

Libro 3: Variables aleatorias continuas

Sitio: Agencia de Habilidades para el Futuro

Curso: Estadística y probabilidades para el desarrollo del soft 1° D

Libro: Libro 3: Variables aleatorias continuas

Imprimido por: Eduardo Moreno

Día: lunes, 5 de mayo de 2025, 15:20

Tabla de contenidos

1. Variables aleatorias continuas

2. Función de densidad de probabilidad

2.1. Introducción

2.2. Definición

3. Función de distribución acumulada

4. Momentos de la variable aleatoria continua

4.1. Esperanza

4.2. Varianza

5. Modelos de variables aleatorias continuas

5.1. Distribución uniforme continua

5.2. Distribución normal

5.3. Distribución gamma y exponencial



Variables aleatorias continuas

En el libro anterior, definimos qué es una variable aleatoria y desarrollamos en profundidad las de tipo discretas. Ahora bien, veamos el siguiente ejemplo:

Sea X la variable aleatoria definida como el tiempo que pasa, en minutos, para que un radar detecte entre conductores sucesivos a los que exceden los límites de velocidad. ¿Qué valores puede tomar la variable aleatoria? ¿Es discreta?

Si pensamos en datos puntuales, podría darse que entre un auto y el siguiente hayan pasado 2 minutos, 0, 5 minutos, 5, 35 minutos, incluso 0 minutos si dos autos son detectados simultáneamente. ¿Qué sucede con los valores de esta variable en comparación a las trabajadas en el libro anterior? Sucede que la variable puede tomar cualquier valor dentro de un intervalo, ($x \geq 0$ en nuestro caso), por lo que no podremos contar su conjunto de resultados posibles.

Esto motiva a definir otro tipo de variables aleatorias que se denominan **continuas**.

Definición

Cuando una variable aleatoria puede tomar cualquier valor dentro de un intervalo, se la denomina **variable aleatoria continua**.

En la mayoría de los problemas prácticos, las variables aleatorias continuas representan datos medidos, como serían todos los posibles pesos, alturas, temperaturas, distancias, tiempo, entre otros.

¿Juego con Genially?

Clasificar las siguientes variables aleatorias como discretas o continuas:

X : el número de accidentes automovilísticos que ocurren al año en Argentina. →discreta

Y : el tiempo para jugar un partido de vóley a tres sets. →continua

P : el número de huevos que una gallina pone mensualmente. →discreta

Q : el peso de las uvas producidas en un viñedo específico. →continua



Primeramente, necesitamos esclarecer algunas ideas:

- **¿Cómo hacemos para presentar, en forma tabular, la distribución de probabilidad de una variable aleatoria continua?**

Se acordarán que las tablas fueron una de nuestras principales representaciones para el caso discreto, ¿podríamos seguir utilizándolas acá? Veamos...

Consideremos una variable aleatoria cuyos valores son las estaturas de todas las personas mayores de 18 años de edad. Entre cualesquiera dos valores, digamos 1,5 y 1,6 metros, hay un número infinito de estaturas, una de las cuales es 1,53 metros. La probabilidad de seleccionar al azar a una persona que tenga exactamente 1,53 metros de estatura en lugar de una del conjunto infinitamente grande de estaturas tan cercanas a 1,53 metros que humanamente no sea posible medir la diferencia es remota, por consiguiente, asignamos una probabilidad 0 a tal evento. Por lo tanto, la probabilidad puntual en este tipo de variables ya carece de sentido, y por consiguiente su representación tabular.

Sin embargo, esto no ocurre si nos referimos a la probabilidad de seleccionar a una persona que mida entre 1,53 y 1,55 metros de estatura. Aquí nos referimos a un intervalo en vez de a un valor puntual de nuestra variable aleatoria.

Nos interesamos por el cálculo de probabilidades para varios intervalos de variables aleatorias continuas como $P(a < X < b)$, $P(Y \geq c)$, entre otras. Observe que cuando X es continua,

$$P(a < X \leq b) = P(a < X < b) + P(X = b) = P(a < X < b)$$

debido a que dijimos que una variable aleatoria continua tiene una probabilidad cero de adoptar exactamente cualquiera de sus valores, por lo que $P(X=b)=0$. Esto significa que no importa si incluimos o no un extremo del intervalo. Sin embargo, esto no es cierto cuando X es discreta.

- **La representación algebraica, ¿sí podemos utilizarla?**

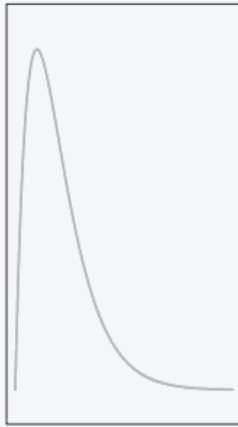
Aunque la distribución de probabilidad de una variable aleatoria continua no se puede representar de forma tabular, sí es posible representarla mediante una fórmula, la cual necesariamente será función de los valores numéricos de la variable aleatoria continua X , y como tal se representará mediante la notación funcional $f(x)$. Cuando se trata con variables continuas, a $f(x)$ por lo general se le llama **función de densidad de probabilidad**, o simplemente **función de densidad de X** .

Obviamente, todo lo trabajado sobre funciones en EAM es aplicable a esta parte de la Teoría de Probabilidades.

- **¿Qué sucede con su representación gráfica?**

Como recordarán, en el libro 2 vimos que las variables aleatorias pueden representarse mediante histogramas. Mientras que aquí, la gráfica será una línea recta (no vertical) o una curva.

La mayoría de las funciones de densidad que tienen aplicaciones prácticas en el análisis de datos probabilísticos son continuas y sus gráficas pueden tomar cualesquiera de varias formas. A manera ilustrativa presentamos las siguientes:



(a)



(b)



(c)

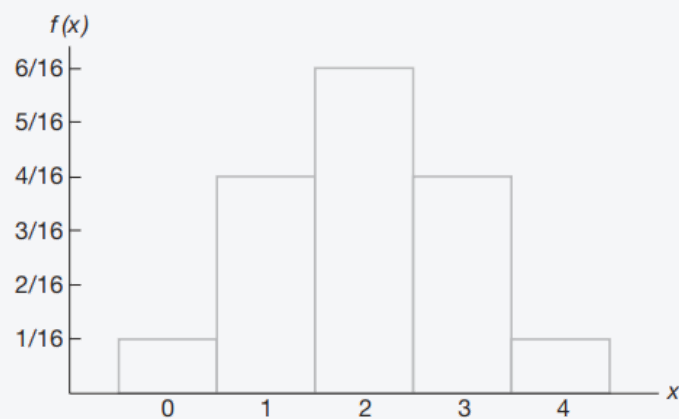


(d)

- **El cálculo de probabilidades como el área de una región.**

Por último, necesitamos retomar la idea que comentamos en las variables aleatorias discretas, sobre que el cálculo de la probabilidad para un valor de la variable, se podía pensar como el cálculo del área del rectángulo que se veía en el histograma. Esto nos va a permitir entender lo que sigue.

Por ejemplo, en el siguiente histograma:



la probabilidad del valor $x = 2$ se puede calcular encontrando el área de su rectángulo correspondiente, que tiene de base 1 unidad y de altura $\frac{6}{16}$. Además, si sumamos todas las áreas, el resultado debe ser 1 como ya se dijo.



