3ª Emisión

Módulo 03

Probabilidad

Dr. Roberto Bárcenas Curtis

DGTIC

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
Dirección General de Cómputo y de Tecnologías de información y Comunicación
Dirección de Docencia en TIC

50 AÑOS
Educación Continua
1971 - 2021

Presentación

En el mundo, los eventos que ocurren se estudian cada día con mayor frecuencia en términos probabilísticos (inciertos), más que deterministas (predecibles).

La **Probabilidad** es la rama de las Matemáticas que estudia las expectativas de que un evento o fenómeno determinado ocurra.

Objetivos

- El participante identificará los principios y enfoques de probabilidad. Además, distinguirá entre un fenómeno aleatorio y un fenómeno determinista.
- En particular, se analizará la noción de variables aleatorias. Se introducirán las distribuciones de probabilidad de variables aleatorias discretas y continuas.
- El participante conocerá las funciones de distribución de probabilidad más relevantes.
- Se concluirá con el estudio conciso de las funciones de distribución de probabilidad para el caso de vectores aleatorios.

Contenido

¿Para qué sirve el estudio de la Probabilidad?

El conocimiento de la teoría de probabilidad y cálculo de probabilidades es de suma importancia, dado que nos proporcionan las reglas para el estudio de los experimentos aleatorios o de azar, que constituyen elementos esenciales para el estudio de la Estadística.

Los modelos son una herramienta esencial en el estudio del comportamiento de diversos fenómenos. Un modelo matemático es una representación simbólica de un evento o fenómeno

Modelos

Determinista

Se pueden representar y operar los factores que intervienen y se controlan sus resultados.

Estocástico

Cuando no se pueden controlar los resultados del fenómeno o experimento

Ejemplos

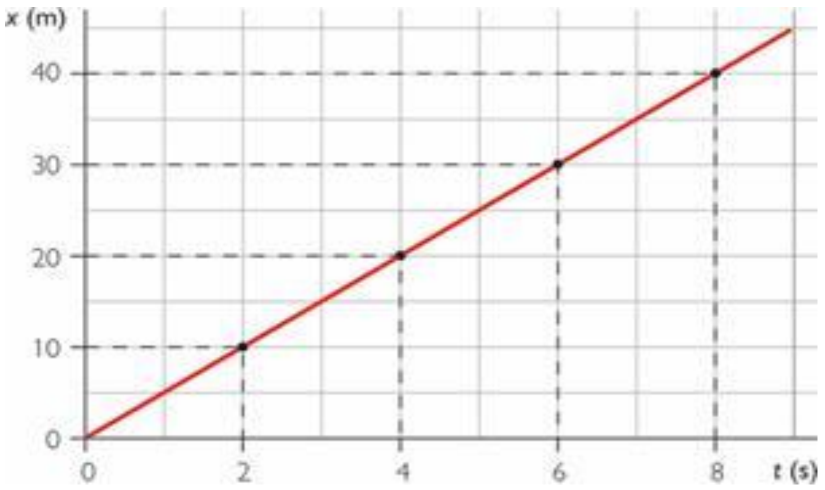
Fenómeno aleatorio

Cualquier acción o proceso que puede producir diferentes observaciones (mediciones), aun cuando se repita siempre de la misma manera. Es un acontecimiento del que no se sabe el resultado que se obtendrá. Está relacionado con el azar.

La **Probabilidad** estudia los fenómenos aleatorios.

Fenómeno determinista

Es el acontecimiento en el cual, de antemano, se sabe cuál será el resultado si las condiciones no cambian.



Movimiento Rectilineo Uniforme

Enfoques probabilidad

Clásico: basado en ciertas simetrías y conteo de posibilidades.

$$P(A) = \frac{n(A)}{n(S)}$$

Frecuentista: considera las frecuencias relativas observadas de eventos repetibles bajo condiciones similares.

Subjetiva: reflejado en juicios personales, criterio y experiencia de sucesos, que pueden justificarse o no. Por ejemplo: “Hoy el cielo está nublado, *creo* que hay un 40% de probabilidad de lluvia”.

Interpretación frecuentista de la probabilidad

Si un experimento aleatorio se realiza *n* veces bajo las mismas condiciones, y *m* veces los resultados son favorables al evento *A.* La probabilidad de que ocurra el evento *A* al realizar nuevamente el experimento, es:

$$P(A) = \lim_{n \to \infty} \frac{m}{n}$$

Experimento

El análisis de la probabilidad se fundamenta en los experimentos, los cuales se definen como:

“*Es cualquier proceso que conduce a algún resultado posible*”

Experimento aleatorio es aquel en el que una vez definidas todas las condiciones bajo las que se realiza, su resultado no queda únicamente determinado.

Álvarez, M. Á. G. (2015). Introducción a la teoría de la probabilidad I. Primer curso. Fondo de Cultura Económica.

Espacio muestral

Al conjunto de todos los posibles resultados de un experimento, lo llamamos espacio muestral del experimento y a cada elemento del espacio muestral, le llamamos punto muestral.

Evento: es una propiedad relativa a un experimento aleatorio (el resultado de un experimento aleatorio).

Si el evento está compuesto por un único elemento, le llamaremos Evento simple.

Si el evento no tiene ningún resultado posible, se le denomina evento nulo o vacío.

El espacio de eventos puede ser finito o infinito y, a su vez, discreto o continuo.

Axiomas de probabilidad

1. Para todo evento A, se tiene que $P(A) \geq 0$. Es decir, toda probabilidad representa a un número positivo.

2. La probabilidad del espacio muestral Ω es igual a 1. Así $P(\Omega) = 1$.

Implicación

Para todo evento A, se cumple la condición $0 \leq P(A) \leq 1$.

Eventos mutuamente excluyentes

Si se tienen dos o más eventos que pertenecen al mismo espacio muestral, en la realización del experimento u ocurrencia del fenómeno, puede ocurrir uno u otro, pero no los dos simultáneamente.

Dos eventos A y B son mutuamente excluyentes o también llamados disjuntos si su intersección es el conjunto vacío, es decir, si $A \cap B = \phi$.

Axiomas de probabilidad

3. Si $A_n \in F$ para $n = 1,2,...$ son eventos disjuntos dos a dos, es decir, tales que $A_n \cap A_m = \varnothing$ si $n \neq m$, entonces

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

Como consecuencia, si *A* y *B* son eventos mutuamente excluyentes, es decir, $A \cap B = \phi$, entonces la probabilidad de que ocurra A o B, escrita como $P(A \cup B)$, es la suma de la probabilidad de A más la probabilidad de B. Así,

$$P(A \cup B) = P(A) + P(B).$$

Regla especial de la adición

Para aplicar esta regla, los eventos deben ser mutuamente excluyentes (cuando un evento ocurre, ninguno de los demás eventos puede ocurrir al mismo tiempo).

Ejemplo: en un experimento de lanzamiento de dados los eventos “un número mayor o igual que 4” y “un número menor que 2”. Si el resultado se encuentra en el primer evento {4, 5 y 6}, entonces no puede estar en el segundo evento {2}.

Otro ejemplo: un producto de la línea de montaje no puede estar defectuoso y en buen estado al mismo tiempo.

Regla especial de la adición

$P(A \cup B) = P(A) + P(B) \rightarrow$ Para dos eventos A y B mutuamente excluyentes

$P(A \cup B \cup C) = P(A) + P(B) + P(C) \rightarrow$ Para tres eventos …

Resultado 1

Sea A un evento. Su evento contrario se denotará como A^c y representa su complemento en el espacio muestral. Entonces la probabilidad de A^c es

$$P(A^c) = 1 - P(A).$$

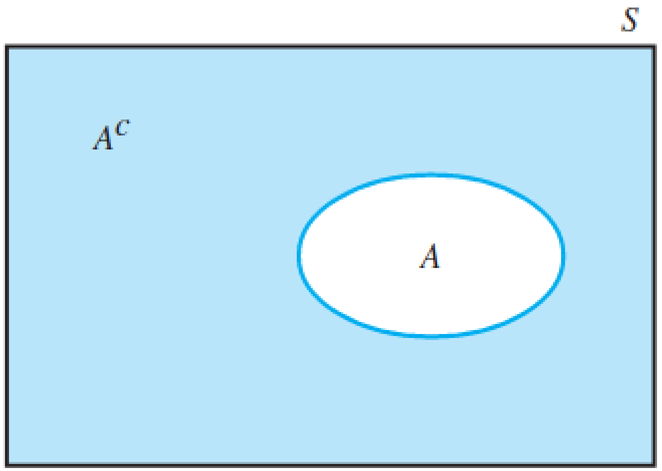
Resultado 2

La probabilidad del evento vacío es cero, es decir, $P(\varnothing) = 0$.

Regla del complemento

Se define como: "la probabilidad de que no ocurra A, es 1 menos la probabilidad de que ocurra A".

$$P(A^C) = 1 - P(A)$$



Resultado 3

Si el evento *A* es un subconjunto de otro evento *B*, es decir, $A \subseteq B$. Entonces $P(A) \leq P(B)$.

Resultado 4

Para todo evento A, se cumple la condición $0 \leq P(A) \leq 1$.

Resultado 5

Para cualesquiera dos eventos *A*, *B* en el espacio muestral se tiene que, la probabilidad de que ocurra *A* o *B* es:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B),$$

donde $P(A \cap B)$ es la probabilidad de que ocurra el evento *A* y *B*.

Ejemplo

Sean *A* y *B* eventos tales que $P(A) = \frac{2}{5}$, $P(B) = \frac{1}{3}$ y $P(A \cup B) = \frac{3}{5}$

¿Son *A* y *B* mutuamente excluyentes?

¿Cuál es el valor de $P(A \cap B)$?

Solución:

Si fueran mutuamente excluyentes se cumpliría que

$P(A \cup B) = P(A) + P(B)$.

Sin embargo, en este caso

$$P(A) + P(B) = \frac{2}{5} + \frac{1}{3} = \frac{6+5}{15} = \frac{11}{15} \neq \frac{3}{5} = P(A \cup B).$$

Por lo tanto, no son mutuamente excluyentes. Ahora para determinar la probabilidad de *A* y *B* se considera la fórmula

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Resolvemos,

$$P(A \cap B) = P(A) + P(B) - P(A \cup B)$$

$$= \frac{2}{5} + \frac{1}{3} - \frac{3}{5} = \frac{6 + 5 - 9}{15}$$

$$= \frac{2}{15}.$$

Entonces la probabilidad de *A* y *B* es igual a $\frac{2}{15}$.

Probabilidad condicional

Considérese el experimento aleatorio de lanzar un dado. Nuevamente, el espacio muestral es $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Sean los eventos

A: cae el número 2

B: cae un número par

Entonces, cómo podemos determinar la probabilidad del evento *A* dado el conocimiento de la ocurrencia del evento *B*. Esto se interpreta como la probabilidad de obtener el número 2 dado que ha caído un número par.

En el ejemplo anterior se observa que el espacio muestral se puede restringir, pues de antemano ya se cuenta con cierta información, es decir, la probabilidad deseada está condicionada a la información disponible.

