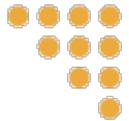


Together for Tomorrow!
Enabling People

Probabilidad y Estadística para la Inteligencia Artificial

SAMSUNG



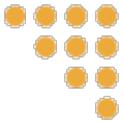
Índice

<u>Módulo 1. Cálculo</u>	<u>5</u>
<u>Módulo 2. Álgebra lineal y geometría analítica</u>	<u>51</u>
<u>Módulo 3. Probabilidad</u>	<u>103</u>
<u>Módulo 4. Estadística</u>	<u>167</u>
<u>Módulo 5. Aplicación a la Inteligencia Artificial</u>	<u>253</u>

Together for Tomorrow!
Enabling People

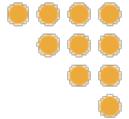
Cálculo

#TecnologíaConPropósito



Índice

Conceptos básicos	5
1. Introducción	7
2. Conceptos básicos	8
2.1. Sistemas numéricicos	8
2.2. Intervalos	11
2.3. Polinomios	12
2.4. Función matemática	13
2.5. Incrementos	14
2.6. Límites	15
Sucesiones y series	17
1. Introducción	19
2. Succesiones	20
2.1. Monotonía y acotamiento	20
2.2. Convergencia	21
3. Series	24
3.1. Propiedades	24
Derivadas	27
1. Introducción	29
2. Derivadas	30
2.1. Función derivada	30
2.2. Continuidad y derivabilidad	31
2.3. Propiedades y reglas de derivación	33



Índice

<u>Integrales</u>	<u>37</u>
<u> 1. Introducción</u>	<u>39</u>
<u> 2. Integrales</u>	<u>40</u>
<u> 2.1. Integrales indefinidas</u>	<u>40</u>
<u> 2.2. Integrales definidas</u>	<u>43</u>

Together for Tomorrow!
Enabling People

Conceptos básicos

#TecnologíaConPropósito



1. Introducción

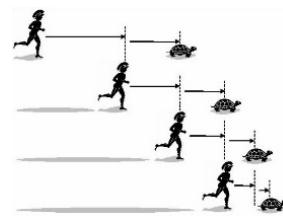
La matemática es la ciencia más ancestral que existe, nacida como un intento por cuantificar todo cuanto rodeaba al humano de la época, hace más de 20 000 años, comenzando por sucesos tan inherentes a la naturaleza como el paso del tiempo o los ciclos menstruales o lunares. Así, surge el concepto que pasaría a ser llamado “número” y cuya representación y manipulación daría origen a la disciplina que hoy conocemos como aritmética.

Sin embargo, tuvieron que pasar muchísimos años para que se descubriese que no todo en la naturaleza podía ser interpretado como un conjunto de sumas, restas, multiplicaciones o divisiones. Muchos fenómenos naturales eran regidos por unas variaciones continuas imposibles de describir mediante estas operaciones, como la aceleración de un objeto en caída libre; otros, conducían a misteriosos valores incommensurablemente grandes o pequeños.

Como respuesta a esto, en el siglo XVII y de la mano de dos grandes pensadores del momento como Isaac Newton (1643-1727) y Gottfried Leibniz (1646-1716) aparece el llamado *cálculo infinitesimal*. Consiguiendo modelar estas infinitas o infinitesimales cantidades, situaciones consideradas paradojas desde la Antigua Grecia, como la de la dicotomía o *Aquiles y la tortuga*, conseguían ser justificadas matemáticamente.

Paradoja de Zenón

Aquiles, el atleta más veloz, capaz de correr los 100 m en 10 segundos, no podrá alcanzar a una lenta tortuga, diez veces menos rápida que él. Ambos disputan una carrera, concediendo Aquiles una ventaja de 100 m a la tortuga. Cuando Aquiles ha cubierto esos 100 m, la tortuga se ha desplazado 10 m. Al cubrir Aquiles esos 10 m., la tortuga se ha desplazado 1 m. Mientras cubre ese metro que le separa de la tortuga, ésta ha recorrido 0'1 m. Y así indefinidamente.





2. Conceptos básicos

Para comprender los principales protagonistas del cálculo, como las derivadas y las integrales, se deben repasar los conceptos básicos de las matemáticas.

2.1. Sistemas numéricos

Primero, se debe conocer la propiedad de los “sistemas cerrados”: si un conjunto es cerrado con respecto a una operación, el resultado de esa operación con dos miembros cualquiera del conjunto es también un miembro del conjunto.

2.1.1. Números naturales

Los números naturales fueron los primeros en usarse, pues son todos aquellos con los que se pueden contar elementos de ciertos conjuntos o realizar algunos de los cálculos elementales. Los números naturales forman un conjunto cerrado con respecto a la suma y la multiplicación. Esto es, de la suma o multiplicación de dos números naturales cualesquiera se obtiene siempre otro número natural.

Existen infinitos números naturales y su conjunto es representado mediante el símbolo \mathbb{N} . Así;

$$\mathbb{N} = \{1, 2, 3, 4, \dots\}$$

La comunidad global matemática no se ha puesto de acuerdo sobre si el número 0 debe pertenecer o no a los números naturales, representado a veces como \mathbb{N}_0

2.1.2. Números enteros

Los números reales están compuestos por los números naturales más el cero (0) y sus opuestos negativos, que son representados con el signo – delante. El conjunto se denota con el símbolo \mathbb{Z} y se representa como:

$$\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$$

Este conjunto es cerrado a la suma, multiplicación y resta.

2.1.3. Números racionales

Este conjunto le debe su nombre a la condición de estar formado por las razones (divisiones, representadas en **fracciones**) de números enteros. Su símbolo es \mathbb{Q} y es representado como;

$$\mathbb{Q} = \left\{ \frac{p}{q} \mid p, q \in \mathbb{Z}; q \neq 0 \right\},$$

donde p es llamado **numerador**, q **denominador** y el símbolo \in se lee como “pertenece a”.



2. Conceptos básicos

Este conjunto lo forman los números enteros, además de los *números decimales*. Éstos pueden ser decimales *exactos*, en caso de que la parte decimal sea finita, o decimal *periódico*, cuando la *fracción* de los enteros genere una cantidad infinita de números decimales formada por secuencias periódicas.

Así, si ejemplos de números racionales exactos pueden ser estos:

$$-\frac{4}{5} = -0,8$$

$$\frac{190}{32} = 5,9375$$

Los números racionales periódicos, se muestran así:

$$\frac{10}{6} = 1,66666 \dots = 1,\hat{6} \quad -\frac{108}{63} = 1,7142857142857 \dots = 1,7\overbrace{14285}$$

Cuando se identifica el periodo de la secuencia decimal, se puede representar encerrando éste con el símbolo $\hat{}$. Por todo esto, este conjunto es cerrado a la división, además de a la suma, resta y multiplicación.

Para continuar, se debe entender primero la propiedad de *equivalencia* en las fracciones. Dos fracciones serán equivalentes si representan el mismo número racional y podrán ser reconocidos si se encuentra un número natural tal que, multiplicando o dividiendo tanto a numerador como denominador de la primera fracción, dé como resultado la otra fracción.

Al realizar esta operación con el mismo número en el denominador y en el numerador, realmente se está multiplicando la fracción por la unidad, 1, ya que cualquier fracción del tipo (x/x) es igual a 1 y por la propiedad matemática del elemento unidad, cualquier número multiplicado o dividido por 1 permanece invariable.

Pizza

Se encargan dos pizzas familiares para cuatro amigos. Una se divide en cuatro porciones y la otra en doce.



¿Por qué debo comer una porción de la primera, pero tres de la segunda para comer la misma cantidad en cada una?



Porque $1/4$ y $3/12$ son equivalentes. La porción de una pizza es el triple que una de la otra

$$\frac{1^{(x3)}}{4^{(x3)}} = \frac{3}{12}$$



2. Conceptos básicos

Este fenómeno ocurre debido a la propia condición de ‘decimal’ que, al igual que el sistema numérico, está basado en diez dígitos. Por este motivo, cada vez que se intente fraccionar un número entre otro distinto o no equivalente de 10 o múltiplo de 10 (esto es, 10, 20, ..., 100, etc.), la división producirá un número racional periódico.

En el caso anterior,

$$\frac{1}{4} * \frac{3}{3} = \frac{3}{12} * \frac{5}{5} = \frac{25}{100} \quad \leftrightarrow \quad \frac{1}{4} * 1 = \frac{3}{12} * 1 = \frac{25}{100}$$

2.1.4. Números irracionales

Si los números racionales tenían la cualidad de expresarse como una razón de números enteros, los irracionales serán aquellos que no puedan. Es decir, es el conjunto de números caracterizados por tener infinitas cifras decimales no periódicas. El conjunto se representa con el símbolo \mathbb{I} .

De entre los números irracionales más conocidos, se encuentra el número π , que fue descubierto para relacionar la longitud de una circunferencia con su diámetro.

$$\pi = 3,141592653589 \dots$$

Otros conocidos son el número e , que aparece en procesos naturales de crecimiento o la desintegración radiactiva, y Φ , llamado áureo o número de oro, obtenida de una proporción encontrada en elementos naturales como las espirales de los caracoles y las piñas o en los pétalos de las flores.

El 14 de marzo de 2019 (el día de π), Emma Haruka Iwao, empleada de Google, calculó el valor más preciso para π conocido hasta la fecha. Usando 25 máquinas virtuales y 121 días desglosó el número en 31,4 billones de decimales.

2.1.5. Números reales

Finalmente, estos números reales forman un conjunto denotado como \mathbb{R} y formados por los números racionales y los irracionales. Así,

$$\mathbb{R} = \mathbb{Q} \cup \mathbb{I},$$

donde el símbolo U representa la unión.

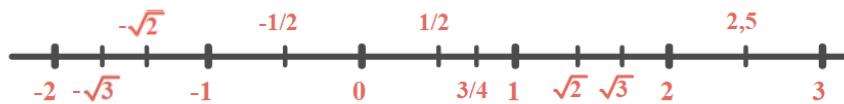
Los números reales no comprenden todos los números existentes, ya que faltarían los llamados números complejos o imaginarios, pero no son objeto de estudio aquí.



2. Conceptos básicos

2.2. Intervalos

Un intervalo es la representación de un subconjunto de números reales. Dicho de otro modo, si los números reales son todos aquellos que pueden ser representados en una misma recta, la *recta real*, un intervalo es un segmento de esta recta.



Un **intervalo abierto** (a, b) es aquel subconjunto formado por todos los números reales **estrictamente mayores** que a y **estrictamente menores** que b . Estas *desigualdades estrictas* se denotan con los símbolos $<$ y $>$ y el intervalo que encierra se marcan con paréntesis. Así,

$$(a, b) = \{x \in \mathbb{R} \mid a < x < b\}$$

Gráficamente, se pintan sobre una recta con los extremos a y b redondeados sin relleno. En la recta anterior, un intervalo abierto podría ser el mostrado abajo e incluiría a todos los números reales entre -1 y 1 , pero **no a estos**.



Por su parte, un **intervalo cerrado** $[a, b]$ es el subconjunto formado por todos los números reales **mayores o iguales** que a y **menores o iguales** que b . Estas *desigualdades amplias*, \leq y \geq , se denominan *desigualdades amplias*, incluyen a los extremos del intervalo y son representados mediante corchetes.

$$[a, b] = \{x \in \mathbb{R} \mid a \leq x \leq b\}$$

En este caso, los extremos cerrados del intervalo son representados sobre la recta como los círculos rellenos:



Existen intervalos en los que un extremo es cerrado, y por lo tanto incluye a su valor en él, y el otro extremo es abierto, por lo que queda fuera del intervalo; son llamados *intervalos semiabiertos*. En cada caso, se indicará el lado abierto en el propio nombre.



2. Conceptos básicos

Así, un **intervalo semiabierto por la izquierda** sería representado como:

$$(a, b] = \{x \in \mathbb{R} \mid a < x \leq b\}$$

cuya recta en el ejemplo anterior quedaría así:



Y los **intervalos semiabiertos por la derecha** como:

$$[a, b) = \{x \in \mathbb{R} \mid a \leq x < b\}$$



Finalmente, los **intervalos infinitos** son aquellos que contienen el valor infinito, ∞ , en uno o ambos extremos. Debe prestarse atención al hecho de que el extremo infinito siempre deberá ser abierto, por su misma condición de inalcanzable.

$$(a, \infty) = \{x \in \mathbb{R} \mid x > a\}$$



$$(-\infty, b] = \{x \in \mathbb{R} \mid x \leq b\}$$



2.3. Polinomios

Para poder continuar, es necesario introducir mínimamente al lector en el álgebra. Y es que esta rama de las matemáticas nace por el interés en conocer ciertos valores, cantidades, números, a partir de reglas y/o relaciones conocidas. Por esta razón, además de operadores y números, se usan letras para indicar valores desconocidos, incógnitas o **variables**.

Estas reglas están formadas, en su forma más básica, por los llamados polinomios, con su versión unitaria, el *monomio*, que tiene la siguiente forma:

$$-3x^2,$$

estando compuesto por un *signo*, un valor constante llamado *coeficiente*, una *variable* y un *exponente*.



2. Conceptos básicos

Debe entenderse que estos elementos enumerados estarán siempre presentes en un monomio ya que:

- si no aparece ningún signo, éste será un signo positivo,
- si no aparece coeficiente, éste será igual a 1 (elemento identidad en la multiplicación),
- si no aparece variable, ésta será x^0 , ya que cualquier número elevado a 0 es igual a 1
- y si no aparece exponente, éste será igual a 1, pues un número elevado a 1 es él mismo.

Un polinomio está formado por varios monomios unidos por sumas o restas y denominados términos; el caso particular de un polinomio de dos términos es llamado *binomio*. Así;

$$4x - 3x^2,$$

sería un binomio formado por un primer término con signo positivo, coeficiente igual a 4, variable x y exponente 1; y un segundo término de signo negativo, coeficiente igual a 3, variable x y exponente 2.

Un turista está de viaje en los Estados Unidos, encuentra un termómetro urbano y ve que todo está en grados Fahrenheit. De vuelta en el hotel le pregunta sobre ese símbolo al recepcionista y le explica que son grados Fahrenheit y que restándole 32 y multiplicándolo todo por cinco novenos obtiene los grados Celsius; se lo apunta para la próxima vez que salga:

$$(x - 32) * \frac{5}{9} = \frac{5}{9}x - \frac{5 * 32}{9} = 0,5x - 17,7$$

Cuando vuelve a pasar por un termómetro y lee 35.6°F descubre que hace... $0,5 * 35,6 - 17,7 = 2^\circ C!$

2.4. Función matemática

En el ejemplo anterior, el turista en Estados Unidos descubre la manera de calcular los grados Celsius *en función* de la temperatura en grados Fahrenheit por lo que la temperatura en grados Celsius es la incógnita. Pero hasta el momento en que observa la temperatura actual en grados Fahrenheit, ¡Ésta también es una incógnita!



2. Conceptos básicos

La temperatura en grados Fahrenheit es una incógnita denominada **variable independiente**, pues depende tan sólo de la *observación*. Por su parte, los grados Celsius dependen de la fórmula que lo relaciona con la primera, por lo que se denomina **variable dependiente**. De esta relación surge la **función matemática** como la correspondencia que asocia a cada elemento de un determinado subconjunto de números reales (*observación*), llamado dominio, otro número real. Esto se representa matemáticamente como:

$$f: D \rightarrow \mathbb{R}$$

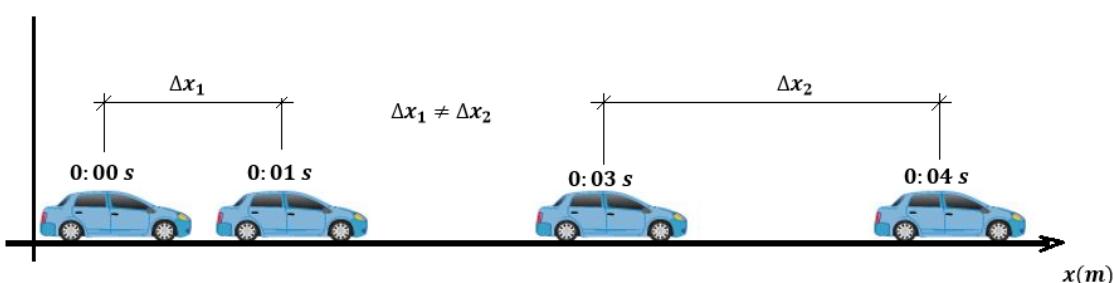
Si la variable independiente es representada como x y el conjunto de valores que puede adoptar se denomina dominio, y la variable dependiente como y y el conjunto de valores que adopte es llamado **imagen**, se dice que y es *función de x* y la fórmula anterior quedaría así;

$$y = f(x) = (x - 32) * \frac{5}{9}$$

2.5. Incrementos

En ocasiones, la incógnita a resolver no será la variable dependiente en sí, sino el incremento que sufre o haya sufrido esa variable. Para estos casos existen los *incrementos*, los cuales se definen con la letra griega delta, Δ . En muchos casos, las variaciones de una variable no serán constantes, sino que dependerán de otra variable y por lo tanto no se podrá generalizar la variación de ésta a partir de un incremento puntual.

Como ejemplo, cuando se pisa el acelerador de un coche, éste acelera provocando que el incremento de la distancia x , Δx , sufrido en un intervalo de tiempo T_1 sea distinto (menor si está acelerando) que, en el mismo intervalo en un instante posterior, T_2 . Éste es el motivo por el que surge el **diferencial**, δx , y se define como un incremento infinitesimal; el ocurrido, para el ejemplo del coche, en un intervalo infinitamente pequeño de tiempo.





2. Conceptos básicos

2.6. Límites

Finalmente, el límite de una función $f(x)$ en x_0 es el valor al que se acercan las imágenes cuando los puntos del dominio se aproximan a x_0 . Se representa como L y se denomina límite de $f(x)$

$$L = \lim_{x \rightarrow x_0} f(x)$$

Así, en los ejemplos

$$\lim_{x \rightarrow \infty} \frac{3}{x} = 0 \quad \text{y} \quad \lim_{x \rightarrow \infty} x^2 = \infty$$

se busca el valor al que se aproxima el monomio $\frac{3}{x}$ cuando x tiende a infinito, lo que es lo mismo, es un número inmenso; o al que se aproxima x^2 .

Los límites matemáticos son una herramienta muy útil para representar ‘incongruencias’ matemáticas dentro del cálculo infinitesimal, como las *formas indeterminadas*, así como diversos métodos para calcularlas, pero no son objeto de estudio en este tema.

La siguiente función conduce a una forma indeterminada cuando $x = 1$

$$f(x) = \frac{x^3 - 1}{x - 1} \rightarrow f(1) = \frac{1^3 - 1}{1 - 1} = \frac{0}{0}$$

Pero si se estudia el caso como la tendencia de x a acercarse al valor 1:

x	$f(x)$
0,9	$\frac{0,9^3 - 1}{0,9 - 1} = 2,71$
0,99	$\frac{0,99^3 - 1}{0,99 - 1} = 2,9701$
0,999	$\frac{0,999^3 - 1}{0,999 - 1} = 2,997001$



2. Conceptos básicos

A medida que el valor de la variable independiente x se aproxima a 1, su imagen o variable dependiente lo hace a 3, por lo que

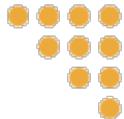
$$\lim_{x \rightarrow 1} f(x) = \lim_{x \rightarrow 1} \frac{x^3 - 1}{x - 1} = 3$$

Gracias a estos límites matemáticos, pudo ser resuelta la paradoja de Zerón, dos mil años después de ser planteada. Aunque Aquiles siempre pueda recorrer una distancia menor que la anterior antes de alcanzar a la tortuga, fue demostrado que la suma de infinitos números puede tener un valor finito cuando los elementos de esta suma decrecen.

Together for Tomorrow!
Enabling People

Sucesiones y series

#TecnologíaConPropósito

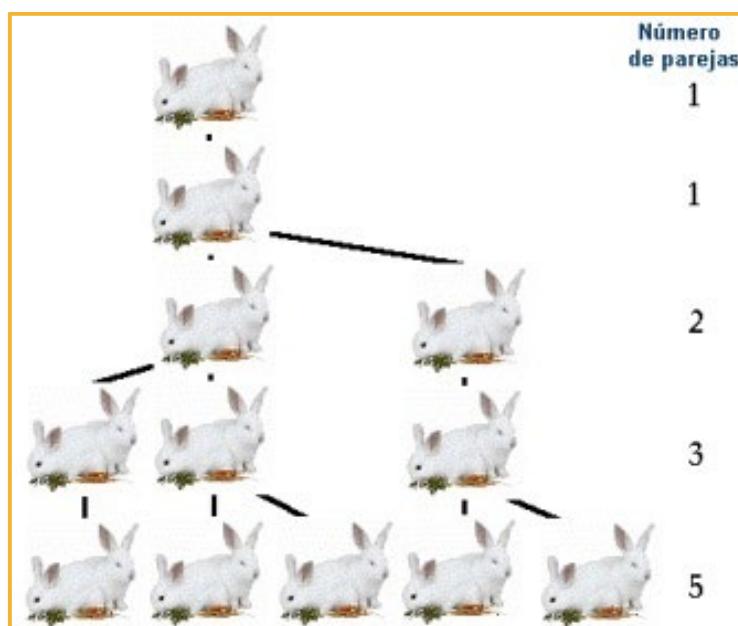


1. Introducción

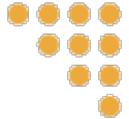
Uno de los principales impulsores de las matemáticas fue la curiosidad inherente a la condición del ser humano que necesita entender y modelar la realidad que lo rodea. A partir de la observación de los procesos de la naturaleza, se buscaba algún tipo de similitud o relación entre estos y, en muchas ocasiones, se encontraban patrones secuenciales que debían de estar regidos por alguna regla.

De este modo, un matemático italiano del siglo XII, Leonardo de Pisa (1170 - 1240), observando la cría de conejos intentó modelar esta reproducción en el tiempo, tomando como referencia parejas: comenzó teniendo 1 pareja que cruzó y al mes siguiente daba luz a la segunda pareja, momento en el que cruzaba de nuevo a la primera pareja. Al siguiente mes, la primera pareja volvía a dar a luz y ya sumaban 3 parejas; se volvía a cruzar la primera pareja y, esta vez, la segunda también. Al mes siguiente, la primera y la segunda pareja daban a luz y sumaban dos parejas más, lo que hacían 5 parejas; se cruzaban la primera, segunda y tercera pareja...

Esto conducía a una secuencia del tipo 1, 2, 3, 5, 8, 13, 21, ... cuyo elemento en una posición determinada se obtenía mediante la suma de los dos elementos anteriores. Esto es $5 + 8 = 13$. Fue llamada la sucesión de Fibonacci y es una de las sucesiones matemáticas más importantes que existen, con aplicaciones en matemáticas, teoría de juegos o ciencias de la computación.



No solo eso, el cociente de dos números consecutivos de Fibonacci, cuando la posición de estos elementos tiende a infinito, se aproxima a uno de los números irrationales más importantes y presentes en la naturaleza, la relación áurea.



2. Sucesiones

Se llama **sucesión de números reales** a toda función del tipo

$$f: \mathbb{N} \rightarrow \mathbb{R},$$

en el que a todo número natural n le corresponde un número real. Esta sucesión, como cualquier otra función, guarda una relación matemática entre los diferentes valores naturales y sus imágenes reales. Una sucesión cualquiera podría ser

$$1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \dots$$

en el que

$$a(1) = 1, \quad a(2) = \frac{1}{2}, \quad a(3) = \frac{1}{3}, \quad a(4) = \frac{1}{4}, \dots$$

Cada elemento de la sucesión es llamado *término*. El enésimo término de la sucesión se denomina *término general de la sucesión*.

En lugar de $a(1), a(2), \dots$, las sucesiones suelen ser representadas como

$$\{a_1, a_2, a_3, \dots, a_n\} \text{ ó } \{a_n\}$$

Entonces, la sucesión anterior $\{a_n\}$ es representada por su término general como $a_n = 1/n$

2.1. Monotonía y acotamiento

Se define una **sucesión monótona** como aquella que lleva una progresión creciente o decreciente. Es decir

$$\{a_n\} \text{ es monótona creciente} \leftrightarrow a_n \leq a_{n+1} \quad \forall n$$

$$\{a_n\} \text{ es monótona decreciente} \leftrightarrow a_n \geq a_{n+1} \quad \forall n$$

$$\{a_n\} \text{ es monótona estrictamente creciente} \leftrightarrow a_n < a_{n+1} \quad \forall n$$

$$\{a_n\} \text{ es monótona estrictamente decreciente} \leftrightarrow a_n > a_{n+1} \quad \forall n$$

donde el símbolo \forall debe ser leído como 'para cualquier'.

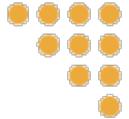
Así, la sucesión anterior $a_n = 1/n$ es monótona estrictamente decreciente.

Una sucesión $\{a_n\}$ es **acotada** si existe un número real tal que el conjunto de valores de la sucesión nunca alcanza. Es decir,

$$\{a_n\} \text{ está acotada superiormente} \leftrightarrow \exists M \in \mathbb{R} / a_n \leq M \quad \forall n$$

$$\{a_n\} \text{ está acotada inferiormente} \leftrightarrow \exists m \in \mathbb{R} / a_n \geq m \quad \forall n$$

$$\{a_n\} \text{ está acotada} \leftrightarrow \exists K \in \mathbb{R}^+ / |a_n| \leq K \quad \forall n$$



2. Sucesiones

donde el símbolo \exists significa ‘existe’, / significa ‘tal que’ y el superíndice + en \mathbb{R}^+ hace referencia a los valores reales positivos. El operador $|x|$ es el *valor absoluto* y convierte el valor encerrado en positivo. Esto es, por ejemplo, $|4 - 7| = |-3| = 3$

La sucesión $\{a_n\} = 1, \dots, n$ es acotada inferiormente por $m = 1$. Sin embargo, no es acotada superiormente.

La sucesión $\{a_n\} = 1, \frac{1}{2}, \dots, \frac{1}{n}$ es acotada superiormente por $M = 1$. Sin embargo, no es acotada inferiormente.

La sucesión $\{a_n\} = 2, \frac{3}{2}, \frac{4}{3}, \dots, \frac{n+1}{n}$ es acotada, con $K = 2$. Más exactamente, puede acotarse entre los valores 1 y 2

La sucesión $\{a_n\} = 2, -4, 8, -16, \dots, (-1)^{n-1} * 2^n$ **no** es acotada.

2.2. Convergencia

Definición: Sea $\{a_n\}$ una sucesión de números reales y sea $l \in \mathbb{R}$. $\{a_n\}$ converge a l y se escribe $\{a_n\} \rightarrow l$ cuando, para cada número real y positivo ε , existe un número natural m , de forma que se tenga $|a_n - l| < \varepsilon$ para cualquier n natural que verifique $n \geq m$. Esto es

$$\{a_n\} \rightarrow l \Leftrightarrow \forall \varepsilon \in \mathbb{R}^+ \exists m \in \mathbb{N} / n \geq m \rightarrow |a_n - l| < \varepsilon$$

Dicho de otro modo, cuando una sucesión converge a l se está diciendo que siempre habrá un número natural m a partir del cual la sucesión esté siempre a una distancia del número l menor que ε ; es decir, $\{a_n\}$ se *aproxima a l* .

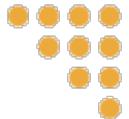
Usando como ejemplo la sucesión dada por $a_n = \frac{n+1}{n}$, se tiene

$$\{a_n\} = 2, 1.5, 1.\hat{3}, 1.25, 1.2, 1.1\hat{6}, 1.14, \dots, 1.000000111, \dots$$

De aquí es fácilmente observable que la sucesión va acercándose hacia el 1. Aunque no es demotración suficiente, se puede hacer la prueba eligiendo aleatoriamente algunos valores para ε y ver qué sucede:

Si se establece, por ejemplo, $\varepsilon = 4$, se puede comprobar cómo la desigualdad $|a_n - 1| < 4$ se cumple para toda la sucesión, esto es, para $m = 1$

Si se establece $\varepsilon = 0.3$, $|a_n - 1| < 0.3$ se cumple a partir del cuarto término, $m = 4$



2. Sucesiones

Si se establece $\varepsilon=0.0000001$ debería existir un valor m tal que, a partir del m -ésimo término (de a_m) se cumpliese $|a_n - 1| < 0.0000001$. Mediante el cálculo de la inecuación

$$|a_n - 1| = \left| \left(\frac{n+1}{n} \right) - 1 \right| = \left| \frac{n+1}{n} - \frac{n}{n} \right| = \left| \frac{1}{n} \right| < 0.0000001 \rightarrow n > 10000000$$

por lo que se obtiene $m = 10000001$.

Resumiendo, una sucesión será convergente si sus términos van aproximándose más y más a un valor sin llegar jamás a alcanzarlo.

A partir de esta última definición la condición de convergencia puede reescribirse como

$$\exists l \in \mathbb{R} / \lim_{n \rightarrow \infty} \{a_n\} = l$$

Una sucesión es **convergente** siempre que posee **límite finito**.

Asimismo, se dice que una sucesión $\{a_n\}$ es **divergente** si posee límite infinito:

$$\lim_{n \rightarrow \infty} |a_n| = +\infty$$

Si no cumple ninguna de las dos condiciones anteriores, es decir, no es ni convergente ni divergente, se le denomina **oscilante**.

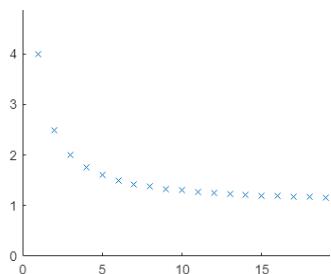
Teorema de Acotación

Toda sucesión $\{a_n\}$ convergente es acotada.

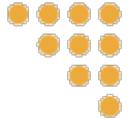
Teorema de Weierstrass

Toda sucesión $\{a_n\}$ monótona y acotada es convergente. Toda sucesión monótona y no acotada es divergente.

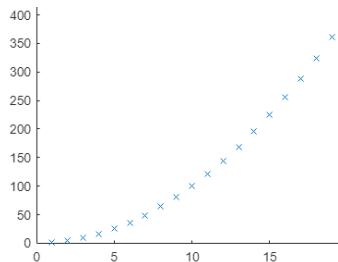
Pero ¿cómo puede una sucesión estar acotada sin converger? Las siguientes representaciones muestran algunos ejemplos.



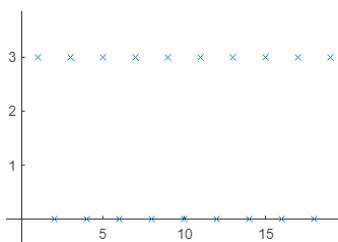
La representación en la gráfica de la izquierda muestra la sucesión $\{a_n\} = 1 + \frac{3}{n}$. Esta sucesión es monótona decreciente y acotada superiormente por $M = 4$ e inferiormente por $m = 1$. Además, convergente a m .



2. Sucesiones



La siguiente gráfica muestra la sucesión $\{a_n\} = n^2$, la cual es monótona creciente, está acotada sólo inferiormente en $m = 0$ y es divergente.



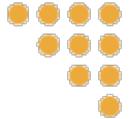
La última gráfica corresponde con la sucesión $\{a_n\} = 3 * [1 + (-1)^n]$, acotada superiormente por $M = 3$ e inferiormente por $m = 0$ pero no es convergente ni divergente, por lo que es oscilante.

Infinitésimo

Toda sucesión $\{a_n\}$ es un infinitésimo si $\lim_{n \rightarrow \infty} \{a_n\} = 0$

Infinito

Toda sucesión $\{a_n\}$ es un infinito si $\lim_{n \rightarrow \infty} \{a_n\} = \pm\infty$



3. Series

Dada una sucesión de números reales $\{a_n\}$ se define su **serie** de números reales como

$$\sum_{n=1}^{\infty} a_n = a_1 + a_2 + a_3 + \cdots + a_n + \cdots$$

la suma infinita de los términos de la sucesión.

El símbolo Σ se llama sumatorio y $\sum_{n=i}^k n$ representa la suma de todos los términos desde el entero i hasta el k . Por ejemplo,

$$\sum_{n=4}^6 3n^5 = 3 * 4^5 + 3 * 5^5 + 3 * 6^5$$

Su **suma parcial n -ésima** es

$$S_n = a_1 + a_2 + \cdots + a_n$$

por lo que es fácilmente deducible que

$$\lim_{n \rightarrow \infty} S_n = \sum_{n=1}^{\infty} a_n$$

También, el **resto n -ésimo** de la serie $\sum_{n=1}^{\infty} a_n$ es

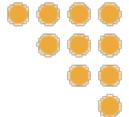
$$R_n = a_{n+1} + a_{n+2} + \cdots = \sum_{k=1}^{\infty} a_{n+k}$$

y así

$$\sum_{k=1}^{\infty} a_k = S_n + R_n$$

3.1. Propiedades

- (1) Dada una serie $\sum_{n=1}^{\infty} a_n$, la supresión o inclusión de un número finito de términos no altera su condición convergente o divergente.
- (2) Dadas las series $\sum_{n=1}^{\infty} a_n$ y $\sum_{n=1}^{\infty} b_n$ convergentes, que $\sum_{n=1}^{\infty} a_n$ converge al número real A y $\sum_{n=1}^{\infty} b_n$ converge al número real B , entonces
 - $\sum_{n=1}^{\infty} (a_n \pm b_n) = \sum_{n=1}^{\infty} a_n \pm \sum_{n=1}^{\infty} b_n = A \pm B$



3. Series

- $\sum_{n=1}^{\infty} (a_n b_n) \neq (\sum_{n=1}^{\infty} a_n)(\sum_{n=1}^{\infty} b_n) = A * B$
- $\sum_{n=1}^{\infty} (\lambda a_n) = \lambda * \sum_{n=1}^{\infty} a_n = \lambda * A$

(3) Condición **necesaria** (pero no suficiente) de convergencia

Si $\sum_{n=1}^{\infty} a_n$ es convergente entonces $\lim_{n \rightarrow \infty} a_n = 0$

(4) Criterio de convergencia **del cociente**

Dada la serie $\sum_{n=1}^{\infty} a_n$ que cumple:

$$\lim_{n \rightarrow \infty} \frac{a_n}{a_{n-1}} = L \quad \text{ó} \quad \lim_{n \rightarrow \infty} \frac{a_{n+1}}{a_n} = L$$

Entonces si

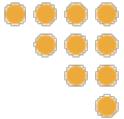
- Si $L < 1 \rightarrow$ la serie $\sum_{n=1}^{\infty} a_n$ es convergente
- Si $L > 1 \rightarrow$ la serie $\sum_{n=1}^{\infty} a_n$ es divergente

Gracias a la última propiedad mencionada, conocer la convergencia de una serie es posible. Tan solo habrá que resolver el límite presentado como condición necesaria y suficiente. Pero ¿cómo se tratan límites que tienden a infinito? A priori deben entenderse varias asunciones que son propiedades del infinito (o de infinitésimos):

Propiedades en límites		
$\infty \pm k = \infty$	$\infty + \infty = \infty$	$\infty - \infty = Ind$
$\infty * k = \infty \quad (si \ k \neq 0)$	$\infty * \infty = \infty$	$0 * \infty = Ind$
$0/k = 0$	$0/\infty = 0$	$0/0 = Ind$
$k/0 = \infty$	$k/\infty = 0$	$\infty/\infty = Ind$
$\infty/k = \infty$	$\infty/0 = \infty$	$0^0 = Ind$
$0^k = 0 \quad (si \ k > 0)$	$0^\infty = 0$	$1^\infty = Ind$
$0^k = \infty \quad (si \ k < 0)$	$k^\infty = \infty \quad (si \ k > 1)$	$\infty^0 = Ind$
$k^0 = 1$	$k^\infty = 0 \quad (si \ 0 < k < 1)$	

'Ind' se refiere a forma indeterminada

El método más aceptado para la resolución de límites es conocido como **regla de L'Hôpital**, ya que da una solución generalizada a límites en estas **formas indeterminadas**. Sin embargo, para trabajar la regla de L'Hôpital es necesario conocer el cálculo de derivadas, por lo que en este tema se usará un método que es factible para los casos más simples.



3. Series

Dada la sucesión $\{a_n\}$ de término general

$$a_n = \frac{2^n}{n!},$$

si se requiere conocer su convergencia, se debe hacer uso del criterio arriba mencionado

$$\lim_{n \rightarrow \infty} \frac{a_{n+1}}{a_n} = L,$$

Antes de continuar, para el lector que no esté familiarizado con el operador $!$ y/o con las fracciones compuestas, se debe saber que:

- $!$ es la operación *factorial* y se define como el producto de todos los números enteros positivos desde el 1 hasta el valor con el que se opera. Esto es, por ejemplo

$$5! = 5 * 4 * 3 * 2 * 1 = 120$$

excepción de $0!$, que es siempre igual a 1.

- Una fracción compuesta tipo $\frac{a/b}{c/d}$ cumple lo siguiente

$$\frac{\frac{a}{b}}{\frac{c}{d}} = \frac{a * d}{b * c}$$

Entonces, aplicando el criterio al caso presentado, queda como

$$\lim_{n \rightarrow \infty} \frac{a_{n+1}}{a_n} = \lim_{n \rightarrow \infty} \frac{\frac{2^{n+1}}{(n+1)!}}{\frac{2^n}{n!}} = \lim_{n \rightarrow \infty} \frac{2^{n+1} * n!}{2^n * (n+1)!}$$

De las propiedades de las potencias, debe saberse que $x^a * x^b = x^{a+b}$. Por lo tanto, $2^{n+1} = 2^n * 2^1 = 2^n * 2$.

Asimismo, si

$$(n+1)! = (n+1) * \underbrace{n * (n-1) * (n-2) * \dots * 1}_{= n!} = (n+1) * n!$$

Todo esto junto a la aplicación de alguna de las propiedades de infinito mostradas en la tabla anterior, se tiene

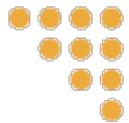
$$\lim_{n \rightarrow \infty} \frac{2^{n+1} * n!}{2^n * (n+1)!} = \lim_{n \rightarrow \infty} \frac{2 * 2^n * n!}{2^n * n! * (n+1)} = \lim_{n \rightarrow \infty} \frac{2}{n+1} = \frac{2}{\infty + 1} = \frac{2}{\infty} = 0$$

Entonces, se tiene que $L = 0$ y, puesto que $L < 1$, se pueden concluir diciendo que la serie $a_n = \frac{2^n}{n!}$ converge.

Together for Tomorrow!
Enabling People

Derivadas

#TecnologíaConPropósito



1. Introducción

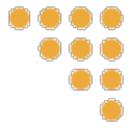
Siguiendo la línea de estudio del cálculo infinitesimal, que no solo trata infinitos e infinitesimales sino también las variaciones instantáneas —la rapidez o lentitud con la que una magnitud está variando en un determinado instante—, aparece el concepto estrella de las matemáticas, la derivada.

Esta idea abrió una nueva puerta en la ciencia, especialmente en la física, donde, mediante cálculo diferencial, era posible estudiar el comportamiento mecánico de los elementos de la naturaleza; desde la caída de un objeto desde una cierta altura, hasta el vaivén de las olas o la afluencia del agua en distintos tramos de un río.

Las derivadas permitieron finalmente ver más allá del básico análisis estático de la realidad, de correlacionar adecuadamente el espacio y el tiempo y así poder hablar de términos tan corrientes como la velocidad o la aceleración. Del mismo modo, es una pieza fundamental en la resolución de problemas de optimización y aprendizaje automático, donde el principal objetivo es encontrar el momento en que el valor de una función es el máximo posible.

Si el análisis estático permite conocer los valores de las imágenes de una función ante determinados valores en su(s) variable(s) de entrada dentro del dominio, mediante las derivadas es posible conocer la dinámica que rige el comportamiento de estas imágenes. ¿Crecen siempre? ¿Cuándo dejan de crecer para comenzar a disminuir? ¿Tiene más pendiente la curva en el punto *A* o en el *B*? ¿Dónde se encuentra el máximo valor? ¿Y el mínimo?

Todas estas preguntas pueden ser respondidas gracias al cálculo de derivadas.

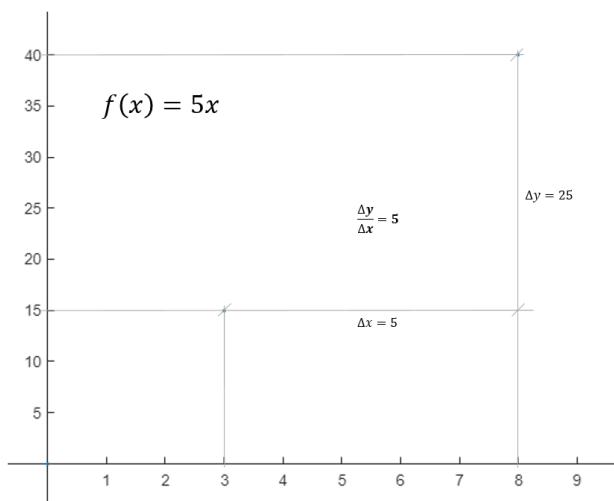


2. Derivadas

2.1. Función derivada

La derivada puede ser interpretada como el cálculo de la *variación media* o incremento unitario, $\Delta y/\Delta x$, que sufre una función en un intervalo infinitesimal del dominio de x . La variación media es aquella que experimenta y ante un cambio unitario de x .

Así, si el incremento de las imágenes de una función $f(x) = 5x$ en el intervalo $x \in [x, x + \Delta x] = [3, 8]$ es igual a $\Delta y = 25$, su variación media será $VM = \frac{\Delta y}{\Delta x} = \frac{25}{5} = 5$



Pero la variación media no es información suficiente para conocer cómo se comportan las imágenes y dentro del dominio de x . La función $f(x) = 5x$ es lineal, —forma una recta—, y por lo tanto su VM no variará a lo largo de su dominio. Pero si se analizase, por ejemplo, la función polinómica

$$f(x) = x^3 - 5x^2 + 5,$$

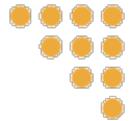
se tendrían las imágenes $f(0) = 5, f(4) = -11$, lo que resultaría en

$$VM = \frac{\Delta y}{\Delta x} = \frac{-11 - 5}{4 - 0} = \frac{-16}{4} = -4$$

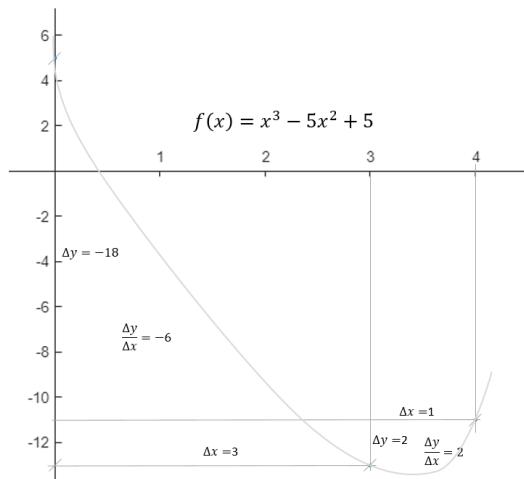
en el intervalo $[0,4]$. Pero si se dividiese ahora el intervalo $[0,4]$ en dos, $[0,3] \cup [3,4]$, los intervalos tendrían, respectivamente, como variación media

$$VM^{[0,3]} = \frac{\Delta y}{\Delta x} = \frac{f(3) - 5}{3 - 0} = \frac{-13 - 5}{3} = -6$$

$$VM^{[3,4]} = \frac{\Delta y}{\Delta x} = \frac{-11 - f(3)}{4 - 3} = \frac{-11 - (-13)}{1} = 2$$



2. Derivadas



Por esta falta de linealidad, se debe reducir el intervalo en x al mínimo para conocer la variación que sufre una función en un punto determinado x_0 . Esto es

$$\lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} \Big|_{x_0} = \lim_{\Delta x \rightarrow 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

Esta variación instantánea se denomina **derivada de $f(x)$** y se representa como $f'(x)$

Si la derivada de una función a su vez también puede variar en función del valor de x sobre el que se esté analizando, quiere decir que la función derivada de x , $f'(x)$, también puede tener derivada. Así, la función derivada de una función derivada es llamada *derivada segunda de x* y se denota por $f''(x)$; la función de ésta es la *derivada tercera de x* , $f'''(x)$. Para derivadas de orden superior a 3, se escribe $f^{(n)}(x)$, aunque también se puede encontrar n en números romanos.

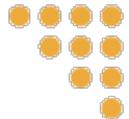
2.2. Continuidad y derivabilidad

Definición: Dada una función $f(x)$, se dice que es **continua** en el punto $x = x_0$ si, y solo si, cumplen lo siguiente:

- La función está definida en x_0 . Esto es, en $x = x_0$, existe un valor para $f(x_0)$.
- Los límites laterales de $f(x)$ cuando $x \rightarrow x_0$ existen y son iguales. Esto es, cuando x se aproxima a x_0 por la izquierda (desde valores ligeramente menores) y cuando lo hace por la derecha (desde valores ligeramente mayores) se obtiene el mismo valor límite L .

$$\lim_{x \rightarrow k} f(x) = \lim_{x \rightarrow k^+} f(x) = \lim_{x \rightarrow k^-} f(x) = L$$

- El anterior límite L de $f(x)$ cuando x tiende a x_0 sea igual a la imagen de x en x_0 , $y = f(x_0)$.



2. Derivadas

Para aclarar la anterior definición, se aplica al caso $f(x) = \frac{1}{x}$ en los puntos $x = 2$ y $x = 0$.

Para el primer caso en el que $x = 2$, la primera condición se cumple ya que $f(x)$ existe en $x = 2$, $f(2) = \frac{1}{2} = 0,5$. El límite de $f(x)$ cuando x tiende a 1 por la izquierda y cuando lo hace por la derecha puede ser estimado como sigue:

$$\lim_{x \rightarrow 2^-} f(x) \sim \frac{1}{1,99999} = 0,5000025 \approx 0,5$$

$$\lim_{x \rightarrow 2^+} f(x) \sim \frac{1}{2,00001} = 0,4999975 \approx 0,5$$

por lo que este segundo punto también se cumple. Como tanto la imagen en $x = 2$ como los límites laterales son iguales, se puede concluir que la función es **continua en $x = 2$** .

Para el segundo caso, donde $x = 0$, la primera condición conduce a $f(1) = \frac{1}{0}$, forma indeterminada y, por lo tanto, sin valor. Incluso si se evaluase la segunda condición, los límites laterales

$$\lim_{x \rightarrow 0^-} f(x) \sim \frac{1}{-0,001} = -1000, \quad \lim_{x \rightarrow 0^-} f(x) \sim \frac{1}{-0,00001} = -100000 \rightarrow$$

$$\lim_{x \rightarrow 0^-} f(x) = -\infty$$

$$\lim_{x \rightarrow 0^+} f(x) \sim \frac{1}{0,001} = 1000, \quad \lim_{x \rightarrow 0^+} f(x) \sim \frac{1}{0,00001} = 100000 \rightarrow$$

$$\lim_{x \rightarrow 0^+} f(x) = +\infty$$

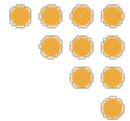
no se cumple la igualdad

$$\lim_{x \rightarrow 1^-} f(x) \neq \lim_{x \rightarrow 1^+} f(x),$$

por lo que la función **no es continua en $x = 0$** .

Definición. Dada una función $f(x)$, se dice que es **derivable** en un punto $x = x_0$ si:

- Es continua en x_0
- Existe el límite $\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0}$
- Los límites laterales de $\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0}$ existen y son iguales



2. Derivadas

Si, a modo de ejemplo, se aplican estas condiciones sobre $f(x) = |x|$ en $x_0 = 0$ se puede concluir en la no derivabilidad de esta:

- $f(x) = |x|$ en $x_0 = 0$ es igual a $f(0) = 0$, por lo que cumple la primera condición de continuidad.
- Los límites laterales de $f(x)$ cuando x tiende a 0

$$\lim_{x \rightarrow 0^-} f(x) \sim |-0,000001| = 0,000001 \approx 0$$

$$\lim_{x \rightarrow 0^+} f(x) \sim |0,000001| = 0,000001 \approx 0$$

son también 0, por lo que $f(x) = |x|$ en $x_0 = 0$, cumpliendo así la primera condición de derivabilidad

- El límite $\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0}$ existe, ya que

$$\lim_{x \rightarrow 0} \frac{f(x) - f(x_0)}{x - x_0} \sim \frac{f(0,000001) - f(0)}{0,000001 - 0} = \frac{0,000001 - 0}{0,000001} = 1$$

- Ahora bien, sus límites laterales

$$\lim_{x \rightarrow 0^+} \frac{f(x) - f(x_0)}{x - x_0} \sim \frac{f(0,000001) - f(0)}{0,000001 - 0} = \frac{0,000001 - 0}{0,000001} = 1$$

$$\lim_{x \rightarrow 0^-} \frac{f(x) - f(x_0)}{x - x_0} \sim \frac{f(-0,000001) - f(0)}{-0,000001 - 0} = \frac{0,000001 - 0}{-0,000001} = -1$$

no coinciden, por lo que **no es derivable** en $x_0 = 0$.

2.3. Propiedades y reglas de derivación

La aplicación de la definición formal de la función derivada representada en límite puede ser de ayuda para análisis de derivabilidad de una función en un punto concreto o para resolver algunas derivadas simples.

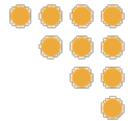
Sin embargo, existe una serie de reglas y propiedades de aplicación a las funciones para obtener su función derivada de una manera mucho más sencilla:

- **Derivada de constante.** Si la función de x es constante, del tipo $f(x) = a$, siendo a un número real, entonces su derivada es siempre cero:

$$f(x) = a \rightarrow f'(x) = 0, \quad \forall a \in \mathbb{R}$$

- **Derivada suma.** La derivada de la suma de dos funciones es la suma de sus derivadas:

$$(f(x) + g(x))' = f'(x) + g'(x)$$



2. Derivadas

- **Derivada de una potencia.** La función de un monomio del tipo $f(x) = x^a$ es:

$$f'(x) = ax^{a-1}$$

- **Derivada con coeficiente.** La derivada del producto de una constante por una función es la derivada de la función por la constante:

$$(af(x))' = af'(x)$$

- **Derivada del producto.** La derivada del producto de dos funciones es igual al producto de la derivada de la primera por la segunda función más la primera función por la derivada de la segunda:

$$(f(x) * g(x))' = f'(x) * g(x) + f(x) * g'(x)$$

- **Derivada del cociente:**

$$\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x) * g(x) - f(x) * g'(x)}{g^2(x)}$$

- **Derivada de la composición.** La derivada de $g(f(x))$ es $f'(x) * g'(f(x))$.

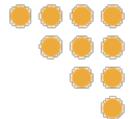
Además, existen un grupo de funciones que por su carácter especial e irregular conviene memorizar.

Estas son:

$$\begin{aligned} f(x) &= \sin(x) \rightarrow f'(x) = \cos(x) \\ f(x) &= \cos(x) \rightarrow f'(x) = -\sin(x) \\ f(x) &= \tan(x) \rightarrow f'(x) = 1 + \tan^2(x) \\ f(x) &= e^x \rightarrow f'(x) = e^x \\ f(x) &= \ln(x) \rightarrow f'(x) = \frac{1}{x} \end{aligned}$$

Sabiendo esto, podrás ser capaz de calcular la derivada de cualquier función.

En cualquier caso, es importante que se muestren varios ejemplos para conseguir una idea más clara y práctica sobre cómo aplicar las anteriores propiedades.



2. Derivadas

Ejemplo 1

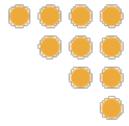
Se tiene la función $f(x) = x^5 - 6x^3 + x - 28$. Para calcular su derivada, se debe aplicar, en primer lugar, la regla de la suma.

Así, para derivar ese polinomio, que puede ser reescrito como $f(x) = f_1(x) + f_2(x) + f_3(x) + f_4(x) = x^5 - 6x^3 + x - 28$, habrá que derivar cada uno de sus términos:

- $f_1(x) = x^5$, aplicando la regla de la potencia, se tiene $f'_1(x) = 5x^4$
- De $f_2(x) = -6x^3$ se sabe que
 - o por la regla con coeficiente, $f'_2(x) = -6g'(x)$ siendo $g(x) = x^3$
 - o por la regla de la potencia, $g'(x) = 3x^2$
por lo que $f'_2(x) = -6 * 3x^2 = -18x^2$
- De $f_3(x) = x$, por la regla de la potencia y sabiendo que $x = x^1$, se tiene
$$f'_3(x) = 1x^0 = 1 * 1 = 1$$
- De $f_4(x) = -28$, por la regla de la constante, se tiene $f'_4(x) = 0$

Por lo tanto, la función derivada de $f(x) = x^5 - 6x^3 + x - 28$ es

$$f'(x) = 5x^4 - 18x^2 + 1$$



2. Derivadas

Ejemplo 2

Se tiene en este caso la función:

$$f(x) = \frac{x^2 + 3x^{3/4} + \operatorname{sen}(x^2)}{\ln(x)}$$

La cual puede descomponerse de este modo:

$$f(x) = \frac{x^2 + 3x^{3/4} + \operatorname{sen}(x^2)}{\ln(x)} = \frac{F(x)}{G(x)} = \frac{f_1(x) + f_2(x) + f_3(h(x))}{G(x)}$$

- $f_1(x) = x^2, \quad f_1'(x) = 2x$
- $f_2(x) = 3x^{5/4}, \quad f_2'(x) = 3 * \frac{5}{4}x^{(5/4-1)} = \frac{15}{4}x^{1/4}$
- $f_3(h(x)) = \operatorname{sen}(x^2)$ es la composición de $f_3(x) = \operatorname{sen}(x)$ y $h(x) = x^2$. Su derivada, por la regla de la composición, se calcula como $(f_3(h(x)))' = h'(x) * f_3'(h(x))$. La derivada de $f_3(x) = \operatorname{sen}(x)$ es $f_3'(x) = \cos(x)$ y la de $h(x) = x^2, \quad h'(x) = 2x$. Por lo tanto,

$$(f_3(h(x)))' = f_3'(h(x)) * h'(x) = 2x * \cos(x^2)$$

- Finalmente, $G(x) = \ln(x)$, se sabe que $G'(x) = 1/x$

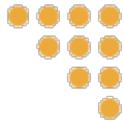
Si se recompone ahora la derivada de la función a partir de su representación más general, $f(x) = \frac{F(x)}{G(x)}$, haciendo uso de la regla del cociente, ésta queda como:

$$\begin{aligned} f'(x) &= \frac{F'(x) * G(x) - F(x) * G'(x)}{G^2(x)} = \\ &= \frac{[2x + \frac{15}{4}x^{1/4} + 2x \cos(x^2)](\ln(x)) - [x^2 + 3x^{5/4} + \operatorname{sen}(x^2)](1/x)}{(\ln(x))^2} \end{aligned}$$

Together for Tomorrow!
Enabling People

Integrales

#TecnologíaConPropósito

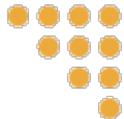


1. Introducción

De la misma manera que la función derivada consigue extraer las variaciones instantáneas que ocurren en cada intervalo infinitamente pequeño del dominio de x , llamadas diferencial o derivadas de x , $-dx$ ó ∂x –, las integrales se encargan de sumar todos esos intervalos infinitesimales dx y “reconstruir” la función original, también denominada *función primitiva*. Es por esto por lo que a las integrales se las conoce como operaciones inversas a las derivadas, ya que al *integrar* una función derivada $f'(x)$ se obtiene la función desde la que se derivó, $f(x)$.

No solo constituye una herramienta fundamental en la disciplina de cálculo infinitesimal, sino también en geometría, al calcular superficies bajo curvas $f(x)$, o volúmenes encerrados en superficies en rotación, $f(x, y)$. Más allá de las matemáticas, no sería posible hablar de funciones de distribución, crucial en problemas de probabilidad y estadística, o de redes neuronales convolucionales en inteligencia artificial sin hacer uso del cálculo integral.

En el caso del cálculo de integrales, la materia es algo más extensa y compleja que en el caso de derivadas. A lo largo de esta unidad, se presentan los distintos tipos de integrales, así como sus diferencias y aplicabilidad; también se describen los métodos de resolución de integrales simples dejando fuera del alcance de la unidad la resolución de las más complejas.



2. Integrales

Definición. Se denomina **función primitiva de $f(x)$** a aquella función cuya derivada es la misma función dada, $f(x)$.

Si se tiene la función $F(x) = x^4 - 4x^3 + 6x - 12$, su derivada será otra función $F'(x) = f(x) = 4x^3 - 12x^2 + 6$. De aquí se tiene que $F(x)$ es la función primitiva de $f(x)$.

Es importante notar que, si se tiene en cuenta la regla de derivación de una constante, en la que el término constante de un polinomio siempre es cero, la función primitiva de cualquier función $f(x)$ será cualquiera del tipo $F(x) + C$ con $C \in \mathbb{R}$. Independientemente del valor de C , se cumplirá:

$$(F(x) + C)' = F'(x) + C' = F'(x) + 0 = F'(x)$$

Haciendo uso del ejemplo anterior, si se altera el término constante -12 por cualquier otro número, $F_2(x) = x^4 - 4x^3 + 6x + 3432$, la derivada sigue siendo la misma función $f(x) = F'_2(x) = 4x^3 - 12x^2 + 6$ y, por lo tanto, esta nueva $F(x)$ también sería función primitiva de $f(x)$. Por lo tanto, una función $f(x)$ tiene infinitas funciones primitivas.

2.1. Integrales indefinidas

Definición. Se denomina **integral indefinida de $f(x)$** al conjunto de todas las funciones primitivas de ésta. Se representa como:

$$\int f(x)dx = F(x) + C, \quad \forall C \in \mathbb{R}$$

Nota: La simbología $\forall C \in \mathbb{R}$ significa “para todo valor de C perteneciente al conjunto de números reales”.

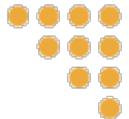
2.1.1. Propiedades

Las propiedades de las integrales indefinidas son obtenidas de las propiedades de derivación:

- **Integral suma:**

$$\int [f(x) + g(x)]dx = \int f(x)dx + \int g(x)dx$$

Nótese que toda propiedad aplicada a la suma, lo es también a la resta, la cual es una suma con números negativos $1 - 1 = 1 + (-1)$.



2. Integrales

- **Integral con coeficiente:**

$$\int a * f(x) dx = a \int f(x) dx, \quad \forall a \in \mathbb{R}$$

- **Integral de una constante.** La integral de toda función constante es igual a la constante convertida en coeficiente de x (constante multiplicada por x).

$$\int adx = ax + C, \quad \forall a, C \in \mathbb{R}$$

- **Integral de una potencia.** Si toda función tipo $f(x) = x^a$ se deriva como $f'(x) = ax^{a-1}$, la operación inversa para calcular su integral será;

$$\int x^a dx = \frac{1}{a+1} x^{a+1} + C, \quad \forall C \in \mathbb{R}$$

Efectivamente, puede comprobarse haciendo el cálculo por el camino opuesto. Si se tiene $f(x) = \frac{1}{a+1} x^{a+1}$, su derivada es:

$$f'(x) = \frac{1}{a+1} (a+1)x^{(a+1)-1} = \cancel{\frac{a+1}{a+1}}^= x^a$$

Del mismo modo que para las derivadas, existen algunos tipos de funciones cuyas integrales conviene memorizar:

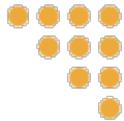
- $\int \sin(x) dx = -\cos(x) + C$
- $\int \cos(x) dx = \sin(x) + C$
- $\int e^x dx = e^x + C$
- $\int \frac{1}{x} dx = \ln(x) + C$

2.1.2. Integrales inmediatas

Son llamadas así debido a que se obtienen directamente de la propia definición de integral. La función presentada puede asociarse, directamente o mediante alguna transformación sencilla del integrando, con alguna de las transformaciones directas de derivación. Así, las integrales:

- $\int 18dx = 18x + C$
- $\int -\frac{9}{x} dx = -9 \int \frac{1}{x} dx = -9 \cdot \ln(x) + C$
- $\int 3x^2 dx = x^3 + C$

pueden calcularse inmediatamente sabiendo que la derivada de x siempre es igual a 1 ($\int 1 * dx$), que la derivada de $\ln(x)$ es $\frac{1}{x}$, o que la derivada de una potencia siempre es ax^{a-1} .



2. Integrales

En los casos que se requiere alguna transformación previa del integrando, se debe partir de la regla de la derivación de composición para integrales, es decir, la inversa de la regla. Así, si la regla de derivación de composición se escribía como:

$$[f(g(x))]' = g'(x) \cdot f'(g(x))$$

para integrar, se deberá descomponer el integrando de tal manera que se obtenga la derivada de la función interna, $g'(x)$, por un lado, y la función compuesta derivada, por el otro, $f'(g(x))$. Esto se puede ver más claramente con los siguientes ejemplos:

- $\int \frac{3x^2}{x^3} dx = \ln(x^3) + C$
- $\int 3x^2 e^{(x^3+23)} dx = e^{(x^3+23)} + C$

En el primer caso, se puede ver con facilidad que se tiene un integrando del tipo $\frac{f(x)}{g(x)}$, en el que la función f es la derivada de g . Esto es $g'(x) = (x^3)' = 3x^2$, por lo que tenemos una integral de tipo $\int \frac{g'(x)}{g(x)} dx$.

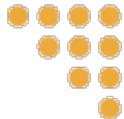
De la derivación inmediata $f(x) = \ln(x) = \frac{1}{x}$ se obtiene la de $f'(g(x)) = \ln'(g(x)) = \frac{1}{g(x)}$ y permite reescribir lo anterior como $\int g'(x) \frac{1}{g(x)} dx = \int g'(x) \cdot f'(g(x)) dx = f(g(x)) + C = \ln(x^3) + C$.

En el segundo ejemplo, aparece una función exponencial, siendo la integral completa del tipo $\int g(x)f(x)dx$, con $g(x) = 3x^2$ y $f(x) = e^{(x^3+23)}$. Es sabido que la derivada de la función exponencial es ella misma, $f'(x) = (e^x)' = e^x$, y que, para el caso compuesto quedaría como $[f(g(x))]' = (e^{g(x)})' = g'(x)e^{g(x)}$.

Al comprobar que la exponencial está siendo multiplicada por la derivada de su potencia, esto es, $g'(x) = (x^3 + 23)' = 3x^2$, se tiene que $g'(x)e^{g(x)} = 3x^2e^{(x^3+23)} = [e^{(x^3+23)} + C]'$.

Resumiendo, un primer análisis eficaz en el cálculo de una integral con fracción será analizar la posibilidad de que el numerador contenga la derivada del denominador, para trabajar sobre la hipótesis de una función primitiva tipo $f(x) = \ln(x)$.