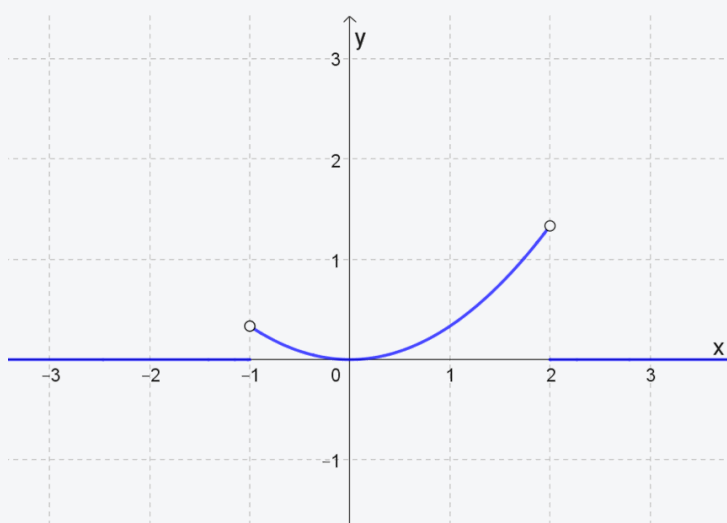
Actividad inicial para "Función de densidad de probabilidad"

Suponga que el error en la temperatura de reacción, en °C, en un experimento de laboratorio controlado, es una variable aleatoria continua (X) que tiene la función de densidad de probabilidad

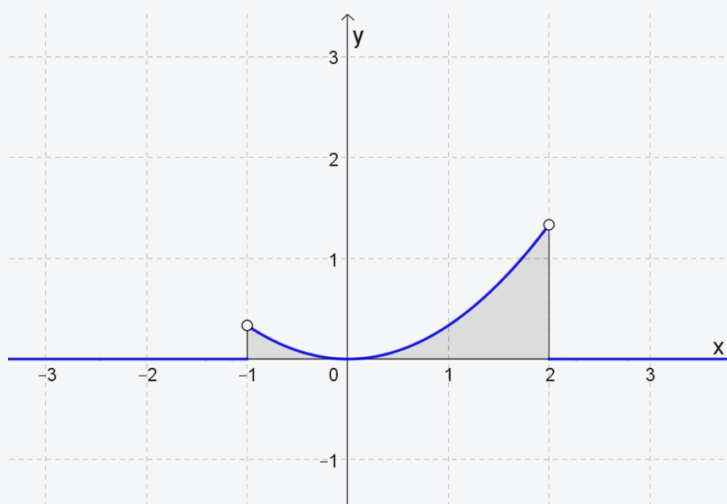
$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2 \\ 0 & \text{en otro caso} \end{cases}$$

Ya vamos a definir luego las condiciones que deben cumplirse para que sea una función de densidad.

Lo primero que haremos será representar gráficamente nuestra función $f(x)$ como lo hicimos en EAM. Dicho gráfico nos queda:



Según las ideas que estuvimos construyendo antes, la probabilidad coincide con el área, en este caso, con el que está debajo de la curva que corresponde a la función $f(x)$. En la gráfica se indica con gris.



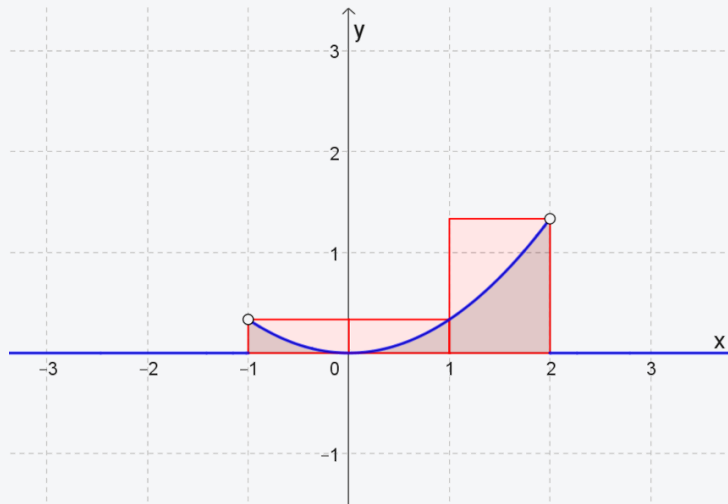
También, por lo que dijimos en libros anteriores, en la gráfica vemos a todos los valores posibles de la variable aleatoria. Por lo tanto, ese área debe ser igual a (1) . Pero... ¿cómo hacemos para calcular ese área debajo de la curva?

Cálculo de áreas de Formas irregulares

El estudio sobre el cálculo de áreas de figuras irregulares como la nuestra, ya movilizó a los matemáticos hace mucho tiempo. La idea de base era dividir un problema complejo en otros más sencillos. Así fue como surgió el método griego del "agotamiento". Por ejemplo, para calcular el área de un círculo, lo que hicieron fue insertar triángulos cada vez más pequeños, y transformar el área del círculo como la suma de las áreas de esos triángulos, sabiendo que, cuántos más triángulos se añadieran, más exacto sería el cálculo del área.

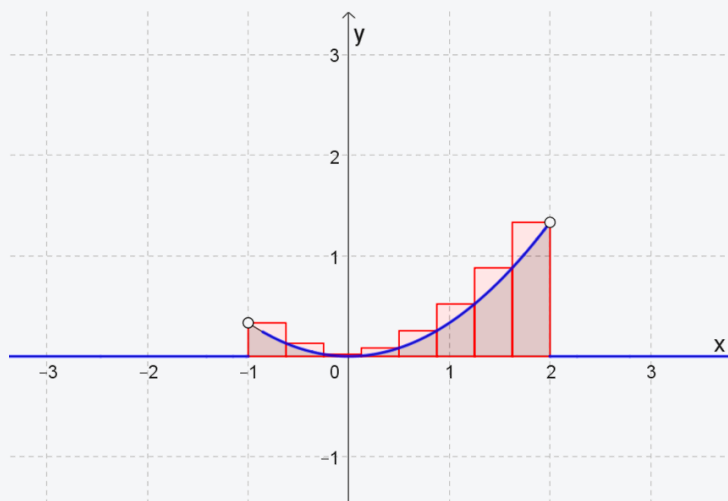
La idea griega se trasladó al cálculo de áreas de otras figuras irregulares: el área de una figura compleja se calcularía intentado transformar la figura irregular en formas regulares.

Entonces, volviendo a nuestro cálculo del área. Podríamos realizar la siguiente construcción:

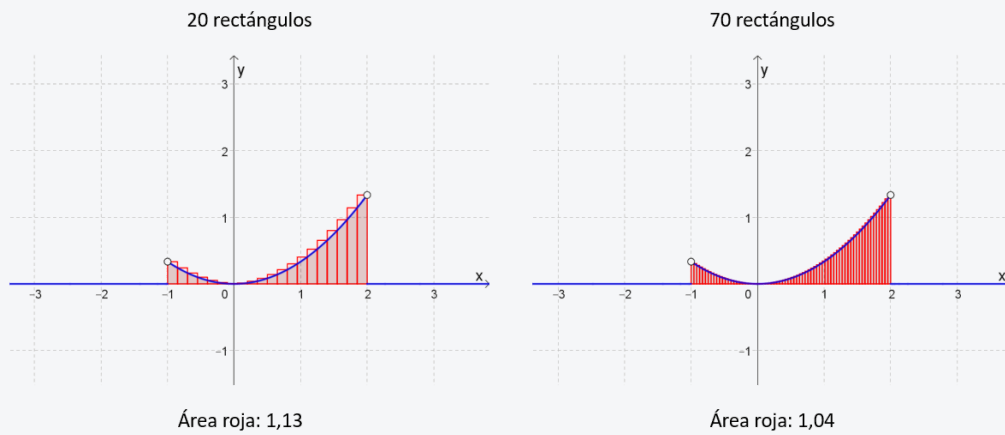


Hemos dividido el intervalo entre (-1) y (2) en tres subintervalos de una unidad cada uno. En cada uno de ellos, construimos rectángulos cuya altura coincida con el valor máximo que toma la función en ese subintervalo (también se podría hacer con el mínimo).

