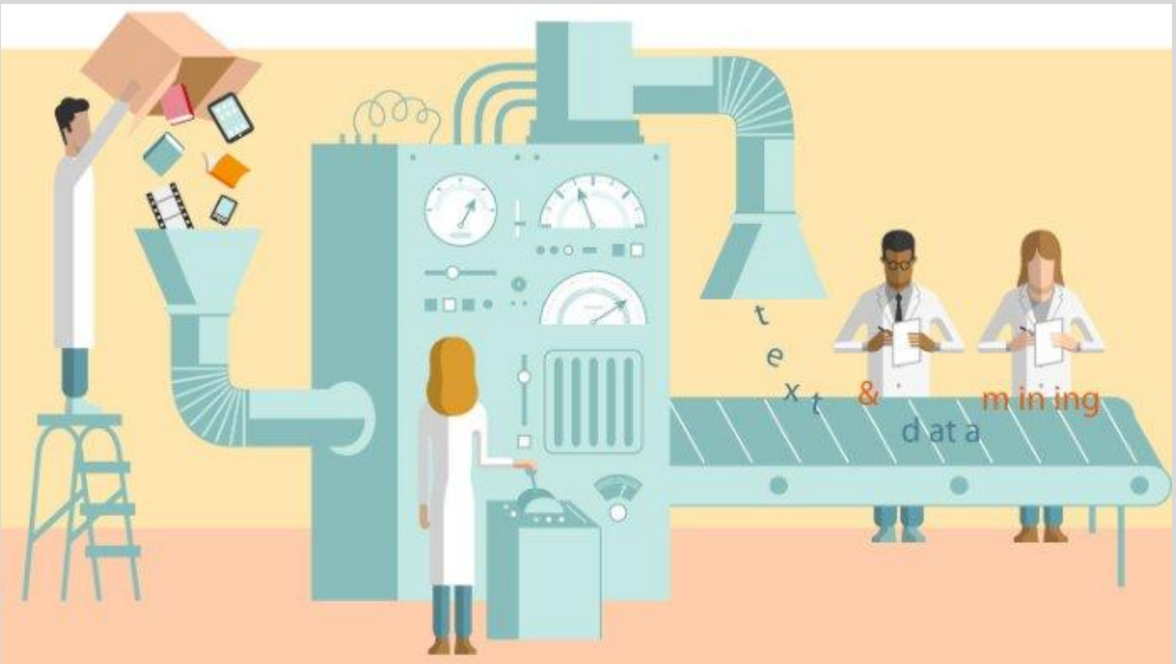


DigitalHouse >
Coding School

DATA SCIENCE

UNIDAD 1

Conceptos de
Probabilidad

AGENDA

1. Experimento Aleatorio - Probabilidad Clásica
2. Probabilidad Condicional - Regla de Bayes
3. Variables Aleatorias - Funciones de Distribución

- Definimos como **experimento** a cualquier proceso que produce una **observación** o resultado.

- Un **experimento determinístico**: cada vez que se realiza bajo condiciones similares, arroja el mismo resultado. Por ejemplo, conociendo la altura desde la que se arroja un objeto es posible saber exactamente el tiempo que tardará en llegar al suelo en condiciones de vacío.

- Un **experimento aleatorio** es un experimento tal que no se tiene certeza de su resultado hasta que no se lo ejecute.
  - Es posible conocer a priori el conjunto de todos los posibles resultados del experimento pero no puede predecirse con exactitud el resultado de un experimento en particular.
  - Ejemplo: lanzamiento de dados, lanzamiento de una moneda, extracción de una carta de una baraja, etc.

- **Espacio Muestral**
  - Conjunto de **todos los resultados posibles** de un experimento aleatorio.
  - Se lo denota con la letra griega Ω
  - A un resultado particular del experimento se lo denota con ω

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

- Ejemplo

Un experimento aleatorio consiste en lanzar un dado y observar el número que sale. El espacio muestral es claramente Ω = {1, 2, 3, 4, 5, 6}.

Como ejemplo de evento, podríamos definir el conjunto A = {1, 3, 5}, que corresponde al suceso de obtener como resultado un número impar.

Si al lanzar el dado se obtiene un 3, entonces decimos que se observó la ocurrencia del evento A.

Pero si lanzo el dado y sale 2, entonces no se observa la ocurrencia del evento A

- Ejemplo

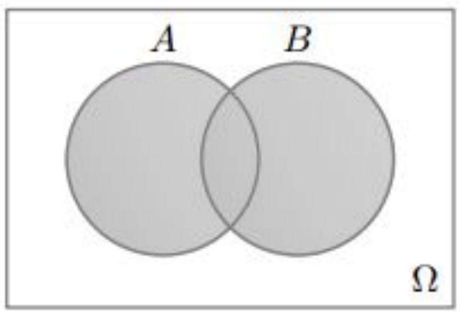
Consideremos el caso de un juego de lotería. Hay un millón de números y un jugador participa con un boleto. Sólo hay un número ganador.

Naturalmente el espacio muestral que surge es el de todos los números hasta un millón, **$\Omega = \{1, 2, ..., 10^6\}$**.

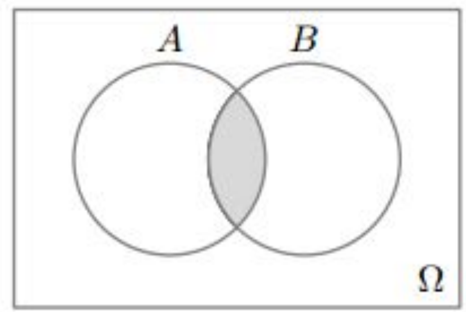
Los resultados posibles se pueden agrupar en eventos arbitrarios. Al jugador le interesa conocer su suerte en este juego. Entonces, también se pueden agrupar los resultados como definir los eventos **"ganar" o "perder" dado el número de lotería que tiene el apostador.**

Sean A y B dos subconjuntos cualesquiera de Ω. Recordamos a continuación las operaciones básicas de unión, intersección, diferencia y complemento.

$$A \cup B = \{\omega \in \Omega : \omega \in A \ \text{ o } \ \omega \in B\},$$
$$A \cap B = \{\omega \in \Omega : \omega \in A \ \text{ y } \ \omega \in B\},$$

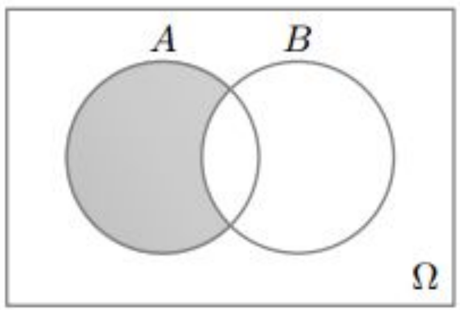


$A \cup B$

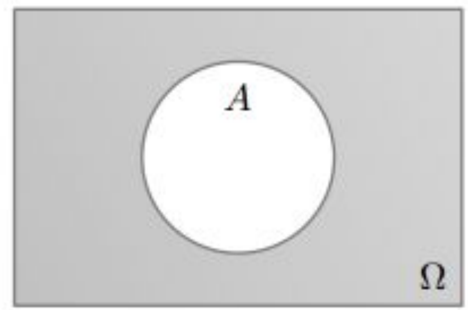


$A \cap B$

$$A - B = \{\omega \in \Omega : \omega \in A \text{ y } \omega \notin B\},$$
$$A^c = \{\omega \in \Omega : \omega \notin A\}.$$



$A - B$

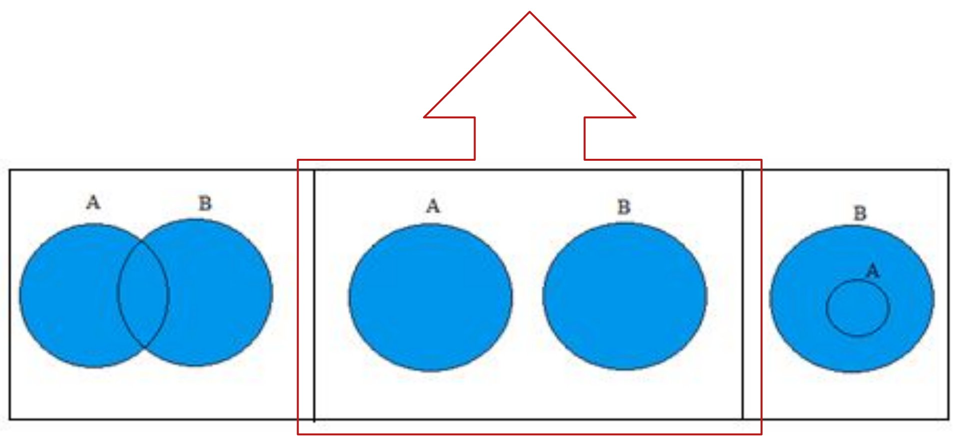


A^c

- Conjuntos disjuntos

Cuando dos conjuntos no tienen ningún elemento en común se dice que son disjuntos..

$$A \cap B = \varnothing.$$



- Ejercicio

Sean A, B y C tres eventos de un experimento aleatorio.
Exprese las siguientes oraciones en términos de estos conjuntos.

A. Ocurre A o B, pero no C.
B. Ninguno de estos tres eventos ocurre.
C. Sólo uno de ellos ocurre.
D. Exáctamente dos de ellos ocurre.
E. Por lo menos uno de ellos.
F. A lo sumo dos de ellos ocurren.

Definición 1.2 Sea A un subconjunto de un espacio muestral Ω de cardinalidad finita. Se define la probabilidad clásica del evento A como el cociente

$$P(A) = \frac{\#A}{\#\Omega},$$

en donde el símbolo #A denota la cardinalidad o número de elementos del conjunto A.

a) el espacio muestral es finito.

b) todos los elementos del espacio muestral tienen el mismo "peso".

Ejemplo 1.10 Considere el experimento aleatorio de lanzar un dado equilibrado. El espacio muestral es el conjunto $\Omega = \{1, 2, 3, 4, 5, 6\}$. Si deseamos calcular la probabilidad (clásica) del evento A, correspondiente a obtener un número par, es decir, la probabilidad de $A = \{2, 4, 6\}$, entonces

$$P(A) = \frac{\#\{2,4,6\}}{\#\{1,2,3,4,5,6\}} = \frac{3}{6} = \frac{1}{2}.$$

Fácilmente se pueden distinguir algunas propiedades

a) $P(\Omega) = 1$.

b) $P(A) \geqslant 0$ para cualquier evento A.

c) $P(A \cup B) = P(A) + P(B)$ cuando A y B son ajenos.

Cuando tenemos la posibilidad de repetir N veces un experimento aleatorio, podemos registrar la cantidad de veces que ocurre el suceso A. Esta información puede ser utilizada para calcular la probabilidad de A.

Definición 1.4 Sea n_A el número de ocurrencias de un evento A en n realizaciones de un experimento aleatorio. La probabilidad frecuentista del evento A se define como el límite

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

Como no es humanamente posible realizar un experimento infinitas veces, nos aproximamos:

$$P(A) \approx \frac{n_A}{n}.$$

https://seeing-theory.brown.edu/basic-probability/

- Usted va a comprar algo en Internet:
  - ¿Mira la calificación del vendedor?
  - ¿Su evaluación de la calidad del vendedor es la misma si éste tiene muchos años en el sitio y/o alto porcentaje de valoraciones positivas que si no lo tiene?