A este cálculo de probabilidad se le conoce como **probabilidad condicional** de un evento. A continuación, veamos su definición.

Probabilidad condicional

La probabilidad condicional de un evento A dado un evento B, que se denota por $P(A|B)$ está dada por:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

si $P(B) > 0$.

De ejemplo, con $A=\{2\}$ y $B=\{2,4,6\}$. Podemos calcular $A\cap B=\{2\}$

De la fórmula de probabilidad condicional, obtenemos

$P(A\cap B)=\frac{1}{6}$ y $P(B)=\frac{3}{6}$. Así, podemos calcular

$$P(A|B)=\frac{P(A\cap B)}{P(B)}=\frac{\frac{1}{6}}{\frac{3}{6}}=\frac{1}{3}.$$

Independencia

Un evento A se dice que es independiente de otro evento B, si la probabilidad condicional del evento A dado B no depende del evento B, es decir, A es independiente de B si

$$P(A \mid B) = P(A).$$

Si se toma esta definición y se observa la fórmula original de probabilidad condicional

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Podemos reescribirla como

$$P(A) = \frac{P(A \cap B)}{P(B)}$$

De donde, se llega al resultado

$$P(A \cap B) = P(A)P(B)$$

Por lo tanto, una consecuencia de la independencia es que, si los eventos *A* y *B* son independientes, la probabilidad de *A* y *B* es el producto de la probabilidad de *A* y la probabilidad de *B*.

Ejemplo

Diariamente el precio de la acción de una empresa sube con probabilidad 0.50, permanece sin cambios con probabilidad 0.30 y baja con probabilidad de 0.20. Suponiendo que entre cada día las probabilidades se comportan de manera independiente

a) ¿Cuál es la probabilidad de que suba los tres días?

b) ¿Cuál es la probabilidad de que durante tres días el comportamiento sea: baja el primer día, sube el segundo y el tercero permanece igual?

c) ¿Cuál es la probabilidad de que el primer día no baje, el segundo no quede igual y el tercero no suba?

Solución

a) Usamos la independencia entre las probabilidades de cada día. Entonces, para conocer la probabilidad de que suba los tres días, hacemos el producto de las probabilidades

$$(0.50)(0.50)(0.50) = (0.50)^3 = 0.125$$

La probabilidad de que la acción suba consecutivamente los tres días es de 12.5%

b) Suponiendo que entre cada día los eventos son independientes, la probabilidad deseada es el producto de las probabilidades

$$(0.20)(0.50)(0.30) = 0.03$$

La probabilidad es de 3%.

c) Para calcular la probabilidad de que no baje, no suba y no quede igual, respectivamente; usamos la propiedad de eventos complementarios. Así, el producto es

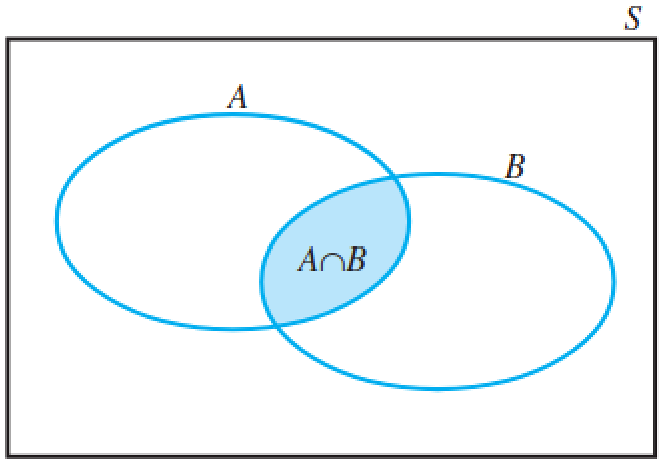
$$(1 - 0.20)(1 - 0.30)(1 - 0.50) = (0.80)(0.70)(0.50) = 0.28.$$

La probabilidad es de 28%.

Probabilidad conjunta

La probabilidad conjunta es un tipo de medida que se encuentra calculando la probabilidad de que dos eventos ocurran juntos. En otras palabras, es la probabilidad de que el evento *X* ocurra al mismo tiempo que el evento *Y*, como una intersección de dos eventos.

$$P(A \cap B) = P(A) * P(B)$$



Regla especial de la multiplicación

La regla especial de la multiplicación implica que dos eventos, A y B, sean independientes; si el hecho de que uno ocurra no altera la probabilidad de que el otro suceda.

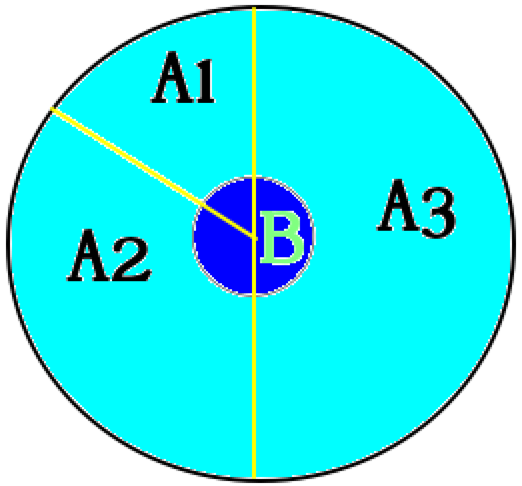
INDEPENDENCIA: si un evento ocurre, no tiene ningún efecto sobre la probabilidad de que otro evento ocurra:

$$P(A \cap B) = P(A) * P(B)$$

Probabilidad total

La regla de la probabilidad total es útil para determinar la probabilidad de un suceso B, en el que su probabilidad depende de otros sucesos (A_1, A_2 y A_3), que son incompatibles dos a dos y cubren todo el espacio muestral. Si en estas condiciones conocemos las probabilidades condicionadas, podemos hallar la probabilidad de B.

Sean A_1, A_2 ... A_n sucesos en el espacio muestral (Ω) excluyentes dos a dos, tales que $\Omega = A_1 \cup A_2 \cup \ldots \cup A_n$. Entonces,

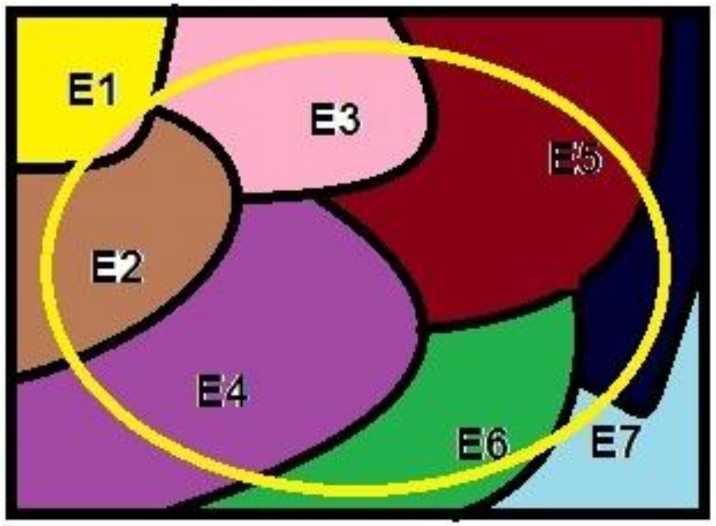


$$P(B) = P(A_1) \cdot P(B|A_1) + P(A_2) \cdot P(B|A_2) + \ldots + P(A_n) \cdot P(B|A_n)$$

Teorema de Bayes

Si $E_1, E_2, \ldots, E_n$ constituyen una partición del espacio muestral Ω y supóngase que la probabilidad del evento E_i es positiva, es decir, $P(E_i) > 0$, para $i = 1, \ldots, n$. Ahora, si A es un evento, es posible escribirlo como la unión de n eventos mutuamente excluyentes, a saber

$$A = (A \cap E_1) \cup (A \cap E_2) \cup \cdots \cup (A \cap E_n).$$



Formula de Bayes

Entonces para cualquier evento A de Ω tal que $P(A) > 0$, la probabilidad "posterior" del evento E_k dado el evento A es

$$P(E_k \mid A) = \frac{P(E_k \cap A)}{P(A)} = \frac{P(A \mid E_k) \cdot P(E_k)}{\sum_{i=1}^{n} P(A \mid E_i) \cdot P(E_i)}$$

Ejemplo

Durante la pandemia, en la Ciudad de México se han registrado 1,170,000 de casos de la enfermedad COVID-19. Supóngase que se tiene una prueba cuya efectividad es de 98% para pacientes enfermos. Es decir, si se tiene la enfermedad la prueba es positiva en el 98% de los casos. En cambio, da un 5% de falsos positivos, esto es, aunque no se tiene la enfermedad, con una probabilidad de 5%, la prueba de cualquier forma es positiva.

Si se considera una población total de 9,000,000, al efectuar una prueba en una persona aleatoriamente, ¿cuál es la probabilidad de que realmente se tenga la enfermedad dado que esta fue positiva?

Ejemplo

Vamos a usar la siguiente notación particular de eventos.

E : *estar enfermo,* E^c : *no estar enfermo,* + : la prueba es positiva.

De esta manera, a partir de la información sabemos que, dado que se tiene la enfermedad, la prueba tiene una probabilidad de 98% de ser positiva, escrita como

$$P(+ \mid E) = 0.98$$

Luego, si no se tiene la enfermedad es positiva con probabilidad de 5%, es decir,

$$P(+ \mid E^c) = 0.05$$

Ejemplo

La probabilidad deseada es que dada la prueba es positiva, se esté enfermo $P(E \mid +)$, la cual puede expresarse a través del Teorema de Bayes:

$$P(E \mid +) = \frac{P(+ \mid E)P(E)}{P(+)}$$

Nótese que aparecen dos probabilidades desconocidas, $P(E)$ y $P(+)$ interpretadas cada una como la probabilidad de estar enfermo y la probabilidad de que la prueba sea positiva.

Ejemplo

En el primer caso, implícitamente conocemos el número personas enfermas y considerando una población total, la probabilidad de tener la enfermedad será la proporción de enfermos respecto al total de la población:

$$P(E) = \frac{1{,}170{,}000}{9{,}000{,}000} = 0.13$$

Ejemplo

Por otra parte, la prueba puede dar positivo en dos casos: cuando se tiene la enfermedad y cuando no se tiene. Se aplica la fórmula de probabilidad total

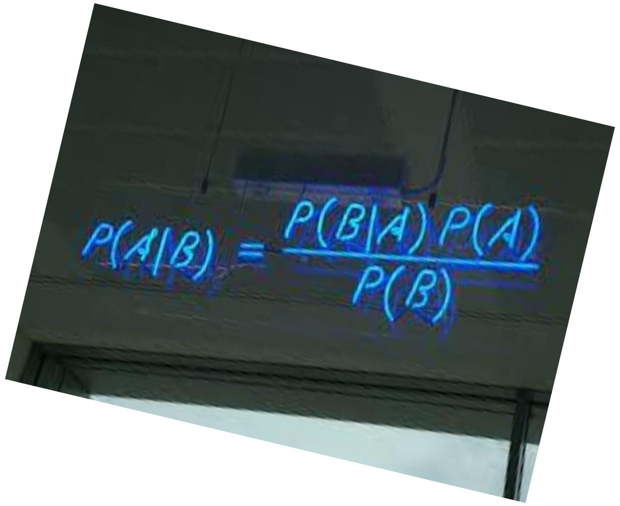
$$P(+) = P(+ \mid E)P(E) + P(+ \mid E^C)P(E^C),$$

en donde, tenemos

$P(E) = 0.13$ y $P(E^c) = 1 - 0.13 = 0.87$

Ejemplo

Por lo tanto, en la fórmula de Bayes se tiene que



$$P(E\,|\,+) = \frac{P(+\,|\,E)P(E)}{P(+)} = \frac{(0.98)(0.13)}{0.1709} = \frac{0.1274}{0.1709} = 0.7455$$

Finalmente, concluimos lo siguiente: Dado que la prueba se ha aplicado en alguien aleatoriamente y es positiva, la probabilidad de que tenga realmente la enfermedad es aproximadamente 75%.

