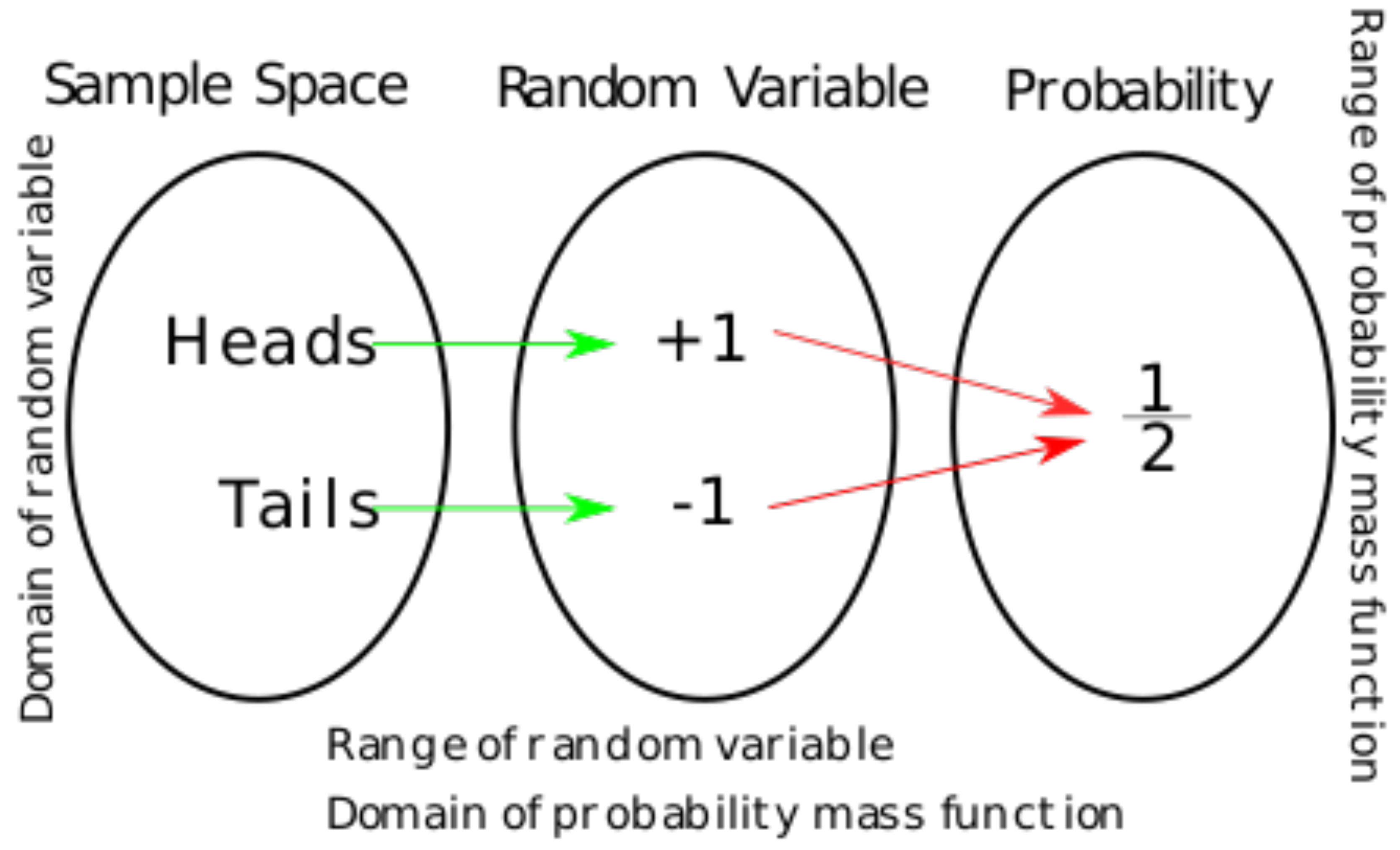
Probabilidad

Maestría en Ciencia de Datos, CUCEA, Universidad de Guadalajara.

Guadalajara, Jal., agosto de 2025

¿Que significa tener algo aleatorio?

- Tener algo aleatorio significa tener algo que carece de un patrón, o que no puede predecirse con exactitud.
- **Ejemplo comunes.** La tirada de una moneda, de un dado, toma una carta de una baraja.



Introducción a la probabilidad

Midiendo la incertidumbre

- La probabilidad de un resultado un fenómeno aleatorio es la proporción de veces que ocurre un resultado en una serie grande de repeticiones. El valor puede ir de 0 a 1.

- Por ejemplo, si tener águila en la tirada de una moneda es considerada un acierto, deberíamos tener un acierto 1/2 del tiempo, i.e., la probabilidad de tener un resultado de acierto en una tirada es de 0.5.

- $P(E_{aguila}) = 0.5$

Introducción a la probabilidad

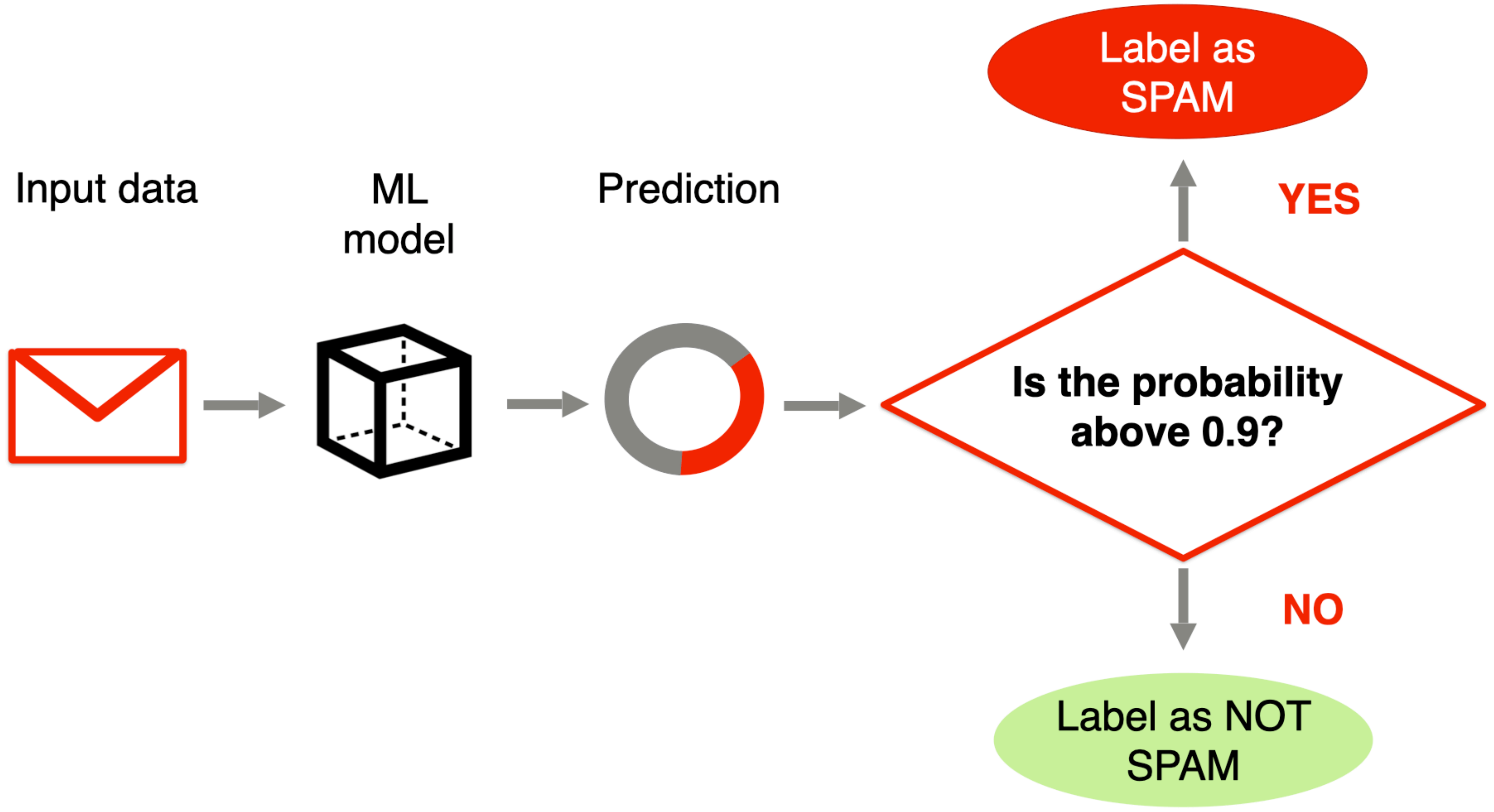
Machine Learning

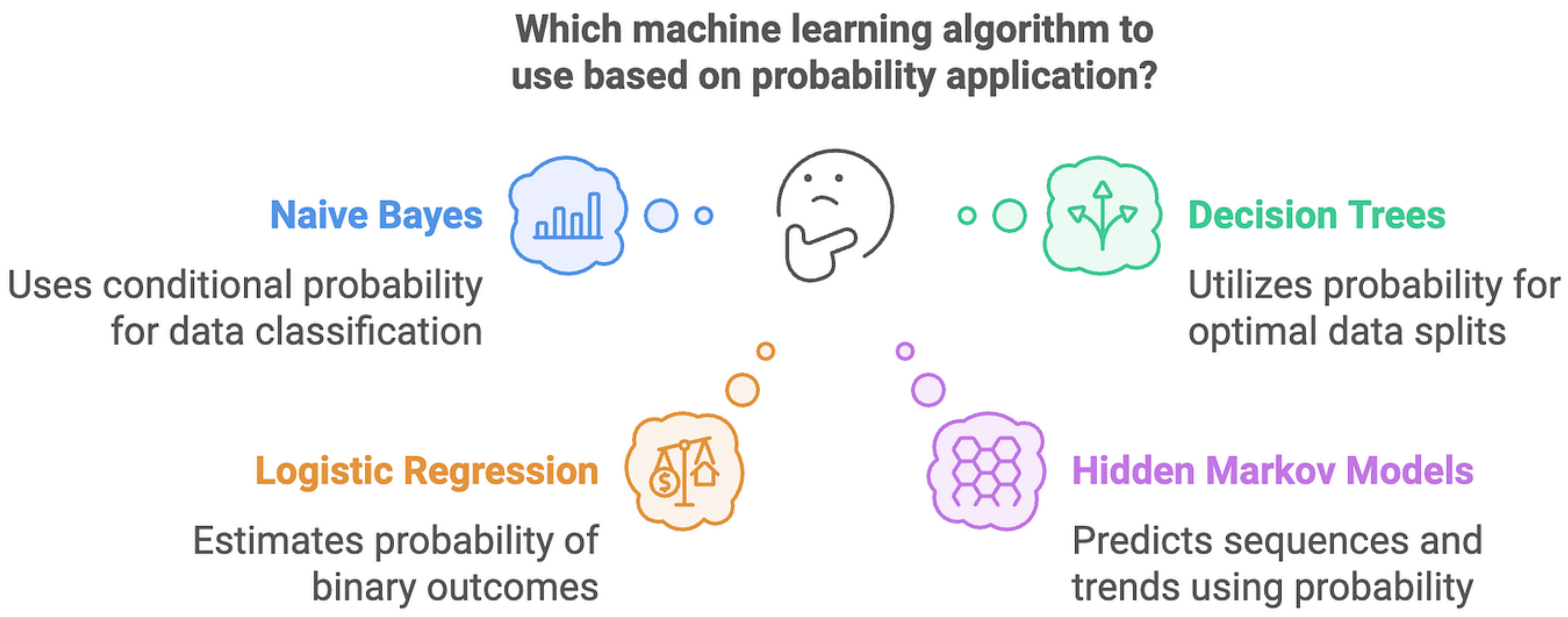
- En Machine Learning trabajamos constantemente con incertidumbre:
  - "¿Qué tan probable es que un correo sea spam?"
  - "¿Qué tan seguro está un modelo en su predicción?"
  - "¿Cómo actualizamos nuestras creencias cuando llega nueva información?"

La probabilidad nos da el lenguaje y las herramientas matemáticas para enfrentar este tipo de preguntas de manera rigurosa.