Calcular probabilidades condicionales es como ese proceso que usted realiza al elegir el vendedor en base a la reputación. En base a la información en la que se condiciona actualizamos nuestra creencia (probabilidad) de que el vendedor sea malo o bueno.

- Supongamos una pequeña población de vendedores en sitios de internet con los siguientes atributos.
- Para cada vendedor se registra si tiene 10 o más años vendiendo en el sitio y se sabe su tipo (si es bueno u oportunista)

	bueno	malo
10 años o más	16	8
menos de 10 años	10	16

- ¿Cuál es la probabilidad de que no sea oportunista dado que lleva 10 o más años vendiendo en el sitio?
    - Condicionar es como **fijar la fila que voy a mirar**.
    - Dado que **sé que el vendedor tiene 10 o más años**, sólo consideramos la proporción de vendedores buenos dentro de la primera fila de la tabla.

	bueno	malo	
10 años o más	16	8	24
menos de 10 años	10	16	26
	26	24	50

- En nuestro ejemplo: La probabilidad de un evento A (no ser vendedor oportunista) dado que ocurrió el evento B (tener 10 años o más vendiendo en el sitio por ejemplo)
    - Sería algo así como *los casos favorables a A (que deben ser aquellos que simultáneamente son favorables a B también porque B ya es conocido) sobre casos posibles (limitandonos a aquellos donde B ocurre).*

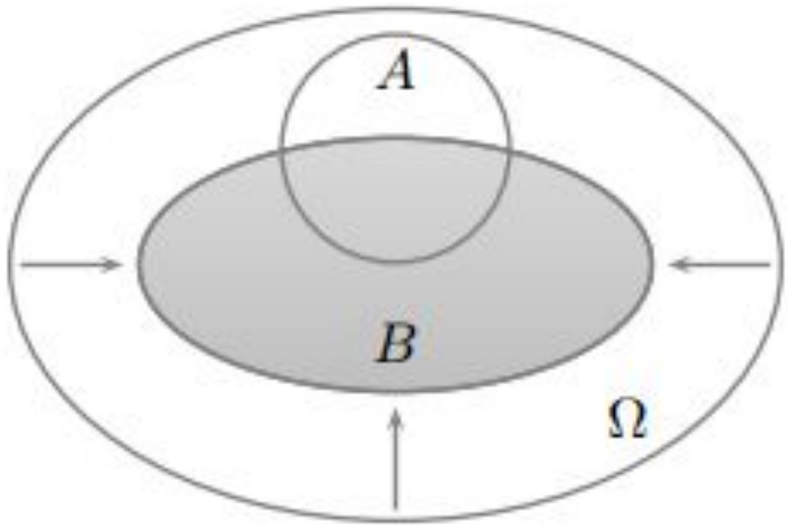
	bueno	malo	
10 años o más	16	8	24
menos de 10 años	10	16	26
	26	24	50

Definición 1.10 Sean A y B dos eventos y supongamos que B tiene probabilidad estrictamente positiva. La probabilidad condicional del evento A, dado el evento B, se denota por el símbolo $P(A \mid B)$ y se define como el cociente

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}. \tag{1.3}$$

El término P(A/B) se lee:
"La probabilidad de A dado B"

Podemos imaginar que **el espacio muestral del experimento aleatorio se ha reducido al evento B** de tal forma que todo lo que se encuentre fuera de este evento tiene probabilidad condicional cero.



- Volviendo al ejemplo de vendedores online.
  - Definamos los eventos A “bueno” y B “10 años o más vendiendo en el sitio”.
  - $P(A \text{ y } B) = 16/50$, $P(B) = 24/50$.
  - $P(B/A)=(16/50)/(24/50)=(16/50)*(50/24)=16/24$

	bueno	malo	
10 años o más	16	8	24
menos de 10 años	10	16	26
	26	24	50

- Ejemplo

Considere el experimento de lanzar un dado equilibrado y defina los eventos

A = {2} = "Se obtiene el número 2"
B = {2, 4, 6} = "Se obtiene un número par"

Es claro que P(A) = ⅙, sin embargo sabiendo que ha ocurrido B, es decir, sabiendo que el resultado es un número par, la probabilidad del evento A es ahora

$$P(A\,|\,B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{2\})}{P(\{2,4,6\})} = \frac{1/6}{3/6} = \frac{1}{3}$$

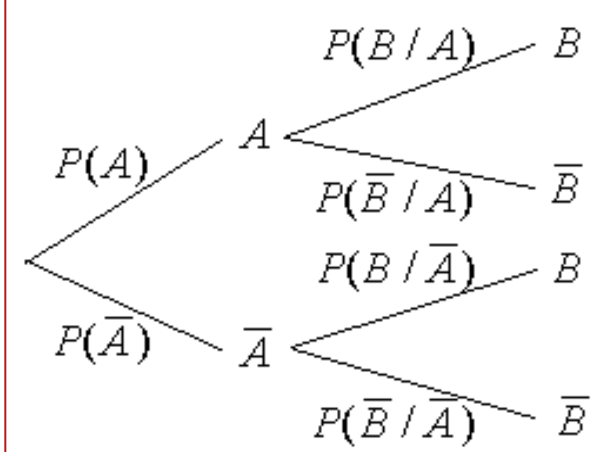
Es decir, la información adicional de la ocurrencia del evento B, ha hecho que la probabilidad de A se incremente de ⅙ a 1/3

Ejemplo

Un grupo de personas está compuesto de 60% hombres y 40% de mujeres. De los hombres, el 30% fuma y de las mujeres, el 20% fuma. Si una persona de este grupo se escoge al azar, encuentre la probabilidad de que

A. sea hombre y fume
B. sea hombre y no fume
C. sea mujer y fuma
D. sea mujer y no fume

¿ Y la probabilidad de que sea hombre dado que fuma?



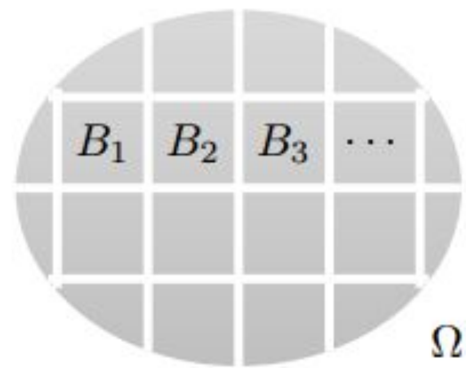
- Partición

Sea Ω el espacio muestral de un experimento aleatorio. Decimos que la colección de eventos $\{B_1, \ldots, B_n\}$ es una partición finita de Ω si se cumplen las siguientes condiciones:

a) $B_i \neq \varnothing, \quad i = 1, \ldots, n.$

b) $B_i \cap B_j = \varnothing \quad \text{para } i \neq j.$

c) $\bigcup_{i=1}^{n} B_i = \Omega.$



Sea $B_1, \ldots, B_n$ una partición de Ω tal que $P(B_i) \neq 0$, $i = 1, \ldots, n$. Para cualquier evento A,

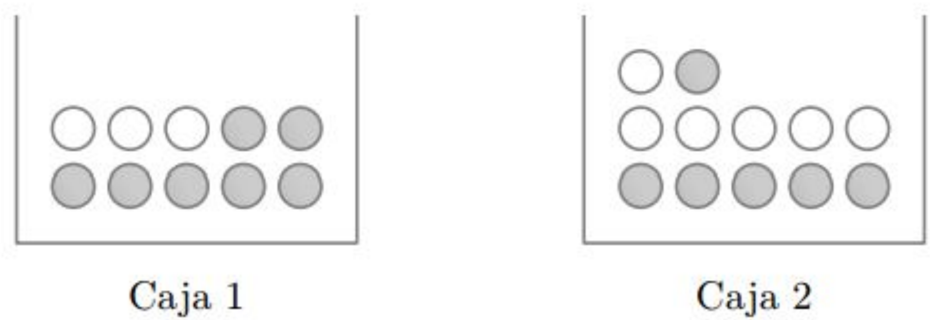
$$P(A) = \sum_{i=1}^{n} P(A \mid B_i) P(B_i).$$

Cuando la partición del espacio muestral consta de únicamente los elementos B y B^c, la fórmula del teorema de probabilidad total se reduce a la expresión

$$P(A) = P(A \mid B)\, P(B) + P(A \mid B^c)\, P(B^c).$$

- Ejemplo

  Suponga que tenemos dos cajas: una con 3 bolas blancas y 7 bolas de color gris, la otra con 6 blancas y 6 grises.
  Si se elije una caja al azar y después se saca una bola al azar, ¿cuál es la probabilidad de que se blanca?



$$\Omega = \{(C_1, B), (C_1, G), (C_2, B), (C_2, G)\},$$

en donde C1 y C2 denotan los eventos en donde las cajas uno y dos fueron escogidas, respectivamente, y B y G denotan los eventos en donde una bola blanca o gris fueron escogidas, respectivamente.

Nos piden calcular la probabilidad de B. Observe que es fácil calcular la probabilidad de este evento cuando se conoce la caja que fue escogida.

Esto sugiere condicionar sobre el resultado de escoger alguna de las dos cajas y aplicar el teorema de probabilidad total, es decir,

$$\begin{aligned} P(B) &= P(B \mid C_1)P(C_1) + P(B \mid C_2)P(C_2) \\ &= (3/10)(1/2) + (6/12)(1/2) \\ &= 2/5. \end{aligned}$$

- Ejemplo

Suponga que en una población humana de igual número de hombres y mujeres, el 4 % de hombres son daltónicos y el 1 % de las mujeres son daltónicas. Una persona es elegida al azar, ¿cuál es la probabilidad de que sea daltónica?

Solución.

Definamos primero los eventos de interés. Sea:

M el evento "La persona escogida es mujer",

H el evento "La persona escogida es hombre",

D el evento "La persona escogida es daltónica".

Deseamos calcular P(D). Por el teorema de probabilidad total,

$$\begin{aligned} P(D) &= P(D \mid M)P(M) + P(D \mid H)P(H) \\ &= (1/100)(1/2) + (4/100)(1/2) \\ &= 1/40. \end{aligned}$$

- Teorema de Bayes

  El resultado interesante que estudiaremos a continuación involucra nuevamente probabilidades condicionales. Fue publicado por primera vez en 1763, dos años después de la muerte de su creador: el matemático y teólogo inglés Thomas Bayes

Thomas Bayes
(Inglaterra 1702–1761)

Teorema 1.2 (Teorema de Bayes) Sea $B_1, \ldots, B_n$ una partición de Ω tal que $P(B_i) \neq 0$, $i = 1, \ldots, n$. Sea A un evento tal que $P(A) \neq 0$. Entonces para cada $j = 1, 2, \ldots, n$,

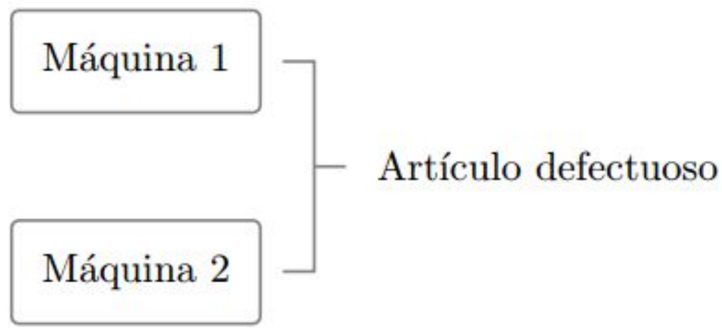
$$P(B_j \mid A) = \frac{P(A \mid B_j)P(B_j)}{\sum_{i=1}^{n} P(A \mid B_i)P(B_i)}.$$

Demostración. Por definición de probabilidad condicional, y después usando el teorema de probabilidad total, tenemos que para cada $j = 1, \ldots, n$,

$$P(B_j \mid A) = \frac{P(A \cap B_j)}{P(A)} = \frac{P(A \mid B_j)P(B_j)}{\sum_{i=1}^{n} P(A \mid B_i)P(B_i)}.$$

- Ejemplos

En una fábrica hay dos máquinas. La máquina 1 realiza el 60% de la producción total y la máquina 2 el 40%. De su producción total, la máquina 1 produce 3% de material defectuoso, la 2 el 5%. El asunto es que se ha encontrado un material defectuoso.
¿Cuál es la probabilidad de que este material defectuoso provenga de la máquina 2?



