MIOTI MAKE BUILD LEARN

20 de febrero de 2020

Estadística para Data Science

Sesión 2: Introducción a la combinatoria y la probabilidad

Jesús Hernando Corrochano

Madrid Internet of Things Institute
Make-build-learn

Estadística para Data Science

www.mioti.es

Programa

	Sesión 1 13/2	Sesión 2 20/2	Sesión 3 27/2	Sesión 4 5/3	Sesión 5 12/3	Sesión 6 19/3	Sesión 7 26/3	Sesión 8 2/3
Introducción a la estadística								
Introducción a la combinatoria y la probabilidad								
Estadística descriptiva								
Regresión y correlación								
Estadística inferencial								
Probabilidad Total. Teorema de Bayes.Test A/B								

Índice

Combinatoria

Combinatoria

Factorial:

El número **n·(n–1)·(n–2)·…·3·2·1** se llama **factorial** de n, y se representa por **n!**, donde n es un número natural.

O! = 1

Permutaciones:

Ordenaciones posibles de un conjunto de n elementos distintos

$$P_n = n!$$

COMBINATORIA:

Parte de la matemática que estudia la formación de subconjuntos o agrupaciones de elementos partiendo de un conjunto dado, teniendo en cuenta la ordenación y el número de esos elementos

Permutaciones con repetición:

posibles ordenaciones de una secuencia de n elementos, entre los que hay algunos repetidos (*uno se repite α veces, otro β veces, otro γ veces… etc*.).

$$P_n^{\alpha,\beta,\gamma\ldots} = \frac{n!}{\alpha!\cdot\beta!\cdot\gamma!\cdots}$$

Combinaciones:

Denominamos combinaciones de n elementos tomados de r en r: a las posibles muestras **sin orden** de r elementos distintos que se pueden extraer de un conjunto de n elementos (r≤n).

$$\binom{m}{n} = \frac{m!}{n!\cdot(m-n)!}$$

Combinaciones con repetición:

de n elementos tomados de r en r: posibles muestras no ordenadas de r elementos no necesariamente distintos que se pueden extraer de un conjunto de n elementos

$$CR_n^r = \binom{n+r-1}{r}$$

Variaciones:

de n elementos tomados de r en r : posibles muestras **ordenadas** de r elementos distintos que se pueden extraer de un conjunto de n elementos (r≤n).

$$V_n^r = n \cdot (n-1) \cdot (n-2) \cdots (n-r+1)$$

Variaciones con repetición:

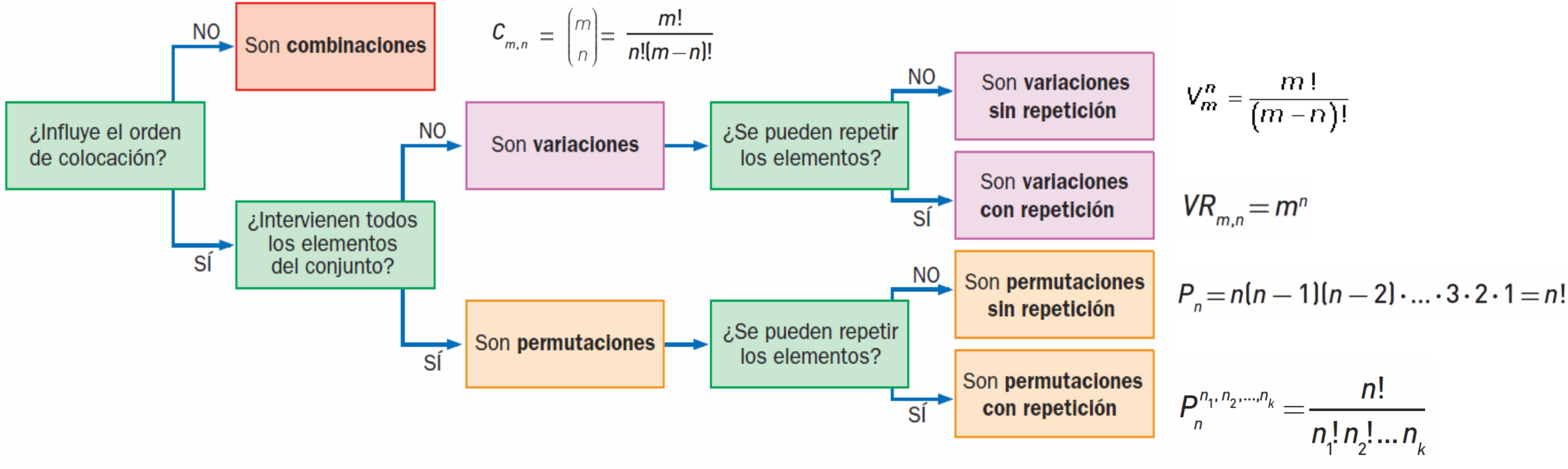
de n elementos tomados de r en r: posibles muestras **ordenadas** de r elementos no necesariamente distintos que se pueden extraer de un conjunto de n elementos