Por otro lado, cuando la integral contiene la exponencial, es interesante analizar la posibilidad de que esté multiplicada por la derivada de su potencia.



2. Integrales

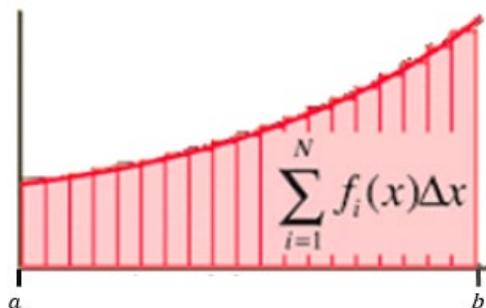
2.2. Integrales definidas

Definición. Se denomina **integral definida entre a y b de $f(x)$** a la suma de todos los diferenciales de x, dx , que encierra el intervalo $[a, b]$. a y b se denominan **límites de integración**. Se representa matemáticamente como:

$$\int_a^b f(x)dx$$

Si de la integral indefinida se obtiene otra función, o conjunto de funciones, en el caso de la integral definida de una función, se busca obtener un valor que gráficamente se traduce en el área encerrada bajo la curva que forma $f(x)$, el eje horizontal de la gráfica y las rectas verticales $x = a$ y $x = b$.

A la hora de calcular áreas irregulares se suele acudir a la simplificación del área en polígonos o formas simples. Por esto, un primer intento para calcular áreas bajo curvas de funciones fue la división del espacio en rectángulos, tal y como se presenta en la siguiente imagen.

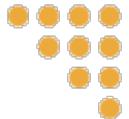


De este modo, el cálculo se reducía a una suma de N rectángulos, de misma base igual a Δx y de altura variable e igual a la imagen de $f(x)$ en algún punto dentro de cada intervalo Δx .

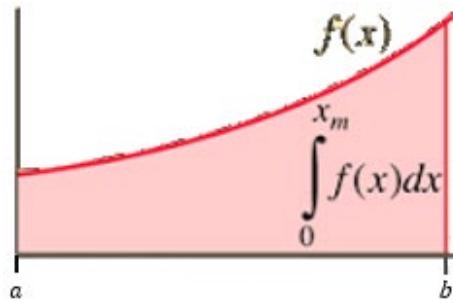
Debido a que la curva ha podido variar dentro del incremento Δx , se puede ver fácilmente cómo cuanto mayor es el incremento Δx (más anchura en los rectángulos), mayor es el error que se comete al hacer esta aproximación.

Por este motivo, conviene reducir la anchura de los rectángulos al máximo, lo que es lo mismo, reducir el incremento al orden infinitesimal.

Esto se puede representar como el límite del sumatorio inscrito en la imagen cuando el incremento de $x, \Delta x$, tiende a cero y es lo que se conoce como integral definida.



2. Integrales



$$\lim_{\Delta x \rightarrow 0} \sum_{i=1}^N f_i(x) \Delta x = \int_a^b f(x) dx$$

Si N era el número de rectángulos $N = \frac{b-a}{\Delta x}$, al aproximar a cero el incremento Δx se aproxima N a infinito, lo que deja ver nuevamente la integral como una suma de infinitos elementos infinitamente pequeños.

2.2.1. Propiedades

- Sea una integral definida del tipo $\int_a^b f(x) dx$, si sus límites de integración coinciden, $a = b$, el valor de la integral es siempre cero.

$$\int_a^a f(x) dx = 0$$

- Sea una función $f(x)$ continua y positiva para todo el intervalo $[a, b]$, su integral definida entre los límites de integración a y b será positiva.

Por el contrario, ante una función $f(x)$ continua y negativa para todo $[a, b]$, su integral definida entre a y b será siempre negativa.

$$f(x) > 0, \quad \forall x \in [a, b] \rightarrow \int_a^b f(x) dx > 0$$

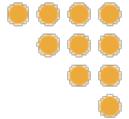
$$f(x) < 0, \quad \forall x \in [a, b] \rightarrow \int_a^b f(x) dx < 0$$

- Sea $f(x)$ y $g(x)$ dos funciones continuas en el intervalo $[a, b]$, y sean las imágenes de $f(x)$ mayores que las imágenes de $g(x)$ en todo el intervalo $[a, b]$, la integral definida de $f(x)$ será siempre mayor que la integral definida de $g(x)$ entre los límites de integración a y b .

$$f(x) \geq g(x), \quad \forall x \in [a, b] \rightarrow \int_a^b f(x) dx \geq \int_a^b g(x) dx$$

- Sea c un valor incluido en el intervalo $[a, b]$, se cumple:

$$\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx, \quad \forall c \in (a, b)$$



2. Integrales

Regla de Barrow. Sea $f(x)$ una función continua en el intervalo $[a, b]$ y sea $F(x)$ una de las funciones primitivas de $f(x)$, la integral definida de $f(x)$ entre los límites de integración a y b se calcula como

$$\int_a^b f(x)dx = F(b) - F(a)$$

Para calcular la integral definida $\int_2^3 2x dx$ es necesario primero calcular su integral indefinida, $F(x)$. Aplicándole la propiedad del coeficiente e integrando como función potencial queda:

$$\int 2x dx = 2 \int x^1 dx = 2 \left(\frac{1}{1+1} x^{1+1} \right) = \frac{2}{2} x^2 = x^2 = F(x)$$

Ahora sólo es necesario calcular las imágenes en $x = 3$ y $x = 2$ y restarlas para obtener el resultado:

$$\int_2^3 2x dx = F(3) - F(2) = 3^2 - 2^2 = 5$$

Del mismo modo, las integrales definidas $\int_0^\pi \sin(x) dx$ y $\int_6^7 \frac{dx}{x}$ quedarían como:

$$\int_0^\pi \sin(x) dx = -\cos(x)]_0^\pi = -[\cos(\pi) - \cos(0)] = -(-1 - 1) = 2$$

$$\int_6^7 \frac{dx}{x} = \ln(x)]_6^7 = \ln(7) - \ln(6) = 0.1541$$

Para calcular una integral definida del tipo $\int_1^3 (3x^4 - 4x + 4) dx$ se debe aplicar la propiedad de la suma y el cociente para conseguir simplificarlo del siguiente modo:

$$\int_1^3 (3x^4 - 4x + 4) dx = 3 \int_1^3 x^4 dx - 4 \int_1^3 x^1 dx + 4 \int_1^3 dx$$

Así factorizado, se aplican los cálculos integrales a cada término y se restan sus imágenes en los límites de integración 3 y 1:

$$\begin{aligned} \cdot \int x^4 dx &= \frac{1}{4+1} x^{4+1} = \frac{x^5}{5} & \cdot \int x^1 dx &= \frac{1}{1+1} x^{1+1} = \frac{x^2}{2} & \cdot \int dx &= x \\ \rightarrow 3 \int_1^3 x^4 dx - 4 \int_1^3 x^1 dx + 4 \int_1^3 dx &= \left[3 \frac{x^5}{5} - 4 \frac{x^2}{2} + 4x \right]_1^3 \\ &= \left(3 \frac{1}{5} (3)^5 - 2(3)^2 + 4(3) \right) - \left(3 \frac{1}{5} (1)^5 - 2(1)^2 + 4(1) \right) = 139.8 - 2.6 \\ &= 137.2 \end{aligned}$$

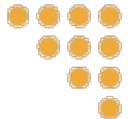
SAMSUNG

BeJob[®]

Together for Tomorrow!
Enabling People

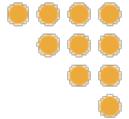
Álgebra lineal y geometría analítica

#TecnologíaConPropósito



Índice

Conceptos básicos	5
1. <u>Introducción</u>	7
2. <u>Espacios vectoriales</u>	8
2.1. <u>Definiciones</u>	8
2.2. <u>Operaciones con vectores y sus propiedades geométricas</u>	9
3. <u>Matrices</u>	12
3.1. <u>Definiciones</u>	12
3.2. <u>Operaciones con matrices</u>	14
3.3. <u>Propiedades operacionales de matrices</u>	16
3.4. <u>Representación matricial de sistemas de ecuaciones</u>	17
Representación de funciones	19
1. <u>Introducción</u>	21
2. <u>Funciones lineales</u>	22
2.1. <u>La ecuación de la recta</u>	22
2.2. <u>Tipos</u>	23
2.3. <u>Paralelismo</u>	25
3. <u>Funciones no lineales</u>	26
3.1. <u>Funciones principales</u>	26
4. <u>Funciones multivariable</u>	31
Operaciones espaciales	33
1. <u>Introducción</u>	35
2. <u>Distancias e intersecciones</u>	36



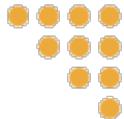
Índice

2.1. Distancia entre puntos	36
2.2. Distancia entre rectas y puntos	36
2.3. Distancia entre rectas paralelas	37
3. Análisis geométrico de funciones	38
3.1. Simetrías	38
3.2. Rectas asintóticas	38
3.3. Derivada de la curva	41
3.4. Integral de la curva	49

Together for Tomorrow!
Enabling People

Conceptos básicos

#TecnologíaConPropósito



1. Introducción

El álgebra es la disciplina de las matemáticas que trata la simbolización de relaciones numéricas y de estructuras matemáticas. En la incesante búsqueda por encontrar sentido a las cosas que ocurrían a su alrededor, el hombre formulaba hipótesis que luego tenía que demostrar mediante la expresión de teoremas y ecuaciones. El tratamiento de ecuaciones polinomiales entra dentro de lo que hoy se conoce como álgebra elemental.

Sin embargo, la palabra álgebra proviene del árabe *al-yabr* que significa “la reducción” y justifica en mayor medida al *álgebra lineal*, que será parte de la materia de este bloque. Y es que, si el álgebra elemental se encarga de la manipulación de ecuaciones relativamente sencillas de primer y segundo orden, el álgebra lineal se encarga de *reducir* sistemas de ecuaciones multivariable mediante herramientas de abstracción, como son las matrices y los vectores, y así facilitar las operaciones.

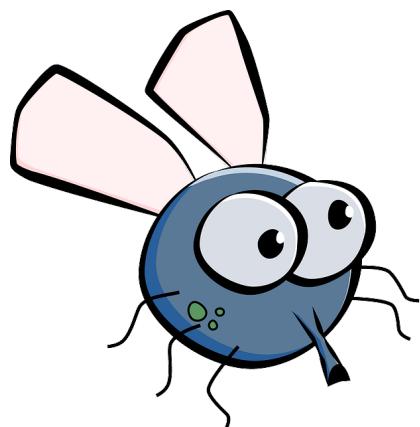
Este avance, junto a otro gran salto en las matemáticas como fue la conexión entre la geometría y esta álgebra incipiente para tratar objetos, formas, figuras, orientaciones, distancias, etc., ahora mediante las ecuaciones que la misma proveía, tuvo lugar en la Francia del siglo XVII por dos grandes científicos de la época, René Descartes y Pierre Fermat.

En esta unidad se tratarán los conceptos básicos del álgebra lineal cuya correcta comprensión será necesaria para su aplicación en las siguientes unidades del bloque.

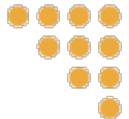
René Descartes y la mosca

René Descartes tuvo que pasar gran cantidad de horas en cama debido a su delicado estado de salud.

En cierta ocasión se quedó observando a una mosca que estaba en la habitación. Mientras tanto se preguntaba si se podría determinar a cada instante la posición que tendría el insecto, por lo que pensó que, si se conociese la distancia a dos superficies perpendiculares, en este caso la pared y el techo, se podría saber.



Entonces se levantó de la cama y cogió un trozo de papel en el que dibujó dos rectas perpendiculares: cualquier punto de la hoja quedaba determinado por su distancia a los dos ejes. A estas distancias las llamó coordenadas del punto: acababan de nacer las **coordenadas cartesianas**, y con ellas, la geometría analítica.



2. Espacios vectoriales

Como en la anécdota de Descartes, seguramente hayas manejado *espacios vectoriales* sin siquiera saberlo, bien usando coordenadas cartesianas, o “flechas” para representar direcciones o magnitudes físicas de movimiento como velocidades o fuerzas.

2.1. Definiciones

Se define un espacio vectorial como:

$$\mathbb{R}^n = \{\vec{u} = (x_1, x_2, \dots, x_n) / x_1, x_2, \dots, x_n \in \mathbb{R}\}$$

Los elementos del espacio, \vec{u} , se denominan **vectores**, los cuales a su vez se componen de n **coordenadas**, x_i . La *dimensión* de un vector \mathbf{v} viene definido por el número de elementos o coordenadas que contenga y se denota por $\dim(\mathbf{v})$.

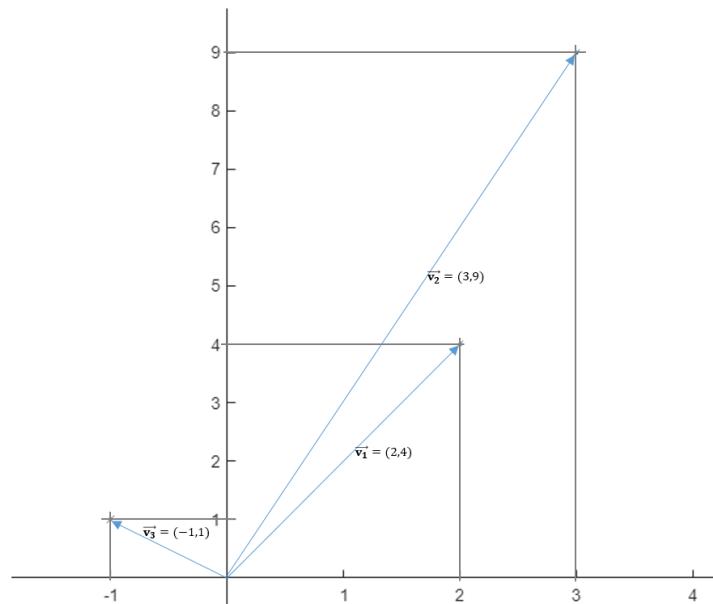
En el caso bidimensional, esto es $n = 2$, se pueden representar las imágenes de una función $f(x)$ en gráficas de tipo XY , las cuales pueden ser interpretadas como espacios \mathbb{R}^2 , que se definen como:

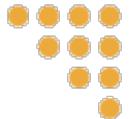
$$\mathbb{R}^2 = \{(x, y) / x, y \in \mathbb{R}\}$$

En geometría, estos espacios son llamados **planos cartesianos** y es un tipo de espacio vectorial.

Así, si las imágenes de $f(x) = x^2$ en $x_1 = 2$, $x_2 = 3$ y $x_3 = -1$ son $f(2) = 4$, $f(3) = 9$, $f(-1) = 1$, éstas se pueden representar en el plano cartesiano por sus vectores:

$$\vec{v}_1 = (x, y) = (2, 4), \quad \vec{v}_2 = (3, 9), \quad \vec{v}_3 = (-1, 1)$$





2. Espacios vectoriales

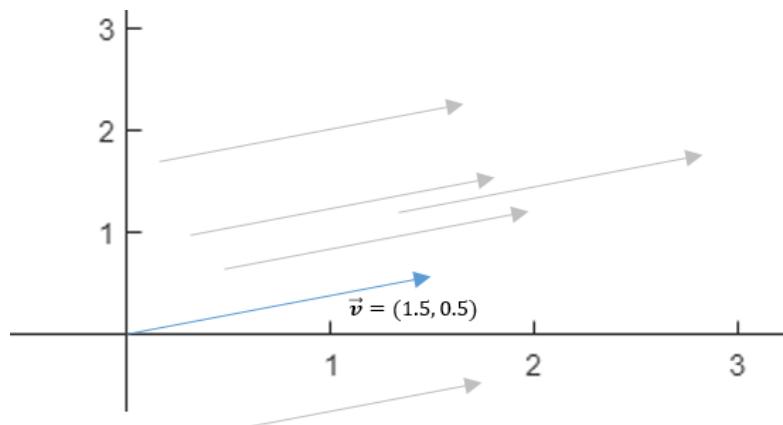
2.2. Operaciones con vectores y sus propiedades geométricas

Sean dos vectores $\vec{u}, \vec{v} \in \mathbb{R}^n$, con $\vec{u} = (x_1, x_2, \dots, x_n)$, $\vec{v} = (y_1, y_2, \dots, y_n)$ y sea el escalar $c \in \mathbb{R}$, se cumplen las siguientes propiedades:

- **Igualdad.** Dos vectores son iguales si todos sus elementos coinciden para mismas posiciones del vector:

$$\vec{u} = \vec{v} \leftrightarrow x_i = y_i, \quad \forall i$$

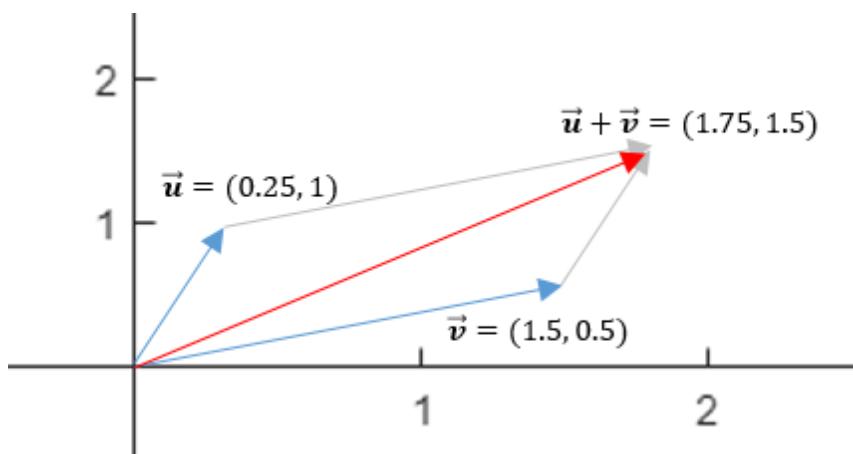
Esto es lo mismo que decir que tengan la misma magnitud y dirección, que sean *paralelas*.

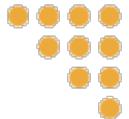


- **Suma de vectores.** Sea \vec{w} el vector suma de \vec{u}, \vec{v} , tal que $\vec{w} = \vec{u} + \vec{v}$, se cumple

$$\vec{w} = (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n)$$

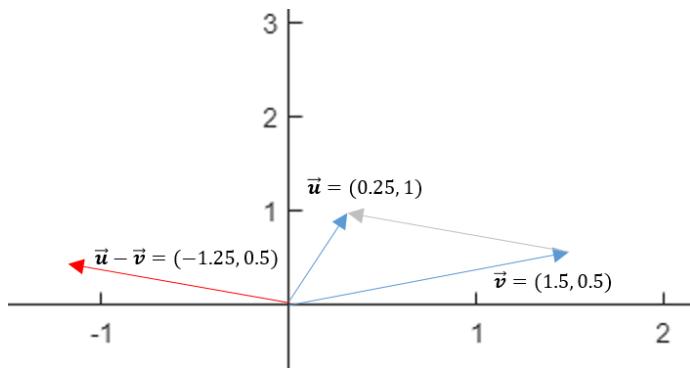
Hay un método geométrico para calcular la suma de dos vectores. Esta es colocar un vector idéntico a \vec{u} al final del vector \vec{v} y viceversa, un vector idéntico a \vec{v} al final del vector \vec{u} . Estos vectores clonados se tocarán en unas coordenadas iguales a la suma $\vec{u} + \vec{v}$.





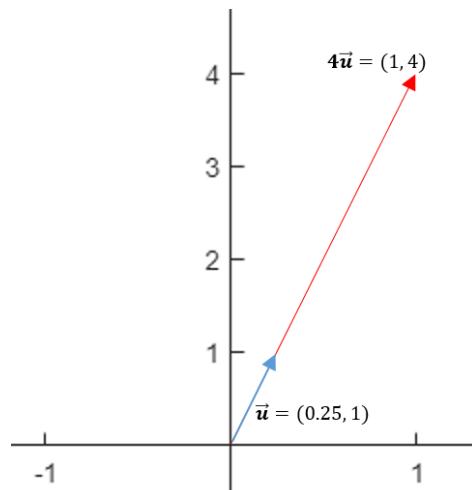
2. Espacios vectoriales

Para el caso de la resta, puede llegarse a la misma conclusión convirtiéndolo el segundo vector a su opuesto $-\vec{v}$. El vector opuesto tiene misma magnitud, dirección, pero sentido contrario. Sin embargo, un método más sencillo aún es crear un vector que empieza donde acaba el segundo término de la resta, el vector \vec{v} , y unirlo al punto donde acaba el primero \vec{u} .



- **Producto de un vector por un escalar.** El resultado de un vector por un escalar c , es otro vector con la misma dirección de \vec{u} y una magnitud c veces la de \vec{u} . Se obtiene multiplicando dicho escalar a todos y cada uno de los elementos de \vec{u} .

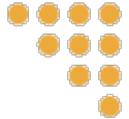
$$c\vec{u} = (cx_1, cx_2, \dots, cx_n)$$



- **Producto escalar.** Se define como producto escalar o *producto interior* de dos vectores como:

$$\vec{u} \cdot \vec{v} = (x_1 \cdot y_1, x_2 \cdot y_2, \dots, x_n \cdot y_n)$$

Es importante remarcar en este punto la importancia de usar el símbolo adecuado, \cdot o $*$, para llevar a cabo productos escalares. Existe otro tipo de operación entre vectores llamado *producto vectorial*, que se denota con el símbolo \times .

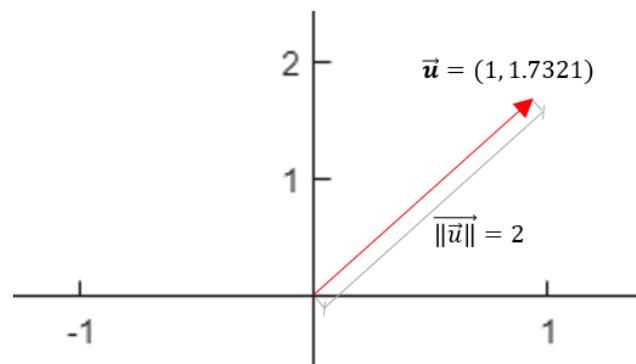


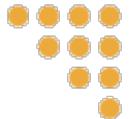
2. Espacios vectoriales

- **Norma de un vector.** La norma, o lo que es lo mismo, la magnitud de un vector \vec{u} se calcula como:

$$\|\vec{u}\| = \sqrt{\sum_{i=1}^n x_i^2} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

El vector da información directa y, sobre todo, dinámica, sobre una medida. Esto es, hacia dónde va esa magnitud o medida, pero es la norma del vector la que informa sobre su valor real. Geométricamente sería la distancia o tamaño de la flecha que representa el vector:





3. Matrices

Las matrices son creadas para facilitar ciertas operaciones numéricas. Representan una ordenación numérica que brinda la posibilidad de trabajar con sistemas de múltiples ecuaciones y variables.

3.1. Definiciones

Definición. Sean $n, m \in \mathbb{R}$, la estructura matemática

$$K^{m \times n} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \quad a_{ij} \in K \quad \forall \{i \in [1, m], j \in [1, n]\}$$

se denomina **matriz de orden $m \times n$** , donde cada elemento a_{ij} es un elemento de la matriz llamado *escalar*.

El orden de una matriz define el número de filas y columnas que tiene. Así, los elementos $a_{i1}, a_{i2}, a_{i3}, \dots, a_{in}$ forman la fila i -ésima y los elementos $a_{1j}, a_{2j}, a_{3j}, \dots, a_{mj}$ forman la columna j -ésima.

Se puede ver también una matriz representada como $K = (k_{ij})$.

Sean A y B dos matrices, son *matrices iguales* si y solo si tienen el mismo orden y todos sus elementos iguales para las mismas posiciones. Esto es:

$$A^{m \times n} = B^{m \times n} \Leftrightarrow a_{ij} = b_{ij} \quad \forall i, j$$

Las matriz $A^{m \times n}$ que cumple $m = n$ se denomina **matriz cuadrada** de orden n . Si una matriz cuadrada cumple $a_{ij} = a_{ji} \quad \forall i, j$ se le llama **matriz simétrica**.

$$\begin{pmatrix} -3 & 2 & 0 \\ 0.54 & 1 & -10 \end{pmatrix}$$

Matriz

$$\begin{pmatrix} 10 & -56 & 1 \\ 29 & 4 & 0 \\ 20 & 18 & 1 \end{pmatrix}$$

Matriz cuadrada

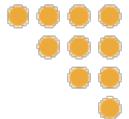
$$\begin{pmatrix} 10 & -56 & 20 \\ -56 & 4 & 18 \\ 20 & 18 & 1 \end{pmatrix}$$

Matriz simétrica

Se denomina **matriz transpuesta** de $A^{m \times n} = (a_{ij})$ a aquella $B^{n \times m} = (a_{ji})$. Es decir, la matriz que resulta de intercambiar las filas por las columnas de A .

Dicha operación se denota con el superíndice A^T . Así,

$$A = \begin{pmatrix} -3 & 2 & 0 \\ 0.54 & 1 & -10 \end{pmatrix} \quad \rightarrow \quad A^T = \begin{pmatrix} -3 & 0.54 \\ 2 & 1 \\ 0 & -10 \end{pmatrix}$$



3. Matrices

Los elementos de las matrices, $a_{ii} \forall i \in [1, \min\{m, n\}]$ se denominan *elementos diagonales* todos ellos forman la **diagonal principal de la matriz**.

La matriz que posee todos sus elementos por debajo de la diagonal principal a cero se denomina *triangular superior*; si esto ocurre con los elementos por encima de la diagonal, se denomina *triangular inferior*.

En el caso particular en que una matriz sólo posee elementos no nulos en su diagonal principal, la matriz se denomina *matriz diagonal*:

$$\begin{pmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -10 \end{pmatrix}$$

Matriz diagonal

$$\begin{pmatrix} -3 & 3 & 5 \\ 0 & 1 & 6 \\ 0 & 0 & 1 \end{pmatrix}$$

Triangular superior

$$\begin{pmatrix} -3 & 0 & 0 \\ 50 & -6 & 0 \\ 10 & 110 & 1 \end{pmatrix}$$

Triangular inferior

La diagonal principal de la matriz anterior A , lo formaría el vector $\vec{d} = (-3 \quad 1)$.

Definición. Toda matriz $A^{m \times n}$ que cumple $n = 1$ o $m = 1$ se denomina **vector**. Para el caso $n = 1$, se tiene un *vector columna*; para el caso $m = 1$, se tiene el *vector fila*.

Nótese que la traspuesta de un vector fila es un vector columna y viceversa.

Por ejemplo, la anterior matriz de ejemplo A , estaría compuesto por los vectores fila

$$a_{1j} = (-3 \quad 2 \quad 0), \quad a_{2j} = (0.54 \quad 1 \quad -10)$$

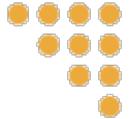
Y los vectores columna

$$a_{i1} = \begin{pmatrix} -3 \\ 0.54 \end{pmatrix}, \quad a_{i2} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad a_{i3} = \begin{pmatrix} 0 \\ -10 \end{pmatrix}$$

De entre las posibles matrices diagonales, existe una que merece especial atención. Aquella matriz diagonal cuyos elementos no nulos son todos igual a 1:

$$I_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

se denomina **matriz identidad de orden n** .



3. Matrices

3.2. Operaciones con matrices

- **Suma de matrices.** Sean dos matrices A y B , con $A = (a_{ij})$ y $B = (b_{ij})$, se cumple:

$$A + B = a_{ij} + b_{ij}, \quad \forall i, j$$

- **Producto por escalar.** Sea la matriz A con $A = (a_{ij})$ y el escalar $c \in \mathbb{R}$, se cumple:

$$cA = ca_{ij}, \quad \forall i, j$$

Es decir, para sumar dos matrices, éstas deberán ser del mismo orden y se sumarán todos los elementos que ocupen posiciones coincidentes ij . En el caso de un producto escalar, será necesario multiplicar todos y cada uno de los elementos de la matriz por el escalar.

Dada las matrices

$$A = \begin{pmatrix} -3 & 2 & 0 \\ 0.54 & 1 & -10 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 7.2 & 2 \\ 0.45 & 6 & 11 \end{pmatrix}, \quad c = 2.5$$

se tiene

$$\begin{aligned} A + B &= \begin{pmatrix} -3 & 2 & 0 \\ 0.54 & 1 & -10 \end{pmatrix} + \begin{pmatrix} 1 & 7.2 & 2 \\ 0.45 & 6 & 11 \end{pmatrix} \\ &= \begin{pmatrix} -3 + 1 & 2 + 7.2 & 0 + 2 \\ 0.54 + 0.45 & 1 + 6 & -10 + 11 \end{pmatrix} = \begin{pmatrix} -2 & 9.2 & 2 \\ 0.99 & 7 & 1 \end{pmatrix} \end{aligned}$$

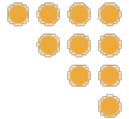
$$\begin{aligned} A - B &= \begin{pmatrix} -3 & 2 & 0 \\ 0.54 & 1 & -10 \end{pmatrix} + \begin{pmatrix} 1 & 7.2 & 2 \\ 0.45 & 6 & 11 \end{pmatrix} \\ &= \begin{pmatrix} -3 - 1 & 2 - 7.2 & 0 - 2 \\ 0.54 - 0.45 & 1 - 6 & -10 - 11 \end{pmatrix} = \begin{pmatrix} -4 & -5.2 & -2 \\ 0.09 & -5 & -21 \end{pmatrix} \end{aligned}$$

$$\begin{aligned} cB &= 2.5 \begin{pmatrix} 1 & 7.2 & 2 \\ 0.45 & 6 & 11 \end{pmatrix} = \begin{pmatrix} 1 * 2.5 & 7.2 * 2.5 & 2 * 2.5 \\ 0.45 * 2.5 & 6 * 2.5 & 11 * 2.5 \end{pmatrix} \\ &= \begin{pmatrix} 2.5 & 18 & 0 \\ 1.125 & 15 & 27.5 \end{pmatrix} \end{aligned}$$

Aunque haya podido parecer una operación muy directa operar una matriz por un escalar, no ocurre así en el caso del producto de matrices.

- **Producto de matrices.** Sean dos matrices A y B , con $A = (a_{ij})$ y $B = (b_{ij})$, se cumple

$$C = A * B = \sum_{k=1}^n a_{ik}b_{kj}, \quad \forall i, j$$

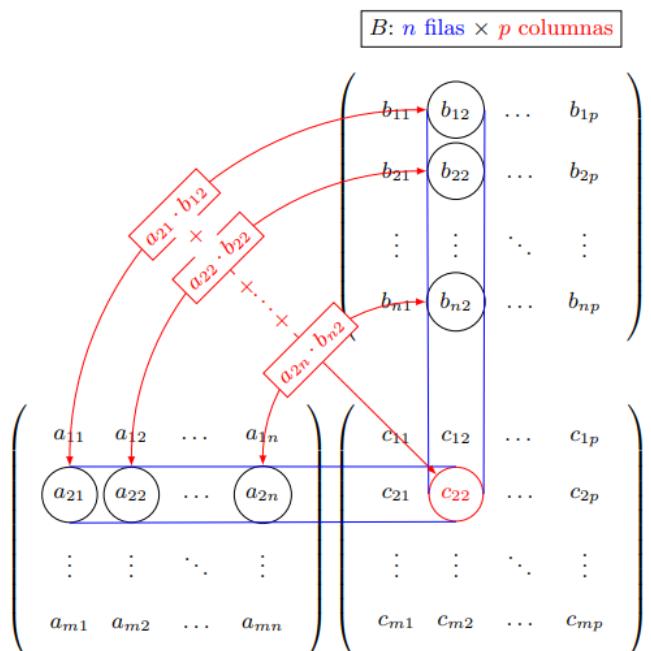


3. Matrices

¿Cómo interpretar esta definición? Para multiplicar dos matrices se deben realizar una serie de sumas entre elementos de ambas matrices para cada elemento del producto. Es decir, cada elemento $C = (c_{ij})$ será el resultado de la suma de los productos de cada par de elementos ordenados de la fila i -ésima de A y la columna j -ésima de B . De este modo, sean dos matrices A y B

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \quad B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix}$$

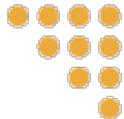
su producto $C = A * B$ se calcula del siguiente modo



$$C = A * B = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix}$$

$$= \begin{pmatrix} (a_{11}b_{11} + a_{12}b_{21} + \cdots + a_{1n}b_{m1}) & c_{12} & \cdots & (a_{11}b_{1n} + a_{12}b_{2n} + \cdots + a_{1n}b_{mn}) \\ (a_{21}b_{11} + a_{22}b_{21} + \cdots + a_{2n}b_{m1}) & c_{22} & \cdots & (a_{21}b_{1n} + a_{22}b_{2n} + \cdots + a_{2n}b_{mn}) \\ \vdots & \vdots & \ddots & \vdots \\ (a_{m1}b_{11} + a_{m2}b_{21} + \cdots + a_{mn}b_{m1}) & c_{m2} & \cdots & (a_{m1}b_{1n} + a_{m2}b_{2n} + \cdots + a_{mn}b_{mn}) \end{pmatrix}$$

Para obtener, por ejemplo, el elemento c_{34} hay que multiplicar, de manera ordenada, los elementos de la fila 3 de A por los de la columna 4 de B .



3. Matrices

Puede parecer complejo, pero se entiende con facilidad a través de ejemplos.

Sean las matrices vistas anteriormente A y la traspuesta de B

$$A = \begin{pmatrix} -3 & 2 & 0 \\ 0.54 & 1 & -10 \end{pmatrix}, \quad B^T = \begin{pmatrix} 1 & 0.45 \\ 7.2 & 6 \\ 2 & 11 \end{pmatrix}$$

Su producto C se calcula así

$$\begin{aligned} C = A * B &= \begin{pmatrix} -3 & 2 & 0 \\ 0.54 & 1 & -10 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0.45 \\ 7.2 & 6 \\ 2 & 11 \end{pmatrix} \\ &= \begin{pmatrix} -3 * 1 + 2 * 7.2 + 0 * 2 & -3 * 0.45 + 2 * 6 + 0 * 11 \\ 0.54 * 1 + 1 * 7.2 + (-10) * 2 & 0.54 * 0.45 + 1 * 6 + (-10) * 11 \end{pmatrix} \\ &= \begin{pmatrix} 11.4 & 10.65 \\ -12.26 & -103.757 \end{pmatrix} \end{aligned}$$

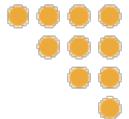
3.3. Propiedades operacionales de matrices

Dada la particular metodología que se sigue para calcular productos de matrices, es importante tener en cuenta que **esta operación no es siempre posible**. Dependerá del orden de ambas matrices. Si, como se ha explicado, para obtener cada elemento c_{ij} de la matriz producto C es necesario operar con los elementos del vector fila i de A con el vector columna j de B , se hace necesario que estos vectores tengan el mismo número de elementos y, por lo tanto, A tenga el mismo número de columnas que B tiene de filas.

Definición. Sean A y B dos matrices de orden $m \times n$ y $k \times l$, su producto sólo será posible cuando el número de columnas de la primera matriz coincida con el número de filas de la segunda matriz, es decir, $n = k$

$$\exists C = A^{m \times n} * B^{k \times l} \leftrightarrow n = k$$

El anterior ejemplo se componía de una matriz $A^{2 \times 3}$ y otra $B^{3 \times 2}$. Su producto se pudo realizar ya que la A tenía 3 columnas y B tenía 3 filas. Esta característica de las multiplicaciones entre matrices conduce a otra muy importante. Y es que **el producto de matrices no es comutativo**. Esto es, $A * B$ no tiene porqué ser igual que $B * A$. Existe un caso en el que sí se cumple.



3. Matrices

Definición. Dada una matriz cuadrada $A^{n \times n}$, se dice que la matriz $B^{n \times n}$ es su **matriz inversa** si se cumple

$$A * B = B * A = I_n$$

donde I_n es la matriz identidad. Esta relación se denota como $A^{-1} = B$. Si existe matriz inversa para A se le llama *regular*; en caso contrario, *singular*.

3.4. Representación matricial de sistemas de ecuaciones

Se ha mencionado anteriormente la importante utilidad que tienen las matrices para representar grandes sistemas de ecuaciones. Así, un sistema de m ecuaciones y n variables o incógnitas se puede escribir como

$$\left\{ \begin{array}{l} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ a_{31}x_1 + a_{32}x_2 + \cdots + a_{3n}x_n = b_3 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{array} \right.$$

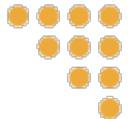
haciendo uso de las operaciones antes descritas, es posible convertir este sistema de ecuaciones en una ecuación matricial del tipo $A\mathbf{x} = \mathbf{b}$

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ a_{31} & a_{32} & \cdots & a_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_m \end{pmatrix}$$

formado por

$$A^{m \times n} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ a_{31} & a_{32} & \cdots & a_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \quad \mathbf{x}^{n \times 1} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix}, \quad \mathbf{b}^{m \times 1} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_m \end{pmatrix}$$

A contendrá siempre todos los coeficientes del sistema de ecuaciones y será de orden $m \times n$, siendo m el número de ecuaciones del sistema y n el número de variables o incógnitas del sistema. Por lo tanto, fila i -ésima contendrá los coeficientes de la ecuación i -ésima del sistema y cada columna j -ésima contendrá todos los coeficientes multiplicados por la misma incógnita j -ésima.



3. Matrices

El vector \mathbf{x} contendrá todas las variables o incógnitas del sistema y \mathbf{b} los términos independientes, aquellos números de la ecuación que no son coeficientes de ninguna incógnita.

Se presenta el sistema de ecuaciones siguiente

$$\begin{aligned}x + 5y - z &= 2 \\4x - 12y + 3z + 1 &= -4 \\x + 4y &= 0 \\-3x + \frac{6}{7}y + 10z &= 39\end{aligned}$$

Para convertir esto a su versión matricial, se debe, en primer lugar, reestructurar el sistema de ecuaciones a la llamada *forma canónica*. Esta contiene todos los términos con incógnitas a la izquierda del signo igual, $=$, y los términos independientes sumados a la derecha. Todas las ecuaciones están en esta forma a excepción de la segunda, que posee el término independiente 1 en la izquierda, por lo que habrá que pasarlo a la derecha restando, de tal manera que la ecuación queda $4x - 5y + 3z = -5$. Ahora ya se pueden obtener \mathbf{b} además de \mathbf{x} tras localizar las incógnitas. Estas son $\{x, y, z\}$, por lo que

$$\mathbf{x}^{3 \times 1} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 2 \\ -5 \\ 0 \\ 39 \end{pmatrix}$$

El sistema tiene 3 incógnitas y 4 ecuaciones por lo tanto la matriz A será de orden 4×3 . Extrayendo cada coeficiente de cada ecuación, se rellena la matriz A como sigue

$$A^{4 \times 3} = \begin{pmatrix} 1 & 5 & -1 \\ 4 & -12 & 3 \\ 1 & 4 & 0 \\ -3 & \frac{6}{7} & 10 \end{pmatrix}$$

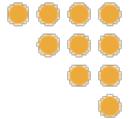
Es posible comprobar el resultado realizando la operación $A\mathbf{x} = \mathbf{b}$

$$\begin{aligned}A\mathbf{x} &= \begin{pmatrix} 1 & 5 & -1 \\ 4 & -12 & 3 \\ 1 & 4 & 0 \\ -3 & \frac{6}{7} & 10 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 * x + 5 * y + (-1) * z \\ 4 * x - 12 * y + 3 * z \\ 1 * x + 4 * y + 0 * z \\ -3 * x + \frac{6}{7} * y + 10 * z \end{pmatrix} \\&= \begin{pmatrix} x + 5y - z \\ 4x - 12y + 3z \\ x + 4y \\ -3x + \frac{6}{7}y + 10z \end{pmatrix} = \begin{pmatrix} 2 \\ -5 \\ 0 \\ 39 \end{pmatrix} = \mathbf{b}\end{aligned}$$

Together for Tomorrow!
Enabling People

Representación de funciones

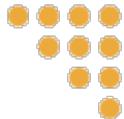
#TecnologíaConPropósito



1. Introducción

La conexión entre lo analítico y lo gráfico ayuda a ambas partes; estas disciplinas se apoyan mutuamente, bien para dar una representación espacial de un modelo matemático, y ser así más comprensible, bien para modelar aquello que puede ser sabido de un modo empírico, pero no está demostrado hasta que no se analiza apropiadamente.

En esta unidad se verá cómo representar una función lineal sobre una gráfica, las componentes de esta y, del mismo modo, llevar una recta representada sobre el espacio bidimensional a su versión analítica en forma de función matemática. Además, se realizará un análisis en profundidad sobre ecuaciones no lineales para conocer, por ejemplo, cómo evoluciona a lo largo del dominio, su convergencia o dónde se sitúan sus máximos y sus mínimos.



2. Funciones lineales

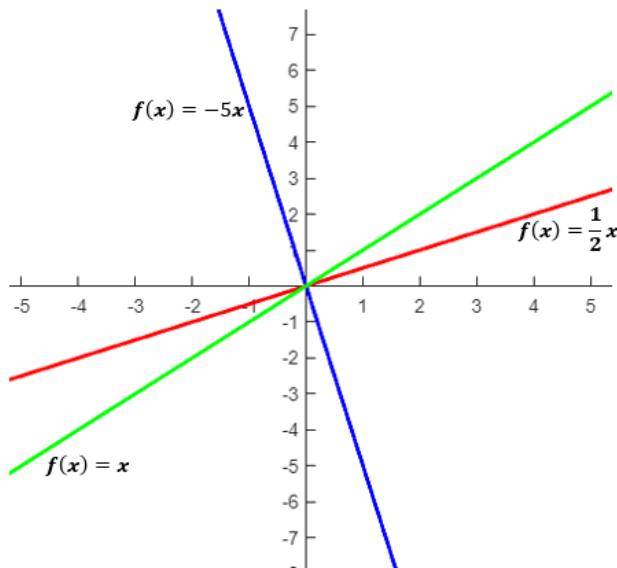
2.1. La ecuación de la recta

Un concepto importante cuyo significado y clasificación es primordial comprender antes de continuar es referido a la linealidad. Una **función lineal**, también llamada *función de proporcionalidad directa*, es toda aquella que no contiene ninguna de sus variables con una potencia mayor de 1, del tipo $f(x) = mx$ o $y = mx$.

Debido a su condición de proporcional derivada de la propia ecuación que la define, vendrá siempre representada en la gráfica como una línea recta, de ahí su nombre y su esencial diferencia de las funciones no lineales.

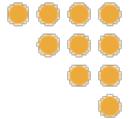
El coeficiente m es llamado *coeficiente de proporcionalidad*. Indica por cuánto se multiplica el valor de entrada de x para obtener su imagen y o, dicho de otro modo, indica cuánto debe incrementar y , Δy , dado un incremento un incremento en x , Δx . Es decir, es la *variación media* de la función que, para el caso lineal, es a su vez la **pendiente** de la recta.

Ejemplos de funciones lineales son $f(x) = x$, $f(x) = -5x$, $f(x) = \frac{1}{2}x$:

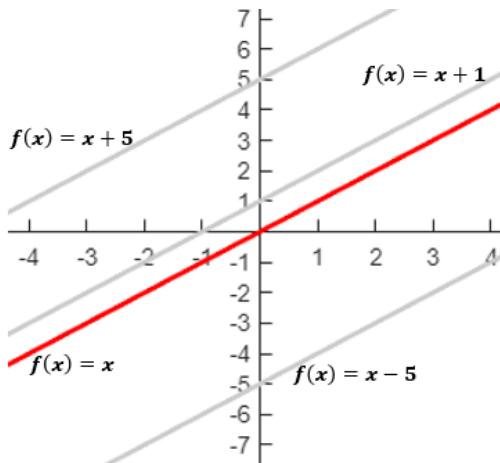


La **función afín** es aquella, también proporcional, que contiene alguna condición inicial que provoca que no pase por el origen.

Son del tipo $f(x) = mx + n$. El nuevo término n se denomina *ordenada en origen*, ya que proporciona el valor de la imagen $f(0) = 0$, es decir, el punto de intersección entre la recta y el eje de ordenadas $x = 0$.



2. Funciones lineales



En la imagen anterior puede comprobarse como, efectivamente, el término independiente de la función de la recta indica el punto de corte con el eje vertical.

2.2. Tipos

La forma de la función $f(x) = mx + n$ es llamada **ecuación explícita de la recta**. Sin embargo, existen otros tipos de representación de esta misma función que es necesario conocer:

- **Ecuación implícita de la recta.** Esta es de la forma

$$Px + Qy + R = 0,$$

y está caracterizada por dejar uno de los miembros de la ecuación a cero.

La ecuación explícita es única para cada recta, pero sin embargo es posible obtener infinitas ecuaciones implícitas de una misma recta.

Por ejemplo, la ecuación

$$3x + 4y - 15 = 0,$$

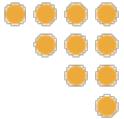
representa la misma que estas tres equivalentes

$$\underbrace{6x + 8y - 30 = 0}_{x2}, \quad \underbrace{-1.5x - 2y + 7.5 = 0}_{x(-0.5)}, \quad \underbrace{45x + 60y - 225 = 0}_{x15}$$

Al estar igualadas a cero, el término de la izquierda puede ser multiplicado por cualquier cociente sin alterar la ecuación.

Para transformar la ecuación de la recta de su versión implícita a la explícita, se debe aislar y en la ecuación. Así, en la anterior

$$3x + 4y - 15 = 0 \rightarrow \frac{3}{4}x + \frac{4}{4}y - \frac{15}{4} = 0 \rightarrow y = -\frac{3}{4}x + \frac{15}{4}$$



2. Funciones lineales

- **Ecuación punto-pendiente.** Ésta se obtiene de la propia definición de la pendiente m como variación media. Así, se obtiene

$$m = VM = \frac{\Delta y}{\Delta x} = \frac{y - y_0}{x - x_0} \rightarrow y - y_0 = m(x - x_0)$$

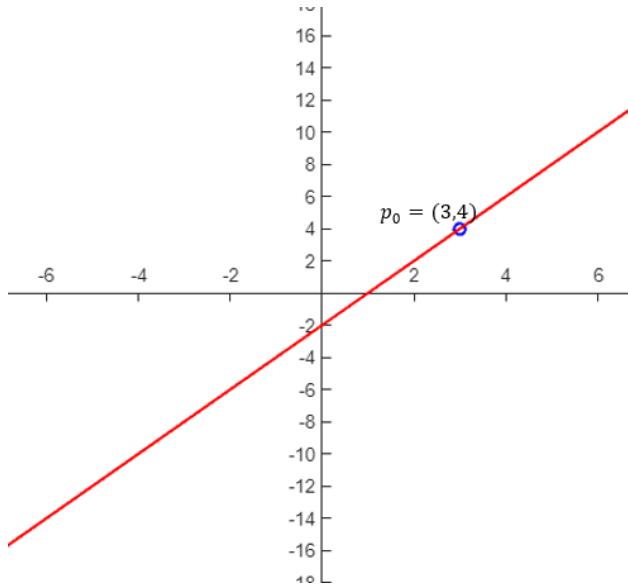
La importancia de esta forma estriba en la posibilidad de poder crear la ecuación de la recta a partir de información como dos puntos de la recta, o un punto y la pendiente.

Supóngase que se busca la función $f(x)$ que modela la recta que pasa por el punto $(3,4)$ y tiene una pendiente de $m = 2$. Existen dos maneras de encontrar esta función.

Como punto inicial se tiene $p_0 = (x_0, y_0) = (3,4)$. Si incluimos esta información y la pendiente en la ecuación *punto-pendiente* se tiene

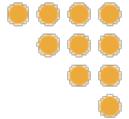
$$\begin{aligned} y - y_0 &= m(x - x_0) \rightarrow y - 4 = 2(x - 3) \rightarrow y = 2x - 2 * 3 + 4 \rightarrow y \\ &= 2x - 2 \end{aligned}$$

Ahora, teniendo el punto p_0 y $n = -2$, se obtiene la recta uniendo estos dos puntos:



Si en lugar de la pendiente, se tuviese inicialmente dos puntos $p_0 = (x_0, y_0)$ y $p_1 = (x_1, y_1)$, a partir de sus coordenadas se obtendría la pendiente.

$$m = \frac{\Delta y}{\Delta x} = \frac{y_1 - y_0}{x_1 - x_0}$$



2. Funciones lineales

2.3. Paralelismo

Se ha visto anteriormente como las funciones afines son representadas mediante rectas paralelas. La condición de paralelismo exige que las rectas nunca se corten entre sí. Todo par de rectas no paralelas, se conocen como *rectas secantes*.

Se resume así; sean dos funciones lineales $f_1(x) = m_1x + n_1$ y $f_2(x) = m_2x + n_2$:

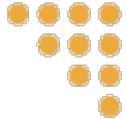
- Si $m_1 = m_2$, las rectas son **paralelas** (o afines).
 - Si además $n_1 = n_2$ son coincidentes
- Si $m_1 = -\frac{1}{m_2}$, las rectas son **perpendiculares**
- Si no se cumple ninguna de las dos anteriores son **secantes** y se puede conocer el punto donde se cortan resolviendo el sistema de ecuaciones

$$y = m_1x + n_1$$

$$y = m_2x + n_2$$

Para saber dónde corta la recta con alguna de los ejes, tan sólo hay que darle a la variable correspondiente el valor nulo. Así, si se quiere determinar si una función corta el eje de abscisas ($y = 0$) tan sólo hay que aplicar esa igualdad a la función de la recta. La recta del anterior ejemplo; $y = 2x - 2$, cortaría en $0 = 2x - 2 \rightarrow x = \frac{2}{2} = 1$.

Del mismo modo, para conocer el punto donde corta al eje de ordenadas, es necesario convertir x en 0; para el caso de la función explícita este valor será n .



3. Funciones no lineales

Se ha visto cómo la simplicidad de las funciones lineales permite reducir el número de características para identificar cualquier recta de forma unívoca tan sólo a dos elementos: la pendiente y el desplazamiento vertical respecto al origen, es decir, la altura a la que cortaba con el eje de ordenadas.

Pero en muchos otros casos prácticos, la relación entre un valor de entrada, x , y su valor de salida, y , no son proporcionales y, por lo tanto, no se representan mediante líneas rectas sino curvas.

Una señal práctica que indica si una función es lineal o no lineal es el valor de la potencia de las variables independientes o de entrada; si éste es 1, será lineal, en cualquier otro caso será una función no lineal.

3.1. Funciones principales

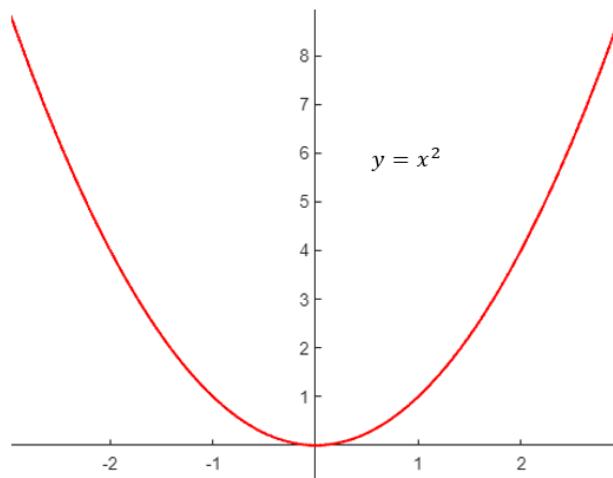
A continuación, se va a describir una serie de funciones básicas a partir de las cuales, mediante operaciones aritméticas o de composición, compondrán el resto de funciones, y por esto merecen especial atención.

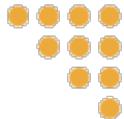
3.1.1. Función cuadrática

Las funciones cuadráticas son funciones polinómicas de la forma:

$$y = ax^2 + bx + c, \quad \forall a \neq 0$$

Que en su versión más simple lleva la forma $y = x^2$. Gráficamente, forman una curva en forma de U llamada **parábola**, forma muy importante y recurrida en la geometría.





3. Funciones no lineales

Las paráolas, analíticamente hablando –tienen otra definición más precisa en la geometría–, representan a aquellas funciones cuyas variaciones medias incrementan de manera constante con el tiempo.

En el caso de la función lineal, la variación media era constante e igual a su pendiente, m . En el caso cuadrático, la función tendrá una variación que incrementa con x . Esto es, para el caso $y = x^2$, para una variación en x de $\Delta x = 1$, y experimenta un incremento de $\Delta y = 1$ (del punto $(0,0)$ al $(1,1)$); sin embargo, cuando $\Delta x = 2$, y incrementa $\Delta y = 4$ (del punto $(0,0)$ al $(2,4)$). Esta variación continua de los propios incrementos de la función ocurre durante todo el dominio de X .

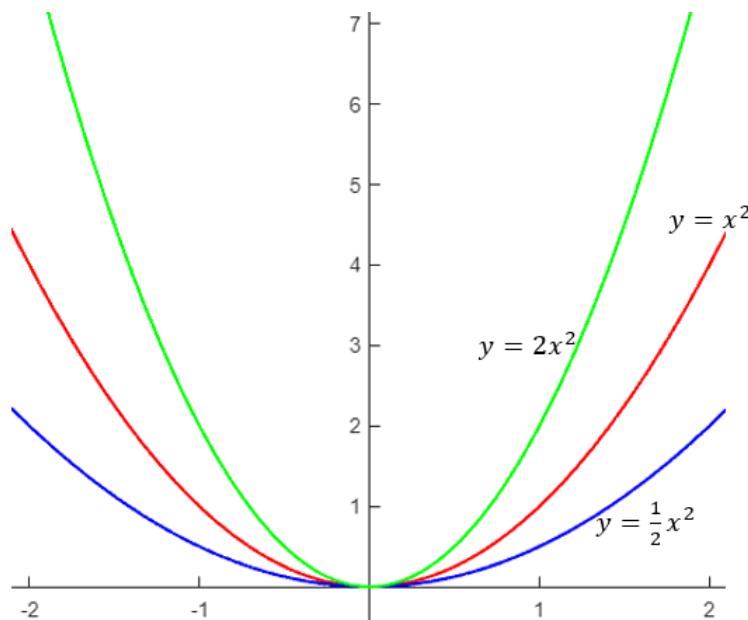
Las propiedades de la función cuadrática

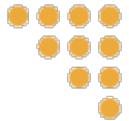
Sea $y = ax^2 + bx + c$ una función cuadrática o parábola

- Si $a > 0$ la función es **convexa** (la curva es de la forma U)
- Si $a < 0$ la función es **cóncava** (la curva es de la forma ∩)
- La función es **simétrica** y su eje vertical de simetría se sitúa en $x = -\frac{b}{2a}$. En el caso particular $y = x^2$, está en $x = 0$.

La función cuadrática puede ser reescrita de la forma $y = (ax + P)^2 + Q$ de la que podemos sacar información acerca de la curva de forma más directa.

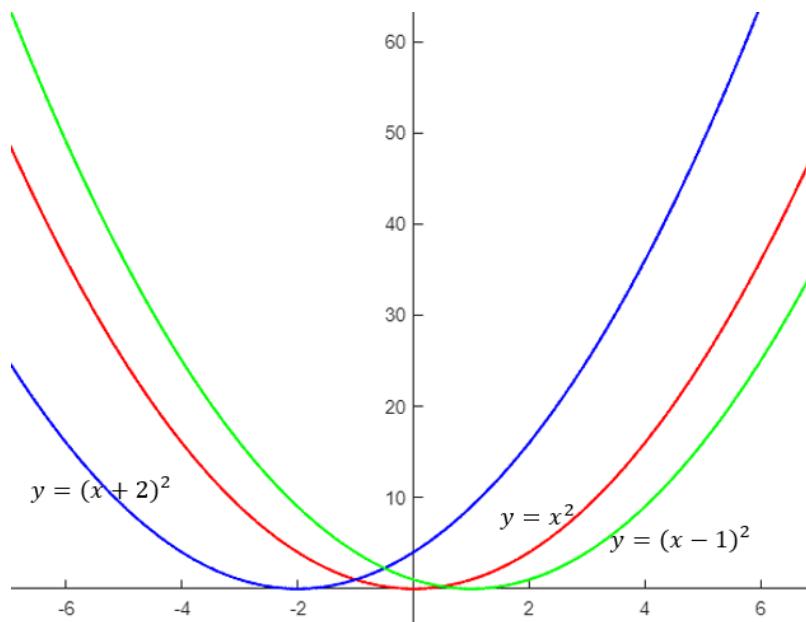
- a define la pronunciación de la curva. Esto es, a mayor a , mayor es el incremento de y ante incrementos de x :



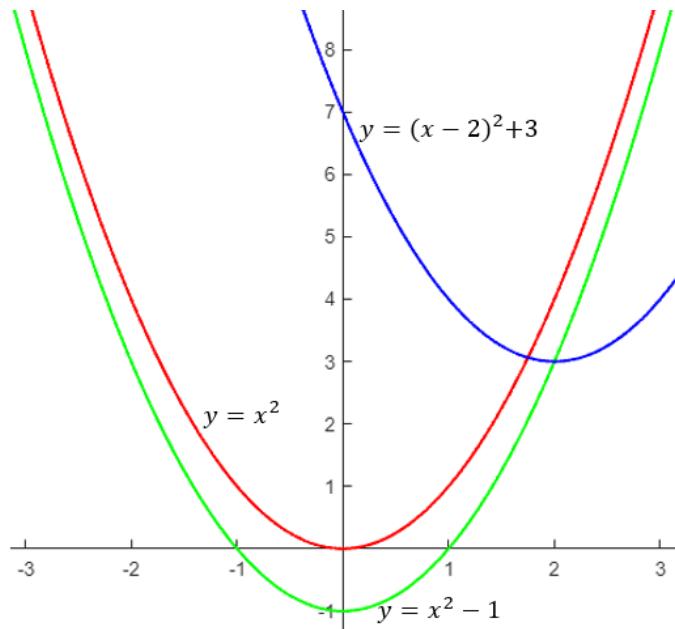


3. Funciones no lineales

- P define el desplazamiento a lo largo del eje de abscisas:



- Q define el desplazamiento a lo largo del eje de ordenadas:

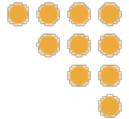


3.1.2. Funciones de variación inversa

En este otro caso particular la variable y lleva una proporcionalidad inversa a x . Es decir, cuanto más crece en x , más se aproxima y a 0, y lo contrario. Toma la forma

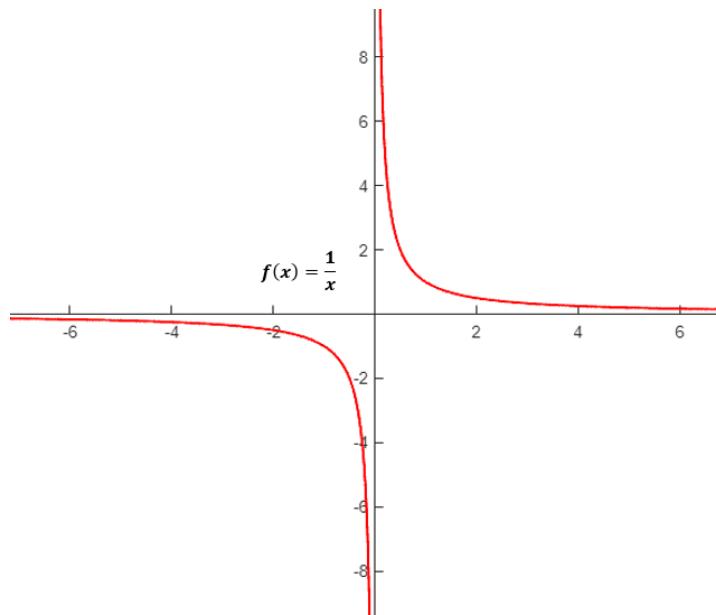
$$y = \frac{k}{x}, \quad \forall k \neq 0$$

siendo su versión más simple y conocida $y = \frac{1}{x}$.

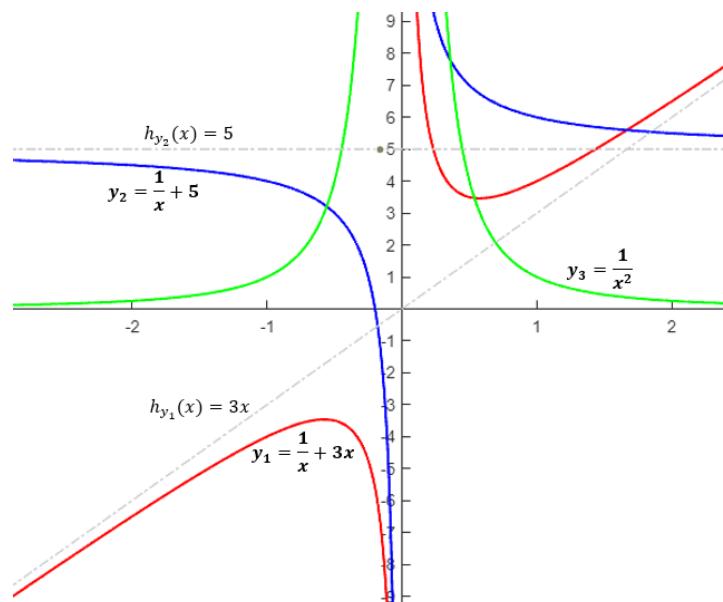


3. Funciones no lineales

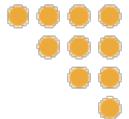
En la gráfica, forman otro tipo curvas llamadas **hipérbolas**, caracterizadas por llevar la forma de “variación acelerada” de las paráboles en los puntos cercanos a sus vértices e ir perdiendo su condición de curva a medida que se recorre hacia sus extremos, aproximándose a la forma de una recta sin nunca llegar a serlo. Este comportamiento es el mismo que siguen las sucesiones y series convergentes.



Se dice que una recta $h(x)$ es asintótica a la función $f(x)$ cuando los valores que toma y convergen hacia $h(x)$. En el caso de arriba se tienen dos rectas asintóticas coincidentes con los ejes $y = 0$ y $x = 0$.