Variable aleatoria

Una variable aleatoria es una función que asocia un número real a los elementos del espacio muestral.

Formalmente, es una función de valores reales definida sobre el espacio muestral Ω, denotada por $X : \Omega \rightarrow \mathbb{R}$, de tal forma que para un experimento aleatorio, si ω es un resultado que pertenece a Ω, al aplicarle la función X se obtiene el número real $X(\omega) = x$.

Ejemplo

Consideremos el experimento de lanzar un volado. Se ha visto que

$\Omega = \{\text{Cara, Cruz}\}$ es el espacio muestral. Entonces, sea la función X tal que asigna el valor 1 si cae *Cara* y asigna el valor 0 si cae *Cruz*. Con la notación anterior

$$X(\text{"}Cara\text{"}) = 1,$$

$$X(\text{"}Cruz\text{"}) = 0.$$

Dada esta relación, se tiene que X es una variable aleatoria.

Distribución de probabilidad

Al cálculo

$$P(X \in B) = P\left(X^{-1}(B)\right) = P\left(\{\omega \in \Omega : \mathrm{X}(\omega) \in B\}\right),$$

se le conoce como distribución o ley de X, también denotada como $P_X(B)$.

En particular, es de interés encontrar la distribución de conjuntos *B en R* de la forma $(-\infty, x]$, como

$$P\left(X \in (-\infty, x]\right),$$

también denotada como

$$P(X \leq x).$$

Esto permite calcular probabilidades de subconjuntos de $\mathbb{R}$. En el ejemplo,

$$P\left(X \in (-\infty, 0]\right) = \frac{1}{2}$$

La formulación anterior permite definir a la función de distribución de la variable aleatoria *X* como

$$F(x) = P(X \leq x) = P\left(X \in (-\infty, x]\right).$$

Tipos de variables

Dada la naturaleza de la función de distribución las **variables aleatorias** dividen en dos clases.

1. **<u>Continuas.</u>** Toman valores reales en algún intervalo $(a, b) \subset \mathbb{R}$

Ejemplos:

- El tiempo de llegada de un origen a un lugar de destino.
- La medición del peso de un grupo de personas.
- El tiempo que tarda una persona en hablar por teléfono en una llamada.
- Los rendimientos de un activo financiero.
- La distancia recorrida por un repartidor diariamente.

2. **<u>Discretas:</u>** Toman valores asociados a los números naturales y son utilizadas en procesos de conteo. Así, una variable aleatoria es discreta cuando el conjunto de valores que toma es un conjunto discreto, es decir, corresponde a un conjunto contable, esto es, finito o numerable.

Ejemplos:

- El número de personas que llegan a una sucursal bancaria.
- El número de autos que vende una agencia automotriz.
- El número de nacimientos o defunciones en una ciudad en un año determinado.
- El número de huracanes en una temporada.
- El número de pacientes recuperados en un tratamiento.

Variables aleatorias discretas

Una variable aleatoria que es discreta, toma valores en un conjunto finito $\{x_1, x_2, \ldots, x_n\}$ o numerable $\{x_1, x_2, \ldots\}$ y existen valores $P(X = x_1), P(X = x_2), \ldots,$ que representan sus respectivamente sus probabilidades, sobre las cuales se define la función de densidad o de probabilidad

$$f(x) = \begin{cases} P(X = x) \\ 0 \end{cases}$$

En este caso f_X es llamada función de probabilidad de la variable aleatoria X.

Función de densidad y distribución

Propiedades

1. $f(x) \geq 0$, para toda $x \in \mathbb{R}$.
2. $\sum_i f(x_i) = 1$, donde la suma se hace sobre todas las $\{x_1, x_2, ..., x_n\}$ o bien $\{x_1, x_2, ...\}$.

Así, su función de distribución está dada por

$$F(x) = \sum_{x \leq u} f(u)$$

Ejemplo

Veamos el caso de distribuciones discretas con probabilidades uniformes. E particular, consideremos una uniforme discreta en el conjunto $\{1, 2, 3, 4\}$, de ta forma que su función de probabilidad es

$$P(X = j) = \frac{1}{4}, \text{ para } j = 1,2,3,4.$$

De acuerdo con esa función de probabilidad, su función de distribución es

$$F(x) = \begin{cases} 0 & \text{si } -\infty < x < 1 \\ \frac{1}{4} & \text{si } 1 \leq x < 2 \\ \frac{1}{2} & \text{si } 2 \leq x < 3 \\ \frac{3}{4} & \text{si } 3 \leq x < 4 \\ 1 & \text{si } x \geq 4 \end{cases}$$

Variable aleatoria discreta

Por ejemplo, si en el grupo de “Probabilidad y estadística” hay 73 estudiantes entonces el número de ausentes a la clase el día lunes sólo puede ser 0, 1, 2, 3, ..., 73. Una variable discreta suele ser el resultado de realizar un conteo.

Variables aleatorias continuas

Una variable aleatoria es continua, si existe una función no negativa e integrable f tal que

- $f(x) \geq 0$ para todo x,
- $\int_{-\infty}^{\infty} f(x) dx = 1$.

La función f es llamada función de densidad de la variable aleatoria X.

Propiedades

Su función de distribución es

- $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du,$ para todo $x \in \mathbb{R}$

Notar que si F es diferenciable, por el Teorema Fundamental del Cálculo se tiene la siguiente relación

$$f(x) = \frac{d}{dx} F(x) = F'(x), \quad -\infty < x < \infty.$$

Propiedades

Es importante aclarar que $f_X(x)$ no tiene la interpretación de una probabilidad. Por lo tanto, no necesariamente sucede que $0 \leq f_x(x) \leq 1$.

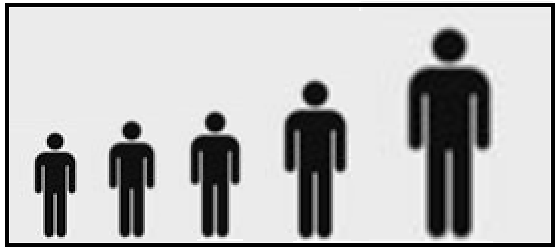
Lo que sí tiene una interpretación probabilística es la densidad de X para en algún intervalo (a, b) de $\mathbb{R}$ en donde

$$P\left(X \in (a, b)\right) = \int_a^b f(x) dx.$$

Esto es, la probabilidad de que X tome un valor dentro del intervalo (a, b) es el área bajo la función de densidad en dicho intervalo.

Interpretación

Por otra parte, si la variable aleatoria es continua, es una distribución de probabilidad continua. Si mide algo, como la anchura de una recámara, la estatura de una persona o la presión de la llanta de un automóvil, se trata de una variable aleatoria continua. Se puede suponer una infinidad de valores, dentro de un intervalo determinado.



Esperanza o media

Se define el valor esperado, esperanza o media de la variable aleatoria *X como*

$$E\left[X\right] = \sum_{j=1}^{\infty} x_j \cdot P\left(X = x_j\right)$$

Para *X* discreta con valores $x_1, x_2, \ldots,$ y

$$E\left[X\right] = \int_{-\infty}^{\infty} x \cdot f\left(x\right) dx$$

Si *X* es continua con función de densidad *f*.

Considérese una variable aleatoria *X* y una función *g* con ciertas propiedades. Puede probarse que *g(X)* también es variable aleatoria y además se define el valor esperado de la función *g(X)* como

$$E\left[g\left(X\right)\right] = \sum_{j=1}^{\infty} g\left(x_j\right) P\left(X = x_j\right)$$

si *X* es discreta con valores $x_j, \; j \geq 1$. Y como

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

si *X* es continua con función de densidad f.

Media

La media constituye un valor típico para representar la localización central de una distribución de probabilidad.

Se trata de un promedio ponderado, en el que los posibles valores de una variable aleatoria se ponderan con sus correspondientes probabilidades de ocurrir. La media de una distribución de probabilidad discreta, se calcula con la fórmula:

$$\boldsymbol{\mu} = \boldsymbol{\Sigma}[xP(x)]$$

Momentos

El k-ésimo momento de una variable aleatoria *X* denotado por μ_k está dado por la cantidad $E\left(X^k\right)$, con $k = 1, 2, \ldots$

Nótese que, si $k = 1$, se tiene el primer momento, el cual corresponde a la esperanza de *X*, es decir, $E\left[X\right] = \mu = \mu_1$.

Para alguna función $g\left(x - \mu_1\right)^k$, se tiene que el k-ésimo momento central de la variable aleatoria *X*, para $k = 1, 2, \ldots$, es

$$E\left[\left(X - \mu_1\right)^k\right].$$

En algunas referencias se denota por μ_k'.

Varianza y desviación estándar

El segundo momento central dado por $\mu_2' = E\left[(X - \mu)^2\right]$ se llama *varianza* de la variable aleatoria *X* y puede denotarse por $Var(X)$ o bien por el símbolo σ_X^2 o únicamente por σ^2.

Mide la dispersión de la distribución de probabilidad de la variable alrededor de la media.

A la raíz cuadrada positiva de la varianza se le llama ***desviación estándar*** y se denota como σ_X o como σ.

Si X es discreta con función de probabilidad $f(x)$ con esperanza finita μ, si existe, su varianza se calcula como

$$Var[X] = \sum_{x} (x - \mu)^2 P(X = x) = \sum_{x} (x - \mu)^2 f(x).$$

O bien, si X es continua con función de densidad $f(x)$ y esperanza finita μ, su varianza es

$$Var(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx .$$

Cálculo de la varianza

La varianza satisface una relación fundamental con su valor esperado

$$Var\left(X\right) = E\left(X^{2}\right) - \left(E\left(X\right)\right)^{2}.$$

Esto permite calcular la varianza de su definición o usando esta fórmula.

Propiedades

Sean *X* y *Y* variables aleatorias y *c* una constante. Entonces

$$Var(X) \geq 0$$

$$Var(X) = E\left(X^2\right) - \left(E(X)\right)^2$$

$$Var(c) = 0$$

$$Var(cX) = c^2 Var(X)$$

$$Var(c + X) = Var(X)$$

Ejemplo

Encontrar el valor de *c* que hace que la función

$$f(x) = \begin{cases} cx & \text{si } x \in (2,4) \\ 0 & \text{e.o.c.} \end{cases}$$

sea una función de densidad.