Introducción a la probabilidad

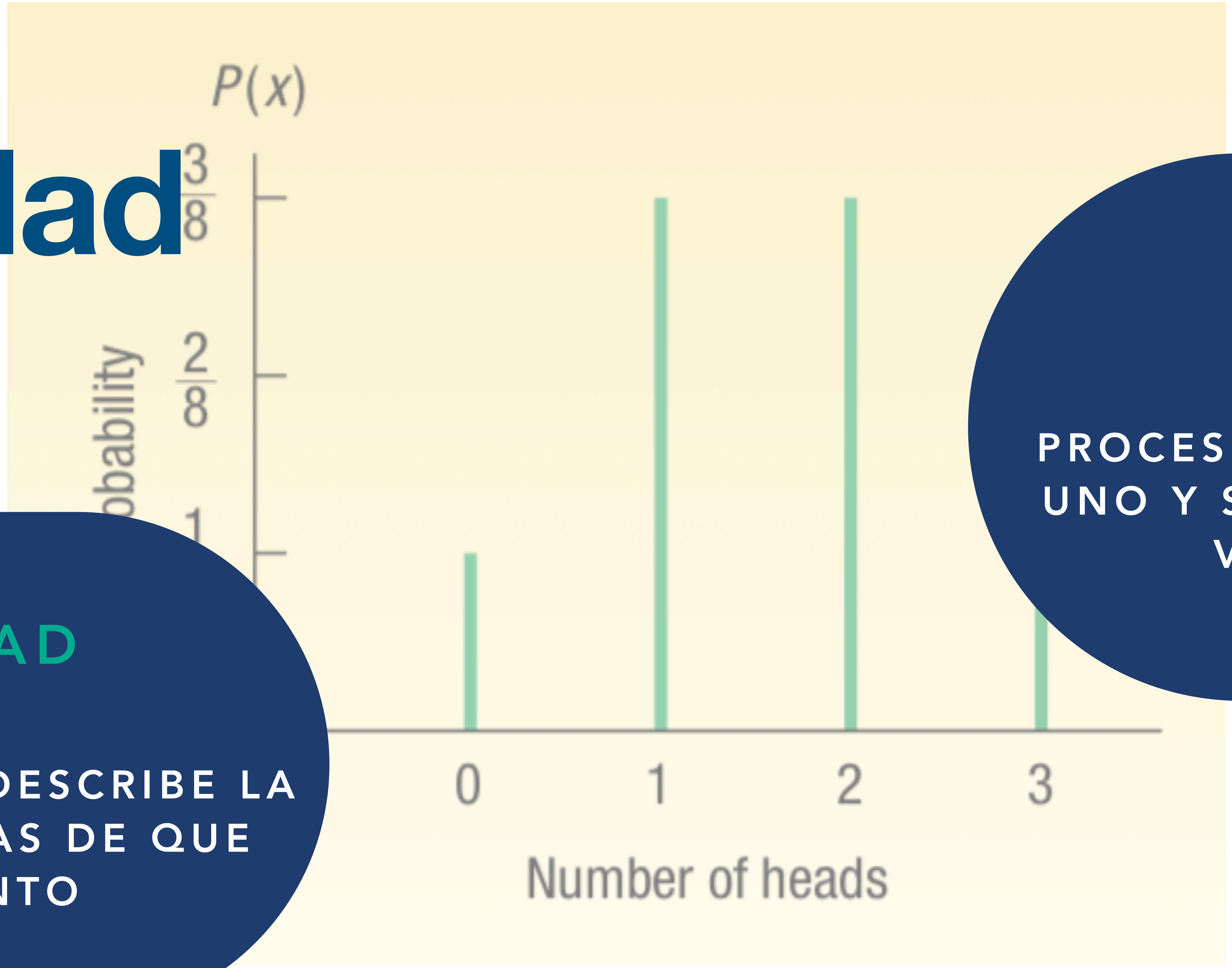
Machine Learning





Probabilidad

Terminología



ENSAYO

PROCESO QUE INDUCE A QUE OCURRA UNO Y SOLO RESULTADO UNA DE LAS VARIAS OBSERVACIONES

PROBABILIDAD

VALOR ENTRE 0 Y 1 QUE DESCRIBE LA PROBABILIDAD RELATIVAS DE QUE OCURRA UN EVENTO

EXPERIMENTO

PUEDE INCLUIR LA REALIZACIÓN DE UNO O VARIOS ENSAYOS

RESULTADO

EL RESULTADO PARTICULAR DE UN EVENTO

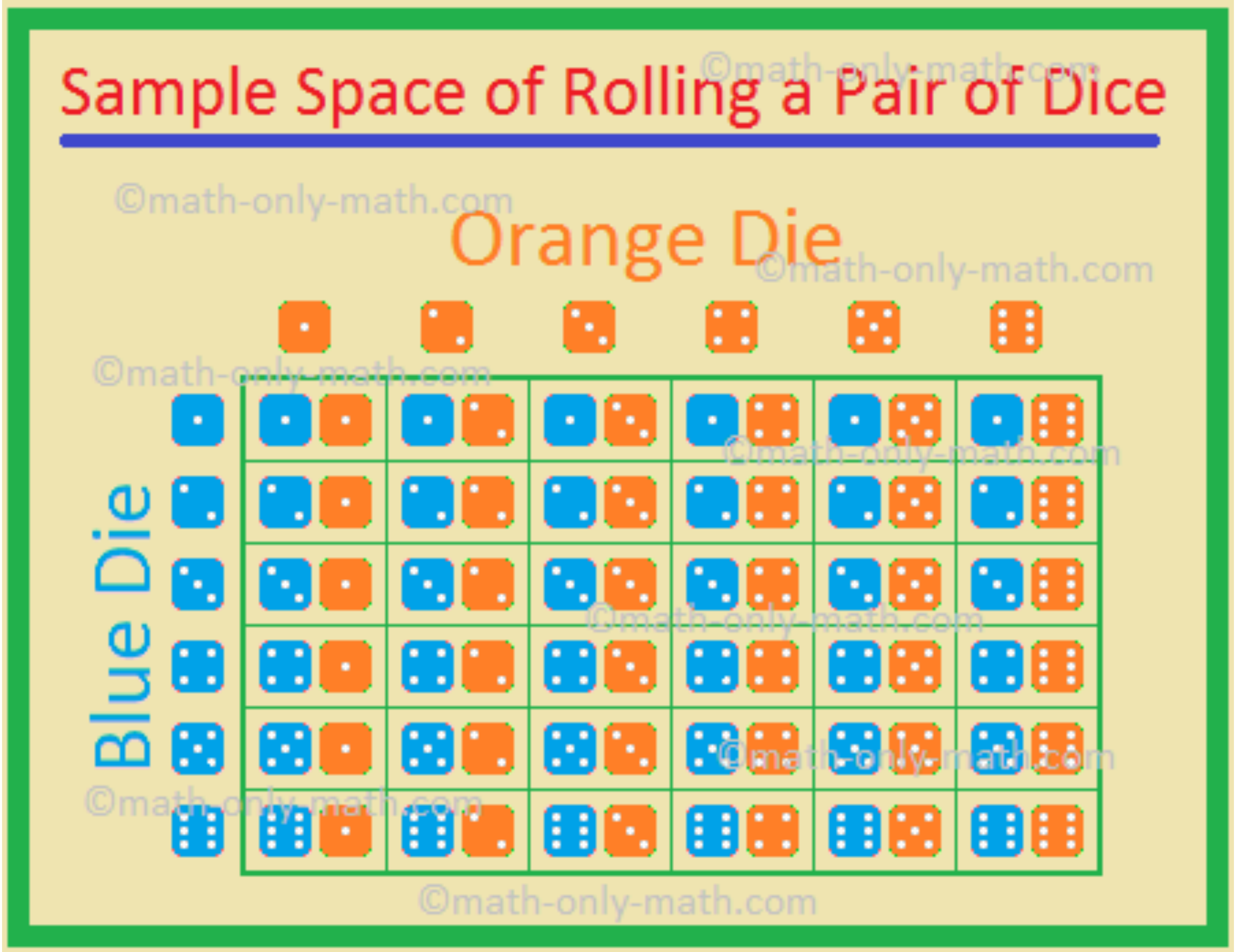
EVENTO

CONJUNTO DE UNO O MÁS RESULTADOS DE UN EXPERIMENTO

Probabilidad

Espacio muestral

- Un espacio muestral es una colección o conjunto de todos los resultados posibles de un experimento aleatorio. Normalmente es representado por S Ejemplo: $S = \{1,2,3,4,5,6\}$, ó $S = \{\text{Aguila, Sello}\}$.



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

Probabilidad

Probabilidad clásica

- Basada en la suposición de que todos los resultados de un experimento son igualmente probables.
- $$\text{Probabilidad de un evento} = \frac{\text{Numero de resultados favorables}}{\text{Numero Total de resultados posibles}}$$
- Ejemplo: Probabilidad de tener un número par. En este caso los resultados “favorables” son 2,4,6, y la probabilidad de este evento sería 0.5

Probabilidad

Probabilidad empírica

- **Ley de grandes números.** Para experimentos independientes, mientras aumenta el número de éstos, la frecuencias relativa de eventos repetidos se acerca cada vez más a un valor singular. Ejemplo: Al tirar una moneda varias veces es dónde se aprecia la prob. 0.5

- Si observamos repetidamente un resultado de un evento, entonces la fracción de veces que se tiene dicho acierto es la **probabilidad empírica**.

- Al tirar una sola vez una moneda, la frecuencia relativa está entre 0 y 1. Al tirar varias veces se aproxima a 0.5

$$\text{Frecuencia Relativa } (A) = \frac{\text{Numero de veces que ocurre A}}{\text{Total de experimentos}}$$

Probabilidad

Probabilidad subjetiva

- Este tipo de probabilidad es subjetiva
- Se basa en la experiencia de una persona (puede ser un experto), y a veces es basado en una ocurrencia similar del mismo evento en el pasado.
- Ejemplo: Como está el clima. Hay probabilidad 30% de que llueva.

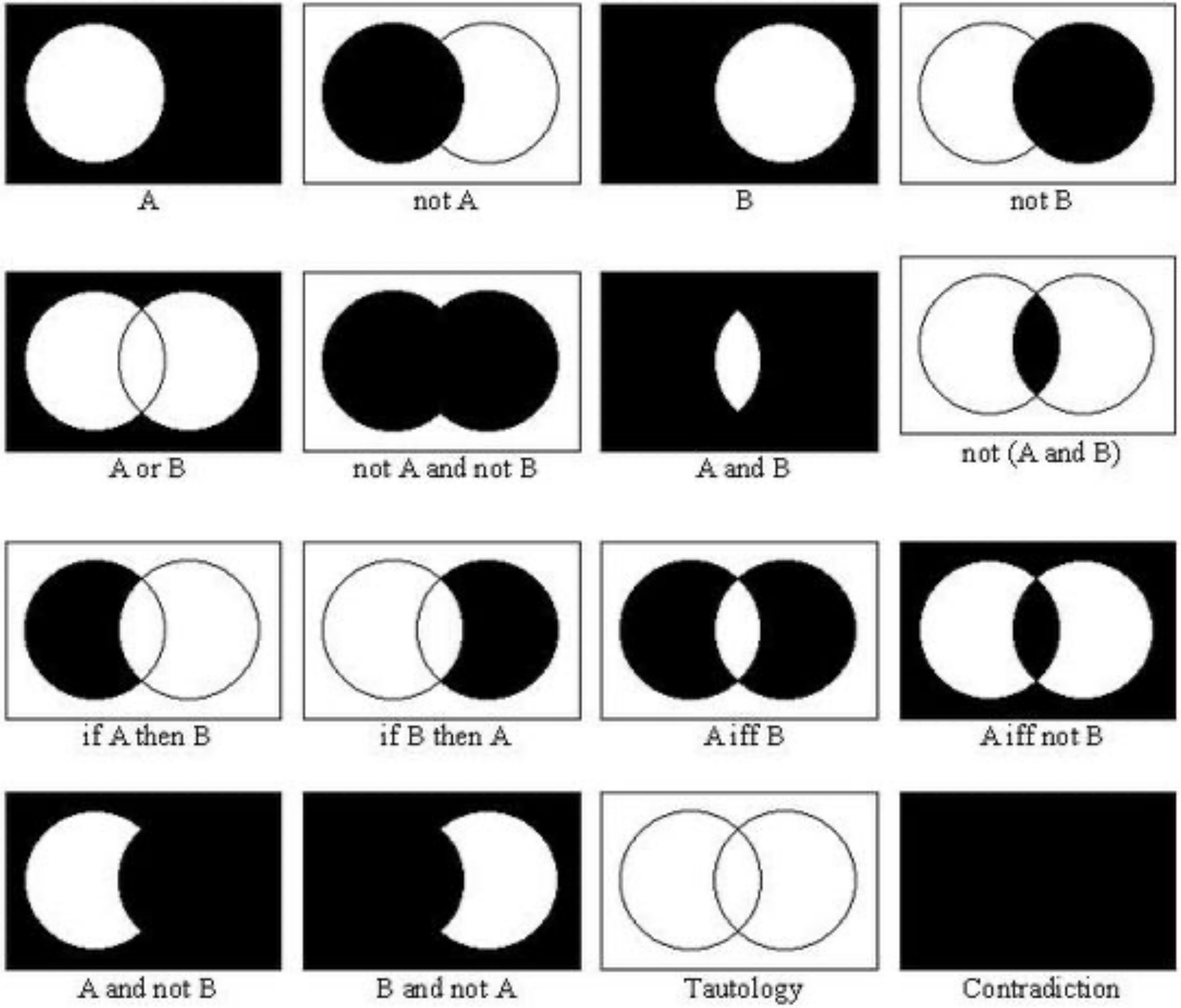
Probabilidad

Kolmogorov's axioms

- $P(C) \geq 0$, para todos los eventos C. La probabilidad de un resultado particular está entre 0 y 1.
- $P(S) = 1$. La lista es exhaustiva, la suma de probabilidades de los varios eventos es 1.
- Para una colección de eventos mutuamente exclusivos $\hat{C} = \{C_1, C_2, \ldots\}$, se tiene que $P\left(\bigcup_{C \in \hat{C}} C\right) = \sum_{C \in \hat{C}} P(C)$