En la imagen anterior se presentan distintas hipérbolas con las funciones $h_{y_i}(x)$ de sus rectas asintóticas



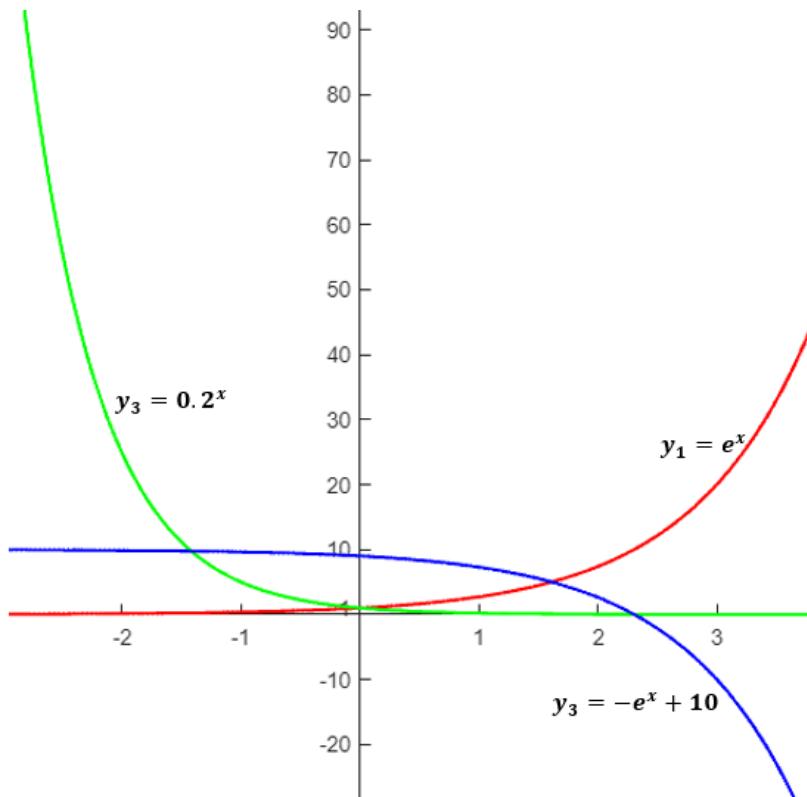
3. Funciones no lineales

3.1.3. Funciones exponenciales

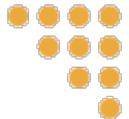
Un último tipo de funciones no lineales que merece ser descrita son las exponenciales. Estas tienen la forma general

$$y = ab^x, \quad \forall a, b \neq 0$$

aunque se conoce formalmente por su versión $b = e$, siendo e el número irracional de *Euler*, debido a que aparece en numerosos procesos físicos y bioquímicos de la naturaleza. En general, las funciones exponenciales son ampliamente usadas en disciplinas estadísticas relacionadas con crecimientos poblacionales como biología, epidemiología, economía o física. De hecho, se ha expandido el término *exponencial* al uso cotidiano para referirse a un crecimiento o evolución altamente acelerado, descontrolado o explosivo. Observando la gráfica de abajo se puede entender fácilmente el motivo.



La función $y = e^x$ evoluciona desde valores infinitesimales cuando x tiende a $-\infty$, cruza el eje de ordenadas en $y = a$, ya que $b^0 = 1, \forall b \neq 0$ y crece con una “aceleración” muy elevada a medida que x crece a lo largo de valores positivos.



4. Funciones multivariable

Hasta ahora se ha trabajado sobre funciones de una sola variable, $y = f(x)$ en las cuales, para cada valor de x se obtiene una única imagen de $f(x)$. Sin embargo, en la mayoría de los casos reales en los que se estudian dependencias entre la información disponible y la información que se pretende obtener, esta información disponible no recae sobre una sola magnitud. Es decir, en muchos casos el valor a calcular dependerá no solo de un factor o variable de entrada, sino de varios.

Definición. Se denomina **función multivariable** con dominio en $D \subset \mathbb{R}^n$, con $n > 1$, a cualquier función de la forma

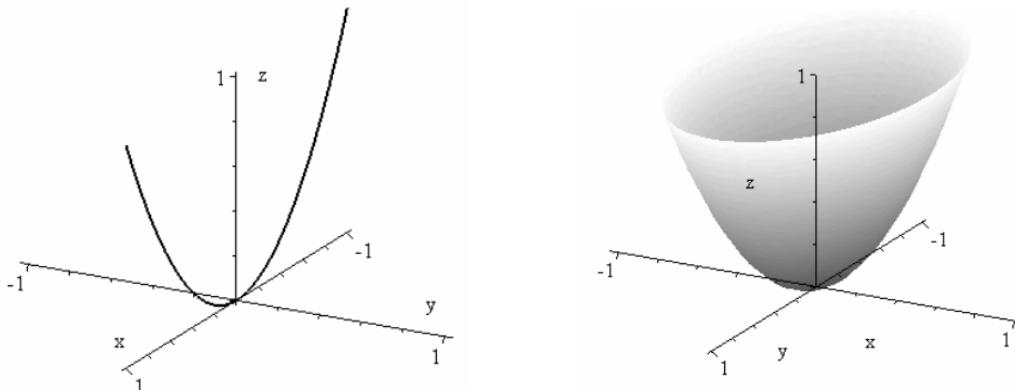
$$f: D \subset \mathbb{R}^n \rightarrow \mathbb{R}$$

y se expresa, en su *forma explícita*, como

$$y = f(x_1, x_2, \dots, x_n)$$

La representación gráfica de estas funciones vendrá determinada por la dimensión de esta, que no es más que el número de elementos, variables o incógnitas que aparecen en la función. Así, si en funciones de una sola variable, $y = f(x)$, se usaba el plano cartesiano XY para su representación, en el caso de dos variables, tipo $z = f(x, y)$ las funciones se representan en el **espacio tridimensional XYZ** .

En el caso de una función univariable se podían representar puntos, rectas, formas poligonales que encerraban superficies; en el caso de funciones con dos variables, podemos representar planos y formas poliédricas que encierran volúmenes.



$$f(x) = x^2$$

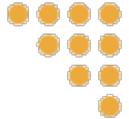
$$f(x, y) = x^2 + 3y^2$$

Para los casos de más de 2 variables (3 dimensiones), no existe representación como en los casos univariable y bivariable con espacios cartesianos, pero sí un término hipotético para este supuesto espacio multidimensional llamado **hiperespacio**.

Together for Tomorrow!
Enabling People

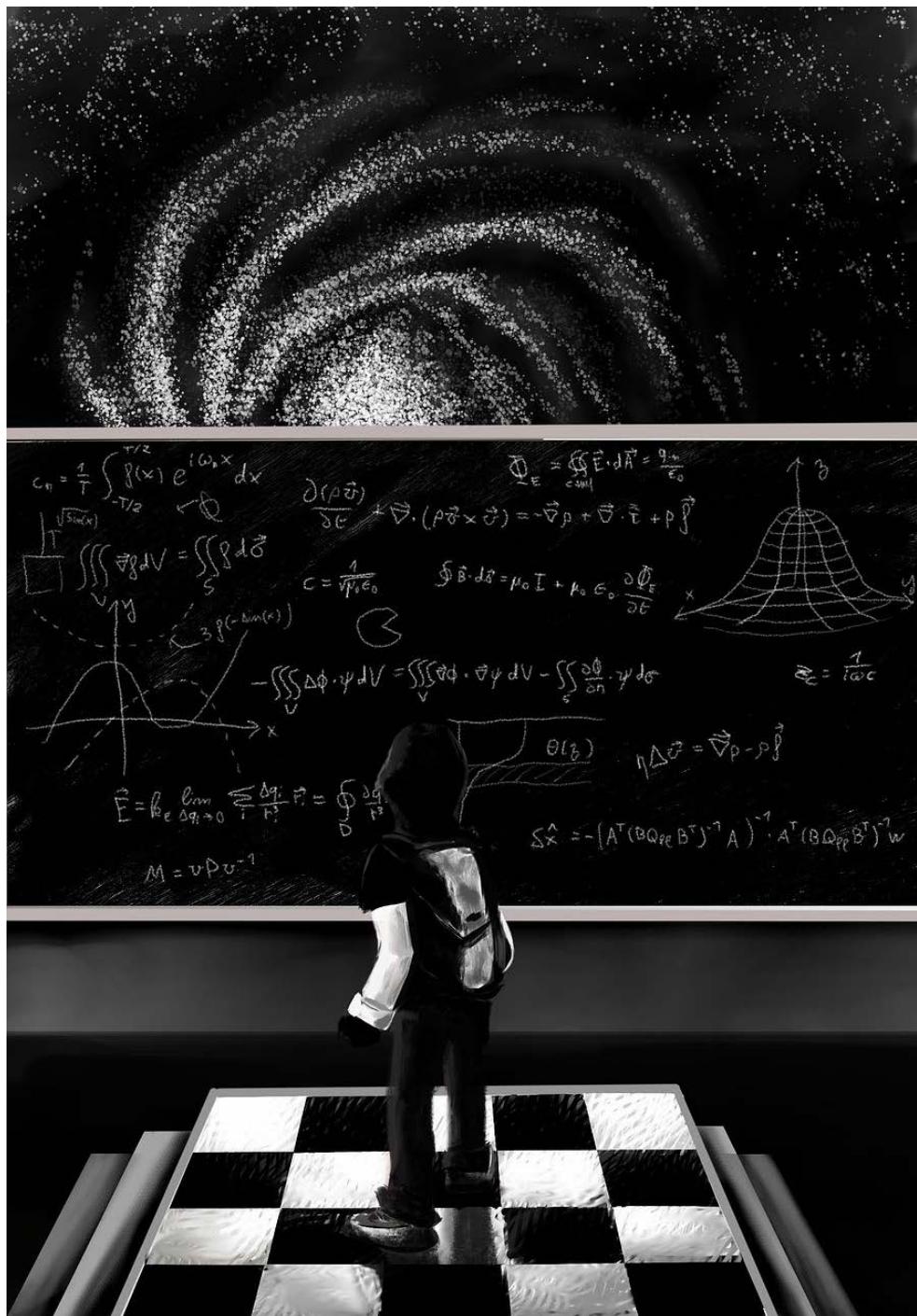
Operaciones espaciales

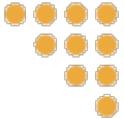
#TecnologíaConPropósito



1. Introducción

Interpretar y manipular ecuaciones y estudiar su morfología es una herramienta clave en el análisis del modelo y, por tanto, en la resolución de problemas matemáticos en disciplinas como estadística, economía, ingeniería o inteligencia artificial. Desde el cálculo de rectas de regresión o problemas de flujo máximo-coste mínimo, hasta diseño de modelos predictivos, evaluaciones de máquinas de soporte vectorial o entrenamientos de redes neuronales, todas ellas herramientas muy potentes de la inteligencia artificial, requieren de conocimiento y de métodos.





2. Distancias e intersecciones

Las distancias se pueden calcular entre puntos, entre rectas y puntos o entre rectas paralelas.

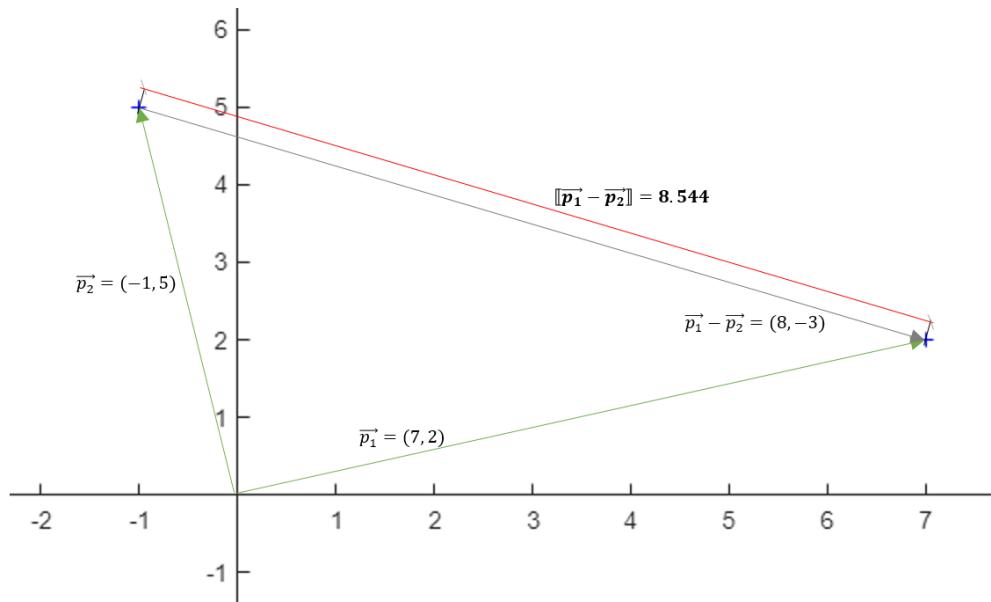
2.1. Distancia entre puntos

Definición. Sean dos vectores $\vec{p}_1, \vec{p}_2 \in \mathbb{R}^2$ asociados a las coordenadas $p_1 = (x_1, y_1)$ y $p_2 = (x_2, y_2)$. La distancia entre ellos se calcula como la norma del vector resta de sus vectores:

$$d_{p_1 p_2} = \|\vec{p}_1 - \vec{p}_2\|$$

Se tienen los puntos $p_1 = (7, 2)$ y $p_2 = (-1, 5)$ la distancia entre ellos es

$$d_{p_1 p_2} = \sqrt{(7 - (-1))^2 + (2 - 5)^2} = \sqrt{73} = 8.544$$

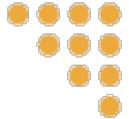


2.2. Distancia entre rectas y puntos

Definición. Sean una recta r de la forma $r \equiv Px + Qy + R = 0$ y un punto $p_0 = (x_0, y_0)$, la distancia entre ellos es la norma del vector que va desde p_0 hasta el punto de intersección con la recta en dirección perpendicular a ésta, y se calcula como

$$d_{p_0 r} = \frac{|P * x_0 + Q * y_0 + R|}{\sqrt{P^2 + Q^2}}$$

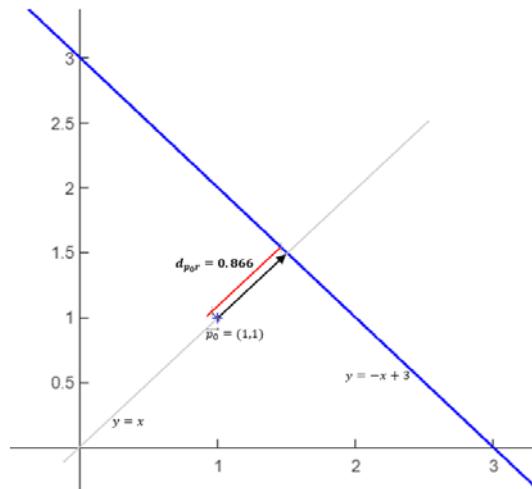
Nota: El símbolo \equiv significa 'idéntico a' y se usa para referirse a elementos definidos por una ecuación, como una recta.



2. Distancias e intersecciones

Se tienen el punto $p_0 = (1,1)$ y la recta $r_1 \equiv y = -x + 3$. Su forma implícita se obtiene dejando a cero uno de los miembros de la ecuación, $x - y - 3 = 0$, y de aquí se obtiene $P = 1, Q = -1, R = 3$. Así, se tiene:

$$d_{p_0r} = \frac{|1 * 1 + (-1) * 1 + 3|}{\sqrt{1^2 + 1^2}} = \frac{3}{\sqrt{2}}$$

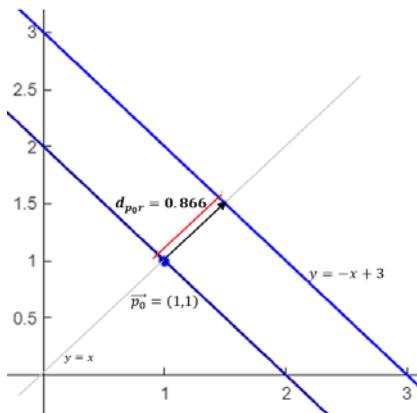


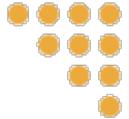
2.3. Distancia entre rectas paralelas

El caso del cálculo de distancia entre dos rectas paralelas es un caso concreto del cálculo entre una recta y un punto, dando la libertad de elegir el punto entre los infinitos que pasan por una de las rectas.

Similar al ejemplo anterior, si se tuviesen ahora las rectas $r_1 \equiv y = -x + 3$ y $r_2 \equiv y = -x + 2$, se podría elegir de r_2 el punto $p_0 = (1,1)$ y el problema seguiría exactamente igual que el anterior.

Para encontrar un punto aleatorio por la que pase una recta, tan solo habrá que sustituir un valor aleatorio de x y despejar la y de la ecuación para obtener la coordenada del punto.





3. Análisis geométrico de funciones

El análisis geométrico de funciones se hace a través de las simetrías, las rectas asintóticas, las derivadas e integrales de la curva.

3.1. Simetrías

3.1.1. Funciones pares

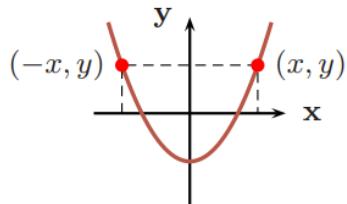
Definición. La función $f(x)$ es una función par y por lo tanto **simétrica con respecto al eje de ordenadas** si, para todo x dentro del dominio de la función, las imágenes para ese valor y para su opuesto son coincidentes. Esto es:

$$f(-x) = f(x), \quad \forall x \in D$$

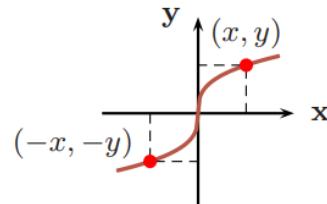
3.1.2. Funciones impares

Definición. La función $f(x)$ es una función impar y por lo tanto **simétrica con respecto al origen de coordenadas** si, para todo x dentro del dominio de la función, la imagen del valor opuesto de x es igual a la imagen opuesta de x . Esto es:

$$f(-x) = -f(x), \quad \forall x \in D$$



Función par: $f(-x) = f(x)$



Función impar: $f(-x) = -f(x)$

3.2. Rectas asintóticas

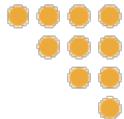
Una recta es asintota de una función $f(x)$ si los valores de ésta última se acercan infinitamente a la recta.

3.2.1. Asintota horizontal

Sea la recta horizontal $y = a$ y la función $f(x)$, y será asintota de $f(x)$ si el límite de $f(x)$ cuando x tiende a $+\infty$ o $-\infty$ es a . Es decir, se cumple

$$\lim_{x \rightarrow \pm\infty} f(x) = a$$

Si esto ocurre cuando x tiende a $-\infty$, se le llama *asíntota horizontal por la izquierda*; si ocurre cuando x tiende a $+\infty$, será una *asíntota horizontal por la derecha*.



3. Análisis geométrico de funciones

3.2.2. Asíntota vertical

Sea la recta vertical $x_0 = a$ y la función $f(x)$, x_0 es asíntota vertical de $f(x)$ si el límite de $f(x)$ por la derecha o izquierda de a tiende a infinito. Esto es

$$\lim_{x \rightarrow a^\pm} f(x) = \pm\infty$$

Puede ser recta asíntota por la izquierda, por la derecha o por ambos lados en función de si existe el límite sólo cuando x tiene a por la izquierda, sólo cuando lo hace por la derecha o en ambos casos.

Las rectas asintóticas verticales suelen ser aquellas rectas $x_0 = a$ que coinciden con valores no definidos en f .

3.2.3. Asíntota oblicua

Sea la recta vertical $y = mx + n$ y la función $f(x)$, la recta y es asíntota oblicua de $f(x)$ si se cumple

$$\lim_{x \rightarrow \pm\infty} f(x) - (mx + n) = 0$$

Una vez más, la recta puede ser asíntota por la izquierda, por la derecha o por ambos lados. Para encontrar la recta, se puede buscar a partir de sus componentes

$$m = \lim_{x \rightarrow \pm\infty} \frac{f(x)}{x}, \quad n = \lim_{x \rightarrow \pm\infty} [f(x) - mx]$$

Se tienen las funciones $f_1(x) = e^x$, $f_2(x) = \frac{1}{x^2}$, $f_3(x) = \frac{1}{x} + 5x + 1$ y se quiere comprobar si tiene rectas asintóticas. En primer lugar, se buscan rectas asintóticas horizontales

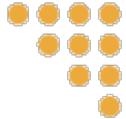
Para la primera función $f_1(x) = e^x$:

- $\lim_{x \rightarrow +\infty} e^x = e^\infty = \infty$.
- $\lim_{x \rightarrow -\infty} e^x = e^{-\infty} = 0 \rightarrow y = 0$ es una asíntota horizontal por la izquierda de $f_1(x)$.

Para la segunda $f_2(x) = \frac{1}{x^2}$:

- $\lim_{x \rightarrow +\infty} \frac{1}{x^2} = \frac{1}{\infty} = 0 \rightarrow y = 0$ es una asíntota horizontal por la derecha de $f_2(x)$.
- $\lim_{x \rightarrow -\infty} \frac{1}{x^2} = \frac{1}{-\infty} = 0 \rightarrow y = 0$ es una asíntota horizontal por la izquierda de $f_2(x)$.

Continúa ⇒



3. Análisis geométrico de funciones

Continuación

Y para la tercera $f_3(x) = \frac{1}{x} + 5x + 1$:

- $\lim_{x \rightarrow +\infty} (\frac{1}{x} + 5x + 1) = \frac{1}{\infty} + 5 * \infty + 1 = 0 + \infty + 1 = +\infty$
- $\lim_{x \rightarrow -\infty} (\frac{1}{x} + 5x + 1) = \frac{1}{-\infty} + 5 * (-\infty) + 1 = 0 - \infty + 1 = -\infty$

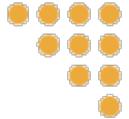
Ahora se pasa a buscar posibles rectas asintóticas verticales:

- $f_1(x) = e^x$ es continua para todo el dominio de X , por lo que no tiene rectas asintóticas verticales.
- $f_2(x) = \frac{1}{x^2}$ no está definida para $x_0 = 0$, así que se comprueba si la recta vertical $X = 0$ es asintota vertical de $f_2(x)$:
 - $\lim_{x \rightarrow 0^-} \frac{1}{x^2} = \infty \rightarrow X = 0$ es una asintota vertical por la izquierda.
 - $\lim_{x \rightarrow 0^+} \frac{1}{x^2} = \infty \rightarrow X = 0$ es una asintota vertical por la derecha.
- $f_3(x) = \frac{1}{x} + 5x + 1$ tampoco está definida para $x_0 = 0$. Se analiza en ese punto:
 - $\lim_{x \rightarrow 0^-} (\frac{1}{x} + 5x + 1) = \infty \rightarrow X = 0$ es una asintota vertical por la izquierda.
 - $\lim_{x \rightarrow 0^+} (\frac{1}{x} + 5x + 1) = \infty \rightarrow X = 0$ es una asintota vertical por la derecha.

Por último, se buscan asíntotas oblicuas:

- Para la primera función $f_1(x) = e^x$:
 - $m = \lim_{x \rightarrow +\infty} \frac{e^x}{x} = \frac{e^x}{x} = \infty$.
 - $m = \lim_{x \rightarrow -\infty} \frac{e^x}{x} = \frac{e^x}{x} = 0 \rightarrow$ Cuando $m = 0$ es asintota horizontal.
- Para la segunda $f_2(x) = \frac{1}{x^2}$:
 - $m = \lim_{x \rightarrow +\infty} \frac{\frac{1}{x^2}}{x} = \lim_{x \rightarrow -\infty} \frac{1}{x^3} = 0 \rightarrow$ Asintota horizontal ya obtenida.
 - $m = \lim_{x \rightarrow -\infty} \frac{\frac{1}{x^2}}{x} = \lim_{x \rightarrow -\infty} \frac{1}{x^3} = 0 \rightarrow$ Asintota horizontal ya obtenida.

Continúa ⇨



3. Análisis geométrico de funciones

Continuación

Y finalmente para la tercera $f_3(x) = \frac{1}{x} + 5x + 1$:

- $m = \lim_{x \rightarrow +\infty} \frac{\left(\frac{1}{x}+5x+1\right)}{x} = \lim_{x \rightarrow +\infty} \frac{1+5x^2+x}{x^2} = \frac{\infty}{\infty} = 5^*$
- $m = \lim_{x \rightarrow -\infty} \frac{\left(\frac{1}{x}+5x+1\right)}{x} = \lim_{x \rightarrow -\infty} \frac{1+5x^2+x}{x^2} = \frac{-\infty}{-\infty} = 5^*$

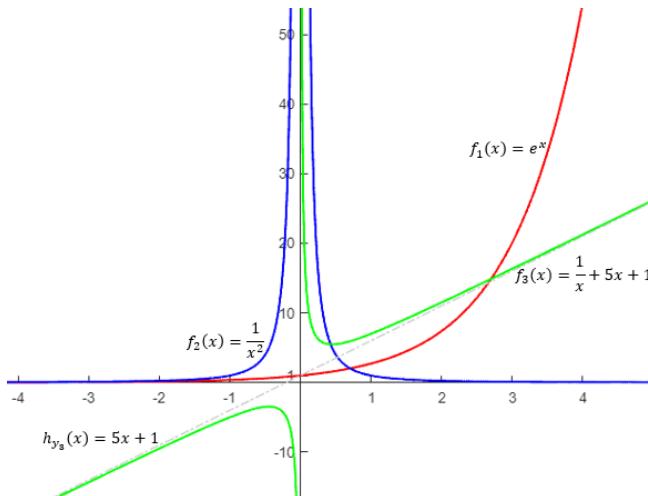
*Forma indeterminada resuelta por la regla de L'Hopital, que indica que se deben derivar por separado las funciones del numerador y denominador hasta encontrar el valor

$$\frac{[1+5x^2+x]'}{[x^2]'} = \frac{[10x+1]'}{[2x]'} = \frac{10}{2} = 5$$

Teniendo m es posible calcular n

$$n = \lim_{x \rightarrow \pm\infty} [f(x) - mx] = \lim_{x \rightarrow \pm\infty} \left(\frac{1}{x} + 5x + 1 - 5x \right) = \lim_{x \rightarrow \pm\infty} \frac{1}{x} + 1 = 1$$

Luego, la asíntota oblicua por la izquierda y la derecha de $f_3(x)$ es $h_{y_3}(x) = 5x + 1$

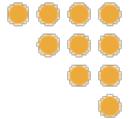


3.3. Derivada de la curva

Las derivadas de funciones $f(x)$ pueden interpretarse con las funciones $f'(x)$ que definen, para cada x , las rectas tangentes a la curva en ese punto.

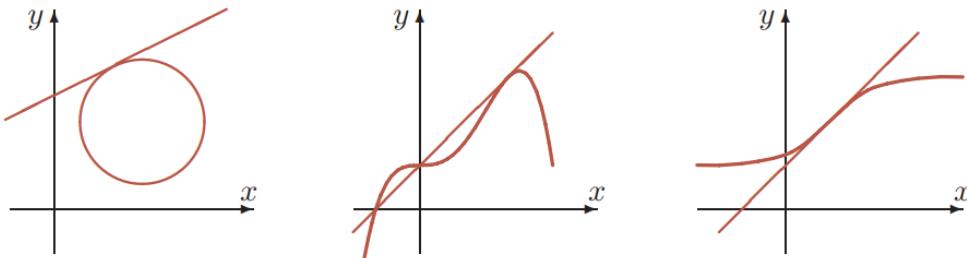
3.3.1. Recta tangente

Formalmente, la tangente a una curva en un punto x_0 es la recta que toca a la curva solo en dicho punto. Esto es fácil de ver en una circunferencia, pero al generalizar



3. Análisis geométrico de funciones

la definición para cualquier tipo de curva, surgen cuestiones que esta definición no es capaz de resolver.



Definición. Sea la función de la curva $y = f(x)$ y un punto x_0 dentro del dominio de f , la recta tangente a $f(x)$ en x_0 es aquella que pasa por el punto $(x_0, f(x_0)) = (x_0, y_0)$ con la misma dirección que la propia curva y y cuya pendiente es igual a $f'(x_0)$, derivada de la función en x_0 . Se define con la ecuación:

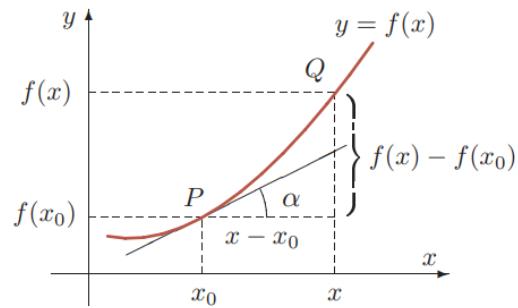
$$y - f(x_0) = f'(x_0)(x - x_0)$$

Es sabido que la derivada de la función se define como el límite:

$$f'(x) = \lim_{\Delta x \rightarrow 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} = \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

Y, finalmente, del mismo nombre, es fácil comprender que es posible obtener el ángulo de esta pendiente a partir de:

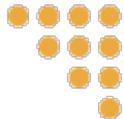
$$f'(x) = \tan(\alpha)$$



3.3.2. Monotonía, máximos y mínimos relativos

A partir de la ecuación derivada de una función $f(x)$, es posible conocer la pendiente de la curva de dicha función para cualquier punto x_0 , ¿Qué información práctica proporciona esto?:

- Si $f'(x_0) > 0 \rightarrow$ La curva es **creciente** en x_0 .
- Si $f'(x_0) < 0 \rightarrow$ La curva es **decreciente** en x_0 .



3. Análisis geométrico de funciones

Efectivamente, si la recta tangente a una curva en un punto es aquella que pasa por el mismo punto con la **misma dirección** a la curva si esa dirección es positiva, la pendiente de la recta tangente, $f'(x_0)$, lo será también; al estar hablando sobre variaciones instantáneas, se tendrá la certeza de que el próximo punto infinitesimal de x será superior a x_0 y el anterior será inferior, definición de *crecimiento*.

El **máximo absoluto** de una función $f(x)$ es aquel punto x_0 que tiene el mayor valor $f(x_0)$ para todo el dominio de la función. Del mismo modo, el **mínimo absoluto** es aquel punto x_0 que tiene el menor valor $f(x_0)$ de todo el dominio de la función. Expresado matemáticamente:

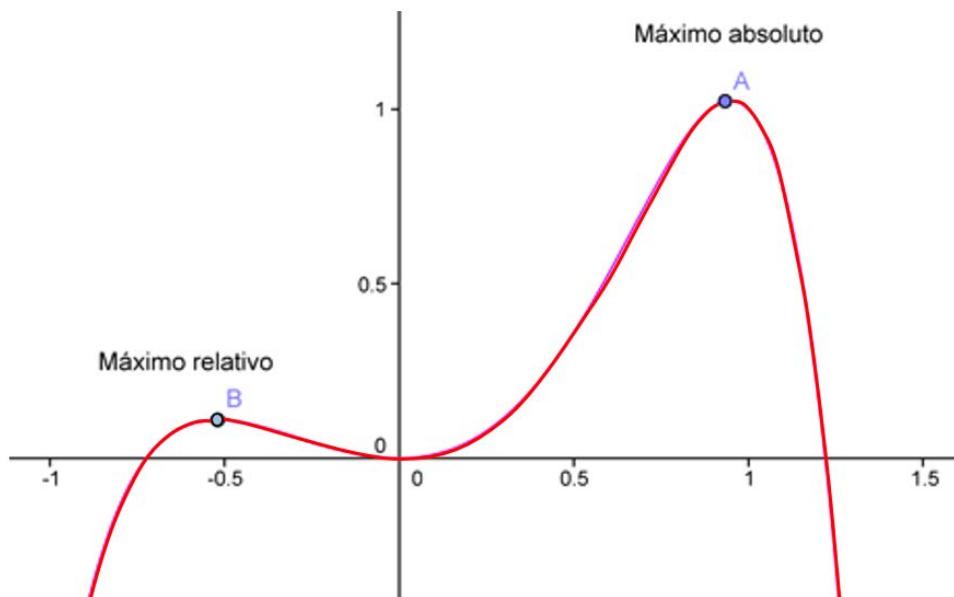
$$\text{Máximo absoluto} \rightarrow f(x_0) \geq f(x), \forall x \in D: f$$

$$\text{Mínimo absoluto} \rightarrow f(x_0) \leq f(x), \forall x \in D: f$$

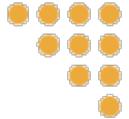
Los **máximos y mínimos relativos** son aquellos que cumplen una condición parecida a los máximos y mínimos absolutos pero restringida tan sólo a sus vecinos infinitesimales inmediatos. Es decir:

$$\text{Máximo relativo} \rightarrow f(x_0 - \Delta x) \leq f(x_0) \geq f(x_0 + \Delta x), \Delta x \rightarrow 0$$

$$\text{Mínimo relativo} \rightarrow f(x_0 - \Delta x) \geq f(x_0) \leq f(x_0 + \Delta x), \Delta x \rightarrow 0$$

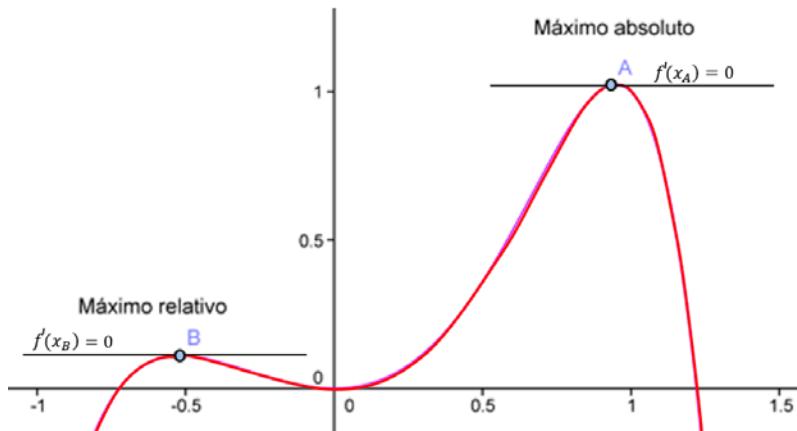


Pero ¿es posible calcular analíticamente los máximos y mínimos relativos de una función $f(x)$? Si la derivada de una función en un punto determina la dirección de la curva en ese punto o, dicho de otro modo, su condición creciente o decreciente, ¿qué dirección lleva la curva en un máximo o un mínimo relativo?



3. Análisis geométrico de funciones

La tangentes a una curva en sus puntos máximos y mínimos forman rectas horizontales y, por lo tanto, de pendiente $m = 0$.

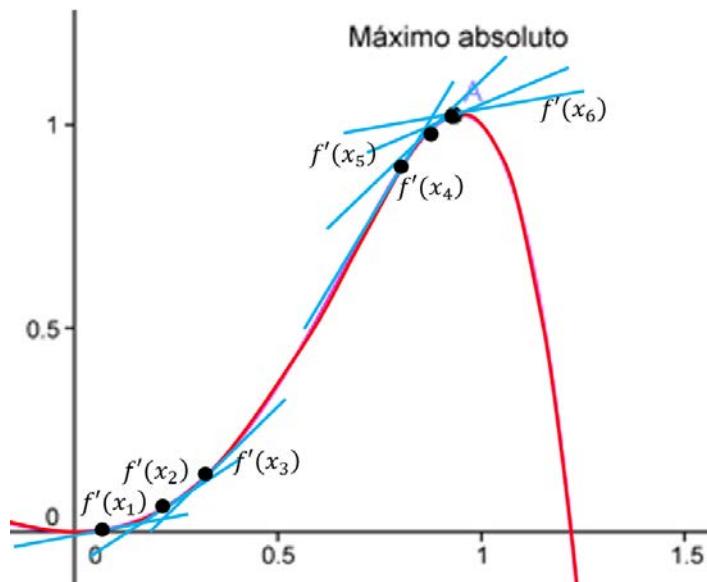


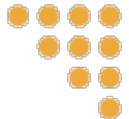
Para conocer los puntos máximos o mínimos de una función $f(x)$ se deben encontrar los puntos x_i que cumplen:

$$f'(x) = 0$$

3.3.3. Criterio de Concavidad y punto de inflexión

Si se presta atención a la imagen anterior, es posible diferenciar dos subintervalos dentro del intervalo creciente \overline{OA} , un primer intervalo en el cual la curva parece ir ‘acelerando’ y un segundo intervalo en el que desacelera. Si fueran calculadas las derivadas en varios puntos del intervalo, podría comprobarse cómo durante el primer intervalo el valor de estas derivadas crece con x alcanza un máximo en algún punto intermedio \overline{OA} y comienza a disminuir hasta llegar a cero en $x = A$, momento a partir del cual la función comienza a decrecer y por lo tanto su derivada pasa a ser negativa.





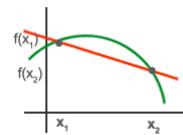
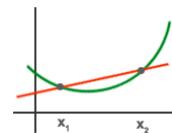
3. Análisis geométrico de funciones

Pero ¿qué herramienta analítica podría usarse para estudiar la variación media instantánea que lleva, a su vez, la propia variación de la función? Realmente, se pretende medir la variación instantánea de la variación instantánea de la función, la derivada de la derivada, es decir, **la derivada segunda, $f''(x)$** .

El comportamiento creciente o decreciente de la derivada $f'(x)$ de una función, indica la **concavidad o convexidad** de la función primitiva, $f(x)$.

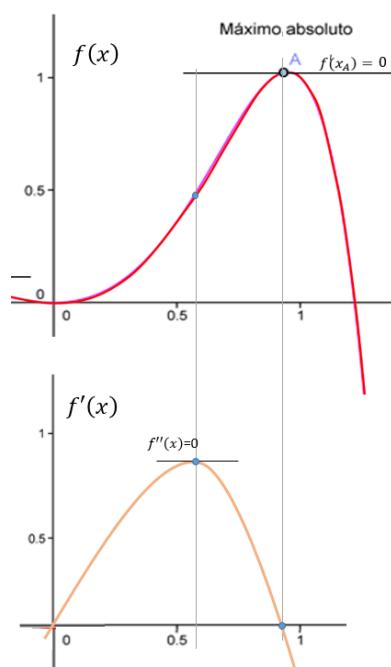
La concavidad/convexidad son términos que se definen geométricamente así:

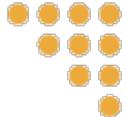
- La curva de una función $f(x)$ es **convexa** si, al unir dos puntos cualesquiera de la misma con un segmento, éste queda siempre por encima de la curva.
- La curva de una función $f(x)$ es **cóncava** si, al unir dos puntos cualesquiera de la misma con un segmento, éste queda siempre por debajo de la curva.



Si se vuelve a inspeccionar el intervalo \overline{OA} la imagen anterior, se identifica con facilidad al primer intervalo cuyas derivadas crecían como convexo y al segundo, cuyas derivadas disminuían, como cóncavo. Esto ayuda a comprender el siguiente criterio de convexidad:

- Si en un intervalo $(a, b) \in D$: $f''(x) > 0 \rightarrow$ la curva en (a, b) es convexa.
- Si en un intervalo $(a, b) \in D$: $f''(x) < 0 \rightarrow$ la curva en (a, b) es cóncava.





3. Análisis geométrico de funciones

Estos puntos donde la curva pasa de ser cóncava a convexa o al contrario se denominan **puntos de inflexión**. Con lo explicado hasta el momento se pueden identificar intervalos de concavidad y convexidad, puntos de inflexión, así como localizar máximos y mínimos relativos. Pero ¿cómo discriminar entre si el punto x_0 donde $f'(x_0) = 0$ es un máximo o un mínimo? Observando la imagen de arriba, donde se muestra la equivalencia de la función $f(x)$ con su derivada $f'(x)$, podemos encontrar detalles que diferencian los máximos y los mínimos.

Las curvas de máximos relativos siempre son cóncavas y las curvas donde se encuentran los mínimos relativos siempre convexas; es decir, en un máximo relativo x_0 de $f(x)$ su derivada $f'(x)$ es decreciente y, por lo tanto, su derivada segunda $f''(x)$ menor que cero.

$$f(x) \left\{ \begin{array}{l} f'(x_0) > 0 \rightarrow f(x) \text{ Creciente en } x_0 \\ f'(x_0) < 0 \rightarrow f(x) \text{ Decreciente en } x_0 \\ f'(x_0) = 0 \left\{ \begin{array}{l} f''(x_0) < 0 \rightarrow \text{Máximo relativo en } x_0 \\ f''(x_0) > 0 \rightarrow \text{Mínimo relativo en } x_0 \end{array} \right. \\ \rightarrow f''(x_0) = 0 \rightarrow \text{Punto de inflexión en } x_0 \end{array} \right.$$

Como ejemplo, se va a analizar la función $\rightarrow f(x) = -x^3 + 15x^2 - 3x + 1$

1. Se comienza buscando los tramos crecientes y decrecientes extrayendo los máximos y mínimos relativos. Para eso se calcula su derivada y se iguala a cero:

$$f'(x) = -3x^2 + 30x - 3 = 0 \quad \rightarrow \quad x_1 = 0.101, \quad x_2 = 9.899$$

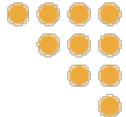
Esto quiere decir que en los puntos x_1 y x_2 hay máximos o mínimos relativos, y que el resto de la función en su dominio será estrictamente creciente o estrictamente decreciente. En estos puntos la función vale $f(x_1) = 0.849$, $f(x_2) = 471.151$.

2. Como $x_1 < x_2$, se sabe que en el intervalo (x_1, x_2) la función es creciente.

Se podría ahora calcular la derivada $f'(x)$ en cualquier punto de los intervalos $(-\infty, x_1)$ y (x_2, ∞) y conocer, en función de su signo si es creciente o decreciente.

$$f'(-5) = -228, \quad f'(10) = -3 \rightarrow \text{en ambos intervalos decrece.}$$

Continúa ⇒



3. Análisis geométrico de funciones

Continuación

3. Ahora, se calcula la derivada segunda $f''(x)$ y se iguala a cero para buscar puntos de inflexión.

$$f''(x) = -6x + 30 = 0 \rightarrow x_3 = 5$$

En $x_3 = 5$ se encuentra un punto de inflexión.

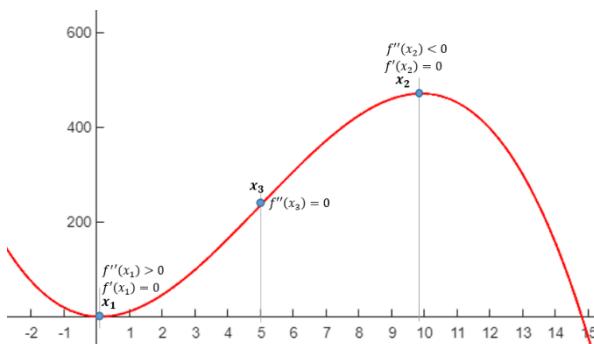
4. Finalmente, para saber cuál de x_1 y x_2 era máximos y cuál mínimo relativo, se calcula su valor en su derivada segunda para conocer su signo.

$$f''(x_1) = 29.394 > 0, \quad f''(x_2) = -29.394 < 0$$

Por lo tanto, en el punto crítico x_1 la curva es convexa y en x_2 es cóncava.

Resumiendo, se tiene:

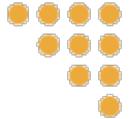
- En $x_1 = 0.101$ hay un mínimo relativo; en $x_2 = 9.899$ un máximo relativo.
- Los intervalos $(-\infty, x_1)$ y (x_2, ∞) la función decrece; en (x_1, x_2) crece.
- En $x_3 = 5$ hay un punto de inflexión.



3.3.4. Funciones multivariable. Derivadas parciales

Es importante resaltar, que todo lo visto hasta ahora se puede aplicar a casos de funciones de más de una variable. Para el caso tridimensional, los intervalos crecientes y decrecientes pasarían a ser superficies curvas con total libertad e independencia entre sus variables de entrada x e y .

En el caso univariable, un incremento podía ser medido de una única forma, a lo largo del eje x , única variable independiente; ahora el incremento podría ir en cualquier dirección del espacio vectorial XY . Esto quiere decir que, partiendo de un punto $p_0 = (x_0, y_0, z_0)$, el comportamiento creciente o decreciente dependería de en qué dirección se llevaría a cabo el incremento. Lo que es lo mismo, sobre qué variable se deriva la función, en estos casos se habla de **derivada parcial**.



3. Análisis geométrico de funciones

Definición. Dada una función de dos variables $f(x, y): D \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$, se define la derivada parcial de $f(x, y)$ con respecto a x en el punto $\mathbf{p} = (x_0, y_0)$ como el valor del límite:

$$\frac{\partial f}{\partial x}(x_0, y_0) = \lim_{\Delta x \rightarrow 0} \frac{f(x_0 + \Delta x, y_0) - f(x_0, y_0)}{\Delta x}$$

Y, del mismo modo, se define la derivada parcial de $f(x, y)$ con respecto a y en \mathbf{p} , como el siguiente límite:

$$\frac{\partial f}{\partial y}(x_0, y_0) = \lim_{\Delta y \rightarrow 0} \frac{f(x_0, y_0 + \Delta y) - f(x_0, y_0)}{\Delta y}$$

En la práctica, ante una función $f(x, y)$ su derivada parcial sobre x deberá realizarse como si el resto de las variables (en este caso y) fuese un valor constante.

En estos casos, para encontrar extremos relativos, ambas derivadas parciales $\frac{\partial f}{\partial x}$ y $\frac{\partial f}{\partial y}$ deberían ser cero en el mismo punto. Ya se sabe que, para saber si los extremos encontrados son máximos o mínimos, hay que estudiar la derivada, en este caso parcial, segunda para descubrir su signo. Pero, en estos casos tridimensionales existe un nuevo caso singular.

Se tiene la función $f(x, y) = x^2 - y^2$.

Para determinar sus extremos relativos, se calculan sus derivadas parciales. En derivada parcial respecto a x , el término $-y^2$ se trata como una constante, como si fuese el número 3, por ejemplo. En la derivada parcial respecto y , se hará lo mismo con x :

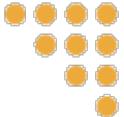
$$\frac{\partial f}{\partial x}(x, y) = 2x + 0, \quad \frac{\partial f}{\partial y}(x, y) = 0 - 2y$$

Al igualar estas derivadas parciales a 0 se encuentra el punto:

$$2x = 0 \rightarrow x = 0, \quad 2y = 0 \rightarrow y = 0 \quad \rightarrow \quad \mathbf{p} = (0, 0)$$

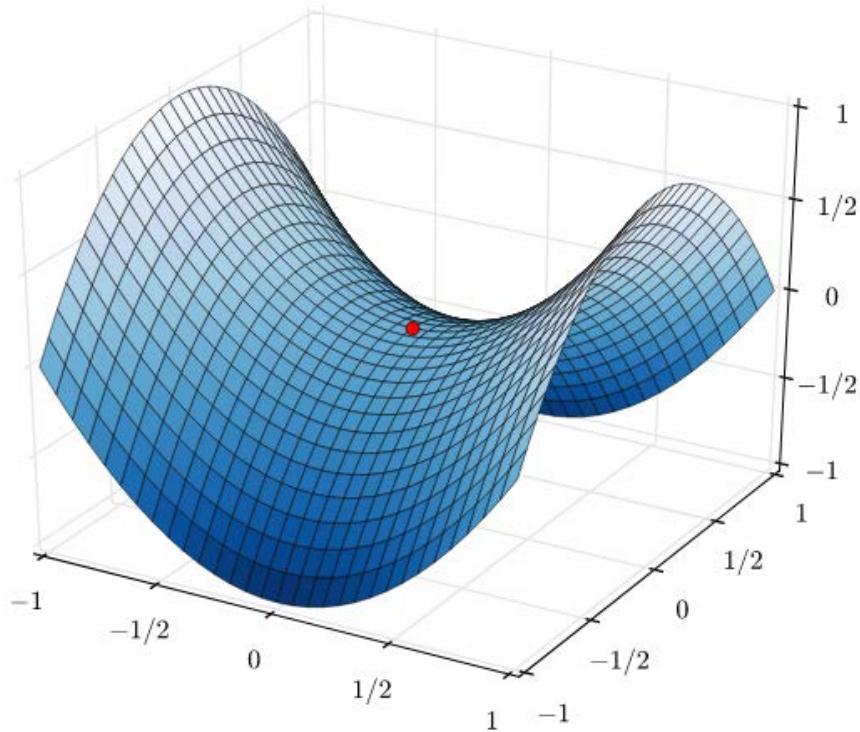
Cuando se vuelve a derivar para conocer, mediante el signo, su concavidad y, de ahí, si es máximo o mínimo, se descubre que:

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



3. Análisis geométrico de funciones

Esto quiere decir que el punto $\mathbf{p} = (0,0)$ es un mínimo relativo en el eje x , pero máximo relativo en el eje y . Estos puntos singulares se denominan **punto de silla**.



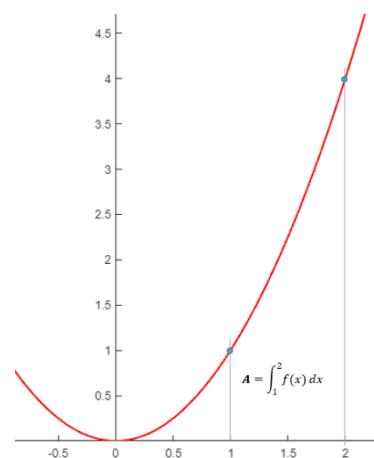
3.4. Integral de la curva

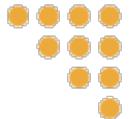
Se debe recordar que la integral se puede interpretar como la recomposición o suma de porciones infinitesimales de un cuerpo. En el caso de una función $\int f(x)dx$, se sumarán todas las barras verticales de altura $f(x)$ y base infinitamente pequeña, dx . Pero ¿y en el caso de una función $f(x, y)$?

3.4.1. Área bajo la curva

Es ya sabido que es posible calcular el área bajo la curva de una función $f(x)$ entre los intervalos a y b mediante la *integral definida* de la función entre los límites de integración a y b . Así, es posible calcular, por ejemplo, el área que encierra la función $f(x) = x^2$ entre $a = 1$ y $b = 2$ mediante:

$$A = \int_1^2 x^2 dx = \left[\frac{x^3}{3} \right]_1^2 = \frac{(2)^3}{3} - \frac{(1)^3}{3} = \frac{7}{3} = 2.3$$

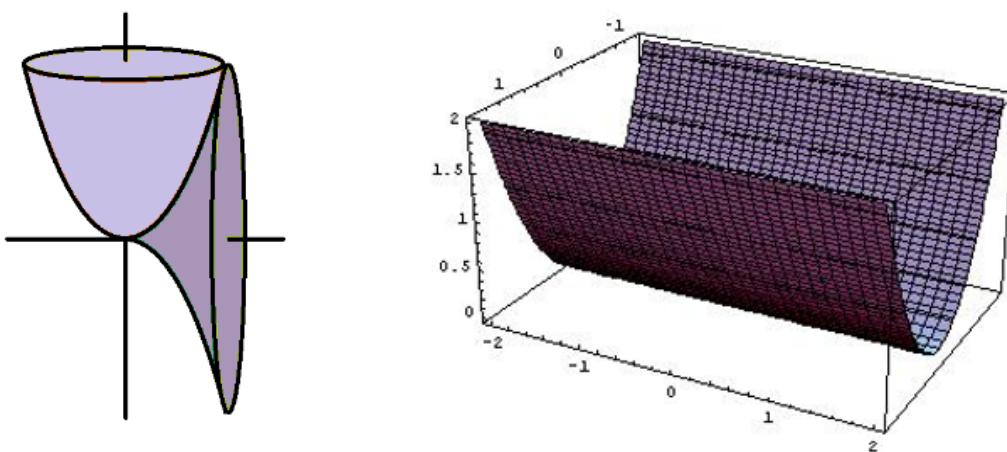




3. Análisis geométrico de funciones

3.4.2. Volumen encerrado

De un modo similar, cuando se trabaja sobre funciones multivariable, es posible obtener distintos volúmenes de la función anterior $f(x) = x^2$. Por un lado, podría extenderse la curva a lo largo de un nuevo eje mediante $\int f(x)dz$, o podrían desarrollarse lo que se conoce como volúmenes de revolución sobre el eje OX o OY .



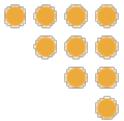
SAMSUNG

BeJob[™]

Together for Tomorrow!
Enabling People

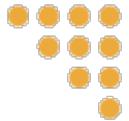
Probabilidad

#TecnologíaConPropósito



Índice

Fundamentos de la probabilidad	5
1. <u>Introducción</u>	7
2. <u>Sucesos aleatorios</u>	8
2.1. <u>Experimento aleatorio</u>	8
2.2. <u>Probabilidad clásica</u>	9
2.3. <u>Tipos de sucesos</u>	9
3. <u>Probabilidad frecuentista</u>	11
Teoría de conjuntos	13
1. <u>Introducción</u>	15
2. <u>Representación de conjuntos</u>	16
3. <u>Operaciones de conjuntos</u>	17
3.1. <u>Tipos de conjuntos</u>	17
3.2. <u>Relaciones de conjuntos</u>	18
3.3. <u>Operaciones con sucesos</u>	19
4. <u>Cálculo de probabilidades</u>	23
4.1. <u>Teoremas</u>	23
4.2. <u>Leyes de Morgan</u>	24
Condicionalidad	27
1. <u>Introducción</u>	29
2. <u>Sucesos dependientes</u>	30
3. <u>Sucesos independientes</u>	32
Regla de la Multiplicación, Probabilidad Total y Teorema Bayes	37
1. <u>Introducción</u>	39



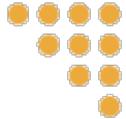
Índice

2. Regla de la Multiplicación	40
3. Teorema de la Probabilidad Total	42
4. Teorema de Bayes	44
Técnicas de conteo	47
1. Introducción	49
2. Fundamentos del conteo	50
 2.1. Terminología	50
 2.2. Herramientas del conteo	51
 2.3. Principios del conteo	54
3. Variaciones	56
 3.1. Variaciones sin repetición	56
 3.2. Variaciones con repetición	57
4. Permutaciones	58
 4.1. Permutaciones sin repetición	58
 4.2. Permutaciones con repetición	59
5. Combinaciones	60
 5.1. Combinaciones sin repetición	60
 5.2. Combinaciones con repetición	61
6. Resumen	62

Together for Tomorrow!
Enabling People

Fundamentos de la probabilidad

#TecnologíaConPropósito



1. Introducción

La innata curiosidad del ser humano ha hecho que desde siempre se haya interesado tanto por el motivo por el que ocurren los fenómenos como por adivinar lo que deparará el futuro.

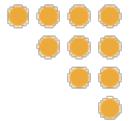
En muchas ocasiones, la comprensión del fenómeno era completa, de manera que el problema podía ser modelado y obtener de este modelo el resultado o estado final, dadas unas determinadas condiciones iniciales. Sin embargo, en otras ocasiones los fenómenos escapan al determinismo y no parece posible someterlos a las leyes descubiertas hasta el momento, y por tanto imposibilitan ante una determinada situación o experiencia concluir un resultado determinado. Es tal el número de factores de los que de partida dependen que se hace inviable la búsqueda de reglas que la determinen.



La probabilidad nace a partir de una jugada de dados que obsesionaba a un antiguo escritor y jugador francés del siglo XVII, Antoine Gombaud (1607-1684), amigo del también matemático francés Blaise Pascal (1623-1662), al cual pedía consejo respecto a las garantías de éxito que ofrecía dicha jugada.

“En el fondo, la teoría de probabilidades es solo sentido expresado con números.”

Pierre-Simon Laplace, astrónomo, físico y matemático francés (1749-1827)



2. Sucesos aleatorios

Se presentan dos experimentos. En el primero se deja caer desde cierta altura un objeto y se anota el tiempo que tarda en tocar el suelo; en el segundo se lanza una moneda al aire y se anota si sale cara o cruz. ¿Qué diferencia elemental puede apreciarse entre estos dos experimentos? Este epígrafe tratará esta cuestión, y otras al respecto, cuya respuesta forma los fundamentos de la probabilidad.

2.1. Experimento aleatorio

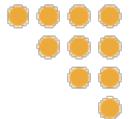
Un **fenómeno o experimento determinista** es aquel que, cuando se reproduce en las mismas condiciones, es posible predecir con certeza cuál será el resultado, la observación de en qué derivará. Por el contrario, el **fenómeno o experimento aleatorio** es aquel que, en cada manifestación, aunque se repita en condiciones idénticas, derivará en un resultado impredecible y solo conocido tras su realización. Estos resultados son llamados **sucesos o eventos**.

La **incertidumbre** que implican estos experimentos aleatorios es el objeto primero de la Estadística y la Probabilidad, que se encarga de cercar los posibles resultados al mínimo espacio de categorías potenciales. Este conjunto de resultados posibles se denomina **espacio muestral, E** , y los elementos que lo componen se denominan **sucesos elementales, s** .

Pero definir el espacio muestral con los posibles resultados en los que puede acabar un fenómeno aleatorio no es la única forma de reducir el nivel de incertidumbre; no sería lo mismo lanzar un dado común y esperar que salga un cinco, que lanzar un dado que tuviera cinco caras con el valor cinco y una con el valor dos, o lanzar un dado en el que saliera el valor cuatro en la mayoría de los lanzamientos por algún defecto físico en su geometría suele salir el. De todo esto se deduce la necesidad de medir, cuantificar, la incertidumbre que se tiene sobre cada suceso posible, y esta medición se lleva a cabo a través de la **probabilidad**.

Por ejemplo; el tiempo que tarda un objeto en tocar el suelo desde una determinada altura está regido por las leyes gravitacionales de la física clásica. Es un fenómeno modelable y, por lo tanto, conociendo cierta información acerca de la condición inicial del fenómeno —la altura este caso— se puede conocer el resultado del fenómeno (el suceso) antes de llevarlo a cabo, y aunque se repitiese cientos de veces seguiría dando el mismo resultado.

Sin embargo, cuando se lanza una moneda al aire no es posible predecir con exactitud cuál será el resultado y, en caso de resultar “cara”, esto no proporcionaría



2. Sucesos aleatorios

ninguna información adicional que redujese la incertidumbre de los siguientes experimentos si se repitiesen.

Los sucesos elementales $S_1 = \text{cara}$ y $S_2 = \text{cruz}$ forman el espacio muestral del experimento $E = \{S_1, S_2\}$.

2.2. Probabilidad clásica

La probabilidad de un suceso se define como el cociente entre el número de casos favorables y el número total de casos, siempre que todos sean igualmente posibles.

Si un experimento aleatorio consiste en el lanzamiento de un dado, su espacio muestral estará integrado por seis sucesos elementales, uno por cada cara. Conocida la igualdad de posibilidades para cada una de las caras del dado que componen el espacio muestral y sabiendo que para cada experimento tan solo una de las caras del dado puede salir como resultado, se tiene que la probabilidad de que resulte un suceso elemental cualquiera, pongamos que salga el número cuatro, $S = 4$, es de

$$P(S_4) = \frac{\text{suceso}}{\text{espacio muestral}} = \frac{1}{6}$$

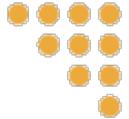
2.3. Tipos de sucesos

Los sucesos elementales que componen un espacio muestral pueden, a su vez, formar un **suceso compuesto**, es decir, un subconjunto del espacio muestral. Así, si en el experimento aleatorio del lanzamiento de un dado interesa ahora el suceso de que caiga un número par cualquiera, entonces el suceso compuesto estará formado por los sucesos elementales $\{2,4,6\}$. Entonces,

$$P(S_{par}) = \frac{\text{suceso compuesto}}{\text{espacio muestral}} = \frac{\{2,4,6\}}{\{1,2,3,4,5,6\}} = \frac{3}{6} = 0.5$$

Ahora se toma como caso el experimento o fenómeno de obtener un número concreto de la ruleta o rueda de la fortuna. La ruleta se compone de 37 números, por lo que la probabilidad del suceso (el 0 no se cuenta como número par en este juego) sería

$$P(S_{par}) = \frac{\text{suceso}}{\text{espacio muestral}} = \frac{\{2,4,6, \dots, 36\}}{\{0,1,2, \dots, 36\}} = \frac{18}{37} = 0.486$$



2. Sucesos aleatorios

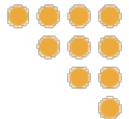
Dado un experimento, aquel suceso compuesto que contiene la totalidad de sucesos elementales del espacio muestral se denomina **suceso cierto**. Si, por el contrario, el suceso consiste en que ocurra algo que no incluye el espacio muestral será un **suceso imposible**.

Usando ejemplos anteriores, el suceso de que salga algún número entre al 0 y el 50 en la ruleta, sería un suceso cierto, ya el intervalo $[0,50]$ encierra todos los posibles valores (los 37 primeros enteros positivos). Que salga un número mayor que 40 es un suceso imposible.

De las definiciones anteriores se obtiene una propiedad básica de la probabilidad, y es que esta estará siempre acotada entre 0 y 1.

$$0 \leq P(S) \leq 1$$

- Suceso cierto $P(S) = 1, \quad S = E$
- Suceso imposible $P(S) = 0, \quad S \notin E$



3. Probabilidad frecuentista

La definición anterior de probabilidad se basa en un enfoque de experimentos aleatorios puntuales, en la probabilidad de que un experimento aleatorio aislado derive en un determinado suceso. Ante la imposibilidad de asumir la necesaria condición que presuponen como igual de probables todos los sucesos elementales de un espacio muestral, se hace necesario un enfoque diferente.

Son dos premisas las que derivan en el concepto *frecuentista* de la probabilidad:

- En primer lugar, la llamada estabilidad de las frecuencias o *regularidad estadística* supone que, a pesar del comportamiento irregular o azaroso de los resultados individuales, los resultados promedios, en largas sucesiones de experimentos aleatorios, muestran una sorprendente regularidad.
- Como segunda premisa, se tiene la objetividad de la probabilidad, la cual puede ser directamente una propiedad física de los objetos, como su peso o su densidad y, por tanto, medible.

La frecuencia absoluta de un suceso en el desarrollo de un experimento aleatorio, repetido de forma independiente N veces, es el número n de veces que ocurre el suceso. Por tanto, la frecuencia relativa es el cociente entre esta frecuencia absoluta y el número total de veces que se ha repetido el experimento.

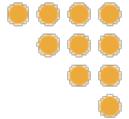
$$f = \frac{n}{N}$$

De nuevo, el experimento aleatorio del lanzamiento del dado se realiza ahora con un dado real, -es decir, con todos los defectos físicos que pueda tener, lo que puede conducir a una situación mínimamente desigual de oportunidades entre los distintos elementos del espacio muestral- y el suceso de interés consiste en obtener un 6.

El experimento se repite 10 veces y en ninguna ocasión se obtiene un 6, lo que resulta en una frecuencia relativa de 0. Realizamos diez veces más el experimento hasta llegar a veinte y el suceso ocurre dos veces; esto hace una frecuencia relativa de $\frac{2}{10} = 0,1$.

Aumentando el número de veces que se lanza el dado a 100 veces se cuentan 17 veces las que sale un 6, por lo que la frecuencia relativa esta vez es de $\frac{17}{100} = 0,17$.

Aumentando el número de experimento la frecuencia va variando 0,173, 0,16, 0,162, ..., 0,168,...



3. Probabilidad frecuentista

Esta teoría frecuentista admite que aumentando indefinidamente en cada experimento el número de veces que se lanza el dado se obtiene la probabilidad de obtener un 6 con ese dado concreto.

