada al origen de coordenadas en Z, tal cual una distribución normal estándar.

Una vez que una variable aleatoria con distribución normal ha sido estandarizada es posible utilizar el cuadro 1 del anexo para calcular las probabilidades asociadas.

✎ *Ejemplo*

Nuevamente, considere el ejemplo de la figura 5 referido a la estatura de los adultos chilenos, las cuales distribuyen N(1,7; 0,02). ¿Qué porcentaje de los adultos chilenos mide menos de 1,65 m?



Observe que la variable aleatoria continua X definida como "estatura de los adultos chilenos" no es una distribución normal estándar, de modo que no es posible ocupar el cuadro 1 del anexo para obtener la probabilidad deseada. Como consecuencia, resulta imperativo estandarizar dicha distribución.

De acuerdo con el enunciado, $\mu = 1,7$ m, $\sigma = 0,02$ m y $x = 1,65$ m. Reemplazando dichos valores en (2) se obtiene:

$$z = \frac{x - \mu}{\sigma} = \frac{1,65 - 1,7}{0,02} = -2,5$$

Luego, el problema se reduce a determinar la probabilidad de que una variable aleatoria Z que distribuye N(0, 1) tome valores menores que $-2,5$.

Para ello, tal como se enunció previamente, hacemos uso de la simetría de la distribución normal estándar, de tal forma que:

$$P(Z < -2,5) = 1 - P(Z < 2,5)$$

Reemplazando en la expresión anterior los valores proporcionados por el cuadro 1 del anexo:

$$P(Z < -2,5) = 1 - 0,99379$$

$$P(Z < -2,5) = 0,00621$$

Así, aproximadamente el 0,6 % de los adultos chilenos mide menos de 1,65 m.

Problemas similares se resuelven de la misma manera: se realiza la estandarización de la distribución mediante el cambio de variable, se calculan los valores de interés y se busca la probabilidad asociada a cada uno según el cuadro 1 del anexo.

✍ *Ejercicios*

1. En un colegio mixto, 220 son los estudiantes de cuarto medio: 80 hombres y 140 mujeres. La masa de los estudiantes varones distribuye N(80, 0,3), mientras que la masa de las damas distribuye N(69, 0,4), ambos en kilogramos.

   a) ¿Cuántos estudiantes varones pesan (masan) más de 80 kg?
   b) ¿Cuántas estudiantes mujeres pesan (masan) menos de 69 kg?
   c) ¿Qué porcentaje de los varones pesa (masa) entre 79,4 kg y 80,6 kg?
   d) ¿Qué porcentaje de las damas pesa (masa) entre 68,6 kg y 69,4 kg?
   e) ¿Qué porcentaje de los varones pesa (masa) menos de 79,4 kg?
   f) ¿Qué porcentaje de las damas pesa (masa) más de 69,4 kg?

2. Considere una variable aleatoria que distribuye N(30, 2). ¿Qué porcentaje de la distribución es menor que 30, 28, 26 y 24? ¿Qué porcentaje es mayor que 30, 32, 34 y 36?

3. El tiempo que emplean los estudiantes de ingeniería matemática en rendir un examen distribuye N(300, 35), en minutos. Al respecto:

   a) ¿Cuál es el tiempo promedio empleado por los estudiantes en rendir el examen?
   b) ¿Qué porcentaje de los estudiantes emplea exactamente 5 horas en rendir el examen?
   c) ¿Qué porcentaje de los estudiantes emplea más de 6 horas en rendir el examen?
   d) ¿Qué porcentaje de los estudiantes emplea menos de 4 horas en rendir el examen?
   e) ¿Qué porcentaje de los estudiantes emplea entre 4 y 6 horas en rendir el examen?

4. Considere una variable aleatoria que distribuye N(0, 1).

   a) ¿Cuál es la expresión algebraica de la función de densidad?
   b) ¿Cuál es la probabilidad de que la variable tome valores entre $-1,24$ y $3,02$?