Teoría de conjuntos

Conjuntos
Overview

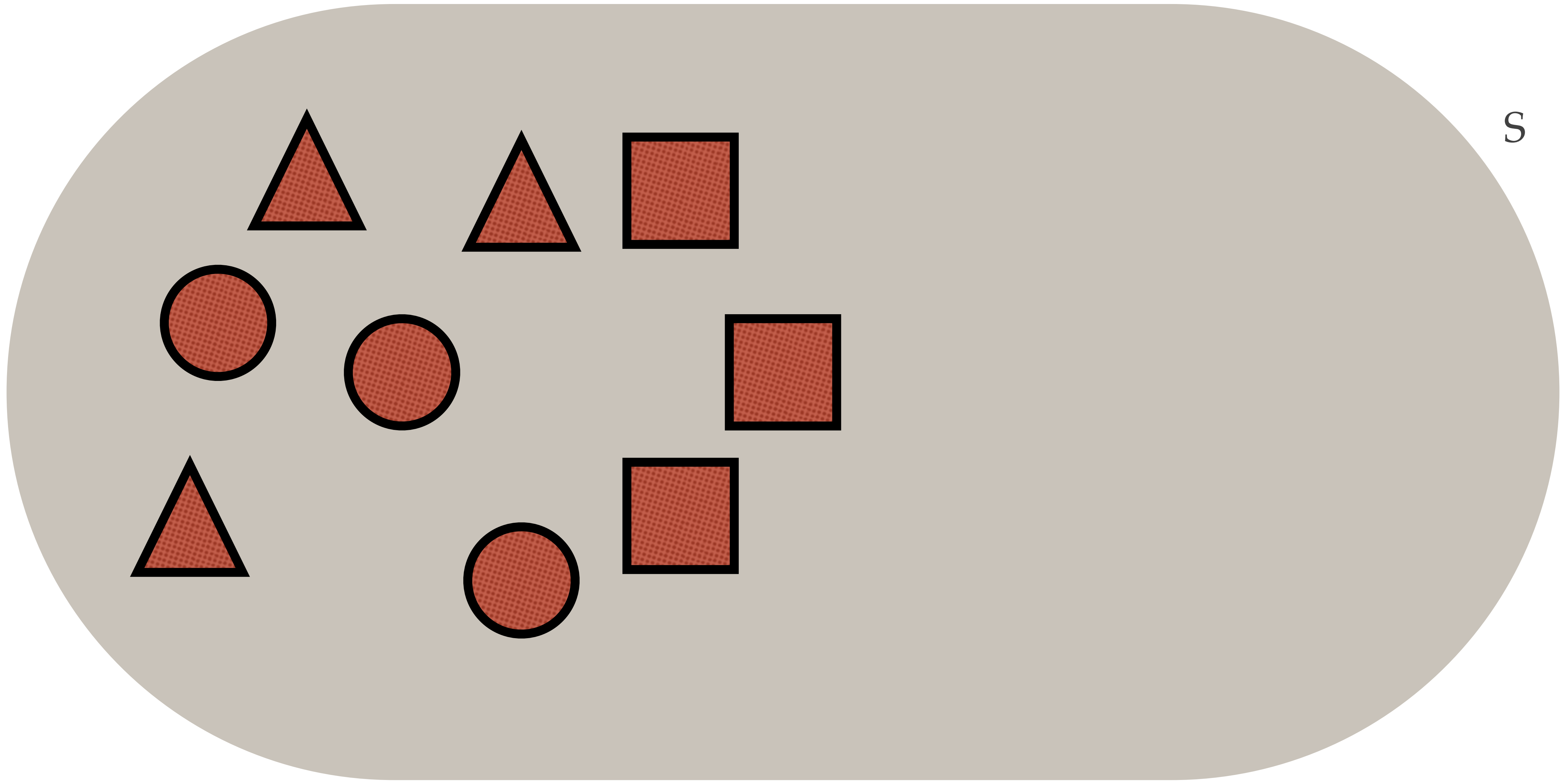


Conjuntos

Intersección

- En probabilidad, una intersección describe el espacio muestra donde dos eventos ocurren simultáneamente.
- Consideremos una caja con pelotas de color y algunas formas.

- 9 de las figuras son rojas

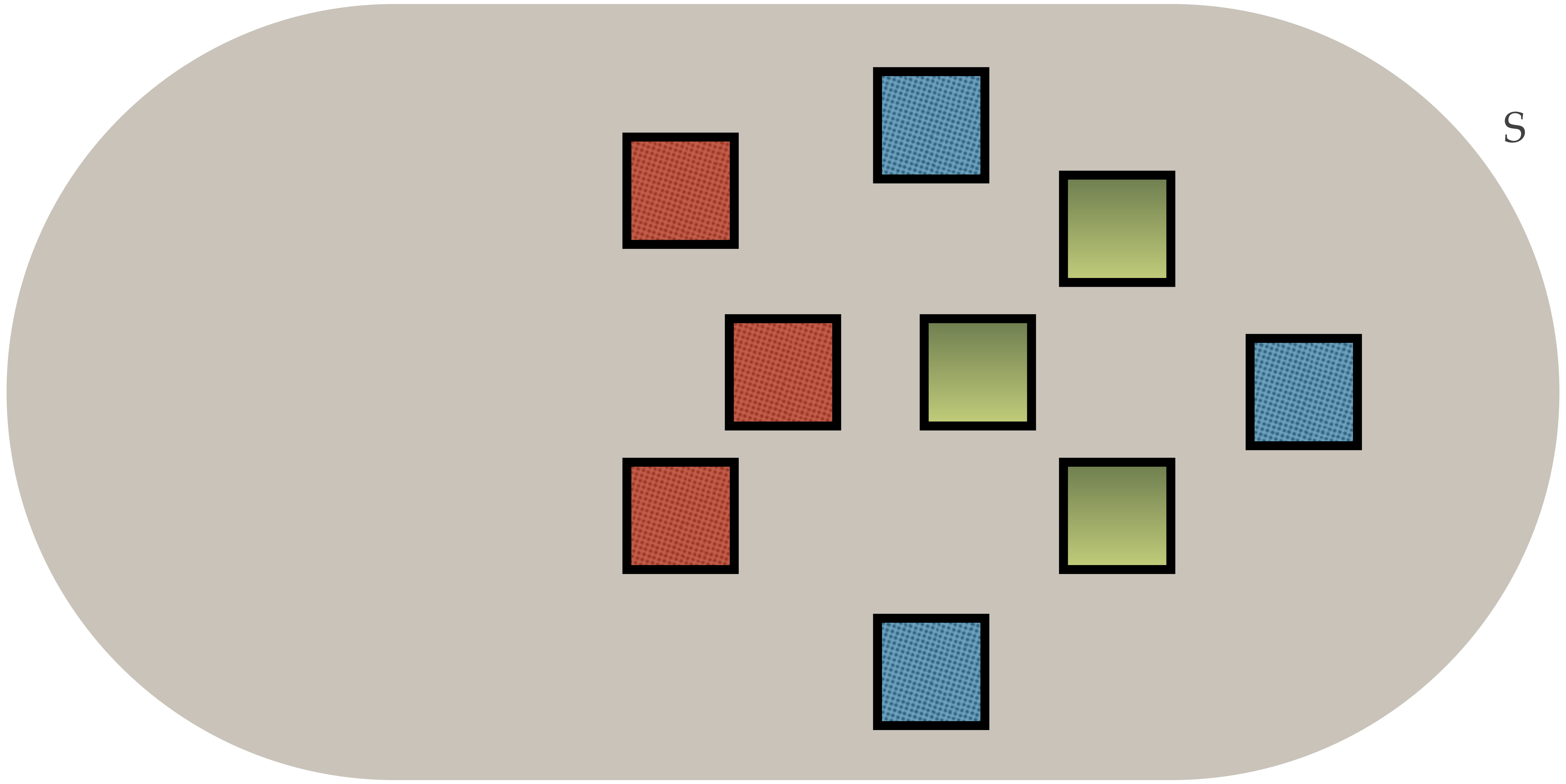


Conjuntos

Intersección

- En probabilidad, una intersección describe el espacio muestra donde dos eventos ocurren simultáneamente.
- Consideremos una caja con pelotas de color y algunas formas.

- 9 de las figuras son cuadrados.

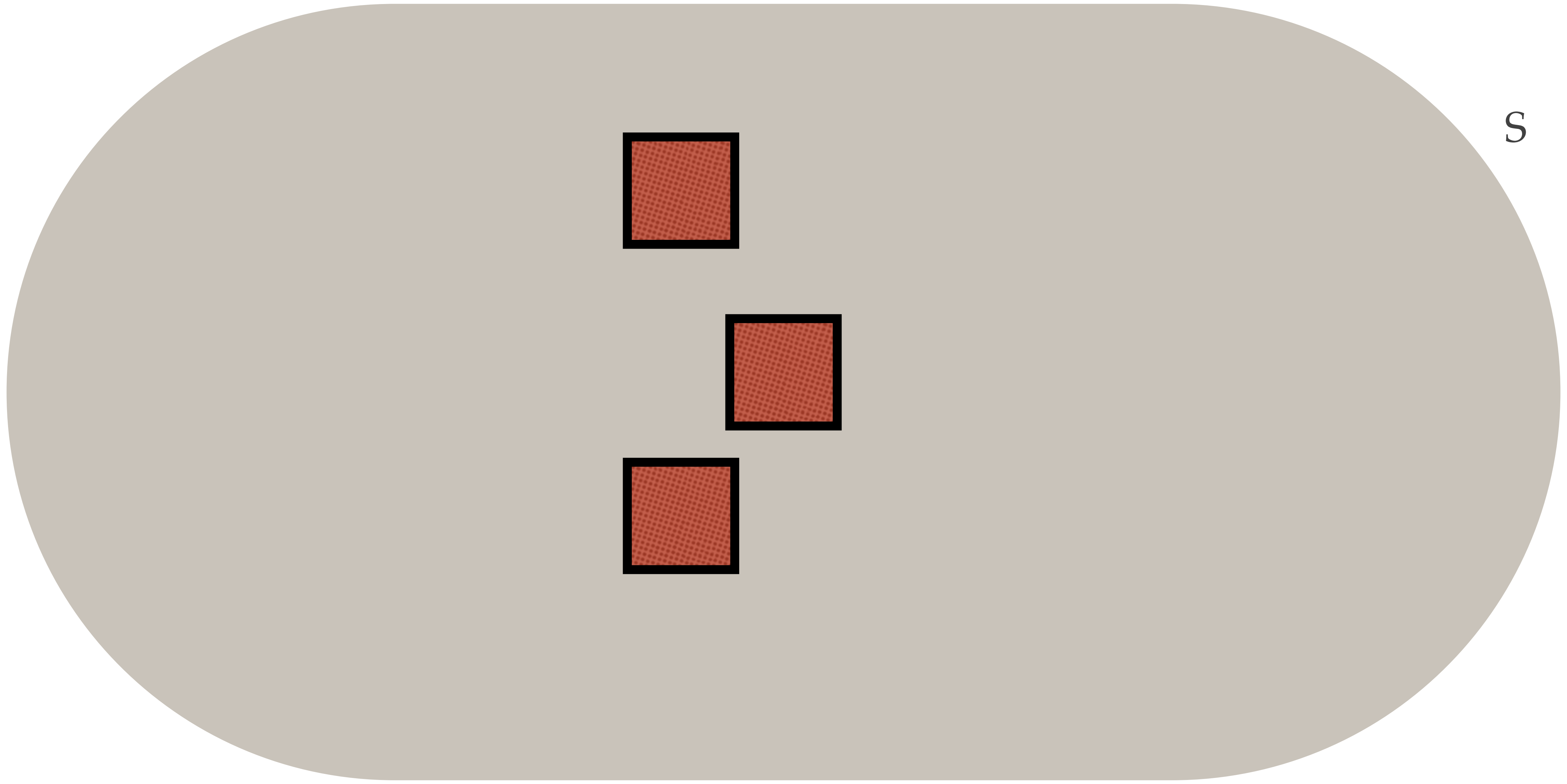


Conjuntos

Intersección

- En probabilidad, una intersección describe el espacio muestra donde dos eventos ocurren simultáneamente.
- Consideremos una caja con pelotas de color y algunas formas.