Si n es la frecuencia absoluta del suceso S y N el número total de veces que se repite el experimento aleatorio entonces

$$\lim_{N \rightarrow \infty} \frac{n}{N} = P(S)$$

Propiedades

- La frecuencia relativa de un suceso está comprendida entre 0 y 1

$$0 \leq \frac{n}{N} \leq 1$$

- Si un suceso S es la unión de un número finito de sucesos disjuntos S_1, S_2, \dots, S_k , su frecuencia absoluta n es la suma de las frecuencias absolutas de cada uno de los n sucesos, n_i , se tiene

$$f(S) = \frac{n}{N} = \frac{\sum_{i=1}^k n_i}{N} = \frac{n_1}{N} + \dots + \frac{n_k}{N} = f(S_1) + \dots + f(S_k)$$

Y, por lo tanto, se tiene que la probabilidad es aditiva entre los sucesos

$$P(S) = P(S_1) + \dots + P(S_k)$$

Si el suceso de interés al lanzar el dado fuese obtener un número par se tendría un suceso compuesto por $S = \{S_2, S_4, S_6\} = \{2, 4, 6\}$. Si el experimento repetido realizado anteriormente para el cálculo de que ocurra el suceso S_3 se repite para el caso S_2 y S_4 , y se obtiene $P(S_2)$ y $P(S_4)$, después de 100 experimentos para cada suceso elemental se obtendría

$$P(S_2) = 0,1593$$

$$P(S_4) = 0,1644$$

$$P(S_6) = 0,168$$

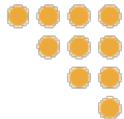
La anterior propiedad dice que sería posible calcular la probabilidad de S como la suma de las probabilidades individuales de cada uno de los sucesos elementales por separado

$$P(S) = P(S_1) + P(S_1) + P(S_1) = 0,4917$$

Together for Tomorrow!
Enabling People

Teoría de conjuntos

#TecnologíaConPropósito

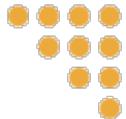


1. Introducción

El uso del concepto **conjunto** ha acompañado al ser humano a lo largo de todas sus edades. Es el estudio inherente a la capacidad asociativa que tenemos para darle significado o sentido a lo que nos rodea. Solemos agrupar todas las ideas y relacionarlas en función a sus similitudes, y la teoría de conjuntos es la ciencia que formaliza esta práctica.

El desarrollo de esta disciplina se le debe a Georg Cantor (1845-1918), matemático, físico y filósofo alemán, de origen ruso, del siglo XIX. Georg asienta las definiciones de conjunto, de elementos y sus reglas y operaciones, pero, a medida que profundiza en cuestiones conjuntistas del infinito, más y más paradojas van apareciendo con intenciones de invalidar sus bases de la teoría de conjuntos. Esta continua lucha por mantener la consistencia en sus definiciones, ante las contradicciones que iban apareciendo, llevan a Cantor a una profunda depresión, lo que hace que pase los últimos años de su vida en un centro psiquiátrico.

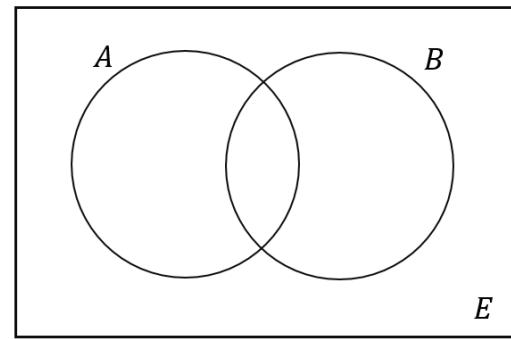
Hoy la teoría de conjunto es ampliamente estudiada por numerosas disciplinas de muy variada naturaleza, como la filosofía, la lingüística, la estadística o la electrónica.



2. Representación de conjuntos

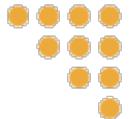
Toda agrupación, colección o reunión de objetos definidos y distinguibles que puedan concebirse como una unidad mediante el cumplimiento de cierta propiedad o regla se denomina **conjunto**. Así, la teoría de conjuntos es la rama de las matemáticas que trabaja sobre unas bases conceptuales que tratan estas entidades y colecciones y sus posibles relaciones y operaciones. Dada tal abstracción en la interpretación de elementos y conjuntos de elementos se hacen necesarias herramientas gráficas que ayuden a dar sentido a los distintos resultados que sus interacciones derivan. Existen dos herramientas principales de este tipo (aunque a lo largo de la unidad solo se usará una) los **diagramas de Venn** y los **diagramas de Carroll**.

Los **diagramas de Venn** consisten en figuras geométricas planas y cerradas, dentro de las cuales se encierran cada uno de los elementos o subconjuntos que lo forman. Por lo general, los conjuntos y subconjuntos se representarán mediante circunferencias; los elementos que lo forman serán representados mediante puntos, cuando sea necesario. Además, el diagrama siempre aparece, a su vez, encerrado en un rectángulo que representa el *conjunto universal*.



Los **diagramas de Carroll** tienen la misma utilidad, dándole más prioridad a la visualización de los atributos que definen los conjuntos. Estos atributos son de tipo binario (dan respuestas tipo sí/no). El diagrama consiste en dos columnas y dos filas; las filas categorizan los elementos de uno de los atributos y las columnas hacen lo mismo del segundo atributo.

	Primos	No primos
Pares	2,	4, 6, 8, 10, 12, 14, 16, 18,
No pares	3, 5, 7, 11, 13, 17, 23, 29,	1, 9, 15, 21, 25, 27, 33, 35,



3. Operaciones entre conjuntos

Para realizar operaciones con conjuntos de deberán conocer los tipos de conjuntos, sus relaciones y las operaciones con sucesos.

3.1. Tipos de conjuntos

Por definición los conjuntos agrupan elementos que se relacionan por alguna característica o propiedad. Si un conjunto posee un número determinado de elementos se puede denominar **conjunto finito**, término usado para diferenciarlo del *conjunto infinito*, que poseen infinitos elementos, como por ejemplo el conjunto de números reales o el número de puntos que posee una recta.

El término usado para definir a ese otro conjunto que alberga a todos los existentes es **conjunto universal, E** .

$$S = E$$

Este conjunto posee ciertas propiedades:

- El conjunto universal no es único, sino referencial. Existe un conjunto universal para cada grupo de conjuntos definidos.
- Todo conjunto pertenece a un conjunto universal, $S \subset E, \forall S$.

En los casos en que un conjunto S no contiene ningún elemento se denomina **conjunto vacío** o conjunto nulo, denotado como \emptyset ,

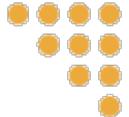
$$S = \emptyset$$

Se pueden extraer las siguientes propiedades del conjunto vacío:

- El conjunto vacío es único (no existe ‘un’ conjunto vacío, sino el conjunto vacío).
- El conjunto vacío pertenece a todo conjunto, $\emptyset \subset S, \forall S$.
- Si un conjunto pertenece al conjunto vacío, es el conjunto vacío, $S \subset \emptyset \leftrightarrow S = \emptyset$.

Se conocen como **conjunto unitario** y **conjunto binario** a aquellos que tienen uno y dos elementos, respectivamente. Los sucesos elementales que forman un espacio muestral podrían ser interpretados como conjuntos unitarios, S_i , pertenecientes al conjunto universal, E .

En caso de ser sucesos compuestos, serían conjuntos finitos pertenecientes a E que contendrían los distintos sucesos elementales, como conjuntos unitarios. Finalmente, un suceso imposible se definiría con el conjunto vacío.

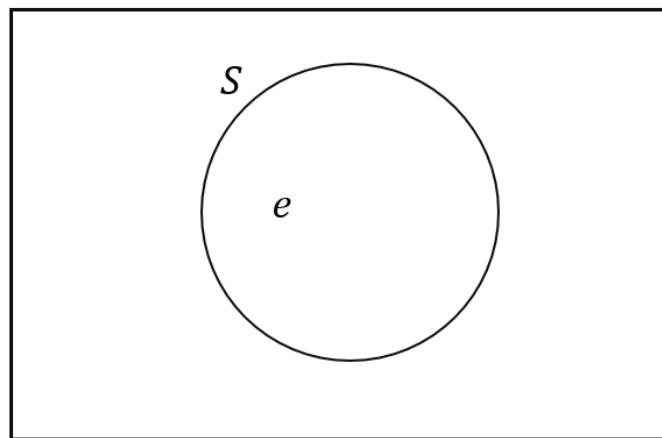


3. Operaciones entre conjuntos

3.2. Relaciones de conjuntos

Pertenencia. Sea S un conjunto cualquiera y e un elemento cualquiera, si e es un elemento de S , es decir, pertenece a S , se simboliza,

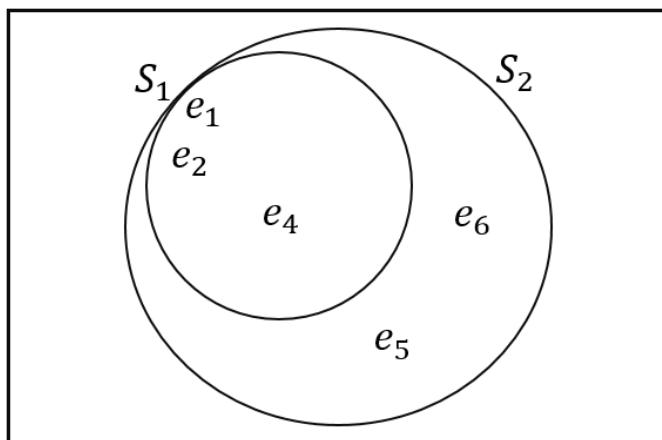
$$e \in S$$



Inclusión. Un conjunto S_1 es un subconjunto de otro conjunto S_2 , lo que se denota como $S_1 \subset S_2$, si cada elemento de S_1 pertenece a S_2 pero no todos los elementos de S_2 pertenecen a S_1 . Esto es,

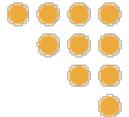
$$S_1 \subset S_2 \leftrightarrow \forall e (e \in S_1 \rightarrow e \in S_2) \wedge \exists e \setminus (e \in S_2 \rightarrow e \notin S_1)$$

que se lee como “ S_1 es un subconjunto de S_2 si, y solo si, para todo elemento e , si pertenece a S_1 pertenece también a S_2 y, a la vez, existe, al menos, un elemento e el cual, si pertenece a S_2 no lo hace a S_1 .”

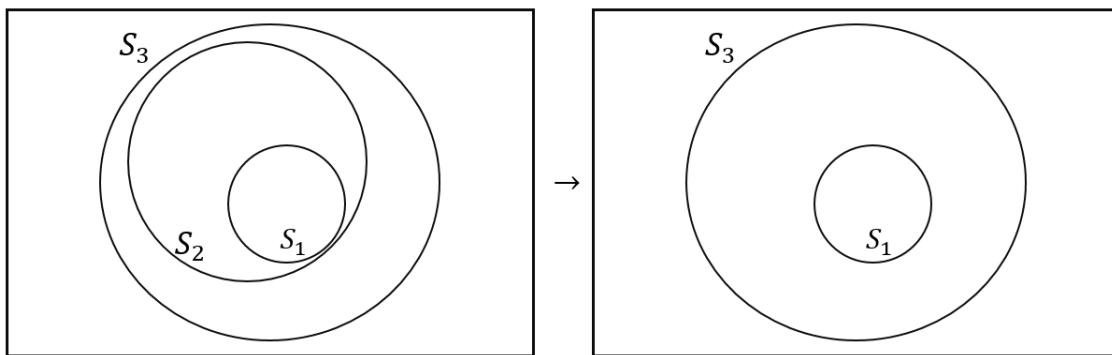


En la inclusión se tienen las propiedades:

- Todo conjunto es subconjunto de sí mismo $S_1 \subseteq S_1$
- Si $(S_1 \subset S_2) \wedge (S_2 \subset S_3) \rightarrow S_1 \subset S_3$

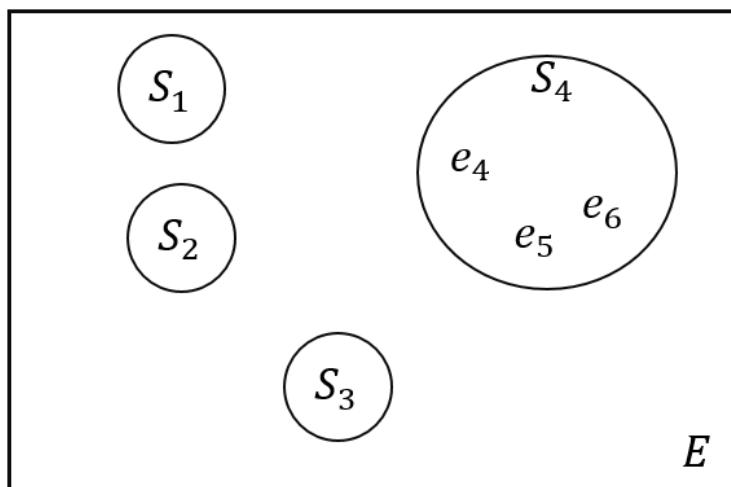


3. Operaciones entre conjuntos



Igualdad. Dos conjuntos son considerados iguales $S_1 = S_2$ si cada elemento de S_1 está contenido en S_2 , $S_1 \subset S_2$, y cada elemento de S_2 estaría contenido en S_1 , $S_2 \subset S_1$. Es decir,

$$S_1 = S_2 \leftrightarrow \forall e (e \in S_1 \leftrightarrow e \in S_2)$$



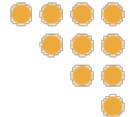
3.3. Operaciones con sucesos

A la hora de realizar operaciones con sucesos se ha de tener en cuenta la unión de sucesos, la intersección de sucesos, los sucesos complementarios, las diferencias de sucesos y las diferencias simétricas de sucesos.

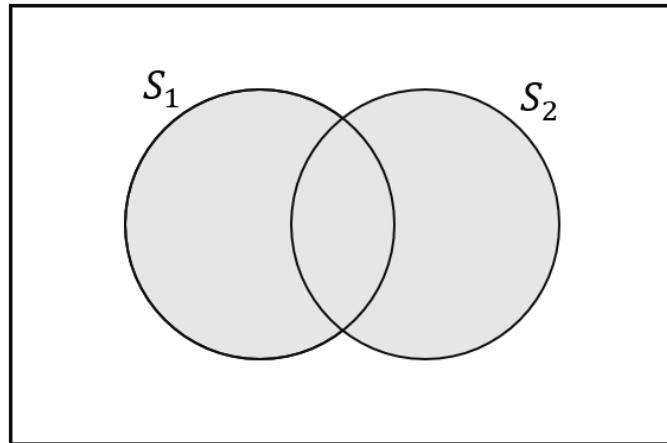
3.3.1. Unión de sucesos

Dado un conjunto de n sucesos S_1, S_2, \dots, S_n , la unión de éstos es otro suceso formado por los sucesos elementales comunes y los no comunes de todos ellos.

$$\bigcup_{i=1}^n S_i = S_1 \cup S_2 \cup \dots \cup S_n$$



3. Operaciones entre conjuntos



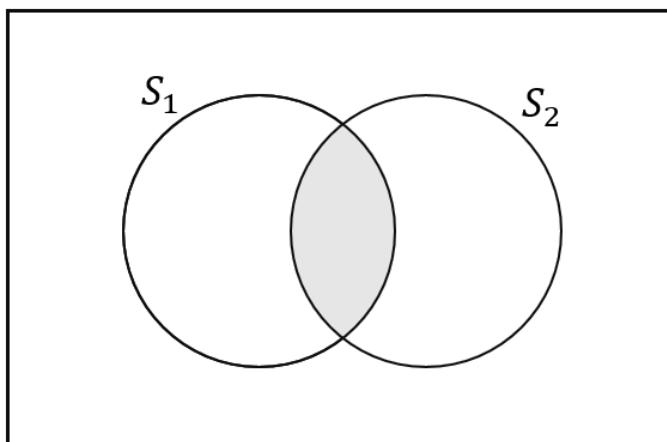
En la lógica proposicional esta unión se denomina **disyunción**, puesto que son operaciones que responden a relaciones entre conjuntos de tipo ($S_1 \text{ o } S_2 \text{ o } \dots \text{ o } S_n$). También se puede encontrar en otras disciplinas como la ingeniería, la electrónica o las ciencias de la computación como **suma lógica** u operación 'OR'.

Como casos particulares existen:

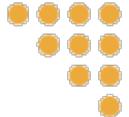
- $S \cup \emptyset = S$
- $S \cup E = E$
- $\emptyset \cup \emptyset = \emptyset$

3.3.2. Intersección de sucesos

Dado un conjunto de n sucesos S_1, S_2, \dots, S_n , la intersección de éstos es otro suceso formado por los sucesos comunes a todos ellos. Equivalen a relaciones conjunción en lógica proposicional, es decir, un suceso que tiene lugar cuando se dan simultáneamente S_1 y S_2 y ... y S_n . Son llamadas **multiplicación lógica** u operación 'AND'.



$$\bigcap_{i=1}^n S_i = S_1 \cap S_2 \cap \dots \cap S_n$$



3. Operaciones entre conjuntos

Se verifican los siguientes casos particulares:

- $S \cap \emptyset = \emptyset$
- $S \cap E = S$
- $\emptyset \cap \emptyset = \emptyset$

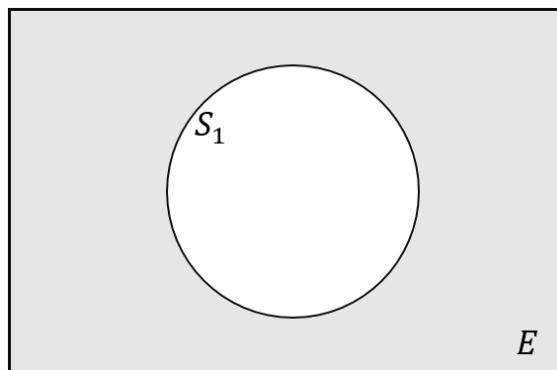
Las anteriores operaciones tienen las propiedades que se describen a continuación:

Comutativa	Asociativa	Distributiva
$S_1 \cup S_2 = S_2 \cup S_1$	$S_1 \cup (S_2 \cup S_3) = (S_1 \cup S_2) \cup S_3$	$S_1 \cup (S_2 \cup S_3) = (S_1 \cup S_2) \cup (S_1 \cup S_3)$
$S_1 \cap S_2 = S_2 \cap S_1$	$S_1 \cap (S_2 \cap S_3) = (S_1 \cap S_2) \cap S_3$	$S_1 \cap (S_2 \cap S_3) = (S_1 \cap S_2) \cap (S_1 \cap S_3)$

3.3.3. Suceso complementario

El complemento de un suceso S , denominado S^* , es el suceso compuesto por los elementos de E que no pertenecen a S . Es también conocido como **negación**. Por la misma definición, se entienden los siguientes casos:

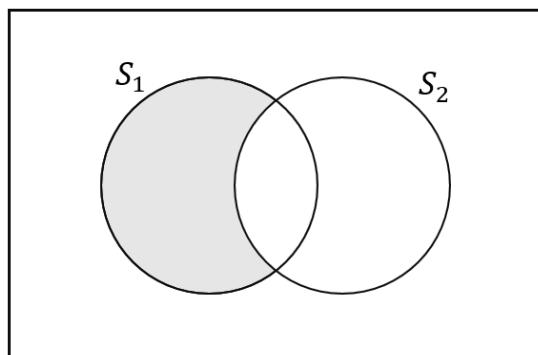
- $S \cup S^* = E$
- $S \cap S^* = \emptyset$
- $(S^*)^* = S$

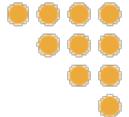


3.3.4. Diferencia de sucesos

La diferencia entre dos conjuntos S_1 y S_2 son aquellos elementos que contiene S_1 pero no S_2 . Simbólicamente quedan definidos como:

$$S_1 - S_2 = \{e \mid e \in S_1 \text{ y } e \notin S_2\}$$



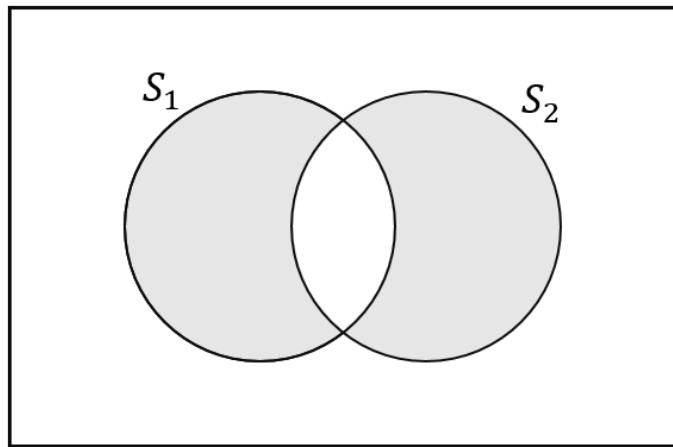


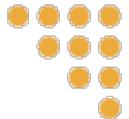
3. Operaciones entre conjuntos

3.3.5. Diferencia simétrica de sucesos

Es aquella que, dados dos sucesos S_1 y S_2 , produce otro suceso formado por los elementos no comunes de ambos sucesos:

$$S_1 \oplus S_2 = \{e \mid e \in (S_1 - S_2) \vee e \in (S_2 - S_1)\}$$





4. Cálculo de probabilidades

Mediante lo desarrollado en los apartados anteriores, junto a ciertos teoremas y reglas que serán explicadas a continuación, será posible solucionar problemas de cálculo de probabilidades mediante la aplicación de la teoría de conjuntos.

4.1. Teoremas

Existen 5 teoremas:

Teorema 1

La probabilidad del suceso imposible es cero.

$$P(\emptyset) = 0$$

Teorema 2

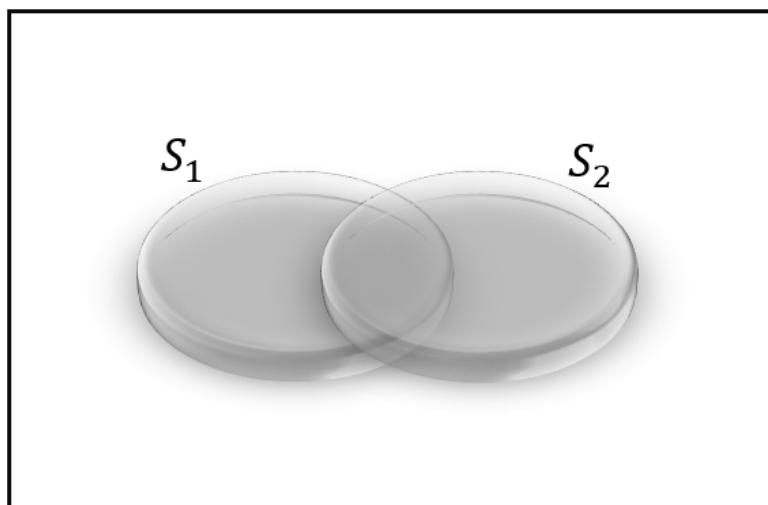
La probabilidad de la unión de n sucesos disjuntos S_1, S_2, \dots, S_n es igual a la suma de las probabilidades individuales de cada suceso.

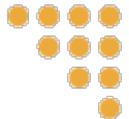
$$P\left(\bigcup_{i=1}^n S_i\right) = \sum_{i=1}^n P(S_i)$$

Teorema 3

La probabilidad de la unión de dos sucesos cualesquiera es igual a la suma de sus probabilidades menos la intersección de ambas

$$P(S_1 \cup S_2) = P(S_1) + P(S_2) - P(S_1 \cap S_2)$$





4. Cálculo de probabilidades

A partir del diagrama de Venn para $S_1 \cup S_2$, es posible deducir que la simple suma de los sucesos tiene en cuenta la zona de intersección, $S_1 \cap S_2$, dos veces, motivo por el cuál es necesario restar esa ‘porción’ de alguno de los sucesos.

Para el caso de tres sucesos, lo anterior se desarrolla:

$$P\left(\bigcup_{i=1}^3 S_i\right) = P(S_1) + P(S_2) + P(S_3) - P(S_1 \cap S_2) - P(S_1 \cap S_3) - P(S_2 \cap S_3) \\ + P(S_1 \cap S_2 \cap S_3)$$

Y, generalizando a n sucesos:

$$P\left(\bigcup_{i=1}^n S_i\right) = \sum_{i=1}^n P(S_i) - \sum_{\substack{i,j=1 \\ i < j}}^n P(S_i \cap S_j) + \sum_{\substack{i,j,k=1 \\ i < j < k}}^n P(S_i \cap S_j \cap S_k)$$

Teorema 4

Si un suceso S_1 está contenido en otro, es decir, $S_1 \subset S_2$, se cumple

$$P(S_1) \leq P(S_2)$$

Teorema 5

La probabilidad del suceso complementario, S^* es:

$$P(S^*) = 1 - P(S)$$

4.2. Leyes de Morgan

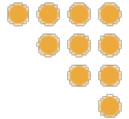
Existen un par de reglas de transformación en la lógica de conjuntos que permite la transformación de conjunciones en disyunciones y viceversa, por medio de la negación o complementario:

1. La negación de la conjunción es la disyunción de las negaciones; o, lo que es lo mismo, el complementario de la unión de dos conjuntos es igual a la intersección de los complementarios de los mismos conjuntos:

$$(S_1 \cup S_2)^* = S_1^* \cap S_2^*$$

2. La negación de la disyunción es la conjunción de las negaciones; o, lo que es lo mismo, el complementario de la intersección de dos conjuntos es igual a la unión de los complementarios de los mismos conjuntos:

$$(S_1 \cap S_2)^* = S_1^* \cup S_2^*$$



4. Cálculo de probabilidades

Volviendo a hacer uso del fenómeno aleatorio de lanzamiento de dados no perfectos, esta vez lanzando dos dados, se presenta el siguiente suceso: consiste en que salgan los números 5 y 6 en una tirada. Esto es,

$$S = \{S_5, S_6\}$$

Como el suceso requiere que ocurra el suceso 5 **y** el suceso 6 en el experimento, se traduce como la búsqueda de la intersección de estos dos sucesos elementales.

$$P(S_5 \cap S_6)$$

Previamente se ha calculado, mediante probabilidad frecuentista que, con el dado empleado, la probabilidades de que ocurran estos sucesos, así como la probabilidad de que no salga ni el 5 ni el 6, son

$$P(S_5) = 0,1599, \quad P(S_6) = 0,1680, \quad P(S_5^* \cap S_6^*) = 0,6989$$

Un modo de calcularlo puede ser el siguiente: usando el teorema de la probabilidad complementaria junto a la segunda ley de Morgan, se obtiene

$$P(S_5 \cap S_6) = 1 - P[(S_5 \cap S_6)^*] = 1 - P(S_5^* \cup S_6^*)$$

Esta transformación , sería lo mismo que plantear lo siguiente: la probabilidad de que salgan un 5 y un 6 en el mismo lanzamiento de dos dados es la contraria o complementaria a que salga cualquier otra combinación posible que no sea esa, $P[(S_5 \cap S_6)^*]$, como salir dos cincos, dos seises, un uno y un 4, etc. A su vez, esperar que la combinación cinco y seis no salgan en el mismo experimento, es lo mismo que la unión de esperar que no salga el cinco junto a que no salga el 6. Usando ahora la fórmula para la unión de sucesos se tiene

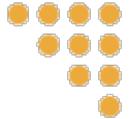
$$P(S_5^* \cup S_6^*) = P(S_5^*) + P(S_6^*) - P(S_5^* \cap S_6^*)$$

$$\begin{aligned} P(S_5 \cap S_6) &= 1 - [P(S_5^*) + P(S_6^*) - P(S_5^* \cap S_6^*)] \\ &= 1 - [(1 - 0,1599) + (1 - 0,168) - 0,6989] = \mathbf{0,0268} \end{aligned}$$

Together for Tomorrow!
Enabling People

Condisionalidad

#TecnologíaConPropósito

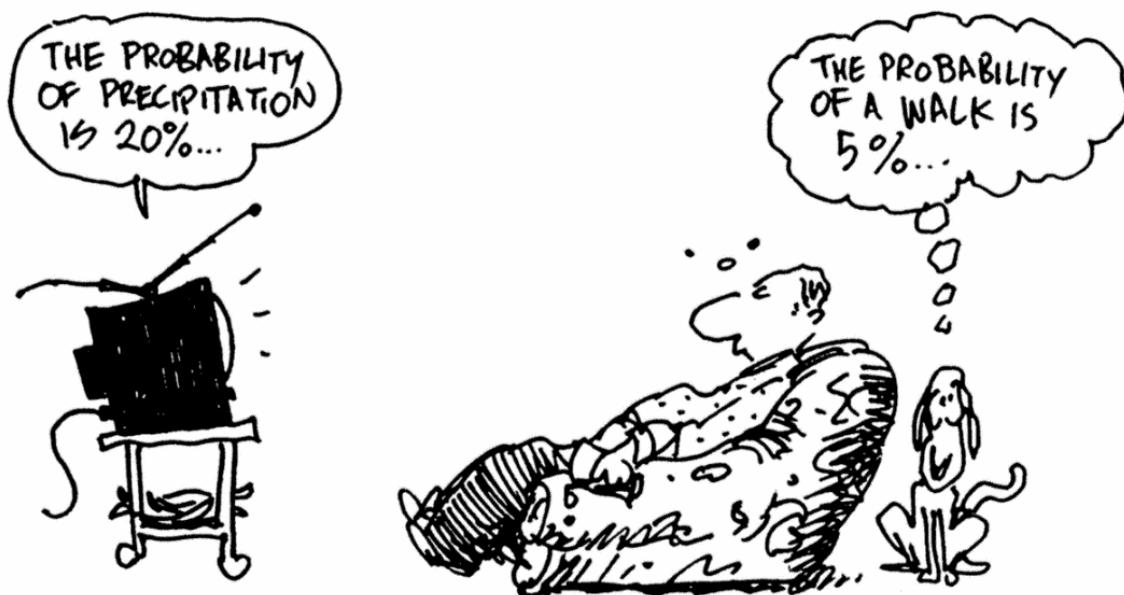


1. Introducción

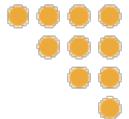
Debe tenerse claro que, en la mayoría de casos, a mayor información de calidad sobre un asunto, mayor probabilidad de tomar decisiones correctas, de acertar. Si la probabilidad busca reducir, mediante la información disponible, la llamada incertidumbre acerca del resultado de un fenómeno aleatorio, es fácilmente entendible esta relación entre información disponible y probabilidad de acierto.

Pero, ¿cuálquier información es útil? ¿Conocer la persona que lanza los dados impactaría en el resultado?

La búsqueda de relaciones y patrones entre información de entrada y la salida que se pretende predecir define uno de los primeros pasos en cualquier proyecto o tarea dentro del mundo de la ciencia de los datos y, una buena forma de introducirse en este tipo de análisis es mediante la llamada probabilidad condicional. Según la existencia o ausencia de la condicionalidad entre distintos sucesos, los problemas deben ser tratados de distinto modo y ambas serán descritas a lo largo de esta unidad, que está estructurada acorde a esta misma división.



Fuente: *Probability of a walk* del libro [Cartoon Guide to Statistics](#)



2. Sucesos dependientes

Considérense los dos sucesos siguientes ante un mismo experimento del lanzamiento de dados: que salga un seis y que salga un múltiplo de tres.

A simple vista, parece más probable el segundo suceso ya que es un suceso compuesto de mayor tamaño. Siendo un dado ideal, la probabilidad de que ocurra el primer suceso es

$$P(S_6) = \frac{1}{6},$$

pero si ya se sabe que el segundo suceso se ha cumplido, es decir, se sabe que el número que ha salido es múltiplo de 3, la probabilidad del suceso anterior, el primero, pasa a ser

$$P(S_6) = \frac{1}{2},$$

porque dentro del espacio muestral $E = \{1, 2, 3, 4, 5, 6\}$ sólo dos números, el 3 y el 6, son múltiplos de 3.

Esta información adicional ha provocado que sea posible reducir el espacio muestral y así aumentar la probabilidad del suceso.

El conocimiento del suceso ‘es múltiplo de 3’ **condiciona** el resultado del siguiente suceso.

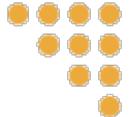
Esta información adicional es llamada **evidencia** o **condicionante**.

El suceso que es ayudado por la evidencia, se denomina **hipótesis** o **condicionado**.

Definición. Dados dos sucesos S_1 y S_2 , el suceso S_1 está condicionado por el suceso S_2 si la probabilidad de que suceda S_1 depende de que haya sucedido S_2 , y la **probabilidad condicional**, en este caso, viene definida como

$$P(S_1|S_2) = \frac{P(S_1 \cap S_2)}{P(S_2)},$$

siempre que $P(S_2) > 0$



2. Sucesos dependientes

Usando esta definición en el ejemplo inicial, se tendría

$$P(S_6 | \text{multiplo } 3) = \frac{P(S_6 \cap \text{multiplo } 3)}{P(\text{multiplo } 3)}$$

La intersección $P(S_6 \cap \text{multiplo } 3)$ encierra los casos en los que es múltiplo de 3 y es 6. En este caso, el primer suceso pertenece al segundo. Por otro lado, ya se dijo arriba que sólo dos números dentro del espacio muestral es múltiplo de 3, por lo que se cumple

$$P(S_6 \cap \text{multiplo } 3) = P(S_6) = \frac{1}{6}$$

$$P(\text{multiplo } 3) = \frac{2}{6}$$

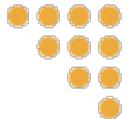
Finalmente, tenemos

$$P(S_6 | \text{multiplo } 3) = \frac{P(S_6 \cap \text{multiplo } 3)}{P(\text{multiplo } 3)} = \frac{\frac{1}{6}}{\frac{2}{6}} = \frac{1}{2}$$

Esta definición de probabilidad condicional se extiende, fácilmente, a más de dos sucesos. Para el caso de tres sucesos, es posible establecer los siguientes casos condicionales $\{S_1 | S_2 \cap S_3\}$ y $\{S_1 \cap S_2 | S_3\}$. Sus fórmulas quedarían como

$$P(S_1 | S_2 \cap S_3) = \frac{P(S_1 \cap S_2 \cap S_3)}{P(S_2 \cap S_3)},$$

$$P(S_1 \cap S_2 | S_3) = \frac{P(S_1 \cap S_2 \cap S_3)}{P(S_3)},$$



3. Sucesos independientes

Considérese una bolsa con bolas, seis bolas rojas y cuatro bolas negras. El suceso planteado es el de sacar dos bolas de manera consecutiva y que la segunda salga negra.

A partir de aquí, hay dos maneras de continuar con este experimento compuesto por varios consecutivos, extraer la bola *con reemplazamiento* o *sin reemplazamiento*. Si se lleva a cabo con reemplazamiento, la bola extraída la primera vez volverá a la bolsa tras conocer su color para los siguientes experimentos; en el caso sin reemplazamiento, esta bola nunca volverá a entrar. Conocer el método empleado es crucial, ya que cambia por completo el suceso.

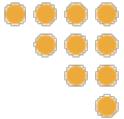
En el primer caso, durante la primera extracción se tiene un espacio muestral formado por diez elementos (bolas), seis rojas y cuatro negras. Tras sacar una, comprobar el resultado y volver a meterla se procede a la segunda extracción, la cual vuelve a tener un espacio muestral de seis bolas rojas y cuatro negras, diez en total. El fenómeno no varía a medida que las distintas extracciones tienen lugar y, por lo tanto, las distintas probabilidades son las mismas.

En el segundo caso, la primera bola extraída se queda fuera y en la segunda extracción se tiene un espacio muestral que depende estrechamente del resultado del primer suceso. Si la bola sacada en la primera extracción fue roja, el espacio muestral en la segunda extracción será de 9 bolas, cinco rojas y cuatro negras; si esta bola sacada fue negra el espacio muestral ahora será también de 9 bolas, pero compuesta por seis rojas y tres negras.

Esto impacta directamente en la probabilidad de obtener una bola determinada, ya que si la probabilidad de obtener, por ejemplo, una bola roja en la primera extracción era de $\frac{6}{10} = 0,6$, en la segunda extracción puede ser, o bien de $\frac{5}{9} = 0,55$ en el caso de que se hubiese obtenido una roja en el primer experimento, o bien de $\frac{6}{9} = 0,67$, si se hubiese sacado una bola negra.

En el primer caso, con reemplazamiento, los distintos experimentos consecutivos no afectaban a su siguientes ya que la bola volvía a entrar en la bolsa, y por lo tanto en el espacio muestral.

En estos casos, se dice que los **sucesos** son **independientes**. En el segundo caso, son **sucesos dependientes**.



3. Sucesos independientes

Esta independencia ayuda a verificar, para estos casos, la igualdad:

$$P(S_1) = P(S_1|S_2)$$

Haciendo uso de la anterior fórmula de condicionalidad, esta vez con la aplicación de esta igualdad anterior, se puede obtener la condición que se debe cumplir entre sucesos cuando son independientes. Y es que, para que se cumpla esto:

$$P(S_1) = P(S_1|S_2) = \frac{P(S_1 \cap S_2)}{P(S_2)}$$

Es decir, para que el cociente entre la probabilidad de la intersección de los sucesos y la del suceso S_2 sea igual a la probabilidad de S_1 , se deberá cumplir:

$$\text{si } P(S_1) = \frac{P(S_1 \cap S_2)}{P(S_2)} \rightarrow P(S_1 \cap S_2) = P(S_1) \cdot P(S_2)$$

En el experimento de ahora se hace uso de un dado y también de una moneda, siempre perfectos, —misma igualdad de oportunidades para cada elemento del espacio muestral—. El suceso se define como el resultado de obtener el número 3 en el dado, sea cual sea el resultado de la moneda.

El espacio muestral está compuesto por las posibles combinaciones de cada uno: {cara, cruz} y {1, 2, 3, 4, 5, 6}; cada número puede salir con cara o cruz, esto suman 12 posibilidades en el espacio muestral.

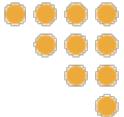
El suceso planteado se compone de los casos siguientes: $S_{c2} \equiv \{\text{cara} + 2\}$ y $S_{x2} \equiv \{\text{cruz} + 2\}$. Por esto, se tiene

$$P(S_2) = P(S_{c2}) \cup P(S_{x2}) = \frac{1}{12} + \frac{1}{12} = \frac{1}{6}$$

Aplicando la fórmula de sucesos condicionados a supuestos sucesos individuales como $S_x = \{\text{cruz}\}$ y $S_2 = \{2\}$, se tiene

$$P(S_2|S_x) = \frac{P(S_2 \cap S_x)}{P(S_x)} = \frac{1/12}{6/12} = \frac{1}{6} = P(S_2)$$

Verificándose así la condición de independencia. Se puede probar para cualquier otra combinación del experimento y comprobar que es siempre independiente.



3. Sucesos independientes

Es importante notar que el experimento general no determina la dependencia o independencia del experimento, sino los sucesos que se presentan. Así, si un experimento fuese el lanzamiento de dos dados y el suceso de interés obtener sólo múltiplos de 3, se tendría el siguiente espacio muestral, los sucesos elementales de interés redondeados.

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

$S_{d1} \equiv$ dado 1 obtiene multiplo de 3

$$P(S) = P(S_{d1} \cap S_{d2}) = \frac{4}{36}, \quad P(S_{d1}) = P(S_{d2}) = \frac{2}{6}$$

Al estudiar la condicionalidad de los sucesos de cada dado

$$P(S_{d1}|S_{d2}) = \frac{P(S_{d1} \cap S_{d2})}{P(S_{d2})} = \frac{4/36}{2/6} = \frac{1}{3} = \frac{2}{6} = P(S_{d1})$$

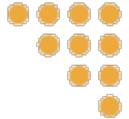
Por lo que son sucesos **independientes** $P(S_{d1} \cap S_{d2}) = P(S_{d1}) * P(S_{d2})$

Es fácil deducir esta condición ya que el suceso no implica ninguna relación entre dados, tan sólo condiciones que no implican mutualidad entre ellos: sacar un número múltiplo de 3 en el dado 1 o no hacerlo no implica que al tirar el segundo dado vayas a tener más probabilidades o menos de obtener ese resultado.

Pero si, ante el experimento de lanzar dos dados al aire, se presenta como suceso la posibilidad de que salgan, en el mismo lanzamiento, un 5 y un 6, la probabilidad del suceso se representa como

$$P(S) = P(S_5 \cap S_6)$$

Continúa ⇒



3. Sucesos independientes

Continuación

El evento de conseguir un 5 y un 6 está formado por dos sucesos fuertemente relacionados ya que, uno de los dados puede sacar un 5 o un 6, lo que le daría dos sucesos elementales dentro de su suceso compuesto. Sin embargo, una vez el dado saque un 5, el otro sólo puede sacar un 6, y viceversa, para tener éxito por lo tanto serán sucesos **dependientes**. En cualquier caso, se estudiará su condicionalidad.

En la tabla de abajo se muestran todos los elementos del espacio muestral del experimento. Además, se encierran en rectángulos rojos todos los sucesos elementales que forman el suceso de que salga un 5 en cualquiera de los dados, S_5 , y rodeado con círculos azules los sucesos elementales que forman el suceso buscado, que salga un 5 y un 6.

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

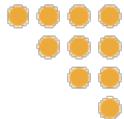
$$P(S) = P(S_5 \cap S_6) = \frac{2}{36}, \quad P(S_5) = P(S_6) = \frac{11}{36}$$

$$P(S_5|S_6) = \frac{P(S_5 \cap S_6)}{P(S_6)} = \frac{2/36}{11/36} = \frac{2}{11} \neq P(S_5)$$

Together for Tomorrow!
Enabling People

Regla de la Multiplicación, Probabilidad Total y Teorema Bayes

#TecnologíaConPropósito



1. Introducción

Las probabilidades condicionales nos ayudan en prácticamente todas las decisiones y razonamientos que realizamos, sin siquiera ser conscientes de ello. ¿Ayer llovió y hoy el cielo sigue cubierto? Me llevo hoy un paraguas. ¿Me han contado que en este restaurante no suelen dar buen servicio y llegó a cerrarse un tiempo por problemas de higiene?, no voy a ir porque es probable que mi experiencia no sea mejor hoy.

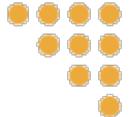
Razonamos continuamente, ya sea deduciendo hechos particulares sobre ciertos fenómenos a partir de conocimiento generalista que tenemos al respecto, o lo contrario, reuniendo información puntual sobre distintos casos para obtener un conocimiento generalista sobre algo relacionado con todos ellos, por *razonamiento inductivo*.

Tanto en una dirección como en otra se hace uso de la información o conocimiento que se tiene sobre un suceso para reducir la incertidumbre sobre sucesos que aún están por ocurrir o de los cuales aún no tenemos información sobre el resultado.

Del estudio de disciplinas como la lógica formal filosófica y el método científico, que tomaban cada vez más importancia en el mundo a medida que el empirismo se formalizaba, grandes nombres de la época moderna como Francis Bacon (1561-1626), filósofo, político, abogado y escritor inglés; John Locke (1632-1704), filósofo y médico inglés; Daniel Bernoulli (1700-1782), matemático, estadístico, físico y médico suizo; o Thomas Bayes (1702-1761), matemático británico, daban forma a lo que se conoce como **estadística inferencial**, llegando a definir teoremas que daban solución a la conocida como probabilidad inversa.

A partir de los principios de la probabilidad condicional, en esta unidad se desarrollan los tres teoremas principales de la teoría de la probabilidad condicional, alguno de los cuales impulsan métodos que se siguen usando para resolver problemas de clasificación, predicción o diagnóstico.





2. Regla de la Multiplicación

En el lanzamiento de dos dados el suceso de sacar un 5 y un 6 está compuesto por la intersección de los sucesos elementales $S_5 = \{\text{sacar un 5 en un dado}\}$ y $S_6 = \{\text{sacar un 6 en un dado}\}$, sucesos dependientes entre sí.

Por esta razón, según la definición de probabilidad condicional se cumple:

$$P(S_6|S_5) = \frac{P(S_5 \cap S_6)}{P(S_5)}.$$

La igualdad anterior puede reescribirse como:

$$P(S_5 \cap S_6) = P(S_5)P(S_6|S_5)$$

Si se descubre ahora un nuevo suceso o evidencia S_x la cual da información al resultado final, es decir, un suceso dependiente de éste podría calcularse la probabilidad como:

$$P(S_6|S_5 \cap S_x) = \frac{P(S_5 \cap S_x \cap S_6)}{P(S_5 \cap S_x)}.$$

A la intersección de las evidencias que aparece ahora en el denominador se le puede aplicar la igualdad extraída más arriba quedando así:

$$P(S_6|S_5 \cap S_x) = \frac{P(S_5 \cap S_x \cap S_6)}{P(S_5)P(S_x|S_5)},$$

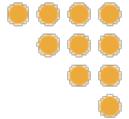
Lo cual se podría reescribir como:

$$P(S_5 \cap S_x \cap S_6) = P(S_5) \cdot P(S_x|S_5) \cdot P(S_6|S_5 \cap S_x)$$

Definición. Dados n sucesos S_1, S_2, \dots, S_n , todos dependientes, se verifica

$$P(\bigcap_{i=1}^n S_i) = P(S_1)P(S_2|S_1)P(S_3|S_1 \cap S_2) \dots P(S_n|S_1 \cap S_2 \cap \dots \cap S_{n-1});$$

es la llamada **Regla de la multiplicación**.



2. Regla de la Multiplicación

El experimento se basa en extraer hasta tres cartas de una baraja española sin reemplazo, es decir, sin devolver la carta extraída a la baraja. El suceso consiste en obtener dos figuras en las primeras dos cartas y cualquier otra que no sea figura en la tercera carta.

La propia condición de no reemplazo en el experimento provoca que se reduzca el espacio muestral en cada extracción sucesiva y por lo tanto se presentan sucesos condicionales. Estos son:

$$S_1 \equiv \text{se extrae figura en 1ª extracción}$$

$$S_2 \equiv \text{se extrae figura en 2ª extracción}$$

$$S_3 \equiv \text{no se extrae figura en 3ª extracción}$$

La probabilidad de que esto ocurra será la intersección de estos sucesos elementales y, aplicando la regla de la multiplicación, se tiene

$$P(S_1 \cap S_2 \cap S_3) = P(S_1)P(S_2|S_1)P(S_3|S_2 \cap S_1)$$

Analizando ahora el problema tenemos distintos sucesos del tipo *figura/nofigura* en una baraja española que consta de 40 cartas totales y 12 figuras. Así, la probabilidad de que, en una primera extracción se consiga figura es de

$$P(S_1) = \frac{\{S_1\}}{E} = \frac{12}{40} = 0,3$$

Suponiendo que el primer suceso hubiese sido exitoso, en el nuevo espacio muestral habría una carta menos, la cual sería figura. Así

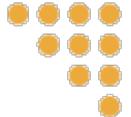
$$P(S_2|S_1) = \frac{\{S_2|S_1\}}{E_2} = \frac{11}{39} = 0,282$$

Finalmente, el tercer suceso conociendo que los dos anteriores se han cumplido, significaría conseguir una carta distinta de las restantes figuras que quedan en la baraja.

$$P(S_3|S_2 \cap S_1) = \frac{\{S_3|S_2 \cap S_1\}}{E_3} = \frac{38 - 10}{38} = 0,737$$

Y ya se puede calcular la probabilidad del suceso pedido

$$P(S_1 \cap S_2 \cap S_3) = P(S_1)P(S_2|S_1)P(S_3|S_2 \cap S_1) = 0,3 \cdot 0,282 \cdot 0,737 = \mathbf{0,062}$$

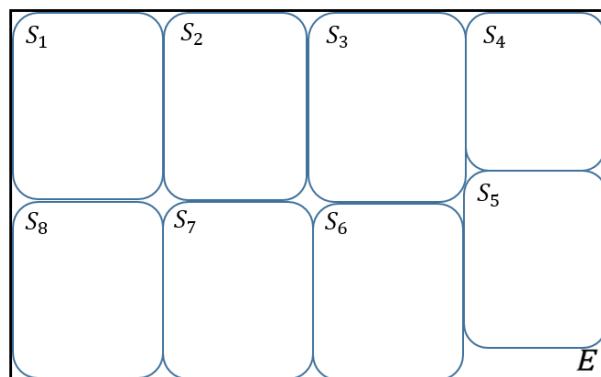


3. Teorema de la Probabilidad Total

En el ejemplo anterior, el experimento estaba formado por tres ensayos que consistían en extraer una carta. Cada suceso correspondía al resultado de cada una de las extracciones. Esto propiciaba que los sucesos no fuesen excluyentes entre sí, es decir, existía la posibilidad de que se cumpliesen todos esos sucesos en el experimento. Esto puede simbolizarse como

$$S_1 \cap S_2 \cap S_3 \neq \emptyset$$

Existen otros casos en los que esta desigualdad no se da debido a que el mismo hecho de cumplirse un suceso S_1 hace imposible que se pueda cumplir el resto de sucesos, normalmente porque forman parte del espacio muestral del mismo experimento. Sobre los sucesos que cumplen esta disyunción suele decirse que son **disjuntos dos a dos**.

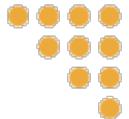


Una extracción de carta en la que el suceso fuese $S = \{\text{sacar carta } x\}$ sería un claro ejemplo de lo mencionado arriba; cada suceso sería una carta, todas juntas ocuparían el espacio muestral y jamás sería posible sacar una carta a la vez que sacas otra en la misma extracción. Bajo este tipo de casos se desarrolla la siguiente definición

Definición. Dados un suceso A y n sucesos S_1, S_2, \dots, S_n , disjuntos dos a dos, tales que $S_i \cap S_j = \emptyset$, $\cup_{i=1}^n S_i = E$ y $(A \cap S_i) \neq \emptyset$, se cumple que

$$P(A) = \sum_{i=1}^n P(A|S_i)P(S_i)$$

y define el llamado **Teorema de la Probabilidad Total**.



3. Teorema de la Probabilidad Total

Ejemplo. Existe un gen hereditario particular en los gatos que determina un pelaje muy característico. Los que poseen este gen nacen con un pelaje tricolor.

Está presente en el cromosoma X (femenino); sin embargo, en raras ocasiones se puede encontrar en el cromosoma Y (masculino).

Recogiendo datos de distintas fuentes fiables, se sabe que el 0,5879% de todos los gatos poseen esta característica y que las hembras tienen una probabilidad del 1,15 % de ser tricolor. Además, en el reino animal de los gatos, el 51,08 % son hembras. ¿Sería posible a partir de aquí conocer la probabilidad de que un gato macho sea tricolor?

Analizando lo enunciado se identifican los sucesos de acertar el sexo del gato como dos sucesos disjuntos dos a dos, ya que la probabilidad de nacer macho es complementaria a nacer hembra. Así, usando $m \equiv \text{macho}$ y $h \equiv \text{hembra}$

$$P(h) = 0,5108, \quad P(m) = 1 - P(h) = 0,4892$$

Por otro lado, el suceso de nacer con pelaje tricolor puede intersectar con cualquier de los anteriores sucesos ($(A \cap S_i) \neq \emptyset$), por lo que podemos hacer, usando $tr \equiv \text{tricolor}$,

$$P(tr) = 0,005879, \quad P(tr|h) = 0,0115$$

Lo que se pretende encontrar en este problema es la probabilidad de que, teniendo como evidencia el sexo masculino del gato, éste sea tricolor

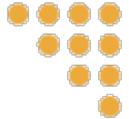
$$P(tr|m) = x$$

$$P(tr) = \sum_{i=1}^n P(tr|S_i)P(S_i) = P(tr|m)P(m) + P(tri|h)P(h)$$

$$P(tr|m) = \frac{P(tr) - P(tri|h)P(h)}{P(m)} = \frac{0,005879 - 0,0115 \cdot 0,5108}{0,4892} = 9,812 \cdot 10^{-6}$$

Es decir, aproximadamente 9,8 ppm* de los gatos machos son tricolor.

* ppm \equiv partes por millón



4. Teorema de Bayes

Uno de los más importantes métodos de cálculo de probabilidades es el desarrollado a partir de la inducción o razonamiento inductivo.

En este, al contrario que en razonamiento tradicional deductivo en el que se infiere conclusión particular a partir de premisas generales, se consiguen inferir en conclusiones generales a partir de premisas particulares.

En el ejemplo anterior se sabía la probabilidad de gatos machos y de hembras, también la probabilidad de que, si una gata es hembra sea también tricolor y, si se hubiese conocido también la probabilidad de ser tricolor siendo gato, se hubiera podido hacer inferencia sobre la probabilidad de ser tricolor. Es decir,

X gatos son machos + Y gatos machos son tricolor $\rightarrow Z$ gatos o gatas son tricolor

Se puede *deducir* la probabilidad de que un gato cualquiera sea tricolor conociendo datos generales sobre cuantos machos hay y cuantos machos y hembras suelen ser tricolor.

¿Y si se necesitase conocer la probabilidad de que un gato siendo tricolor sea también macho? Habría que *inducir* el resultado.

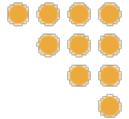
Para estos tipos de cálculos se tiene el **Teorema de Bayes** que vincula la probabilidad de un suceso A general y perteneciente a espacio muestral disjunto, como ser macho o hembra, dado otro suceso B no disjunto y que intersecta con todos los sucesos del primer espacio muestral, como ser tricolor.

Definición. Dados n sucesos S_1, S_2, \dots, S_n , disjuntos dos a dos, tales que $S_i \cap S_j = \emptyset$, $\bigcup_{i=1}^n S_i = E$ y $(A \cap S_i) \neq \emptyset$, y un suceso A del que se conocen las probabilidades condicionales o evidencias $P(A|S_i)$, se cumple que

$$P(S_i|A) = \frac{P(A|S_i)P(S_i)}{P(A)}$$

y define el llamado **Teorema de la Probabilidad Total**. Aplicando ahora la regla de la probabilidad total al denominador, se obtiene la **Regla de Bayes**.

$$P(S_i|A) = \frac{P(A|S_i)P(S_i)}{\sum_{i=1}^n P(A|S_i)P(S_i)}$$



4. Teorema de Bayes

Ejemplo. Un sistema automático de extracción de datos obtiene información de una web de previsiones meteorológicas. Sobre un día en particular se tienen tres posibilidades con respecto a las precipitaciones. Existe una probabilidad de lluvia del 30 %, una probabilidad de nieve del 25 % y, finalmente, un 45 % de probabilidades de que no haya precipitaciones. De otra fuente se tiene que la probabilidad habitual del 1,9 % de que ocurra un accidente de tráfico se incrementa en +0,4 % cuando llueve y en un +1,2 % en casos de nieve.

Se investiga un accidente concreto ocurrido en esa fecha, de la que no se tiene información sobre el tiempo que realmente hizo, si llovió, nevó, o ninguna de las dos. La información sobre las probabilidades de la previsión meteorológica, en cuanto a las precipitaciones, son llamadas *probabilidades a priori*. Se pretende inferir en la probabilidad de que, dado que tuvo el accidente, llovía nevaba o ninguna de las dos, las llamadas *probabilidades a posteriori*.

De la información que tenemos, se reconoce fácilmente un conjunto de sucesos disjuntos que completan el espacio muestral de posibilidades, la previsión. Así, si se usa $acc \equiv \text{accidente}$, $ll \equiv \text{lluvia}$, $snw \equiv \text{nieve}$ y $np \equiv \text{no precipitación}$, se tiene

$$P(ll) = 0,3, \quad P(snw) = 0,25, \quad P(np) = 0,45$$

También se conocen las probabilidades condicionales de tener un accidente bajo cada una de las hipótesis anteriores

$$\begin{aligned} P(acc|np) &= 0,019, & P(acc|ll) &= 0,019 + 0,004 = 0,0194, \\ P(acc|sn) &= 0,019 + 0,012 = 0,031 \end{aligned}$$

De toda esta información, se puede saber la probabilidad de que aquel día no estuviese lloviendo ni nevando, $P(np|acc)$, por ejemplo.

$$\begin{aligned} P(np|acc) &= \frac{P(acc|np)P(np)}{P(acc|np)P(np) + P(acc|ll)P(ll) + P(acc|snw)P(snw)} \\ &= \frac{0,019 \cdot 0,45}{0,019 \cdot 0,45 + 0,0194 \cdot 0,3 + 0,031 \cdot 0,25} = 0,3865 \end{aligned}$$

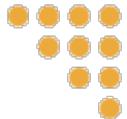
Se deduce de aquí que lo más probable es que aquel día no hubiese precipitaciones, con un 38,65 % de probabilidades.

* ppm \equiv partes por millón

Together for Tomorrow!
Enabling People

Técnicas de conteo

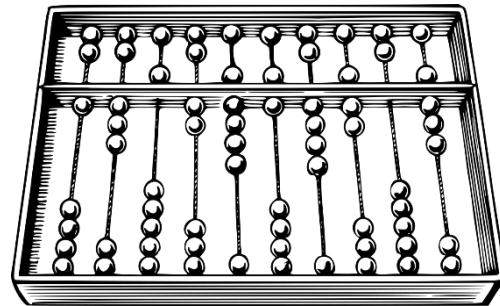
#TecnologíaConPropósito



1. Introducción

La disciplina que estudia la combinatoria se ha desarrollado en paralelo al de otras ramas de las Matemáticas, como el álgebra, teoría de números, y probabilidad.

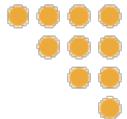
Algunos de los problemas de combinatoria a las que se han enfrentado los matemáticos desde tiempo inmemorial se encuentran algunos como el problema de los **cuadrados mágicos** aparece en un libro chino fechado 2200 a. C., los **coeficientes binominales** conocidos en el siglo XII o el **triángulo de Pascal** desarrollado en el siglo XIII.



En Europa se considera que la combinatoria surge en el siglo XVII con los trabajos de Blaise Pascal (1623-1662), matemático, físico y filósofo francés, y de Pierre de Fermat (1607-1665), jurista y matemático francés, sobre la **teoría de juegos de azar**. Estos trabajos, que formaron los fundamentos de la teoría de la probabilidad, contenían asimismo los principios para determinar el número de combinaciones de elementos de un conjunto finito, y así se estableció la tradicional conexión entre combinatoria y probabilidad.

El término combinatoria como lo usamos actualmente fue introducido por Wilhelm Leibniz (1646-1716), filósofo, matemático, teólogo y jurista alemán en su primer libro, y también tesis de habilitación, *Disertación acerca del arte combinatorio* (1666).

Si para conocer la probabilidad de que un suceso tenga éxito en un experimento aleatorio es necesario tener bien definido el espacio muestral con todas las posibles alternativas, conocer cómo de grande será este espacio muestral ante cualquier tipo de experimento debe ser igual de importante y las técnicas de conteo y combinatoria desarrolladas en esta unidad, proveerá al lector de esta habilidad.



2. Fundamentos del conteo

El conteo o combinatoria es la rama de las matemáticas discretas cuyo estudio permite contar elementos, en la mayoría de casos de interés serán opciones, alternativas o estados posibles, sin enumerar directamente todos los casos posibles que pueden darse en un suceso determinado; proporciona las herramientas y métodos de conteo necesarios para calcular de cuántas maneras distintas puede darse un suceso, es decir, para calcular el tamaño del espacio muestral de un experimento o fenómeno dado.

Si el experimento presentado es el lanzamiento de un dado es fácil calcular que hay seis posibles sucesos; si se lanzan dos dados, también se deduce rápidamente que serán 12 los posibles resultados.

Pero ¿y si se pretende calcular la cantidad de números de 4 cifras que es posible formar con las posibles combinaciones de números que salgan de esos dos dados? ¿O cuantos números de cuatro cifras se pueden formar con todos los valores de un dado?

Estas respuestas ya no son tan inmediatas y requieren de estas técnicas mencionadas para poder conocerlas. De hecho, estos dos ejemplos representan casos distintos y, por lo tanto, será necesario hacer uso de distintas técnicas según la naturaleza del problema.

Antes de entrar a clasificar cada estas técnicas, es necesario conocer algunos cálculos y operadores particulares, así como principios básicos del conteo sobre los que se apoyarán las técnicas de la combinatoria.

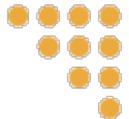


2.1. Terminología

Es necesario conocer las definiciones de las expresiones más comunes del conteo:

- **Conjunto base.** Aquel que contiene todos los elementos de los que se formarán los diferentes subconjuntos a partir de la técnica combinatoria correspondiente.

Si, por ejemplo, se parte de cualquier problema combinatorio con todos los posibles valores de un dado, el conjunto base estaría formado por los números del 1 al 6.



2. Fundamentos del conteo

- **Base o cardinal.** Es el número que indica la cantidad de elementos que existen en el conjunto base. Siguiendo con el dado la base sería seis.

Sin embargo, si el problema combinatorio fuese conocer todos los posibles números que puedan formarse con los obtenidos en el lanzamiento de dos dados, la base sería 2. Dado un conjunto A , se representa como $|A|$.

- **Orden.** Es el número de elementos que contienen los subconjuntos que forman el conjunto base. Esto es, si el ejemplo anterior fuese conocer los posibles números a formar con los obtenidos en el lanzamiento de dos dados sin que estos puedan repetirse, la base y el orden coincidirían.

Sin embargo, si se buscasen todos los posibles números de cinco cifras que se puedan formar con esos dos números, en ese caso, la base seguiría siendo 2, indicando los números con los que combinar, pero el orden sería cinco.

2.2. Herramientas del conteo

Existen dos operadores matemáticos muy importantes en el análisis combinatorio debido a su amplio uso, por lo que se deben conocer y saber usar. Estos son el **factorial** y los **coeficientes binomiales**.

2.2.1. Factorial

Para resolver el problema de análisis combinatorio habrá que conocer un operador llamado **factorial**.