5. Considere una variable aleatoria con distribución normal estándar. A partir del cuadro 1 del anexo verifique que se cumplen las propiedades de la figura 9, esto es, verifique que se cumplen los porcentajes asociados a cada uno de los intervalos dentro de la distribución.

Respuestas:
Ejercicio 1: a) 40, b) 70, c) 95,45 %, d) 68,26 %, e) ≈2,28 % y f) 15,87 %.
Ejercicio 2: 50 %, 15,87 %, 2,28 % y 0,14 % para ambos casos respectivamente.
Ejercicio 3: a) 5 horas, b) 0 %, c) ≈ 4 %, d) ≈ 4 % y e) ≈ 92 %.
Ejercicio 4: a) $f(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$ y b) 89,125 %

Anexo

Cuadro 1: Tabla de probabilidad ($P(X < z)$) para una variable aleatoria continua X con distribución normal estándar $N(0, 1)$.

x	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0,5	0,50399	0,50798	0,51197	0,51595	0,51994	0,52392	0,5279	0,53188	0,53586
0,1	0,53983	0,5438	0,54776	0,55172	0,55567	0,55962	0,56356	0,56749	0,57142	0,57535
0,2	0,57926	0,58317	0,58706	0,59095	0,59483	0,59871	0,60257	0,60642	0,61026	0,61409
0,3	0,61791	0,62172	0,62552	0,6293	0,63307	0,63683	0,64058	0,64431	0,64803	0,65173
0,4	0,65542	0,6591	0,66276	0,6664	0,67003	0,67364	0,67724	0,68082	0,68439	0,68793
0,5	0,69146	0,69497	0,69847	0,70194	0,7054	0,70884	0,71226	0,71566	0,71904	0,7224
0,6	0,72575	0,72907	0,73237	0,73565	0,73891	0,74215	0,74537	0,74857	0,75175	0,7549
0,7	0,75804	0,76115	0,76424	0,7673	0,77035	0,77337	0,77637	0,77935	0,7823	0,78524
0,8	0,78814	0,79103	0,79389	0,79673	0,79955	0,80234	0,80511	0,80785	0,81057	0,81327
0,9	0,81594	0,81859	0,82121	0,82381	0,82639	0,82894	0,83147	0,83398	0,83646	0,83891
1	0,84134	0,84375	0,84614	0,84849	0,85083	0,85314	0,85543	0,85769	0,85993	0,86214
1,1	0,86433	0,8665	0,86864	0,87076	0,87286	0,87493	0,87698	0,879	0,881	0,88298
1,2	0,88493	0,88686	0,88877	0,89065	0,89251	0,89435	0,89617	0,89796	0,89973	0,90147
1,3	0,9032	0,9049	0,90658	0,90824	0,90988	0,91149	0,91308	0,91466	0,91621	0,91774
1,4	0,91924	0,92073	0,9222	0,92364	0,92507	0,92647	0,92785	0,92922	0,93056	0,93189
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,6	0,9452	0,9463	0,94738	0,94845	0,9495	0,95053	0,95154	0,95254	0,95352	0,95449
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,9608	0,96164	0,96246	0,96327
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97128	0,97193	0,97257	0,9732	0,97381	0,97441	0,975	0,97558	0,97615	0,9767
2	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,9803	0,98077	0,98124	0,98169
2,1	0,98214	0,98257	0,983	0,97882	0,98382	0,98422	0,98461	0,985	0,98537	0,98574
2,2	0,9861	0,98645	0,98679	0,97882	0,98745	0,98778	0,98809	0,9884	0,9887	0,98899
2,3	0,98928	0,98956	0,98983	0,97882	0,99036	0,99061	0,99086	0,99111	0,99134	0,99158
2,4	0,9918	0,99202	0,99224	0,97882	0,99266	0,99286	0,99305	0,99324	0,99343	0,99361
2,5	0,99379	0,99396	0,99413	0,97882	0,99446	0,99461	0,99477	0,99492	0,99506	0,9952
2,6	0,99534	0,99547	0,9956	0,97882	0,99585	0,99598	0,99609	0,99621	0,99632	0,99643
2,7	0,99653	0,99664	0,99674	0,97882	0,99693	0,99702	0,99711	0,9972	0,99728	0,99736
2,8	0,99744	0,99752	0,9976	0,97882	0,99774	0,99781	0,99788	0,99795	0,99801	0,99807
2,9	0,99813	0,99819	0,99825	0,97882	0,99836	0,99841	0,99846	0,99851	0,99856	0,99861
3	0,99865	0,99869	0,99874	0,97882	0,99882	0,99886	0,99889	0,99893	0,99896	0,999
3,1	0,99903	0,99906	0,9991	0,97882	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929
3,2	0,99931	0,99934	0,99936	0,97882	0,9994	0,99942	0,99944	0,99946	0,99948	0,9995
3,3	0,99952	0,99953	0,99955	0,97882	0,99958	0,9996	0,99961	0,99962	0,99964	0,99965
3,4	0,99966	0,99968	0,99969	0,97882	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976
3,5	0,99977	0,99978	0,99978	0,97882	0,9998	0,99981	0,99981	0,99982	0,99983	0,99983
3,6	0,99984	0,99985	0,99985	0,97882	0,99986	0,99987	0,99987	0,99988	0,99988	0,99989
3,7	0,99989	0,9999	0,9999	0,97882	0,99991	0,99991	0,99992	0,99992	0,99992	0,99992
3,8	0,99993	0,99993	0,99993	0,97882	0,99994	0,99994	0,99994	0,99995	0,99995	0,99995
3,9	0,99995	0,99995	0,99996	0,97882	0,99996	0,99996	0,99996	0,99996	0,99997	0,99997
4	0,99997	0,99997	0,99997	0,97882	0,99997	0,99997	0,99998	0,99998	0,99998	0,99998