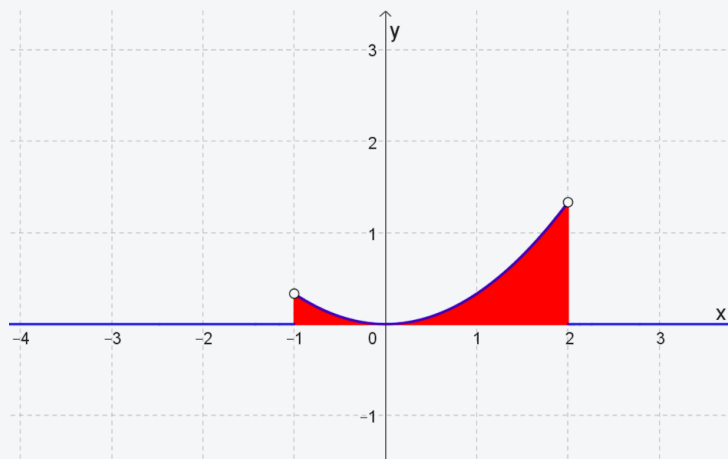
Para calcular el área roja, la encontramos sumando los tres productos entre la base y altura (ambos valores conocidos) de cada rectángulo. Si hacemos estos cálculos, van a encontrar que el resultado les da (2) . Sin embargo, habíamos dicho que tenía que darnos (1) , esto se debe a que claramente el área roja supera el área gris. ¿Cómo solucionamos esto? Dividiendo el intervalo en más subintervalos y generando así, más rectángulos. Por ejemplo, si dividimos en (8) nos queda:



El cálculo del área roja ahora nos da $(1,33)$. Observen cómo hemos mejorado la aproximación, esto se debe a que hemos reducido el "sobrante". Como podrán advertir, podríamos seguir agregando más y más rectángulos...



Ya con 501 rectángulos, el software marca que el área roja es igual a 1 y queda casi perfectamente comprendida debajo de la curva:



¿Qué nos dice esto? Que cuando la cantidad de rectángulos que construyamos tiende al infinito, la suma de las áreas de esos infinitos rectángulos, coincidirá con el área bajo la curva deseado. Este concepto recibe el nombre de **integral definida**. Para nuestro caso, lo anotamos de la siguiente manera:

$$\int_{-1}^2 f(x) \, dx = 1$$

Por lo tanto, la integral definida es la suma de las áreas de un conjunto de rectángulos cuyas alturas vienen dadas por los valores de una función ($f(x_i)$) y cuyas bases, tienen longitudes infinitesimales (Δx_i). Por lo tanto, en símbolos nos queda:

$$\int_a^b f(x) \, dx = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(x_i) \Delta x_i$$

¿Cómo vamos a calcular las integrales?

Utilizaremos calculadora científica o algún software específico como el GeoGebra o el Symbolab. Esto lo iremos explicando cuando sea oportuno.



Definición

A partir de lo que vimos antes, estamos en condiciones de definir la función de densidad de probabilidad.

Definición

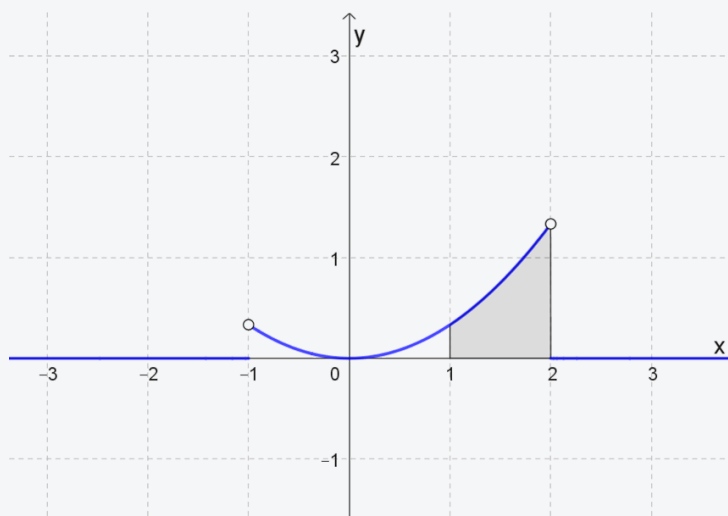
La función $f(x)$ es una función de densidad de probabilidad para la variable aleatoria continua (X) , definida en el conjunto de números reales, si

1. $f(x) \geq 0$, para toda $x \in \mathbb{R}$
2. $\int_{-\infty}^{\infty} f(x) \, dx = 1$
3. $P(a < X < b) = \int_a^b f(x) \, dx$

Si retomamos el ejemplo anterior, vemos que las dos primeras condiciones las cumple para que sea una función de densidad de probabilidad. Noten que la parte que teníamos función cero, no fue de nuestro interés, pues allí no tenemos área bajo la curva.

¿Cómo calculamos $P(1 < X < 2)$?

De acuerdo a lo que vimos antes, deberíamos calcular el área bajo la curva de la función $f(x)$, pero esta vez solamente entre 1 y 2 , es decir:



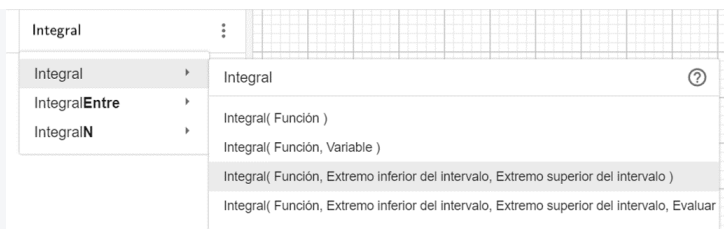
Usando el concepto de la integral definida, nos queda:

$$P(1 < X < 2) = \int_1^2 f(x) \, dx = \frac{7}{9}$$

¿Cómo calcular probabilidades de variables aleatorias continuas utilizando el GeoGebra?

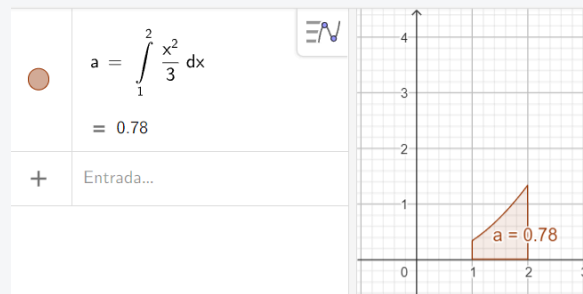
Supongamos que queremos hacer el cálculo anterior en GeoGebra.

1. Abrimos GeoGebra: <https://www.geogebra.org/classic>
2. En la parte de entrada, escribimos la función de densidad de probabilidad.
3. También en la parte de entrada, anotamos: "integral". Se nos va a desplegar un menú y seleccionamos lo siguiente:



4. Rellamos los espacios con la función correspondiente (utilizar la letra asignada por GeoGebra) y los extremos, que son los valores entre los que está la variable aleatoria (X) . En caso de ser necesario utilizar el infinito, escribir "inf" o "-inf", según corresponda.

Para nuestro ejemplo, nos queda de la siguiente manera:



El resultado que nos muestra es el área debajo de la curva de $(f(x))$, entre los valores (1) y (2) de (X) , es decir, $(P(1 < X < 2))$.



Función de distribución acumulada

Ya vimos en detalle en el libro 2, a qué se le llama función de distribución acumulada, pero particularizamos únicamente en las variables aleatorias discretas. Ahora extendemos el concepto también para las continuas.