Solución.

Es conveniente definir los siguientes eventos:

D = "El material escogido es defectuoso",

M_1 = "La máquina 1 produjo el material escogido",

M_2 = "La máquina 2 produjo el material escogido"

La pregunta planteada se traduce en encontrar $P(M_2|D)$ y observamos que la información que tenemos es $P(D|M_2)$. Por el teorema de Bayes,

$$\begin{aligned} P(M_2 \mid D) &= \frac{P(D \mid M_2)P(M_2)}{P(D \mid M_1)P(M_1) + P(D \mid M_2)P(M_2)} \\ &= \frac{(5/100)(40/100)}{(3/100)(60/100) + (5/100)(40/100)} \\ &= 10/19. \end{aligned}$$

Comprobar que $P(M_1|D) = 9/19$

DigitalHouse > Coding School

- Ejemplos

En un laboratorio se descubrió una prueba para detectar cierta enfermedad, y sobre la eficacia de dicha prueba se conoce lo siguiente: si se denota por E el evento de que un paciente tenga la enfermedad y por N el evento de que la prueba resulte negativa, entonces se sabe que

$$\begin{aligned} P(N^c \mid E) &= 0.95, \\ P(N \mid E^c) &= 0.96, \\ P(E) &= 0.01\,. \end{aligned}$$

Observe que esta información corresponde al caso cuando se conoce la situación médica del paciente, es decir, si está enfermo o no lo está. Con únicamente estos datos, uno podría pensar que la prueba es muy buena, sin embargo calcularemos las probabilidades $P(E \mid N)$ y $P(E \mid N^c)$ para saber la efectividad de la prueba cuando una persona recibe sus resultados. Usando el teorema de Bayes tenemos que

$$
\begin{aligned}
P(E \mid N) &= \frac{P(N \mid E)P(E)}{P(N \mid E)P(E) + P(N \mid E^c)P(E^c)} \\
&= \frac{0.05 \times 0.01}{0.05 \times 0.01 + 0.96 \times 0.99} \\
&= 0.000526\,.
\end{aligned}
$$

probabilidad de falso negativo (bajo es bueno)

$$
\begin{aligned}
P(E \mid N^c) &= \frac{P(N^c \mid E)P(E)}{P(N^c \mid E)P(E) + P(N^c \mid E^c)P(E^c)} \\
&= \frac{0.95 \times 0.01}{0.95 \times 0.01 + 0.04 \times 0.99} \\
&= 0.193\,.
\end{aligned}
$$

probabilidad de verdadero positivo (bajo es malo)

- Independencia de Eventos

  La independencia entre eventos es equivalente a la situación cuando la ocurrencia de un evento no afecta la probabilidad de ocurrencia de otro evento.

  Es un concepto importante que en algunos casos nos simplificará considerablemente el cálculo de probabilidades conjuntas.

Definición 1.11 Se dice que los eventos A y B son independientes si se cumple la igualdad

$$P(A \cap B) = P(A)P(B). \tag{1.4}$$

Bajo la hipótesis adicional de que P(B) > 0, la condición de independencia puede escribirse como

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

Esto significa que la ocurrencia del evento B no afecta la probabilidad del evento A. Análogamente, cuando P(A) > 0, la condición se puede escribir como

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

es decir, la ocurrencia del evento A no cambia a la probabilidad de B. Veamos algunos ejemplos

- Ejemplos

  Considere un experimento aleatorio con espacio muestral equiprobable $\Omega = \{1, 2, 3, 4\}$

  A. Los eventos $A = \{1, 2\}$ y $B = \{1, 3\}$ son independientes pues un tanto $P(A \cap B)$ como $P(A)P(B)$ coinciden en el valor $1/4$.

  B. Los eventos $A = \{1, 2, 3\}$ y $B = \{1, 3\}$ no son independientes pues $P(A \cap B) = 1/2$, mientras que $P(A)P(B) = (3/4)(1/2) = 3/8$

- ## Independencia vs disjuntos

Considere un evento $A \neq \varnothing$ y el espacio muestral Ω. Recuerde que $P(\Omega)=1$.

Es claro que A y Ω son independientes pues $P(A \cap \Omega) = P(A)=P(A)^{*}1=P(A)^{*}\ P(\Omega)$.

Sin embargo, $A \cap \Omega = A \neq \varnothing$.

Por lo tanto, el hecho de que dos eventos sean independientes no implica necesariamente que sean disjuntos.

- Ejemplo

Considere el experimento aleatorio de lanzar un dado equilibrado y defina los eventos A como obtener un número par y B como obtener un número impar. Es claro que los eventos A y B son disjuntos, sin embargo no son independientes pues

$$0 = P(A \cap B) \neq P(A)P(B) = (1/2)(1/2).$$

Por lo tanto, dos eventos pueden ser disjuntos y ello no implica necesariamente que sean independientes

La definición de independencia de dos eventos puede extenderse al caso de tres eventos y, más generalmente, para cualquier colección finita de eventos de la manera siguiente

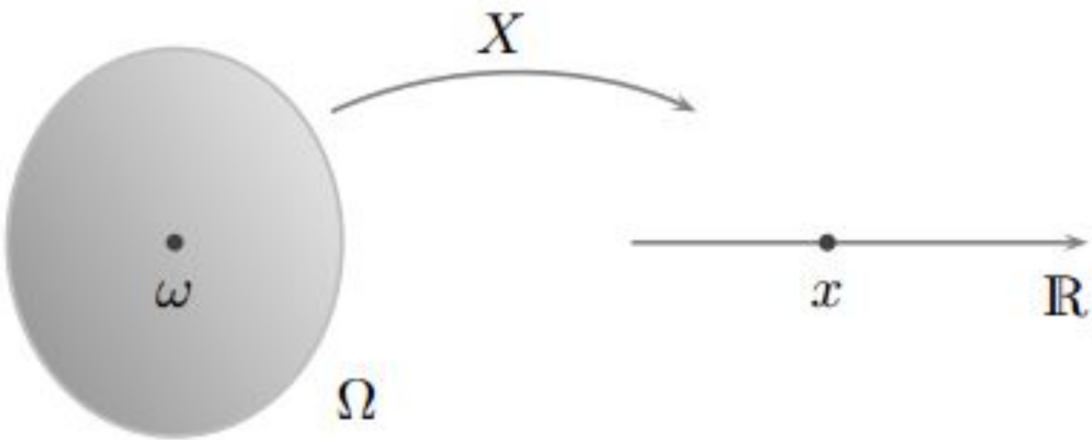
Definición 1.12 Decimos que n eventos $A_1, \ldots, A_n$ son (mutuamente) independientes si se satisfacen todas y cada una de las condiciones siguientes:

$$
\begin{aligned}
P(A_i \cap A_j) &= P(A_i)P(A_j), \quad i, j \text{ distintos.} && (1.5)\\
P(A_i \cap A_j \cap A_k) &= P(A_i)P(A_j)P(A_k), \quad i, j, k \text{ distintos.} && (1.6)\\
&\vdots \\
P(A_1 \cap \cdots \cap A_n) &= P(A_1) \cdots P(A_n). && (1.7)
\end{aligned}
$$

Variables Aleatorias

Una variable aleatoria es una transformación X del espacio de resultados Ω al conjunto de números reales, esto es,

$$X : \Omega \to \mathbb{R},$$



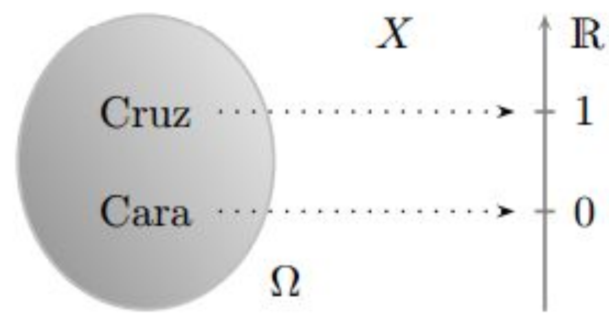
Nuestro interés en el estudio de los experimentos aleatorios se trasladará al estudio de las distintas variables aleatorias

- Ejemplo

Suponga que un experimento aleatorio consiste en lanzar al aire una moneda equilibrada y observar la cara superior una vez que la moneda cae.

Denotemos por “Cara” y “Cruz” los dos lados de la moneda. Entonces, claramente, el espacio muestral es el conjunto Ω = {“Cruz” , “Cara”}. Defina la variable aleatoria $X : \Omega \rightarrow \mathbb{R}$ de la forma siguiente.

$$\begin{aligned} X(\text{“Cruz”}) &= 1, \\ X(\text{“Cara”}) &= 0. \end{aligned}$$



De este modo podemos suponer que el experimento aleatorio tiene dos valores numéricos: 0 y 1. Observe que estos números son arbitrarios pues cualquier otro par de números puede ser escogido como los valores de la variable aleatoria.

Se muestran a continuación algunos ejemplos de eventos de esta variable aleatoria y sus correspondientes probabilidades

a) $P(X \in [1,2)) = P(\{\text{“Cruz”}\}) = 1/2.$

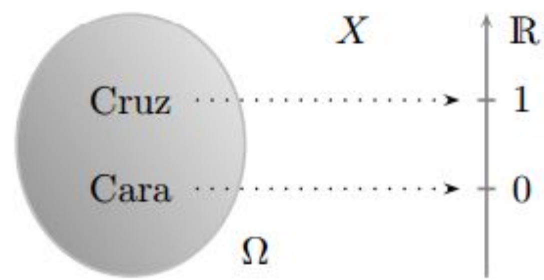
b) $P(X \in [0,1)) = P(\{\text{“Cara”}\}) = 1/2.$

c) $P(X \in [2,4]) = P(\emptyset) = 0.$

d) $P(X = 1) = P(\{\text{“Cruz”}\}) = 1/2.$

e) $P(X \leqslant -1) = P(\emptyset) = 0.$

f) $P(X \geqslant 0) = P(\Omega) = 1.$



Variable Aleatoria Discreta

Cuando los valores que toma una variable aleatoria son *finitos o infinitos numerables* se dice que es **discreta**.