- 3 de las figuras son rojas y cuadrados.

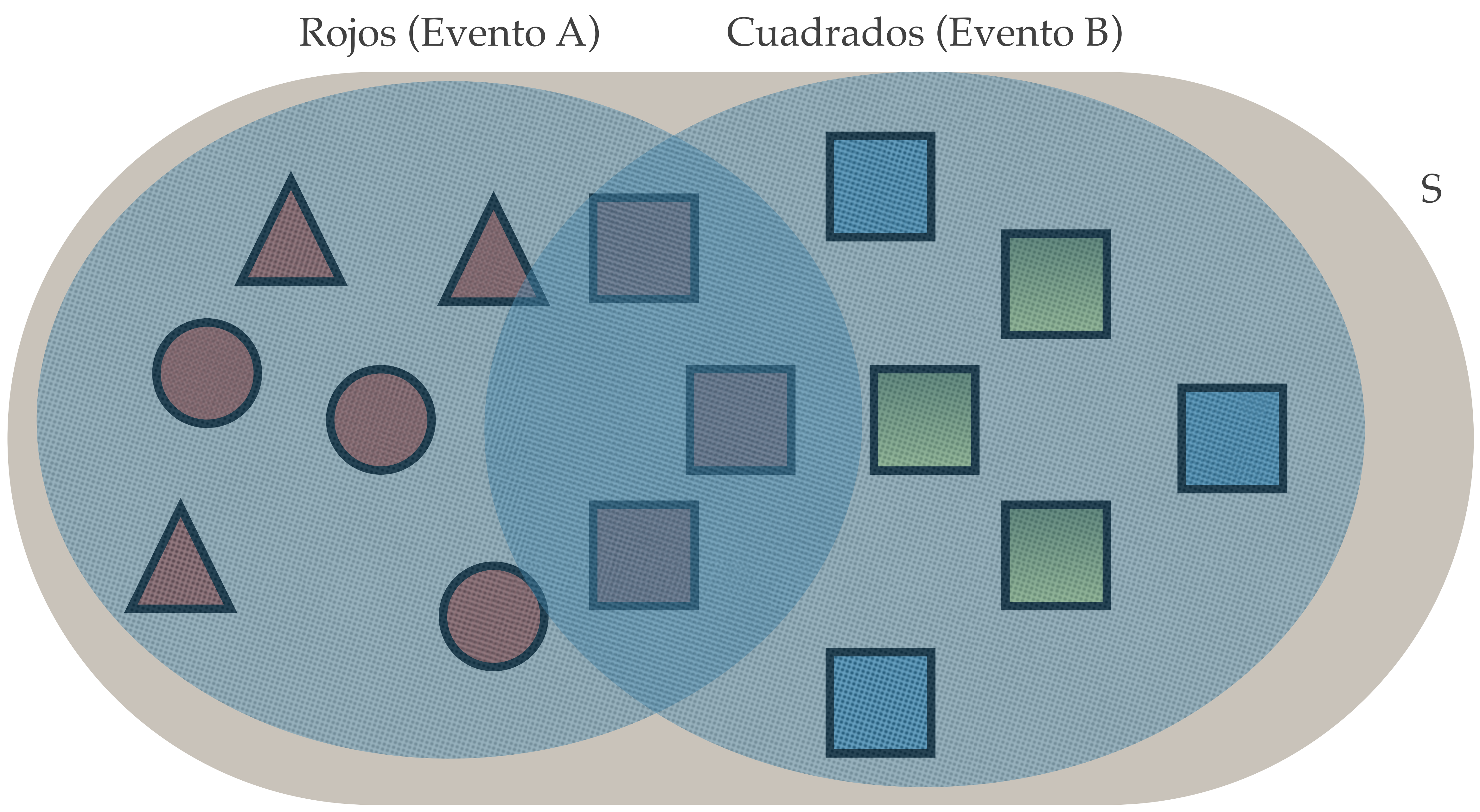


Conjuntos

Intersección

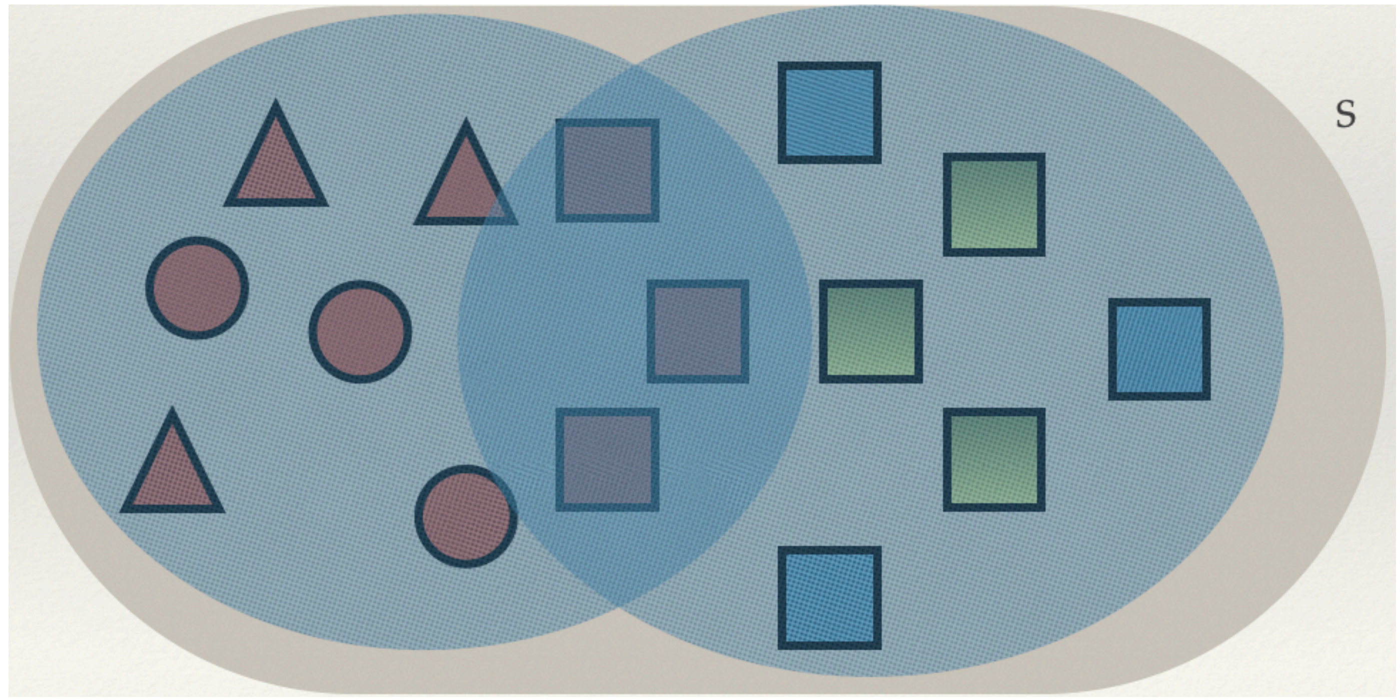
- En probabilidad, una intersección describe el espacio muestra donde dos eventos ocurren simultáneamente.
- Consideremos una caja con pelotas de color y algunas formas.

- Cuál es la probabilidad de tener un cuadrado rojo?



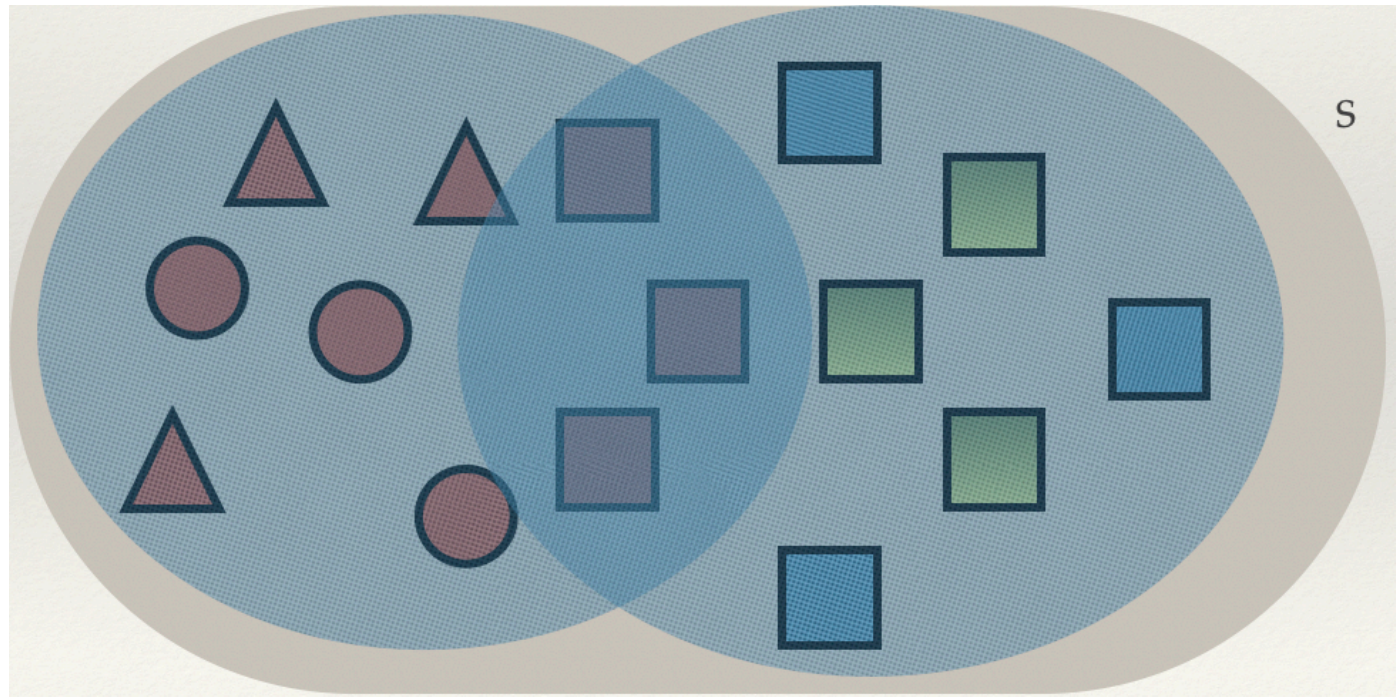
Conjuntos

Intersección



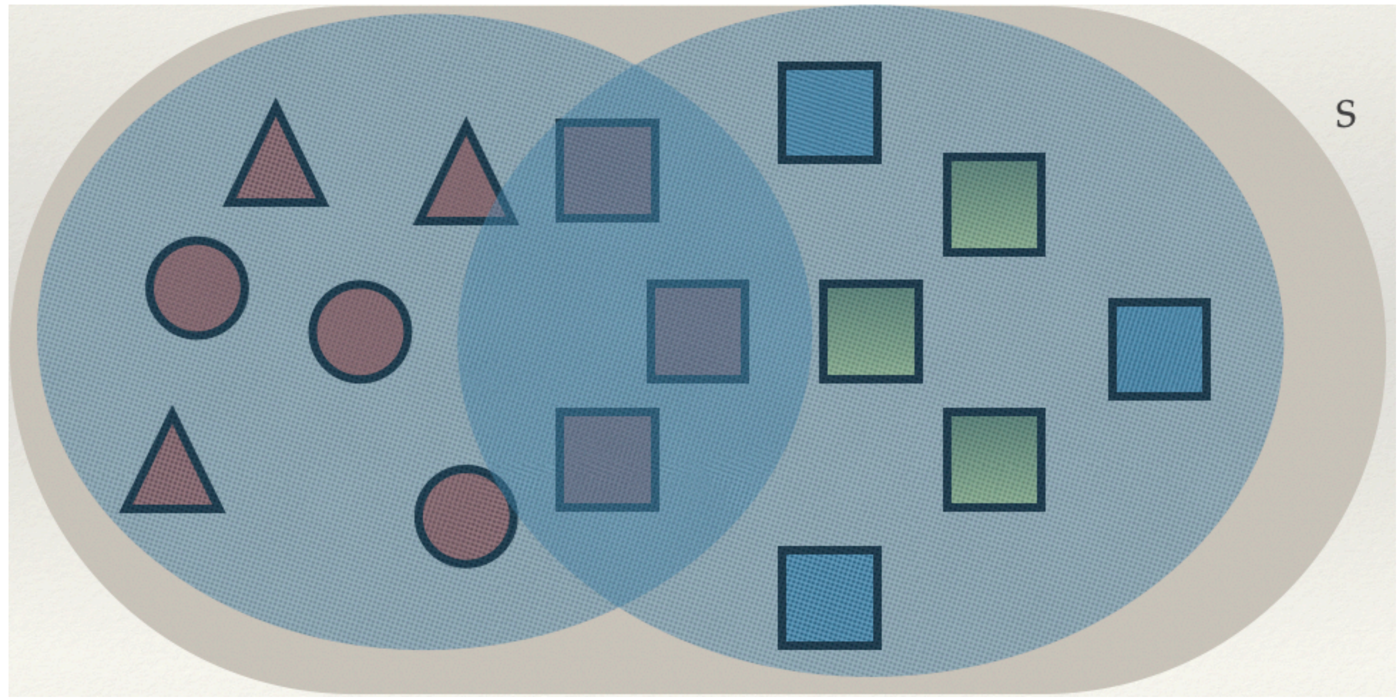
- Si asignamos A como el evento de elegir objetos rojos.
- Asignamos B al evento de elegir objetos cuadrados.
- La intersección de A con B está dada por:
- $A \cap B$
- El orden aquí no es importante (un evento no afecta el otro).
- La probabilidad de A ($P(A)$) y B ($P(B)$) se denota por $P(A \cap B)$.
- En este caso se interpreta como:
- $P(A \cap B) = \frac{3}{15} = 0.2$