Definición

La función de distribución acumulada $F(x)$, de una variable aleatoria continua (X) con función de densidad $f(x)$, es

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt, \text{ para } (-\infty < x < \infty)$$

Ejemplo

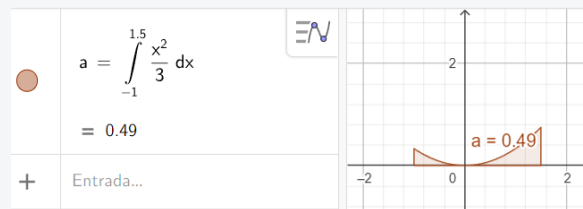
Tomando la función de nuestro ejemplo anterior, supongan que ahora nos piden calcular la probabilidad de que como máximo el error en la temperatura de reacción debe ser de $(1,5)$. Por lo tanto, debemos encontrar $P(X \leq 1,5)$, es decir, $F(1,5)$.

Recordemos que:

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2 \\ 0 & \text{en otro caso} \end{cases}$$

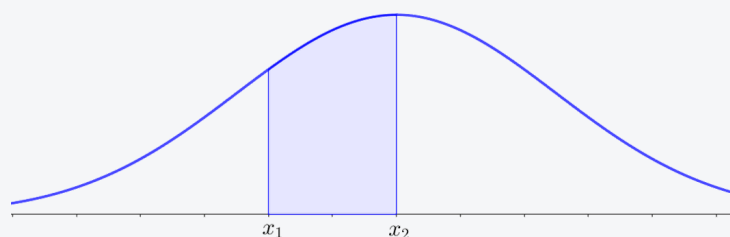
Luego, el cálculo de la función de distribución acumulada nos queda:

$$F(1,5) = P(X \leq 1,5) = \int_{-\infty}^{-1} 0 dx + \int_{-1}^{1,5} \frac{x^2}{3} dx = 0 + 0,49 = 0,49$$

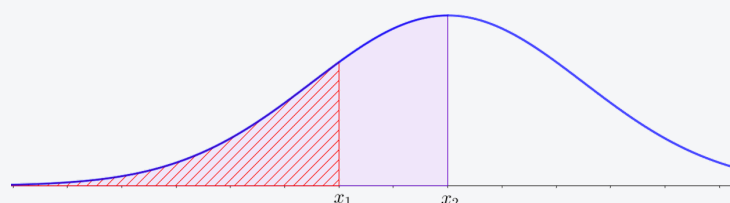


Búsqueda de relaciones entre áreas debajo de la curva

La siguiente imagen representa $P(x_1 < X < x_2)$:



Ahora bien, esta otra representa las probabilidades acumuladas $F(x_1)$ y $F(x_2)$:



donde $F(x_1)$ está diferenciado con rayas rojas y $F(x_2)$, de color violeta.

Si comparamos ambas imágenes, vemos que si a $F(x_2)$ le quitamos $F(x_1)$, obtenemos $P(x_1 < X < x_2)$, es decir:

$$P(x_1 < X < x_2) = F(x_2) - F(x_1)$$

Otra consecuencia inmediata de la definición de función de densidad acumulada es:

$$f(x) = \frac{dF(x)}{dx},$$

si existe la derivada.



Momentos de la variable aleatoria continua

Vamos a extender las definiciones dadas para las variables aleatorias discretas, pero ahora para las continuas. Además, resolveremos algunos ejercicios a modo de ejemplo.

En caso de ser necesario, retomar las conceptualizaciones dadas en el libro 2.



Esperanza

Definición

Sea (X) una variable aleatoria continua con distribución de probabilidad $(f(x))$. La media o valor esperado de (X) es

$$(\mu = E(X) = \int_{-\infty}^{\infty} xf(x) \, dx)$$

Ejemplo

Sea (X) la variable aleatoria que denota la vida en horas de cierto dispositivo electrónico. La función de densidad de probabilidad es

$$(f(x) = \begin{cases} \frac{20000}{x^3}, & x > 100, \\ 0, & \text{en otro caso} \end{cases})$$

Calcular la vida esperada para esta clase de dispositivo.

Tomando la definición de valor esperado para una variable aleatoria continua, tenemos:

$$(\mu = E(X) = \int_{100}^{\infty} x \cdot \frac{20000}{x^3} \, dx = 200)$$

Este cálculo lo podemos hacer en GeoGebra o en Symbolab, aquí se deja la utilización de este último a modo de ejemplo:

$$\int_{100}^{\infty} \left(x \cdot \frac{20000}{x^3} \right) dx$$

Pasos Ejemplos

$$\int_{100}^{\infty} \left(x \cdot \frac{20000}{x^3} \right) dx$$

Solución

200

Por lo tanto, esperamos que este tipo de dispositivo dure en promedio 200 horas.

Teorema

Sea (X) una variable aleatoria continua con distribución de probabilidad $(f(x))$. El valor esperado de la variable aleatoria $(g(X))$ es

$$(\mu_{g(X)} = E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x) \, dx)$$

Ejemplo

Sea (X) una variable aleatoria con función de densidad

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{en otro caso} \end{cases}$$

Calcular el valor esperado de $(g(X) = 4X + 3)$.

Según este teorema, tenemos:

$$\mu_{g(X)} = E[g(X)] = \int_{-1}^2 (4x+3) \cdot \frac{x^2}{3} \, dx = 8$$



Varianza

Sea (X) una variable aleatoria continua con distribución de probabilidad $(f(x))$ y media (μ) . La varianza de (X) es

$$\sigma^2 = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

Al igual que antes, la raíz cuadrada positiva de la varianza, (σ) , se llama **desviación estándar de (X)** .

Ejemplo

La demanda semanal de una bebida para una cadena local de vinotecas, en miles de litros, es una variable aleatoria continua (X) que tiene la siguiente densidad de probabilidad:

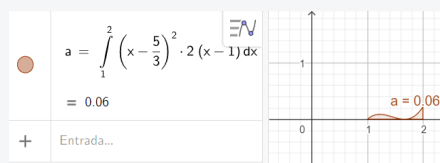
$$\begin{cases} 2(x-1), & 1 < x < 2, \\ 0, & \text{en otro caso} \end{cases}$$

Vamos a calcular la media y la varianza de (X) , como se muestra a continuación:

$$\mu = E(X) = \int_1^2 x \cdot 2(x-1) dx = \frac{5}{3}$$



$$\sigma^2 = E[(X - \frac{5}{3})^2] = \int_1^2 (x - \frac{5}{3})^2 \cdot 2(x-1) dx = \frac{1}{18}$$



Hasta el momento la varianza o la desviación estándar solo tienen significado cuando comparamos dos o más distribuciones que tienen las mismas unidades de medida. Por lo tanto, podemos comparar las varianzas de las distribuciones de contenido, medido en litros, de botellas de jugo de naranja de dos empresas, y el valor más grande indicaría la empresa cuyo producto es más variable o menos uniforme.

Extenderemos ahora nuestro concepto de varianza de una variable aleatoria (X) para incluir también variables aleatorias relacionadas con (X) , como se hizo para el caso discreto.

Teorema

Sea (X) una variable aleatoria continua con distribución de probabilidad $(f(x))$. La varianza de la variable aleatoria $(g(X))$ es

$$\sigma^2_{g(X)} = E\{[g(X) - \mu_{g(X)}]^2\} = \int_{-\infty}^{\infty} [g(x) - \mu_{g(X)}]^2 f(x) dx$$

Ejemplo

Sea (X) una variable aleatoria con función de densidad

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{en otro caso} \end{cases}$$

Calcular la varianza de $g(X) = 4X + 3$.