$$VR_n^r = n \cdot n \cdot n \cdots n = n^r$$

Combinatoria

m = número de elementos

n = de cuánto en cuánto se toman



Probabilidad

Las probabilidades constituyen una rama de las matemáticas que se ocupa de medir o determinar cuantitativamente la posibilidad de que un suceso o experimento produzca un determinado resultado.

Una variable aleatoria es una variable estadística cuyos valores se obtienen de mediciones en algún tipo de experimento aleatorio.

VARIABLE ALEATORIA DISCRETA:

El conjunto de posibles valores es numerable.

Suelen estar asociadas a experimentos en los que se mide el número de veces que sucede algo.

VARIABLE ALEATORIA CONTINUA:

El conjunto de posibles valores es no numerable.

Puede tomar todos los valores de un intervalo. Son el resultado de medir.

ESPACIO MUESTRAL

El *espacio muestral* es el conjunto de todos los posibles resultados de un experimento aleatorio y se suele representar como E (o bien como omega, Ω, del alfabeto griego).

$$Dim(E)=2^n$$

PUNTO MUESTRAL / SUCESO ELEMENTAL

Denominamos punto muestra a cada uno de los elementos del experimento

SUCESO

Un suceso es cualquier subconjunto del espacio muestral.

PROBABILIDAD *es una medida de la posibilidad que tiene un suceso de ocurrir*

*La **probabilidad** está basada en el estudio de la combinatoria y es fundamento necesario de la estadística, además de otras disciplinas como matemática, física u otra ciencia*

*La **Probabilidad** es la mayor o menor **posibilidad** de que ocurra un determinado suceso. En otras palabras, su noción viene de la necesidad de medir o determinar cuantitativamente la certeza o duda de que un suceso dado ocurra o no.*

*La **teoría de la probabilidad** es un modelo matemático que se ocupa de <u>analizar</u> los fenómenos aleatorios; esto implica la contraposición respecto de los fenómenos ya determinados, que son aquellos en los cuales el resultado del experimento que se realiza, atendiendo a determinadas condiciones, produce un resultado único y previsible, que se repetirá la cantidad de veces que éste vuelva a hacerse, siempre y cuando se respeten las mismas condiciones*

ESPACIO MUESTRAL

El *espacio muestral* es el conjunto de todos los posibles resultados de un experimento aleatorio y se suele representar como E (o bien como omega, Ω, del alfabeto griego).

PUNTO MUESTRAL / SUCESO ELEMENTAL

Denominamos punto muestra a cada uno de los elementos del experimento

SUCESO

Un suceso es cualquier subconjunto del espacio muestral. El mismo espacio muestral es un suceso llamado suceso seguro y el conjunto vacío, $\emptyset$, es el suceso imposible.

ESPACIO DE SUCESOS (S)

Es el conjunto de todos los sucesos aleatorios

Si E tiene un número finito de elementos, n, el número de elementos de S es 2n

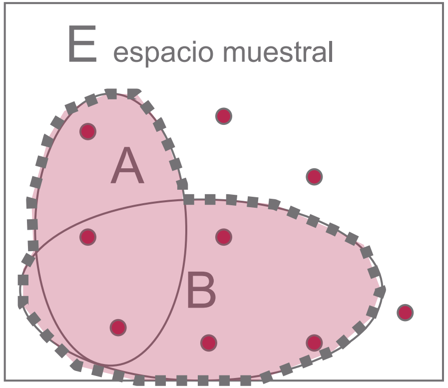
$Dim(S)=2^n$

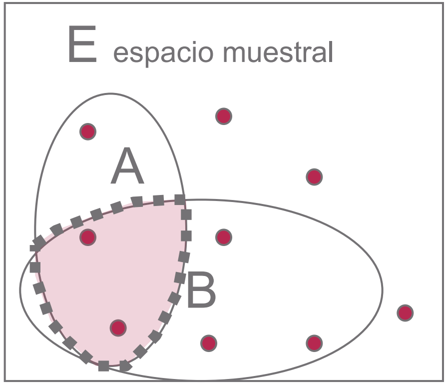
PROBABILIDAD es una medida de la posibilidad que tiene un suceso de ocurrir

Se llama **suceso contrario** (complementario) de un suceso A, A’ , al formado por los elementos que no están en A

Se llama **suceso unión** de A y B, AUB, al formado por los resultados experimentales que están en A o en B (incluyendo los que están en ambos.