Conjuntos
Unión



- La unión de dos eventos considera si A **o** B ocurren (una **u** la otra), y se denota por $A \cup B$.
- Otra vez, aquí el orden no es importante.
- La probabilidad de A **o** B está dada por:
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- En este caso:
- $P(A \cup B) = \dfrac{9}{15} + \dfrac{9}{15} - \dfrac{3}{15} = 1.0$

Conjuntos
Complemento



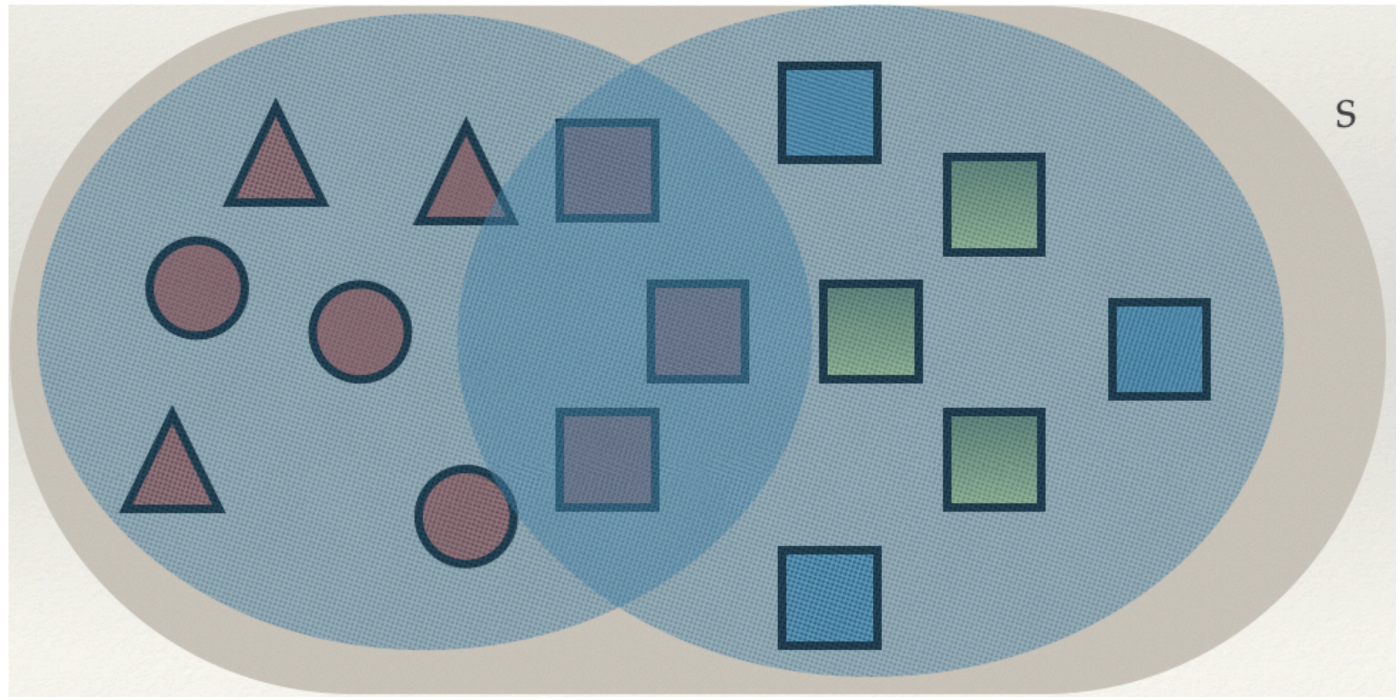
- El complemento de un evento considera todo resultado exterior al evento, denotado por $\bar{A}$.

- La probabilidad de no tener un evento A está dado por:

- $P(\bar{A}) = 1 - P(A) = \dfrac{15}{15} - \dfrac{9}{15} = \dfrac{6}{15} = 0.4$

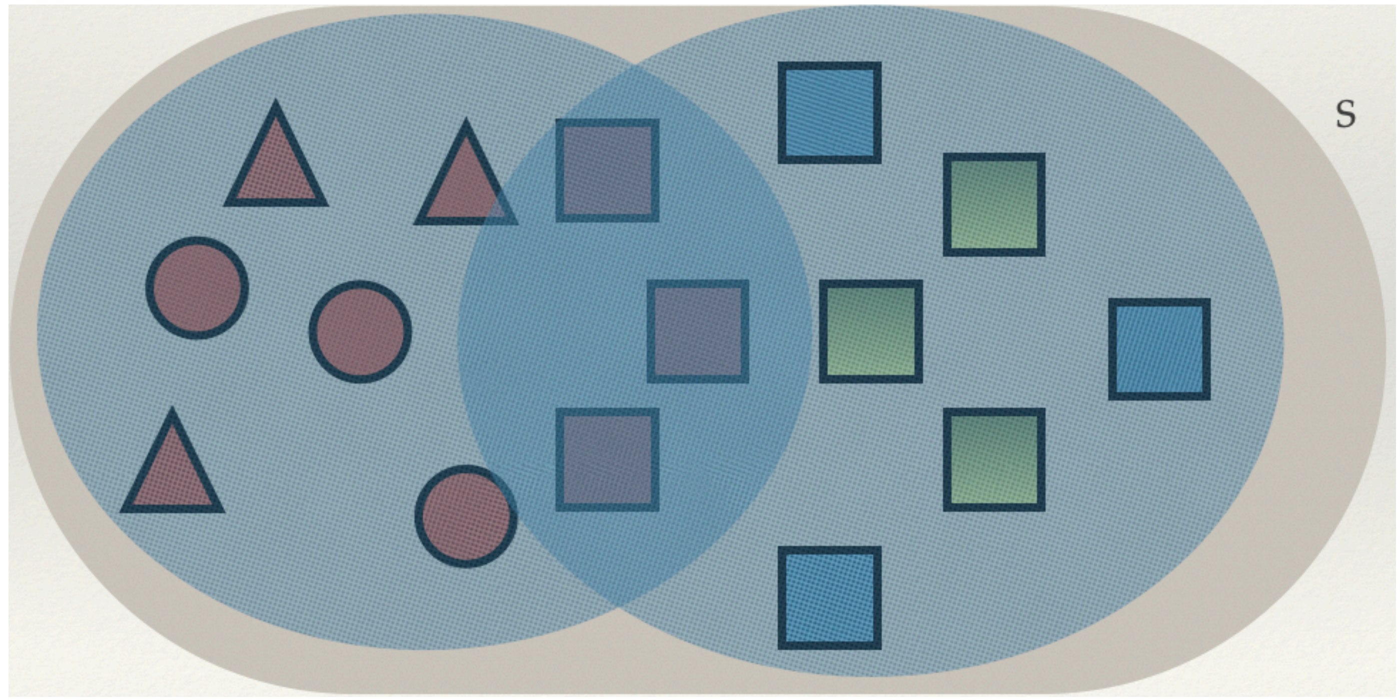
Conjuntos

Ejemplo



- La probabilidad de que una carta escogida al azar de una baraja sea rey **O** corazón.
- Hay 52 cartas totales, 13 corazones, 13 espada, 13 trébol, 13 diamantes. De esas, 4 son reyes.
- A (rey), B (corazón)
- $P(A) = \dfrac{4}{52}; \quad P(B) = \dfrac{13}{52}; \quad P(A \cap B) = \dfrac{1}{52},$

Conjuntos
Ejemplo



- La probabilidad de que una carta escogida al azar de una baraja sea rey **O** corazón.
- Hay 52 cartas totales, 13 corazones, 13 espada, 13 trébol, 13 diamantes. De esas, 4 son reyes.
- A (rey), B (corazón)
- $P(A) = \dfrac{4}{52}; \quad P(B) = \dfrac{13}{52}; \quad P(A \cap B) = \dfrac{1}{52},$

$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 4/52 + 13/52 - 1/52$

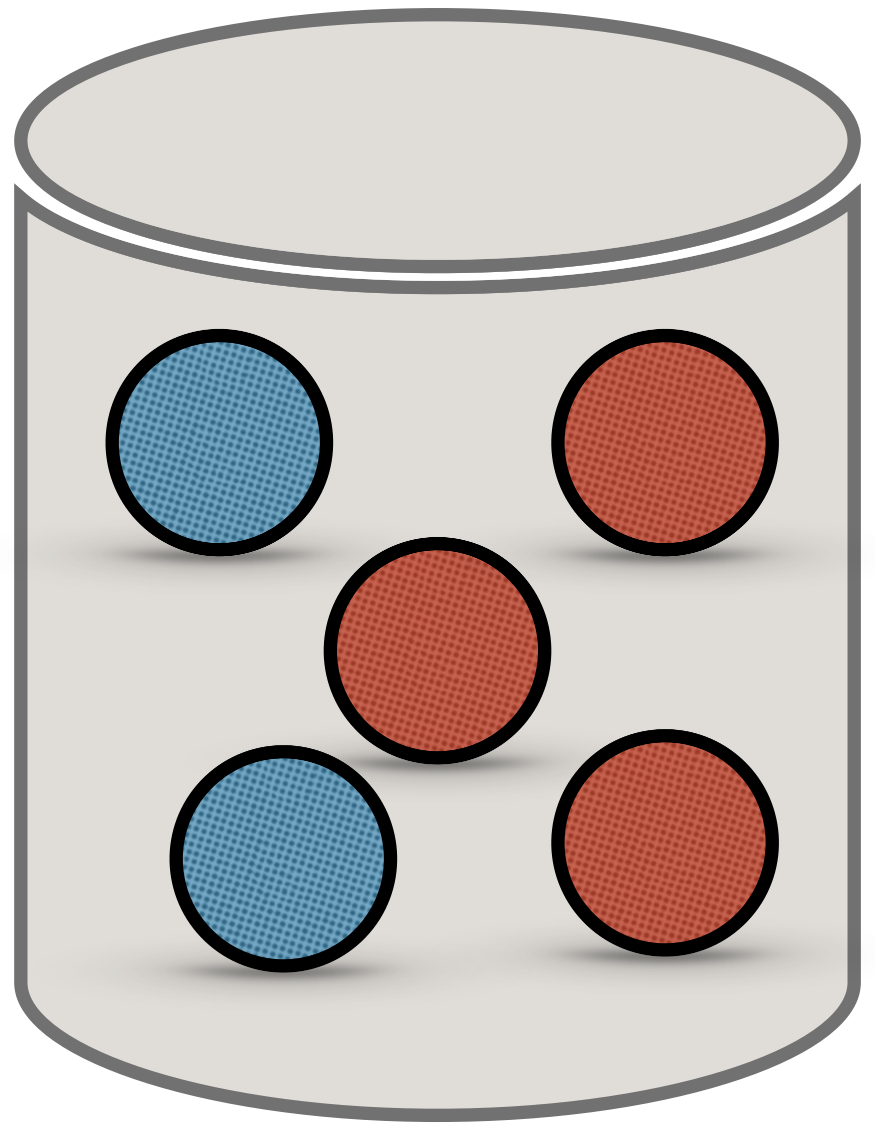
Eventos dependientes e independientes

Eventos dependientes e independientes

- Una serie de eventos **independientes** ocurren cuando el resultado de un evento no afecta el resultado de otro evento.
- Ejemplo: tirar una moneda dos veces. La probabilidad en cualquiera de las tiradas sigue siendo 1/2 para sello.
- Un evento **dependiente** ocurre cuando el resultado del un evento anterior afecta la probabilidad de un evento posterior.
- Consideremos por ejemplo simple un jarro con canicas.