Anteriormente habíamos calculado que $\mu_{g(X)} = 8$. Entonces:

$$\sigma^2_{4X+3} = E\left\{[(4X+3)-8]^2\right\} = E[(4X-5)^2]$$

$$\sigma^2_{4X+3} = \int_{-1}^2 (4x-5)^2 \cdot \frac{x^2}{3} \, dx = \frac{51}{5}$$

$$\begin{aligned} a &= \int_{-1}^2 (4x-5)^2 \cdot \frac{x^2}{3} \, dx \\ &= 10.2 \end{aligned}$$



Modelos de variables aleatorias continuas

Al igual que hicimos con las variables aleatorias discretas, con las continuas también podemos identificar experimentos aleatorios que tienen el mismo tipo general de comportamiento. En consecuencia, las variables aleatorias continuas asociadas con estos experimentos se pueden describir esencialmente con la misma distribución de probabilidad y, por lo tanto, es posible representarlas usando una sola fórmula.

En esta sección estudiaremos la distribución uniforme, normal, gamma y exponencial.



Distribución uniforme continua

Una de las distribuciones continuas más simples es la **distribución uniforme continua**. Esta distribución se caracteriza por una función de densidad que es constante, por lo cual la probabilidad es la misma en un intervalo cerrado, digamos $[a, b]$.

Definición

La función de densidad de la variable aleatoria uniforme continua (X) en el intervalo $[a, b]$ es

$$f(x; a, b) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b, \\ 0, & \text{en otro caso} \end{cases}$$

Observación: el intervalo no siempre debe ser cerrado, también podría ser (a, b) .

Ejemplo

Se sabe que el peso de las uvas se distribuye siempre de forma uniforme entre (6) y (10) gramos, incluidos estos valores.

Los valores (6) y (10) , corresponden a (a) y (b) en nuestra fórmula, respectivamente. Entonces, construimos la función de densidad de la siguiente manera:

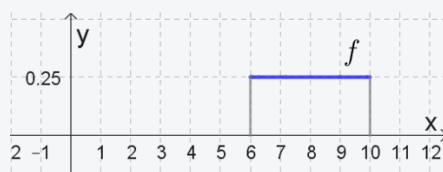
$$f(x; 6, 10) = \begin{cases} \frac{1}{10-6}, & 6 \leq x \leq 10, \\ 0, & \text{en otro caso} \end{cases}$$

Lo que de forma simplificada nos queda:

$$f(x; 6, 10) = \begin{cases} \frac{1}{4}, & 6 \leq x \leq 10, \\ 0, & \text{en otro caso} \end{cases}$$

¿Qué Forma tiene la gráfica de esta Función?

Si graficamos esta función en un sistema de ejes cartesianos, nos queda:



La función de densidad forma un rectángulo con base $(b-a)$ y altura constante $(\frac{1}{b-a})$, en nuestro caso, (4) de base y $(0,25)$ de altura. Como resultado, la distribución uniforme a menudo se conoce como **distribución rectangular**.

Observen que el área de ese rectángulo da como resultado (1) , que coincide con la probabilidad total a la que debemos llegar, ya que tenemos todos los valores de la variable aleatoria continua.

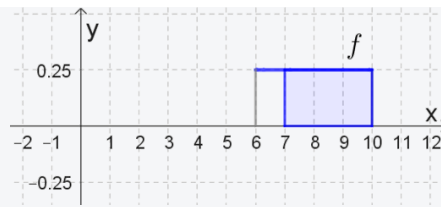
Volviendo al ejemplo de las uvas...

¿Cuál es la probabilidad de que cualquier uva pese al menos (7) gramos?

En este caso debemos calcular la probabilidad de que $(X \geq 7)$, lo cual hacemos como se explicó antes:

$$P(X \geq 7) = \int_7^{10} \frac{1}{4} dx = \frac{3}{4}$$

¿Tiene lógica no? Gráficamente hemos calculado el siguiente área:



Media y varianza

La media y la varianza de la distribución uniforme continua son:

$$\mu = \frac{a+b}{2} \quad \text{y} \quad \sigma^2 = \frac{(b-a)^2}{12}$$

Para nuestro ejemplo, la media y la varianza son:

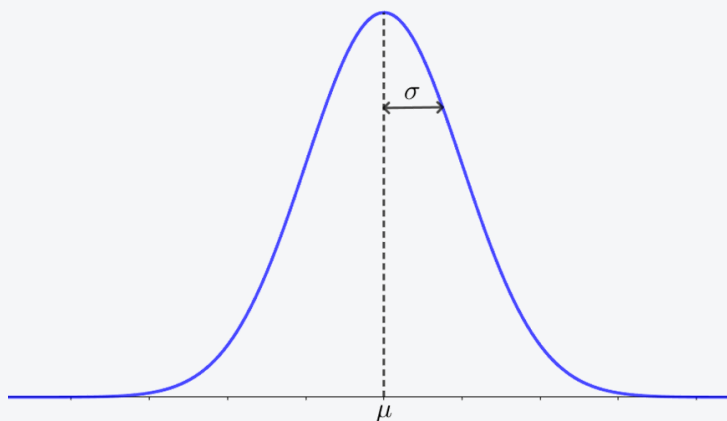
$$\mu = \frac{6+10}{2} = 8 \quad \text{y} \quad \sigma^2 = \frac{(10-6)^2}{12} = \frac{4}{3}$$



Distribución normal

Introducción

La distribución de probabilidad continua más importante en todo el campo de la Probabilidad y la Estadística es sin dudas la **distribución normal**. Su gráfica, denominada **curva normal**, es la curva con forma de campana como la que se muestra a continuación:



la cual describe de manera aproximada muchos fenómenos que ocurren en la naturaleza, la industria y la investigación. Por ejemplo, las mediciones físicas en áreas como los experimentos meteorológicos, estudios de la precipitación pluvial y mediciones de partes fabricadas, a menudo se explican más que adecuadamente con una distribución normal. Además, los errores en las mediciones científicas se aproximan muy bien mediante ella.

Una variable aleatoria continua (X) que tiene la distribución en forma de campana como se vio antes, se denomina **variable aleatoria normal**. La ecuación matemática para la distribución de probabilidad de la variable normal depende de su media y desviación estándar.

Un poco de historia...

En 1733, Abraham DeMoivre desarrolló la ecuación matemática de la curva normal, la cual sentó las bases sobre las que descansa gran parte de la teoría de la estadística inductiva. La distribución normal a menudo se denomina distribución gaussiana en honor a Karl Friedrich Gauss (1777-1855), quien también derivó su ecuación a partir de un estudio de errores en mediciones repetidas de la misma cantidad.