- Resultado obtenido al lanzar un dado
  {1,2,3,4,5,6}
- Número de veces que hay que lanzar una moneda hasta obtener una CARA
  {1,2,3,4, ...}

Distribución de probabilidad

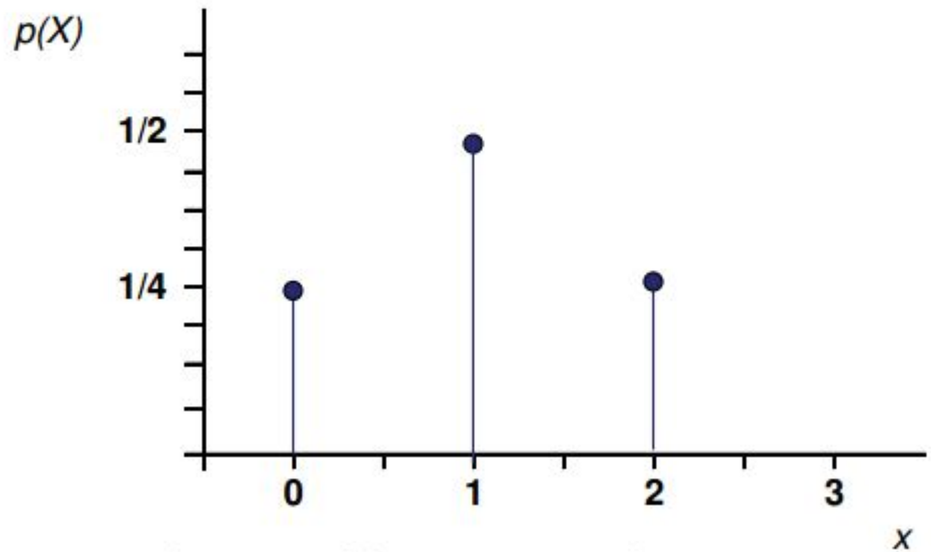
Sea { x_1 , x_2 , ..., x_n } los valores que puede tomar la variable aleatoria *X*. Se denomina **distribución de probabilidad** de la variable aleatoria a *P(X=x_i)* que cumple:

- $P(X = x_i) \geq 0$
- $\Sigma_{i=1} P(X = x_i) = 1.$

Nº de Caras al lanzar 2 monedas

x		$P(X=x)$
0	$\rightarrow$	1/4
1	$\rightarrow$	1/2
2	$\rightarrow$	1/4

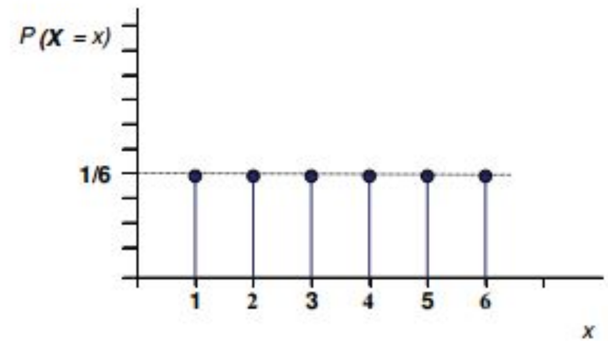
Distribución de probabilidad



Nº de Caras al lanzar 2 monedas

Lanzamiento de un dado

x	$P(X = x)$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6



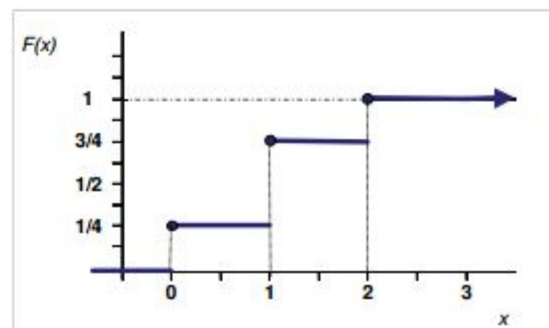
Función de distribución

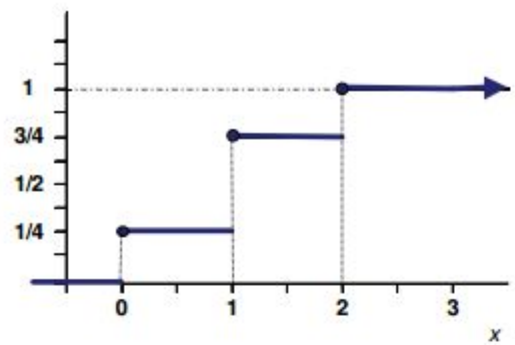
La función de distribución $F_X(x)$ de una variable aleatoria X se define para todo número real x como:

$$F_X(x) = P_X(X \leq x).$$

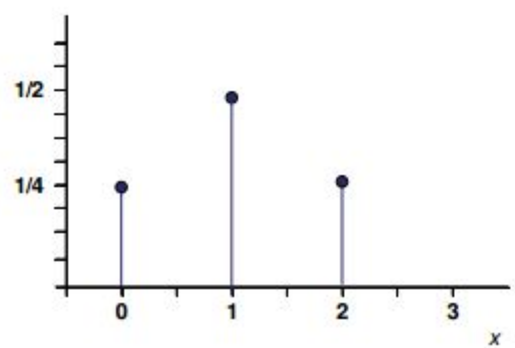
Ejemplo. X = Número de caras al lanzar 2 monedas

x	$F_X(x)$
$(-\infty,0)$	0
$[0,1)$	1/4
$[1,2)$	3/4
$[2,\infty)$	1



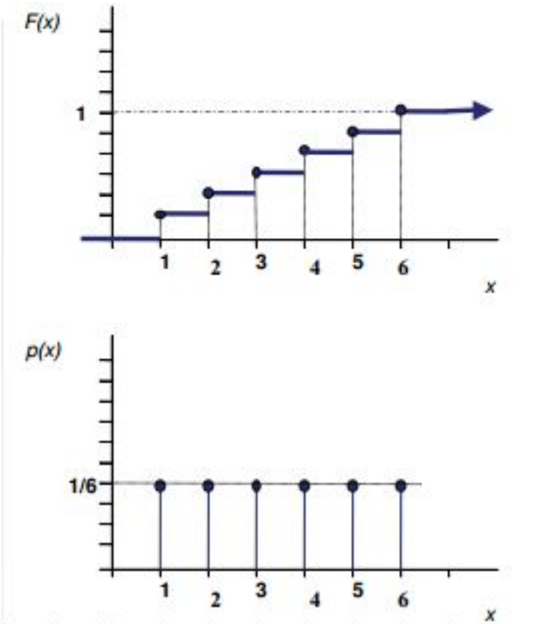


Función de Distribución



Distribución puntual de probabilidad

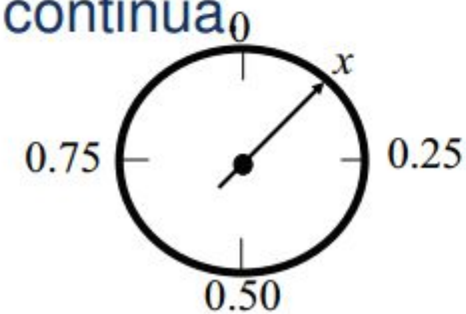
Lanzamiento de un dado

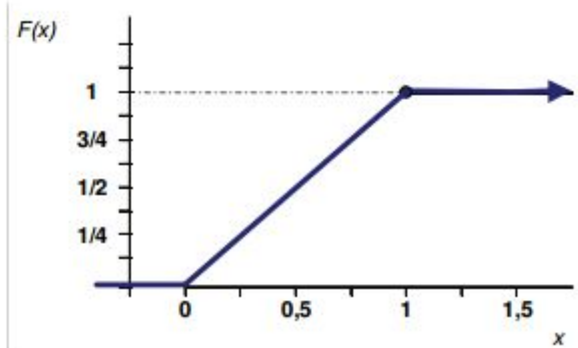


x	$P(X=x)$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

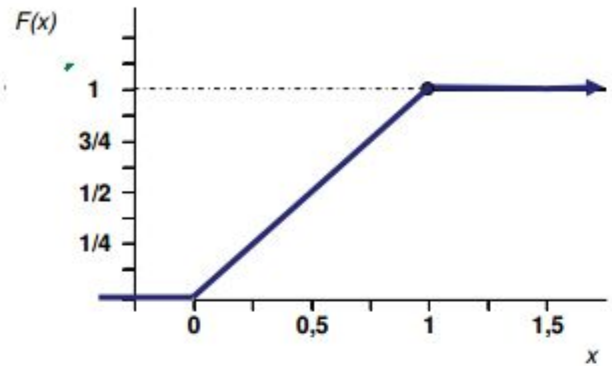
Variable aleatoria continua

Una variable aleatoria X es continua si su función de distribución $F_X(x)$ es continua



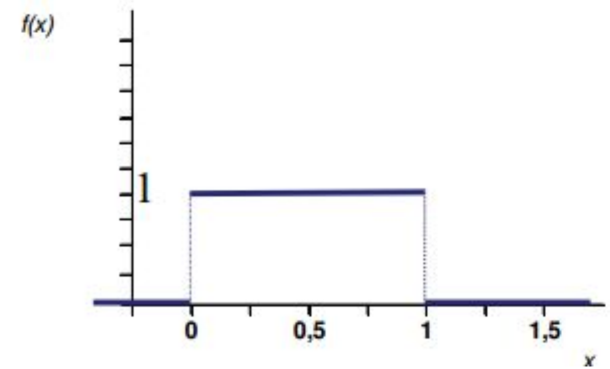


$$F_X(x) = x, \quad x \in [0,1)$$



Función de distribución

$$F_X(x) = x, \;\; x \in [0,1)$$



Función de densidad

$$f_X(x) = 1, \;\; x \in [0,1]$$

- Estudiaremos ahora algunas distribuciones de probabilidad particulares.

- Empezaremos con las **distribuciones de tipo discreto** y continuaremos con las de **tipo continuo**.

- Observaremos además que las distribuciones pueden depender de uno o más **parámetros**, es decir, para cada valor de estos parámetros se tiene una distribución de probabilidad. A veces se usa el término "**familias de distribuciones paramétricas**" para referirse a estas distribuciones.

Decimos que una variable aleatoria X tiene una distribución **uniforme discreta** sobre el conjunto de *n* números $\{x_1, ..., x_n\}$ si la probabilidad de que X tome cualquiera de estos valores es **constante 1/*n***.

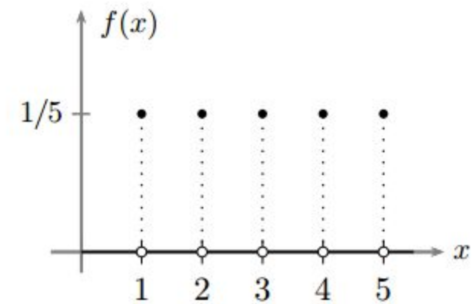
Esta distribución surge en **espacios de probabilidad equiprobables**, esto es, en situaciones en donde tenemos *n* resultados diferentes y todos ellos tienen **la misma probabilidad de ocurrir**.

Los juegos de lotería son un ejemplo donde puede aplicarse esta distribución de probabilidad.

Se escribe $X \sim$ unif $\{ x_1, ..., x \}_n$, en donde el símbolo "∼" se lee "se distribuye como" o "tiene una distribución".