Desafíos resueltos

✓ Desafío 1:

Para que la gráfica de la figura 4 sea función de densidad se deben cumplir dos condiciones: f debe ser positiva en todo su dominio y su área bajo la curva debe ser 1. La primera condición se cumple, pues $f(x) \geq 0 \; \forall x \in \text{Dom}\{f\}$ (la función es mayor o igual a 0 para todos los elementos pertenecientes al dominio). Luego, se debe verificar el cumplimiento de la segunda condición.

Para ello, considere el triángulo de vértices A $= (-1, 0)$, B $= (0, h)$ y C $= (0, 0)$, de base $\overline{\text{AC}} = |0 - (-1)| = 1$ y altura $\overline{\text{BC}} = |h - 0| = h$. Luego, el área del triángulo ABC es:

$$\text{área } \triangle\text{ABC} = \frac{\text{base} \cdot \text{altura}}{2} = \frac{1 \cdot h}{2} = \frac{h}{2}$$

Del mismo modo, considere el triángulo de vértices B $= (0, h)$, C $= (0, 0)$ y D $= (2, 0)$, de base $\overline{\text{CD}} = |2 - 0| = 2$ y altura $\overline{\text{BC}} = |h - 0| = h$. Así, el área del triángulo BCD es:

$$\text{área } \triangle\text{BCD} = \frac{\text{base} \cdot \text{altura}}{2} = \frac{2 \cdot h}{2} = h$$

Finalmente, puesto que el área bajo la curva de f corresponde a la suma de las áreas de los triángulos ABC y BCD, se debe cumplir que:

$$\text{área } \triangle\text{ABC} + \text{área } \triangle\text{BCD} = 1$$

$$\frac{h}{2} + h = 1$$

$$\frac{3}{2}h = 1$$

$$h = \frac{2}{3}$$

Así, el valor de h tal que la gráfica de la figura 4 sea función de densidad es $\frac{2}{3}$.