Se llama **suceso intersección** de A y B, A∩B o simplemente AB, al formado por los elementos que están en A y B





CONCEPTOS

La probabilidad es una aplicación (función) que asigna a cada suceso A de un espacio muestral E un número real que debe verificar las siguientes propiedades:

$$E \to R$$

$$A \to P(A)$$

1.- La probabilidad del suceso seguro es 1:

$$P(E) = 1.$$

2.- La probabilidad de cualquier suceso siempre es un número no negativo:

$$P(A) \geq 0, \text{ para todo } A.$$

3.- Si dos sucesos son incompatibles entonces la probabilidad de la unión es la suma de sus probabilidades:

$$\text{Si } A \cap B = \varnothing \text{ entonces } P(A \cup B) = P(A) + P(B).$$

CONCEPTOS

Ley de Laplace: $P(A) = \dfrac{\text{Número de casos favorables}}{\text{Número de casos posibles}}$

Suceso seguro E: $P(E)=1$

Suceso imposible $\varnothing$: $P(\varnothing)=0$

Suceso opuesto: $P(\overline{A}) = 1 - P(A)$

Sucesos incompatibles: $P(A \cap B) = 0$

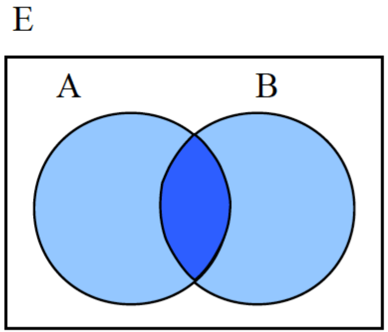
Sucesos independientes: $P(A/B)=P(A)$

Unión de sucesos incompatibles: $P(A \cup B) = P(A) + P(B)$

Unión de sucesos compatibles: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Intersección de sucesos independientes: $P(A \cap B) = P(A) \cdot P(B)$

"Si los sucesos elementales son equiprobables, la probabilidad de un suceso A es el número de casos favorables dividido por el número de casos posibles".



CONCEPTOS

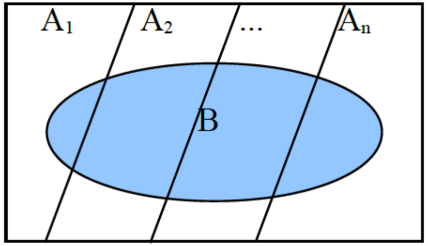
Intersección de sucesos dependientes (Probabilidad compuesta): $P(A \cap B) = P(A) \cdot P(B / A)$

Diferencia de sucesos: $P(A - B) = P(A) - P(A \cap B)$

Leyes de De Morgan:

$$P(\overline{A \cup B}) = P(\overline{A} \cap \overline{B})$$

$$P(\overline{A \cap B}) = P(\overline{A} \cup \overline{B})$$



Probabilidad condicionada: $P(A / B) = \dfrac{P(A \cap B)}{P(B)}$

Teorema de la Probabilidad total:

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

Teorema de Bayes: $P(A_i / B) = \dfrac{P(A_i) \cdot P(B / A_i)}{P(B)}$

Enunciado del teorema de la probabilidad total

Sean $\{A_1, A_2, \ldots, A_n\}$ un sistema completo de sucesos incompatibles dos a dos, con probabilidades no nulas, suma de probabilidades 1. Sea B otro suceso del que conocemos las probabilidades condicionadas: $P(B/A_i)$. Entonces:

$$P(B) = \sum_{k=1}^{n} P(B/A_k) \cdot P(A_k)$$

Enunciado del teorema de Bayes

Sean $\{A_1, A_2, \ldots, A_n\}$ un sistema completo de sucesos incompatibles dos a dos, con probabilidades no nulas, suma de probabilidades 1. Sea B otro suceso del que conocemos las probabilidades condicionadas: $P(B/A_i)$. Entonces:

$$P(A_i / B) = \frac{P(B/A_i) \cdot P(A_i)}{P(B)} = \frac{P(B/A_i) \cdot P(A_i)}{\sum_{k=1}^{n} P(B/A_k) \cdot P(A_k)}$$