Eventos

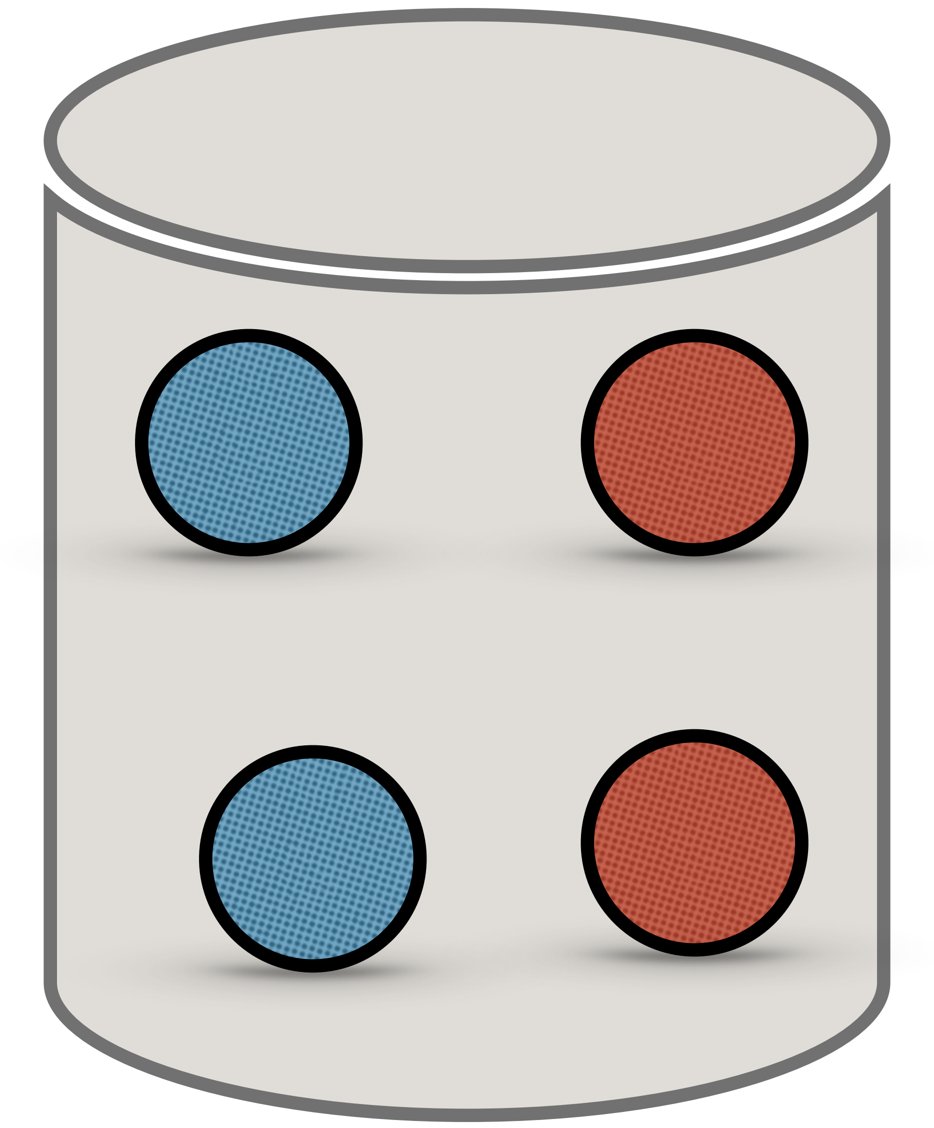
Eventos dependientes



- Consideramos eventos donde se toma una canica, sin reemplazo, i.e., sin devolverla en el frasco después de tomarla.
- **¿Si se toman dos canicas del frasco, cuál es la probabilidad de que ambas sean rojas?** $P(R_1 \cap R_2)$
- El color obtenido del primer intento afecta la probabilidad de obtener un color en el segundo intento.
- En el primer intento, la probabilidad de obtener una canica roja es simple $P(R_1) = \dfrac{3}{5}$.

Eventos

Eventos dependientes (Probabilidad condicional)



- La probabilidad de tomar una segunda canica roja **dada que la primera canica tomada fue roja** se denota por:
- $P(R_2 | R_1)$
- Esto es llamado una **probabilidad condicional.**
- Después de remover una canica roja en el primer intento entonces la probabilidad de la segunda es:
- $P(R_2 | R_1) = \dfrac{2}{4}$

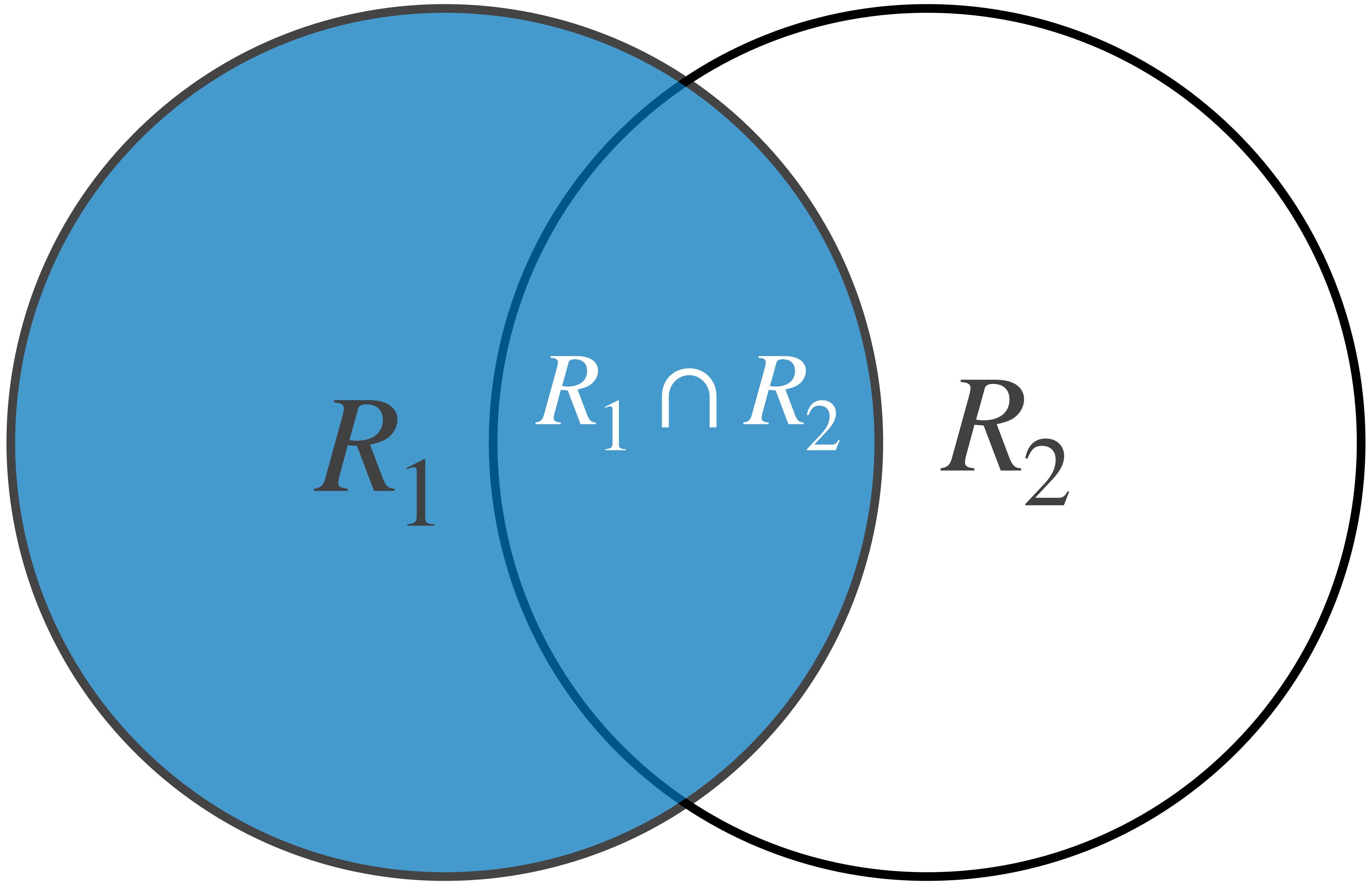
Eventos

Eventos dependientes (Probabilidad condicional)

- Al considerar esto, sabemos que R_1 ya ocurrió, lo que quiere decir que **nuestro espacio muestral se reduce** de lo que ya teníamos antes a sólo aquellos resultados presentes en R_1.

- En este caso las posibilidades del segundo se ven alteradas, y podemos ver que la probabilidad de tener una canica roja en el segundo intento estaría dado por:

- $\dfrac{P(R_1 \cap R_2)}{P(R_1)} = P(R_2 \mid R_1)$



Eventos dependientes

Probabilidad condicional

- De lo anterior, podemos hacer un despeje.

$$P(R_1 \cap R_2) = P(R_1) \cdot P(R_2 | R_1) = \frac{3}{5} \times \frac{2}{4} = 0.3$$

- Si R_2 fuera independiente de $R_1 \Rightarrow P(R_2 | R_1) = P(R_2)$

$$P(R_1 \cap R_2) = P(R_1) \cdot P(R_2)$$

- En general, la probabilidad de tener dos eventos que son dependientes está dada por:

$$P(A \cap B) = P(A) \cdot P(B | A) \quad ; \quad P(B | A) = P(B)$$

- Se puede extender a varios eventos

$$P(A \cap B \cap C) = P(A) \cdot P(B | A) \cdot P(C | A \cap B)$$

Probabilidad condicional

Regla general de la multiplicación

- En general, la probabilidad de tener dos eventos que son dependientes está dada por:

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

- Se puede extender a varios eventos

$$P(A \cap B \cap C) = P(A) \cdot P(B|A) \cdot P(C|A \cap B)$$

- En el caso de tener sólo eventos independientes, esto se reduce a la regla especial de la multiplicación.

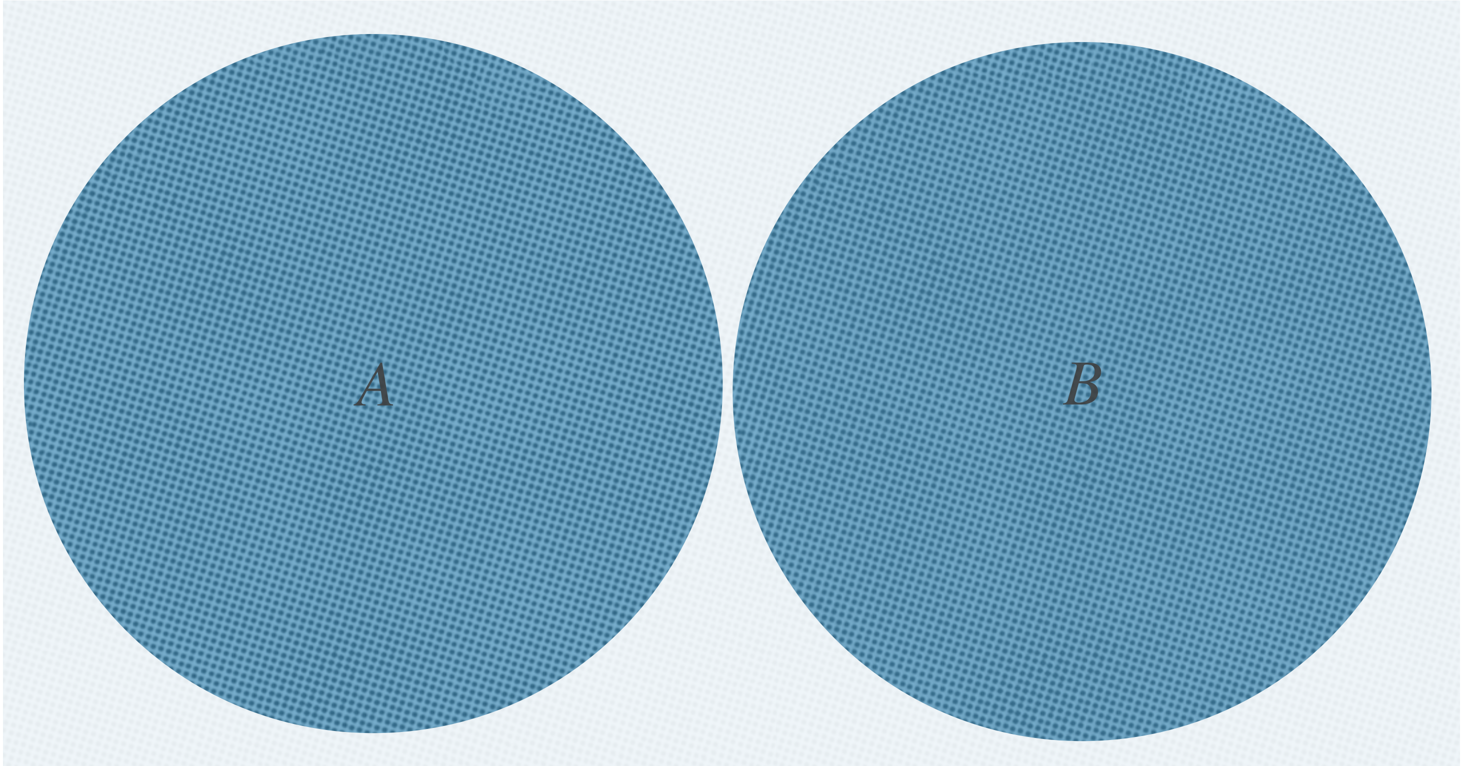
$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$$

Probabilidad

Regla general de adición (eventos mutuamente excluyentes)

- Los eventos mutualmente excluyentes son eventos que no pueden ocurrir al mismo tiempo.
- Para eventos independientes mutualmente excluyentes la regla de adición también se puede simplificar, puesto que $A \cap B = 0$.