La distribución de probabilidad de esa variable aleatoria es:

$$f(x) = \begin{cases} 1/n & \text{si} \quad x = x_1, \ldots, x_n, \\ 0 & \text{en otro caso.} \end{cases}$$



Un ensayo Bernoulli[1] se define como aquel experimento aleatorio con únicamente dos posibles resultados, llamados genéricamente: éxito y fracaso. Supondremos que las probabilidades de estos resultados son p y $1-p$, respectivamente. Si se define la variable aleatoria X como aquella función que lleva el resultado éxito al número 1 y el resultado fracaso al número 0, entonces decimos que X tiene una distribución Bernoulli con parámetro $p \in (0,1)$ y escribimos $X \sim \text{Ber}(p)$. La función de probabilidad se puede escribir de la siguiente forma.

$$f(x) = \begin{cases} 1-p & \text{si} \quad x = 0, \\ p & \text{si} \quad x = 1, \\ 0 & \text{en otro caso.} \end{cases}$$

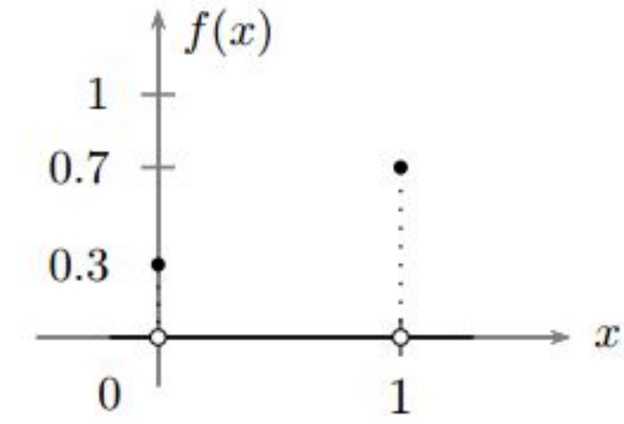
O bien de manera compacta,

$$f(x) = \begin{cases} p^x(1-p)^{1-x} & \text{si} \quad x = 0, 1, \\ 0 & \text{en otro caso.} \end{cases}$$

from scipy.stats import bernoulli

**generamos 1000 números aleatorios con la distribución mostrada en el gráfico.*

```
r = bernoulli.rvs(p=0.7, size=1000)
```



- Distribución Binomial.

Mide el número de éxitos en una secuencia de *n* ensayos independientes entre sí, con una probabilidad fija *p* de ocurrencia del éxito entre los ensayos.

  - Ejemplo. Se lanzará una moneda tres veces. ¿Cuál es la probabilidad de sacar exactamente una cara?
  - 8 posibles resultados: HHH, HHT, HTH, THH, HTT, THT, TTH, TTT
  - 3 casos con exactamente una cara: HTT, THT, and TTH por lo que la probabilidad es 3/8=0.375

- Distribución Binomial.
    - Vamos a pensar una forma para generalizar este cálculo (con n grande sería engorroso contar los casos).
    - ¿Cuántas formas hay de poner 3 éxitos (que sale cara) en 5 repeticiones del experimento? Tengo 5 elementos (las pruebas) y quiero armar un grupo de 3 donde no importa el orden de caras y cecas (solo importa el total de caras).
    - ¿Cuál es la probabilidad de observar 3 caras en cualquier orden? Llamemos p a la probabilidad de cara, 1-p es la probabilidad de no cara. Entonces la probabilidad buscada es $p^*p^*p^*(1-p)^*(1-p)$.
    - Entonces la probabilidad buscada es $p^*p^*p^*(1-p)^*(1-p)^*5!/((5-3)!3!)$

- Distribución Binomial.
    - Sea n los intentos, x los éxitos que buscamos y p la probabilidad de éxito en cada prueba (independientes entre sí).

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

Sea la variable aleatoria discreta X el número de veces que determinado evento ocurre en un intervalo de tiempo o espacio. Por ejemplo, autos que pasan por una intersección o personas que entran a un banco, errores de tipeo en un slide, etc.

Entonces X puede ser una variable Poisson que toma valores $x=0,1,2,\ldots$ si cumple con las siguientes condiciones:

1. El número de eventos que ocurren en períodos de tiempo , sin superposición entre períodos, es independiente.
2. La probabilidad de exactamente un evento en un intervalo de tiempo corto de duración h=1/n es aproximadamente h*λ donde n es la cantidad de intervalos dentro del período considerado.
3. La probabilidad de exactamente dos o más eventos en un intervalo de tiempo corto es esencialmente cero.

Si se cumplen esas condiciones *X es una variable aleatoria que sigue un proceso de Poisson aproximado con parámetro* λ > 0 por lo que *su función de probabilidad puntual es*

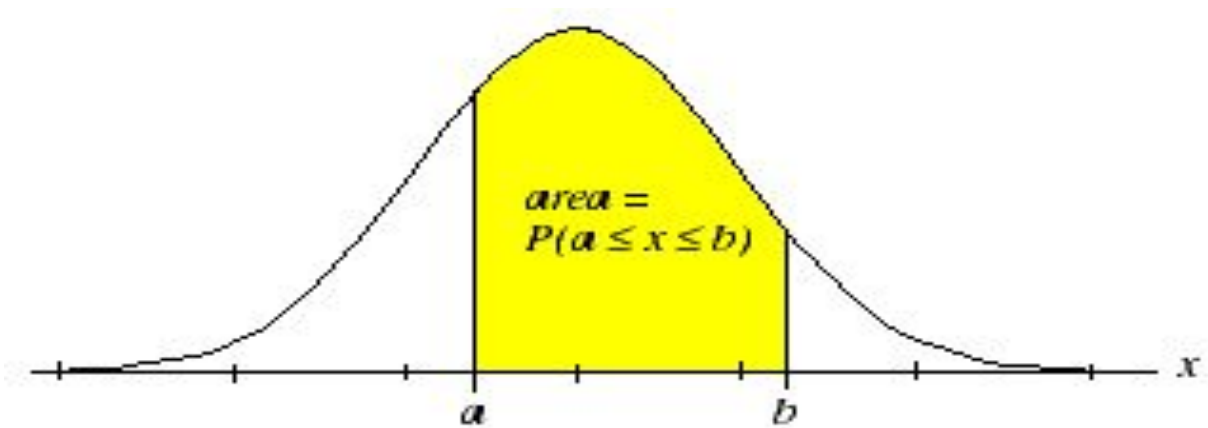
$$f(k;\lambda) = \Pr(X = k) = \frac{\lambda^k e^{-\lambda}}{k!},$$

Se puede mostrar que λ es la media y la varianza de una variable Poisson.

- Variables aleatorias continuas.
  - Imagine que una variable aleatoria puede tomar cualquier valor real incluido en entre los números 0 y 1. ¿Cuántos valores posibles puede tomar esta variable aleatoria?
  - El concepto de probabilidad como frecuencia relativa de determinado resultado puntual (casos favorables sobre casos posibles) carece de sentido ya que la cantidad de valores diferentes que puede tomar la variable aleatoria es infinita en este caso.
  - Si X es variable aleatoria continua entonces la probabilidad de que X=x (cualquier valor particular) es cero.
  - Podemos calcular la probabilidad de que X se ubique dentro de cierto intervalo de valores.

- Variables aleatorias continuas.
  - La “función de densidad” es el concepto similar a la “función de probabilidades”, excepto que no indica una probabilidad (f(x) no es prob(X=x).
  - La densidad es más bien una medida del cambio en la probabilidad a medida que nos movemos sobre el eje horizontal.

- Variables aleatorias continuas.
  - En el siguiente gráfico imagine que aumentamos el intervalo moviéndonos un poco más hacia la derecha del punto b. f(X) mide la sensibilidad de la probabilidad acumulada en ese intervalo dado el aumento marginal en el valor del borde derecho que lo define.

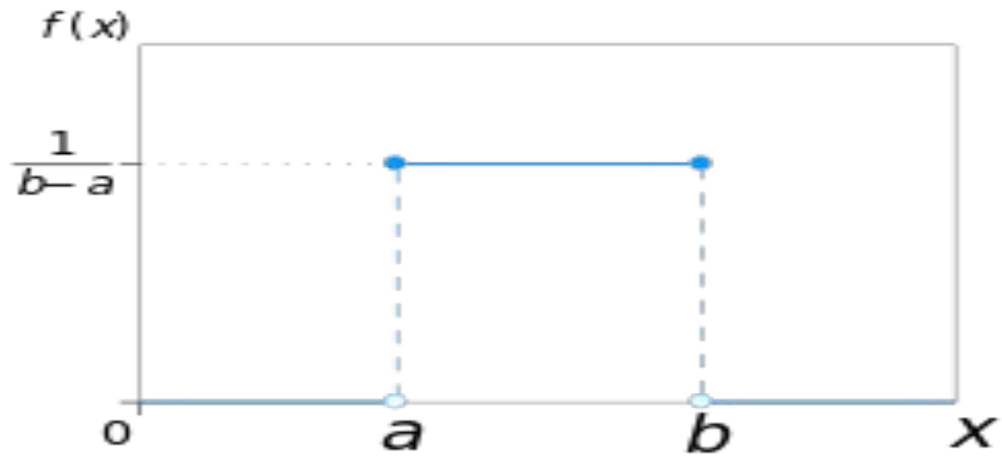


Si una variable aleatoria continua tiene distribución uniforme en el intervalo (a,b), donde a y b son números reales, su función de densidad es

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & a < x < b \\ 0 & \text{en otro caso} \end{cases}$$

- La uniforme asigna probabilidad positiva sólo a valores de la variable aleatoria en determinado rango pero sin diferenciar entre ellos.
- Por ejemplo cuando queremos simular y creemos que cierto parámetro está en determinado rango pero no tenemos motivos para asignar probabilidades distintas a los diferentes rangos dentro del intervalo (a,b).

DISTRIBUCIÓN UNIFORME CONTÍNUA



$$F(x) = \begin{cases} 0 & \text{para } x < a \\ \frac{x-a}{b-a} & \text{para } a \le x < b \\ 1 & \text{para } x \ge b \end{cases}$$

- La distribución exponencial describe el tiempo entre eventos en un proceso de Poisson

$$f(x) = P(x) = \begin{cases} \lambda e^{-\lambda x} & \text{para } x \geq 0 \\ 0 & \text{en caso contrario} \end{cases}$$

Su función de distribución acumulada es:

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{para } x < 0 \\ 1 - e^{-\lambda x} & \text{para } x \geq 0 \end{cases}$$

Donde e representa el número e.