El factorial de un número determinado n es el producto de toda la secuencia desde n hasta 1 y usa el símbolo ! como operador. Es decir,

$$n! = n * (n - 1) * (n - 2) * \dots * 2 * 1$$

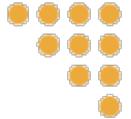
Por ejemplo,

- $3! = 3 * 2 * 1 = 6$
- $9! = 9 * 8 * 7 * 6 * 5 * 4 * 3 * 2 * 1 = 362880$

Propiedades

Como excepción es importante conocer el caso

- $0! = 1$



2. Fundamentos del conteo

Además, cuando se trabaja con factoriales, no es posible aplicar directamente sumas, rectas, multiplicaciones o simplificaciones de fracciones. Es decir

- $5! + 3! \neq 8!$
- $4! * 3! \neq 12!$
- $5! - 3! \neq 2!$
- $\frac{5!}{3!} \neq \frac{10!}{6!}$

Sí que se sigue cumpliendo la igualdad

$$\frac{5!}{5!} = 1$$

y por esto será muy útil simplificar de la siguiente manera cuando se tengan casos del tipo $\frac{n!}{n-x!}$

$$\frac{n!}{m!} = \prod_{i=m+1}^n i, \quad \forall m < n$$

Efectivamente, esto puede demostrarse con el siguiente ejemplo en el que $n = 19$ y $m = 17$.

$$\frac{19!}{17!} = \frac{19 * 18 * 17 * 16 * \dots * 2 * 1}{17!} = \frac{19 * 18 * 17!}{17!} = 19 * 18$$

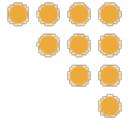
2.2.2. Coeficientes binomiales

Una vez conocido el factorial, se presenta ahora una herramienta de representación de factoriales llamada coeficiente binomial o números combinatorios, en general. Concretamente, estos coeficientes cumplen

$${n \choose r} = \frac{n!}{r!(n-r)!}, \quad \forall r < n$$

Ejemplos:

- ${10 \choose 7} = \frac{10!}{7!(10-7)!} = \frac{10*9*8}{3!} = 120$
- ${6 \choose 4} = \frac{6!}{4!(6-4)!} = \frac{6*5}{2!} = 15$



2. Fundamentos del conteo

Propiedades

- $\binom{n}{r} = \binom{n}{n-r}$, $\binom{6}{4} = \binom{6}{2} = 15$
- $\binom{n}{0} = \binom{n}{n} = 1$
- $\binom{n}{r} + \binom{n}{r+1} = \binom{n+1}{r+1}$, $\binom{6}{4} + \binom{6}{5} = 15 + \frac{6!}{5!*1!} = 21 = \frac{7!}{5!*2!} = \binom{7}{5}$

Triángulo de Pascal

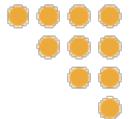
El triángulo de Pascal es otro tipo de representación de coeficientes binomiales en la que números con una determinada relación forman un triángulo el cual, posee por cada fila un número más que en la fila superior. Comenzando, de manera irregular en la punta superior del triángulo siempre con dos unos, sigue creciendo mediante la colocación de unos en sus extremos y la suma de los dos números que quedan justo por encima de cada posición, formando el triángulo de la imagen.

1	1													
1	2	1												
1	3	3	1											
1	4	6	4	1										
1	5	10	10	5	1									
1	6	15	20	15	6	1								
1	7	21	35	35	21	7	1							
1	8	28	56	70	56	28	8	1						
1	9	36	84	126	126	84	36	9	1					
1	10	45	120	210	252	210	120	45	10	1				
1	11	55	165	330	462	462	330	165	55	11	1			
1	12	66	220	495	792	924	792	495	220	66	12	1		
1	13	78	286	715	1287	1716	1716	1287	715	286	78	13	1	
1	14	91	364	1001	2002	3003	3432	3003	2002	1001	364	91	14	1

Siguiendo esta regla, la cuarta fila queda formada por 1,4,6,4,1: unos en los extremos y la suma de los dos números que quedan por encima en las posiciones interiores. Arriba de cada 4 hay un 1 y un 3, sobre el 6 hay dos 3.

Pero la relación que hace verdaderamente interesante este triángulo es que la n -ésima fila del triángulo, comenzando desde arriba, está formada por los resultados de $\binom{n}{0}, \binom{n}{1}, \dots, \binom{n}{n}$. Efectivamente,

$$\langle 1, 4, 6, 4, 1 \rangle = \langle \binom{4}{0}, \binom{4}{1}, \binom{4}{2}, \binom{4}{3}, \binom{4}{4} \rangle = \langle 1, \frac{4!}{1! * 3!}, \frac{4!}{2! * 2!}, \frac{4!}{1! * 3!}, 1 \rangle$$



2. Fundamentos del conteo

2.3. Principios del conteo

Es importante no olvidar durante el estudio de esta unidad, que se está tratando con conjuntos.

Concretamente, se pretende, mediante conteo, conocer el tamaño de un conjunto dado, el número de subconjuntos que lo componen; o, dicho de otro modo, el número de sucesos que existen dentro de un espacio muestral.

Como tal, se pueden seguir ciertos principios de la teoría de conjuntos:

- **Principio de adición.** Este principio dice que si A y B son dos conjuntos disjuntos, entonces la unión de sus conjuntos es igual a la suma de estos.

$$\text{Si } A \cap B = \emptyset \rightarrow |A \cup B| = |A| + |B|$$

Extrapolado al tema de la combinatoria, el principio dice que si el suceso A puede ocurrir de m distintas maneras y B puede ocurrir de n distintas formas, las posibles opciones que tiene el suceso A ó B es $m + n$.

Se quiere llegar a un destino concreto y se plantean dos alternativas iniciales: transporte público o privado.

Además, si se elige transporte público, es posible llegar tanto en autobús, como en metro o taxi; en transporte privado se puede ir en moto, o en coche. ¿Cuántas maneras existen de llegar a este destino?

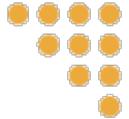
Aquí, un subconjunto sería el transporte público y, el otro, transporte privado, una con tres opciones y el otro con dos. Por lo tanto, el número posible de maneras de llegar son cinco.

- **Principio de multiplicación.** Este caso se aplica a dos sucesos A y B no disjuntos y subconjuntos del suceso compuesto que lo forma, $A \cap B$.

En este caso, esta intersección de sucesos puede desembocar en $|A| * |B|$ distintas posibilidades.

$$\text{Si } A \cap B \neq \emptyset \rightarrow |A \cap B| = |A| * |B|$$

Estos casos pueden verse como consecutivos, que ocurren uno a continuación del otro.



2. Fundamentos del conteo

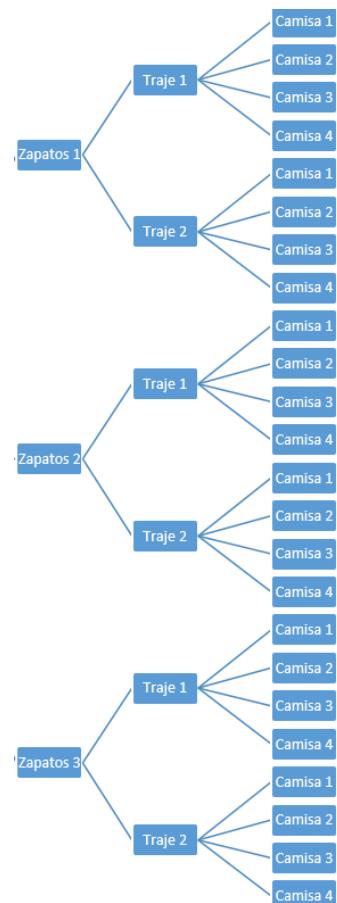
Se plantean todas las posibilidades de vestuario para una reunión importante.

Sin importar lo bien o mal que combinen estilísticamente hablando, se pretende saber cuantas posibilidades hay de ir vestido ante la combinación de 3 pares de zapatos distintos, 2 trajes de chaqueta y 4 camisas.

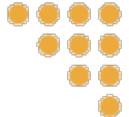
Un par de zapatos cualquiera podría combinarse con ambos trajes de chaqueta lo que da 2 posibilidades.

Esto, para cada uno de los 3 pares de zapatos son $3 * 2 = 6$ posibilidades entre zapatos y traje que podrían, a su vez, combinarse con cada una de las 4 camisas distintas.

Resumiendo, se tendrían los sucesos no disjuntos $A \equiv \text{zapatos}, B \equiv \text{trajes}, C \equiv \text{camisas}$ y



$$|A * B * C * D| = 3 \cdot 2 \cdot 4 = 24 \text{ conjuntos posibles}$$



3. Variaciones

Una de las situaciones anteriormente comentadas consistía en el cálculo de la cantidad de números posibles que se podrían formar con los obtenidos tras el lanzamiento de dos dados. Es determinante recalcar que los números tendrían, en este caso, las mismas cifras que el conjunto base proporcionado por los dos dados; es decir, la base del problema sería dos. Si saliesen el número 5 y el 1 se tendrían las posibilidades 11, 15, 51 y 55. Existirían cuatro posibilidades.

3.1. Variaciones sin repetición

El mismo tipo de problema sería si se usasen 6 dados y tuviese que conocerse el número de números de cuatro cifras posibles a formar con esos 6 números. Cada número se podría usar solamente una vez, es decir, la *repetición* no estaría permitida.

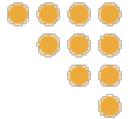
Se denomina **variación ordinaria o sin repetición** de n elementos tomados de k en k al número total de grupos que se pueden formar escogiendo k elementos distintos de los n elementos del conjunto base de tal manera que cada grupo pueda ser diferenciado no sólo por los elementos que lo forman, sino también por el orden en el que se disponen.

$$V_{n,k} = \frac{n!}{(n - k)!}$$

Se tiene una baraja de cartas española y se van a extraer, una a una, tres cartas. Si las extracciones tuviesen lugar con reemplazo, es decir, si la carta extraída fuese devuelta a la baraja antes de la siguiente extracción, se tendrían 40 cartas candidatas cada vez. En este problema, por la regla de la multiplicación, se sabría que existen $40 * 40 * 40 = 64000$ combinaciones posibles de esas tres cartas. Al no haber reemplazo en esta ocasión, las posibilidades se reducirían a una menos en cada extracción, siendo $40 * 39 * 38 = 59280$.

Este cálculo se generaliza como

$$40! \quad 40 * 39 * 38 * 37!$$



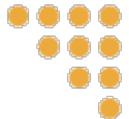
3. Variaciones

3.2. Variaciones con repetición

Se ha visto antes el caso de la baraja española con reemplazo, en la que sacando tres cartas se tendrían $40 * 40 * 40 = 40^3 = 64000$ posibles combinaciones de trios de cartas.

Se denomina **variación con repetición** de n elementos tomados de k en k al número total de grupos que se pueden formar escogiendo k elementos iguales o distintos de los n elementos del conjunto base de tal manera que cada grupo pueda ser diferenciado no sólo por los elementos que lo forman, sino también por el orden en el que se disponen.

$$VR_{n,k} = n^k$$



4. Permutaciones

La propia definición de permutación enuncia *variar la disposición u orden en que estaban dos o más cosas*. Las permutaciones son un caso particular de las variaciones, ya que son los grupos que se pueden formar usando todos los elementos n del conjunto base a la vez, simplemente cambiándolos de orden.

4.1. Permutaciones sin repetición

Se denomina **permutación sin repetición** de n elementos al número total de ordenaciones posibles que se pueden formar con estos elementos sin que falte ninguno.

Estos casos se calculan como

$$P_n = n!$$

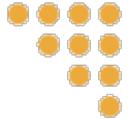
Imagínese que se tuviese de nuevo la baraja de cartas española pero no se realizase ninguna extracción. Esta vez el interés reside en conocer la probabilidad que existe de que, después de barajar las cartas, estas apareciesen perfectamente ordenadas.

Para esto, se debería conocer, en primer lugar, el tamaño del espacio muestral; es decir todas las posibles soluciones que podría encontrarse tras barajar las cartas. Dicho de otro modo, todas las ordenaciones posibles con las 40 cartas.

Comenzando desde la primera a la última: se tendría 40 posibles cartas para salir en primer lugar; una vez saliese elegida la primera, la segunda posición tendría 39 posibilidades ya que una carta estaría escogida para la primera posición. Así, sucesivamente se tendrían 40, 39, 38, 37 ... hasta llegar a la última posición, en la que no cabría duda de que la ocuparía la carta restante.

Por la regla de la multiplicación, se sabe que el número de posibilidades sería igual al producto de toda esta secuencia y, por lo tanto,

$$P_{40} = 40! \cong 8,16 * 10^{47}$$



4. Permutaciones

4.2. Permutaciones con repetición

Se llama **permutación con repetición** de r elementos distintos cada uno repetido k_i veces , de tal manera que formen $\sum_{i=1}^r k_i = n$ elementos en total, al número total de ordenaciones posibles que pueden formarse con estos, de tal manera que cada una de ellas se diferencie del resto en la colocación de sus elementos, excluyendo las reordenaciones entre elementos repetidos.

Estos casos se calculan como

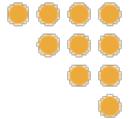
$$PR_n^{k_1, k_2, \dots, k_r} = \frac{n!}{k_1! \cdot k_2! \cdot \dots \cdot k_r!}$$

Las fichas de un ajedrez están formadas por 16 fichas: ocho peones, dos caballos, dos torres, dos alfiles, un rey y una reina. ¿Cuántas configuraciones posibles existirían de colocación de estas fichas como posición inicial para comenzar una partida?

Se distingue que se está ante un problema en el que se busca las agrupaciones con todos los elementos totales, $n = 16$ fichas, los cuales están agrupados en $r = 6$ distintos subconjuntos de elementos indistinguibles, formando $k_1 = 8$ peones, $k_2 = 2$ torres, $k_3 = 2$ alfiles, $k_4 = 2$ caballos, $k_5 = 1$ rey y $k_6 = 1$ reina.

El número total de configuraciones posibles sería:

$$PR_{18}^{8,2,2,2,1,1} = \frac{16!}{8! \cdot 2! \cdot 2! \cdot 2! \cdot 1! \cdot 1!} = 64.864.800 \text{ posibilidades}$$



5. Combinaciones

Hasta ahora, se han visto diferentes escenarios, con sus respectivas técnicas de resolución, pero siempre dando importancia al orden de los elementos. Cuando esto no influye en el cálculo, es decir, cuando distintas configuraciones de los mismos números no sean consideradas como distintas, el problema deberá tratarse como una **combinación**.

5.1. Combinaciones sin repetición

Se llaman **combinaciones ordinarias o sin repetición** de n elementos, tomados de k en k , a los diferentes conjuntos de k elementos distintos, esto es, que un conjunto se diferencie de los demás en, al menos, un elemento (no importa el orden de colocación o selección). El número de elementos se calcula como:

$$C_{n,k} = \frac{V_{n,k}}{P_k} = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

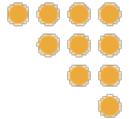
Es decir, si se quiere calcular el número de grupos posibles de k elementos escogidos de entre n elementos sin que importe el orden y sin que pueda repetirse, éste sería igual al número de grupos en caso de que sí importase el orden, es decir, la variación, entre el número de grupos en caso de que se escogiese el total de n elementos, es decir, la permutación.

En una empresa se necesitan, de manera repentina, 3 compañeros expertos en una determinada materia, para dar soporte en un proyecto que lleva cierto retraso. Se llegan a identificar hasta 7 compañeros considerados como expertos en la materia.

Esto es un caso de combinación sin repetición, obviamente porque los 3 compañeros que se elijan no pueden ser la misma persona y porque, al no haber ninguna distinción entre lo que hará cada colaborador, no importa el orden en el que se elijan.

En este caso, el número de posibles tríos de expertos sería de

$$C_{7,3} = \binom{7}{3} = \frac{7!}{3!(7-3)!} = 35 \text{ posibles tríos}$$



5. Combinaciones

5.2. Combinaciones con repetición

En el caso contrario en el que se permita la repetición de los elementos elegidos del conjunto dado pero el orden de colocación de estos siga sin ser importante, se está ante **combinaciones con repetición**.

Se llaman **combinaciones con repetición** de n elementos, tomados de k en k , a los diferentes conjuntos de k elementos iguales o distintos., esto es, que un conjunto se diferencie de los demás en, al menos, un elemento (no importa el orden de colocación o selección). El número de elementos, en este caso, se calcula como:

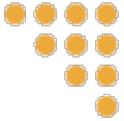
$$CR_{n,k} = \binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!}$$

Se va a comprar para una celebración 5 botellas de alcohol. Pueden ser de alcohol destilado, entre las que están disponibles 2 tipos de wiski, 3 tipos de ginebra y 1 de ron, o vino que puede ser blanco, tinto o rosado. ¿Cuántas posibles combinaciones podrían darse?

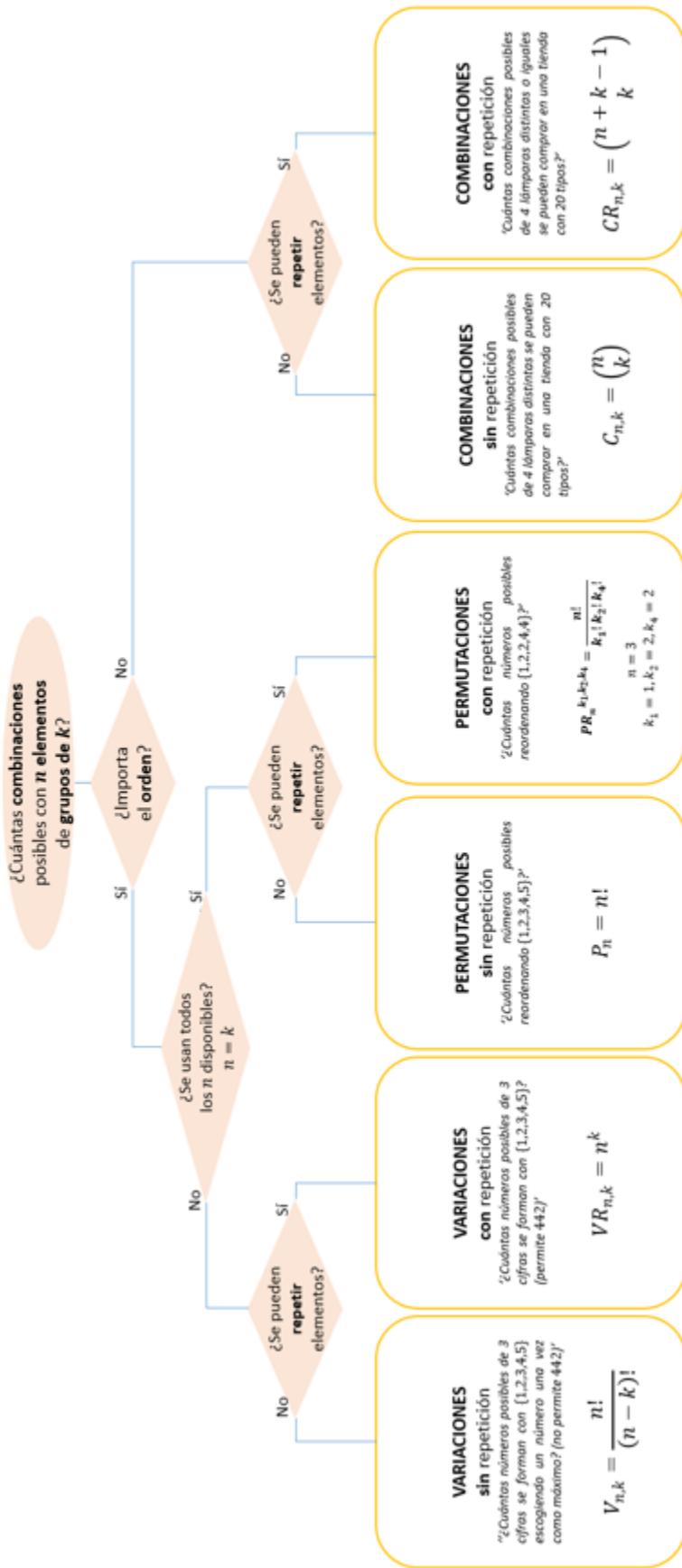
Sumando, se tienen $n = 2wiski + 3gin + 1ron + 3vinos(1blanco + 1tinto + 1rosado) = 9$ tipos, de entre los que se deben coger $k = 5$ botellas, sin importar que aparezcan tipos repetidos (incluso que todos fuesen, por ejemplo vinos rosados).

Las posibles combinaciones, en este caso, sumarían

$$CR_{7,3} = \binom{9+5-1}{5} = \binom{13}{5} = \frac{13!}{5!(13-5)!} = 1\,287 \text{ posibilidades}$$



6. Resumen



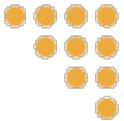
SAMSUNG

BeJob[™]

Together for Tomorrow!
Enabling People

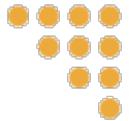
Estadística

#TecnologíaConPropósito



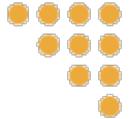
Índice

Conceptos básicos	7
1. Introducción	9
2. Ciencia estadística	10
2.1. Tipos	10
3. Definiciones	11
3.1. Parámetros estadísticos básicos	12
4. Análisis estadístico	17
Sistemas continuos y discretos	19
1. Introducción	21
2. Tipos de datos	22
2.1. Variable discreta	22
2.2. Variable continua	22
3. Representación de los datos	24
3.1. Diagrama de barras	24
3.2. Histogramas	24
3.3. Diagrama de dispersión	25
3.3. Diagrama de Box-Whisker	26
Distribución de frecuencias y <i>Probability Distribution Function (PDF)</i>	31
1. Introducción	33
2. Distribución de frecuencias	34
2.1. Tabla de frecuencias	35
2.2. Función de distribución	37



Índice

2.3. Variables aleatorias discretas	39
2.4. Variables aleatorias continuas	40
3. Estadísticos de la variable aleatoria	44
3.1. Esperanza matemática	44
3.2. Momentos	45
3.3. Estadísticos de dispersión	46
3.4. Estadísticos de forma	47
Tipos de funciones de probabilidad	51
1. Introducción	53
2. Distribuciones discretas	54
2.1. Distribución degenerada	54
2.2. Distribución uniforme discreta	55
2.3. Distribución de Bernoulli o binomial ($1, p$)	56
2.4. Distribución binomial (n, p)	57
2.5. Distribución de Poisson	59
3. Distribuciones continuas	60
3.1. Distribución uniforme	60
3.2. Distribución normal	62
Inferencia estadística	67
1. Introducción	69
2. Tipos de inferencia	70
2.1. La estimación	70
2.2. El contraste de hipótesis	71



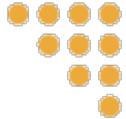
Índice

3.	<u>Muestreo</u>	72
3.1.	<u>Técnicas de muestreo</u>	72
3.2.	<u>Distribuciones muestrales</u>	73
4.	<u>Estimaciones</u>	74
4.1.	<u>Estimación puntual</u>	74
4.2.	<u>Estimación por intervalos de confianza</u>	75
5.	<u>Contraste de hipótesis</u>	81
5.1.	<u>Conceptos básicos</u>	81
5.2.	<u>Tipos de contrastes</u>	82

Together for Tomorrow!
Enabling People

Conceptos básicos

#TecnologíaConPropósito

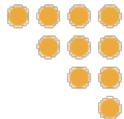


1. Introducción

La estadística comenzó a manifestarse en la civilización como una mera recolección de datos, normalmente de la población. En los primeros estados como Babilonia o el Antiguo Egipto ya se recogían datos sobre sus habitantes para recaudación de impuestos o reclutamiento de guerreros para sus ejércitos. Esta recolección de datos de la población se convierte en sistemática en la Europa del siglo XVII, profundizando en la demografía y, a su vez, focalizando esta información para fines socio-económicos; por el momento, se llevaba a cabo trabajos de recolección de información para obtener conclusiones sobre la misma; no es hasta el siglo XIX que la Estadística da un salto evolutivo para comenzar a inferir esta información a partir de muestras y ser capaz así de prever el resto de información no recogida. Además, se usa estas muestras para realizar predicciones en el tiempo o tomar decisiones en situaciones inciertas. En esta época surgen grandes contribuidores como Carl Friedrich Gauss (1777-1855), matemático, astrónomo y físico alemán; Adolphe Quetelet (1796-1874), astrónomo, naturalista, matemático, sociólogo y estadístico belga; o Karl Pearson (1857-1936), matemático, estadístico, historiador, filósofo, biógrafo y psicólogo inglés, los cuales crean nuevas maneras de representar la información o métodos para conocer la correlación entre distintos tipos de datos y ser capaz de prever información desconocida a partir de los datos disponibles.

Todos estos avances hacen que en el siglo XX la Estadística se introduzca en prácticamente todos los campos científicos y que con mucha probabilidad sea una de las ramas científicas más estudiadas y empleadas.





2. Ciencia estadística

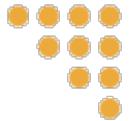
La **estadística** es la rama de la matemática que se encarga de analizar y estudiar datos, así como también buscar las explicaciones de los fenómenos que alteran los resultados.

Es una de las ciencias más extendidas en el mundo debido a que su estudio, uso y aplicación es fundamental para la toma de decisiones de diferentes ámbitos, por lo que es posible sacar rédito de la estadística en todos los campos profesionales; es un proceso que estudia problemas sociales, científicos e industriales. No es una ciencia exacta, por lo que proporciona un resultado estimado, ya sea numérico o categórico, a la vez que ofrece conclusiones que conducen a una solución.

2.1 Tipos

Existen cuatro ramas en la Estadística:

- **Descriptiva o deductiva.** Aquella estadística que se encarga de mostrar el resultado de los datos estudiados de forma específica, sin generalizaciones.
- **Inferencial o inductiva.** Aquella estadística que a diferencia de la descriptiva ofrece resultados junto con datos generales de investigación amplia.
- **Aplicada.** Esta estadística se utiliza después de investigar, estudiar y analizar con los métodos anteriores para proporcionar resultados específicos y generalizados sobre la investigación.
- **Matemática.** Además de realizar los procesos de estadística descriptiva o inferencial, la estadística matemática utilizará el álgebra y ciertos análisis más profundos para ofrecer un punto de vista enfocado y formal.



3. Definiciones

Al introducirse en el mundo estadístico es muy frecuente encontrarse términos usados en la materia y que se dan por sabidos, pero de los cuales quizás no se conozca su significado.

Esto puede convertir el estudio de una disciplina tan interesante como la estadística en intento frustrado y tedioso. Por esto es de vital importancia entender los siguientes conceptos y términos con total claridad antes de continuar con el bloque.

La **estadística** se encarga de estudiar datos, y para cada estudio estos datos pueden definirse por su propia naturaleza u origen. Supóngase que quiera realizar un estudio estadístico de las partidas de tornillos que llegan defectuosos a una fábrica; los elementos a analizar serán los tornillos y se podrá decir entonces que la **población** está formada por tornillos.

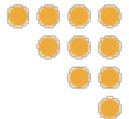
Además, será necesario tener un correcto etiquetado de estos elementos, de tal manera que sea posible asociar ciertos estadísticos con subconjuntos dentro de la población. Por este motivo, son considerados **individuos** los distintos elementos que forman una **población**. En el ejemplo anterior, los individuos serían el tornillo 1, el tornillo 2, el 3, y así hasta formar la población completa.

Habrá casos en los que la población sea de un tamaño muy grande como para estudiar a todos sus individuos uno a uno por lo que, con cierta lógica y método, se deberá escoger un subconjunto de la población de tal manera que éste sea suficientemente representativo con respecto a la población y así no se pierdan características de la población por no haber escogido el subconjunto adecuado. Este subconjunto es llamado **muestra** y el proceso de selección de una muestra de calidad suficientemente representativa, **muestreo**.

El propio objetivo de la estadística convierte el muestreo en uno de los pasos más importantes en un análisis estadístico.

Interesa obtener un subgrupo lo más parecido posible a la población en todas sus características excepto en el número de elementos. Una muestra de poca calidad alterará negativamente los estadísticos y proporcionará *información sesgada* sobre la población.

Es por esto por lo que las **técnicas de muestreo** son herramientas que reciben una extraordinaria atención y estudio.



3. Definiciones

Para realizar un estudio sobre las calificaciones de los alumnos de un instituto, una mañana se extrae una muestra con las calificaciones de seis alumnos de cada clase, elegidos por orden de llegada. Tras realizar el estudio se obtienen unos resultados que nada se asemejan con la realidad, pues las calificaciones parecen ser brillantes en la escuela, mucho mayor de lo que realmente son. El motivo del fracaso del estudio ha venido por la técnica de muestreo usada, ya que no se ha tenido en cuenta, por ejemplo, la posible correlación que guarda el orden de llegada de los alumnos al instituto (puntualidad) con el resultado académico.

En el estudio estadístico deberá existir al menor una característica o propiedad observable que defina y distinga a los individuos y sobre la que se realiza el análisis. Estas características se llaman **variables** y los **datos** son los valores concretos que toma la variable para cada individuo.

Cuando se tiene disponibilidad de una cantidad suficiente de datos o una muestra que los represente fielmente, interesa obtener valores numéricos que definen los datos o estiman características de la población a la que se refieren; estos son llamados **estadísticos o parámetros estadísticos**.

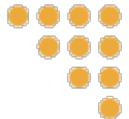
3.1. Parámetros estadísticos básicos

Como se ha dicho anteriormente, un parámetro estadístico es un valor obtenido a partir de los datos y que proporciona información sintetizada o resumida sobre estos. Existen numerosos tipos de parámetros estadísticos que pueden ser clasificados por el tipo de información que dan. Sin embargo, como requieren un tratamiento distinto en función de la naturaleza del dato (tipo de variable), tan sólo se describirán los más principales. Los siguientes parámetros estadísticos se denominan **de centralización, de posición o medidas de tendencia central** y dan información sobre el intervalo o valor en torno a los cuales se distribuyen los datos de la variable estadística observada.

3.1.1. Media aritmética

Dado un conjunto de los N valores de una variable concreta $X = \{x_1, x_2, \dots, x_N\}$, la media aritmética, o promedio, se simboliza como \bar{x} y se calcula mediante la expresión:

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N} = \frac{x_1 + x_2 + \dots + x_N}{N}$$



3. Definiciones

Esta medida proporciona el valor promedio de la muestra, el cual puede ser interpretado como una especie de centro de gravedad de la distribución de los valores. Si se hiciese una repartición perfectamente equitativa entre cada una de las observaciones $\{1, 2, \dots, N\}$ con el monto total de la variable, esta cantidad sería la media aritmética.

3.1.2. Mediana

Si se ordenase el conjunto anterior de la variable X de menor a mayor valor, la mediana sería el valor central de la secuencia. Es decir, independientemente de la distribución de los valores, la mediana será el valor que aparezca en el centro de estos, dejando siempre el mismo número de valores a su izquierda y a su derecha. Se denota por Me .

Pueden darse dos casos: cuando el número de valores es impar, siempre existirá un valor central; sin embargo, cuando el número de valores de los que se va a extraer la mediana es par existirán dos valores centrales y, en ese caso, la mediana será igual al valor promedio de estos dos.

Así, la mediana de los valores $\{2009, 5, 9430, 6, 74\}$ es 74 pero, para calcular la mediana de $\{12, 29, 3, 7\}$, se debe calcular el valor promedio de sus valores centrales $\{12, 29\}$ y por lo tanto la mediana resulta 20,5.

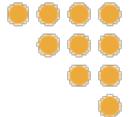
Es importante notar que, a pesar de que la media aritmética se usa más frecuentemente, ésta puede no dar información útil cuando existen datos atípicos, ya que pueden alterar enormemente el valor de la media.

Por ejemplo, una media aritmética de $\bar{x} = 11758$ puede proporcionar una idea acerca de en torno a qué orden de valores se encuentran los datos. Sin embargo, esta media puede ser obtenida de los valores $\{1, 2, 4, 5, 10, 2, 1, 94030\}$, secuencia que no queda definida por su media aritmética debido a la existencia de valores atípicos.

En estos casos, la mediana proporcionaría la información necesaria acerca de la posible existencia de estos valores.

Cuantiles

Al igual que la mediana, son medidas de localización e indican el valor que ocupa una determinada posición. Dividen la distribución en un número determinado de partes, y nombran como cuantiles los valores que ocupan los valores límites entre



3. Definiciones

los intervalos obtenidos de la partición. Según el número de partes, pueden encontrarse distintos tipos, entre los que destacan:

- **Cuartiles.** Dividen la distribución de valores en cuarto partes y definen, por ello, tres medidas correspondientes a los valores que ocupan los límites del 25%, el 50% y el 75%. Se denotan por Q_1, Q_2, Q_3 .
- **Deciles.** Compuestos por 9 medidas que dividen la distribución en diez partes iguales. Se denotan por D_1, D_2, \dots, D_9 .
- **Percentiles.** Dividen la distribución en 100 partes y están formados por 99 medidas denotadas como P_1, P_2, \dots, P_{99} .

Debido al número de medidas, o divisiones realizadas, por cada uno de los estadísticos, el interés por el cálculo de los distintos cuantiles variará según el tamaño de la variación. Así, si se tiene una distribución de 80 valores es imposible obtener los 99 percentiles. Es importante notar la coincidente naturaleza de los cuantiles y de la mediana. De hecho, tanto Q_2 , como D_5 y P_{50} siempre coincidirá con el valor de la mediana, ya que representan una partición equivalente a $\frac{1}{2}$.

Al igual que ocurría con la mediana, si la partición fuese exacta, y esto condujese a la obtención de dos valores para la medida, se recurriría a elegir el valor promedio de estos.

Para la distribución de valores

$$\{4, 2, 2, 3.1, 4, 5, 0, 1, 9, -3, 2, -4, 1, 2, 2, 2, 2, 4, 9, 10\}$$

se podrían obtener los deciles y cuartiles mediante la ordenación de los valores y su posterior partición. Ordenados quedarían como

$$\{-4, -3, 0, 1, 1, 2, 2, 2, 2, 2, 3.1, 4, 4, 4, 5, 9, 9, 10\}$$

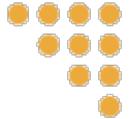
La distribución se compone de 20 elementos, por lo que la división de los cuartiles se posicionaría en la quinta/sexta, décima/undécima y decimoquinta/ decimosexta posición. Estos valores son

$$Q_1 = 1.5, \quad Q_2 = 2, \quad Q_3 = 4$$

Y los deciles

$$D_1 = -1.5, \quad D_2 = 1, \quad D_3 = 2, \quad D_4 = 2, \quad D_5 = 2,$$

$$D_6 = 2.55, \quad D_7 = 4, \quad D_8 = 4.5, \quad D_9 = 9$$



3. Definiciones

3.1.3. Moda

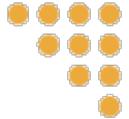
La moda es el valor de una variable que tiene mayor frecuencia absoluta, es decir, el valor más repetido en el conjunto de valores de la variable. Se denota por Mo . Es el único parámetro estadístico que puede estar compuesto por más de un valor, ya que puede darse el caso en el que dos valores coincidan en su frecuencia absoluta. Además, es el único estadístico de los tres presentados que puede tratar variables no numéricas.

A partir de un censo, es posible determinar la moda de la raza de los habitantes de una determinada zona, por ejemplo. Sin embargo, resultaría imposible obtener de este tipo de dato la media o la mediana.

Ante este tipo de dato, de variables, sólo es posible obtener estadísticos calculados a partir de conteo o distribución, es decir, relacionados con la cantidad de veces que sus valores se repiten y cómo se *distribuyen* este valor de repetición; por ejemplo, con la fracción de elementos que poseen el valor, normalmente medido en porcentaje.

Supóngase que trabaja en el departamento de recursos humanos de una gran empresa y, como cada año en estas fechas, es momento de realizar las evaluaciones del personal. A continuación se presenta una tabla de datos de los veinte primeros trabajadores.

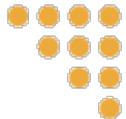
ID	Antigüedad	Sexo	Nota	Calificación	Altura	Peso
1	15	H	5,5	APRO	1,76	81
2	12	H	8,3	NOT	1,62	72
3	1	M	9	SOB	1,63	62
4	2	H	8,8	NOT	1,82	86
5	5	M	7	NOT	1,52	55
6	2	H	6,8	APRO	1,83	90
7	9	H	6,3	APRO	1,8	93
8	3	H	5,9	APRO	1,71	65
9	4	M	7	NOT	1,6	68
10	2	H	8,1	NOT	1,75	63
11	11	M	6,7	APRO	1,74	69
12	5	H	6,2	APRO	1,78	78
13	1	H	5,6	APRO	1,81	86
14	14	M	5,8	APRO	1,76	80
15	3	M	7	NOT	1,67	52
16	3	H	6,5	APRO	1,68	70
17	2	H	7,3	NOT	1,7	71
18	4	H	3,9	SUSP	1,73	80
19	8	M	7,1	NOT	1,58	59
20	1	H	4,5	SUSP	1,79	75



3. Definiciones

En primer lugar, se quiere explotar esta información y extraer estadísticos que proporcionen una idea sobre los recursos humanos de la empresa. Se obtiene la siguiente información:

- La antigüedad media de la empresa es de $\frac{\Sigma(15+12+\dots+8+1)}{20} = 5,35$ años, la moda es de 2 años y la mediana es de 3,5 años.
- En cuanto al sexo de los trabajadores la moda es Hombre, con un 70% de los trabajadores.
- La calificación moda es el *aprobado* con un 45% de los trabajadores. De los 8 trabajadores que obtuvieron *notable* (40%), el 62,5% fueron mujeres.
- La altura media en hombres es de 1,75m y el peso medio 77,7kg; en mujeres es de 1,64m y 63,6kg.



4. Análisis estadístico

Se le llama análisis estadístico o estudio estadístico al proceso formado por un conjunto de técnicas estadísticas que permite, a partir de los datos de una población, encontrar patrones de comportamiento ocultos o difícilmente observables directamente de los datos y, a partir de estos, modelar estos patrones para realizar predicciones, descubrir relaciones entre características, entre otros.

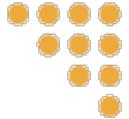
Este proceso puede ser desarrollado en las siguientes fases:

- **Planteamiento del problema.** Como paso primero, debe definirse el objetivo del estudio de tal manera que se tenga claro detalles del mismo como la población objeto de estudio o qué información será necesaria o de relevancia para el objetivo.
- **Recolección de datos.** Esta fase puede, a su vez, dividirse en dos: una primera de definición de metodologías, en la que se tiene que decidir las fuentes de la que se recogerán los datos, cómo se recopilarán estos datos (dataset públicos, entrevistas-encuestas, web-scraping, etc.), qué técnicas de muestreo se llevarán a cabo dado el volumen de datos que se espere, como se interpretarán datos no esperados, etc.; y una segunda fase más práctica en la que se llevan a cabo las acciones anteriormente definidas.
- **Análisis de los datos.** En esta etapa se comienza el trabajo puramente estadístico y se puede clasificar en análisis descriptivo y análisis inferencial. A partir del análisis descriptivo se obtendrán estadísticos que definan y proporcionen la máxima información del bloque completo de datos mediante la adecuada técnica de muestreo. Posteriormente, el análisis inferencial se encargará de ajustar el comportamiento de los datos a cierta distribución conocida o modelar estos.
- **Interpretación de los datos.** Finalmente, con el modelo o las herramientas creadas en base a los datos se podrá interpretar los resultados para poder tomar decisiones en base a ellos, realizar predicciones sobre datos futuros o extraer más información inicialmente no disponible.

Together for Tomorrow!
Enabling People

Sistemas continuos y discretos

#TecnologíaConPropósito



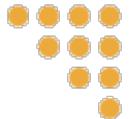
1. Introducción

La estadística descriptiva realmente comienza una vez se han recolectado los datos, mediante adquisición directa, descargándola de internet, comprándola, etc., o mediante entrevistas y encuestas.

A partir de aquí, se debe poner mucha atención en lo que se tiene enfrente: ¿Son fidedignos estos datos? ¿Existe una manera de averiguar cuáles no lo son o tienen mayor probabilidad de que no lo sean? ¿Es posible redistribuir estos datos para que proporcionen más información aún?

Para poder responder estas preguntas, antes hay que conocer qué datos se tiene, cuál es su origen o en qué formato aparecen presentados.

A lo largo de esta unidad, se verán estos pasos previos. Primero distinguiendo correctamente entre los distintos tipos de variables que se puede encontrar en una base o tabla de datos; después dando un repaso a los gráficos más relevantes para tratar los distintos tipos de dato.



2. Tipos de datos

La naturaleza del dato es importante para su tratamiento. Según el formato en que el dato se presente el analista se enfrentará a una información que requerirá distintas herramientas de representación, cálculos y métodos para su análisis e interpretación de resultados.

Estas variables se pueden clasificar según el tipo de dato que recogen:

- **Variable cualitativa.** Son todas aquellas variables que recogen información con valores no numéricos, que no se pueden asociar a números y, por lo tanto, no se puede realizar operaciones algebraicas con ellos. Cuando estas variables no pueden ordenarse se denominan **nominales** (raza de una persona, color de ojos, tipo de un producto, etc.), de lo contrario son **ordinales** (calificaciones cualitativas, grados de satisfacción, etc.).
- **Variable cuantitativa.** Se caracterizan por ser representadas con valores numéricos. Se pueden, a su vez, clasificar en **discretas** o **continuas**. Las primeras sólo pueden tomar valores concretos aislados (número de miembros de la familia, número de piezas por paquete, etc., mientras que las variables continuas pueden tener cualquier valor de un intervalo como distancias, tamaños, etc.

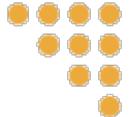
2.1. Variable discreta

Una variable cuantitativa es de tipo discreto cuando su campo de variación se compone de un número finito o numerable de puntos, existiendo en ellos *masa discreta*. Es decir, la variable no puede tener cualquier valor dentro del intervalo, tan solo ciertos, normalmente equidistantes, llamados también *puntos de salto*.

Este tipo de variable está asociado con el conteo o cantidades indivisibles de elementos, como número de piezas, número de hijos, edades. Como ayuda para reconocer una variable discreta y discernir ésta de una variable continua, suelen venir representados con números enteros, aunque no necesariamente, y siempre debe ser posible “contar” o hacer una lista de todos los posibles valores que puede tomar dicha variable.

2.2. Variable continua

Una variable continua es aquella medida dentro de un rango continuo de infinitos valores numéricos. Suelen ser representados con números reales.



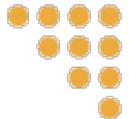
2. Tipos de datos

Dados dos valores medidos de dicha variable, siempre existe la posibilidad de encontrar una muestra con un valor dentro del intervalo que encierran esos valores.

Si la variable es edad, cuando se tienen los valores 30 y 32, es posible encontrar un valor dentro de estos, el 31. Sin embargo, entre 30 y 31 ya no es posible encontrar un nuevo valor entre estos.

En la práctica existirá siempre una resolución máxima a partir de la cual no se pueda encontrar fracciones menores en la medida, causado por las limitaciones propias de las herramientas de medida, lo que dará apariencia discreta a la variable. Sin embargo, estos serán aproximaciones que conducirán a un error de medida y, por lo tanto, no reflejarán el valor real que sí se mueve en un rango continuo.

La estatura es una medida que normalmente viene representada con una resolución de 0,01. Esto es, el salto mínimo entre valores es de 0,01 y una persona puede tener una estatura de 1,78 o 1,79 metros aparentando ser puntos de salto de una variable discreta; pero esto viene limitado por la herramienta de medida y no por la naturaleza de la variable, ya que, de ser posible su medida precisa, podría tener una estatura de 1,785769899 metros.



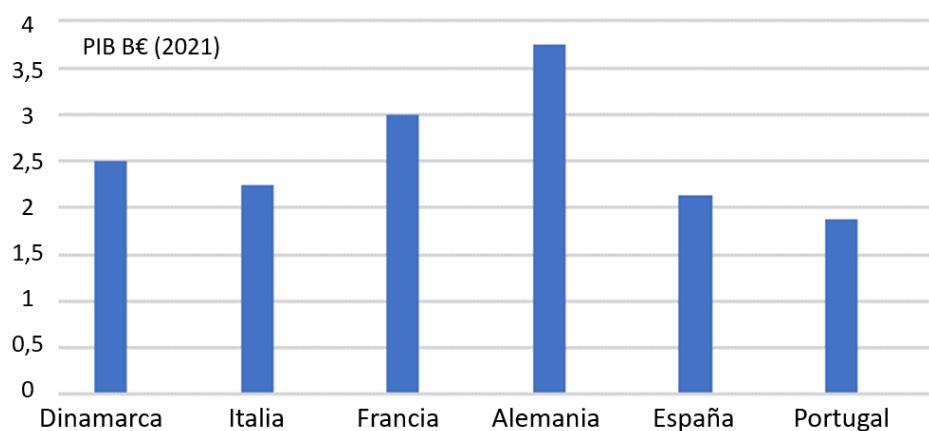
3. Representación de los datos

En un análisis estadístico, la organización de toda la información recogida en la fase de recolección de datos constituye una actividad crucial en el proceso. Una buena organización de los datos, como primer paso en su tratamiento, facilita una representación más comprensible para posteriores cálculos y finales conclusiones o decisiones.

El modo de organizar los datos dependerá de la cantidad de observaciones distintas y de la naturaleza de sus atributos o, lo que es lo mismo, el tipo de variable que contiene. Partiendo de la tabla de datos, considerada como el almacén estándar de la información, existen distintas maneras de representar esta información de manera gráfica. A continuación, se explicarán las principales herramientas gráficas para representar de la mejor manera posible variables discretas, continuas o ambas.

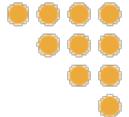
3.1. Diagrama de barras

Los diagramas de barras son herramientas gráficas usadas para representar la frecuencia con la que los distintos valores de una variable cualitativa o cuantitativa discreta se repiten. En un espacio cartesiano, se representan las distintas categorías o valores numéricos encontrados en la variable y sobre el eje vertical, la cuenta de las veces de cada valor. Las barras que crecen de cada posición correspondiente del eje horizontal, alcanzará una altura igual a este valor asociado.



3.2. Histogramas

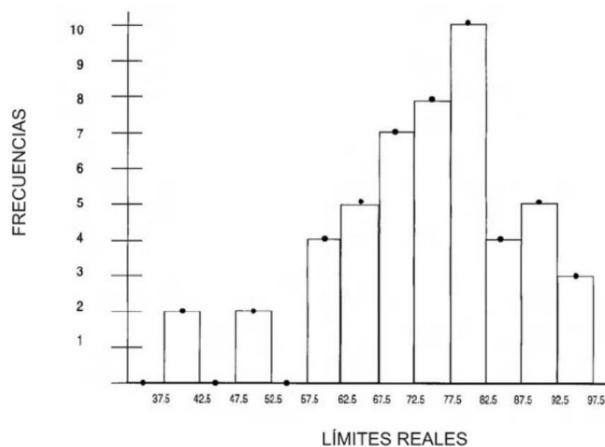
Los histogramas son gráficas ampliamente usadas en la representación de distribuciones de variables continuas. Al contrario que las gráficas de barras, que sirven para mostrar la frecuencia de los distintos valores de una variable categórica, las barras de un histograma muestran el número de observaciones cuyo valor de la



3. Representación de los datos

variable continua representada está dentro del intervalo que la barra representa. El eje horizontal muestra el rango de valores existente en la variable continua, agrupados en intervalos de mismo tamaño, los cuales tendrán una barra asociada; el eje vertical representa la frecuencia de cada intervalo de la variable. Entonces, para un mismo conjunto de datos y variable puede haber infinitas versiones de histogramas que varíen en función del tamaño del intervalo escogido, que impactará en el ancho y, a su vez, en el número de barras.

El motivo de su extendido uso es la rápida información visual que es posible obtener de ella ya que, con la práctica adecuada, se puede aproximar estadísticos asociados a la dispersión, tendencia central o continuidad de los valores de un solo vistazo.

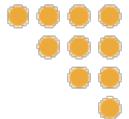


3.3. Diagrama de dispersión

También conocido como gráfico de puntos o diagrama XY, ya que se representa en el plano cartesiano XY y mediante puntos, el diagrama de dispersión es la herramienta gráfica más empleada cuando se necesita analizar la posible relación existente entre dos variables continuas.

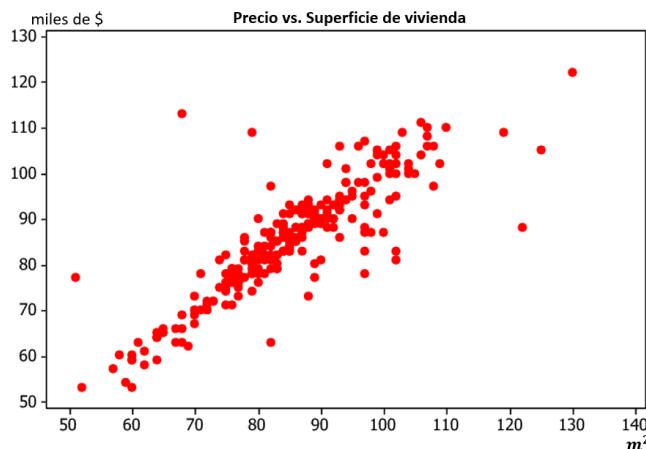
Dadas dos variables con valores continuos para m observaciones, de modo que queda como $X_1 = \{x_{11}, x_{12}, \dots, x_{1m}\}$ y $X_2 = \{x_{21}, x_{22}, \dots, x_{2m}\}$, el diagrama de dispersión se construye mediante los ejes $X_1 X_2$ coordinados correspondientes a cada una de las variables, en los que cada punto i , localizado en $\{x_{1i}, x_{2i}\}$, corresponde a un elemento observado.

Esta gráfica tiene un gran valor visual ya que permite al analista hacerse una idea sobre si existe algún tipo de relación entre ambas variables y, en caso de existir, qué grado de relación. Puede ayudar incluso a realizar predicciones sobre el valor de una variable a partir del valor de la otra.



3. Representación de los datos

Si para realizar un estudio estadístico sobre el mercado inmobiliario de una zona determinada se tiene una lista de pares de datos referentes al precio de viviendas y la superficie de construcción, el diagrama de dispersión quedaría parecido al de la imagen.



Sobre el eje horizontal se representaría el rango de valores de la variable continua de superficie y el eje vertical el precio de la vivienda. Para cada muestra individual u observación, se dibuja un punto localizado en sus valores correspondientes.

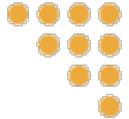
3.4. Diagrama de Box-Whisker

Por sus diferencias, las variables continuas y las discretas poseen también distintas probabilidades. En las variables discretas la frecuencia con la que un valor se repite puede ser una información muy representativa; fenómeno poco frecuente y, por lo tanto, poco útil en el caso de las variables continuas.

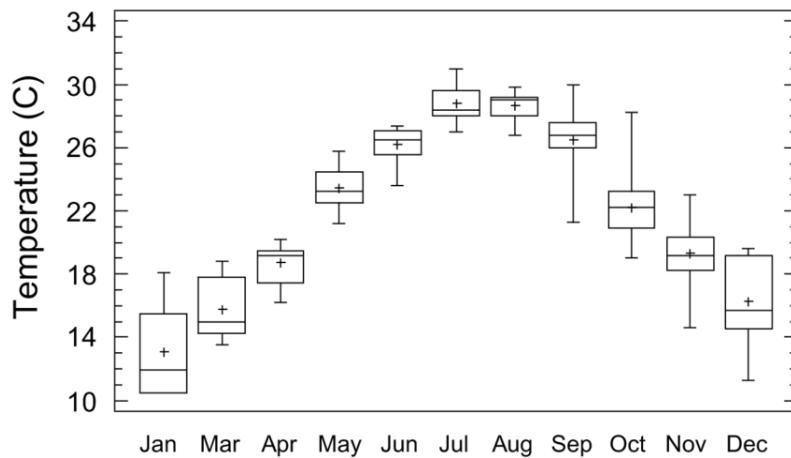
Así, en ocasiones habrá muchas observaciones donde se podrá encontrar una distribución de distintos valores de una variable continua con el mismo valor de otra variable discreta. Una de las herramientas más importantes para representar los distintos valores de una variable continua, agrupados por cada uno de los valores de otra variable, esta vez, discreta es el llamado diagrama de Box-Whisker. De este modo es posible analizar la posible relación entre ambas variables de distinto tipo.

Al igual que en un diagrama de dispersión, se construye un espacio cartesiano con las dos variables. Pero en este caso, en lugar de puntos, la gráfica se completa con las llamadas *cajas y bigotes* (Box-Whisker).

Para cada valor de la variable discreta x habrá un objeto caja-bigote dibujado, representando la distribución de valores continuos encontrados en la variable continua cuando la variable discreta es igual a x .



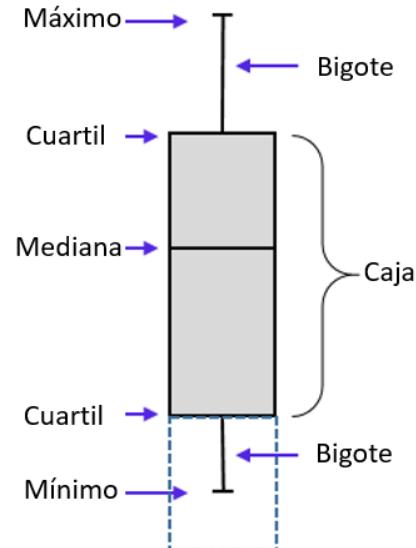
3. Representación de los datos

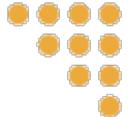


Estas cajas y bigotes representan una herramienta muy completa e informativa ya que, separando el objeto en sus distintas partes, es posible obtener los principales parámetros estadísticos de centralización.

Como se puede ver en la imagen de abajo, el objeto se compone de los siguientes elementos:

- La **caja** encierra el rango en el que se encuentra el 50% de la distribución de valores de la variable continua. Esto quiere decir que el extremo superior indicará el valor del *3^{er} cuartil* y el extremo inferior indicará el valor del *1^{er} cuartil*. Los cuartiles son cuantiles, —parámetros que definen particiones determinadas en una distribución de valores—, que dividen la distribución en cuatro partes.
- La caja queda dividida por una línea que representa la **mediana** de los valores. A veces, se puede encontrar una cruz en el centro de la caja indicando la **media**; en cualquier caso, el centro de la caja siempre estará localizado en el valor promedio.
- De cada extremo de la caja salen los **bigotes**, que representan sendos extremos de la distribución de valores y, por lo tanto, vendrán delimitados por el **máximo** y el **mínimo** de todos los valores de la variable.
- En algunas ocasiones, se pueden encontrar puntos dispersos por arriba y/o debajo de los bigotes, indicando que son valores identificados como **atípicos**.





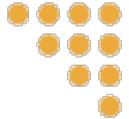
3. Representación de los datos

Se necesita analizar la información de la tabla siguiente. En primer lugar, habrá que saber qué tipos de variables se presentan.

ID	Antigüedad	Sexo	Nota	Calificación	Altura	Peso
1	15	H	5,5	APRO	1,76	81
2	12	H	8,3	NOT	1,62	72
3	1	M	9	SOB	1,63	62
4	2	H	8,8	NOT	1,82	86
5	5	M	7	NOT	1,52	55
6	2	H	6,8	APRO	1,83	90
7	9	H	6,3	APRO	1,8	93
8	3	H	5,9	APRO	1,71	65
9	4	M	7	NOT	1,6	68
10	2	H	8,1	NOT	1,75	63
11	11	M	6,7	APRO	1,74	69
12	5	H	6,2	APRO	1,78	78
13	1	H	5,6	APRO	1,81	86
14	14	M	5,8	APRO	1,76	80
15	3	M	7	NOT	1,67	52
16	3	H	6,5	APRO	1,68	70
17	2	H	7,3	NOT	1,7	71
18	4	H	3,9	SUSP	1,73	80
19	8	M	7,1	NOT	1,58	59
20	1	H	4,5	SUSP	1,79	75

Cada fila en la tabla corresponde a un ejemplo o muestra unitaria. Las distintas variables corresponden a cada una de las columnas de la tabla:

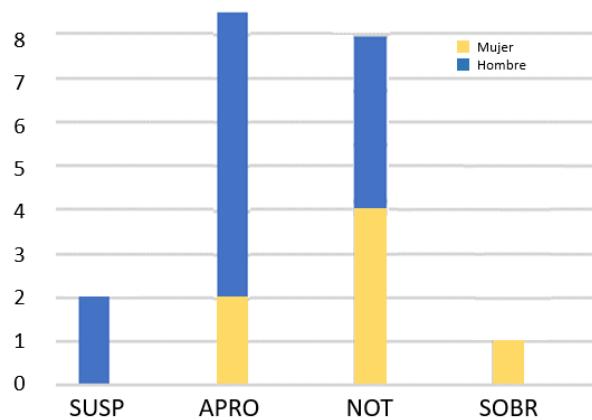
- **ID.** Es el identificador con el que se asocia cada muestra unitaria. Suele introducirse automáticamente y de manera ordenada. No es una variable, es una etiqueta.
- **Antigüedad.** Mostrada en años, es una variable cuantitativa discreta. No es continua ya que tiene una resolución máxima de 1 año.
- **Sexo.** Es una variable cualitativa nominal. Sus valores no son interpretables numéricamente y tampoco pueden ser ordenados. Al ser dos posibles valores, también suelen ser llamados variable cualitativas binarias.
- **Nota.** Es una variable cuantitativa continua
- **Calificación.** Variable cualitativa ordinal. Aun siendo cualitativa si puede interpretarse distintos grados.
- **Altura y peso.** Variables cuantitativas continuas, ya que pueden tomar cualquier valor.



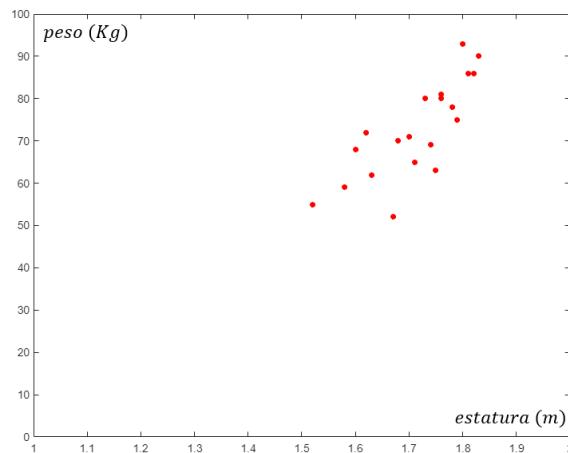
3. Representación de los datos

Conocidos los tipos de variables presentes en la tabla de datos, se puede saber cuál será la mejor manera de representarlo, es decir, cuáles son, de entre las conocidas, las herramientas gráficas adecuadas para cada variable o unión de variables.

Podría mostrarse un diagrama de barras para las variables cualitativas, como el sexo y la calificación. De hecho, podría construirse barras compuestas por ambas variables, por ejemplo, asociando cada barra a una categoría de “Calificación” y diferencia cada una mediante dos colores la cantidad de hombres y mujeres que forman parte de esa categoría.



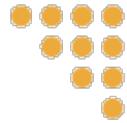
Si, por ejemplo, se quisiese analizar la posible correlación entre la estatura y el peso de los trabajadores, se usaría una tabla de dispersión.



Together for Tomorrow!
Enabling People

Distribución de frecuencias y *Probability Density Function (PDF)*

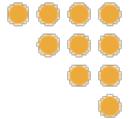
#TecnologíaConPropósito



1. Introducción

Aunque pueda parecer algo distinto, la organización de los datos constituye una fase crítica en un estudio estadístico. Saber cómo reorganizar, redistribuir o adaptar los datos a otro tipo de tablas o gráficos facilita cálculos posteriores y, además, evita posibles confusiones o malinterpretaciones. Por este motivo, esta unidad explica con detalle cómo agilizar la comprensión de, por ejemplo, una larga tabla de datos, mediante la síntesis de su información en tablas de frecuencia.

Tras esto, se pasa a desarrollar formalmente la definición de la variable aleatoria y la que es la herramienta de representación más importante de ésta, la función de distribución. Al acabar la unidad, el lector estará preparado para entender, a grandes rasgos, cualquier función de distribución que se le presente y, al mismo tiempo, analizar la variable aleatoria asociada a la función de distribución a partir de parámetros estadísticos, sabiendo cómo interpretar cada uno de ellos.



2. Distribución de frecuencias

De los conceptos que rodean a los experimentos aleatorios se pueden recoger tres que forman el modelo matemático que explica el resultado de un fenómeno aleatorio. Éstos forman el **espacio de probabilidad** (E, Ω, P): E , como ya es sabido, reúne todos los elementos del espacio muestral, Ω es el espacio de sucesos y P es la función que atribuye probabilidades a los sucesos de Ω , y que toma valores sin salir del intervalo $[0,1]$. Existen técnicas de conteo que cubren ciertos casos en los que es posible conocer el número de elementos que forma el espacio muestral. Sin embargo, conocer sus valores puede, a veces, no ser una tarea inmediata y dificultar su tratamiento matemático.

Ante un experimento aleatorio compuesto, en el cual se repite el mismo experimento n veces, el espacio muestral crecerá haciéndose cada vez menos representativo a medida que n crece. En estos casos puede hacerse interesante tener un número reducido pero suficiente de parámetros que defina este conjunto.

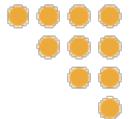
Si, por ejemplo, el experimento consiste en lanzar una moneda, estando así el experimento elemental formado por el espacio muestral $E = \{\text{cara}(c), \text{cruz}(+)\}$ y este experimento se tiene que repetir cien veces, el espacio muestral total estaría formado por un número enorme de secuencias del tipo $(c++cc+++++ccc+...)$. En muchos casos es más importante saber que el *número de caras* ha sido, por ejemplo, cincuenta y ocho que conocer, además, la secuencia exacta con que han ido ocurriendo. Esta asignación o transformación a valores numéricos facilita el manejo matemático.

Una **variable aleatoria** es una cantidad variable cuyos valores dependen del azar y para la cual existe una *distribución de probabilidad*. Asigna un valor a cada elemento del espacio muestral de un experimento aleatorio.

El lanzamiento de una moneda es un experimento aleatorio que posee un espacio muestral formado por $E = \{\text{cara}(c), \text{cruz}(+)\}$. Si se asigna el valor 1 a cara y 0 a cruz, éstos valores formarían el dominio de la variable aleatoria asociada X .

La variable aleatoria vendrá definida por el suceso de interés del experimento. En caso de que el experimento fuese el lanzamiento de un dado, podría presentarse como suceso de interés el conseguir un 5 en el lanzamiento. En este caso se tiene el espacio de probabilidad $(E, \Omega, P) = (\{1,2,3,4,5,6\}, \{5\}, \frac{1}{6})$ y la variable aleatoria sería

$$X(r) = \begin{cases} 1, & r = 5 \\ 0, & r \neq 5 \end{cases}$$



2. Distribución de frecuencias

Sin embargo, al modificar el suceso de interés y analizar ahora el problema de obtener un número par del lanzamiento, también se modifica el espacio de probabilidad y la variable aleatoria. $(E, \Omega, P) = (\{1,2,3,4,5,6\}, \{2,4,6\}, \frac{3}{6} = \frac{1}{2})$ y la variable aleatoria sería:

$$X(r) = \begin{cases} 1, & r \in \{2,4,6\} \\ 0, & r \notin \{2,4,6\} \end{cases}$$

Esta agrupación o categorización de sucesos en otros compuestos, y el impacto que esto provoca en la probabilidad de que ocurran, es lo que se denomina **distribución de frecuencias**.