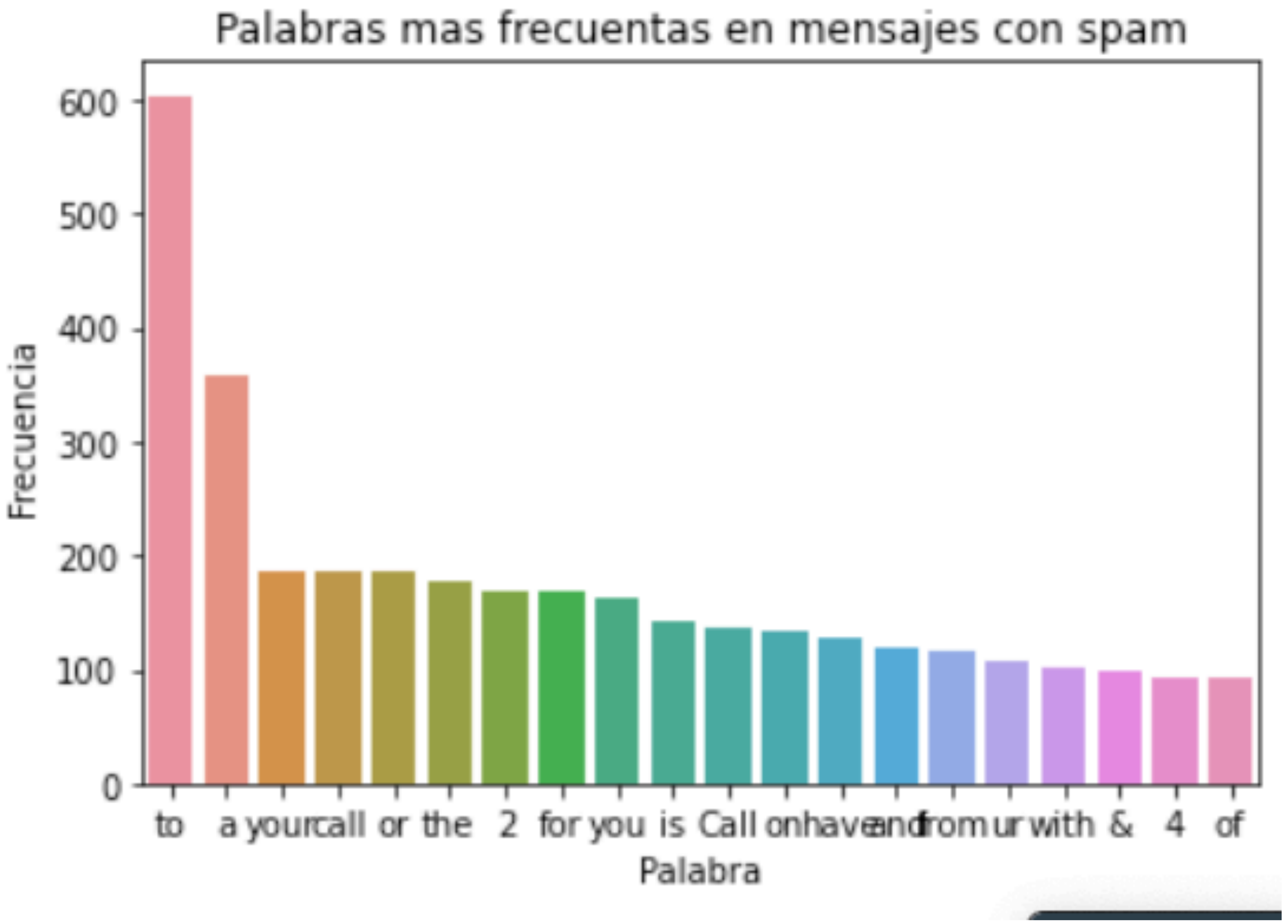
$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B)$$



Teorema de Bayes

$$P(\text{to} \mid \text{ham}) = 0.004$$

$$P(\text{spam} \mid \text{to do}) \propto P(\text{spam}) * P(\text{to} \mid spam) * P(\text{do} \mid spam)$$

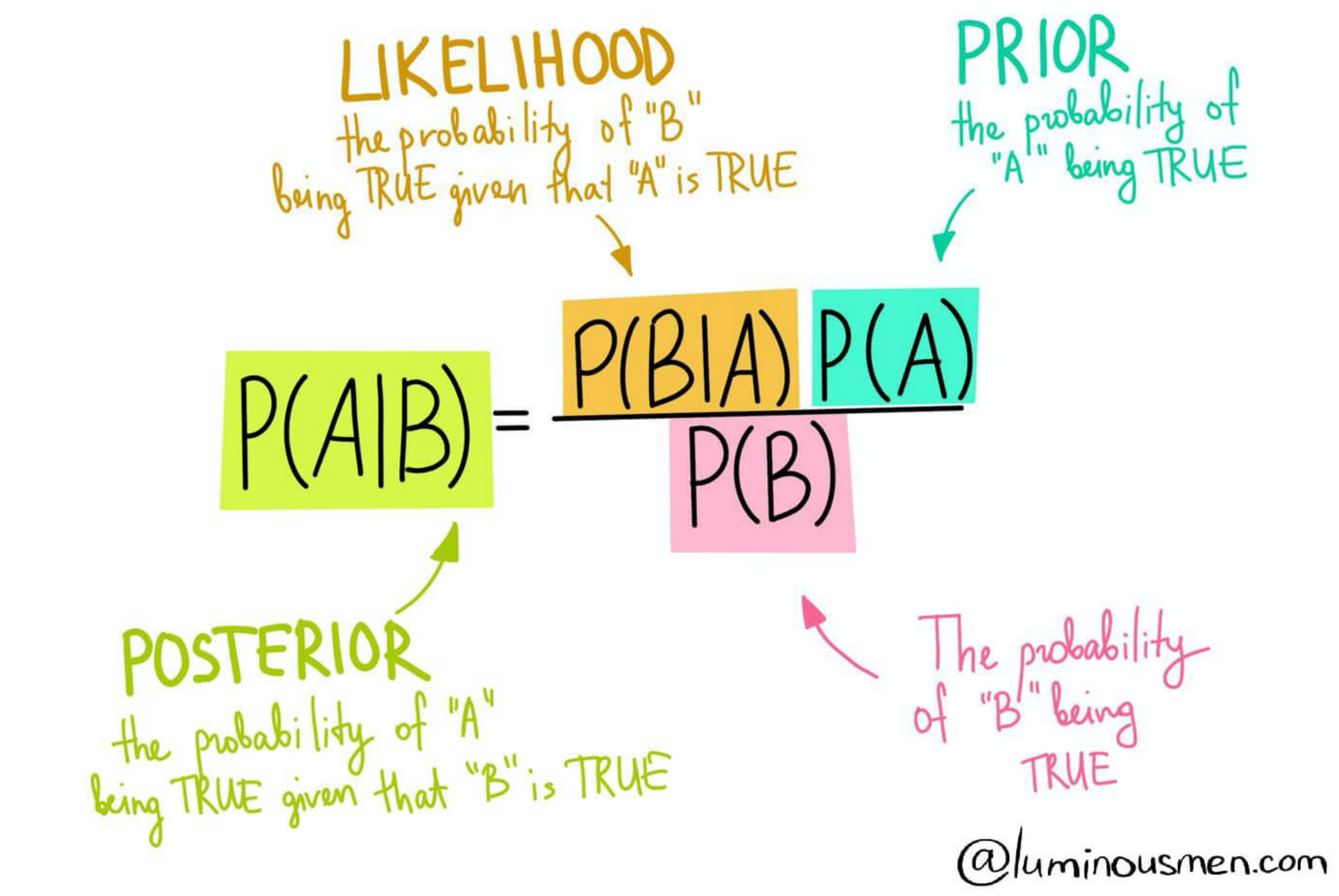


$$P(\text{to} \mid \text{spam}) = 0.0007$$

Teorema de Bayes

Overview

- Sea A el evento de sufrir de la enfermedad, y B el evento de dar un test positivo, sabemos $P(B|A)$ pero lo que queremos calcular es $P(A|B)$. O sea, queremos la probabilidad de que uno sufra de la enfermedad cuando ya se conoce que el test es positivo.

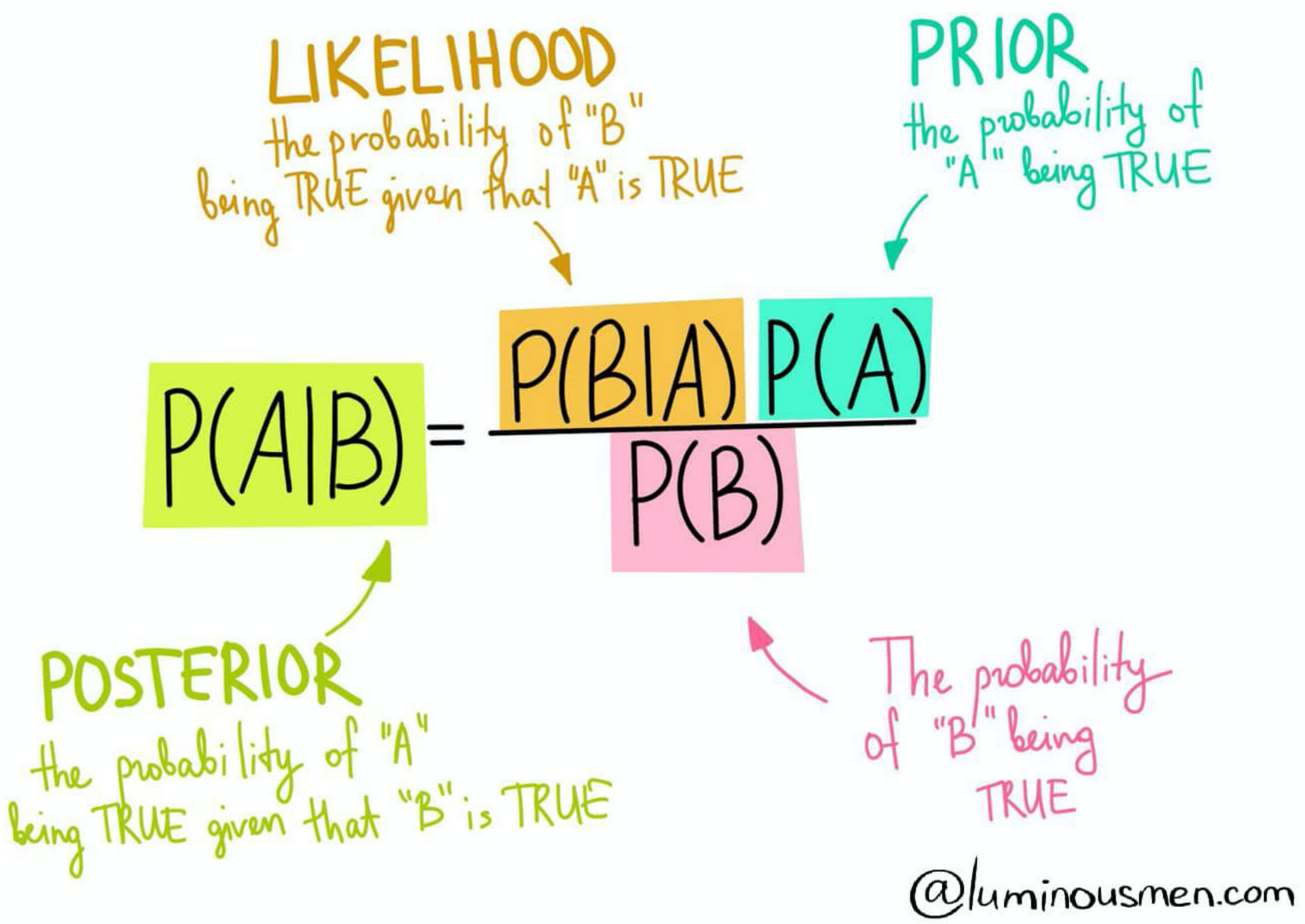


- Ya conocemos como se calcula una probabilidad condicional.
- $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$, suponiendo que $P(B) > 0$.
- En el mismo caso, se tiene que:
- $P(B|A) = \dfrac{P(B \cap A)}{P(A)} = \dfrac{P(A \cap B)}{P(A)}$, suponiendo que $P(A) > 0$.