Solución

Integrando $f(x)$ sobre $(2,4)$, hacemos

$$1 = \int_2^4 cx\,dx = c\int_2^4 x\,dx = c\left.\frac{x^2}{2}\right|_0^2 = c\left(\frac{4^2}{2} - \frac{2^2}{2}\right) = c\left(\frac{16}{4} - \frac{4}{2}\right) = c(6)$$

de donde $c = \frac{1}{6}$.

Por lo tanto, la función de densidad es

$$f(x) = \begin{cases} \frac{1}{6}x & \text{si } x \in (2,4) \\ 0 & e.o.c. \end{cases}$$

Ejemplo

Considere la variable aleatoria *X* cuya función de probabilidad está dada en la tabla.

Determinar el valor de $E[X]$.

x	$f(x) = P(X = x)$
1	$\frac{1}{9}$
2	$\frac{2}{9}$
3	$\frac{3}{9}$
4	$\frac{2}{9}$
5	$\frac{1}{9}$

Solución:

De la definición, se calcula

$$E[X] = \sum_{j=1}^{5} x_j f(x_j) = \sum_{j=1}^{5} x_j P(X = x_j)$$

$$= 1 \cdot \frac{1}{9} + 2 \cdot \frac{2}{9} + 3 \cdot \frac{3}{9} + 4 \cdot \frac{2}{9} + 5 \cdot \frac{1}{9}$$

$$= \frac{1}{9} + \frac{4}{9} + \frac{9}{9} + \frac{8}{9} + \frac{5}{9} = \frac{27}{9} = 3$$

Ejemplo

Sea X una variable aleatoria tal que su densidad es $f(x) = \frac{2}{3}x, 1 < x < 2$. Calcular $Var(X)$.

Solución

La relación $Var(X) = E\left(X^2\right) - \left(E(X)\right)^2$ es útil, pues con un cálculo permite hallar la varianza deseada.

Así,

$$
\begin{aligned}
Var(X) &= E\left(X^2\right) - \left(E(X)\right)^2 \\
&= \int_1^2 x^2 \cdot \frac{2}{3}\, xdx - \left(\int_1^2 x \cdot \frac{2}{3}\, xdx\right)^2 \\
&= \frac{2}{3}\int_1^2 x^3 dx - \left(\int_1^2 \frac{2}{3}\, x^2 dx\right)^2 \\
&= \frac{2}{3} \cdot \left.\frac{x^4}{4}\right|_1^2 - \left(\frac{2}{3} \cdot \left.\frac{x^3}{3}\right|_1^2\right)^2 \\
&= \frac{2}{3} \cdot \frac{15}{4} - \left(\frac{14}{9}\right)^2 \\
&= \frac{31}{12} - \frac{196}{81} = \frac{13}{162}
\end{aligned}
$$

Función generadora de momentos

Dada una variable X , con función de distribución F , se define la *función generadora de momentos* como

$$M_X(t) = E\left(e^{tX}\right)$$

Se cumple

$$M_X'(0) = E(X)$$

$$M_X''(0) = E\left(X^2\right)$$

En general,

$$M_X^{(n)}(0) = E\left(X^n\right)$$

Observaciones

- Su existencia no está garantizada en todas las distribuciones.
- Si existe determina de manera única a la distribución asociada, en tanto se puede utilizar en variables aleatorias discretas y continuas.
- Notar el requisito de su existencia en una vecindad no trivial alrededor de cero.
- Cuando existe y también existe la función generadora de probabilidad, estas se relacionan por la igualdad $M_X(t) = \phi\left(e^t\right)$

Distribuciones Discretas

- Distribución Bernoulli
- Distribución Discreta Uniforme
- Distribución Binomial
- Distribución Geométrica
- Distribución Poisson

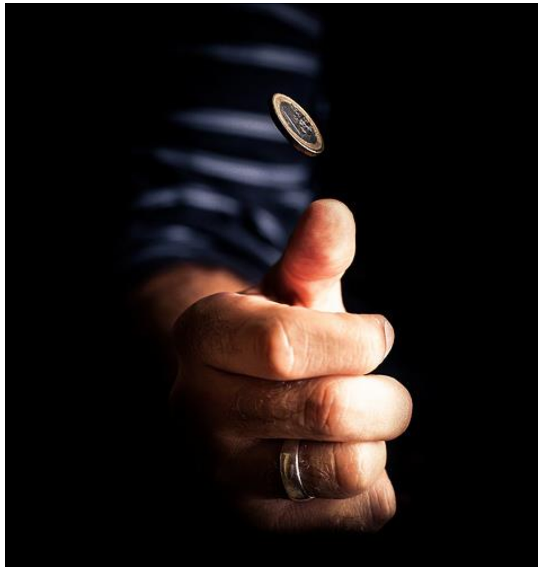
Distribuciones Continuas

- Distribución Uniforme
- Distribución Exponencial
- Distribución Gamma
- Distribución Beta
- Distribución Weibull
- Distribución Normal
- Distribución Lognormal

Distribución Bernoulli

Un ensayo de Bernoulli es un experimento aleatorio con dos resultados posibles. Considérese una variable aleatoria X que puede tomar dos valores, a saber: fracaso y éxito; denotados numéricamente como 0 y 1, y con probabilidades p y $1 - p$, respectivamente. Entonces X tiene distribución Bernoulli con parámetro p y su función de probabilidad es

$$f(x) = p^x (1-p)^{1-x}, \text{ para } x = 0, 1.$$



$$f(0) = p^0 (1-p)^{1-0} = 1 \bullet (1-p)^1 = 1-p$$

$$f(1) = p^1 (1-p)^{1-1} = p^1 \bullet (1-p)^0 = p \cdot 1 = p$$

Distribución de probabilidad binomial

Es una distribución de probabilidad discreta. En la distribución binomial sólo hay dos posibles resultados en determinado intento de un experimento. Con frecuencia, se clasifican los dos posibles resultados como éxito y fracaso.

Otra característica clave de la distribución binomial es el hecho de que la variable aleatoria es el resultado de la suma de variables aleatorias Bernoulli. Es decir, se cuenta el número de éxitos X en un total de n ensayos Bernoulli.

Cálculo de probabilidad binomial

$P(x) = nCx\ p^{x}(1 - p)^{n - x}$

Fórmula de combinaciones

$${}_{n}C_{r} = \frac{n!}{(n-r)!r!}$$

Donde:

C representa una combinación

n es el número de pruebas

x es la variable aleatoria, definida como el número de éxitos

p es la probabilidad de un éxito en cada prueba

Ejemplo

US Airways tiene cinco vuelos diarios de Pittsburgh al Aeropuerto Regional de Bradford, Pennsylvania. Suponga que la probabilidad de que cualquier vuelo llegue a tiempo es de 0.80.

¿Cuál es la probabilidad de que ninguno de los vuelos llegue a tiempo hoy?

¿Cuál es la probabilidad de que exactamente uno de los vuelos llegue a tiempo hoy?

La probabilidad de que un vuelo llegue tarde es de 0.20. Así, $\pi = 0.20$. Hay cinco vuelos. Así, $n = 5$, y X, la variable aleatoria, se refiere al número de éxitos.

En este caso, un éxito consiste en que un avión llegue a tiempo. Como se desea que ninguno llegue a tiempo, se tiene $x = 0$.

Entonces:

P(0) = 5C0 0.200(1 - 0.20)5 - 0 = (1)(1)(0.3277) = 0.3477 → Es la probabilidad de que ningún vuelo llegue tarde

P(1) = 5C1 0.201(1 - 0.20)5 - 1 = (5)(0.20)(0.4096) = 0.4096 → Probabilidad de que al menos un vuelo de los cinco llegue a tiempo

Medial de una distribución binomial

$$\boldsymbol{\mu} = np$$

Varianza de una distribución binomial

$$\boldsymbol{\sigma}^2 = np(1 - p)$$

Ejemplo

Play Time Toys, Inc. tiene 50 empleados en el departamento de ensamble. Cuarenta empleados pertenecen a un sindicato, y diez no. Se eligen al azar cinco empleados para formar un comité que hablará con la empresa sobre los horarios de inicio de los turnos. ¿Cuál es la probabilidad de que cuatro de los cinco empleados elegidos para formar parte del comité, pertenezcan a un sindicato?

Datos:

N = 50 (Total de empleados en el departamento de ensamble)

S = 40 (Empleados sindicalizados)

x = 4 (Empleados sindicalizados elegidos)

n = 5 (Número de empleados elegidos para formar un comité)

P(4) = Probabilidad de que 4 de los 5 miembros del comité sean sindicalizados

Ejemplo

$$P(4) = \frac{\dfrac{40!}{(40-4)!4!} * \dfrac{(50-40)!}{((50-40)-(5-4))!(5-4)!}}{\dfrac{50!}{(50-5)!5!}}$$

$$P(4) = \frac{\dfrac{40!}{36!4!} * \dfrac{10!}{9!1!}}{\dfrac{50!}{45!5!}} = \frac{(91390)\,(10)}{2118760} = 0.431$$

$$_{n}C_{x} = \frac{n!}{(n-x)!\,x!}$$

$$P(x) = \frac{({}_{S}C_{x})({}_{N-S}C_{n-x})}{{}_{N}C_{n}}$$

Esto significa que la probabilidad de elegir al azar a 5 trabajadores de ensamble de los 50 trabajadores y encontrar que 4 de 5 son sindicalizados es de 0.431.

Calcular tabla de probabilidad de 0 a 5 trabajadores sindicalizados en el comité.

Distribución de probabilidad de Poisson

La distribución de probabilidad de Poisson describe el número de veces que se presenta un evento durante un intervalo específico. El intervalo puede ser de tiempo, distancia, área o volumen.

La distribución se basa en dos supuestos:

1. La probabilidad es proporcional a la longitud del intervalo
2. Los intervalos son independientes

Distribución de probabilidad de Poisson

$$P(x) = \frac{\mu^x e^{-\mu}}{x!}$$

Donde:

μ = la media de la cantidad de éxitos que presenta un evento en un intervalo particular. (**μ** = np)

e = 2.71828

x = es el número de veces que se presenta un evento

P(x) = probabilidad para un valor específico de x

Media de una distribución de Poisson

$$\mu = np$$

La varianza de una distribución de Poisson es igual a su media.

En una distribución binomial, existe una cantidad fija de pruebas. Por ejemplo, en una prueba de selección múltiple de cuatro preguntas, sólo puede haber cero, uno, dos, tres o cuatro éxitos (respuestas correctas). Sin embargo, la variable aleatoria, X, para una distribución de Poisson, puede adoptar una infinidad de valores; es decir, 0, 1, 2, 3, 4, 5, ….

Ejemplo

En Northwest Airlines, pocas veces no se pierden las maletas. Suponga que una muestra aleatoria de 1000 vuelos arroja un total de 300 maletas perdidas. De esta manera, la media aritmética del número de maletas perdidas por vuelo es de 0.3, que se calcula al dividir 300/1000 (**μ = np**).