La distribución normal con parámetros *μ (media) y σ^2 (varianza)* tiene función de densidad dada por

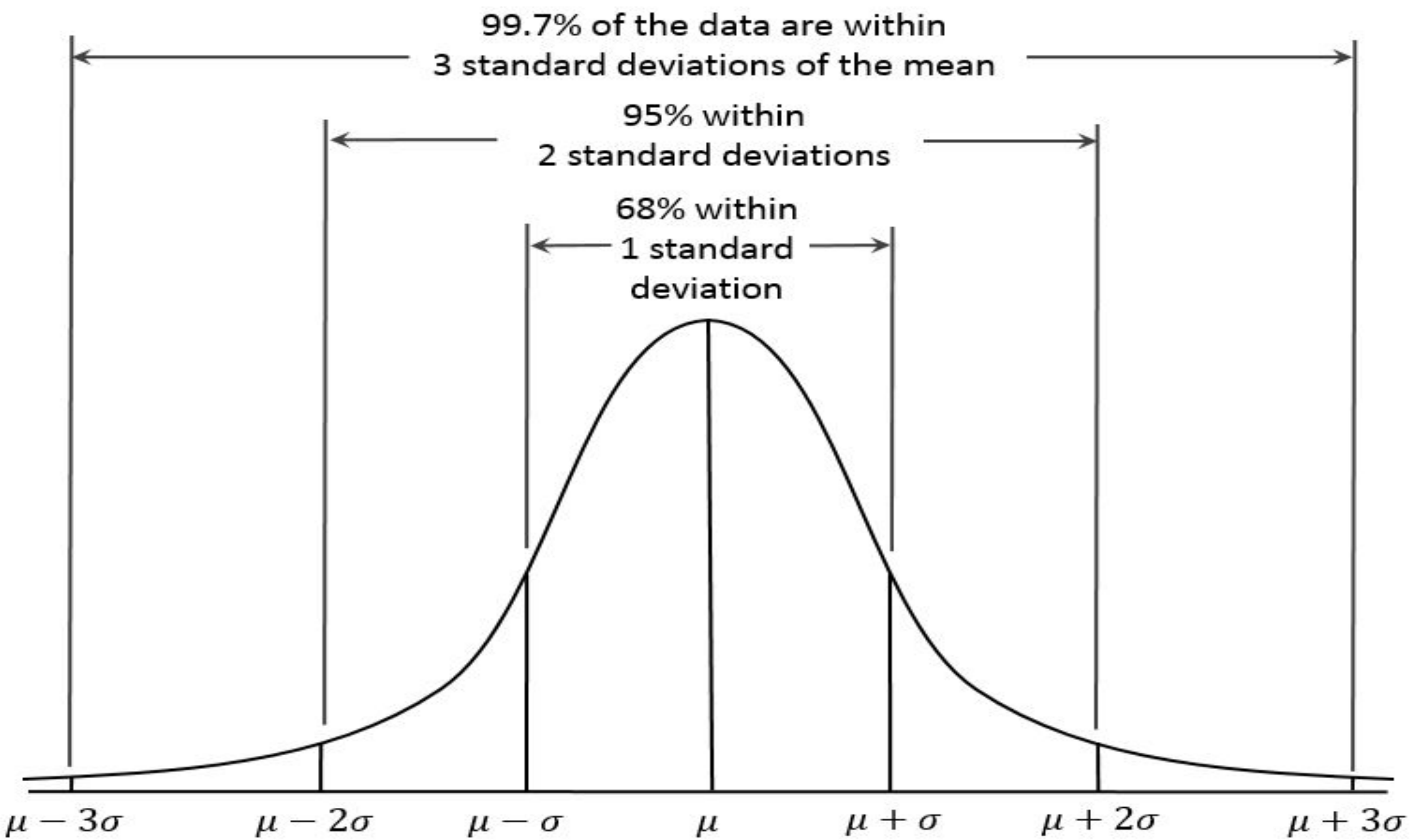
$$f(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

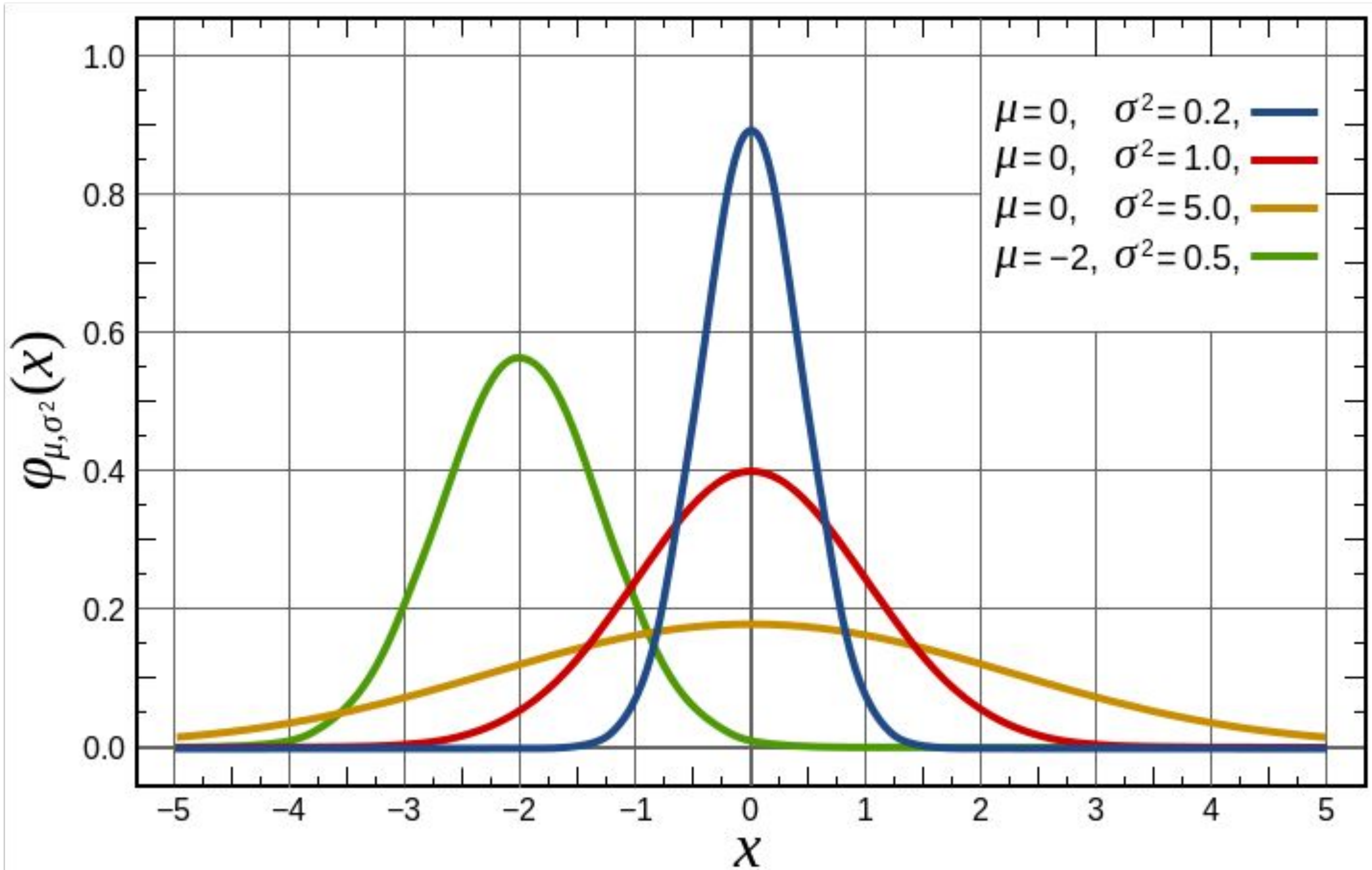
- Los valores posibles de X son los reales entre menos y más infinito.
- El gráfico de la densidad para una variable normal tiene forma de campana.
- La probabilidad de que una variable aleatoria X con distribución normal y parámetros *μ, σ^2 sea menor o igual que un valor a fijo es*

$$\mathrm{Prob}(X \le a) = \int_{-\infty}^{a} \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{(x-\mu)^2}{2\sigma^2}\right)} dx$$

- La distribución normal es simétrica respecto a su media.
- En particular cuando la media es cero y el desvío estándar es igual a 1 decimos se se trata de una normal estándar.
- Decimos que la distribución normal está caracterizada por su media y varianza porque si conozco esos parámetros puedo reemplazarlos en la función densidad y tener caracterizada a la distribución. Por ejemplo puedo calcular probabilidades acumuladas, momentos, quartiles.
- Las medidas de asimetría y curtosis usuales se basan en las características de la distribución normal. Ambas medidas valen cero para la normal (es la distribución de referencia).

DISTRIBUCIÓN NORMAL




$\mu=0, \quad \sigma^2=0.2,$
$\mu=0, \quad \sigma^2=1.0,$
$\mu=0, \quad \sigma^2=5.0,$
$\mu=-2, \sigma^2=0.5,$
1.0
0.8
0.6
0.4
0.2
0.0
$\varphi_{\mu,\sigma^2}(x)$
−5
−4
−3
−2
−1
0
1
2
3
4
5
x

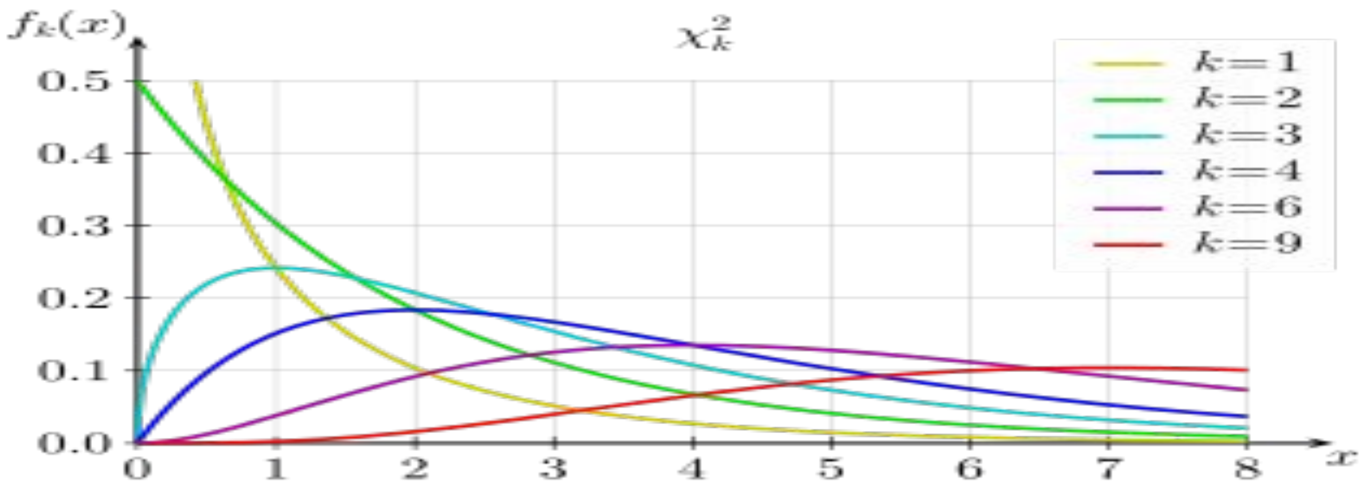
Algunas propiedades más técnicas

- Cualquier función lineal de una variable aleatoria normal también es normal. Si X es una variable aleatoria normal con parámetros *μ y σ^2 podemos definir Y=a+bX con b distinto de cero (¿Por qué?). Entonces Y también tendrá distribución normal (¿Con qué parámetros?)*
- La variable aleatoria generada por una suma finita de variables aleatorias independientes es normal.
- En el caso de que se sumen infinitas variables aleatorias, independientes y tales que tanto la suma de las medias como de las varianzas sean finitas también vale la propiedad.

- Sean Z1, Z2,..Zk variables aleatorias independientes normales con media 0 y varianza 1. Definamos una nueva variable aleatoria

$$Q = \sum_{i=1}^{k} Z_i^2 ,$$

- Q tiene distribución Chi-Cuadrado con n grados de libertad.