CONCEPTOS

Sucesos	Al realizar un experimento aleatorio existen varios posibles resultados o **sucesos posibles**. Un **suceso** es un subconjunto del conjunto de posibles resultados.	Tiramos un dado. Posibles resultados = {1, 2, 3, 4, 5, 6} Suceso *obtener múltiplo de* 3 = {3, 6}
Asignación de probabilidades	Una medida Límite al que tienden las frecuencias relativas. Regla de Laplace: Si los sucesos elementales son equiprobables entonces: p = casos favorables / casos posibles.	$P(5) = 1/6$. P(sacar múltiplo de 3) = 2/6
Axiomática de Kolmogorov	*1*. $P(E) = 1$. 2. $P(A) \geq 0$, para todo A. *3*. Si $A \cap B = \varnothing$ entonces $P(A \cup B) = P(A) + P(B)$.	
Teoremas de Probabilidad	Suceso contrario: $P(X) + P(noX) = 1$. Probabilidad de intersección: $P(A \cap B) = P(A) \cdot P(B/A)$. Probabilidad de unión: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.	P(no 5) = 1 – 1/6 = 5/6. $P(5 \cup$ múl. 3) = 1/6 + 2/6 =3/6 P sacar primero un 5 y luego múltiplo de 3 =1/6·2/6 =2/36
Teorema de *Bayes*	$P(A_i / B) = \dfrac{P(B/A_i) \cdot P(A_i)}{P(B)} = \dfrac{P(B/A_i) \cdot P(A_i)}{\sum_{k=1}^{n} P(B/A_k) \cdot P(A_k)}$	

CONCEPTOS

Una **función de probabilidad o de masa de probabilidad**, es una función que asocia a cada punto de su espacio muestral X su probabilidad:

talque p_i es la probabilidad del suceso $X = x_i$. $P(x_i) = p_i$

Por definición de probabilidad:

$$\sum_{i=1}^{k} P(x_i) = 1.$$

Función de Distribución es la probabilidad de que la variable tome valores iguales o inferiores a x:

$$F(x) = p(X \leq x)$$

Madrid Internet
of Things Institute
Make-build-learn

Distribuciones Discretas de Probabilidad

Distribuciones *DISCRETAS* de probabilidad

Una **variable aleatoria** da una descripción numérica de los resultados de un experimento. La distribución de probabilidad de una variable aleatoria describe cómo se reparten las probabilidades entre los valores que toma dicha variable. En toda variable aleatoria discreta, x, su distribución de probabilidad se define mediante una función de probabilidad, que se denota f(x) y la cual da la probabilidad que corresponde a cada valor de la variable aleatoria.

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

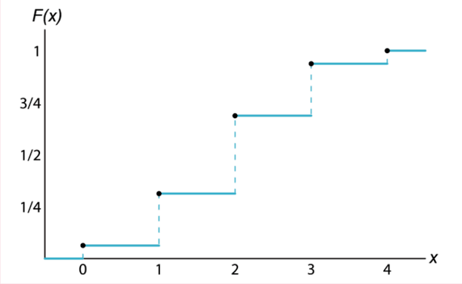
2. $\sum_x f(x) = 1,$

3. $P(X = x) = f(x)$

Distribuciones *DISCRETAS* de probabilidad

La **función de distribución acumulada $F(x)$** de una variable aleatoria discreta X, con distribución de probabilidad $f(x)$ es:

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t), \qquad para \quad -\infty < x < \infty$$



Distribuciones *DISCRETAS* de probabilidad: binomial

La distribución binomial

Un **experimento binomial** es el que contiene las siguientes características:

1. El experimento consiste en ***n* intentos idénticos**.
2. Cada intento resulta en uno de dos resultados: éxito o fracaso.
3. La probabilidad de éxito en un solo intento es igual a *p* y es de un intento a otro. La probabilidad de fracaso es igual a (*1*-*p*)= *q*.
4. Los intentos son independientes.
5. Estamos interesados en *x,* el número de éxitos observados durante los *n* intentos, para *x*= 0, 1, 2, ..., *n*.

n= número de ensayos o experimentos

q= probabilidad de fracaso

$$f(x) = \binom{n}{x} p^x q^{n-x}$$

x= número de éxitos

p= probabilidad de éxito

Distribuciones *DISCRETAS* de probabilidad: poisson

Distribución de Poisson

Se dice que una variable aleatoria *X* tiene una **distribución de Poisson** con parámetro λ (λ > 0) si la función de masa de probabilidad de X es:

$$p(x; \lambda) = \frac{e^{-\lambda} \lambda^x}{x!}$$

x= 0, 1, 2, 3…

El valor de λ es con frecuencia un valor por unidad de tiempo o por unidad de área.