Nota: para efectos de los cálculos previos, las unidades de medida de longitud y área se expresan en "unidades" y "unidades cuadradas" respectivamente.
Volver

✓ Desafío 2:

Al igual que en el desafío 1, se debe verificar que las imágenes de f sean positivas o cero en todo su dominio y que el área bajo la curva sea 1. Veamos cada uno de los casos:

- $f(x) = 0,5$ con $x \in [-1, 1]$:

La función es una constante positiva en todo su dominio, por lo que cumple con la primera condición. Para verificar la segunda condición calculamos el área bajo la curva de f mediante:

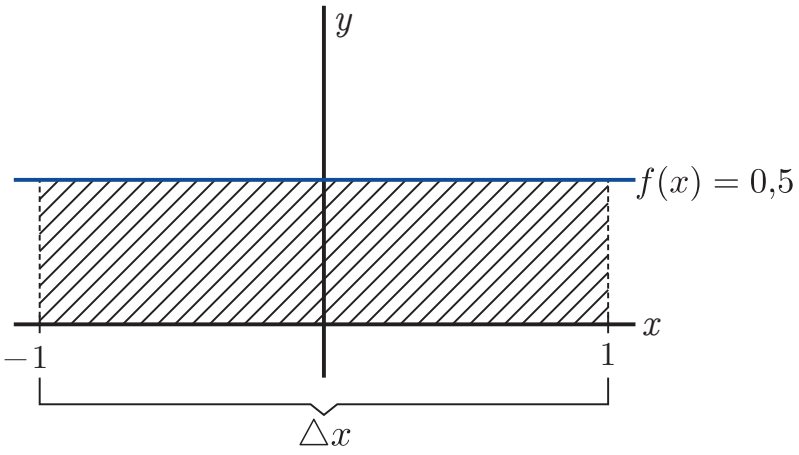
$$\Delta x \cdot f(x)$$

Donde Δx corresponde a la longitud del intervalo que define el dominio de la función y cuyo valor es 2, verifíquelo.

Reemplazando valores:

$$2 \cdot 0,5 = 1$$

Luego, f es función de densidad.



• $f(x) = x - 1$ con $x \in [1, 3]$:

La función toma valores comprendidos entre 0 y 2, de modo que cumple con la primera condición. Esto se verifica fácilmente reemplazando en la función los valores extremos del dominio:

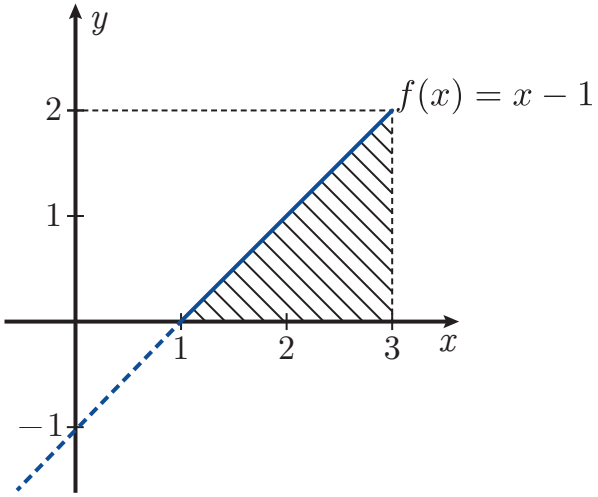
$$f(1) = 1 - 1 = 0 \qquad \text{y} \qquad f(3) = 3 - 1 = 2$$

El procedimiento anterior resulta siempre verdadero puesto que la función que se analiza es una recta de pendiente positiva, de modo que no se pierde información referente a los elementos intermedios del dominio.

Por otra parte, el área bajo la curva de f corresponde al área de un triángulo de base y altura igual a 2 unidades, esto es:

$$\frac{2 \cdot 2}{2} = 2$$

Luego, dado que el área bajo la curva de f es distinto de 1, **no** corresponde a una función de densidad.