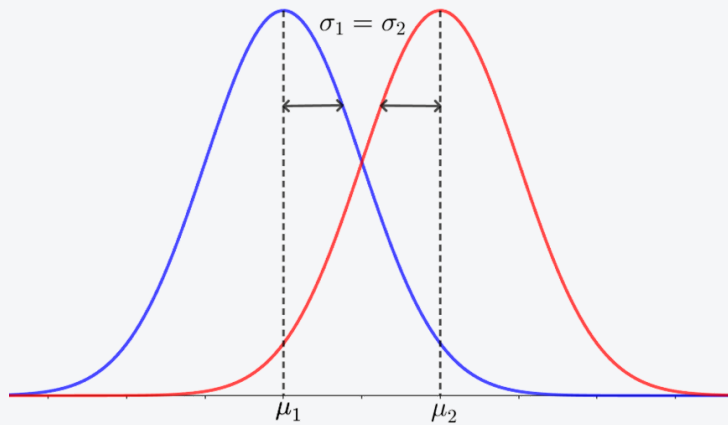
Definición

La densidad de la variable aleatoria normal (X) , con media (μ) y varianza (σ^2) , es

$$f(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, -\infty < x < \infty$$

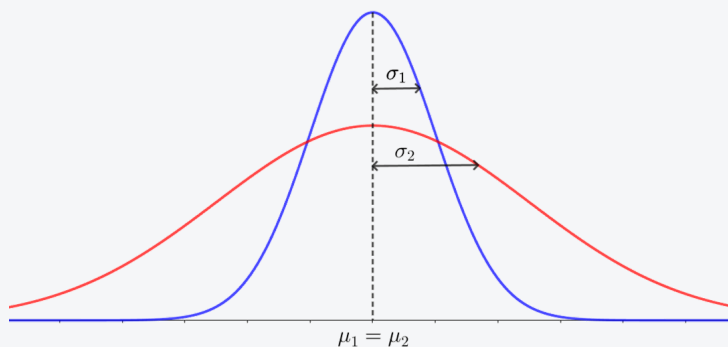
Analicemos algunas gráficas:

- **Misma desviación estándar pero diferentes medias:**



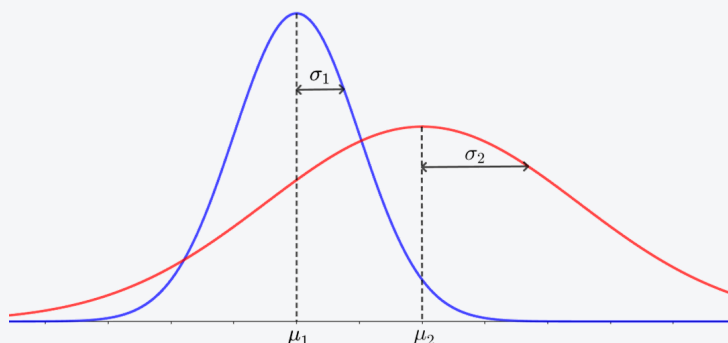
Las dos curvas son idénticas en forma, pero están centradas en diferentes posiciones a lo largo del eje horizontal.

- **Misma media pero con desviaciones estándar diferentes:**



Aquí se observa que las dos curvas están centradas exactamente en la misma posición sobre el eje horizontal, pero la curva con la mayor desviación estándar es más baja y más extendida. Debemos recordar que el área bajo una curva de probabilidad debe ser igual a (1) y, por lo tanto, cuanto más variable sea el conjunto de observaciones, más baja y más ancha será la curva correspondiente.

- **Diferentes medias y diferentes desviaciones estándar.**



Claramente están centradas en posiciones diferentes sobre el eje horizontal y sus formas reflejan los dos valores diferentes de (σ) .

Con base a lo que observamos en las gráficas anteriores, detallamos algunas propiedades de la curva normal:

1. La moda, que es el punto sobre el eje horizontal donde la curva tiene su punto máximo, ocurre en $(x = \mu)$.
2. La curva es simétrica alrededor de un eje vertical a través de la media $(x = \mu)$.

3. La curva normal se aproxima al eje horizontal de manera asintótica, conforme nos alejamos de la media en cualquier dirección.
4. El área total bajo la curva y sobre el eje horizontal es igual a uno.

Media y varianza

La media y la varianza de $n(x, \mu, \sigma)$ son (μ) y (σ^2) , respectivamente. Por lo tanto, la desviación estándar es (σ) .

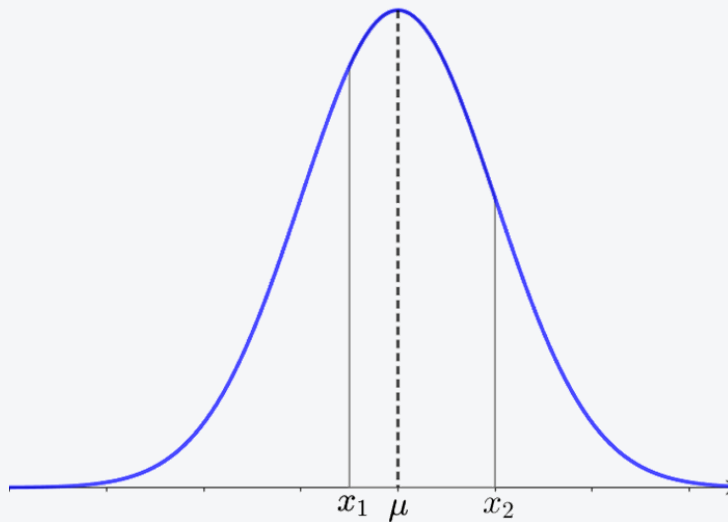
En este libro supondremos que se conocen estos dos parámetros, quizás a partir de investigaciones anteriores.

Área bajo la curva normal

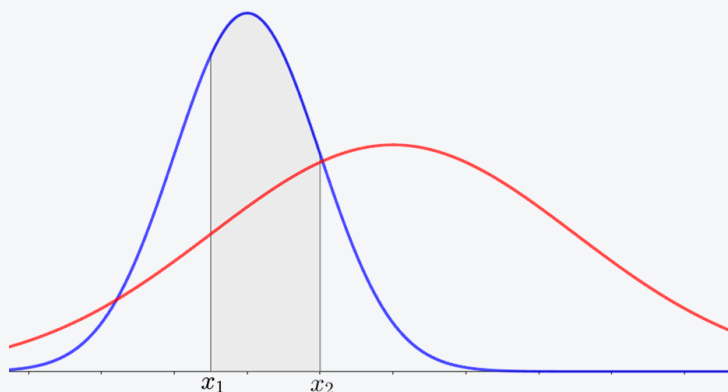
La curva de cualquier función de densidad se construye de manera que el área bajo la curva limitada por (x_1) y (x_2) sea igual a la probabilidad de que la variable aleatoria (X) tome un valor entre (x_1) y (x_2) . Por lo tanto

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} n(x, \mu, \sigma) dx$$

es representada por el área de la región sombreada



En las gráficas anteriores, vimos cómo la curva normal depende de la media y de la desviación estándar de la distribución que se está estudiando. El área bajo la curva entre cualesquiera dos valores de la variable aleatoria, también debe depender de los valores (μ) y (σ) . Esto es evidente en la siguiente imagen, donde sombrearemos las regiones que corresponden a $(P(x_1 < X < x_2))$ para dos curvas con medias y varianzas diferentes.



El cálculo "a mano" de este tipo de distribuciones, exige tabular las áreas de la curva normal para una referencia rápida. Sin embargo, sería inútil tratar de establecer tablas separadas para cada posible valor de (μ) y (σ) . Por fortuna, podemos transformar todas las observaciones de cualquier variable aleatoria normal (X) en un nuevo conjunto de observaciones de una variable aleatoria normal (Z) con media (0) y varianza (1) .

Sin embargo, aquí no realizaremos este tipo de transformaciones, ya que los cálculos los haremos mediante software específicos.

Distribución normal estándar

La distribución de una variable aleatoria normal con media (0) y varianza (1) se llama distribución normal estándar. Tal fórmula nos queda:

$$f(x, 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, -\infty < x < \infty$$

Ejemplo 1

Dada una distribución normal estándar, calcular el área bajo la curva que se localiza:

- a) a la derecha de $(1,84)$ y
- b) entre $(-1,97)$ y $(0,86)$.

Para el apartado (a), debemos calcular $(P(X \geq 1,84) = 1 - P(X < 1,84))$. Es decir:

$$P(X \geq 1,84) = 1 - \int_{-\infty}^{1,84} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = 1 - 0,9671 = 0,0329$$

En cambio, para el ítem (b), el cálculo es:

$$P(-1,97 < X < 0,86) = \int_{-1,97}^{0,86} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = 0,7807$$

Ejemplo 2

Dada una variable aleatoria (X) que tiene una distribución normal con $(\mu = 50)$ y $(\sigma = 10)$. ¿Cuál es la probabilidad de que (X) tome un valor entre (45) y (62) ?