Algunas propiedades más técnicas..

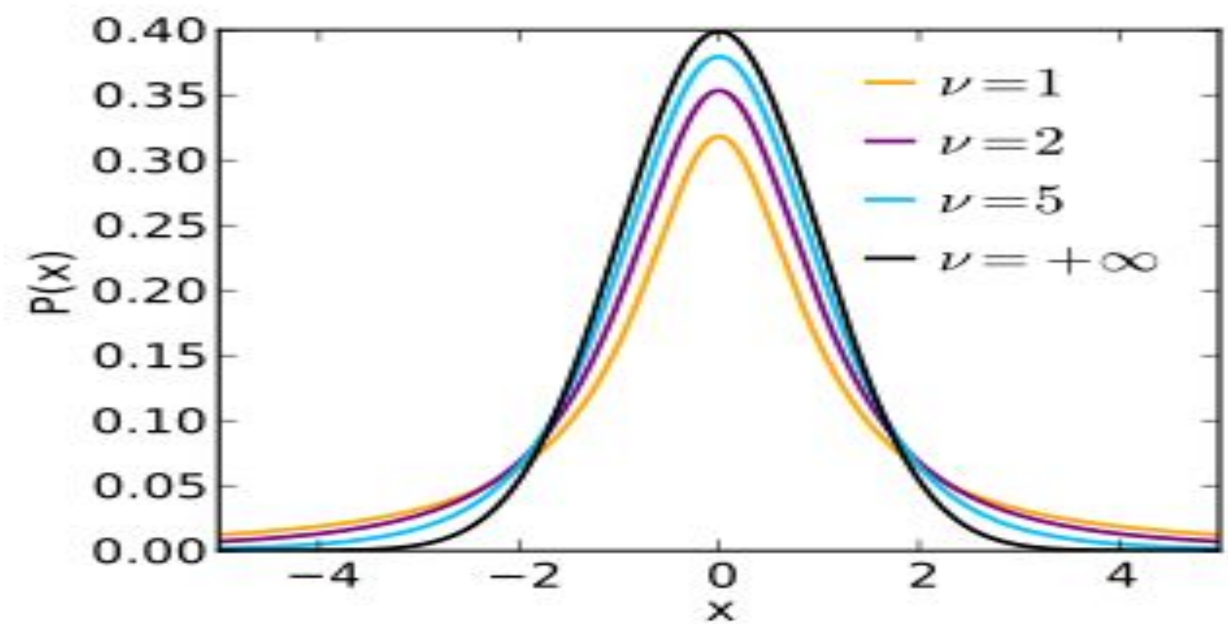
- L a distribución Chi-Cuadrado tiene asimetría positiva.
- A medida que aumentan los grado de libertad se reduce la asimetría.
- Cuando los grados de libertad tienden a infinito la distribución se aproxima a la normal.

- Sean Z una variable aleatoria normal con media 0 y varianza 1.
- Sea V con distribución Chi-Cuadrado con n grados de libertad independiente de Z. Definamos una nueva variable aleatoria

$$T = \frac{Z}{\sqrt{V/n}}$$

- T tiene distribución T-Student con n grados de libertad.

- ¿Es simétrica la distribución T- Student? ¿Qué la diferencia de la normal?

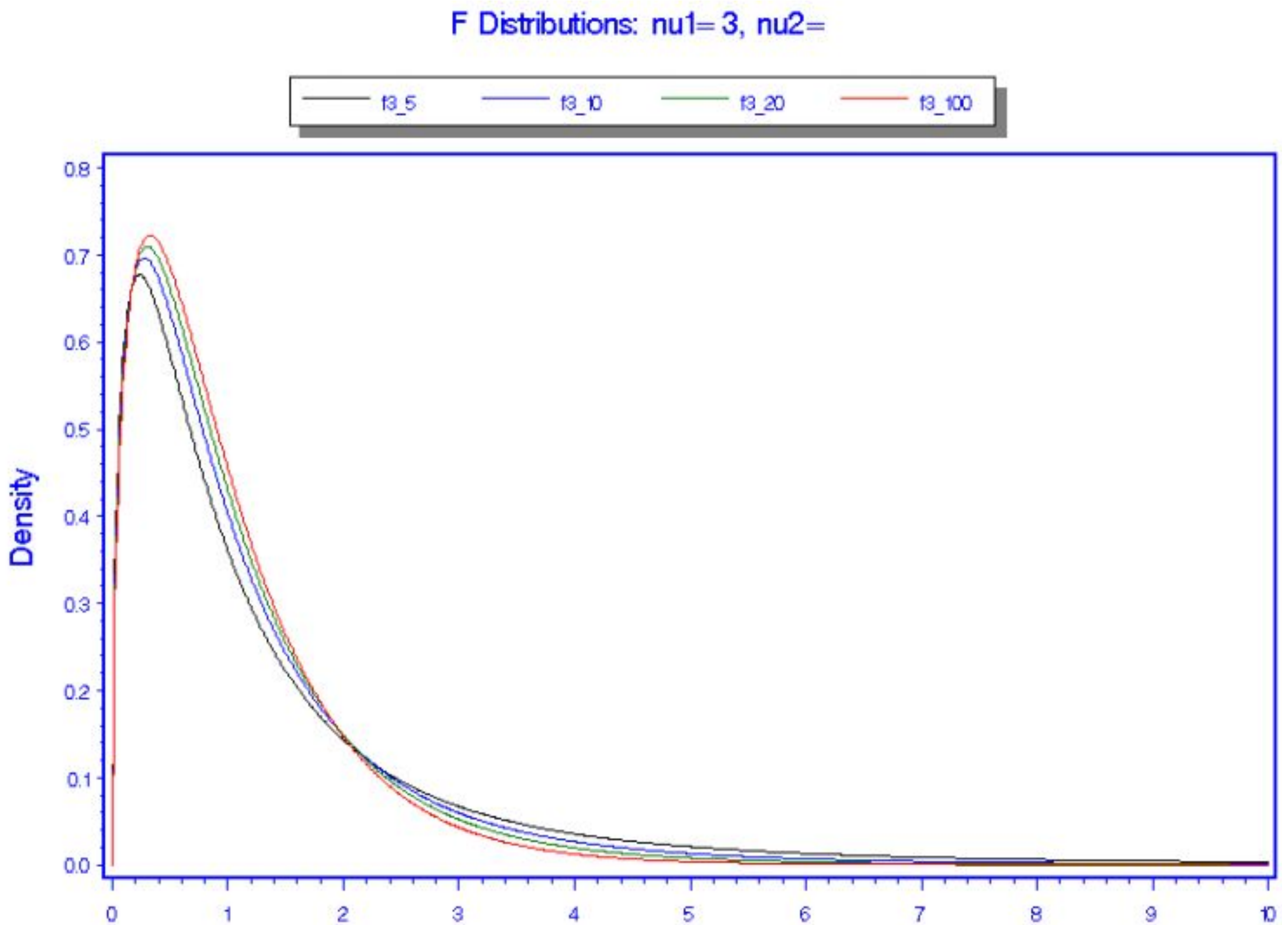


- Sea con U_1 distribución Chi-Cuadrado con d_1 grados de libertad.
- Sea U_2 con distribución Chi-Cuadrado con d_2 grados de libertad independiente de U_1. Definamos una nueva variable aleatoria

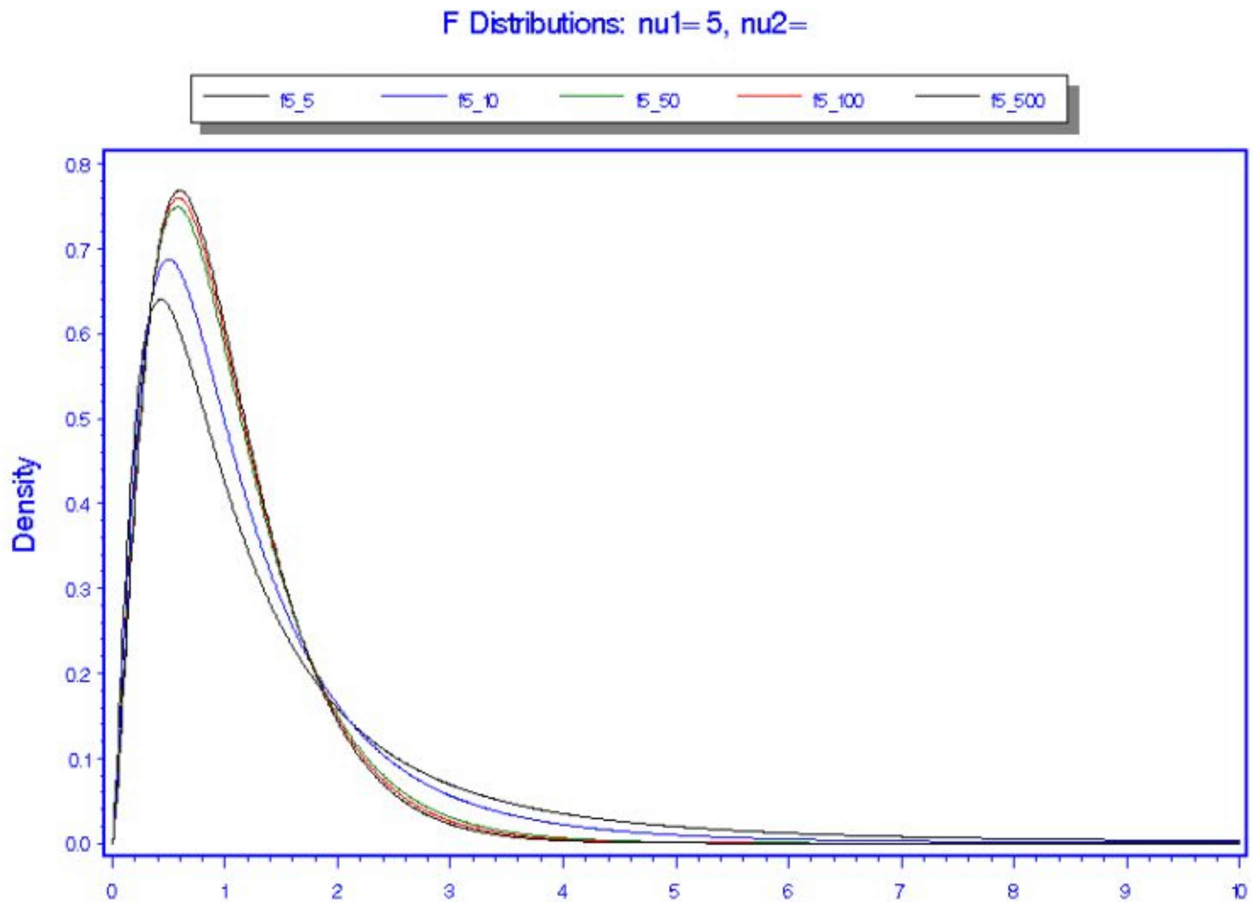
$$X = \frac{U_1/d_1}{U_2/d_2}$$

- X tiene distribución F con grados de libertad d_1 y d_2.
- La variable aleatoria X con distribución F toma valores en rango rango de 0 a infinito.
- X tiene además 2 parámetros: los grados de libertad del numerador (d_1) y del denominador (d_2)
- Para d_1 y d_2 fijos la distribución no es simétrica.
- La forma de la distribución varía mucho de acuerdo a los valores de d_1 y d_2.