- $f(x) = -1$ con $x \in [-1, 0]$:

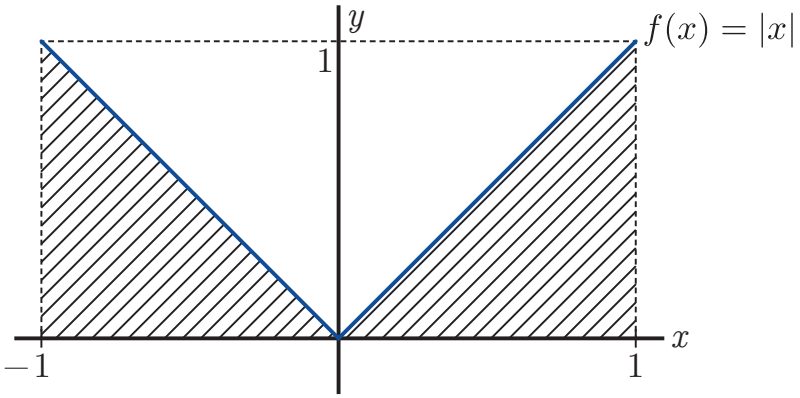
La función no cumple con la primera condición porque es negativa en todo su dominio. Luego, **no** es función de densidad.

- $f(x) = |x|$ con $x \in [-1, 1]$:

Dado que la función se define como un valor absoluto, es positiva en todo su dominio. Por otra parte, f define dos triángulos de base y altura iguales cuya longitud es 1 unidad. Luego, el área bajo la curva de f es:

$$2 \cdot \frac{1 \cdot 1}{2} = 1$$

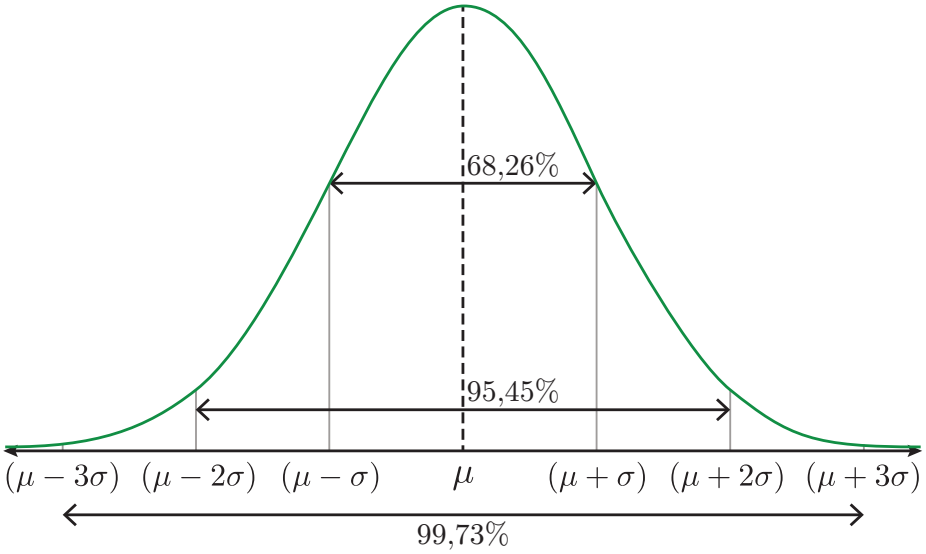
Así, puesto que el área bajo la curva de f es igual 1, sí es función de densidad.



Volver

✓ Desafío 3:

Para dar solución al problema vamos a considerar la figura 9 y nos aprovecharemos de que la distribución normal es simétrica respecto de la media aritmética de los datos.



Veamos cada una de las probabilidades solicitadas:

• $P(X > (\mu - 3\sigma))$:

De acuerdo con la figura 9, el 99,73 % de los datos se encuentra entre $\mu - 3\sigma$ y $\mu + 3\sigma$. Dado que la distribución es simétrica con respecto a la media aritmética, se concluye que aproximadamente el 49,9 % de los datos (la mitad del 99,73 %) se encuentra entre $\mu - 3\sigma$ y μ.