Debemos calcular $(P(45 < X < 62))$, donde $(X \sim n(x, 50, 10))$. Dicho procedimiento se muestra a continuación:

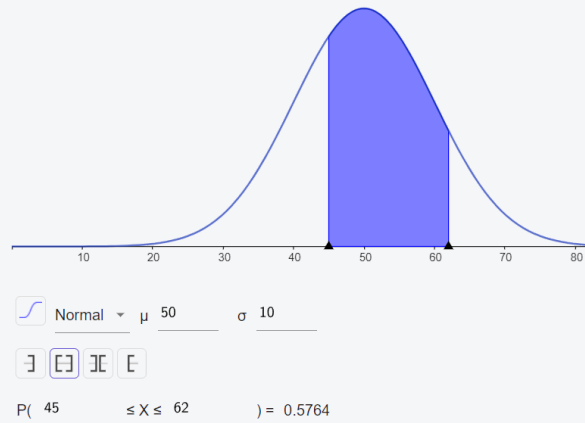
$$P(45 < X < 62) = \int_{45}^{62} \frac{1}{\sqrt{2\pi \cdot 10^2}} e^{-\frac{1}{2 \cdot 10^2}(x - 50)^2} dx = 0,5764$$

¿Cómo calcular probabilidades de variables aleatorias normales utilizando GeoGebra?

Para calcular probabilidades de variables aleatorias normales utilizando GeoGebra seguiremos los siguientes pasos:

1. Abrimos GeoGebra.
2. Hacemos clic en las tres rayitas de la parte superior izquierda. Seleccionamos "apariencias" y luego, "probabilidad".
3. Por defecto aparece la distribución normal, pero podrán ver que hay muchas otras.
4. Seleccionamos los valores para (μ) y (σ) .
5. Luego, los valores para la variable aleatoria (X) . En caso de necesitar el infinito, colocar "-inf" o "inf", según corresponda.

Para nuestro último ejemplo, nos queda:



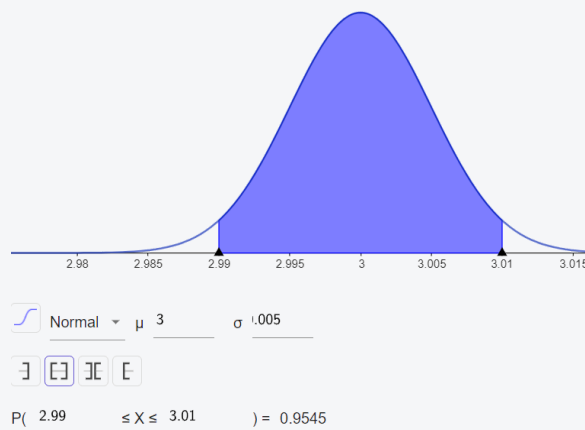
Como observarán, también nos muestra la curva y el área correspondiente.

Ejemplo 3

En un proceso industrial, el diámetro de un rodamiento de bolas es una medida importante. El comprador establece que las especificaciones en el diámetro sean $(3 \pm 0,01)$ cm. Esto implica que no se aceptará ninguna parte que no cumpla estas especificaciones.

Se sabe que en el proceso, el diámetro de un rodamiento tiene una distribución normal con media $(\mu = 3)$ y una desviación estándar $(\sigma = 0,005)$. En promedio, ¿cuántos de los rodamientos fabricados se descartarán?

Si llamamos (X) a la variable aleatoria continua que describe el diámetro de cada rodamiento. Entonces, debemos calcular la probabilidad de que la variable aleatoria (X) tome un valor inferior a los $(2,99)$ $(3 - 0,01)$ o un valor superior a los $(3,01)$ $(3 + 0,01)$, ya que en esos valores, se debería descartar el rodamiento. Es decir, $(P(X < 2,99))$ o $(P(X > 3,01))$, lo cual es equivalente a calcular $(1 - P(2,99 < X < 3,01))$.



Con el GeoGebra, podemos ver que $(P(2,99 < X < 3,01) = 0,9545)$. Por ende, $(1 - P(2,99 < X < 3,01) = 0,0455)$.

Por lo tanto, en promedio se descartarán $(4,55 \%)$ de los rodamientos fabricados.



Distribución gamma y exponencial

Introducción

Aunque la distribución normal se puede utilizar para resolver muchos problemas de ingeniería y ciencias, aún hay numerosas situaciones que requieren diferentes tipos de funciones de densidad, entre las que encontramos a la **distribución gamma** y la **distribución exponencial**.

Resulta que la distribución exponencial es un caso especial de la distribución gamma, y ambas tienen un gran número de aplicaciones. Las dos desempeñan un papel importante en la teoría de las líneas de espera dentro de un sistema y en problemas de confiabilidad. Los tiempos entre llegadas en instalaciones de servicio y los tiempos de operación antes de que partes componentes y sistemas eléctricos empiecen a fallar, a menudo se representan bien mediante la distribución exponencial.

¿De dónde proviene el nombre "gamma"?

La distribución gamma deriva su nombre de la bien conocida función gamma, que se estudia en muchas áreas de la matemática, la cual se define como

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx, \text{ para } (\alpha > 0)$$

Si bien nosotros no trabajamos con esta función en EAM, los cálculos los haremos con software específico.

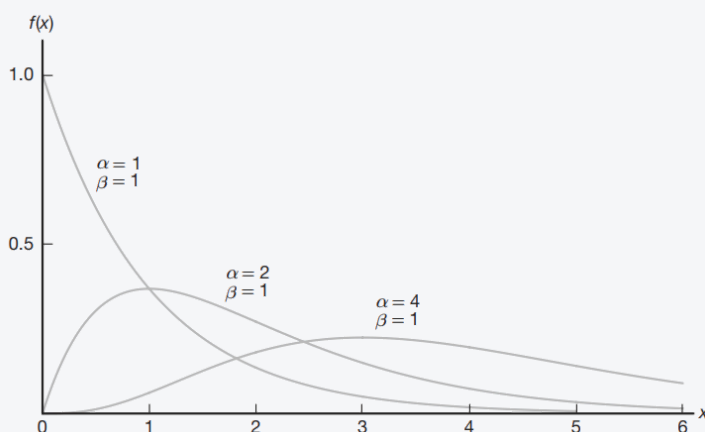
Distribución gamma

La variable aleatoria continua X tiene una distribución gamma, con parámetros α y β , si su función de densidad está dada por

$$f(x, \alpha, \beta) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, & x > 0 \\ 0, & \text{en otro caso} \end{cases}$$

donde $\alpha > 0$ y $\beta > 0$.

En la siguiente imagen se muestran gráficas de varias distribuciones gamma para ciertos valores específicos de los parámetros α y β . La distribución gamma especial para la que $\alpha = 1$ se llama **distribución exponencial**.



Distribución exponencial

La variable aleatoria continua (X) tiene una distribución exponencial, con parámetro (β) , si su función de densidad es dada por

$$f(x, \beta) = \begin{cases} \frac{1}{\beta} e^{-x/\beta}, & x > 0, \\ 0, & \text{en otro caso,} \end{cases}$$

donde $(\beta > 0)$.