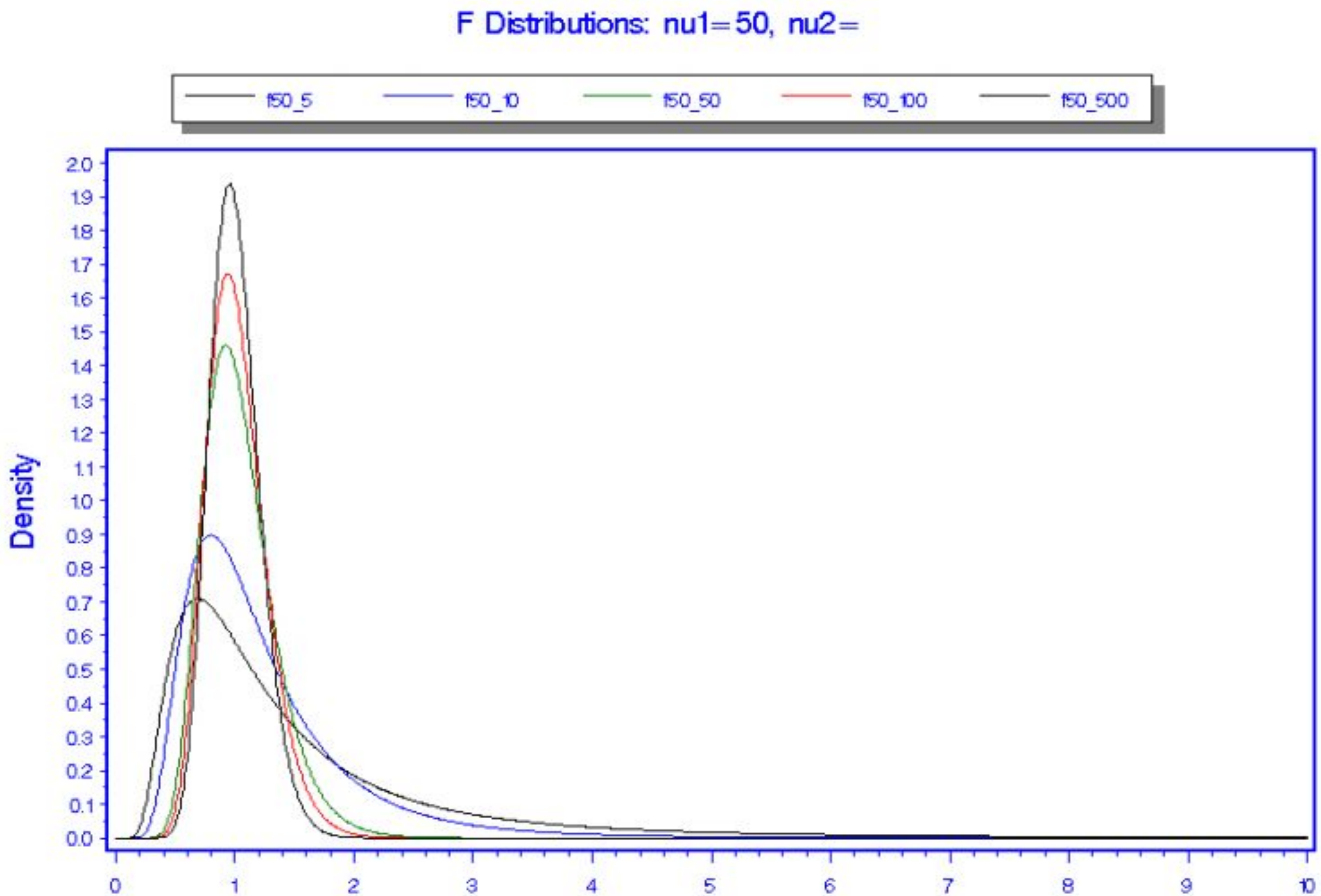
DISTRIBUCIÓN F



DISTRIBUCIÓN F



DISTRIBUCIÓN F



2. PRÁCTICA GUIADA PROBABILIDAD

- Algunos recursos
  - https://www.random.org/randomness/
  - https://pypi.python.org/pypi/randomdotorg/
  - https://pypi.python.org/pypi/quantumrandom/
  - https://en.wikipedia.org/wiki/Inverse_transform_sampling
  - `rng=random.SystemRandom() para usar información del sistema para generar números aleatorios`

OPTATIVO

- Juego del cumpleaños.
  - ¿Qué probabilidad hay de encontrar 2 personas en este curso de Data Science que hayan nacido el mismo día?
  - ¿Cree que esta probabilidad es baja?
  - Consideremos el caso con 2 personas.

- Juego del cumpleaños.
  - El experimento lo podemos describir como “sacar aleatoriamente” una fecha de nacimiento para la primer persona y luego otra para la segunda persona.
  - La probabilidad buscada será entonces el producto de las probabilidades de “éxito” en cada parte del experimento.
  - Buscamos entonces la probabilidad de que la primer persona nazca un día cualquiera y que la segunda nazca un día cualquiera excepto el que nació la primera.
  - La probabilidad de que no hayan nacido el mismo día se calcula como:

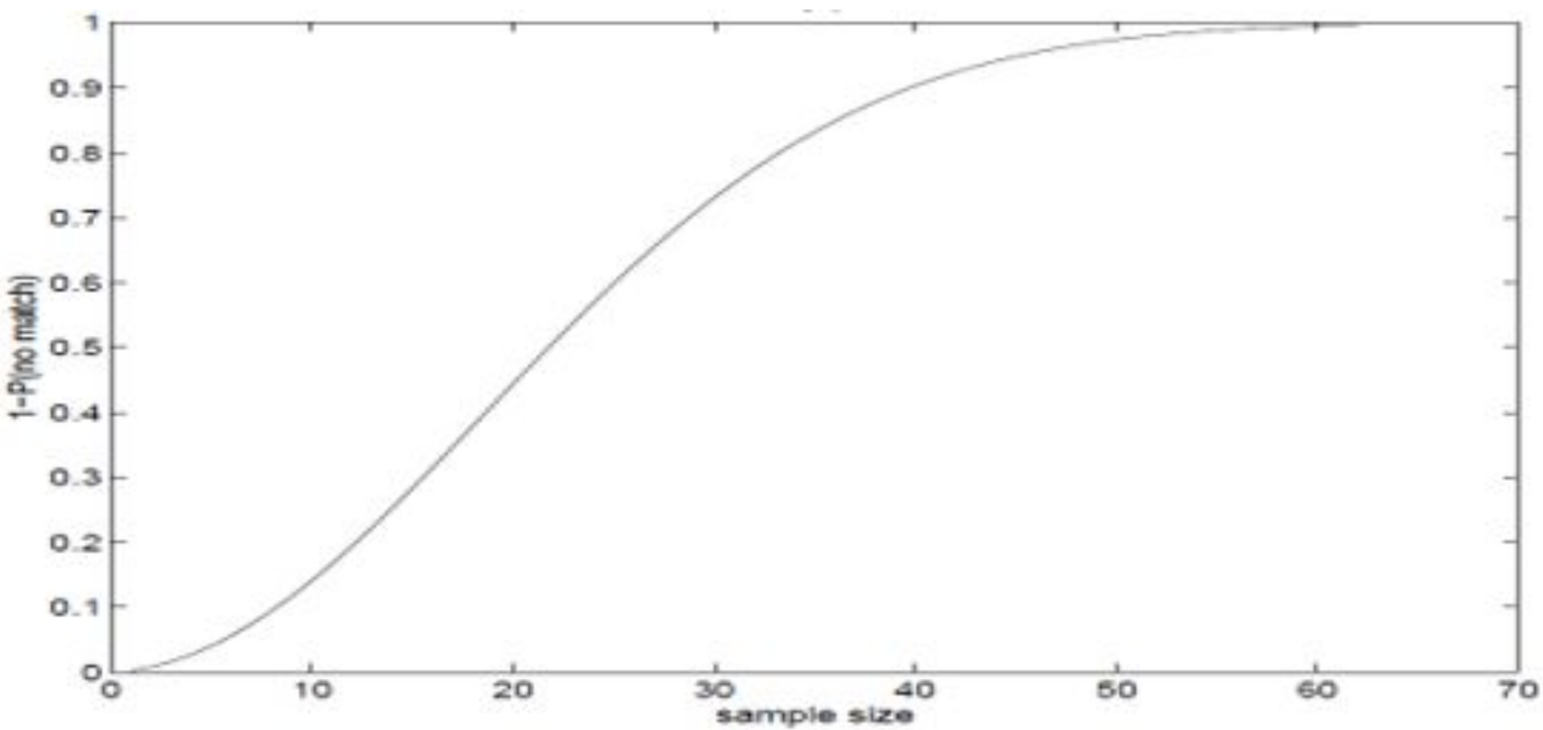
(365/365) * (364/365) = (364/365) = 0,99726

- Juego del cumpleaños.
  - Si hubiera tres personas (365/365) * (364/365) *(363/365) = 0,99726
  - Que se puede escribir como (365 * 364 * 363)/365^3
  - Para n personas...

    ((365-1)/365) * ((365-2)/365) * ((365-3)/365) * . . . * ((365-n+1)/365).
  - Que se puede reescribir como 365! / ((365-n)! * 365^n).

- Juego del cumpleaños.
  - ¿Qué pasa si hay más de 365 personas? ¿Puedo asignar los días del año para que no haya dos personas con la misma fecha de nacimiento? No!
  - Por eso la fórmula anterior daría 0 en ese caso.
  - En este ejemplo la probabilidad pedida parecería muy compleja de calcular. El problema se simplifica al ver las posibles fechas como la característica a asignar entre las personas.

- Juego del cumpleaños.



- Factoriales.
  - Recordar n!=1*2*3*…*(n-1)*n y por definición 0!=1
  - Se interpreta como la cantidad de formas diferentes de ordenar n elementos distintos
  - Ejemplo: usted tiene 3 libros de Python. ¿De cuántas formas puede ponerlos en un estante para impresionar a sus amigos?
    1. Al poner el primero tiene 3 opciones de libros para ese lugar.
    2. Una vez elegido el primero le quedan 2 opciones para libro que va a ir en el segundo lugar.
    3. Una vez elegidos los primeros 2 libros (en posiciones 1 y 2) le queda solo una opción de libro para el tercer lugar.

- Portafolio.
  - Las acciones A y B tiene igual probabilidad de subir o bajar un 10% siendo estadísticamente independientes.
  - Un portfolio igualmente ponderado de las acciones A y B tiene como retornos posibles el conjunto {10%, 0, -10%}. El retorno será 10% si los dos suben 10%, 0% si A baja y B sube o bien A sube y B baja mientras que será -10% si ambas bajan

Acción	A	B
	x	p(x)
Baja	-10%	1/2
Sube	10%	1/2

- Portafolio.
  - Los posibles resultados son

Portafolio	x	p(x)
Baja Baja	-10%	1/4
Baja Sube	0	1/4
Sube Baja	0	1/4
Sube Sube	10%	1/4

- Portafolio.
  - La distribución de probabilidad de los retornos es entonces

Retorno del Portafolio (x)	p(x)
-10%	1/4
0%	2/4
10%	1/4