2.1. Tablas de frecuencias

Habitualmente, se tienen grandes bases de datos o tablas de datos con numerosas observaciones. En ocasiones, cuando se pretende estudiar con detenimiento una variable en concreto, conviene reorganizar esta variable agrupando todos los valores encontrados y registrar la *frecuencia* con la que estos valores se repiten. Así, se crean las llamadas **distribuciones de frecuencias**.

Las distribuciones de frecuencias son ordenaciones en forma de tabla de los datos estadísticos de una o más variables, en los que se asigna la frecuencia de repetición a cada valor.

2.1.1. Tipos de frecuencias

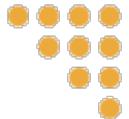
Existen varios tipos de frecuencias:

- **Frecuencia absoluta.** Se forma con el número de veces que un determinado valor aparece repetido. Se denota por f_i y suele aparecer en la segunda columna de una tabla de frecuencias, junto al valor de la variable asociada.

Suponga un estudio estadístico sobre los destinos vacacionales favoritos en España para los ingleses. Tras una larga encuesta, se recoge la respuesta de 2200 personas de origen inglés, número que se reduce a 2000 tras filtrar los resultados a los diez destinos favoritos, por agilizar el análisis.

En este caso, la variable estadística bajo estudio sería el destino vacacional, $X = \text{ciudad}$, la cual, tras el filtro, está constituido por $N = 2000$ observaciones, X_1, X_2, \dots, X_N , que toman $K = 10$ valores distintos. Para realizar la tabla, se deben registrar los K distintos valores en filas, y a la

Ciudades	f_i
Benidorm	97
Madrid	432
Ibiza	69
Mallorca	256
Málaga	102
Sevilla	458
Barcelona	491
Bilbao	15
Valencia	21
Granada	22
Gran Canaria	37
N	2000



2. Distribución de frecuencias

derecha de cada valor, registrar el número de veces que se repite, es decir, su frecuencia absoluta.

La suma de esta columna siempre deberá coincidir con el número de observaciones. Esto es, el número de filas de la tabla de la que se extrajo.

$$\sum_{i=1}^K f_i = N$$

- **Frecuencia relativa.** Se obtiene tras dividir la frecuencia absoluta de un valor determinado y el número total de datos, N . Se denota por fr_i o n_i , se formula como:

$$fr_i = \frac{f_i}{N}$$

Y debe cumplir:

$$\sum_{i=1}^k f_i = 1$$



- **Frecuencia acumulada.** Esta última frecuencia registra la acumulación de frecuencias absolutas de todos los valores inferiores o iguales al valor bajo estudio. Es decir, si se disponen los distintos valores de menor a mayor, se obtiene de la suma de todos los valores vistos (filas arriba) hasta el valor actual. Viene dada por la expresión

$$F_k = \sum_{i=1}^{k \neq K} f_i$$

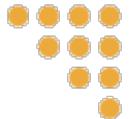
De este modo se tendría $F_1 = f_1$, $F_2 = f_1 + f_2$, $F_k = f_1 + f_2 + \dots + f_k$

También podría añadirse una nueva columna con la llamada *Frecuencia relativa acumulada*.

$$Fr_k = \sum_{i=1}^{k \neq K} fr_i$$

Continuando el ejemplo anterior, se completa la tabla con las nuevas frecuencias:

Ciudades	f_i	fr_i	F_k
Benidorm	97	0,0485	97
Madrid	432	0,216	529
Ibiza	69	0,0345	598
Mallorca	256	0,128	854
Málaga	102	0,051	956
Sevilla	458	0,229	1414
Barcelona	491	0,2455	1905
Bilbao	15	0,0075	1920
Valencia	21	0,0105	1941
Granada	22	0,011	1963
Gran Canaria	37	0,0185	2000
	Σ	2000	1



2. Distribución de frecuencias

2.1.2. Distribuciones de frecuencias por agrupación

Habrá veces en las que, debido a la gran variedad de valores encontrados en los datos de una variable, conviene agrupar estos valores en intervalos representativos con el fin de realizar un recuento o extraer parámetros más informativos. Es una manera de discretizar una variable continua.

Está estandarizado el formato de intervalo del tipo $(a, b]$, a excepción del primer intervalo, que será cerrado por ambos extremos, $[a, b]$.

A la hora de establecer el número de intervalos existen varios métodos de los que dos destacan por su amplio uso.

El **método de la raíz**, por el cual se determina el número de intervalos k como la raíz cuadrada del número total de observaciones, N . $k = \sqrt{N}$

La **fórmula de Sturges**, que calcula el número de intervalos k como

$$k = 1 + \log_2 N$$

2.2. Función de distribución

Una variable aleatoria, ξ , representa un experimento aleatorio y está definida por el espacio muestral (E) de dicho experimento, o *campo de variación*, y el conjunto de probabilidades con que toma valores en cada uno de los elementos de ese espacio muestral.

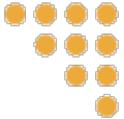
Cuando la variable es discreta, y existe un número finito de posibles valores en el espacio muestral, la función que rige la transformación entre estos elementos y sus probabilidades asociadas en el experimento, denominada **función de probabilidad**, es fácil de entender y se define por

$$p_i = P(\xi = x_i), \quad \forall x_i \in E$$

En términos generales, la probabilidad del suceso en que la variable aleatoria toma cualquier valor menor o igual que x de entre todos los posibles, esto es, $\xi \in (-\infty, x]$, se denomina **función de distribución** de la variable aleatoria

$$F(x) = P(\xi \leq x)$$

Si se ordenasen de menor a mayor todos los posibles valores de un experimento aleatorio, esta función definiría la probabilidad de que ocurra el suceso compuesto asociado a todos los valores a la izquierda de x , incluido él mismo.



2. Distribución de frecuencias

2.1.1. Propiedades

1. $F(-\infty) = 0, \quad F(\infty) = 1$

Por lo explicado anteriormente, $F(-\infty)$ definiría la probabilidad de que la variable aleatoria cumpliese $\xi \in (-\infty, -\infty]$. Esto es lo mismo que buscar la probabilidad del suceso encerrado por un conjunto vacío \emptyset o, dicho de otro modo, la probabilidad de un suceso imposible.

Análogamente, $F(\infty)$ se asociaría con $\xi \in (-\infty, \infty]$, formando un conjunto universal que define un suceso seguro.

2. $P(x_1 < \xi \leq x_2) = F(x_2) - F(x_1)$

Efectivamente, de la teoría de conjuntos, es sabido que $S_1 = \{\xi \leq x_1\}$ y $S_2 = \{x_1 < \xi \leq x_2\}$ son sucesos disjuntos, es decir, no existe intersección entre ellos ($S_1 \cap S_2 = \emptyset$), por lo que $P(S_1 \cup S_2) = P(S_1) + P(S_2)$. Es fácil ver que esta suma es igual a $P(\{\xi \leq x_2\})$.



Se tiene entonces

$$\begin{aligned} P(S_2) &= P(S_1 \cup S_2) - P(S_1) \rightarrow P(x_1 < \xi \leq x_2) = P(\xi \leq x_2) - P(\xi \leq x_1) \\ &= F(x_2) - F(x_1) \end{aligned}$$

3. *La función es monótona no decreciente*

Si se sabe que la probabilidad siempre es mayor o igual que cero, junto a lo visto en el anterior punto se tiene que, para $x_1 \leq x_2$, se cumple siempre $F(x_1) \leq F(x_2)$.

$$\text{si } 0 \leq P(x_1 < \xi \leq x_2) = F(x_2) - F(x_1) \rightarrow F(x_1) \leq F(x_2)$$

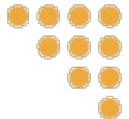
4. *La función de distribución es continua por la derecha*

Por la propia condición de continuidad para una función, la función de distribución debe cumplir

$$\lim_{\Delta x \rightarrow 0} |F(x + \Delta x) - F(x)| = 0$$

Efectivamente, si $F(x + \Delta x) - F(x) = P(x < \xi \leq x + \Delta x)$ el conjunto encerrado por el intervalo contiene tan solo a Δx , por lo que tiene que ser el conjunto vacío,

$$\lim_{\Delta x \rightarrow 0} |F(x + \Delta x) - F(x)| = P(\emptyset) = 0$$



2. Distribución de frecuencias

Si se analiza la continuidad por la izquierda, $\lim_{\Delta x \rightarrow 0} |F(x - \Delta x) - F(x)| = 0$ significaría que $P(x - \Delta x < \xi \leq x) = 0$. El intervalo, en este caso, contiene tan solo a x y no necesariamente será igual a cero.

2.3. Variables aleatorias discretas

Una variable aleatoria es de tipo discreto cuando su campo de variación se compone de un número finito de puntos, un espacio muestral donde los distintos valores posibles pueden ser listados e identificados. En estos casos, cada posible valor lleva asociado una probabilidad discreta.

Como se ha explicado anteriormente, la función que gobierna esta asignación de probabilidades a los valores de una variable aleatoria discreta se denomina función de probabilidad o función de cuantía. Dicha función es del tipo

$$p_i = P(\xi = x_i)$$

en el que cada punto p_i es llamado **punto de salto**. Según las propiedades de la probabilidad, se debe cumplir:

$$\sum_{i=1}^{\infty} P(\xi = x_i) = \sum_{i=1}^{\infty} p_i = 1$$

y, más importante, se relaciona con la función de distribución de la siguiente manera:

$$F(x_i) = P(\xi \leq x_i) = \sum_{x_i \leq x} P(\xi = x_i) = \sum_{k=0}^{k=i} p_k$$

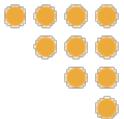
Se habrá notado ya que se está hablando de la *frecuencia relativa* y *frecuencia acumulada*, en términos de probabilidad frecuentista.

Supóngase que se tiene un dado imperfecto, por lo que las probabilidades varían entre las distintas caras. Se llevan a cabo 50 lanzamientos y se anotan los resultados en una tabla

cara	1	2	3	4	5	6
n	5	20	5	10	4	6

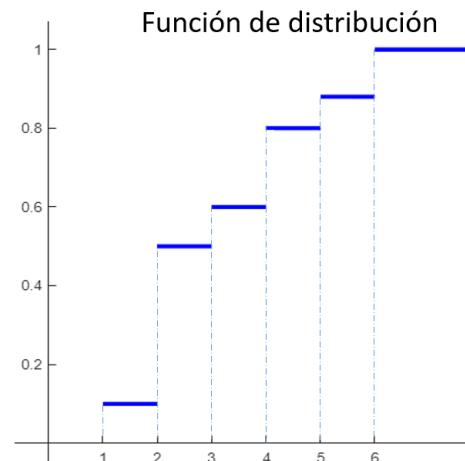
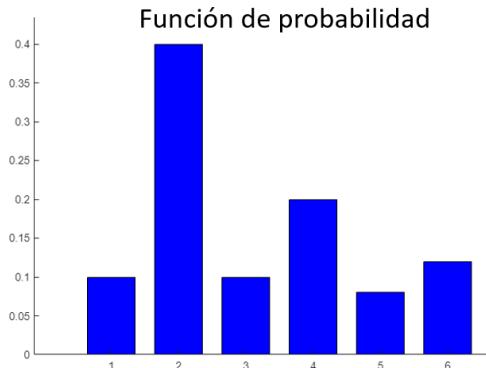
Mediante probabilidad frecuentista es posible calcular las probabilidades de cada cara mediante $P(x_i) = n/N$ siendo n el número de veces que el dado cayó sobre esa cara y N el número total de lanzamientos. A partir de estas probabilidades, se puede definir la función acumulada, como el resultado de la función de dsitribución asociada. Así,

cara	1	2	3	4	5	6
n	5	20	5	10	4	6
p _i	0,10	0,40	0,10	0,20	0,08	0,12
$\sum p_i$	0,10	0,50	0,60	0,80	0,88	1,00



2. Distribución de frecuencias

De este modo, la función de probabilidad o cuantía y la de distribución quedarían así:



2.4. Variables aleatorias continuas

La función de distribución vista en el ejemplo anterior, a pesar de ser continua por la derecha, no era derivable en ciertos puntos debido a los saltos discontinuos entre los distintos valores discretos de x . Cuando la variable aleatoria es continua su función de distribución, $F(x)$, es continua y derivable, y su derivada es también continua (en estos casos la función se denomina *absolutamente continua*).

Es muy importante entender lo siguiente para comprender así las distribuciones de probabilidades en variables continuas. Si se analiza el intervalo infinitesimal $(x; x + \Delta x]$ se tienen las funciones de distribución

$$F(x + \Delta x) = P(\xi \leq x + \Delta x)$$

$$F(x) = P(\xi \leq x)$$

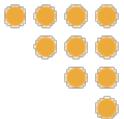
que al ser restadas queda

$$F(x + \Delta x) - F(x) = P(x < \xi \leq x + \Delta x)$$

Anteriormente se demostró que el intervalo definido por $x < \xi \leq x + \Delta x$ es igual a Δx , por lo que

$$P(x < \xi \leq x + \Delta x) = P(\xi = x) = F(x + \Delta x) - F(x) = 0$$

Esto quiere decir que, cuando se tratan variables aleatorias continuas, la probabilidad de que la variable tome un valor exacto es nula.



2. Distribución de frecuencias

Puede parecer difícil de comprender, pero si se piensa en la condición inherente de continuidad de valores en un intervalo, que conlleva tener infinitos valores diferenciados por una variación infinitamente pequeña, se hace obvio que la probabilidad de que ocurra un suceso entre infinitos sea muy próxima a cero.

Si la diferencia anterior, $F(x + \Delta x) - F(x)$, que era igual a la probabilidad de que la variable caiga en el intervalo $(x; x + \Delta x]$, se divide por la amplitud del propio intervalo Δx , se tiene una especie de probabilidad media distribuida a lo largo del intervalo. Si se reescribe como el límite $\Delta x \rightarrow 0$

$$\lim_{\Delta x \rightarrow 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} = F'(x) = f(x)$$

se tiene que el límite es la definición de la derivada de la función $F(x)$, $F'(x)$. Esta derivada se denomina **función de densidad** de la variable aleatoria o, en su forma acortada, **Probability Density Function (PDF)**, y se representa por $f(x)$. Podría interpretarse, a modo orientativo, como la función de probabilidad en casos de variables aleatorias continuas.

2.4.1. Función de densidad de probabilidad o PDF

Esta función de densidad $f(x)$ tampoco, al igual que ocurre con la probabilidad, podrá ser negativa.

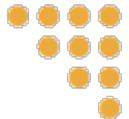
$$f(x) \geq 0$$

Debido a la condición de derivación que determina la condición de creciente a la función primitiva de una función derivada positiva, se tiene esta propiedad, ya que $F(x)$ es creciente.

En el caso de variables aleatorias discretas, la función de distribución era calculada a partir del sumatorio de las probabilidades discretas de cada uno de los posibles valores de la variable, es decir, la acumulación de los valores de la función de probabilidad o función de cuantía. Al estar tratando ahora variaciones continuas en la variable, se tienen infinitos posibles valores, por lo que este cálculo consistirá en la suma de infinitos intervalos infinitesimales. Como ya es sabido, esto se calcula a partir de la **integral definida**.

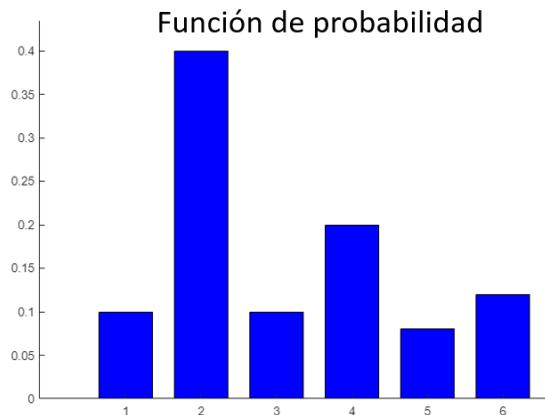
Obviamente, se cumple que:

$$\text{si } F'(x) = f(x) \rightarrow \int f(x) = F(x)$$



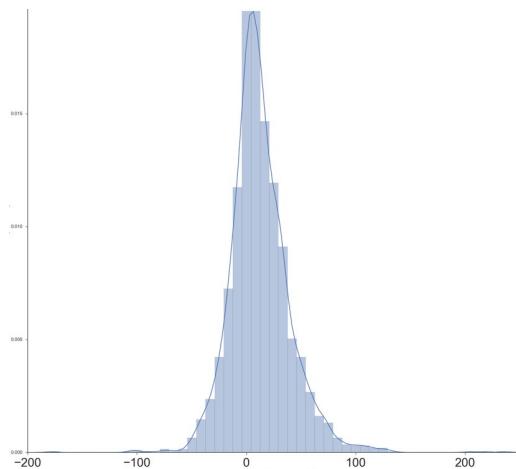
2. Distribución de frecuencias

Antes, el diagrama de barras que mostraba la función de probabilidad de la variable aleatoria discreta asociada al experimento del lanzamiento de un dado imperfecto quedaba como

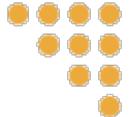


Dónde el número de barras era igual al de posibles valores que la variable aleatoria discreta podía tomar.

Si ahora se construye un histograma con una variable aleatoria continua, éste puede tener tantas barras como agrupaciones permita el tamaño de los intervalos elegidos. Para el caso de intervalos infinitesimales $\Delta x \rightarrow 0$, se tendrían infinitas barras infinitamente estrechas, llegando a tal punto de convertir el histograma en una aproximación a curva, como lo mostrado en la siguiente imagen, la cual ilustra una función de densidad.



La función de densidad, al igual que los histogramas, son herramientas con mucha información visual con respecto a la distribución de la probabilidad de una variable aleatoria.



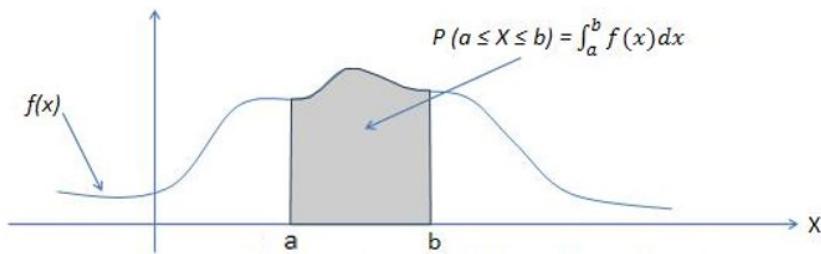
2. Distribución de frecuencias

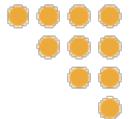
A partir de las propiedades de las derivadas y las integrales, se deduce que la función de distribución en un punto x , es decir, la probabilidad de que la variable toma cualquier valor menor de x podrá ser calculado como **el área bajo la curva** de la función de densidad en el intervalo $(-\infty, x_i]$, lo cual, a su vez se calcula como la integral definida de $f(x)$ entre $-\infty$ y x_i . Efectivamente, si la integral definida entre dos puntos a y b se calcula como

$$\int_a^b f(x) dx = F(b) - F(a)$$

y ya quedó demostrado que $F(-\infty) = 0$, entonces

$$\int_{-\infty}^{x_i} f(x) dx = F(x_i) - F(-\infty) = F(x_i)$$





3. Estadísticos de la variable aleatoria

Comprendiendo ya los conceptos de frecuencias, variable aleatoria, así como todo lo que supone la diferencia entre una variable aleatoria continua y discreta en cuanto a su tratamiento, es posible pasar a la siguiente fase en un análisis estadístico: la correcta extracción de información de valor sintetizada en parámetros.

3.1. Esperanza matemática

La esperanza matemática es a menudo relacionada erróneamente con la media aritmética. Esto es debido a que se busca la misma información cuando se hace uso de ellos, pero en enfoques o situaciones distintas. Para casos en los que tenemos experimentos aleatorios compuestos por sucesos equiprobables, como el lanzamiento de un dado perfecto, o se recogen los datos de una variable sin tener en cuenta su repetición o frecuencia relativa de los mismos, la media aritmética coincide con la esperanza matemática.

La esperanza matemática podría interpretarse como el valor medio de la distribución teórica de probabilidades de un fenómeno, coincidente con la media aritmética cuando el número de observaciones tendiese a infinito. La esperanza matemática de la distribución de probabilidad de una variable aleatoria ξ se denota por $E(\xi)$.

Para variables discretas se expresa como:

$$E(\xi) = \sum_i x_i p_i,$$

Esto es, el sumatorio de los valores que puede tomar la variable multiplicado por sus respectivas probabilidades. Dicho de otro modo, la suma de los productos de cada variable por su frecuencia relativa.

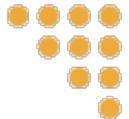
Supóngase una variable aleatoria que toma los valores 1, 2 y 3 en 10, 15 y 25 ocasiones, respectivamente, de un total de 50 experimentos.

Para obtener sus probabilidades, habrá que usar la fórmula de la frecuencia relativa, dadas sus frecuencias absolutas y el total de experimentos. Esto es

$$p_i = fr_i = \frac{f_i}{N}$$

Su esperanza matemática entonces es

$$E(\xi) = \xi_1 * \frac{f_1}{N} + \xi_2 * \frac{f_2}{N} + \xi_3 * \frac{f_3}{N} = 1 * \frac{10}{50} + 2 * \frac{15}{50} + 3 * \frac{25}{50} = 2,3$$



3. Estadísticos de la variable aleatoria

Para distribuciones de tipo continuo se calcula como

$$E(\xi) = \int_{-\infty}^{\infty} xf(x)dx$$

Donde $f(x)$ es la función de densidad de la variable aleatoria.

Se tiene ahora una variable aleatoria continua que se ha definido por la función de densidad como

$$f(x) = 12x^2, \quad 0 \leq x \leq 1$$

Su esperanza matemática entonces es

$$E(\xi) = \int_{-\infty}^{\infty} xf(x)dx = \int_0^1 x \cdot 12x^2 dx = 12 \int_0^1 x^3 dx = 12 \left[\frac{x^4}{4} \right]_0^1 = 3$$

3.2. Momentos

Estos son parámetros estadísticos primitivos, debido a que de ellos se extraen la mayoría de los más importantes parámetros estadísticos dentro del estudio de curvas de distribución, especialmente enfocadas a analizar, como se verá más adelante, sesgos y formas.

3.2.1. Momentos respecto al origen

La función de la variable aleatoria ξ , $f(\xi) = \xi^r$, la expresión $E[f(\xi)] = E(\xi^r)$ se denomina **momento respecto al origen de orden r** de la variable aleatoria ξ . Se suele denotar por α_r .

Al igual que como cualquier otro cálculo de esperanza matemática, para distribuciones de tipo discreto se tiene

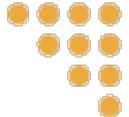
$$\alpha_r = E(\xi^r) = \sum_i x_i^r p_i,$$

y para el caso continuo

$$\alpha_r = E(\xi^r) = \int_{-\infty}^{\infty} x^r f(x)dx$$

Aunque por su definición pueden existir momentos de cualquier grado, los de uso común corresponden a los primeros 4 grados. Así,

- $\alpha_1 = E(\xi) \equiv \text{esperanza matemática}$
- $\alpha_2 = E(\xi^2)$
- $\alpha_3 = E(\xi^3)$
- $\alpha_4 = E(\xi^4)$



3. Estadísticos de la variable aleatoria

3.2.2. Momentos respecto a la media

También llamados *momentos centrados*, responden en este caso a $E[f(\xi)]$ cuando $f(\xi) = (\xi - \mu)^r$, siendo μ la media o esperanza matemática. Estos momentos se expresan mediante μ_r .

Para variables aleatorias discretas, se calcula como

$$\mu_r = E[(\xi - \mu)^r] = \sum_i (x_i - \mu)^r p_i$$

Para variables aleatorias continuas, se calcula como

$$\mu_r = E[(\xi - \mu)^r] = \int_{-\infty}^{\infty} (x - \mu)^r f(x) dx$$

3.3. Estadísticos de dispersión

La **dispersión** puede ser definida como la variabilidad que presentan entre sí los valores de una variable aleatoria alrededor del valor medio que se tome como representativo o resumen de dicha distribución.

3.3.1. Desviación media respecto a la media

La desviación media respecto a la media es la medida en términos absolutos de la dispersión en torno al valor medio, es decir

$$D(\xi) = E(|\xi - \mu|)$$

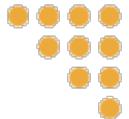
Cuando $D(\xi)$ es relativamente pequeña, los valores de la variable aleatoria estarán cerca del valor medio μ y, por lo tanto, la dispersión será pequeña.

3.3.2. Varianza

La varianza, $V(\xi)$ o σ^2 , es la *media cuadrática de la dispersión* y es igual al momento de segundo orden respecto a la media.

$$\sigma^2 = \mu_2 = E(\xi - \mu)^2$$

Cuando la varianza es pequeña, las desviaciones de la variable aleatoria en torno a su valor medio son también pequeñas, por lo que la media será muy parecida a los valores de la variable y, por lo tanto, muy representativa del conjunto de valores de la distribución y una distribución pequeña.



3. Estadísticos de la variable aleatoria

En caso contrario, cuando σ^2 es grande la dispersión es grande también y la media de la variable no es muy representativa.

3.3.3. Desviación típica

La desviación típica o *desviación estándar* es otra medida de dispersión más acorde a su comparación con los propios datos y con la media. Se representa por σ , ya que

$$\sigma = \sqrt{\sigma^2} = \sqrt{E(\xi - \mu)^2}$$

Su mayor utilidad comparado con la varianza radica en el hecho de estar en las mismas unidades que la media μ de la variable aleatoria, lo que facilita su valoración de la dispersión que se está midiendo.

3.3.4. Coeficiente de variación

Denotado por C_V , coeficiente de variación es un término para referirse a la *desviación típica relativa*, ya que se calcula como el cociente entre la desviación típica y la media o esperanza matemática

$$C_V = \frac{\sigma}{\mu}$$

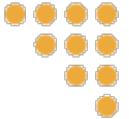
Para valores muy grandes, da mayor información de la que puede dar una desviación típica ya que no sería lo mismo, por ejemplo, una variable aleatoria puede tener una desviación típica de $\sigma_1 = 200$ sin llegar a sufrir, realmente, mayor dispersión que otra variable con desviación típica $\sigma_2 = 80$, si, por ejemplo $\mu_1 = 80230$ y $\mu_2 = 100$.

Esto es debido a que, en el primer caso se está ‘variando’ aproximadamente un 2,49% alrededor de su valor medio y en el segundo caso un 80%.

$$C_{V1} = \frac{\sigma_1}{\mu_1} = \frac{200}{80230} = 0,00249, \quad C_{V2} = \frac{\sigma_2}{\mu_2} = \frac{80}{100} = 0,8$$

3.4. Estadísticos de forma

Se han visto hasta ahora características estadísticas de una variable aleatoria que definen su posición central y la dispersión con respecto a ese valor medio. Pero existen otras características también muy importantes que informan sobre el perfil o forma que sigue una determinada distribución.



3. Estadísticos de la variable aleatoria

3.4.1. Asimetría

Una variable aleatoria ξ puede presentar una distribución **simétrica** respecto a un eje vertical en un determinado punto t si

$$P(\xi \geq t + x) = P(\xi \leq t - x), \quad \forall x$$

Si se está ante la distribución de una variable aleatoria continua y $f(x)$ es su función de densidad, dicha distribución será simétrica cuando

$$f(t + x) = f(t - x), \quad \forall x$$

En caso contrario, la distribución se considera **asimétrica**. Lo habitual es que la medida de simetría se analice sobre el eje coincidente con la media, esto es, $t = \mu$.

Ocurre que, ante una distribución simétrica con respecto a la media, las desviaciones hacia valores menores que la media serán del mismo peso que las contrarias. Puesto que los momentos centrados se calculan como la suma (discretas o continuas) de las diferencias de cada valor u observación de la variable con la media elevado a una potencia igual al orden del momento, es decir,

$$\mu_r = \sum_i (x_i - \mu)^r p_i \quad o \quad \mu_r = \int_{-\infty}^{\infty} (x - \mu)^r f(x) dx,$$

y como se sabe que los valores negativos elevados a potencias pares resultan siempre valores positivos, se tiene que una distribución es simétrica si se cumple que

$$\mu_{2r+1} = E[(\xi - \mu)^{2r+1}] = 0.$$

Es decir, cualquier momento de orden impar deberá ser cero. Puesto que

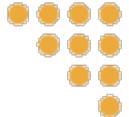
$$\mu_1 = E(\xi - \mu) = E(\xi) - E(\mu) = \mu - \mu = 0$$

sea simétrica o no la distribución, se suele usar el momento impar siguiente, μ_3 .

Así, si $\mu_3 = 0$ la distribución será simétrica, mientras que si $\mu_3 > 0$, las desviaciones positivas (de valores por encima del valor medio) tendrán mayor peso que las negativas y se denominará *distribución asimétrica positiva* o *distribución asimétrica a la derecha*. En caso contrario, será una *distribución asimétrica negativa* o *a la izquierda*.

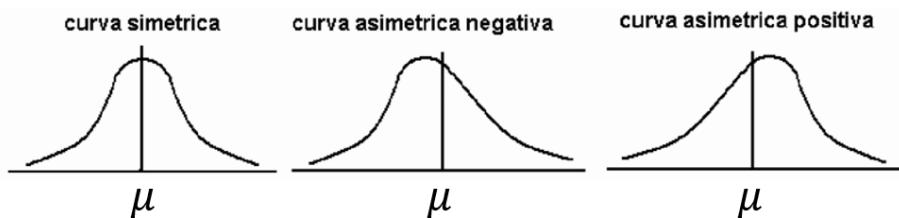
Al igual que con el coeficiente de desviación, existe un **coeficiente de asimetría**, adimensional e invariante a cambios de origen y escala, expresado por

$$\gamma = \frac{\mu_3}{\sigma^3},$$



3. Estadísticos de la variable aleatoria

Y, al igual que con μ_3 , si el coeficiente es nulo, la distribución será simétrica; si es positiva, la distribución lo será también, al igual que en caso contrario.



3.4.2. Curtosis

La curtosis es el último parámetro estadístico que se tratará en la unidad y es muy útil para una distribución muy usada en estadística inferencial y que será desarrollada en temas posteriores. Su representación es conocida como *campana de Gauss*.

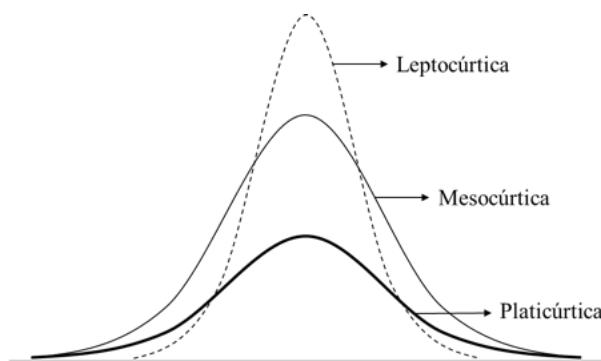
La curtosis, también llamada apuntamiento o achatamiento, mide la “pronunciación” de esta campana, es decir, para una distribución cualquiera, mide su apuntamiento con respecto a la campana de Gauss. Por este motivo, la distribución de probabilidad debe tener un perfil llamado *campaniforme* para que el valor de la curtosis sea en algún modo indicativo.

Se puede decir que la curtosis mide la concentración de valores en la zona cercana a su valor medio. Este **coeficiente de curtosis** se expresa mediante

$$\gamma_1 = \frac{\mu_4}{\sigma^4} - 3,$$

cuyo resultado se interpreta del siguiente modo:

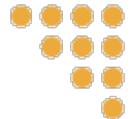
- Cuando $\gamma_1 = 0$, la distribución tiene un perfil de la curva de Gauss, de una *distribución normal*. Se le denomina entonces **mesocúrtica**.
- Cuando $\gamma_1 > 0$, la distribución es más pronunciada o apuntada que la distribución normal, se le denomina **leptocúrtica**.
- Cuando $\gamma_1 < 0$, la distribución es menos pronunciada o más achatada que la distribución normal, se le denomina **plasticúrtica**.



Together for Tomorrow!
Enabling People

Tipos de funciones de probabilidad

#TecnologíaConPropósito

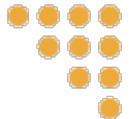


1. Introducción

Un modelo de distribución de probabilidad puede ser elaborado mediante la adquisición de datos ya disponibles o, de ser un experimento aleatorio determinado, mediante la realización de éste un gran número de veces. La limitación es que, en ambos casos, es necesario una gran cantidad de datos; cuantos más datos o resultados se tienen disponibles del experimento, menos error en la estimación de la función de distribución del fenómeno.

Como solución a esto, es posible realizar aproximaciones a modelos construidos a priori y bien estudiados, de tal manera que, por similitudes en la distribución de probabilidades del propio fenómeno con la distribución del modelo propuesto, definido por su función de cuantía o función de densidad de probabilidad (*Probability Density Function, PDF*), según sea variable discreta o continua, sea posible estudiar el fenómeno sin la necesidad de la recolección de datos reales anteriormente mencionado; todo esto a costa, por supuesto, del error que conlleva la simplificación del fenómeno a un modelo predefinido.

Esta unidad está dedicada a analizar los más frecuentes modelos de distribución de probabilidad subyacentes en los fenómenos aleatorios, comenzando con los que impliquen una variable aleatoria discreta y siguiendo por aquellos donde la variable sea continua.



2. Distribuciones discretas

A continuación, se desarrollan los principales modelos estándar que involucran variables aleatorias discretas. Comenzando por las más básicas, cada una de ellas irá introducida por la definición que la identifica mediante casos generales aplicables.

2.1. Distribución degenerada

Esta distribución está modelada para los casos en los que la variable aleatoria ξ concentra toda su masa de probabilidad en un único punto c . Se dice entonces que tiene una *distribución degenerada en c* . En este caso se cumple que la función de cuantía es

$$P(\xi = x) = \begin{cases} 1, & x = c \\ 0, & x \neq c \end{cases}$$

Su función de distribución es

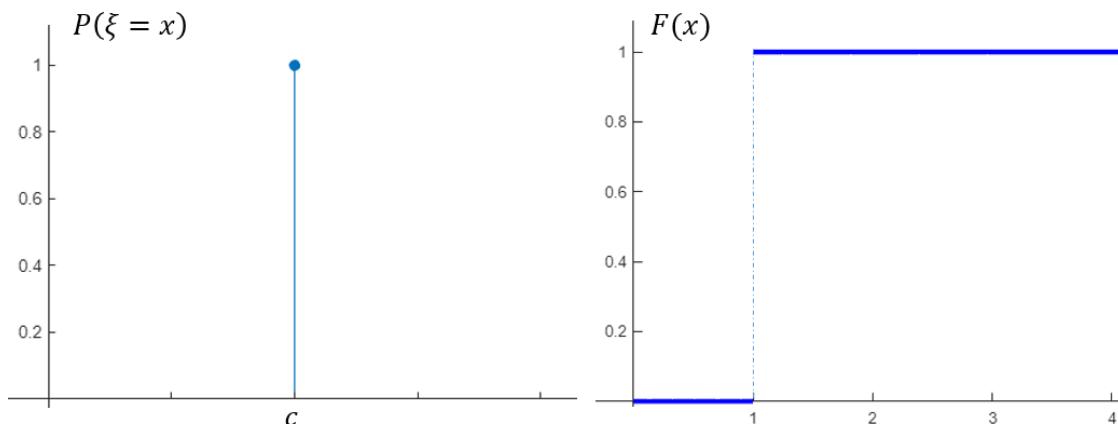
$$F(x) = \begin{cases} 0, & x < c \\ 1, & x \geq c \end{cases}$$

y sus momentos

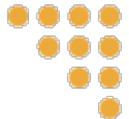
$$\alpha_r = c^r, \quad \mu_r = 0, \quad \forall r \in \mathbb{R}$$

de lo cual se destaca

$$E(\xi) = c, \quad \sigma^2 = 0$$



Esta distribución es la más básica y, a la vez, particular de todas ya que trata los fenómenos cuyo suceso $\xi = c$, independientemente del espacio muestral, siempre resultará vencedor. Si se imagina un fenómeno aleatorio de este tipo repetido n veces, las n observaciones corresponderían con el valor c , por lo que es fácilmente deducible que la media sea siempre c y la varianza igual a cero, ya que no habría dispersión en los datos.



2. Distribuciones discretas

2.2. Distribución uniforme discreta

El anterior caso en el que toda la probabilidad caía sobre un solo valor se puede generalizar para el caso en el que la probabilidad está perfectamente repartida en n valores. La distribución que describe el comportamiento de una variable aleatoria discreta que puede tomar un número finito de valores con la misma probabilidad se denomina **distribución uniforme discreta**. Esto es debido a que la probabilidad total se reparte *uniformemente* a lo largo de todos los elementos del espacio muestral, por lo que su **función de cuantía** queda definida como

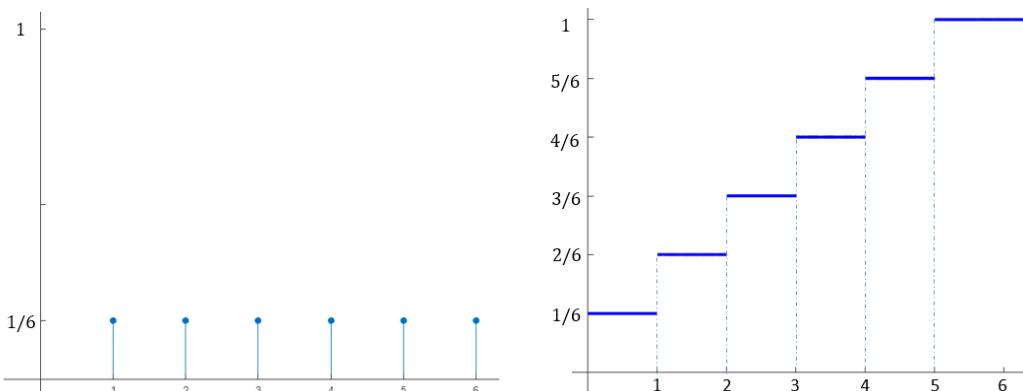
$$p_i = P(\xi = x_i) = \frac{1}{n}, \quad i = 1, 2, \dots, n$$

Su **función de distribución** es

$$F(x_k) = P(\xi \leq x_k) = \sum_{i=1}^k \frac{1}{n} = \frac{k}{n}$$

La **esperanza matemática** y **varianza** serán

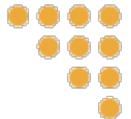
$$E(\xi) = \sum_i x_i p_i = \frac{1}{n} \sum_i x_i, \quad \sigma^2 = \frac{1}{n} \sum_i (x_i - \mu)^2$$



El caso más representativo que sigue este modelo de distribución es el lanzamiento de un dado perfecto.

Efectivamente, en este caso la variable ξ puede tomar cualquier entero del 1 al 6 con probabilidad $1/6$ en cada valor. Siguiendo el modelo de distribución uniforme discreta, se tiene

$$\begin{aligned} E(\xi) &= \sum_i x_i p_i = \frac{1}{n} \sum_{i=1}^6 i = \frac{1}{6} \cdot 21 = 3,5 \\ \sigma^2 &= \frac{1}{n} \sum_i (x_i - \mu)^2 = \frac{1}{6} \sum_{i=1}^6 (i - 3,5)^2 \\ &= \frac{1}{6} [(1 - 3,5)^2 + (2 - 3,5)^2 + \dots + (6 - 3,5)^2] = 2,9167 \end{aligned}$$



2. Distribuciones discretas

2.3. Distribución de Bernoulli o binomial (1, p)

Se tiene una distribución binomial (1, p) en aquellos casos que presentan dos alternativas y éstas son complementarias. Es decir, el experimento está formado por un espacio universal o espacio muestral compuesto por dos sucesos disjuntos que forman el espacio completo. Si se binarizan estos sucesos complementarios S y S^* con los valores 1 y 0, se tiene

$$P(\xi \in S) = P(\xi = 1) = p$$

$$P(\xi \in S^*) = P(\xi = 0) = q = 1 - p$$

En términos generales, su **función de cuantía** se expresa como

$$P(\xi = x) = p^x \cdot q^{1-x}, \quad x \in \{0,1\}$$

y la **función de distribución**

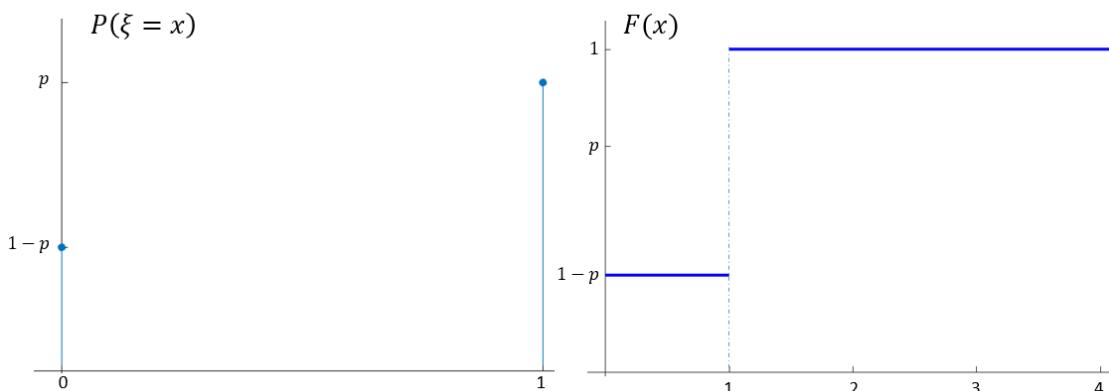
$$F(x) \begin{cases} 0, & x < 0 \\ q, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$$

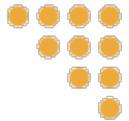
Su **esperanza matemática** y **varianza** vienen definidas por

$$E(\xi) = \sum_{i=1}^2 x_i p_i = 0 \cdot (1-p) + 1 \cdot p = p$$

$$\sigma^2 = E(\xi - \mu)^2 = E(\xi^2) - [E(\xi)]^2 = \sum_{i=1}^2 x_i^2 p_i - p^2 = p(1-p)$$

Estos tipos de fenómenos son denominados *procesos dicotómicos* o de éxito-fraile, puesto solo existen dos sucesos. Cualquier caso que se pueda relacionar con acertar/no acertar, se ajusta a esta distribución.





2. Distribuciones discretas

Estudios han demostrado que 3 de cada 20 000 ciudadanos se contagian con COVID19. Esto es una probabilidad de

$$p = \frac{3}{20\,000} = 0,00015$$

Trasladado a la distribución de Bernoulli, se tienen los dos sucesos

$$S \equiv \{\text{contagiarse}\} = 1, \quad S^* \equiv \{\text{no contagiarse}\} = 0$$

Su esperanza matemática y varianza resultan

$$E(\xi) = p = 0,00015, \quad \sigma^2 = p(1 - p) = 0,0001499775$$

2.4. Distribución binomial (n, p)

Se conoce como distribución binomial (n, p) aquella que describe a conjuntos de n experimentos compuestos de tipo binomial ($1, p$).

Es decir, si el lanzamiento de una moneda perfecta podría asociarse a una distribución de Bernoulli o binomial ($1, p$) con $p = 0,5$, para analizar la secuencia completa de 20 lanzamientos de la misma moneda, se haría uso de la distribución binomial ($20, 0,5$).

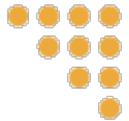
De hecho, el tipo de experimento podría cambiar entre los 20 sucesos, la condición que debe cumplirse es que las variables aleatorias de cada experimento sean independientes y sigan la misma distribución binomial ($1, p$); por ejemplo, podrían alternarse experimentos de lanzamiento de moneda con los de elegir una carta de entre dos disponibles.

Los valores que vayan tomando cada uno de los subexperimentos formarán el valor final de la variable aleatoria:

$$\xi = \xi_1 + \xi_2 + \dots + \xi_n$$

Siguiendo la nomenclatura binaria de los dos posibles valores $\{1, 0\}$, el fenómeno puede tomar dos valores extremos: si cada una de las variables ξ_i toma el valor cero, la variable ξ también será cero; por el contrario, si cada variable ξ_i es igual a uno, la variable ξ tomará su máximo igual a n .

En definitiva, la variable ξ podrá tomar cualquier valor entero entre 0 y n , y su valor indicará el número de veces que el suceso S ocurrió.



2. Distribuciones discretas

La probabilidad, entonces, de que se obtengan x veces el suceso S , dado que cada suceso es independiente, se obtiene a partir de las propiedades de sucesos independientes

$$P(S_1 \cap S_2 \cap \dots \cap S_n) = P(S_1) \cdot P(S_2) \cdot \dots \cdot P(S_n)$$

$$\text{si } P(S) = p \text{ y } P(\bar{S}) = q \rightarrow$$

$$P(S_1 \cap S_2 \cap \dots \cap S_x \cap S_{x+1}^* \cap \dots \cap S_n^*) = \prod_{i=1}^x p_i * \prod_{j=x+1}^n q_j = p^x \cdot (1-p)^{n-x}$$

Esta sería la probabilidad de que la secuencia mencionada arriba se cumpliese en ese orden, pero no es necesario que se dé ese orden, tan solo que sean x las veces que se cumple S , por lo que habrá que calcular el número de posibles combinaciones con los mismos números disponibles repetidos cambiando su disposición, por lo que se está ante una *permutación con repetición*

$$P_n^{x,n-x} = \frac{n!}{x!(n-x)!} = \binom{n}{x}$$

Entonces, la **función de cuantía** resulta

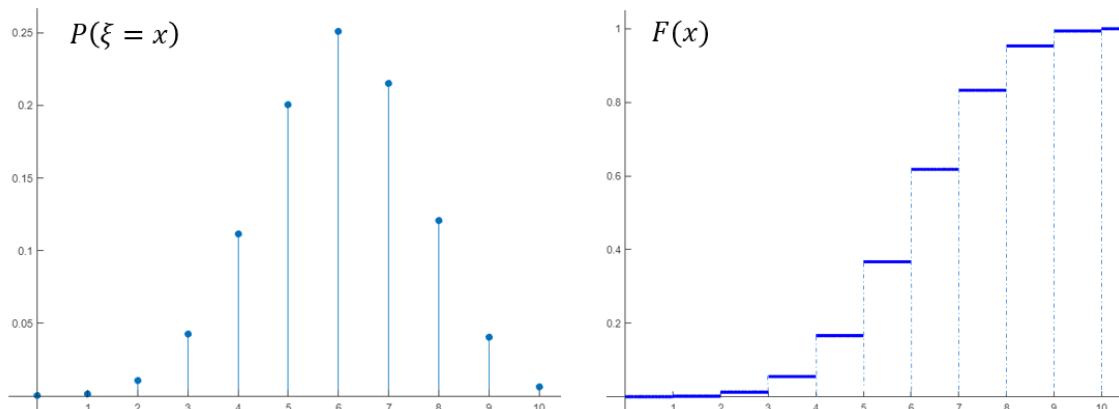
$$P(\xi = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

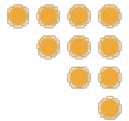
A modo de prueba, el lector podría comprobar que se cumple la condición necesaria de la probabilidad $\sum P(\xi = x) = 1$.

Su **esperanza matemática** y **varianza** son

$$E(\xi) = E(\xi_1 + \dots + \xi_n) = E(\xi_1) + \dots + E(\xi_n) = p + p + \dots + p = np$$

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2 = p(1-p) + p(1-p) + \dots + p(1-p) = np(1-p)$$





2. Distribuciones discretas

Si ahora el estudio anterior se amplía a 10 ciudadanos elegidos de puntos geográficos lo más dispersos posible, de manera que se obtenga información global, pero también para poder considerar las variables individuales como independientes. Si la probabilidad de contagiarse era de $p = 0,00015$, su esperanza matemática y varianza resultan

$$E(\xi) = np = 10 \cdot E(\xi_i) = 0,0015$$
$$\sigma^2 = np(1 - p) = 10 \cdot \sigma_i^2 = 0,001499775$$

2.3. Distribución de Poisson

Esta distribución fue desarrollada como una extensión de la distribución binomial (n, p) cuando $n \rightarrow \infty$ y, por lo tanto, $p \rightarrow 0$. Es por eso que esta distribución también es conocida como distribución *de los sucesos raros*. Básicamente, se aplica a casos prácticos en los que la variable aleatoria ξ puede tomar cualquier valor natural, junto al cero, y se conoce su esperanza matemática a priori.

La **función de cuantía** de esta distribución es

$$p_i = P(\xi = x) = \frac{e^{-\lambda} \lambda^x}{x!},$$

donde

$$E(\xi) = \sigma^2 = \lambda$$

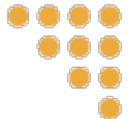
Sobre el volumen de ventas de móviles de una tienda, tan solo se conoce el número medio de móviles vendidos, que es de 8 diarios. Con esa información, se pretende conocer la probabilidad de que se vendan 5 en un día.

La información disponible es el número medio, es decir,

$$E(\xi) = \lambda = 8$$

Se está pidiendo, en este caso, $P(\xi = 5)$.

$$P(\xi = 5) = \frac{e^{-\lambda} \lambda^x}{x!} = \frac{e^{-8} 8^5}{5!} = 0,0916$$



3. Distribuciones continuas

En este apartado se verán con detalle las principales distribuciones de probabilidad continuas, que no son más que aquellas cuya función de distribución es continua. Es decir, se pasará de definir funciones de cuantía a función de densidad, de n posibles valores a intervalos.

De entre las que se describirán, se profundizará en la que es representada con la que puede ser la curva más representativa, incluso icónica, de la estadística, la campana de Gauss.

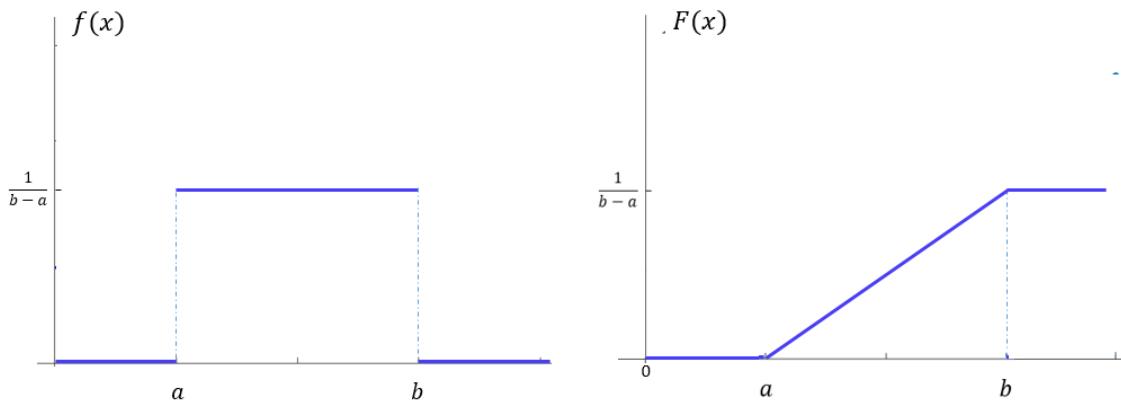
3.1. Distribución uniforme

Una variable aleatoria ξ sigue una distribución uniforme en $[a, b]$, también nombrada como *rectangular* y denotada por $U(a, b)$, si su **PDF** es

$$f(x) = \frac{1}{b-a}, \quad a \leq x \leq b$$

Y la **función de distribución**

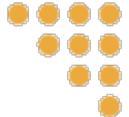
$$F(x) = P(\xi \leq x) = \int_a^x f(x) dx = \frac{1}{b-a} x \Big|_a^x = \frac{x-a}{b-a}$$



Tal y como ocurría en su homónima discreta, la probabilidad del suceso se encuentra repartida de manera uniforme a lo largo del intervalo $[a, b]$, por lo que depende exclusivamente de la amplitud del intervalo estudiado y no de su posición.

La **esperanza matemática** y la **varianza** de esta variable vienen dadas por

$$E(\xi) = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2}, \quad \sigma^2 = \frac{(b-a)^2}{12}$$



3. Distribuciones continuas

Tras el tratamiento térmico a una pieza en una fábrica, se sabe que la longitud final sigue una distribución uniforme entre 12 489,69 cm y 12 490,07 cm. Se quiere conocer la media y la varianza de la longitud de las piezas, así como el número de piezas que saldrán mayores de 12 490 cm de una partida de 100. También se requiere localizar el cuartil Q_3 de la distribución.

En primer lugar, sabiendo que se tiene una distribución uniforme en [12489,69, 12490,07], la media y la varianza vendrán dadas por

$$E(\xi) = \frac{a + b}{2} = 12489,88 \text{ cm}, \quad \sigma^2 = \frac{(b - a)^2}{12} = 0,012 \text{ cm}$$

Conocer el número de piezas que superará una medida determinada de entre un monto total significa calcular la probabilidad de que una de las 100 piezas lo cumpla. Si usamos $t \equiv 12490$ y $N = 100$

$$\frac{n_{\xi>t}}{N} = P(\xi > 12490) = 1 - P(\xi \leq 12490) = 1 - F(12490)$$

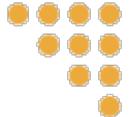
Por lo que hay que calcular la función de distribución a partir de la función de densidad

$$F(x) = \frac{x - a}{b - a} = \frac{x - 12489,69}{0,38}$$

$$P(\xi > 12490) = 1 - F(12490) = 1 - \frac{12490 - 12489,69}{0,38} = 0,184$$

Nótese que, al estar uniformemente distribuido, el porcentaje de piezas que caerá en un intervalo correponde con el porcentaje que ese intervalo supone al intervalo total $[a, b]$. Por esta misma relación, se podría automáticamente deducir que el cuartil Q_3 , que marca la partición del 75 %, es decir, el valor que solo es superado por el 25% de las piezas, será

$$Q_3 = \frac{75}{100} = \frac{x - a}{b - a} \rightarrow x = 0,75(b - a) + a = 12489,957 \text{ cm}$$



3. Distribuciones continuas

3.2. Distribución normal

Esta es, sin lugar a dudas, la distribución más importante de la estadística, pues es la distribución límite de multitud de sucesos, ya sean continuos o discretos.

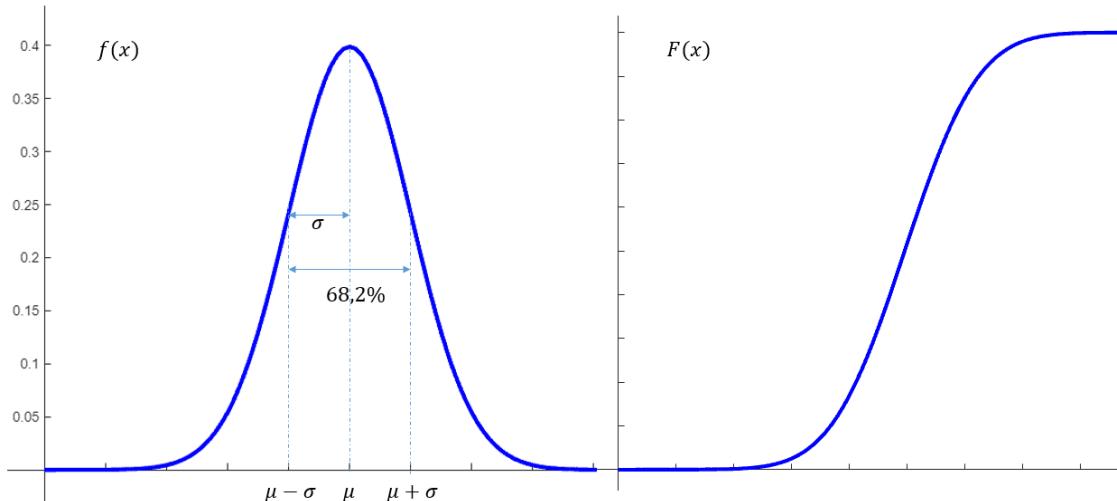
3.2.1. Distribución $N(0, 1)$

Las variables aleatorias que siguen este tipo de distribución pueden tomar cualquier valor real y su **PDF** queda definido por

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

y su **función de distribución** por

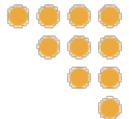
$$F(x) = P(\xi \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{x^2}{2}} dx$$



La propia nomenclatura $N(0, 1)$ define la esperanza y la varianza, entre paréntesis. Así:

$$E(\xi) = 0, \quad \sigma^2 = 1$$

La función de densidad describe la campana de Gauss, una curva simétrica con respecto a la media, por lo que todos sus momentos de orden impar serán nulos, y cuyo intervalo $[\mu - \sigma, \mu + \sigma]$ encierra el 68,1% de los datos.



3. Distribuciones continuas

3.2.2. Distribución $N(\mu, \sigma^2)$

Es la generalización del caso anterior para cualquier media y varianza. En este caso, su **función de densidad** queda expresado como

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

La función de distribución

$$F(x) = P(\xi \leq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

La esperanza matemática y la varianza, en estos casos, debe venir definida en la propia declaración de la distribución.

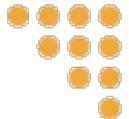
Con respecto a esta, la distribución $N(0,1)$ toma el nombre de *distribución normal tipificada o estándar*.

3.1.3. Propiedades de la distribución normal

1. Tanto la moda, Mo , como la mediana, Me , coinciden con la media μ .
2. Siendo una distribución simétrica, se cumple que $F(-x) = 1 - F(x)$
3. El intervalo $[\mu - \sigma, \mu + \sigma]$ encierra el 68,2% de la distribución; el intervalo $[\mu - 2\sigma, \mu + 2\sigma]$, el 95,4 %; el intervalo $[\mu - 3\sigma, \mu + 3\sigma]$ encierra el 99,74 %.
4. Sean a y b dos números reales y la variable aleatoria ξ , la cual sigue una distribución normal $N(\mu, \sigma^2)$, entonces cualquier otra variable de tipo $a * \xi + b$ sigue también una distribución normal $N(a\mu + b, \sqrt{a^2\sigma^2})$
5. Si dos variables independientes ξ_1 y ξ_2 siguen la distribución normal, de tal modo que $N(\mu_1, \sigma^2_1)$ y $N(\mu_2, \sigma^2_2)$, entonces la variable $\xi = \xi_1 + \xi_2$ seguirá una distribución $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

3.1.4. Aproximación a $Z \sim N(0, 1)$

A la hora de calcular probabilidades y valores para casos que siguen distribuciones normales $N(\mu, \sigma^2)$, gracias a la propiedad 3 es posible transformar esta distribución a una del tipo $N(0,1)$ para la simplificación en sus cálculos.



3. Distribuciones continuas

Mediante la transformada $\eta = \mu + \sigma Z$, siendo $Z \sim N(0,1)$ y $\eta \sim N(\mu, \sigma^2)$

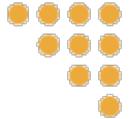
$$P(\eta \leq y) = P(N(\mu, \sigma^2) \leq y) = P(\mu + \sigma Z \leq y) = P\left(Z \leq \frac{y - \mu}{\sigma}\right)$$

Por este motivo la distribución $N(0,1)$ es llamada tipificada o estandarizada: para cualquier distribución normal de tipo $N(\mu, \sigma^2)$, se realiza la transformación anterior y se consulta la *tabla de la normal estándar* o **tabla Z**.

Desv. normal x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010

La siguiente tabla indica las probabilidades $P(Z > x)$, donde x se busca en la primera columna con una precisión de 0,1 y se refina hasta una precisión de 0,01 mediante la primera fila.

Es decir, la probabilidad $P(Z > 2,34)$ sería igual a 0,0096, extraído de la fila 2,3 y la columna 0,04.



3. Distribuciones continuas

En Biología estadística se estudia la reproducción celular, la cual está distribuida normalmente con un tiempo medio de 60 minutos y una desviación estándar de 5 minutos. Con esta información se pretende calcular la probabilidad de que esta reproducción o mitosis ocurra pasados los 65 minutos. Asimismo, se necesita saber cuánto tardará el último 1 % de las células, dada un conjunto de células.

Asignando a $\xi \equiv \text{tiempo}$, variable aleatoria que sigue una distribución de tipo $N(60,5)$, se necesita obtener, en primer lugar, $P(\xi > 65)$ y después el valor de la variable para $P(\xi > x) = 0,01$

Transformando la variable aleatoria para aproximarla a la $N(0,1)$ se tiene

$$P(\xi > 65) \rightarrow P\left(Z > \frac{65 - 60}{5}\right) = P(Z > 1)$$

En la tabla Z, aparece que para esa variable se tiene 0,1587, por lo que

$$P(\xi > 65) = 0,1587$$

Después, abordando la segunda cuestión $P(\xi > x) = 0,01$ se tiene

$$P(\xi > x) = 0,01 \rightarrow P\left(Z > \frac{x - 60}{5}\right) = 0,01$$

En este caso, se busca la variable, no la probabilidad por lo que buscamos el valor que más se asemeja a 0,01. Aparece el valor 0,0099 en la fila marcada con 2,3 y la columna marcada con 0,03. Por lo tanto

$$P\left(Z > \frac{x - 60}{5}\right) 0,01 \rightarrow Z = 2,33$$

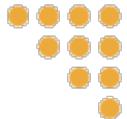
$$\rightarrow x = 2,33 * 5 + 60 = 71,65 \text{ minutos}$$

El 15,87 % de las células tardarán más de 65 minutos y, pasados 71,65 minutos, solo el 1 % permanecerá aún sin haberse dividido o reproducido.

Together for Tomorrow!
Enabling People

Inferencia estadística

#TecnologíaConPropósito



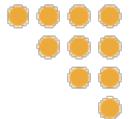
1. Introducción

La estadística descriptiva lo conforma el conjunto de métodos, herramientas visuales e indicadores y medidores que permiten extraer conocimiento sobre los datos estudiados. Por su parte, la teoría de la probabilidad provee al estadista de las herramientas necesarias para contabilizar posibilidades que tienen ciertos eventos determinados de ocurrir, reduciendo así la incertidumbre sobre estos sucesos de interés. La unión de estas dos disciplinas crea la base de una tercera, a partir de la cual es posible realizar estimaciones o hipótesis acerca de sucesos futuros o desconocidos, que involucran datos no disponibles o no incluidos en los datos analizados, e incluso dar un valor numérico a la seguridad con la que se está afirmar esa hipótesis.

Es posible decir que mediante la estadística inferencial se obtiene información sobre un conjunto poblacional a partir de una parte de ésta llamada muestra por medio de inducción para, posteriormente, obtener conclusiones de valor y tomar decisiones útiles a partir de las predicciones, estimaciones o hipótesis contrastadas llevadas a cabo por deducción aplicada a observaciones futuras o desconocidas.

La unidad se desarrollará a lo largo de los dos elementos en los que se divide la inferencia estadística. Estos son la estimación y el contraste de hipótesis.

La idea clave es, a partir de la aplicación de la teoría de la probabilidad a los parámetros estadísticos extraídos de una muestra representativa y aleatoria, realizar estimaciones o contrastes poblacionales.



2. Tipos de inferencia

Las aplicaciones que tiene la **inferencia estadística** son:

- La estimación.
- El contraste de hipótesis.