Por otra parte, por definición se cumple que el intervalo comprendido entre μ y ∞ contiene el 50 % de los datos. Luego, la probabilidad de que la variable aleatoria tome valores mayores que $\mu - 3\sigma$, es:

$$\begin{aligned} P(X > (\mu - 3\sigma)) &= P(\mu - 3\sigma < X < \mu) + P(X > \mu) \\ P(X > (\mu - 3\sigma)) &= 49,9\,\% + 50\,\% \\ P(X > (\mu - 3\sigma)) &= 99,9\,\% \end{aligned}$$

• $P(X > (\mu - 2\sigma))$:

Procediendo de manera análoga al caso anterior, se verifica que aproximadamente el 47,7 % de los datos se encuentra entre $\mu - 2\sigma$ y μ. Así:

$$\begin{aligned} P(X > (\mu - 2\sigma)) &= P(\mu - 2\sigma < X < \mu) + P(X > \mu) \\ P(X > (\mu - 2\sigma)) &= 47,7\,\% + 50\,\% \\ P(X > (\mu - 2\sigma)) &= 97,7\,\% \end{aligned}$$

• $P(X > (\mu - \sigma))$:

Procediendo de manera análoga al primer caso, se verifica que aproximadamente el 34,1 % de los datos se encuentra entre $\mu - \sigma$ y μ. Así:

$$\begin{aligned} P(X > (\mu - \sigma)) &= P(\mu - \sigma < X < \mu) + P(X > \mu) \\ P(X > (\mu - \sigma)) &= 34,1\,\% + 50\,\% \\ P(X > (\mu - \sigma)) &= 84,1\,\% \end{aligned}$$

• $P(X > (\mu + \sigma))$:

Por los cálculos previos sabemos que aproximadamente el 34,1 % de los datos se encuentra entre μ y $\mu + \sigma$ (o equivalentemente entre $\mu - \sigma$ y μ). Por otra parte, dado que el intervalo comprendido entre μ y ∞ contiene el 50 % de los datos, se concluye que la probabilidad buscada es:

$$\begin{aligned} P(X > (\mu + \sigma)) &= P(X > \mu) - P(\mu < X < \mu + \sigma) \\ P(X > (\mu + \sigma)) &= 50\,\% - 34,1\,\% \\ P(X > (\mu + \sigma)) &= 15,9\,\% \end{aligned}$$

- $P(X > (\mu + 2\sigma))$:

Por los cálculos previos se verifica que aproximadamente el 47,7 % de los datos se encuentra entre μ y $\mu + 2\sigma$ (o equivalentemente entre $\mu - 2\sigma$ y μ). Luego:

$$\begin{aligned} P(X > (\mu + 2\sigma)) &= P(X > \mu) - P(\mu < X < \mu + 2\sigma) \\ P(X > (\mu + 2\sigma)) &= 50\,\% - 47,7\,\% \\ P(X > (\mu + 2\sigma)) &= 2,3\,\% \end{aligned}$$

- $P(X > (\mu + 3\sigma))$:

Análogamente al caso anterior, se verifica que aproximadamente el 49,9 % de los datos se encuentra entre μ y $\mu + 3\sigma$ (o equivalentemente entre $\mu - 3\sigma$ y μ). Luego:

$$\begin{aligned} P(X > (\mu + 3\sigma)) &= P(X > \mu) - P(\mu < X < \mu + 3\sigma) \\ P(X > (\mu + 3\sigma)) &= 50\,\% - 49,9\,\% \\ P(X > (\mu + 3\sigma)) &= 0,1\,\% \end{aligned}$$

Ahora calculemos la probabilidad de que la variable aleatoria sea **menor** que un valor dado. Para ello vamos hacer uso de los cálculos previos y de la propiedad referida a sucesos complementarios:

$$P(X < z) = 1 - P(X > z)$$

- $P(X < (\mu - 3\sigma))$:

$$\begin{aligned} P(X < (\mu - 3\sigma)) &= 100\,\% - P(X > (\mu - 3\sigma)) \\ P(X < (\mu - 3\sigma)) &= 100\,\% - 99,9\,\% \\ P(X < (\mu - 3\sigma)) &= 0,1\,\% \end{aligned}$$

- $P(X < (\mu - 2\sigma))$:

$$\begin{aligned} P(X < (\mu - 2\sigma)) &= 100\,\% - P(X > (\mu - 2\sigma)) \\ P(X < (\mu - 2\sigma)) &= 100\,\% - 97,7\,\% \\ P(X < (\mu - 2\sigma)) &= 2,3\,\% \end{aligned}$$