Si el número de maletas perdidas por vuelo se rige por una distribución de Poisson, ¿cuál es la probabilidad de no perder ninguna maleta?

$$P(x) = \frac{\mu^x e^{-\mu}}{x!} \quad \rightarrow \quad P(0) = \frac{(0.3)^0 (e^{-0.3})}{0!} = 0.7408$$

Ejemplo 2

Coastal Insurance Company asegura propiedades frente a la playa, a lo largo de Virginia, Carolina del Norte y del Sur, y las costas de Georgia. El cálculo aproximado es que, cualquier año, la probabilidad de que un huracán de categoría III (vientos sostenidos de más de 177 Km/h) o más intenso azote una región de la costa (la isla de St. Simons, Georgia, por ejemplo) es de 0.05. Si un dueño de casa obtiene un crédito hipotecario de 30 años por una propiedad recien comprada en St. Simons, ¿cuáles son las posibilidades de que el propietario experimente por lo menos un huracán durante el periodo del crédito?

Calculamos la media: μ = np= (30)(0.05) = 1.5

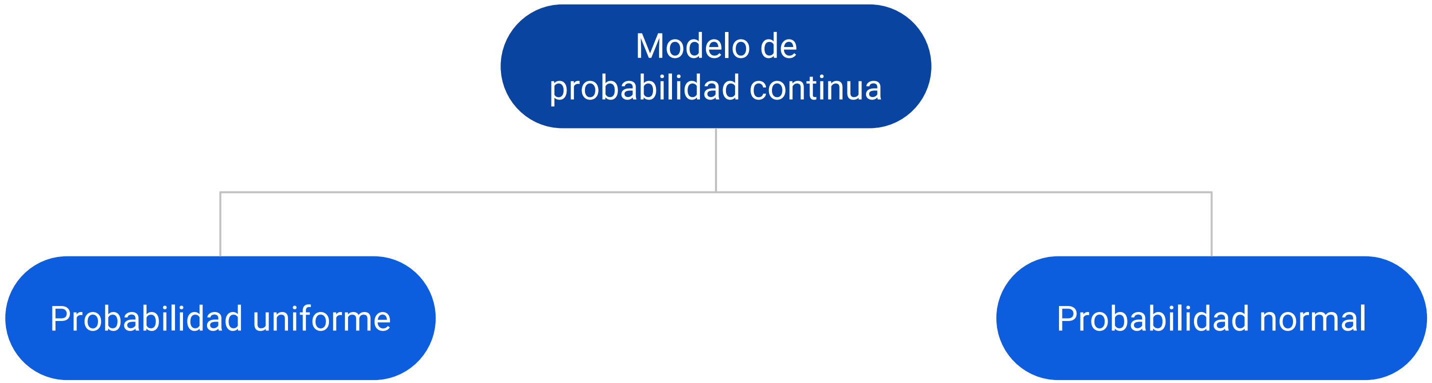
n es el número de años, 30 en este caso

p es la probabilidad de que toque tierra un huracán que se ajuste al criterio

μ es la media o número esperado de tormentas en un periodo de 30 años

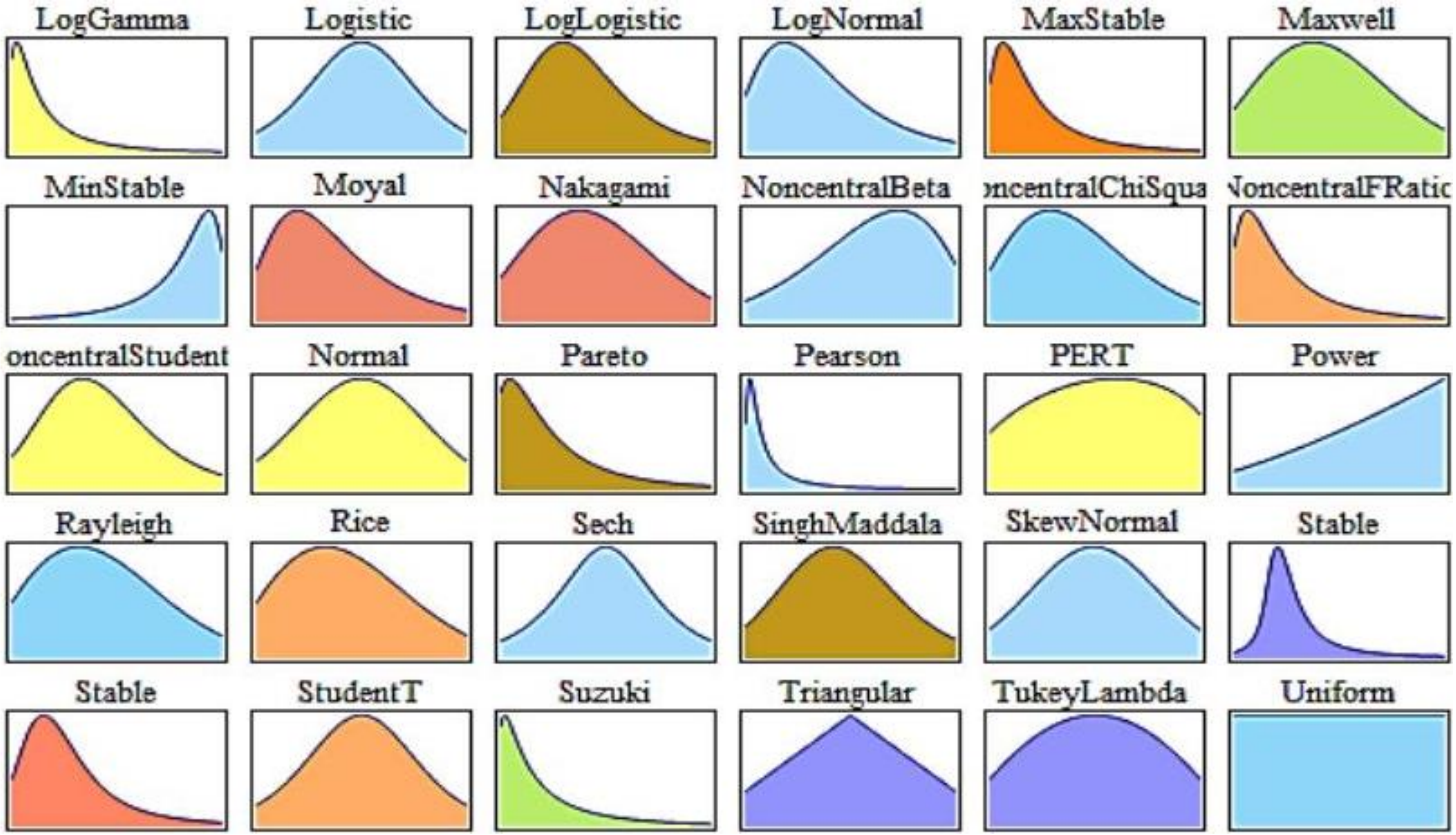
$$P(x) = \frac{\mu^x e^{-\mu}}{x!} \quad \rightarrow \quad P(x \geq 1) = 1 - P(x = 0) = 1 - \frac{\mu^0 e^{-1.5}}{0!} = 1 - .2231 = .7769$$

Distribuciones continuas



Una distribución de probabilidad continua resulta de medir algo, que se presenta dentro de cierto intervalo. Es importante señalar que una variable aleatoria continua tiene un número infinito de valores dentro de cierto intervalo particular.

Distribuciones de probabilidad continuas

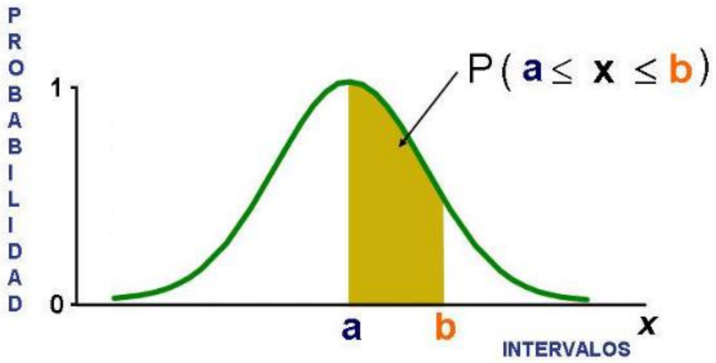


Distribución continua

Sea una variable X que puede tomar un conjunto continuo de valores entre *x=a* y *x=b*. Si la variable aleatoria es continua, hay infinitos valores posibles de la variable entre dos valores.

Si *X* tiene función de densidad *f(x)* y *a<b*, entonces la probabilidad de que *X* se encuentre en el intervalo *[a, b]* es

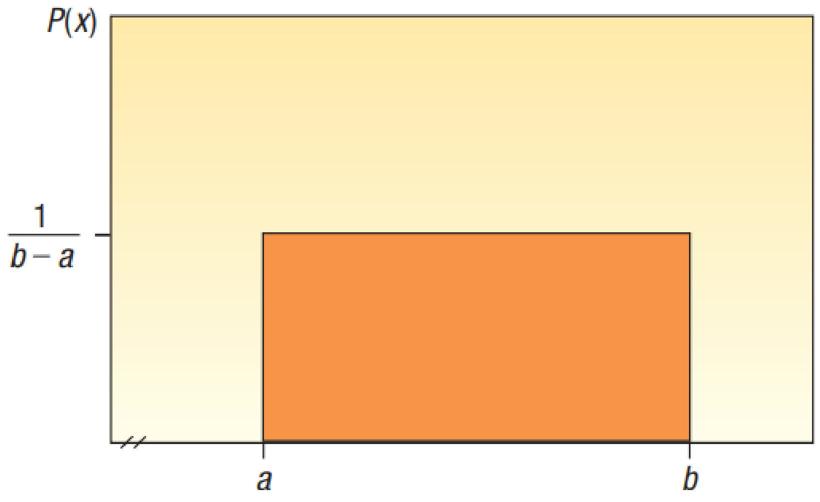
$P(a \leq X \leq b) = \int_a^b f(x)\, dx$



Distribución uniforme continua

La distribución tiene forma rectangular y queda definida por valores mínimos y máximos.

La distribución posee un valor mínimo *a* y un máximo *b*. La altura de la distribución es constante o uniforme para todos los valores entre *a* y *b*.