2.1. La estimación

La **estimación** es el proceso de aproximar un valor sobre algo, dar una respuesta aproximada acerca de un fenómeno que será usada para algún propósito determinado.

Las estimaciones se realizan cuando los datos de entradas no dan la información suficiente como para obtener un modelo exacto a partir del cual calcular el valor buscado, o estos datos de entrada están incompletos, son inciertos o inestables. En estos casos una estimación es el valor más fidedigno encontrado para dar respuesta al problema, pues proviene de la mejor información posible.

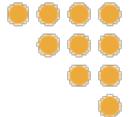
Hay dos tipos de estimación de parámetros:

- La **puntual**. Consiste en encontrar una buena aproximación, a un valor buscado del parámetro en cuestión, basándose en los datos de la muestra extraída.
- El **intervalo de confianza**. Consiste en obtener, a partir de un requisito inicial en forma de probabilidad mínima llamado **nivel de confianza**, un intervalo que encierre a todos los posibles valores que puede adoptar el parámetro con la probabilidad requerida. Es decir, proporciona una solución de tipo “existe una probabilidad de x de que el valor buscado estará en el intervalo (a, b) ”.

Además, la precisión de una estimación depende de la cantidad de elementos que integran la muestra (el tamaño de la muestra), que a su vez depende de múltiples factores como el dinero y el tiempo disponibles para el estudio, la importancia del tema analizado, la confiabilidad que se espera de los resultados, las características propias del fenómeno analizado, etcétera.

Así, a partir de la muestra seleccionada se realizan algunos cálculos y se estima el valor de los parámetros de la población tales como la media, la varianza, la desviación estándar, o la forma de la distribución...

El estudio muestral no es un tema que entre a formar parte de este tema, pero si necesitaremos una serie de conceptos necesarios para el desarrollo del tema.



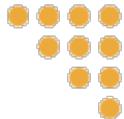
2. Tipos de inferencia

2.2. El contraste de hipótesis

En la misma dirección, pero sentido contrario, se encuentra el otro método inferencial, el **contraste de hipótesis**.

En lugar de averiguar qué valor tomará determinado parámetro o qué valores es necesario tener en cuenta como posibles candidatos cuando se requiere una mínima certidumbre al respecto, se propone una hipótesis, la cual debe ser contrastada, mediante la asignación de una probabilidad, que permitirá aceptar o rechazar dicha hipótesis.

- Consultor: *Tras un exhaustivo análisis de la situación de la empresa junto a los históricos de años anteriores, podemos afirmar, con un 99% de confianza en acertar, que las pérdidas este año estarán entre los 6000 y 13500 euros.*
- CFO: *¿Dices que superarán los 10 000 euros?!*
- Consultor: *No podemos rechazar esa hipótesis.*



3. Muestreo

Se denomina **muestreo** al proceso de elegir una porción de una población, mediante el método correspondiente, para conseguir una muestra aleatoria y representativa.

Las muestras deben cumplir ciertas condiciones para asegurar su máxima similitud con respecto a la población total.

Si la muestra se elige sin tener en cuenta esto, se obtienen errores al realizar inferencia; estos errores se denominan **sesgo** y se encuentran en una **muestra sesgada**.

3.1. Técnicas de muestreo

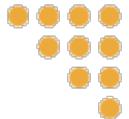
Para conseguir esto existen diversas técnicas de muestreo:

- **Muestreo no probabilístico.** La muestra no es elegida aleatoriamente sino mediante determinados criterios subjetivos.
- **Muestreo probabilístico.** En caso de elegirse de manera aleatoria es posible distinguir entre los siguientes:
 - **Muestreo aleatorio simple.** Cada individuo de la población tiene igualdad de oportunidades para ser escogido en la muestra. Es el más empleado.
 - **Muestreo sistemático.** Solo el primero es escogido al azar y, a partir de él, se eligen los demás hasta completar la muestra mediante saltos en intervalos de tamaño definido.
 - **Muestreo estratificado.** Existe una primera clasificación de los datos que los divide en estratos.

La muestra se forma con un número de individuos de cada grupo formado proporcional al tamaño de cada grupo.

- **Muestreo por conglomerados.** En esta técnica se crean conglomerados, clasificación de la población según algún parámetro que no sea de interés para el estudio, de tal manera que siga siendo un grupo heterogéneo y representativo.

Por ejemplo, si se estudiase alguna característica en la población española totalmente independiente de su localización, los conglomerados podrían ser ciudades o provincias del país y extraer directamente así una muestra de una ciudad determinada.



3. Muestreo

3.2. Distribuciones muestrales

A pesar de que pueda interpretarse como contradictorio, las muestras deben llevar un tratamiento ligeramente modificado al propio de la población. Es por ello que se hace necesario conocer las distribuciones y estadísticos muestrales, es decir, las variables aleatorias asociadas al muestreo, pues están basados en el comportamiento de las propias muestras. Obviamente, la distribución muestral que siga una muestra dependerá, en primer lugar, de la variable aleatoria bajo análisis, pero también del tamaño de la muestra.

A continuación, se muestran las fórmulas de los principales parámetros estadísticos cuando se habla de muestras:

- **Media muestral**, definida por \bar{x} , se supone bien aproximada a la media poblacional o esperanza matemática cuando la muestra está bien escogida y tiene un tamaño suficientemente grande. Se expresa como

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i,$$

y se distribuye siguiendo la normal

$$\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

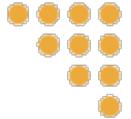
- **Cuasivarianza**. Si para estimar la media poblacional se parte de la media de la muestra, la media muestral, puede parecer lógico usar la varianza de la muestra extraída para estimar la varianza de la población. Cuando se extraen muestras y se comparan con los estadísticos de éstos con los de la población, se descubre un sesgo entre ambos que desaparece mediante el uso de la cuasivarianza. La única diferencia entre cuasivarianza y varianza es que la primera divide las diferencias cuadráticas por $n - 1$. Se denota por s^2

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2,$$

y sigue una distribución de Pearson χ^2 , la cual deriva de la normal y puede consultarse en las tablas de Pearson.

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

La distribución χ^2 se podrá aproximar a una normal en muestras de un tamaño mínimo, $n > 30$.



4. Estimaciones

A continuación, se define la estimación y se desarrollan los dos tipos de estimaciones posibles, profundizando en detalle en las metodologías usadas en cada caso.

4.1. Estimación puntual

La estimación puntual se lleva a cabo cuando, mediante una función que explota los datos obtenidos de una muestra, se consigue aproximar el valor de un parámetro buscado asociado a la población que la muestra representa.

Esta aproximación es llamada **estimador** y se denota por el símbolo $\hat{\theta}$ colocado sobre el símbolo del parámetro a estimar. Así, por ejemplo la media muestral puede ser considerado como el estimador de la media de la población

Definición. Sea X_1, X_2, \dots, X_n una muestra aleatoria simple de tamaño n , $\hat{\theta}$ es el estimador de θ si el estadístico que se emplea para conocer dicho parámetro desconocido es ese.

4.1.1. Propiedades

Podrían ser considerados condiciones o principios de selección, también, pues son necesarios para seleccionar una muestra de calidad:

- **Insesgadez.** La condición para un estimador $\hat{\theta}$ de θ no sea un estimador sesgado, es que la esperanza matemática del estimador coincida con el valor del parámetro. Esto es:

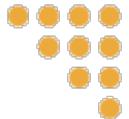
$$E(\hat{\theta}) = \theta$$

- **Eficiencia.** Sean dos estimadores $\hat{\theta}_1$ y $\hat{\theta}_2$ del mismo parámetro θ , $\hat{\theta}_1$ es el más eficiente cuando su varianza es menor.

$$V[\hat{\theta}_1] < V[\hat{\theta}_2]$$

- **Suficiencia.** Un estimador $\hat{\theta}$ es suficiente cuando ha sido calculado a partir de toda la información disponible de la muestra.
- **Consistencia.** El estimador $\hat{\theta}$ es consistente si converge en probabilidad al valor verdadero de θ cuando la muestra tiende a infinito.

$$\lim_{n \rightarrow \infty} V[\hat{\theta}] = 0$$



4. Estimaciones

Existen tres métodos, considerados como los más importantes, para realizar estimaciones puntuales:

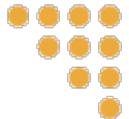
- **Método de los momentos.** Este método se basa en que, dada una muestra de una variable aleatoria ξ que sigue una distribución dada por $f(x, \theta)$, la media muestral \bar{x} debe ser similar a la media poblacional $E_\theta(\xi)$ y ésta es función de θ . Por lo tanto, se obtiene de despejar θ en la igualdad $E_\theta(X) = \bar{x}$.
- **Método de mínimos cuadrados.** Transforma el problema de estimación en uno de optimización. Encuentra el estimador como aquel que minimiza una función de coste que se rige en función del error cometido. Supone el método en el que se basa la regresión lineal.
- **Método de máxima verosimilitud.** Es el más empleado, elige como estimador aquel valor que sea más probable. Obtiene gran consistencia y eficiencia.

4.2. Estimación por intervalos de confianza

De este tipo de estimación se obtiene un margen de variabilidad que se le da al valor estimado y así afirmar que, bajo una determinada probabilidad (o confianza), el valor verdadero no quedará fuera de sus límites.

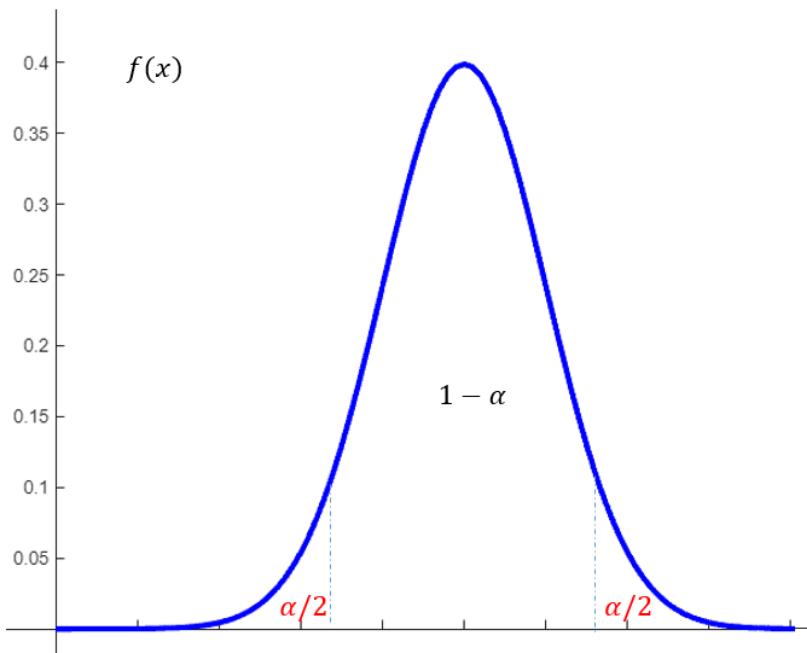
En procesos de estimación por intervalos de confianza aparecen nuevos parámetros que son necesarios conocer:

- El **error de la estimación** es una medida de la precisión de la estimación y no es más que la amplitud del intervalo de confianza obtenido. Así, si se proporciona $\theta_1 < \theta < \theta_2$, se tiene el **intervalo de confianza** $[\theta_1, \theta_2]$, con un **error de estimación** $\epsilon = \theta_2 - \theta_1$.
- El **nivel de confianza** define, habitualmente expresado en niveles porcentuales, la probabilidad de que el verdadero valor del parámetro se encuentre dentro del intervalo de confianza. Se denota por $(1 - \alpha)$. Son habituales los valores 95% y 99%. α representa el **nivel de significación** y cuantifica la **probabilidad de fallo en la estimación dada**. Obviamente supone la probabilidad del suceso complementario a acertar la estimación.
- El llamado **valor crítico**, $Z_{\alpha/2}$, es el valor de la probabilidad que el intervalo de confianza deja a cada lado de una curva, aproximando la estimación a la normal. Es decir, el área bajo la curva de una distribución estándar que deja a



4. Estimaciones

su derecha un área igual a $\alpha/2$. Así, si la distribución que sigue la población de la variable cuyo parámetro se quiere estimar sigue una normal $N(0,1)$, para obtener el valor crítico de un intervalo de confianza de 95%, se debería buscar en la tabla el valor más cercano a $\alpha/2 = 0,025$, y localizar a qué valor de Z se le asocia, en este caso sería $Z_{\alpha/2} = 1,96$



Los intervalos de confianza pueden ser:

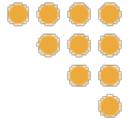
- **Unilaterales.** Un intervalo de confianza unilateral estima un límite, a modo de umbral, inferior o superior, a partir del cual el valor verdadero se sitúa. Es decir, proporciona un intervalo infinito y serán estimaciones del tipo *mayor que* o *menor que*.
- **Bilaterales.** El intervalo de confianza bilateral proporciona un intervalo definido en ambos extremos.

Se extrae la muestra de cierta población y se calcula estima que la media es 7 con un error en la estimación de 0,3 y un nivel de significación 0,01.

De aquí se deduce que el intervalo de confianza estará en torno al valor de la estimación con una amplitud de 0,3, es decir, $\pm 0,15$. Entonces [6,85, 7,15].

Por otro lado, el nivel de confianza será del 99%, ya que

$$1 - \alpha = 0,01 \rightarrow \alpha = 0,99$$



4. Estimaciones

4.2.1. Intervalo de confianza para la media, conocida la varianza

Para una variable aleatoria ξ que sigue una distribución normal, $\xi \sim N(\mu, \sigma)$, a partir de la media muestral \bar{x} como estimador, se puede aproximar:

$$\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Estandarizando la distribución, es decir, centrandola y aproximandola a una normal, se tiene la transformación:

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Así, el intervalo de confianza, con nivel de confianza $1 - \alpha$ para el estimador de la media con varianza conocida σ , viene dado por:

$$(1 - \alpha) = P\left[-Z_{\frac{\alpha}{2}} < \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} < Z_{\frac{\alpha}{2}}\right] = P\left[-Z_{\frac{\alpha}{2}} \cdot \frac{\sqrt{n}}{\sigma} - \bar{x} < -\mu < Z_{\frac{\alpha}{2}} \cdot \frac{\sqrt{n}}{\sigma} - \bar{x}\right]$$

cambiando todos los signos para tener la media positiva, también, se tiene

$$(1 - \alpha) = P\left[\bar{x} + Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} > \mu > \bar{x} - Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

El intervalo resulta:

$$\left[\bar{x} - Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \quad \bar{x} + Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

4.2.2. Intervalo de confianza para la media, desconocida la varianza con $n > 30$

Ya se ha explicado que, para casos de medias de tamaño importante, esto es, $n > 30$, cualquier distribución puede aproximarse a la normal.

Y puesto que no se tiene la varianza, será necesario calcular la cuasivarianza de la muestra. Así,

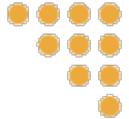
$$\bar{x} \sim N\left(\mu, \frac{s}{\sqrt{n}}\right)$$

Transformando para trabajar sobre la normal centrada

$$Z = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \sim N(0,1)$$

Al poder aproximar la distribución a la normal, debido al tamaño de la muestra, el problema toma la misma forma que el anterior, por lo que el intervalo será, entonces:

$$\left[\bar{x} - Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}, \quad \bar{x} + Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}\right]$$



4. Estimaciones

4.2.3. Intervalo de confianza para la media, desconocida la varianza con $n < 30$

En este caso no será posible aproximar la distribución del estimador a la normal y habrá que asumir que la variable aleatoria se distribuye como una *t-Student* con $n - 1$ grados de libertad, de la forma:

$$\frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}$$

Del mismo modo que en el caso de distribuciones de Pearson, la distribución *t-Student* deriva de la normal y puede ser consultadas en las tablas homónimas.

De la misma forma que anteriormente, se obtiene el intervalo:

$$\left[\bar{x} - t_{n-1; \frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}, \quad \bar{x} + t_{n-1; \frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \right]$$

Se quiere realizar un estudio personalizado a los usuarios de una conocida web de nutrición que se han inscrito en un curso. Para ello, se quiere conocer el porcentaje de grasa de los usuarios, a los que se les hace rellenar un formulario donde pueden incluir este dato; sin embargo, no todo el mundo tiene esa información o las herramientas para medirse el porcentaje de grasa. Consigue tan solo las medidas de 80 inscritos y, de estas, se obtuvo que el porcentaje promedio era $\bar{x} = 25,6\%$ y la desviación estándar muestral $\bar{\sigma} = 1,896\%$. Se quiere encontrar un rango de posibles valores que se puedan asumir como media del porcentaje de grasa del curso con una seguridad o confianza del 95%.

Se tiene, entonces, que $\alpha = 0,05$ y $n = 80$. Puesto que se va a trabajar con una muestra, es necesario calcular la cuasivarianza, por lo que se debe hacer la transformada:

$$s = \frac{n}{n-1} \bar{\sigma} = 1,92,$$

ya que la desviación estándar muestral está dividida por el tamaño muestral n y la cuasivarianza se calcula en base a la corrección $n - 1$. Calculado queda:

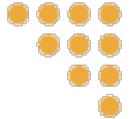
$$\left[\bar{x} - Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \quad \bar{x} + Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \right] = [25,6 - \frac{1,92}{\sqrt{80}} Z_{0,05; \frac{1}{2}}, \quad 25,6 + \frac{1,92}{\sqrt{80}} Z_{0,05; \frac{1}{2}}]$$

Se debe buscar el valor $Z_{0,05; \frac{1}{2}}$, que resulta 1,96, por lo que el intervalo de confianza pedido es:

$$[25,179\%, \quad 26,021\%]$$

Si las medidas obtenidas hubieran sido extraídas de 30 inscritos, como no supera la condición $n > 30$, se debería haber trabajado con la distribución de *t-Student* de $n - 1$ grados de libertad y entonces $t_{n-1; \frac{\alpha}{2}} = 2,0452$, por lo que quedaría:

$$[25,161\%, \quad 26,039\%]$$



4. Estimaciones

4.2.4. Intervalo de confianza para la proporción poblacional

Pasando ahora a poblaciones dicotómicas, en las que un individuo puede cumplir cierto suceso o no, la variable aleatoria asociada sigue la distribución binomial

$$X \sim B(n, p),$$

que, con el tamaño de muestra suficientemente grande, se puede realizar la aproximación a una normal de media igual a la esperanza matemática de la binomial y con la varianza establecida para este tipo de distribuciones.

$$X \sim N(np, \sqrt{np(1-p)})$$

Si el estudio se quiere centrar en una proporción de la población. Si de una muestra se extrae su proporción o su probabilidad muestral, realizando el cambio de variable $\bar{p} = \frac{x}{n}$ y haciendo $q = 1 - p$, ésta \bar{p} se distribuye como

$$\bar{p} \sim N\left(\frac{np}{n}, \frac{\sqrt{npq}}{n}\right) = N\left(p, \sqrt{\frac{pq}{n}}\right)$$

Cuando se tipifica esta distribución queda como

$$Z = \frac{\bar{p} - p}{\sqrt{\frac{pq}{n}}} \sim N(0,1)$$

Y entonces, el intervalo de confianza para el estimador proporción viene dado por

$$\left[\bar{p} - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{pq}{n}}, \quad \bar{p} + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{pq}{n}} \right]$$

De una muestra de 200 personas de un poblado, 40 han dado positivo en una virus que se conoce como fiebre centroafricana. Se necesita, con un nivel de confianza del 99 %, el intervalo de posibles valores para la proporción de habitantes del poblado con el virus contagiado.

Se tiene la proporción muestral $\bar{p} = \frac{40}{200} = 0,2$ y se pide el intervalo para $\alpha = 0,01$. El intervalo para una proporción se obtiene de

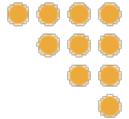
$$\left[\bar{p} - Z_{\frac{0,01}{2}} \cdot \sqrt{\frac{pq}{n}}, \quad \bar{p} + Z_{\frac{0,01}{2}} \cdot \sqrt{\frac{pq}{n}} \right] = \left[0,2 - 0,0283 \cdot Z_{\frac{0,01}{2}}, \quad 0,2 + 0,0283 \cdot Z_{\frac{0,01}{2}} \right]$$

Consultando en la tabla tipificada Z , se obtiene

$$Z_{\frac{0,01}{2}} = Z_{0,005} = 2,575$$

Por lo que el intervalo buscado resulta ser

$$[0,1271, \quad 0,2729]$$



4. Estimaciones

4.2.5. Intervalo de confianza para la varianza

Para el caso de la varianza σ^2 , teniendo la cuasivarianza de la muestra s^2 , éste sigue una distribución:

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

Por lo que se tendría, para un nivel de confianza de $1 - c$,

$$P \left[\chi^2_{n-1;1-\frac{\alpha}{2}} < \frac{(n-1)s^2}{\sigma^2} < \chi^2_{n-1;\frac{\alpha}{2}} \right] = 1 - \alpha$$
$$\rightarrow P \left[\frac{1}{\chi^2_{n-1;1-\frac{\alpha}{2}}} > \frac{\sigma^2}{(n-1)s^2} > \frac{1}{\chi^2_{n-1;\frac{\alpha}{2}}} \right] = P \left[\frac{(n-1)s^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}} > \sigma^2 > \frac{(n-1)s^2}{\chi^2_{n-1;\frac{\alpha}{2}}} \right] = 1 - \alpha$$

El intervalo de confianza buscado queda como

$$\left[\frac{(n-1)s^2}{\chi^2_{n-1;\frac{\alpha}{2}}}, \quad \frac{(n-1)s^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}} \right]$$

De la misma muestra de ese mismo poblado se tienen los siguientes valores sobre la estatura

$$\bar{x} = 175 \text{ cm}, \quad s = 10,1 \text{ cm}$$

Si se necesita calcular un intervalo de confianza con $\alpha = 0,05$ para la varianza, hay que aproximar el valor de la cuasivarianza calculada a la distribución de Pearson de $n - 1$ grados de libertad. Así, se necesita calcular

$$\left[\frac{(n-1)s^2}{\chi^2_{n-1;\frac{\alpha}{2}}}, \quad \frac{(n-1)s^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}} \right] = \left[\frac{20300}{\chi^2_{n-1;\frac{\alpha}{2}}}, \quad \frac{20300}{\chi^2_{n-1;1-\frac{\alpha}{2}}} \right]$$

De la tabla de chi-cuadrado, o Pearson, se obtiene

$$\chi^2_{200;0,0025} = 260,7350, \quad \chi^2_{180;0,0025} = 237,8548$$

$$\chi^2_{200;0,9975} = 148,4262, \quad \chi^2_{180;0,9975} = 131,3050$$

Para encontrar el valor para 199 grados de libertad, se debe obtener asumiendo una variación lineal entre los valores en 200 y 180. Así

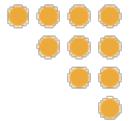
$$\text{si } \frac{199 - 180}{200 - 180} = \frac{\chi^2_{199;0,0025} - \chi^2_{180;0,0025}}{\chi^2_{200;0,0025} - \chi^2_{180;0,0025}} \rightarrow \chi^2_{199;0,0025} = 259,591$$
$$\chi^2_{199;0,9975} = 146,7141$$

y el intervalo buscado es

$$\sigma^2 \in [78,1999, \quad 138,3643]$$

Que, llevado a desviación típica, para compararlo con la cuasivarianza calculada, se tiene

$$\hat{\sigma} \in [8,8431, \quad 11,7628]$$



5. Contraste de hipótesis

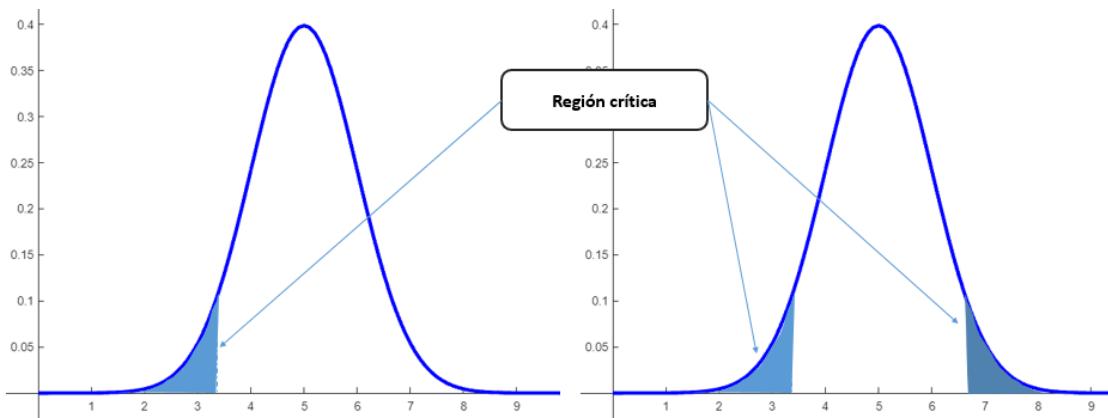
La segunda herramienta fundamental que nos brinda la inferencia estadística para la extracción de conocimiento o conclusiones, o toma de decisiones, es el **contraste de hipótesis**, también llamado *test de hipótesis estadístico*. Es una prueba estadística llevada a cabo sobre una proposición, la hipótesis, acerca de una población determinada que determina si la proposición debe ser aceptada o rechazada.

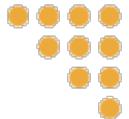
Concretamente, la hipótesis se lanza acerca de una función de probabilidad o densidad de una o varias variables aleatorias, basada en la información extraída de la muestra. De esta manera, al rechazar una hipótesis de este tipo, se está declarando el sesgo o falsedad en los datos de la muestra obtenida.

5.1. Conceptos básicos

Existen dos hipótesis posibles a analizar en un contraste de hipótesis: la **hipótesis nula**, denotada por H_0 , y la **hipótesis alternativa**, H_1 . La muestra extraída sobre la que se formula la hipótesis, que seguirá cierta distribución, se dividirá en dos áreas según estas dos hipótesis, estas son la **región de rechazo o crítica**, formada por los valores que apoyarían el rechazo de la hipótesis nula, y la **región de aceptación o no rechazo**, formada por los valores que reforzarían la decisión de aceptar la hipótesis nula. Los puntos límite entre ambas zonas se denominan **puntos críticos**. De aquí se entiende que la hipótesis nula será la que se desee contrastar y la hipótesis alternativa aquella que se cumplirá en los casos en que la hipótesis nula sea rechazada.

Para llevar a cabo la hipótesis, es necesaria una **regla de decisión** establecida que será usada como criterio a cumplir para poder aceptar la hipótesis. Esta regla de decisión es la encargada de definir los puntos críticos que crearán la zona de rechazo y la zona de aceptación.





5. Contraste de hipótesis

Nuevamente, existirá un valor α denominado **nivel de riesgo**, representando la probabilidad de que un valor concreto del estadístico de contraste caiga en la zona de rechazo, lo que conduciría a rechazar H_0 . El valor $1 - \alpha$ se llama **nivel de confianza**.

Existen contrastes unilaterales, como el ejemplo de la imagen anterior en la gráfica de la izquierda, donde el espacio asociado a α estará concentrado en uno de los dos extremos de la curva. En contrastes bilaterales, se tendrá $\frac{\alpha}{2}$ a cada extremo.

Un ejemplo de **hipótesis unilateral y bilateral** podría ser del tipo:

- **Unilateral** $\rightarrow H_0 \geq 0.7 \rightarrow H_1 < 0.7$
- **Bilateral** $\rightarrow H_0 = 0.7 \rightarrow H_1 \neq 0.7$

Un parámetro a valorar en el contraste de hipótesis es el error. Al aceptar o rechazar cierta hipótesis es posible no estar en lo cierto y que la hipótesis aceptada, H_0 , sea falsa; o también podría pasar lo contrario, que se rechazase la hipótesis siendo verdadera. De aquí, aparecen **dos tipos de errores**:

- **Error tipo I, α** , cometido cuando se rechaza la hipótesis nula, H_0 , siendo correcta.
- **Error tipo II, β** , cometido cuando se acepta la hipótesis nula, H_0 , siendo falsa.

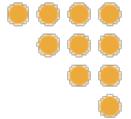
Finalmente, el llamado **p-valor** es una probabilidad que indica con qué seguridad se puede rechazar la hipótesis H_0 . Obviamente, rechazar una hipótesis nula porque el valor considerado ha quedado ligeramente por encima α , es decir, junto al punto crítico, es una decisión más arriesgada que el mismo rechazo cuando el valor considerado está muy al extremo, alejado de la región crítica. Gráficamente, cubre un área de seguridad alrededor del valor α .

5.2. Tipos de contraste

5.2.1. Contraste de hipótesis para la media, conocida la varianza

El mejor estadístico para trabajar hipótesis sobre la media es la **media muestral**, que sigue una distribución normal:

$$\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \rightarrow Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$



5. Contraste de hipótesis

Es posible realizar 3 tipos distintos de contraste:

- $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$, se RECHAZA H_0 si $\left| \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \right| > Z_{\frac{\alpha}{2}}$
- $\begin{cases} H_0: \mu \leq \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$, se RECHAZA H_0 si $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} > Z_\alpha$
- $\begin{cases} H_0: \mu \geq \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$, se RECHAZA H_0 si $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < -Z_\alpha$



5.2.2. Contraste de hipótesis para la media, desconocida la varianza y $n > 30$

En caso de no tener disponible la varianza poblacional, se acude al cálculo de la cuádratica varianza y así

$$\bar{x} \sim N \left(\mu, \frac{s}{\sqrt{n}} \right) \rightarrow \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \sim N(0,1)$$

Es posible realizar 3 tipos distintos de contraste:

- $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$, se RECHAZA H_0 si $\left| \frac{\bar{x}-\mu_0}{s/\sqrt{n}} \right| > Z_{\frac{\alpha}{2}}$
- $\begin{cases} H_0: \mu \leq \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$, se RECHAZA H_0 si $\frac{\bar{x}-\mu_0}{s/\sqrt{n}} > Z_\alpha$
- $\begin{cases} H_0: \mu \geq \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$, se RECHAZA H_0 si $\frac{\bar{x}-\mu_0}{s/\sqrt{n}} < -Z_\alpha$

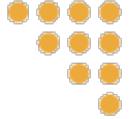
5.2.3. Contraste de hipótesis para la media, desconocida la varianza y $n < 30$

En estos casos no se puede aproximar la media a la normal y se usa la distribución t -Student, $\bar{x} \sim t_{n-1}$

$$\frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}$$

Es posible realizar 3 tipos distintos de contraste:

- $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$, se RECHAZA H_0 si $\left| \frac{\bar{x}-\mu_0}{s/\sqrt{n}} \right| > t_{n-1; \frac{\alpha}{2}}$
- $\begin{cases} H_0: \mu \leq \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$, se RECHAZA H_0 si $\frac{\bar{x}-\mu_0}{s/\sqrt{n}} > t_{n-1; \alpha}$
- $\begin{cases} H_0: \mu \geq \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$, se RECHAZA H_0 si $\frac{\bar{x}-\mu_0}{s/\sqrt{n}} < -t_{n-1; \alpha}$



5. Contraste de hipótesis

5.2.4. Contraste de hipótesis para la proporción

Del mismo modo, esta vez asumiendo que la proporción poblacional sigue una dis-

$$\text{tribución normal, } \bar{p} \sim N\left(p, \sqrt{\frac{pq}{n}}\right)$$

$$\frac{\bar{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} \sim N(0,1)$$

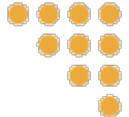
- $\begin{cases} H_0: p = p_0 \\ H_1: p \neq p_0 \end{cases}$, se RECHAZA H_0 si $\left| \frac{\bar{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} \right| > Z_{\frac{\alpha}{2}}$
- $\begin{cases} H_0: p \leq p_0 \\ H_1: p > p_0 \end{cases}$, se RECHAZA H_0 si $\frac{\bar{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} > Z_\alpha$
- $\begin{cases} H_0: p \geq p_0 \\ H_1: p < p_0 \end{cases}$, se RECHAZA H_0 si $\frac{\bar{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} < -Z_\alpha$

5.2.5. Contraste de hipótesis para la varianza

Cuando se realizan hipótesis sobre la varianza, se hace uso de la cuasivarianza y se asume una distribución chi-cuadrado o Pearson,

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$$

- $\begin{cases} H_0: \sigma^2 = \sigma_0^2 \\ H_1: \sigma^2 \neq \sigma_0^2 \end{cases}$, se RECHAZA H_0 si $\frac{(n-1)s^2}{\sigma_0^2} \notin \left(\chi_{n-1; \frac{\alpha}{2}}^2, \chi_{n-1; 1-\frac{\alpha}{2}}^2\right)$
- $\begin{cases} H_0: \sigma^2 \leq \sigma_0^2 \\ H_1: \sigma^2 > \sigma_0^2 \end{cases}$, se RECHAZA H_0 si $\frac{(n-1)s^2}{\sigma_0^2} > \chi_{n-1; \alpha}^2$
- $\begin{cases} H_0: \sigma^2 \geq \sigma_0^2 \\ H_1: \sigma^2 < \sigma_0^2 \end{cases}$, se RECHAZA H_0 si $\frac{(n-1)s^2}{\sigma_0^2} < \chi_{n-1; 1-\alpha}^2$



5. Contraste de hipótesis

En una muestra de 105 comercios elegidos aleatoriamente de una zona de la ciudad se ha observado que 27 han tenido pérdidas durante este mes. Se quiere saber si se puede dar por buena, con un nivel de significación del 5%, la afirmación de un analista que establece que la proporción de comercios en pérdidas es igual o superior al 35%.

Este contraste se puede declarar como

$$\begin{cases} H_0: p \geq 0,35 \\ H_1: p < 0,35 \end{cases}$$

y la proporción muestral obtenida de la observación es

$$\bar{p} = \frac{27}{105} = 0,26$$

De la tabla tipificada se obtiene el cuantil $-Z_{0,05} = -1,645$

El estadístico de contraste es igual a

$$\frac{\bar{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{\overline{0,26} - 0,35}{\sqrt{\frac{0,35(1 - 0,35)}{105}}} = -1,93$$

Puesto que se cumple

$$\frac{\bar{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} < -Z_\alpha,$$

se **rechaza la hipótesis**.

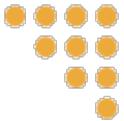
SAMSUNG

BeJob[™]

Together for Tomorrow!
Enabling People

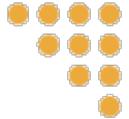
Aplicación a la Inteligencia Artificial

#TecnologíaConPropósito



Índice

Preprocesamiento de los datos	5
1. Introducción	7
2. Limpieza de datos	8
2.1. Duplicidad de los datos	8
2.2. Detección de valores atípicos	9
2.3. Identificación de valores atípicos	11
3. Evaluación de la información	12
3.1. Entropía de la información	12
3.2. Información mutua	13
3.3. Covarianza y Coeficiente de Pearson	13
3.4. Comparativa	15
Regresión lineal	17
1. Introducción	19
2. Planteamiento del problema	20
2.1. Análisis de regresión	20
2.2. Relación entre variables	20
2.3. Requisitos de las variables para regresión lineal	23
3. Método de mínimos cuadrados	24
Función de coste	27
1. Introducción	29
2. Función de coste	30
2.1. Gradiente descendente	31
3. Regresión multivariable	35



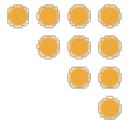
Índice

3.1. Gradiente descendente con múltiples variables	36
3.2. Normalización	37
4. Regresión no lineal	40
4.1. Regresión polinomial	40
4.2. Regresión logística	41
4.3. Métricas de evaluación	45

Together for Tomorrow!
Enabling People

Preprocesamiento de los datos

#TecnologíaConPropósito



1. Introducción

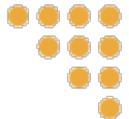
En estos días, está cambiando todo a nuestro alrededor para dar un salto a la nueva era que nos depara y de la que ya, en cierto modo, somos parte. Está aconteciendo una revolución socioeconómica solo comparable con las que trajeron al mundo nuevas Eras. Los motivos de estas revoluciones suelen estar concentrados en uno o varios acontecimientos o descubrimientos. Así, si a partir del descubrimiento de la agricultura, nuestros antepasados pudieron asentarse en un lugar determinado y que esto suscitó el inicio de las civilizaciones, o el desarrollo masivo de la imprenta, durante la transición entre la Edad Media y la Moderna, impactó tanto en el humano que llevó a la sociedad hacia el intelectualismo manifestado en el Humanismo y el Renacimiento, en nuestros días también hay un elemento que juega un rol protagonista en este cambio de Era, que algunos llaman Era Digital y que, mediante el acceso masivo a la información, ha revolucionado tanta crisis en lo establecido, como nuevos frentes esperanzadores hacia el progreso. Estamos hablando del **dato**.

Los datos han llegado a todas partes y con una accesibilidad cercana a lo inmediato proporcionando, prácticamente, cualquier tipo de información; progreso que conlleva no solo ventajas, sino que existe un problema que está creciendo cada vez más con este avance. El ritmo al que el flujo y acceso a la información aumenta supera con creces al ritmo al que la sociedad se está adaptando a este cambio, se prepara para saber tratar tal cantidad masiva de información continua.

Así, la *sobreinformación*, en ocasiones alimentada de *desinformación*, es un fenómeno que parece luchar contra las ventajas que en una situación normal brinda la información, que se resumen en conocimiento para tomar mejores decisiones.

Del mismo modo que afecta a nuestros procesos cognitivos de clasificación o toma de decisiones, sistemas o modelos predictivos basados en aprendizaje de datos pueden verse entorpecidos por información redundante o de baja calidad. Por este motivo, una de las fases primeras y de vital importancia antes de comenzar a explotar datos en un estudio inferencial estadístico o, en general, cualquier proceso dentro del mundo de la ciencia de los datos, es el denominado **preprocesamiento de datos**, que forma parte de la fase de *adquisición de datos* de un estudio estadístico.

Esta unidad proporcionará al lector las herramientas necesarias para preprocesar los datos. Se verá qué debe analizarse en los datos para asegurar su calidad, cómo reconocer que existe información redundante en un conjunto de variables, métodos para detectar valores espurios en grandes conjuntos de datos, así como maneras ordenar la información según su valía para un problema en concreto.



2. Limpieza de datos

Tras la recolección de datos de un problema, antes de comenzar un estudio estadístico o la construcción de un modelo predictivo o prescriptivo, se deben identificar los datos que, por alguna razón, están dañados, duplicados, alterados o, sencillamente, no están disponibles.

Los datos no tratados son denominados **datos crudos** y pasan a ser **datos procesados** una vez se ha preparado la información para ser explotada.

Favorece mucho al posterior tratamiento de los datos identificar estos valores para darle un tratamiento especial y evitar que perjudique el resultado de los cálculos.

2.1. Duplicidad de los datos

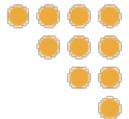
El caso concreto de datos duplicados puede suponer un perjuicio serio hacia el buen resultado esperado de un análisis o estudio estadístico.

Existen aplicaciones concretas en las que la duplicidad de datos se fuerza para conseguir ciertas mejoras en cuanto a tolerancia a fallas o desempeño, como en los sistemas distribuidos, pero igualmente en estos casos, es necesario tener estas duplicidades bien identificadas y localizables.

La única forma en la que la duplicidad de los datos no afecte negativamente a un análisis estadístico sería aquella en que esa misma duplicidad se distribuyese por los datos de la misma forma que lo hacen sus variables asociadas, lo cual es difícil ya que, en la mayoría de los casos, habrá más de una variable asociada a cada dato, y cada uno seguirá distinta distribución. Por lo tanto, es importante llevar a cabo algún proceso que permita eliminar la información duplicada.

Un dato duplicado es aquella observación o muestra cuyos valores de las variables asociadas coinciden exactamente con la de otro dato o muestra. Es importante notar que, si se tiene una tabla de datos con 1 000 muestras de 17 características o variables dos muestras no se consideran duplicadas si existe al menos una característica o variable cuyo valor difieran.

Realizar el proceso de eliminación de duplicados puede parecer una tarea engorrosa, pero asumiendo que los datos se disponen en formato digital, prácticamente cualquier software que almacene datos es capaz de realizar esta tarea.



2. Limpieza de datos

2.2. Detección de valores atípicos

Con tan solo un valor atípico se puede ‘destruir’ por completo un parámetro estadístico como la media o esperanza matemática. Si, por ejemplo, se tiene la siguiente tabla de datos

1	2	3	4	5	6	7	8	9	10	11	12	13
0,126	0,104	0,094	0,11	0,1	0,116	0,121	0,123	0,114	11467	0,107	0,13	0,134

De la que, de una vista rápida, se aprecian valores en torno a 0,1. Sin embargo, al calcular la media aritmética de los valores, se obtiene

$$\bar{x} = \frac{\sum x_i}{n} = 882,2$$

Mirando detenidamente los datos, se detecta que la observación número 10 es, del orden de mil veces superior al resto. Eliminando este dato de la cuenta, se obtiene ya un valor para el parámetro estadístico más realista

$$\bar{x} = \frac{\sum x_i}{n} = 0,11486$$

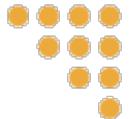
En este ejemplo, es posible detectar rápidamente el origen del problema debido al reducido tamaño de la tabla. Pero en la mayoría de los casos prácticos, en los que se tendrán tablas con muchas variables y una enorme cantidad de observaciones, esta detección se hace intratable manualmente.

2.2.1. Detección por parámetros estadísticos

La distribución que lleven los datos sobre su espacio de variación puede dar suficiente información sobre la posible existencia de datos atípicos. Como ya es sabido, la distribución puede ser representada mediante parámetros estadísticos y mediante gráficas.

Mediante parámetros estadísticos, es posible ver si existe algún desequilibrio en la distribución de los datos para descartar la existencia de datos atípicos.

En el ejemplo anterior, se obtenía una media de $\bar{x} = 882,2$. Como criterio para saber si existen datos atípicos que están alterando la media se puede contrastar este estadístico con la mediana. Se espera que la mediana y la media, aunque distintos, no varíen en gran medida, al menos en el orden, en cuyo caso, la distribución debería tener una varianza acorde al resultado.



2. Limpieza de datos

Es decir, si tras calcular la media, se obtiene una mediana del orden de la media, podría descartarse la existencia de datos atípicos o, al menos, de una cantidad lo suficientemente importante como para alterar seriamente los datos. En caso contrario, para considerar una situación normal (sin valores atípicos) la de una distribución cuya media y mediana varíen sustancialmente, la varianza de la distribución debería ser, como mínimo mayor que la distancia entre estos dos estadísticos.

Para el caso del ejemplo anterior, la mediana resulta ser

$$Me = 0,1149$$

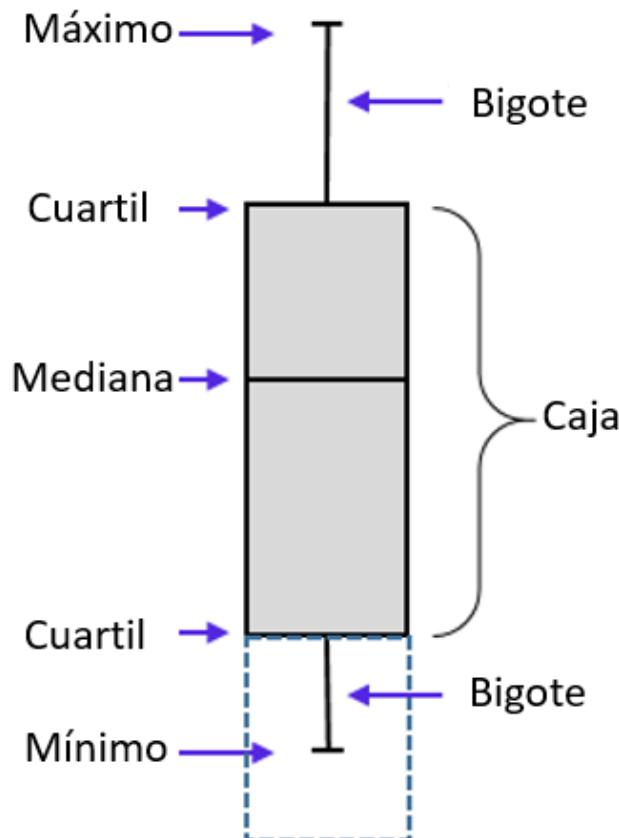
Al diferenciarse tan sustancialmente la mediana de la media se puede suponer que habrá datos atípicos que están alterando la media.

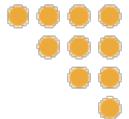
2.2.2. Detección gráfica

Un diagrama Box-Whisker puede resultar muy útil para la detección de valores atípicos.

El conjunto caja-bigotes se componía, como se aprecia en la imagen, de una caja, limitada por los cuartiles Q_1 y Q_3 , indicando que el 50 % de los datos centrales se encuentran dentro de ese intervalo, y una línea divisoria dentro de la caja para indicar la mediana; además, la extensión de los bigotes representa las “colas” del 25 % menor y 25 % mayor de los datos.

Gracias a esta disposición, será fácil identificar la presencia de valores atípicos. Por ejemplo, en casos en que la media se sitúe fuera de la caja; esto indica que la media está desplazada más allá del espacio donde quedan la mitad de los datos más centrados, dejando entrever el desajuste.





2. Limpieza de datos

2.2.3. Detección por normalización

Otro método para detectar la existencia de datos atípicos es mediante la **normalización**. La normalización escala todo el rango de valores de una variable aleatoria al intervalo [0,1], de tal manera que el valor mínimo sea igual a 0 y el máximo a 1 y se mantenga la distribución de los datos en el nuevo intervalo.

Mediante este método, en los casos de presencia de atípicos, habitualmente la mayoría de los valores pasan a ‘comprimirse’ junto a los extremos.

En la tabla del ejemplo anterior, los datos normalizados quedarían así

1	2	3	4	5	6	7	8	9	10	11
0,00000276	0,00000084	0	0,00000143	0,00000052	0,00000019	0,00000023	0,00000025	0,00000017	1	0,0000012

2.3. Identificación de valores atípicos

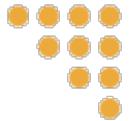
Una vez se demuestra que debe haber presencia de valores atípicos que están alterando la paramétrica de la distribución, es hora de identificarlos. Un método muy usado es identificar como valores atípicos todos aquellos que superen un umbral definido, T , según algún parámetro estadístico de la distribución.

$$T = c \cdot \sigma$$

Siendo σ la desviación típica y c un coeficiente ajustable, que irá acorde a la naturaleza del problema.

Otro método es la **identificación por muestreo**, por la cual se extraen n muestras y se calculan sus medias, las cuales tendrán una alta probabilidad de diferir con respecto la calculada para el total de los datos.

Algunos otros que no se especificarán en esta unidad son la **prueba de Tukey** o las **distancias de Cook**.



3. Evaluación de la información

Es importante también, en aras de la eficiencia, comprobar que los datos disponibles de cierta variable son relevantes para el fin del análisis.

En este caso, probablemente, información no relevante y totalmente independiente del problema no tendrá ningún efecto en el resultado del análisis, ni negativo ni positivo, pero probablemente si afecta a la complejidad de la información y el tiempo de proceso, con la reducción de eficiencia que esto conlleva.

Una variable aleatoria puede no dar ninguna información relevante al caso estudiado, pero también se puede dar el caso que sí de información, pero la información que está aportando es exactamente la misma que está aportando otra variable que también está incluida en el estudio, fenómeno denominado **redundancia**. Cuanto esto ocurre significa que existe una fuerte **correlación** entre estas dos variables.

Se está realizando un análisis para prevención de incendios en una planta de residuos altamente peligrosos. Durante un tiempo determinado, se extraen mediante sensores todo tipo de información ambiental, así como de temperaturas de la maquinaria, estados químicos de los elementos empleados en la planta e incluso datos meteorológicos.

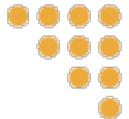
Cuando se tienen todos los datos y se comienza a analizar los datos, el analista se deshace de una columna entera de la tabla referente a la variable *presión atmosférica*.

¿La razón? Entre los datos, también existe otra columna que muestra la información de *temperatura atmosférica*. La temperatura y la presión están fuertemente correlacionadas entre sí, ya que una depende linealmente de la otra.

3.1. Entropía de la información

Para analizar el papel que juega una variable aleatoria en un estudio determinado, es decir, si la información que proporciona es relevante para el análisis se usa un medidor llamado **entropía** o **entropía Shannon**, $H(X)$.

Se puede ver la entropía como una *medida del desorden* lo cual, llevado a la estadística, el desorden es transformado en *incertidumbre*. A partir de su fórmula, se ve que, dada una variable aleatoria ξ , cuanto más descompensando estén las probabilidades entre sus posibles valores, es decir cuanto más conocimiento haya



3. Evaluación de la información

sobre el resultado del fenómeno que mide la variable ξ menos información proveerá.

$$H(\xi) = - \sum_{i=1}^n p(x_i) \log_2 p(x_i)$$

Una variable asociada a un fenómeno muy certero, no proporcionará mucha información, debido a que tendrá el mismo valor prácticamente en todas las observaciones.

Lo anterior se puede demostrar exponiendo tres fenómenos distintos basados en la distribución de Bernoulli. Se tienen tres fenómenos que siguen esta distribución con probabilidades

$$p_1 = 0,5, q_1 = 0,5, \quad p_2 = 0,6, q_2 = 0,4, \quad p_3 = 0,9, q_3 = 0,1$$

La entropía de cada experimento es:

1. $H(\xi_1) = -0,5 \log_2 0,5 - 0,5 \log_2 0,5 = 1$
2. $H(\xi_2) = -0,6 \log_2 0,6 - 0,4 \log_2 0,4 = 0,97$
3. $H(\xi_3) = -0,9 \log_2 0,9 - 0,1 \log_2 0,1 = 0,468$

3.2. Información Mutua $I(x, y)$

La información mutua mide la dependencia mutua entre dos variables. Es decir, la reducción de incertidumbre, es decir de entropía, de una variable, cuando se tiene evidencia del valor de la otra. Se expresa como

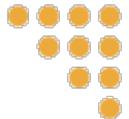
$$MI(x, y) = H(x) - H(x|y) = \sum_{j=1}^m \sum_{i=1}^n p(x_i, y_j) \log_2 \frac{p(x_i, y_j)}{p(x_i)p(y_j)}$$

Es una medida muy empleada en estadística bayesiana ya que, al medir cómo de condicionadas están dos variables, puede ayudar a construir toda una red bayesiana.

3.3. Covarianza y Coeficiente de Pearson

En análisis multidimensional se emplea mucho el término covarianza la cual mide la correlación lineal entre dos variables. Se denota por σ_{xy} y se expresa como

$$\sigma_{xy} = E(XY) - E(X)E(Y),$$



3. Evaluación de la información

que para el caso discreto se puede reescribir como

$$\sigma_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \cdot \bar{y}$$

Al igual que ocurría con la varianza, que no podía trabajarse en las mismas unidades que la media o los propios datos y por ello se transformaba en la variación estándar y de esta en el coeficiente de variación, consiguiendo adimensionalizar la medida, en el caso de la covarianza, existe un coeficiente llamado **coeficiente de Pearson** que no es más que la covarianza estandarizada y, por ello, no posee unidades de medida

$$\rho_{x,y} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

Este parámetro da un valor comprendido entre $[-1, 1]$ donde, donde valores negativos significan que existe una relación lineal opuesta, es decir, cuando más aumenta el valor de x , mas disminuye el valor de la variable y , y viceversa; los valores positivos indican una relación positiva entre ambos. Cuanto más se acerca al -1 o 1 , más perfecta es la correlación, hasta llegar al 0 que indica una falta total de correlación.

3.3.1. Relaciones espurias

Es importante remarcar que este medidor no indica causalidad. Eso quiere decir que, aunque dos variables puedan estar fuertemente correladas, no quiere decir que exista condicionalidad entre ambas, que el hecho de que una ocurra reduzca de ninguna manera la probabilidad de que la otra ocurra.

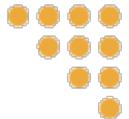
El ejemplo habitual para demostrar esto es el de la venta de helados y la obesidad; durante una temporada estival en una ciudad concreta se observó cómo la tasa de muertes por ahogamiento aumentaba cuando crecía la venta de helados, dando a entender que los helados, en cierto modo, causaban ahogamiento.

Esto era falso, el motivo venía por una tercera variable (llamada para el caso **variable de control o de confusión**)

ambiental que era el aumento de temperaturas, el cual provocaba estas dos que eran totalmente independientes entre sí, aunque correladas. Este fenómeno se denomina **relación espuria** y se puede descubrir mediante la **correlación parcial**.



Consiste en comprobar cómo varían las variables, cuya correlación fue estudiada, al modificar una de ellas, pero dejando la potencial variable de confusión constante. En el ejemplo de los helados, esta consistiría en agrupar todas las observaciones en



3. Evaluación de la información

las que variaron la venta de helados y/o el número de ahogamientos para *una misma temperatura*.

Esta relación se cuantifica mediante la correlación parcial:

$$\rho_{xy,z} = \frac{\rho_{xy} - \rho_{xz}\rho_{yz}}{\sqrt{1 - \rho_{xz}^2}\sqrt{1 - \rho_{yz}^2}}$$

3.4. Comparativa

La información mutua y la covarianza son medidas complementarias, puesto que no describen los mismos aspectos sobre la asociación de dos variables aleatorias. Por un lado, se tiene la gran limitación de la covarianza que “no ve más allá” de la linealidad. Es decir, la covarianza solo analiza correlaciones lineales y se puede tener una covarianza nula mientras las dos variables están fuertemente relacionadas de manera no lineal. Por otro lado, la covarianza puede estimarse de manera muestral, a partir de una muestra de los datos sin la necesidad de conocer las distribuciones reales de probabilidad involucradas en cada una de las variables, mientras que la información mutua requiere esta información. Por esta razón, el empleo de uno u otro parámetro estadístico irá en función del tipo de problema.

En una cadena de restaurantes de comida rápida, se estudian estrategias de expansión. Para ello, se quiere ver qué impacto tienen las distancias entre el domicilio y el restaurante en los que consumen los clientes, con el tipo de pedido que se realiza (delivery o en tienda). Entre otras muchas variables, se dispone de dichas distancias y de número de clientes de la tienda y se pretende saber cómo de correlacionadas están ambas variables.

N. Clientes (00)	8	7	6	4	2	1
Distancias (00 m)	15	19	25	23	34	40

Adjudicando x a los cientos de clientes e y a los cientos de metros, se obtienen las medias muestrales

$$\bar{x} = \frac{\sum x_i}{N} = 4,6, \quad \bar{y} = \frac{\sum y_i}{N} = 26,$$

Con las medias muestrales es posible calcular las tres medidas de varianzas necesarias

$$\sigma_x = \sqrt{\frac{\sum x_i^2}{N} - \bar{x}^2}, \quad \sigma_y = \sqrt{\frac{\sum y_i^2}{N} - \bar{y}^2}, \quad \sigma_{xy} = \frac{\sum x_i \cdot y_i}{N} - \bar{x} \cdot \bar{y}$$
$$\sigma_x = 2,57, \quad \sigma_y = 8,56, \quad \sigma_{xy} = -20,83$$

Con estos valores, el coeficiente de Pearson resulta

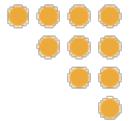
$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = -0,94,$$

lo que indica una fuerte correlación negativa.

Together for Tomorrow!
Enabling People

Regresión lineal

#TecnologíaConPropósito



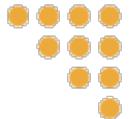
1. Introducción

Un diagrama de dispersión ofrece una idea sobre cómo dos variables están relacionadas. Da información sobre la dispersión existente entre esas dos variables, que puede traducirse de un modo cualitativo en lo bien que aparece definida la forma con la que evolucionan los valores; esto es, la forma de la nube de puntos formada por todas las observaciones. Obviamente, si se estudia una posible correlación lineal, los incrementos entre observaciones deberían poder aproximarse a una misma proporcionalidad. En definitiva, la direccionalidad de la nube de puntos: lo bien que se podría ajustar la nube de puntos a una recta.

El diagrama de dispersión puede dar una idea visual, puede ser bastante útil según el caso, pero nunca permitirá cuantificar, convertir en medible este grado de similitud con una recta. Para realizar esto, el análisis de regresión lineal posee las herramientas para calcular cómo de bien se ajusta un grupo de puntos bidimensionales a una recta, mediante el cálculo del error cometido cuando se asumen todos los puntos pertenecientes a la recta en cuestión.

La importancia del análisis de regresión supone un pilar para numerosas ramas de la ciencia. Es adaptable a una gran variedad de situaciones de tal manera que es de gran aplicación en disciplinas como economía, comercio o en investigaciones bioestadísticas, por nombrar algunos. Y este atractivo es debido a que, si bien se est considerando hasta ahora dos variables de entrada en dicho análisis, es aún más beneficioso realizarlo entre la variable dependiente o de criterio y las llamadas variables independientes o *predictoras*: si es posible encontrar una recta que se ajuste a las variaciones encontradas entre las observaciones de dos variables de este tipo con un error aceptable, se tiene una ecuación (la propia de la recta encontrada) para predecir cualquier valor, desconocido o futuro, dada la entrada de la variable dependiente.

Esta unidad profundizará en detalle todas las fases del proceso de un análisis de regresión lineal, desde la representación de los datos, hasta las métricas usadas para cuantificar el error o como asignar intervalos de confianza al resultado obtenido.



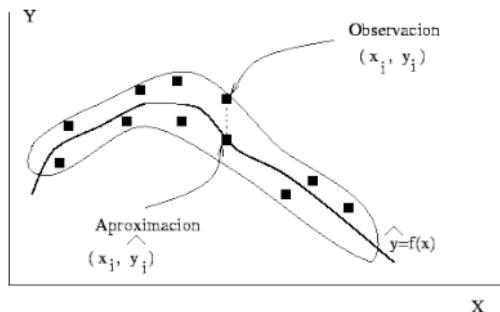
2. Planteamiento del problema

En esta sección, se pretende formalizar el problema de regresión definiendo más detalladamente la información disponible, el objetivo del problema desde una perspectiva geométrica y la forma que toma el resultado esperado.

2.1. Análisis de regresión

En general, el análisis de regresión trata de encontrar la ecuación de la curva que más se asemeja a la evolución de los valores de dos o más variables correlacionadas. Debido a la complejidad que supone el ajuste de la curva que siguen dos variables con una relación fuertemente no lineal a la ecuación que la define, esta unidad se limita al caso lineal. Los casos no lineales suponen soluciones basadas en procesos estocásticos que forman parte mayormente de la disciplina de aprendizaje automático.

$$\hat{Y} = f(X)$$



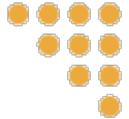
El objetivo principal de la **regresión lineal** es construir un modelo lineal que corrobore la relación lineal existente entre las variables implicadas, y por otro lado sea usado para calcular valores no observados de la variable independiente a partir de las predictoras.

La **regresión lineal simple** simplifica este problema al caso bidimensional en el que se tienen en cuenta dos variables: una independiente o predictora y otra dependiente.

2.2. Relación entre variables

En primer lugar, será interesante describir, también de manera gráfica, los distintos tipos de relación posible entre variables:

- **Relación determinista.** Representa el caso ideal de la relación entre variables. Es la relación existente cuando las variables independientes tienen efecto determinista en la variable dependiente. Ante el mismo estado de valores en las

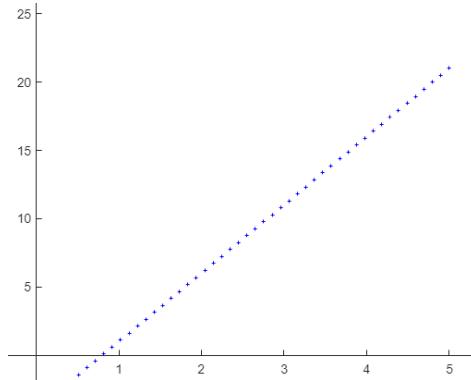


2. Planteamiento del problema

variables independientes, siempre se obtendrá como resultado el mismo valor en la variable dependiente. No existe varianza en las medidas.

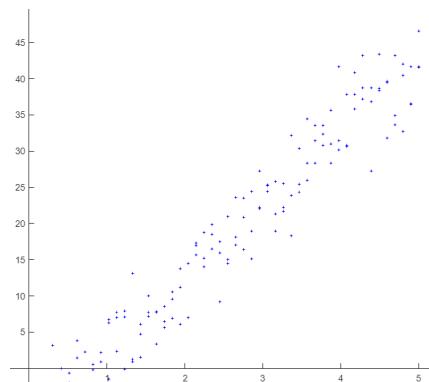
Se puede crear un modelo exacto de esta relación del tipo

$$x_{out} = f(x_1, x_2, \dots, x_n)$$



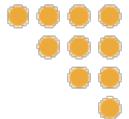
- **Relación estocástica.** Son casos en los que, ante experimentos o mediciones repetidas con condiciones iniciales (valores en variables independientes) exactamente iguales, es posible obtener distintos valores de salida en la variable dependiente. Cuando se modelan estas relaciones, se añade una variable aleatoria llamada *ruido*, r .

$$x_{out} = f(x_1, x_2, \dots, x_n) + r$$



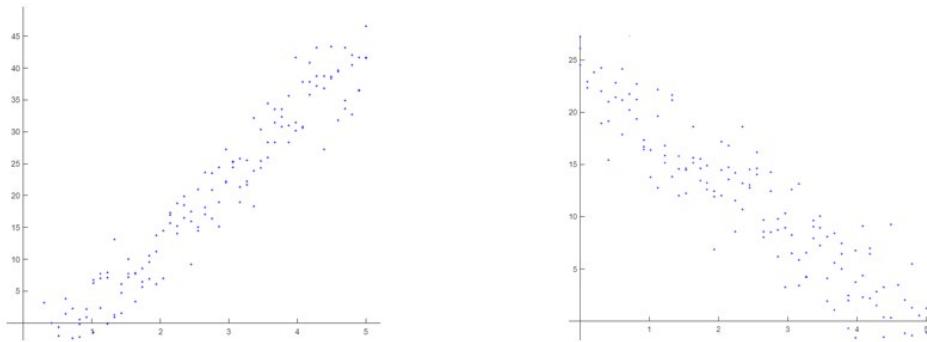
Independientemente, de esta división determinista/estocástico, las relaciones entre variables también se pueden clasificar por linealidad. Así, una **relación lineal** sería cualquiera de las dos anteriores, en las que los valores de las variables visualmente toman forma de línea debido a que la relación que poseen entre ellas puede ser modelada mediante una función lineal para el caso unidimensional del tipo

$$y = ax + b,$$

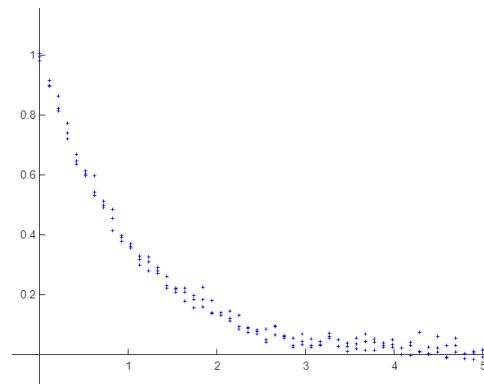


2. Planteamiento del problema

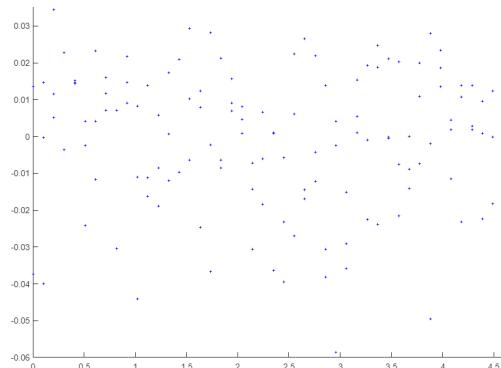
donde el signo del coeficiente de proporcionalidad a (la pendiente de la recta) define el tipo de relación lineal: cuando a toma un valor negativo, el incremento de la variable de entrada x supondrá decremento en la variable de salida y , y se denominará relación lineal negativa; en caso contrario, se denominará relación lineal positiva.