- $P(X < (\mu - \sigma))$:

$$\begin{aligned} P(X < (\mu - \sigma)) &= 100\,\% - P(X > (\mu - \sigma)) \\ P(X < (\mu - \sigma)) &= 100\,\% - 84,1\,\% \\ P(X < (\mu - \sigma)) &= 15,9\,\% \end{aligned}$$

- $P(X < (\mu + \sigma))$:

$$\begin{aligned} P(X < (\mu + \sigma)) &= 100\,\% - P(X > (\mu + \sigma)) \\ P(X < (\mu + \sigma)) &= 100\,\% - 15,9\,\% \\ P(X < (\mu + \sigma)) &= 84,1\,\% \end{aligned}$$

- $P(X < (\mu + 2\sigma))$:

$$\begin{aligned} P(X < (\mu + 2\sigma)) &= 100\,\% - P(X > (\mu + 2\sigma)) \\ P(X < (\mu + 2\sigma)) &= 100\,\% - 2,3\,\% \\ P(X < (\mu + 2\sigma)) &= 97,7\,\% \end{aligned}$$

- $P(X < (\mu + 3\sigma))$:

$$\begin{aligned} P(X < (\mu + 3\sigma)) &= 100\,\% - P(X > (\mu + 3\sigma)) \\ P(X < (\mu + 3\sigma)) &= 100\,\% - 0,1\,\% \\ P(X < (\mu + 3\sigma)) &= 99,9\,\% \end{aligned}$$

En rigor, los cálculos previos no eran necesarios, pues las probabilidades buscadas se podrían haber obtenido a partir de la simetría de la distribución normal. Por ejemplo, por simetría se cumple que:

$$P(X < (\mu + 3\sigma)) = P(X > (\mu - 3\sigma))$$

Verifique la igualdad anterior y establezca las relaciones restantes.
Volver

✓ Desafío 4:

Dado que la variable aleatoria distribuye de forma normal, para el cálculo de las probabilidades debemos utilizar el cuadro 1 del anexo. Por otro lado, advierta que dicho cuadro nos entrega la probabilidad de que la variable aleatoria sea **menor** que un z positivo determinado, razón por la cual debemos encontrar un procedimiento que nos permita utilizarla según nuestro propósito.

Para dar con la probabilidad buscada considere la siguiente igualdad:

$$P(-2,53 < X < a) = P(X < a) - P(X < -2,53)$$

Por otra parte, por simetría $P(X < -2,53)$ es equivalente a $P(X > 2,53)$ y este último a $1 - P(X < 2,53)$. Reemplazando en la igualdad anterior:

$$P(-2,53 < X < a) = P(X < a) - (1 - P(X < 2,53))$$

$$P(-2,53 < X < a) = P(X < a) + P(X < 2,53) - 1$$

Luego, reemplazando valores en la igualdad anterior conforme con el enunciado y con el cuadro 1 del anexo:

$$0,89031 = \mathrm{P}(\mathrm{X} < a) + 0,97882 - 1$$

$$0,89031 = \mathrm{P}(\mathrm{X} < a) - 0,02118$$

$$0,89031 + 0,02118 = \mathrm{P}(\mathrm{X} < a)$$

$$0,91149 = \mathrm{P}(\mathrm{X} < a)$$

Finalmente, para determinar a debemos buscar en el cuadro 1 del anexo aquel valor cuya probabilidad asociada sea exactamente $0,91149$. Dicho valor corresponde a 1,35, verifíquelo. Así, $a = 1,35$.
Volver

Bibliografía

[1] Matemática IV medio, texto del estudiante, *primera edición 2014*, Editorial Santillana. *Gabriel Muñoz Zolotoochin, Viviana Gutiérrez Moris, Sergio Muños Venegas.*

[2] Matemática 2° educación media, *Edición Bicentenario*, Editorial Santillana (2011).