Características

MEDIA DE LA DISTRIBUCIÓN UNIFORME $$\mu = \frac{a+b}{2}$$

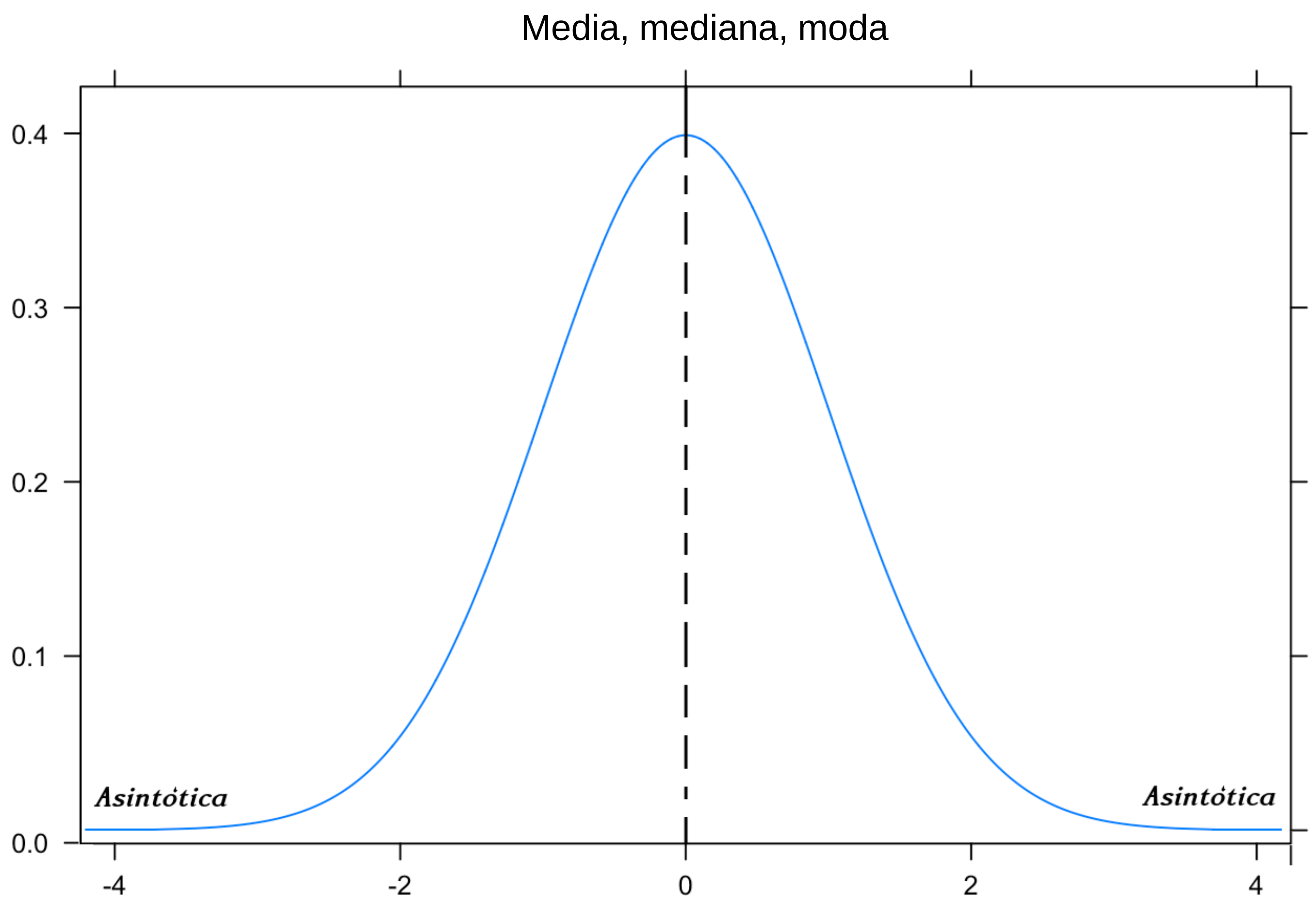
DESVIACIÓN ESTÁNDAR DE LA DISTRIBUCIÓN UNIFORME $$\sigma = \sqrt{\frac{(b-a)^2}{12}}$$

DISTRIBUCIÓN UNIFORME $$P(x) = \frac{1}{b-a} \text{ si } a \leq x \leq b \text{ y 0 en cualquier otro lugar}$$

Distribución Normal: características

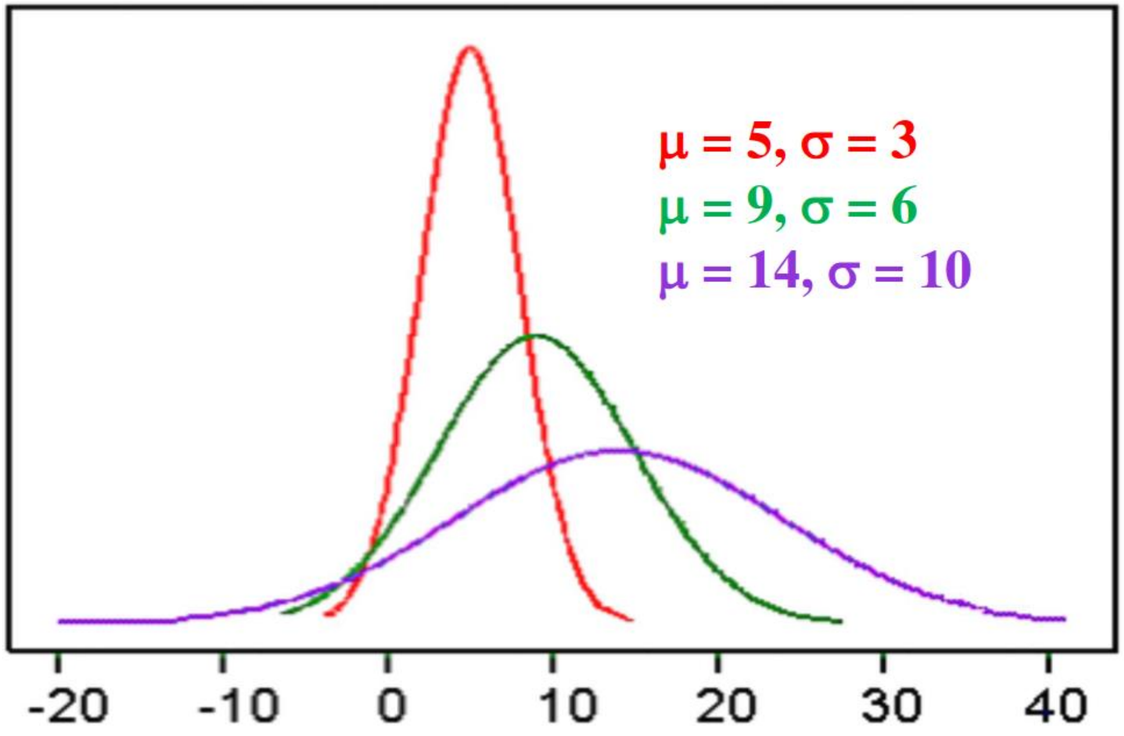
1. La curva normal tiene forma de campana, con un solo pico justo en el centro de la distribución.
2. La media, mediana y moda de la distribución aritmética, son iguales y se localizan en el pico.
3. La mitad del área bajo la curva está a la derecha del pico, y la otra mitad está a la izquierda.
4. La distribución normal es simétrica con respecto a su media.
5. La distribución normal es asintótica -la curva se acerca cada vez más al eje x, pero en realidad nunca llega a tocarlo.

Distribución Normal



Características

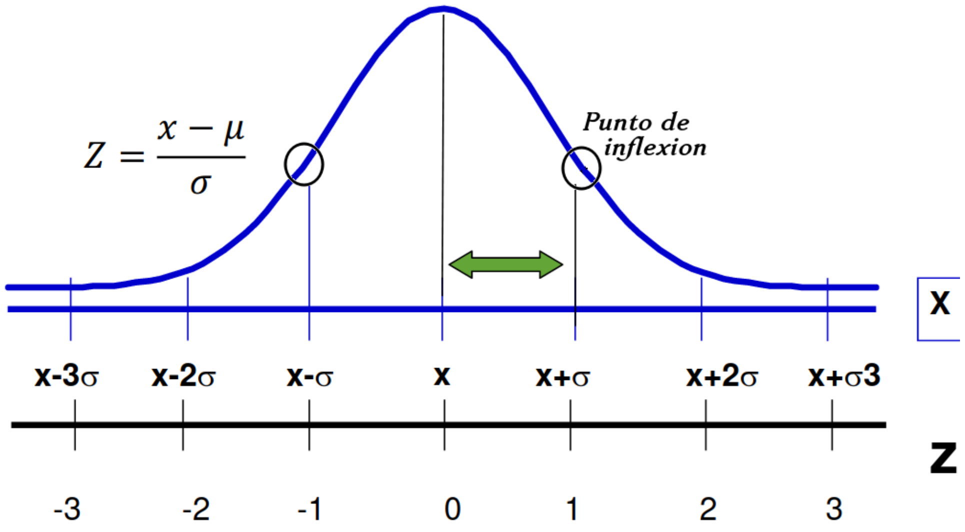
Los valores que toma la distribución normal, dependen del fenómeno que estemos estudiando.



Estandarización de la curva Normal

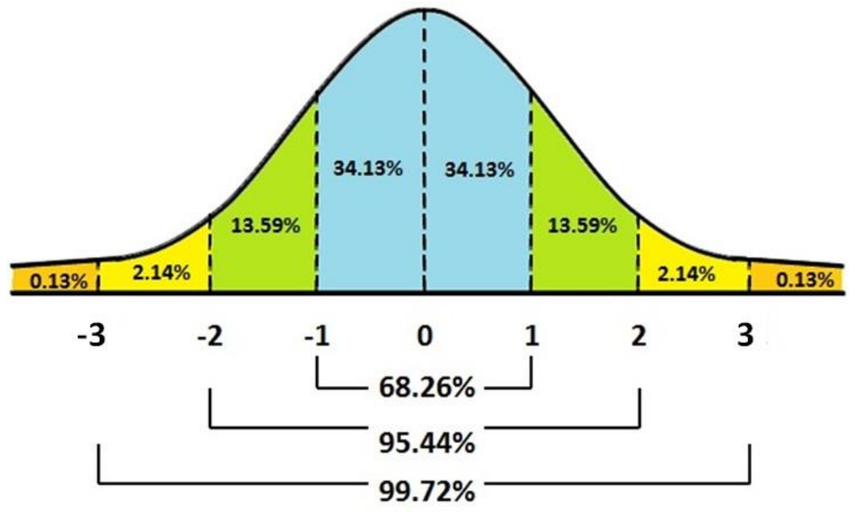
Una distribución normal estándar, que tiene media igual a 0 y desviación estándar igual a 1, se denomina distribución normal estándar.

La desviación estándar sigma representa la distancia de la media al punto de inflexión de la curva normal



Distribución normal estándar: características

1. Cerca de 68.3% del área bajo la curva normal está a menos de una desviación estándar con respecto a la media (μ ± 1σ).
2. Alrededor de 95.5% está a menos de dos desviaciones estándar de la media (μ ± 2σ).
3. 99.7% está a menos de tres desviaciones estándar de la media (μ ± 3σ).