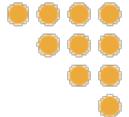


- La relación no lineal, por su parte, es aquella que no puede ser representada mediante una ecuación lineal. Gráficamente siguen una curva, pero esta no es recta.



Cuando no existe ningún tipo de relación entre variables, visualmente no se es posible de encontrar ninguna dirección predominante en la nube de puntos y, por lo tanto, no tendrá ningún sentido intentar ajustar una recta a los datos.





2. Planteamiento del problema

2.3. Requisitos de las variables para la regresión lineal

De manera general, se pueden enumerar una serie de condiciones que dos variables deben cumplir para que sea posible elaborar una recta de regresión lineal simple a partir de ellas:

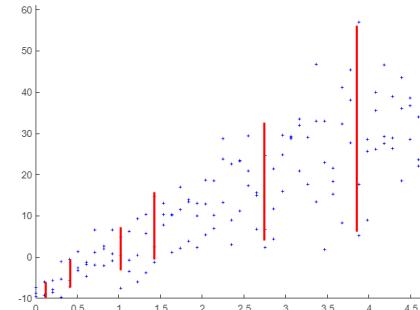
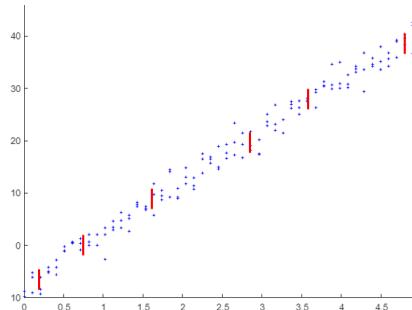
1. **Linealidad.** Definido en el apartado anterior, exige una relación lineal entre las dos variables del tipo

$$f(x) = ax + b$$

2. **Homogeneidad.** Indica que el conjunto de desplazamientos de los valores reales con respecto a los simulados por la recta debe estar centrado en la propia recta. Esto se traduce en que el valor promedio del error calculado debe ser nulo.

$$E(\epsilon_i) = 0$$

3. **Homocedasticidad.** Es la cualidad que caracteriza un conjunto de datos cuya varianza de las distribuciones que siguen las medidas de la variable dependiente es constante a lo largo de las observaciones.

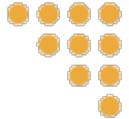


4. **Independencia.** Todas las observaciones empleadas para la construcción de la recta deben ser extraídos de experimentos o medidas independientes.

$$E(y_i y_j) = 0$$

5. **Normalidad.** Los errores de las medidas deben seguir una distribución normal.

$$y_i \sim N(0, \sigma)$$



3. Método de mínimos cuadrados

Existen diversos métodos de ajuste de curvas, pero sin duda alguna, para el caso lineal el método de mínimos cuadrados o mínimos cuadráticos es el más eficaz y más usado.

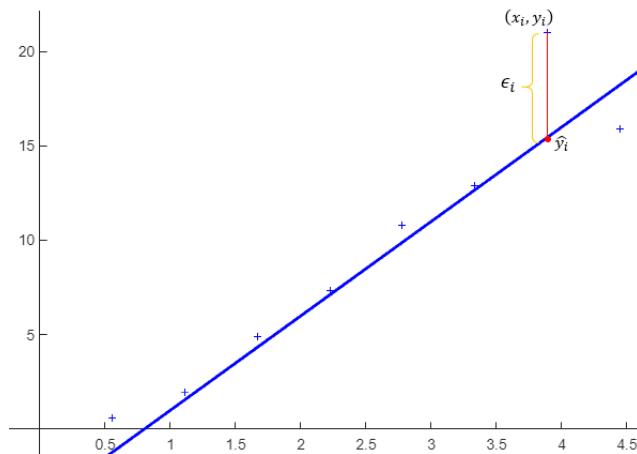
Como se ha dicho anteriormente, el objetivo de la regresión lineal simple es la obtención de la recta más representativa de los datos observados. Es decir, de la expresión general de la función de la recta:

$$f(x) = ax + b,$$

se necesita encontrar los coeficientes a y b que formen la recta más similar a los datos observados.

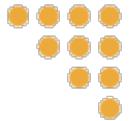
En el caso perfecto e ideal en que se encontrase una recta perfectamente ajustada a los datos, estos datos deberían llevar distribuciones degeneradas en sus medidas, de tal manera que, para cada medida, el valor coincidiese con el estimado por la recta. Gráficamente, todos los puntos de quedarían sobre algún punto de la recta. En la práctica, esto es imposible debido a la condición continua de la variable medida, por lo que normalmente los puntos no quedarán sobre la recta. Existe una manera de medir el error entre el valor de la observación y su valor estimado por la recta.

Dada una observación (x_i, y_i) y la recta de regresión lineal obtenida para el conjunto de observaciones en las variables X e Y , $y = ax + b$, el error cometido entre la estimación ligada a la recta, \hat{y}_i y la observación y_i es igual a la longitud del segmento que va desde el punto (x_i, y_i) a la intersección de la recta vertical que pasa por el mismo punto y la recta vertical (x_i, \hat{y}_i) .



Este error es conocido como **residuo** y se calcula del siguiente modo

$$\epsilon_i = y_i - \hat{y}_i$$



3. Método de mínimos cuadrados

Con esta noción del error puntual cometido, se puede formular la definición del método.

Dada dos variables X e Y , una muestra compuesta por n observaciones (x_1, x_2, \dots, x_n) y (y_1, y_2, \dots, y_n) y una recta de regresión ajustada a las observaciones $\hat{y} = \hat{b}x + \hat{a}$, el **método de mínimos cuadrados** resuelve el problema de optimización definido por

$$\min_{\epsilon} f(\epsilon) = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - (a + bx_i))^2$$

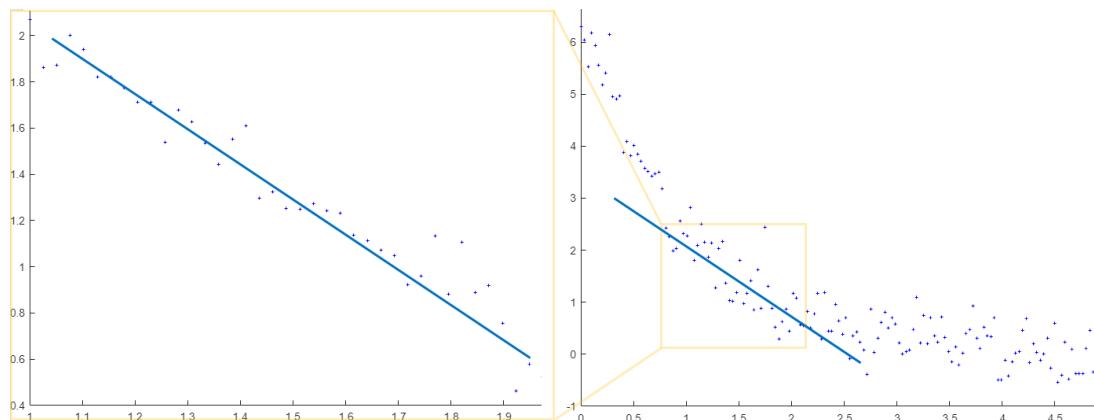
Para la resolución de este problema se llevan a cabo una serie de procesos mediante derivaciones que, finalmente, llegan a la conclusión

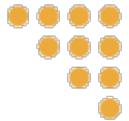
$$\begin{aligned}\hat{b} &= \frac{s_{xy}}{s_x^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ \hat{a} &= \bar{y} - \hat{b}\bar{x}\end{aligned}$$

El coeficiente \hat{b} , que indica la pendiente de la recta de regresión, se denomina **coeficiente de regresión de Y sobre X**.

Una vez más, es importante tener en cuenta la calidad de la muestra pues, si se quiere usar una recta como predictor, se deberá tener un rango de observaciones proporcional a la lejanía de la predicción. Es decir si tenemos N muestras en un intervalo de valores de tamaño n , sería peligroso para la precisión del predictor intentar predecir un punto futuro y cuya distancia con respecto al intervalo de observaciones sea considerable dado el tamaño del intervalo.

Esto es debido a que es posible sufrir un problema de escala y, mientras el rango de valores observados de X se ajusta muy bien a una recta, no resulte igual cuando se multiplica el tamaño del intervalo observado.





3. Método de mínimos cuadrados

De una empresa de un producto alimentario se conocen los siguientes datos referentes al volumen de ventas y gasto en la publicidad.

Ventas (M €)	Gasto publicitario (m€)
10	16
15	32
20	48
22	56
30	64
32	80

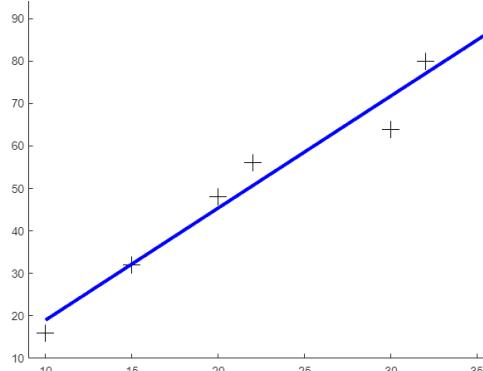
Necesitan conseguir un volumen de ventas para este año de 35 millones de euros para sufragar deudas. ¿Cuánto deberían invertir en gasto publicitario?

Para ello, debe modelarse esta relación mediante una recta de regresión y, dado el planteamiento, se asigna la variable predictora X al volumen de ventas, ya que es el dato de entrada, y la variable Y al gasto, como variable dependiente. Los datos requeridos para la estimación de la recta, $\hat{y} = \hat{m}x + \hat{n}$, son

$$s_x = \sqrt{\frac{\sum x_i^2}{N} - \bar{x}^2} = 7,74, \quad s_{xy} = \frac{\sum x_i \cdot y_i}{N} - \bar{x} \cdot \bar{y} = 158$$
$$\hat{m} = \frac{s_{xy}}{s_x^2} = 2,637, \quad \hat{n} = \bar{y} - \hat{m}\bar{x} = -7,356$$

Se tiene así la recta $\rightarrow \hat{y} = 2,637x - 7,356$

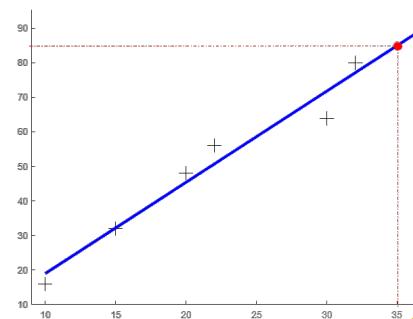
La recta obtenida queda representada así:



Con esta recta, sólo es necesario introducir el valor de entrada en la variable predictora x para obtener la predicción \hat{y} .

$$\hat{y} = 0,363x_o + 3,604 = 2,637(35) - 7,356 = 84'94$$

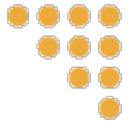
84940 €



Together for Tomorrow!
Enabling People

Función de coste

#TecnologíaConPropósito



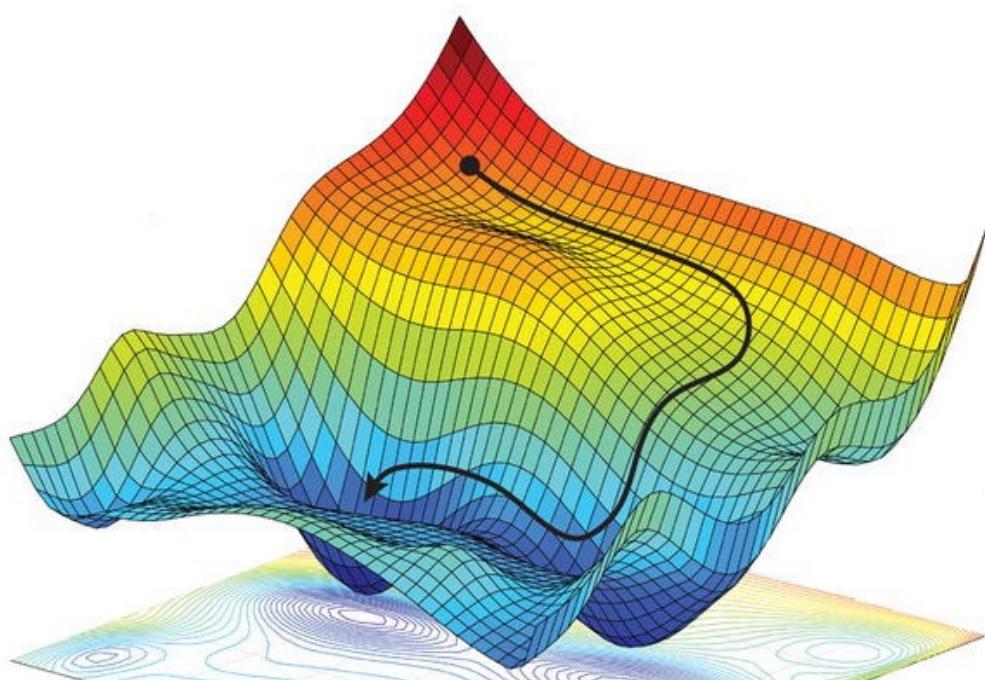
1. Introducción

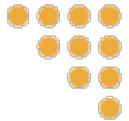
La regresión lineal es una herramienta muy útil a la hora de estimar información a partir de otra disponible correlacionada linealmente con la primera. A modo de aproximación o para sistemas muy simples, puede proporcionar buenos resultados, especialmente cuando se precisan intervalos de confianza sobre la recta estimada.

Sin embargo, en la práctica, rara vez un suceso dependerá exclusivamente de un solo factor y, aunque aislar el sistema a una variable de entrada como única causa de la variable de salida puede permitir realizar aproximaciones de manera sencilla, cuanta más información disponible tenga el sistema, mejor será la precisión de las estimaciones que realice y, por tanto, mejores decisiones se podrán tomar en consecuencia.

Pero el aumento de información de entrada tiene un precio que se refleja en la complejidad del problema. A medida que se avanza en la dirección del modelo predictivo o regresor mejorado, el sistema se aleja de la ideal linealidad y del modelo univariable que la regresión lineal simple es capaz de tratar.

Esta unidad se centra en el cambio de paradigma que supone a la inferencia estadística este aumento en la complejidad del problema del diseño de modelos analíticos, marcado por un escenario fuertemente no lineal y con múltiples variables predictoras.





2. Función de coste

Para el caso lineal simple, la recta de regresión puede ser calculada de manera determinista, dados los datos de una muestra de la población. El método de mínimos cuadrados llega a una fórmula bien definida a partir de la cual se pueden calcular los dos coeficientes necesarios para construir la recta estimada. El proceso que conduce a esta fórmula se transforma en uno de optimización donde, para un espacio de valores $\{m, n\}$, se busca la combinación que reduzca la suma de residuos o errores de cada punto al mínimo.

En general, los problemas de optimización se pueden definir como, dado un objetivo a maximizar o minimizar y unas variables, la búsqueda del estado compuesto por una combinación de todas ellas, tal que la función que define el objetivo se optimice al máximo o mínimo, según el caso. Esta función debe ser cuidadosamente definida y se denomina **función objetivo** o **función de coste**.

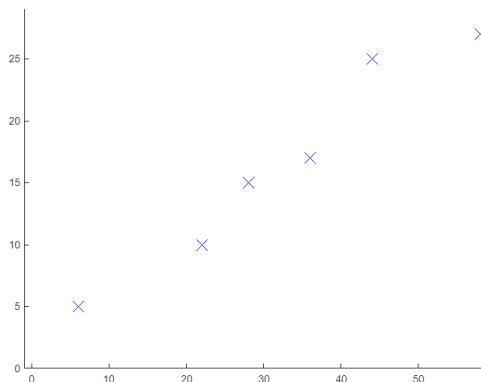
Sin ir más allá del caso lineal, en éste se tiene como objetivo encontrar la recta lineal:

$$h(x) = \theta_1 x + \theta_0$$

más representativa a los datos incluidos en la muestra, donde se han reasignado las variables θ_1 como coeficiente de proporcionalidad de la recta o pendiente y θ_0 , como valor de cruce con el eje de abscisas.

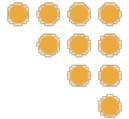
Es decir, dados una tabla de datos con m muestras y n variables con las variables $X = (x_1, x_2, \dots, x_n)$, $Y = (y_1, y_2, \dots, y_n)$, la idea es elegir la pareja θ_0, θ_1 tal que, la suma de todas las distancias cuadráticas entre $h(x_i)$ e y_i sea lo mínima posible.

Se quiere comprobar qué rectas de las tres propuestas se ajusta mejor a los datos.



Se tienen

$$h_1(x) = 0.5x + 5, \quad h_2(x) = \frac{x}{10} + 10, \quad h_3(x) = 3.2x + 4.5$$



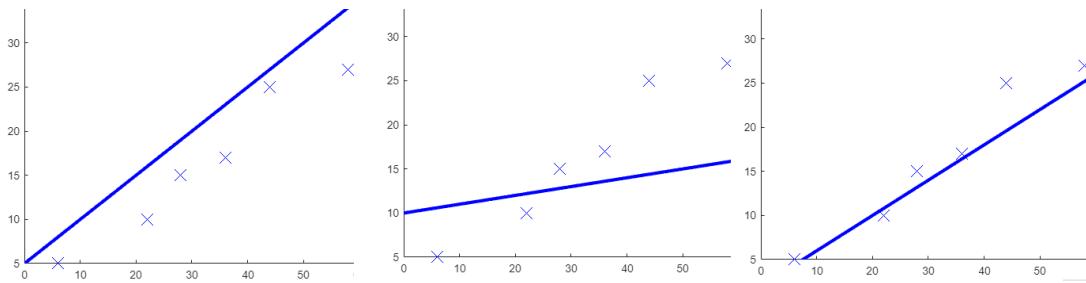
2. Función de coste

El error de un dato con la recta estimada es igual a la distancia vertical entre el punto y la recta. Según el método de mínimos cuadrados (LS), se tiene

$$LS_j = \sum_{i=1}^m (h_j(x_i) - y_i)^2, \quad j = 1, 2, 3$$

$$LS_1 = 150, \quad LS_2 = 290,4, \quad LS_3 = 37$$

La tercera función $f_3(x)$ obtiene el menor error y, por lo tanto, es la que mejor se ajusta a los datos. Efectivamente, si se comprueban sobre la gráfica, se tiene



Se ve cómo, para cada valor de θ_0 y θ_1 se obtiene un indicador del error mínimo cuadrático como salida. Si se reinterpretase estos valores en un nuevo eje sobre un espacio cartesiano $\theta_0\theta_1$, como la función

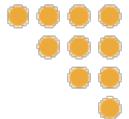
$$LS(\theta_1, \theta_2) = \sum_{i=1}^m (x_i \cdot \theta_1 + \theta_0 - y_i)^2,$$

se tendría una función bivariable a minimizar. Es decir, se tendría un conjunto infinito de valores en el espacio tridimensional con un mínimo que sería el objetivo a buscar. Para el caso de la regresión lineal simple, esa función será la función de coste, la cual suele denominarse por $J(\theta_0, \theta_1)$.

2.1. Gradiente descendente

Como es sabido, la búsqueda de mínimos y máximos locales en una función se lleva a cabo usando la derivada de la función, la cual se hace cero en sus puntos de mínimos y máximos locales. Para casos lineales simples, esto puede resultar sencillo pero, de un modo general que pueda resolver casos desde múltiples variables y no lineales, se usa un algoritmo iterativo llamado **gradiente descendente**.

Este algoritmo, inicializa los coeficientes θ_0 y θ_1 , bien arbitrariamente o mediante algún otro algoritmo de inicialización, y calcula el máximo gradiente negativo como

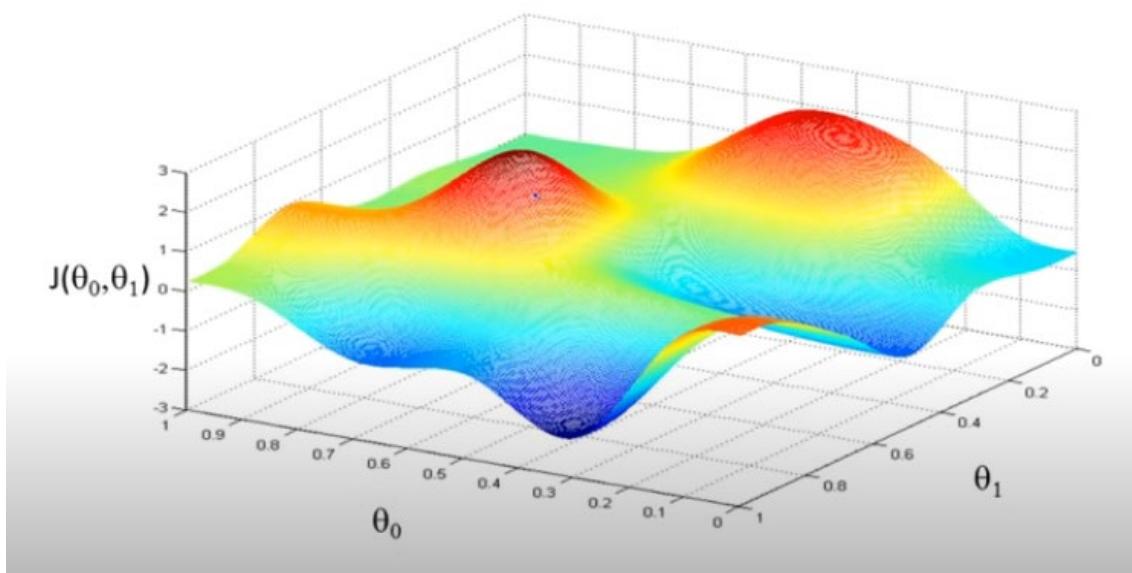


2. Función de coste

la dirección de máximo descenso posible. Este gradiente, junto a otros parámetros, indicará cómo de grande y en qué dirección se avanzará hacia el siguiente punto dentro del espacio de valores θ_0, θ_1 . Tras la actualización de valores, con los nuevos, se calcula el nuevo valor de la función de coste, la cual debe haber disminuido, y se vuelve a empezar el proceso.

Tenga en cuenta que, para el caso univariable, se tiene una sola curva bidimensional en la que no existe direcciones posibles entre las que elegir, tan sólo una que podrá tomarse en ambas direcciones y, si no se ha alcanzado un máximo o un mínimo, la derivada hacia una dirección aumentará el valor de la función de coste y en la otra dirección lo disminuirá.

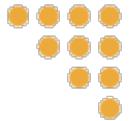
Para el caso de dos variables, cada punto tiene infinitas direcciones posibles hacia las que variar y el gradiente descendente se encarga de encontrar la dirección de máximo descenso, es decir, de mínima pendiente negativa.



Esta dirección de máximo gradiente vendrá representada por el vector compuesto por las derivadas parciales de la función de coste, $J(\theta_0, \theta_1)$, según cada variable de entrada, en este caso, θ_0 y θ_1 . Así, dados unos valores determinados en un paso iterativo cualquiera dentro del algoritmo, la actualización de los nuevos valores se calcula así

$$\theta_0 := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$

$$\theta_1 := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$



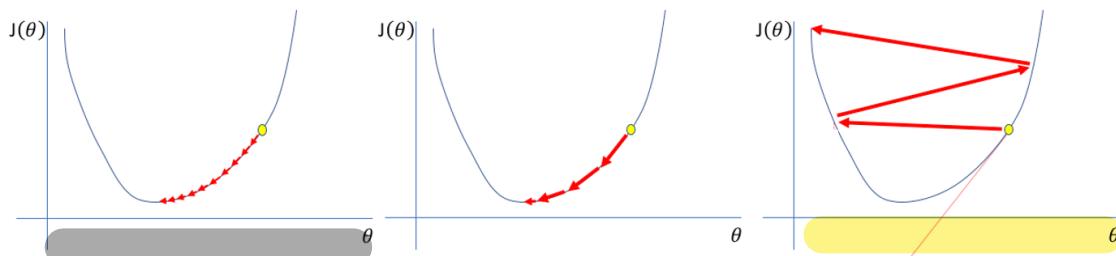
2. Función de coste

Un operador $\frac{\partial}{\partial x} f(x)$ define la derivada parcial de $f(x)$ en función de x . Un parámetro que merece especial atención es el denominado como tasa de aprendizaje, más conocido por su traducción en inglés, **learning rate, α** .

Tasa de aprendizaje

La tasa de aprendizaje decide, junto al valor de la función de coste, la magnitud de la actualización, es decir, una vez encontrada la dirección de mínimo gradiente, α decide cómo de grande será el vector o, lo que es lo mismo, el decremento.

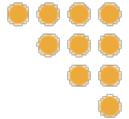
Puede parecer obvia la idea de que, a mayor tasa de aprendizaje, menor número de iteraciones necesarias para alcanzar el mínimo, pero es algo más complejo que eso. Si se interpretan las distintas actualizaciones del algoritmo del gradiente descendente aplicado a una función univariable, se pueden extraer tres posibles escenarios, en función del tamaño de la tasa de aprendizaje.



En la gráfica de la izquierda, se muestra un caso en el que la tasa de aprendizaje fue fijada con un valor muy pequeño. En este caso, el algoritmo necesita muchas iteraciones hasta llegar al mínimo, lo que lo hace un tanto ineficiente.

En la gráfica de la derecha se tiene el caso contrario; al elegir un valor demasiado grande para la tasa de aprendizaje, puede darse el caso en el que se llegue a un punto cercano al mínimo y, debido al tamaño de la tasa de aprendizaje y a la variación de gradientes en cada punto, la siguiente actualización sobreponse el mínimo y comience a incrementar de nuevo la función de coste, alcanzando una inestabilidad que difícilmente conseguirá un buen resultado.

Por otro lado, se entenderá con facilidad por qué el valor de la tasa de aprendizaje deberá ser siempre positivo. La fórmula de actualización de θ_i mostrada anteriormente se lleva a cabo mediante una resta del incremento, forzándolo a ser un incremento negativo o decremento. Si el valor de la tasa de aprendizaje se hace negativo la actualización proporcionaría un valor mayor al anterior y la optimización no se conseguiría nunca.

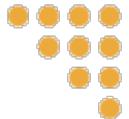


2. Función de coste

Por esta razón, la elección de la tasa de aprendizaje requiere conocimiento previo sobre el problema a tratar y, además de eso, existe en numerosos heurísticos o algoritmos a partir de los cuales se puede obtener tasas de aprendizaje adaptativas a cada iteración.

También, hay que notar que, a medida que los valores de las variables θ_i se acercan al óptimo, el error cometido en la estimación del modelo disminuye y, por lo tanto, las actualizaciones van siendo cada vez de menor tamaño. De hecho, por el principio del cálculo infinitesimal, podría ocurrir que jamás se alcanzase el óptimo debido a que las actualizaciones tomarían valores infinitamente pequeños ante errores infinitamente pequeños.

Es por esto que en la práctica es necesario adoptar una técnica heurística para calcular el **punto de detención**. Típicamente, se establece un valor mínimo en el incremento de las variables θ_i por debajo del cual se considerará alcanzada la convergencia del problema y se detendrá el algoritmo. Se suelen escoger valores de tasa de aprendizaje en variaciones logarítmicas, es decir 0,1, 0,01, 0,001, etc.

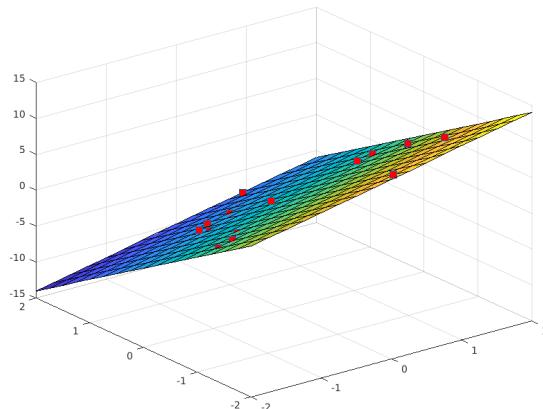


3. Regresión multivariable

Tal y como se adelantaba anteriormente, en la mayoría de casos reales, se querrá aprovechar toda la información disponible para obtener conocimiento y ésta vendrá asociado a distintas variables, por lo que el problema se complica ligeramente.

Por establecer una notación formal, se tiene un conjunto de variables de entrada $X_n = (x_1, x_2, \dots, x_n)$ y una variable de salida y . La dimensión n del espacio vectorial X determinará el número de entradas del sistema.

Si para el caso simple, el modelo de regresión se construye a partir de la función de la recta que regía su comportamiento, para un caso con dos variables de entrada, X_1 y X_2 , se espera obtener un plano.



y la ecuación cuyos coeficientes sería necesario estimar tomaría la forma

$$h(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

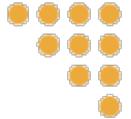
Para una generalización de n variables de entrada, sería intratable mediante espacios cartesianos, ya que el problema encajaría en un hiperespacio n -dimensional, y la función a estimar tomaría la forma

$$h(x) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n = \sum_{i=0}^n \theta_i x_i,$$

donde, por facilitar expresiones, de aquí en adelante se considerará la variable x_0 con valor fijo $x_0 = 1$.

Con el fin de tratar tal tamaño de variables de una manera más ordenada, estas se abstraen en matrices vector de tamaño $(n + 1)$ del siguiente modo

$$\mathbf{x} = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^{n+1}, \quad \boldsymbol{\theta} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix} \in \mathbb{R}^{n+1}$$



3. Regresión multivariable

por lo que la anterior función puede ser reescrita como

$$h(x) = \boldsymbol{\theta}^T \mathbf{x} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix}^T \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = [\theta_0 \ \theta_1 \ \theta_2 \ \dots \ \theta_n] \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$
$$= \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n$$

3.1. Gradiente descendente con múltiples variables

Si se piensa ahora en la función de coste, ésta tomará la forma

$$J(\theta_0, \theta_1, \dots, \theta_n) = J(\boldsymbol{\theta}) = \sum_{i=1}^m (h(x_i) - y_i)^2$$

Entonces, para llevar a cabo el algoritmo del gradiente descendente, será necesario actualizar cada una de las n variables, θ_i , en cada iteración. Esto se expresa como

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1, \dots) = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\boldsymbol{\theta})$$

Por muy complejo que pueda parecer la derivada parcial de una función como la función de coste, realmente la variable sobre la que se deriva cada vez sólo aparece en un solo término. Esto quiere decir que, dado el sumatorio de la función de coste

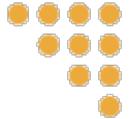
$$J(\boldsymbol{\theta}) = \sum_{i=1}^m (h(x_i) - y_i)^2$$

si se desarrolla, se reescribe como

$$J(\boldsymbol{\theta}) = \sum_{i=1}^m (\theta_0 x_0 + \theta_1 x_1 + \dots + \theta_n x_n - y_i)^2,$$

es decir la ecuación que encierra el sumatorio, sumada una y otra vez para las m observaciones de la muestra. Cuando se deriva una ecuación dentro de un sumatorio, si la progresión de ese sumatorio no involucra la variable a derivar, es decir, el sumatorio no usa a x para su definición, por ejemplo, $\sum_{i=1}^m x_i^2$, el sumatorio seguirá intacto y sólo será necesario operar con la ecuación que encierra.

Además de esto, si se necesita calcular la derivada parcial de $J(\boldsymbol{\theta})$ con respecto a, por ejemplo, θ_3 , el resto de variables $\theta_1, \theta_2, \theta_4, \dots, \theta_n$ se supondrán constantes y, como tal, su derivada será igual a cero.



3. Regresión multivariable

Resumiendo, para calcular la derivada parcial con respecto a θ_j de la función de coste sólo se tendrá en cuenta el término $\theta_j x_j$ y, como la derivada de una composición del tipo $f[g(x)]$ es

$$f[g(x)]' = g'(x)[f(x)]'$$

$$\rightarrow \frac{\partial}{\partial \theta_j} \left[(\theta_0 x_0 + \theta_1 x_1 + \dots + \theta_2 x_2)^2 \right] = x_j \cdot 2(\theta_0 x_0 + \dots + \theta_2 x_2)$$

se concluye que la derivada parcial buscada queda como

$$\frac{\partial}{\partial \theta_j} J(\boldsymbol{\theta}) = 2 \sum_{i=1}^m (h(\mathbf{x}_i) - y_i)^2 \cdot x_{ij}$$

Entonces, en cada iteración, será necesario realizar n actualizaciones como estas

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\boldsymbol{\theta}) = \theta_j - 2\alpha \sum_{i=1}^m (h(\mathbf{x}_i) - y_i) \cdot x_{ij}$$

Tenga en cuenta que la constante 2 no influye en un proceso iterativo donde se está multiplicando por un coeficiente configurable, por lo que no se tendrá en cuenta.

Desde un conjunto de valores iniciales para $\boldsymbol{\theta} = \{\theta_0, \theta_1, \dots, \theta_n\}$, representado como un punto en el hiperespacio n -dimensional, en cada iteración se obtendrán nuevos valores que formarán otro punto más cercano al mínimo buscado.

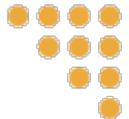
3.2. Normalización

Habrá veces en las que los rangos de valores de las distintas variables de entrada de un problema difieran en escala y, mientras los valores de la variable x_1 se muevan en un rango de valores de $[0,10]$, la variable x_2 lo haga entre uno de $[100,1000]$.

Esto provoca que, durante la construcción del modelo de regresión, se le de más importancia a las variables que manejan valores más grandes.

Gráficamente, el vector obtenido para la actualización en una iteración determinada, llevaría una orientación muy próxima al eje de la variable con rango de valores mayores.

Esto sucede debido a que la magnitud de la modificación que produce la ecuación de actualización a la nuevas variables depende de tres parámetros: la tasa de aprendizaje, α , común para todas; la suma de errores o residuos, también común



3. Regresión multivariable

para todas las variables ; y la variable x_j , propia de cada variable j .

Por lo tanto, si los valores de una determinada variable i son sustancialmente mayores que los de otra variable j , los saltos de valor en cada iteración serán mucho mayores para θ_i que para θ_j , y la orientación final del vector sobre el espacio n -dimensional de valores que apunta hacia el siguiente iteración primará la proyección sobre i por encima de j .

Esto reduce la eficiencia del algoritmo ya que las actualizaciones de valor hacen mover el punto del estado actual hacia otro en una dirección que no es la óptima.

Para corregir esto, se lleva a cabo un escalado de valores llamado **normalización**, la cual transforma los valores de todas las variables a un rango común a todas manteniendo su distribución. Existen varias técnicas de normalización entre las que destacan:

- **MinMax.** Es la más común de todas, escala los valores tomando como extremos el máximo y el mínimo dentro del rango de valores de la variable. Transforma los valores de la variable x mediante

$$z = \frac{x - \min(x)}{\max(x) - \min(x)}$$

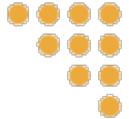
Los valores pasan a tomar valores entre 0 y 1. Se debe tener en cuenta que no proporciona buenos resultados cuando existen valores atípicos, ya que aparecerán la mayoría de valores muy cercanos a alguno de sus extremos.

- **Zscore o normalización estándar.** Esta técnica hace uso de los principales parámetros estadísticos de la distribución de valores de la variable para normalizarlos

$$z = \frac{x - \mu_x}{\sigma_x},$$

siendo μ_x y σ_x la media y desviación estándar de x , respectivamente. Esta técnica transforma los valores de la variable a otros en el intervalo -1 y 1 .

Además, evita el problema de ser distorsionado debido a los valores atípicos y, de hecho, se puede usar como método para rechazar estos valores atípicos, por ejemplo, deshaciéndose de todas las observaciones cuyo valor quede fuera del rango $[-1,1]$.



3. Regresión multivariable

Se tienen los siguientes datos disponibles sobre distintas parcelas en venta

Lejanía (m)	Anchura (m)	Longitud (m)	Calidad suelo	Precio (€)
1300	6,57	8,2	0,56	25000
1750	35,4	90	0,87	20000
2800	135	120	0,4	80500
5000	8	35	0,75	14000
6,00E+08	23	20	0,5	23550
8500	18	15	0,2	8000

La lejanía indica, en metros, la distancia desde el municipio más cercano, la calidad del suelo, por su parte, viene medida a partir de un indicador que va de 0 a 1, en el mejor caso. Hará falta un normalizado para llevar todas las variables a un rango de valores parecido o del mismo orden. La variable *Calidad Suelo* no es necesaria que se transforme, ya que sus valores van entre 0 y 1. Tras un normalizado tipo *MinMax*, se obtiene la siguiente tabla

Lejanía (m)	Anchura (m)	Longitud (m)	Calidad suelo	Precio (€)
0,00000000	0,000	0,000	0,537	25000
0,00000075	0,224	0,732	1,000	20000
0,00000250	1,000	1,000	0,299	80500
0,00000617	0,011	0,240	0,821	14000
1,00000000	0,128	0,106	0,448	23550
0,00001200	0,089	0,061	0,000	8000

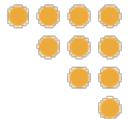
Los valores de la variable *lejanía* han quedado todos desplazados a niveles muy cercanos a cero, a excepción de uno, eso es debido a la presencia de valores atípicos. Si se intenta, ahora, un normalizado tipo *zscore*, los valores quedan como en la siguiente tabla

Lejanía (m)	Anchura (m)	Longitud (m)	Calidad suelo	Precio (€)
-0,408258782	6,57	8,2	0,56	25000
-0,408256945	35,4	90	0,87	20000
-0,408252659	135	120	0,4	80500
-0,408243677	8	35	0,75	14000
2,041241452	23	20	0,5	23550
-0,408229388	18	15	0,2	8000

Es tal la diferencia de los valores atípicos con los regulares que la media se altera lo suficiente como para obtener una desviación estándar muy grande. Sin embargo, la normalización ha servido para detectar los valores atípicos marcando como tales aquellos que superan el rango [-1,1]. Si se vuelve hacer el normalizado *MinMax*, esta vez sin la observación con el valor atípico, se obtiene la tabla

Lejanía (m)	Anchura (m)	Longitud (m)	Calidad suelo	Precio (€)
0,00	0,000	0,000	0,537	25000
0,06	0,224	0,732	1,000	20000
0,21	1,000	1,000	0,299	80500
0,51	0,011	0,240	0,821	14000
1,00	0,089	0,061	0,000	8000

Ya preparada para ser usada en para crear el modelo.



4. Regresión no lineal

Además de añadiendo varias variables al problema, se puede complicar aun más probando con relaciones no lineales. Y es que es aquí donde se aprovecha todo el potencial del gradiente descendente y, de algún modo, la estadística inferencial y el aprendizaje automático, rama de la inteligencia artificial, se solapan.

En principio, pueden existir tantos (y más) tipos de regresión como formas pueda adoptar una función: lineal, exponencial, logarítmica, polinómica, etc. Por su importancia, se explicarán la regresión polinomial y la regresión logística.

4.1. Regresión polinomial

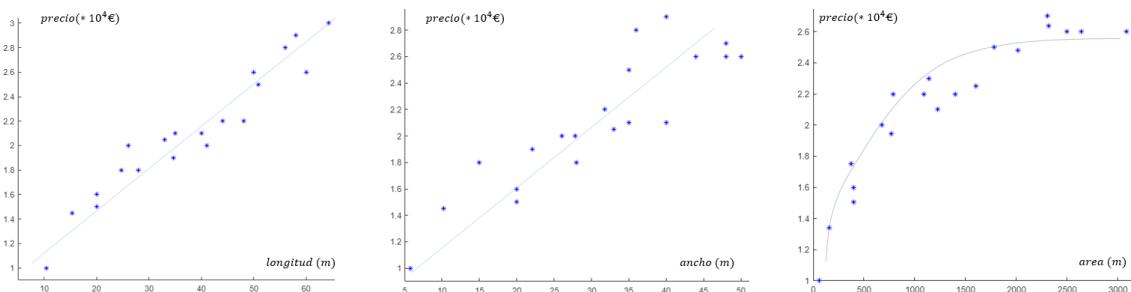
Imagine ahora que el problema del estimador de precios de suelo viene alimentado con los datos del ancho y del largo, en metros. La función que se pretendería ajustar al problema vendría expresado como

$$h(x) = \theta_0 + \theta_1 \cdot (\text{ancho}) + \theta_2 \cdot (\text{largo})$$

Sin embargo, existen ciertos donde la combinación de esas dos variables actúa mejor como variable de entrada que las dos por separado. Si se asigna $x_1 := \text{ancho}$ y $x_2 := \text{largo}$ y $x_1 x_2 := \text{area}$, se puede reescribir la función anterior con el área de la parcela

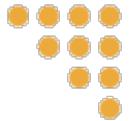
$$h(x) = \theta_0 + \theta_1 x_1 x_2$$

Suponiendo la relación lineal entre el precio de la parcela, $h(x)$ o y , y las variables x_1 y x_2 , por separado, ahora existirá una relación cuadrática entre el precio y el área.



El caso presentado está bien definido y se puede saber a priori el comportamiento cuadrático que va a tener la relación entre precio y área. Sin embargo, cuando esta información no se tiene y se presentan datos como los mostrados en la gráfica de la derecha, es lógico pensar pensar que sea posible ajustar los datos a una ecuación cuadrática del tipo

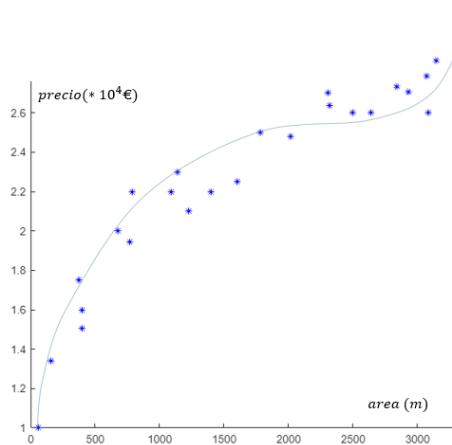
$$h(x) = \theta_0 + \theta_1 x + \theta_2 x^2,$$



4. Regresión no lineal

aunque quizás, al recolectar nuevos datos de áreas mayores, es posible apreciar cómo el precio vuelve a recuperar pendiente y crecer de manera y se complica hacia lo que podría ser la trayectoria seguida por una función cúbica como

$$h(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3$$



Es posible manejar este problema como lineal mediante ciertas transformaciones que proporcionan una rectificación de un vector de variables que encaje en la función cúbica anterior

$$x_1 := x = \text{area}, \quad x_2 := x^2 = (\text{area})^2, \quad x_3 := x^3 = (\text{area})^3,$$

lo que ayudaría transformar la función cúbica en una aparentemente lineal con la que trabajar.

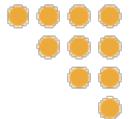
$$h(x) = \theta_0 + \theta_1 x + \theta_2 x_2 + \theta_3 x_3$$

No existe un método universal que indique cómo llevar a cabo las combinaciones entre variables cuando se tienen más de una para obtener esta función, tampoco hay que realizar estas transformaciones si no es necesario.

Si se estudia más en profundidad la regresión y el aprendizaje automático, se descubren ciertas limitaciones o problemas que se pueden encontrar durante el proceso iterativo de la regresión, y estas transformaciones podrían solventarlas. Por supuesto, dado el origen de estas variables, con total seguridad se hará necesaria la normalización de sus valores.

4.2. Regresión logística

La regresión logística abre un campo de posibilidades en la regresión, ya que permite trabajar con variables cualitativas como variables de salida. Esto lo consigue asumiendo la variable como una categórica binaria que sigue una distribución



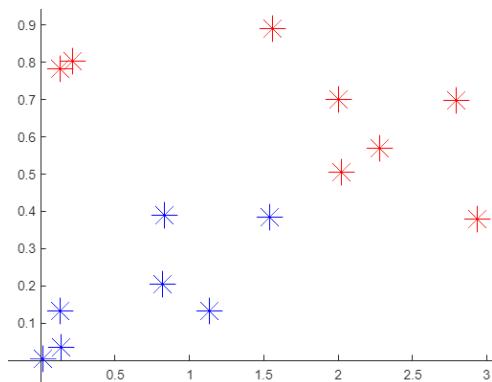
4. Regresión no lineal

binomial. Es decir, la variable de salida tendrá tan sólo dos posibles valores: 1 ó 0, que podrían ser valores asignados a dos variables categóricas. Es por esto que el modelo de regresión logística es ampliamente usado para tareas de **clasificación**.

La regresión logística trata de correlar las probabilidades de una variable cualitativa a partir de un conjunto de variables de entrada. Consigue crear un sistema que proporciona salidas tipo sí/no.

Para un estudio en el que se pretende crear un modelo de clasificación de imágenes de melanomas, se extraen dos características de las fotografías de los melanomas. Estos componen las dos variables de entrada: un indicador de texturas normalizado, representando los melanomas de apariencia más homogénea con un valor cercano a 1, y el radio de la circunferencia circunscrita al melanoma, en centímetros.

La variable de salida definida es $y \in \{1 := tumor, 0 := no\ tumor / benigno\}$. Para una más sencilla representación de los datos disponibles se usan distintos marcadores para la variable y : los valores 1 se representarán con marcas rojas y los valores 0 con marcas azules.

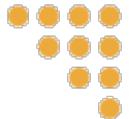


Está estandarizado el denominar al suceso asociado a $y = 1$ **clase positiva**, y al contrario, $y = 0$, **clase negativa**.

En la regresión convencional, se busca una función que nos proporcione nuevos valores desconocidos de la variable de salida a partir de datos de las variables de entrada. Para esto, se lanzaba la hipótesis del espacio de valores de salida contenidos en esta función

$$h(\theta) = \theta^T x = \theta_0 x_0 + \theta_1 x_1 + \dots + \theta_n x_n$$

En clasificación, se necesita una función que divida, de algún modo, el espacio de valores n -dimensional, siendo n el número de variables de entrada, en tantas clases



4. Regresión no lineal

como sucesos posibles, que en caso de clasificación binaria es igual a dos. Además, es necesario que el valor de salida que indica la pertenencia a alguno de los dos conjuntos en los que se divide el espacio de valores se reduzca al intervalo que las probabilidades maneja, es decir, $[0,1]$.

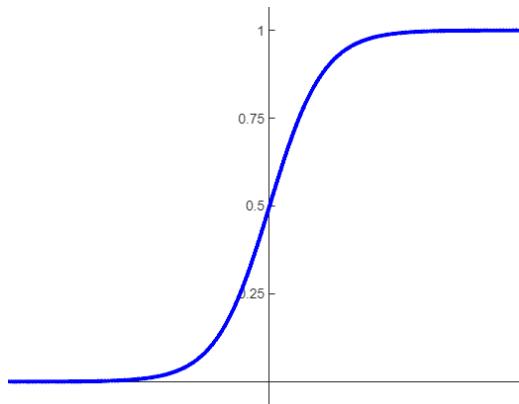
Para esto, se convoluciona la función hipótesis de la regresión, $h(\theta)$, así

$$h(\theta) = g(\theta^T x)$$

Esta función será la **función sigmoide** o **función logística** y toma la forma

$$g(\tau) = \frac{1}{1 + e^{-\tau}} \rightarrow h(\theta) = \frac{1}{1 + e^{-\theta^T x}}$$

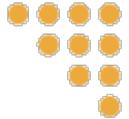
La curva representada por la función sigmoide es



Debido a la asunción de la distribución del problema como una binomial, la función hipótesis en este caso estará proporcionando un valor entre 0 y 1 igual a la probabilidad de la hipótesis nula correspondiente al éxito del suceso $y = 1$. Esto se puede escribir como

$$h(\theta) = P(y = 1|x, \theta)$$

Entonces, para la nube de puntos mostrada en el ejemplo anterior del clasificador de tumores, $h(\theta)$ tomará la forma de un curva en el espacio cartesiano que separará de la mejor manera posible las clases positivas de las negativas. Observando la gráfica, se ve que no sería imposible que una recta separase los dos casos pero no hay que olvidar que, en la práctica, los problemas toman comportamientos altamente no lineales y se tienen en cuenta muchas variables de entrada, a veces del orden de cientos o miles. Sea una recta, una curva, un plano, una superficie, un hiperplano o una hipersuperficie, la función obtenida, $h(\theta)$, ya no será una recta de regresión de donde recoger valores estimados, sino una **frontera de decisión**.



4. Regresión no lineal

4.2.1. Función de coste logística

Del mismo modo que el resto de tipos de regresión, la regresión logística o clasificación también necesita una función con la que pueda autoevaluarse y actualizar sus valores en consecuencia, para la siguiente iteración.

Pero en este caso, debido a la no linealidad de la función logística, $h(\theta)$, cuando se intenta integrar en la función de coste definida para la regresión lineal, la de mínimos cuadráticos,

$$J(\theta) = \sum_{i=1}^m (h(x_i) - y_i)^2 = \sum_{i=1}^m \left(\frac{1}{1 + e^{-\theta_i x_i}} - y_i \right)^2,$$

se obtiene una curva muy oscilante con muchos mínimos locales. Esto, formalmente, es llamado función no convexa y es algo muy problemático a la hora de resolver un problema de optimización. Para evitar esto, y tener una función que garantice en mayor medida la localización del mínimo global, se diseña una función de coste distinta.

$$J(\theta, y) = \begin{cases} -\log(h(x_i)), & y = 1 \\ -\log(1 - h(x_i)), & y = 0 \end{cases}$$

De esta manera, si una observación es realmente 1, un tumor maligno, la función de coste $J(\theta, y)$ aumentará a medida que $h(x_i)$ se aproxime a 0

$$-\log(0.3) = 0.52, \quad -\log(0.0001) = 4, \quad -\log(0.0000001) = 7,$$

y se acercará a 0 a medida que $h(x_i)$ se aproxime al verdadero valor de $y, 1$

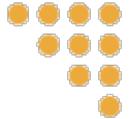
$$-\log(0.7) = 0.15, \quad -\log(0.9) = 0.046, \quad -\log(0.99) = 0.0044.$$

En caso de que se evalúe la observación de un caso benigno o de no tumor, $y = 0$, se tendrá, para estimaciones que se acercan al valor verdadero, un coste decreciente

$$-\log(1 - 0.3) = 0.15, \quad -\log(1 - 0.01) = 0.004, \quad -\log(1 - 0.0001) = 0.00004$$

y para casos que se alejan del valor verdadero, un coste creciente

$$-\log(1 - 0.3) = 0.52, \quad -\log(1 - 0.9) = 1, \quad -\log(1 - 0.99) = 2$$



4. Regresión no lineal

Puesto que resultaría demasiado incómodo tratar con una función a trozos como la de arriba, la función de coste se forma uniendo sus partes y resulta

$$J(\theta, y) = -y \log(h(x_i)) - (1 - y_i) \log(1 - h(x_i))$$

Esta función de coste es conocida como función de coste de **entropía cruzada**, más reconocible por su traducción al inglés, **cross-entropy loss**.

Así, para casos de $y = 0$, la ecuación se reducirá a

$$J(\theta) = -0 \cdot \log(h(x_i)) - (1 - 0) \log(1 - h(x_i)) = -\log(1 - h(x_i)),$$

y para casos $y = 1$, se tendrá

$$J(\theta) = -1 \cdot \log(h(x_i)) - (1 - 1) \log(1 - h(x_i)) = -\log(h(x_i))$$

4.2.2. Gradiente descendente para el caso logístico

El algoritmo optimizador gradiente descendente trabajará del mismo modo que lo hacía con otros tipos de regresión, a diferencia de la función de coste, que habrá cambiado. Sin embargo, tras derivar parcialmente la nueva función de coste, se obtiene exactamente la misma ecuación que con la anterior función de coste.

$$\theta_j := \theta_j - \frac{\partial}{\partial \theta_j} J(\theta) = \theta_j - \alpha \sum_{i=1}^m (h(x_i) - y_i) x_{ij},$$

donde $h(x_i) = \frac{1}{1+e^{-\theta_i x_i}}$

4.3. Métricas de evaluación

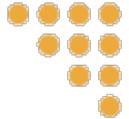
Con lo visto hasta ahora, se tiene una algoritmo que minimiza la función cross-entropy de un problema de regresión logística para obtener un modelo capaz de devolver la probabilidad de pertenencia de la nueva observación de entrada a la clase positiva, esto es, tenga $y = 1$.

4.3.1. Umbral de decisión

Es inmediato pensar que, para considerarlo $\hat{y} = 1$, deba cumplir $h(x) \geq 0.5$. Así,

$$h(x) \geq 0.5 \rightarrow \hat{y} = 1$$

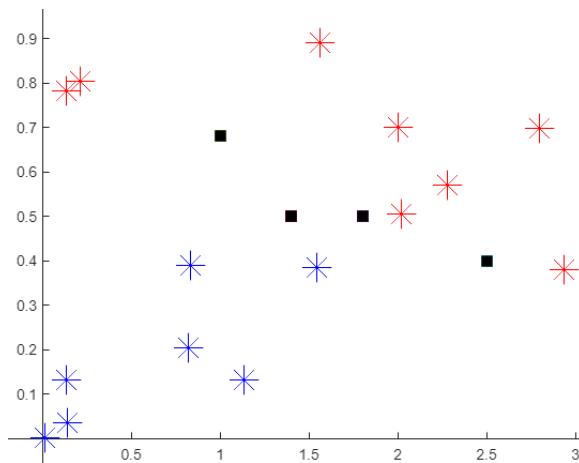
$$h(x) < 0.5 \rightarrow \hat{y} = 0$$



4. Regresión no lineal

Puede parecer lo más razonable, y para el caso genérico lo es, si embargo, no siempre será así. En cualquier caso, es importante conocer cómo impacta en el resultado la variación de este **umbral de decisión**.

Volviendo al ejemplo del clasificador de tumores, donde se tenían 10 datos de resultados conocidos, supóngase que, a partir de estos datos conocidos, se ha construido un modelo clasificador y se quiere testear con los datos que aparecen dibujados en negro sobre la gráfica. Estos no han sido vistos por el modelo durante la optimización del modelo, pero se conocen sus resultados para así poder evaluar la bondad del clasificador.



De las nuevas observaciones x_1, x_2, x_3 y x_4 , se conocen sus verdaderos valores $y_1 = 0, y_2 = 1, y_3 = 1$ y $y_4 = 0$. El modelo devuelve las probabilidades

$$h_1(x) = 0,28, \quad h_2(x) = 0,46, \quad h_3(x) = 0,8, \quad h_4(x) = 0,49$$

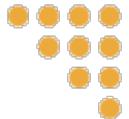
Para un umbral de decisión de 0,5, de dos casos de tumor maligno presentes en el conjunto de test, ha acertado uno, x_3 . En estos casos, puede convenir reajustar el umbral de decisión para suavizarlo y así aumentar la detección.

El caso no detectado, x_2 , obtuvo una probabilidad de ser clase positiva de 0,46, por lo que se prueba reduciendo el umbral de decisión a 0,4. Con este nuevo valor, se tiene

$$h(x) \geq 0,4 \rightarrow \hat{y} = 1$$

$$h(x) < 0,4 \rightarrow \hat{y} = 0$$

$$\hat{y}_1 = 0, \quad \hat{y}_2 = 1, \quad \hat{y}_3 = 1, \quad \hat{y}_4 = 1$$



4. Regresión no lineal

Se ha conseguido así corregir esa falta de precisión y ahora detecta todos los casos de tumor maligno.

Sin embargo, el caso de tumor benigno, x_4 , entra ahora en el grupo de los estimados como detección, lo cual es también erróneo.

Generalmente, siempre ocurrirá este balanceo de resultados: a mayor umbral de decisión, el clasificador será más estricto y probablemente no detecte todos los casos realmente positivos, aunque asegurará no cometer tantas detecciones erróneas; por otro lado, cuando se disminuye el valor del umbral de decisión, se afloja la restricción de detección, se asegura más detecciones válidas, pero a costa de un mayor riesgo de tener detecciones erróneas.

Como se puede comprobar, la tasa de detecciones o aciertos no es el único medidor de un clasificador. Según el problema, habrá que definir si se prioriza la máxima detección posible a costa de dejar pasar clases negativas detectadas como positivas, o aumentar la restricción y priorizar así que el número de aciertos, aunque sea menor, sean todos detecciones de casos reales.

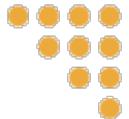
4.3.2. Matriz de confusión

La matriz de confusión sirve al analista para reflejar todos los resultados y poder visualizar con facilidad el desempeño del clasificador. Se denomina matriz de confusión, debido a que separa en cuatro celdas representando los cuatro posibles casos de resultado, dos tipos de aciertos y dos tipos de confusiones.

Está compuesto por dos filas y dos columnas: las dos filas se referirán a los valores reales y las columnas a los valores estimados.

Así, para cada fila, se dividirán las observaciones estimadas entre los dos casos posibles: si se estimó como $\hat{y} = 1$ o como $\hat{y} = 0$. Las cuatro celdas deben reunir en total el número de observaciones estimadas, la suma de los elementos de una fila determina el número total de observaciones pertenecientes a la clase correspondiente a la que indica la fila y la suma de los elementos de una columna es igual al número de observaciones predichas como pertenecientes a la clase correspondiente que indica dicha columna.

El número total de aciertos, positivos o negativos, es igual a la suma de los



4. Regresión no lineal

elementos de la diagonal. Así, la matriz de confusión

		Valor predicho	
		$\hat{y} = 1$	$\hat{y} = 0$
Valor Real	$y = 1$	25	13
	$y = 0$	30	47

indica que, de un total de 115 observaciones testeadas, 38 de clase positiva (fila 1) y 77 de clase negativa (fila 2), ha habido un total de 72 observaciones correctamente clasificadas (suma diagonal): 25 de clase positiva y 47 de clase negativa. Además, del resto de observaciones, 13 observaciones han sido clasificadas como clase negativa cuando realmente son clase positiva y 30 han sido clasificadas como positiva cuando eran pertenecientes a la clase negativa.

Existen nombres para cada uno de estos resultados, además de métricas para obtener una idea más resumida sobre el desempeño del modelo clasificador.

- **Verdaderos positivos (TP, true positives):** Cuenta los casos realmente positivos que fueron clasificados como tal. Corresponde con la celda (1,1) de la matriz.
- **Verdaderos negativos (TN, true negatives):** Aquellos casos realmente negativos que fueron clasificados como tal. Corresponde con la celda (2,2) de la matriz.
- **Falsos positivos (FP, false positives):** Son casos mal clasificados que, siendo realmente negativos que fueron clasificados positivos. También se puede encontrar menciones a este dato como *falsas alarmas*. Corresponde con la celda (2,1) de la matriz.
- **Falsos negativos (FN, false negatives):** Cuenta los casos realmente positivos que fueron clasificados negativos. Corresponde con la celda (1,2) de la matriz.

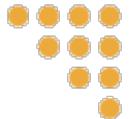
A partir de estos datos, se obtienen una serie de rato que dan distintos puntos de vista acerca de cómo está funcionando el modelo. Los más importantes son

- **Exactitud (Accuracy).** Es el ratio que da la visión más general de todas, ya que calcula el porcentaje de aciertos, tanto positivos cuando eran positivos como negativos cuando eran negativos, entre el total.

$$\text{Exactitud} = \frac{n^{\circ} \text{ pred. correctas}}{\text{total}} = \frac{TP + TN}{TP + TN + FP + FN}$$

Con los datos de la matriz mostrada anteriormente, se tendría una exactitud de

$$\frac{TP + TN}{TP + TN + FP + FN} = \frac{25 + 47}{115} = 62,6\%$$



4. Regresión no lineal

- **Precisión (Precision).** Con esta métrica se obtiene la cantidad de aciertos positivos del total de predicciones positivas. Para el ejemplo de tumores, se obtendría el porcentaje de detecciones de tumores malignos realizados correctamente de entre todos los casos que se detectaron como tumores malignos. Se expresa como

$$\text{Precision} = \frac{n^{\circ} \text{ pred. positivas correctas}}{n^{\circ} \text{ predicciones positivas}} = \frac{TP}{TP + FP}$$

En el caso de la matriz, se tiene

$$\frac{TP}{TP + FP} = \frac{25}{25 + 30} = 45,45\%$$

- **Exhaustividad (Recall).** Esta métrica puede ser considerado como el complementario a la precisión pues, siendo igual de importante, representa el porcentaje de casos positivos detectados entre el total de casos realmente positivos.

Es muy importante pues, en el caso de la detección de tumores malignos, estaría indicando la capacidad del modelo en detectar casos reales.

Se calcula así

$$\text{Exhaustividad} = \frac{n^{\circ} \text{ pred. positivas correctas}}{n^{\circ} \text{ casos positivos}} = \frac{TP}{TP + FN}$$

Para el caso de ejemplo se tendría

$$\frac{TP}{TP + FN} = \frac{25}{25 + 13} = 65,8\%$$

Como se veía antes cuando se manipulaba el umbral de decisión, para valores altos del umbral, el modelo sería más estricto a la hora de declarar como detección una observación y, por lo tanto bajaría la precisión pero, por el mismo motivo, tampoco daría por detecciones positivas tantos casos que realmente no lo fuesen, falsos negativos, lo que haría aumentar la exhaustividad. Si baja el umbral de decisión, el predictor da mas observaciones como predicciones positivas, lo cual incluirá casos realmente positivos, verdaderos positivos y aumentará la precisión, pero también incluirá casos que no son positivos, lo que reducirá la exhaustividad.

El analista debe conocer el problema y priorizar, ¿Qué es más importante, detectar todo o lo máximo posible aunque así detecte como positivas observaciones que no lo son, o no detectar tantos casos con tal de que lo detectado sea realmente positivo? Este análisis es llamado en inglés *precision/recall tradeoff*.

SAMSUNG

BeJob[™]

SAMSUNG

Together for Tomorrow!
Enabling People

BeJobSM