Media y varianza

La media y la varianza de la **distribución gamma** son

$$(\mu = \alpha \beta) \quad \text{y} \quad (\sigma^2 = \alpha \beta^2)$$

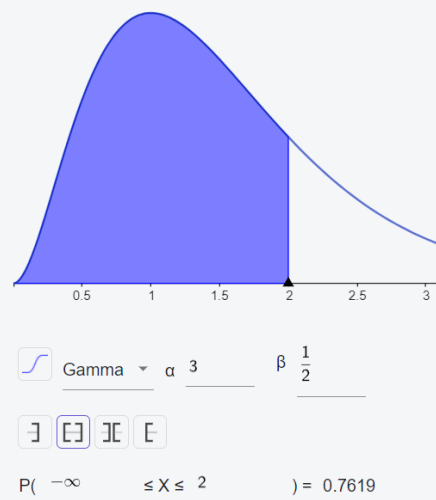
La media y la varianza de la **distribución exponencial** son

$$(\mu = \beta) \quad \text{y} \quad (\sigma^2 = \beta^2)$$

¿Cómo calcular la distribución gamma utilizando GeoGebra?

1. Abrimos GeoGebra.
2. Cambiamos "normal" por "gamma".
3. Seleccionamos los parámetros (α) y (β) deseados, como así también el o los valores de la variable aleatoria (X) .

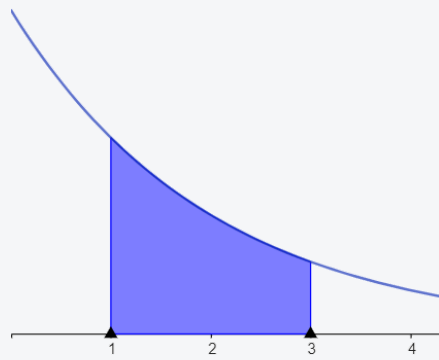
Por ejemplo, $(X \sim f(x, 3, 1/2))$, para calcular $(P(X \leq 2))$ hacemos:



¿Cómo calcular la distribución exponencial utilizando GeoGebra?

1. Abrimos GeoGebra.
2. Seleccionamos "exponencial".
3. Indicamos el parámetro (λ) que equivale a $(1/\beta)$. Además, colocamos el o los valores de la variable aleatoria (X) .

Por ejemplo, $(X \sim f(x, 2))$, para calcular $(P(1 < X < 3))$ hacemos:



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

$P(1 \leq X \leq 3) = 0.3834$

Relación con el proceso de Poisson

Las aplicaciones más importantes de la distribución exponencial son situaciones donde se aplica el proceso de Poisson. Debemos recordar que la distribución de Poisson se utiliza para calcular la probabilidad de números específicos de "eventos" durante un período o espacio particulares. En muchas aplicaciones la variable aleatoria es el tiempo o la cantidad de espacio.

En el libro 2, la distribución de Poisson se desarrolló como una distribución de un solo parámetro (λ), donde λ se interpreta como el número medio de eventos por unidad de "tiempo". Considere ahora la variable aleatoria descrita por el tiempo que se requiere para que ocurra el primer evento. Si utilizamos la distribución de Poisson, vemos que la probabilidad de que no ocurra algún evento, en el periodo hasta el tiempo t , es dada por

$$P(0, \lambda t) = \frac{e^{-\lambda t} (\lambda t)^0}{0!} = e^{-\lambda t}$$

Ahora podemos utilizar lo anterior y hacer que X sea el tiempo para el primer evento de Poisson. La probabilidad de que la duración del tiempo hasta el primer evento exceda x es la misma que la probabilidad de que no ocurra algún evento de Poisson en x . Esto último, por supuesto, es dado por $e^{-\lambda t}$. Como resultado,

$$P(X > x) = e^{-\lambda t}$$

Así, la función de distribución acumulativa para X es dada por

$$P(0 \leq X \leq x) = 1 - e^{-\lambda t}$$

Para poder reconocer la presencia de la distribución exponencial, podemos diferenciar la función de distribución acumulativa anterior con el fin de obtener la función de densidad

$$f(x) = \lambda e^{-\lambda t},$$

que es la función de densidad de la distribución exponencial con $\lambda = \frac{1}{\beta}$.

Aplicaciones

Debemos observar que la media de la distribución exponencial es el parámetro β , el recíproco del parámetro en la distribución de Poisson. Hay que recordar que con frecuencia se dice que la distribución de Poisson no tiene memoria, lo cual implica que las ocurrencias en periodos sucesivos son independientes. El importante parámetro β es el **tiempo promedio entre eventos**. En la teoría de confiabilidad, donde la falla de equipo con frecuencia se ajusta a este proceso de Poisson, β se denomina tiempo medio entre

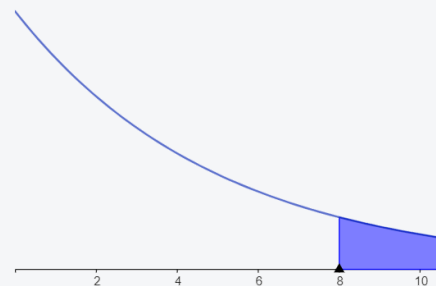
fallas. Muchas roturas de equipo siguen el proceso de Poisson y por ello se aplica la distribución exponencial. Otras aplicaciones incluyen tiempos de supervivencia en experimentos biomédicos y tiempo de respuesta de computadoras.

Ejemplo 1

Suponga que un sistema contiene cierto tipo de componente cuyo tiempo de operación antes de fallar, en años, está dado por (T) . La variable aleatoria (T) se modela bien mediante la distribución exponencial con tiempo medio de operación antes de fallar $(\lambda = 5)$. Si se instalan (5) de estos componentes en diferentes sistemas, ¿cuál es la probabilidad de que al final de (8) años al menos dos aún funcionen?

La probabilidad de que un componente determinado siga funcionando después de 8 años es dada por

$$P(T > 8) = \int_8^{\infty} \frac{1}{5} e^{-t/5} dt \approx 0,2$$



Exponencial $\lambda = \frac{1}{5}$

$P(8 \leq X \leq \infty) = 0.2019$

Luego, representamos con (X) el número de componentes que todavía funcionan después de 8 años. Entonces, utilizando la distribución binomial tenemos:

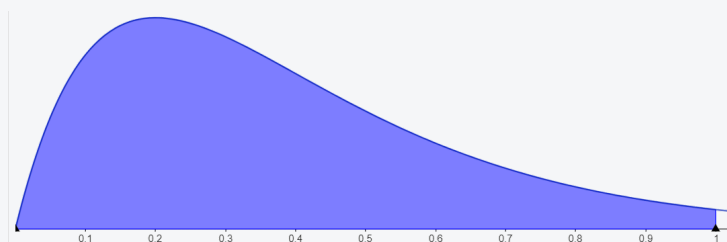
$$P(X \geq 2) = \sum_{x=2}^5 b(x; 5, 0,2) = 0,2627$$

Ejemplo 2

Supongamos que las llamadas telefónicas que llegan a un conmutador particular siguen un proceso de Poisson con un promedio de (5) llamadas entrantes por minuto. ¿Cuál es la probabilidad de que transcurra hasta un minuto en el momento en que han entrado (2) llamadas al conmutador?

Se aplica el proceso de Poisson, con un lapso de tiempo hasta que ocurren (2) eventos de Poisson que sigue una distribución gamma con $(\lambda = 1/5)$ y $(\alpha = 2)$. Denotaremos con (X) el tiempo en minutos que transcurre antes de que lleguen (2) llamadas. La probabilidad que se requiere está dada por

$$P(X \leq 1) = \int_0^1 \frac{1}{2} \left(\frac{1}{5}\right)^2 x e^{-5x} dx = 0,96$$



Gamma $\alpha = 2$ $\beta = \frac{1}{5}$

$P(0 \leq X \leq 1) = 0.9596$