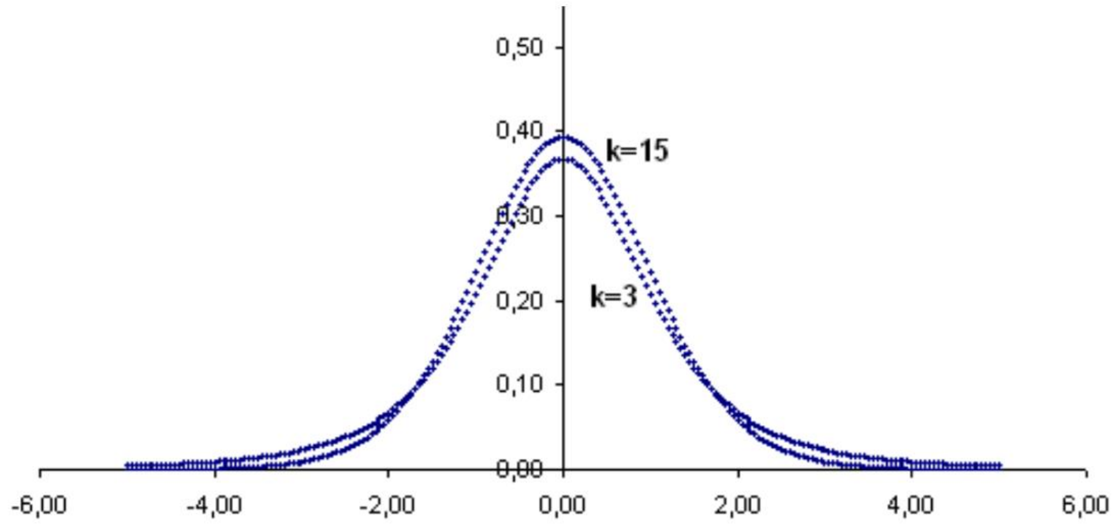
Distribución t de Student

Desarrollada con base en distribuciones de frecuencia empíricas por William Gosset ("Student").

Distribución muestral del promedio se ajusta muy bien a la distribución Normal cuando se conoce σ. Si *n* es grande, esto no presenta ningún problema. Es razonable sustituirla por *s* cuando es desconocida.

Características de la distribución *t* de *student*

1. Cada curva *t* tiene forma localizada en la media.
2. Cada curva *t* está más dispersa que la curva normal estándar.
3. A medida que *k (grados de libertad*)* aumenta, la dispersión de la curva *t* correspondiente disminuye.
4. A medida que $k \rightarrow \infty$, la secuencia de curvas *t* se aproxima a la curva Normal estándar.



*Los grados de libertad son la combinación del número de observaciones de un conjunto de datos que varían de manera aleatoria e independiente menos las observaciones que están condicionadas a estos valores arbitrarios.

En otras palabras, los grados de libertad son el número de observaciones puramente libres (que pueden variar) cuando estimamos los parámetros.

Distribución $\chi 2$ (Chi-cuadrada)

La distribución ji-cuadrada surge de la suma de normales estándar al cuadrado, su definición formal es la siguiente:

Definición: Sean $Z_1, Z_2, \dots, Z_k$ variables aleatorias independientes con distribución normal estándar, entonces la variable aleatoria dada por:

$$X = Z^2_1 + Z^2_2 + \dots, Z^2_k$$

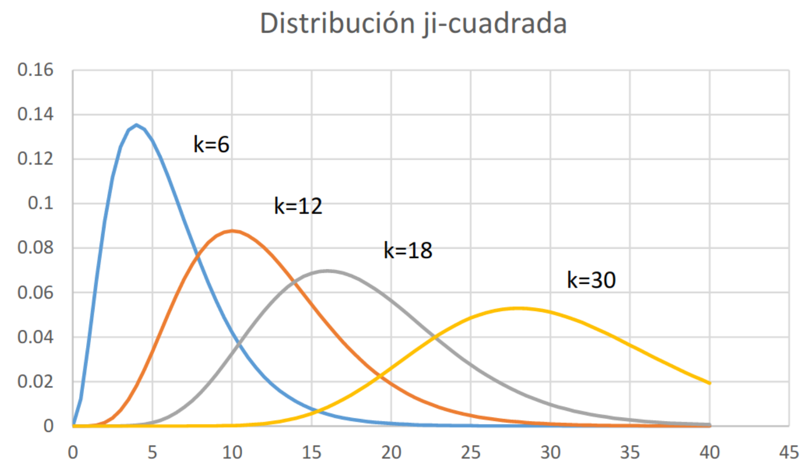
Tiene una distribución conocida como ji-cuadrada con *k* grados de libertad. Con una función de densidad:

$$f_k(x) = \frac{1}{2^{k/2}\Gamma\left(\frac{k}{2}\right)} x^{k/2-1} e^{-x/2} \quad \text{para } x \geq 0; \quad \text{y } 0 \text{ en otro caso}$$

Distribución χ^2 (Chi-cuadrada)

Es una distribución de probabilidad continua con un parámetro *k* (grados de libertad) de la variable aleatoria:

1. La distribución es asimétrica positiva.
2. Dado que la ji-cuadrada es la suma de variables aleatorias independientes, por el teorema del límite central, esta distribución se aproxima a una normal conforme k crece, como se puede ver en la gráfica.
3. Para cada tamaño muestral, se tendrá una distribución $\chi 2$ diferente.
4. El parámetro que caracteriza a una distribución $\chi 2$ son sus grados de libertad (*k*-1), originando una distribución para cada grado de libertad.



Resumen de parámetros de la chi-cuadrada

Resumen	Variable aleatoria ji-cuadrada
Función de densidad	$f_k(x) = \dfrac{1}{2^{k/2}\Gamma\left(\dfrac{k}{2}\right)} x^{k/2-1} e^{-x/2}$ para $x \geq 0$
Media	$\mu = k$
Varianza	$\sigma^2 = 2k$
Función generatriz de momentos	$M_x(t) = (1-2t)^{-k/2}, \qquad t < \dfrac{1}{2}$

Distribuciones de probabilidad discretas

Variables Aleatorias Discretas					
Distribución	$f(x)$	$F(x)$	$E(x)$	$Var(x)$	$M(t)$
$Bernoulli\,(p)$	$p^x q^{1-x}$	0 si $x<0$ q si $0<x<1$ 1 si $x>1$	p	pq	pe^t+q
$Binomial\,(n,p)$	$\binom{n}{x} p^x q^{n-x}$	$\sum_{i=0}^{x} \binom{n}{i} p^i q^{n-i}$	np	npq	$(pe^t+q)^n$
$Poisson\,(\lambda)$	$\frac{\lambda^x e^{-\lambda}}{x!}$	$e^{-\lambda} \sum_{i=0}^{x} \frac{\lambda^i}{i!}$	λ	λ	$e^{\lambda(e^t-1)}$
$Geométrica\ I\,(p)$	pq^{x-1}	$1-q^x$	$\frac{1}{p}$	$\frac{q}{p^2}$	$\frac{pe^t}{1-qe^t}$
$Geométrica\ II\,(p)$	pq^x	$1-q^{x+1}$	$\frac{q}{p}$	$\frac{q}{p^2}$	$\frac{p}{1-qe^t}$
$BinNeg\ I\,(r,p)$	$\binom{x-1}{r-1} p^r q^{x-r}$	$1-I_p(x+1,r)$	$\frac{r}{p}$	$\frac{rq}{p^2}$	$\left(\frac{pe^t}{1-qe^t}\right)^r$
$BinNeg\ II\,(r,p)$	$\binom{x+r-1}{r-1} p^r q^x$	$1-I_p(x+1,r)$	$\frac{rq}{p}$	$\frac{rq}{p^2}$	$\left(\frac{pe^t}{1-qe^t}\right)^r$
$HiperGeo\,(N,m,n)$	$\frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}$	$\sum_{i=0}^{x} \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}$	$\frac{nm}{N}$	$\frac{nm}{N}\left(1-\frac{m}{N}\right)\left(\frac{N-m}{N-1}\right)$	$\emptyset$

Distribuciones de probabilidad continuas

Variables Aleatorias Continuas					
Distribución	$f(x)$	$F(x)$	$E(x)$	$Var(x)$	$M(t)$
$Uniforme\,(a,b)$	$\frac{1}{b-a}$	$\begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & x \in [a,b] \\ 1 & x \geq b \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$
$Exponencial\,(\lambda)$	$\lambda\, e^{-\lambda x}$	$1-e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda-t}$
$Normal\,(\mu,\sigma^2)$	$\frac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\pi}\sigma}$	$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}\,dt$	μ	σ^2	$e^{\,\mu t+\frac{\sigma^2 t^2}{2}}$
$Gamma\,\,(k,\theta)$	$x^{k-1}\frac{e^{-x/\theta}}{\Gamma(k)\,\theta^k}$	$\int_0^x f(u;k,\theta)du$	$k\theta$	$k\theta^2$	$(1-\theta t)^{-k}, t<\frac{1}{\theta}$
$Beta\,(\alpha,\beta)$	$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$	$I_x(\alpha,\beta)$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	$1+\sum_{i=1}^{\infty}\left(\prod_{r=0}^{i-1}\frac{\alpha+r}{\alpha+\beta+r}\right)\frac{t^i}{i!}$
$Weibull\,(\lambda,k)$	$\frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}$	$1-e^{-(x/\lambda)^k}$	$\lambda\,\Gamma\left(1+\frac{1}{k}\right)$	$\lambda^2\,\Gamma\left(1+\frac{2}{k}\right)-\mu^2$	$\sum_{n=0}^{\infty}\frac{t^n\lambda^n}{n!}\Gamma\left(1+\frac{n}{k}\right)$
$Cauchy\,(x_0,\gamma)$	$\frac{1}{\pi\gamma\left(1+\left(\frac{x-x_0}{\gamma}\right)^2\right)}$	$\frac{1}{\pi}\,tan^{-1}\left(\frac{x-x_0}{\gamma}\right)+\frac{1}{2}$	*no existe*	*no existe*	*no existe*
$Paretto\,(x_m,\alpha)$	$\frac{\alpha\, x_m^{\alpha}}{x^{\alpha+1}}, x>x_m$	$1-\left(\frac{x_m}{x}\right)^{\alpha}$	$\frac{\alpha\, x_m}{\alpha-1}, \alpha>1$	$\frac{x_m^2\,\alpha}{(\alpha-1)^2(\alpha-2)}, \alpha>2$	$\alpha(-x_m t)^{\alpha}\,\Gamma(-\alpha,-x_m t)$

Distribuciones conjuntas

Se consideran vectores de dos dimensiones, es decir, vectores aleatorios de la forma (X,Y), en donde cada coordenada es una variable aleatoria.

Sea el vector aleatorio (X,Y), en donde ambas variables aleatorias están definidas en el mismo espacio de probabilidad (Ω, F, P). La función de distribución conjunta de (X,Y) es la función $F_{XY}(x,y):\mathbb{R}^2 \to [0,1]$, por lo general denotada solamente como $F(x,y)$ y definida como

$$F(x,y) = P(X \leq x, Y \leq y).$$

Así, $F(x,y)$ es un número en el intervalo $[0,1]$ que representa la probabilidad de que el vector aleatorio tome algún valor en la región que delimita el rectángulo $(-\infty,x]\times(-\infty,y]$.

Se dice distribución conjunta porque a su vez determina la probabilidad de que la coordenada X sea menor o igual que el valor fijo x y que la coordenada Y sea menor o igual que el valor real fijo y, equivalente a la probabilidad del evento $(X(\omega)\leq x)\cap(Y(\omega)\leq y)$, con $\omega\in\Omega$.