La letra e en p(x;) representa la base del sistema de logaritmos naturales; su valor numérico es aproximadamente 2.71828.

Distribución de Poisson

El hecho de que $\sum_{k=0}^{n} p(x;\lambda) = 1$ es una consecuencia de la expansión de la serie infinita de Maclaurin de e^{λ} , la cual aparece en la mayoría de los textos de cálculo.

$$e^{\lambda} = 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots \sum_{k=0}^{n} \frac{\lambda^x}{x!}$$

Si los dos términos extremos de la expresión se multiplican por $e^{-\lambda}$ y luego $e^{-\lambda}$ se coloca dentro de la suma, el resultado es:

$$1 = \sum_{k=0}^{n} e^{-\lambda} \frac{\lambda^x}{x!}$$

Lo que demuestra que $p(x;\lambda)$ satisface la segunda condición necesaria para especificar una función masa de probabilidad.

Distribuciones *DISCRETAS* de probabilidad: hipergeométrica

Distribución hipergeométrica

La distribución de probabilidad hipergeométrica está estrechamente relacionada con la distribución binomial. Pero difieren en dos puntos: en la distribución hipergeométrica los ensayos no son independientes y la probabilidad de éxito varía de ensayo a ensayo.

En la notación usual en la distribución hipergeométrica, k denota el número de elementos considerados como éxitos que hay en una población de tamaño N, y N-k denota el número de elementos considerados como fracasos que hay en dicha población.

Función de distribución hipergeométrica

La **función de probabilidad hipergeométrica** se usa para calcular la probabilidad de que en una muestra aleatoria de n elementos, seleccionados sin reemplazo, se tengan x éxitos y n-x fracasos. Para que se presente este resultado, debe tener x éxitos de los k éxitos que hay en la población y n-x fracasos de los N-k fracasos. La siguiente función de probabilidad hipergeométrica proporciona $f(x)$, la probabilidad de tener x éxitos en una muestra de tamaño n.

Si X es el número de éxitos (E) en una muestra completamente aleatoria de tamaño n extraída de la población N compuesta de k *éxitos* y (N - k) *fallas*, entonces la distribución de probabilidad de X llamada distribución hipergeométrica, es

$$f(X=x)=\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

con x un entero que satisface máx $(0, n - N + k) \leq x \leq \text{mín}(n, K)$.

Distribuciones *DISCRETAS* de probabilidad

Distribución binomial

$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$

para $x = 0, 1, \ldots, n$.

Parámetros: $n \in \mathbb{N}$, $p \in (0,1)$.

Media: np.

Varianza: $np(1-p)$.

Distribución Poisson

$f(x) = \frac{\lambda^x}{x!} e^{-\lambda}$

para $x = 0, 1, 2, \ldots$

Parámetro: $\lambda > 0$.

Media: λ.

Varianza: λ.

Distribución hipergeométrica

$f(x) = \binom{K}{x}\binom{N-K}{n-x}/\binom{N}{n}$

para $x = 0, 1, \ldots, n$.

Parámetros: N, K, n.

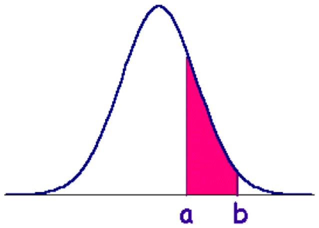
Media: nK/N.

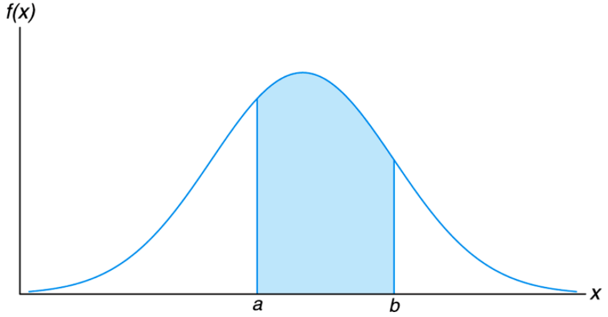
Varianza: $nK(N-K)(N-n)/(N^2(N-1))$.