Conjuntas discretas

Sea el vector aleatorio (X,Y) tal que X y Y toman valores en conjuntos, a lo más numerables. En ese caso se dice que (X,Y) es un vector discreto y su función de probabilidad conjunta es la función $p_{X,Y}(x,y):\mathbb{R}^2\to[0,1]$, se define como

$$p_{X,Y}(x,y)=\begin{cases} P(X=x,Y=y) & si \quad (x,y)\in\{x_1,x_2,\ldots\}\times\{y_1,y_2,\ldots\}, \\ 0 & e.o.c \end{cases}$$

Así, $p_{X,Y}(x,y)$ se interpreta como la probabilidad de que las variables aleatorias X,Y tomen puntualmente los valores x,y.

Esta función de probabilidad conjunta cumple dos condiciones:

1. $p(x, y) \geq 0$.
2. $\sum_{x,y} p(x, y) = 1$.

A partir de la función de probabilidad conjunta se puede obtener la función de distribución conjunta como

$$F(x, y) \doteq F_{X,Y}(x, y) = \sum_{u \leq x} \sum_{v \leq y} p(u, v).$$

Para representar la función de masa de probabilidades conjunta se puede representar usando el siguiente arreglo tabular:

$X \setminus\setminus Y$	y_1	y_2	$\ldots$	y_j	$\ldots$	y_m
x_1	$p(x_1, y_1)$	$p(x_1, y_2)$	$\ldots$	$p(x_1, y_j)$	$\ldots$	$p(x_1, y_m)$
x_2	$p(x_2, y_1)$	$p(x_2, y_2)$	$\ldots$	$p(x_2, y_j)$	$\ldots$	$p(x_2, y_m)$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
x_i	$p(x_i, y_1)$	$p(x_i, y_2)$	$\ldots$	$p(x_i, y_j)$	$\ldots$	$p(x_i, y_m)$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
x_n	$p(x_n, y_1)$	$p(x_n, y_2)$	$\ldots$	$p(x_n, y_j)$	$\ldots$	$p(x_n, y_m)$

Conjuntas continuas

Decimos que la función integrable y no negativa $f_{X,Y}(x,y):\mathbb{R}^2 \to [0,\infty)$ es la función de densidad (conjunta) del vector aleatorio (X,Y) si para toda pareja de valores $(x,y)\in\mathbb{R}^2$ se tiene que

$$F_{X,Y}(x,y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,v)\,dv\,du$$, para $-\infty < x < \infty$, $-\infty < y < \infty$.

Igualdad fundamental

Dada la función de distribución conjunta de dos variables aleatorias continuas $F(x, y)$, es posible determinar la función de densidad conjunta usando el Teorema Fundamental del Cálculo como

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y),$$

siempre y cuando las derivadas parciales existan.

Propiedades

De manera similar al caso discreto, toda función de densidad satisface

1. $f(x, y) \geq 0$.

2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)dxdy = 1$.

Marginales (caso discreto)

Si en el vector aleatorio (X,Y), ambas componentes son discretas, con función de probabilidad $p(x,y)$ las funciones de probabilidad marginales de X y Y están dadas respectivamente por

- $p(x) = p_X(x) = \sum_y p(x,y)$,
- $p(y) = p_Y(y) = \sum_x p(x,y)$.

Marginales (caso continuo)

Sea $f(x,y)$ la función de densidad del vector aleatorio continuo (X,Y). Se definen las funciones de densidad marginales de X , Y , respectivamente como

- $f(x) = f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$.
- $f(y) = f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$.

Independencia

Para el caso bivariado, se tiene que las variables aleatorias son independientes si y sólo si, para cada $(x, y) \in \mathbb{R}^2$, se cumple

$$F_{X,Y}(x, y) = F_X(x) F_y(y).$$

Esta definición es equivalente si se sustituyen la función de probabilidad conjunta y las funciones de probabilidad marginales en el caso de variables aleatorias discretas; y es análogo el caso de variables aleatorias continuas.

Caso discreto y continuo

Si X y Y son variables aleatorias discretas, lo anterior es

$$P(X=x,Y=y)=P(X=x)P(Y=y)\,.$$

En términos de densidades, si X y Y son variables aleatorias continuas, tenemos que

$$f_{X,Y}(x,y)=f_X(x)\,f_Y(y)$$

es la condición de independencia. En caso contrario, si las relaciones anteriores no se cumplen, se dice que X y Y son dependientes, ya sean ambas discretas o continuas.

Distribuciones condicionales (discretas)

Sean X y Y variables aleatorias discretas. La función de probabilidad condicional denotada como $p(x|y)$, cuya interpretación es la probabilidad de la variable aleatoria X dado un valor de $Y=y$, se representa como

$$p(x|y)=P(X=x|Y=y)=\frac{P(X=x,Y=y)}{P(Y=y)}, \text{ si } P(Y=y)\neq 0.$$

Caso discreto

Considerando la función de probabilidad conjunta y la función de probabilidad marginal de Y, puede reescribirse como

$$p(x \mid y) = \frac{p(x,y)}{p(y)} = \frac{p(x,y)}{\sum_{x} p(x,y)}, \text{ si } f(y) \neq 0.$$

Notar que para cada y fijo $p(x \mid y)$ es una función de probabilidad en x, pues para todo y satisface:

1. $p(x \mid y) \geq 0$,
2. $\sum_{x} p(x \mid y) = 1$.

Caso discreto

De esta manera, se define la función de distribución condicional de X dado que $Y = y$, denotada como $F_{X|Y}(x, y)$, por

$$F_{X|Y}(x, y) = \sum_{z \leq x} p(z \mid y) = \frac{1}{p(y)} \sum_{z \leq x} p(z, y)\text{, siempre y cuando } f(y) \neq 0.$$

Considerando la ley de probabilidad total, la función de probabilidad de X puede escribirse como

$$p(x) = P(X = x) = \sum_{y} P(X = x \mid Y = y) P(Y = y) = \sum_{y} p(x \mid y) p(y).$$

Condicionales continuas

Sean X y Y variables aleatorias continuas con función de densidad conjunta $f(x,y)$. Se define la función de densidad condicional de la variable X dado que $Y=y$, como

$$f(x\mid y)=\frac{f(x,y)}{f(y)}, \text{ si } f(y)\neq 0.$$

Caso continuo

Se puede verificar que esta formulación se trata de una ciertamente de un mapeo respecto a x, pues la variable Y toma el valor fijo y, además se trata de función de densidad, a satisfacer entonces, las condiciones:

1. $f(x \mid y) \geq 0,$
2. $\int_{-\infty}^{\infty} f(x \mid y) = 1.$

Distribución condicional

Así, sea (X,Y) un vector aleatorio un vector aleatorio continuo con función de densidad $f(x,y)$, y sea Y una variable aleatoria tal que su densidad satisface $f(y) \neq 0$. La función de distribución condicional de X dado que $Y = y$ está determinada por

$$F(x \mid y) = \int_{-\infty}^{x} \frac{f(u,y)}{f(y)} du.$$

Ejemplo

Consideremos el ejemplo del vector aleatorio (X, Y) que tiene una función de densidad de probabilidad dada por:

$$f(x,y) = \frac{8}{3}x^3 y \text{ , para } 0 \leq x \leq 1,\ 1 \leq x \leq 2.$$

Para las marginales, es necesario integrar sobre el soporte de cada variable

Así,

$$f_X(x) = \int_1^2 \frac{8}{3} x^3 y dy = \left[\frac{8x^3}{3} \cdot \frac{y^2}{2} \right]_1^2 = \left[\frac{8x^3}{3} \cdot \frac{(2)^2}{2} - \frac{8x^3}{3} \cdot \frac{(1)^2}{2} \right]$$

$$= \frac{8x^3}{3} \cdot \frac{4}{2} - \frac{8x^3}{3} \cdot \frac{1}{2} = \frac{24}{6} x^3 = 4x^3 \text{, para } 0 \le x \le 1.$$

Luego, para la marginal de Y, hacemos

$$f_Y(y) = \int_0^1 \frac{8}{3} x^3 y dx = \left[\frac{x^4}{4} \cdot \frac{8y}{3} \right]_0^1 = \frac{(1)^4}{4} \cdot \frac{8y}{3} = \frac{2}{3} y \text{, para } 1 \le y \le 2.$$

Dadas las marginales y usando la definición, las distribuciones condicionales son:

$$f(x \mid y) = \frac{f(x,y)}{f(y)} = \frac{\frac{8}{3}x^3 y}{4x^3} = \frac{2}{3}y, \; 0 < x < 1, \; 1 < y < 2.$$

$$f(y \mid x) = \frac{f(x,y)}{f(x)} = \frac{\frac{8}{3}x^3 y}{\frac{2}{3}y} = 4x^3, \; 0 < x < 1, \; 1 < y < 2.$$

Estas corresponden a las distribuciones (condicionales) de las variables X y Y, con valores fijos $Y = y$ y $X = x$, respectivamente.

Valor esperado

Sea (X,Y) un vector aleatorio y sea $g(X,Y)$ una función (real valuada) tal que $g:\mathbb{R}^2 \to \mathbb{R}$. Es posible probar que $g(X,Y)$ es una variable aleatoria y su valor esperado se puede calcular en los dos casos usuales:

- Discreto

Si X y Y son variables aleatorias discretas,

- Continuo

Si X y Y son variables aleatorias continuas,

$$E[g(x,y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)dxdy .$$

Sea $g(X,Y) = aX + bY$, con $a,b \in \mathbb{R}$, X y Y variables aleatorias con densidad conjunta $f_{X,Y}(x,y)$.

Resultado: $E(X+Y) = aE(X) + bE(Y)$.

Luego, teniendo que $f(x \mid y)$ es continua, por el Teorema Fundamental del Cálculo, esta se puede obtener de la función de distribución condicional de X dado que $Y = y$ como

$$f(x \mid y) = \frac{\partial}{\partial x} F(x \mid y).$$

Además, se tiene la siguiente versión de la ley de probabilidad total para el caso continuo:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) = \int_{-\infty}^{\infty} f_{X|Y}(x \mid y) f_Y(y) dy .$$

Referencias

1. Álvarez, M. Á. G. (2015). Introducción a la teoría de la probabilidad I. Primer curso. Fondo de Cultura Económica.
2. Blitzstein J., Hwang J. (2019) Introduction to probability – Second Edition : Taylor & Francis Group, LLC.
3. Deisenroth M, Faisal A., Ong C. S. (2020) Mathematics for Machine Learning. Cambridge : Cambridge University Press.
4. Doldán, F. D. (2007). Redes bayesianas y riesgo operacional. Revista Galega de Economía, 16, 1-16.
5. Milton, J. S. (2001). Estadística para biología y ciencias de la salud. McGrawHill-Interamericana. México.
6. Wasserman, Larry (2004). All of statistics : a concise course in statistical inference. Springer Verlag -- New York, NY; Springer texts in Statistics.

Contacto

Dr. Roberto Bárcenas Curtis

rbarcenas@ciencias.unam.mx