Distribuciones Continuas de Probabilidad

Distribuciones *CONTINUAS* de probabilidad

Una variable aleatoria que puede tomar cualquier valor en un intervalo se denomina *continua*.





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

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

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

Distribuciones *CONTINUAS* de probabilidad: normal

Distribución normal

Una distribución de probabilidad continua que es muy importante es la distribución normal.

Se dice que una variable aleatoria continua X tiene una distribución normal con parámetros μ y σ (o μ y σ^2), donde -∞ < μ < ∞ y σ> o, si la función de densidad de probabilidad de X es

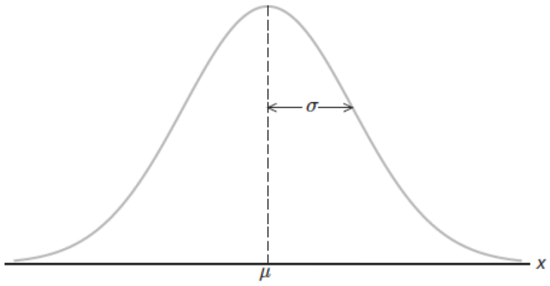
$$f(x;\ \mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{\overline{-(x-\mu)^2}(2\sigma^2)}.$$

- En donde $\mu \in R$ y $\sigma > 0$ son dos parámetros.
- Escribimos entonces $X \sim N(\mu, \sigma^2)$.

Distribución normal

En particular, decimos que la variable aleatoria X tiene una distribución normal estándar si tiene una distribución normal con parámetros $\mu = 0$ y $\sigma^2 = 1$. En este caso la función de densidad se reduce a la expresión:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

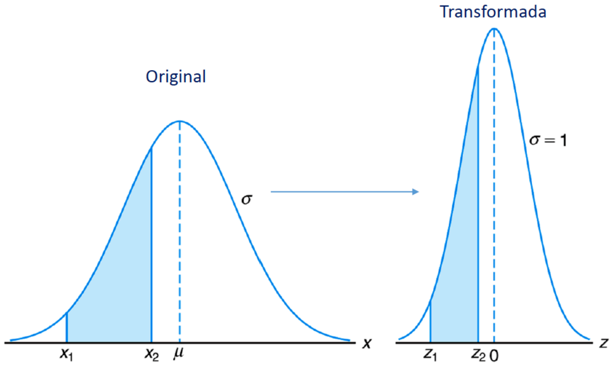


Distribuciones *CONTINUAS* de probabilidad: normal

Estandarización

Cuando no contamos con los parámetros μ= 0 y σ^2=1, tendemos a manipular la curva.

$$Z = \frac{x-\mu}{\sqrt{\sigma^2}} \longleftrightarrow Z = \frac{x-\mu}{\sigma}$$

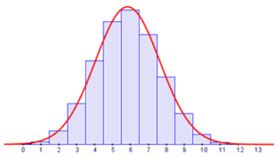


A la operación anterior se le conoce con el nombre de estandarización, y bajo tal transformación se dice que la variable X ha sido estandarizada. De esta manera podemos hablar de una distribución normal estándar $Z \sim n(0,1)$.

Normal & Binomial

Aproximación de la distribución normal a la binomial

Desde un punto de vista teórico, algunas distribuciones convergen a la normal a medida que sus parámetros se aproximan a ciertos limites.



La distribución normal es una distribución de aproximación conveniente, ya que la función de distribución acumulativa se tabula con mucha facilidad. La distribución binomial se aproxima bien por medio de la normal en problemas prácticos cuando se trabaja con la función de distribución acumulativa.

Aproximación de la distribución normal a la binomial

Si X es una variable aleatoria binomial con media $\mu = np$ y varianza $\sigma^2 = npq$, entonces la forma limitante de la distribución de

$$Z = \frac{X - np}{\sqrt{npq}}$$

conforme $n \to \infty$, es la distribución normal estándar $n(z; 0, 1)$.

Madrid Internet
of Things Institute
Make-build-learn

Make
build
learn_

Calle Rufino González 25
28037 Madrid
+34810527241
www.mioti.es

www.mioti.es