Teorema de Bayes

Overview

- Sea A el evento de sufrir de la enfermedad, y B el evento de dar un test positivo, sabemos $P(B \mid A)$ pero lo que queremos calcular es $P(A \mid B)$. O sea, queremos la probabilidad de que uno sufra de la enfermedad cuando ya se conoce que el test es positivo.



- Lo que podemos hacer es relacionar las probabilidades condicional mediante una sustitución.
- $P(A \mid B) = \dfrac{P(B \mid A)P(A)}{P(B)}$, suponiendo que $P(A) > 0$ y $P(B) > 0$.
- Este es el **teorema de Bayes**, utilizado para calcular la probabilidad de un parámetro, dado un determinado evento.

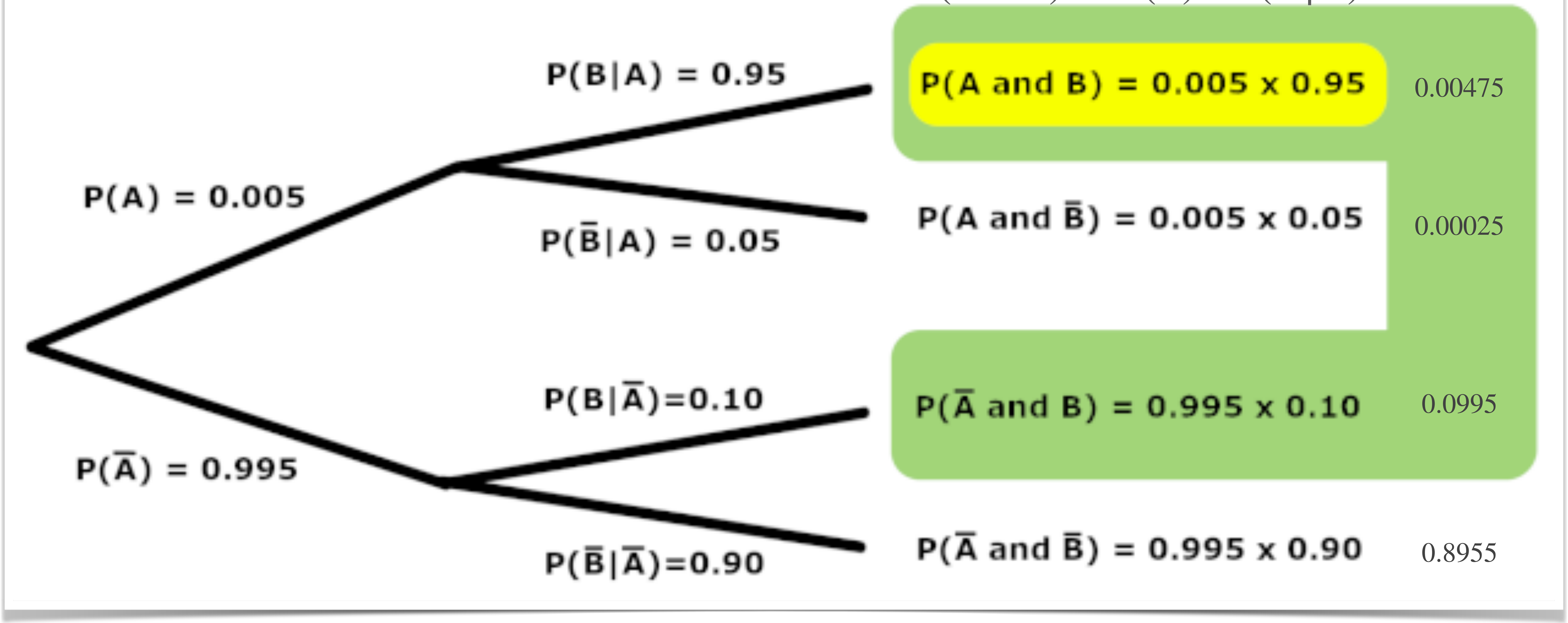
Teorema de Bayes

$$P(A|B)$$

Diagramas de árbol

https://towardsdatascience.com/bayes-theorem-clearly-explained-with-visualization-5083ea5e9b14

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



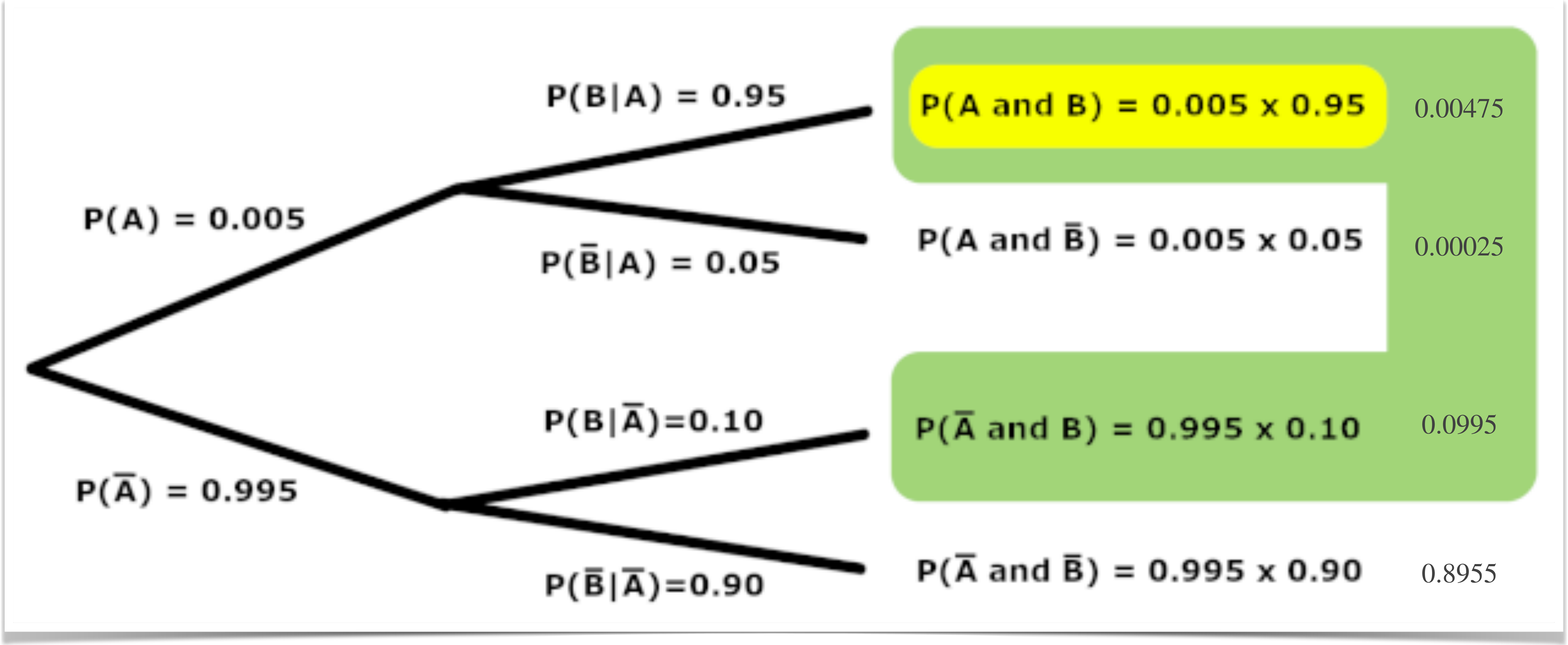
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Teorema de Bayes

Teorema de Bayes

Diagramas de árbol

https://towardsdatascience.com/bayes-theorem-clearly-explained-with-visualization-5083ea5e9b14

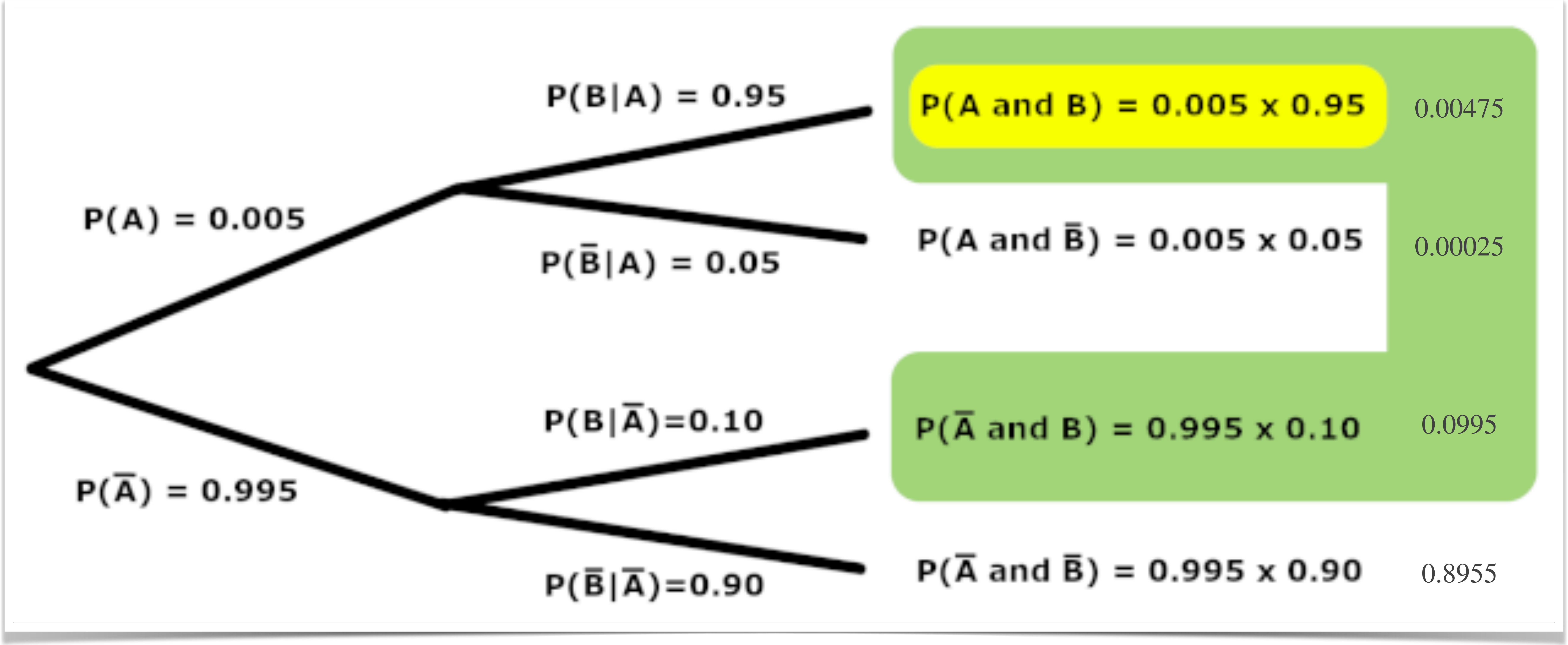


- Para encontrar la probabilidad $P(A\,|\,B)$, nos enfocamos en los casos donde B se sabe que es positivo (test positivo) (En verde).

Teorema de Bayes

Diagramas de árbol

https://towardsdatascience.com/bayes-theorem-clearly-explained-with-visualization-5083ea5e9b14



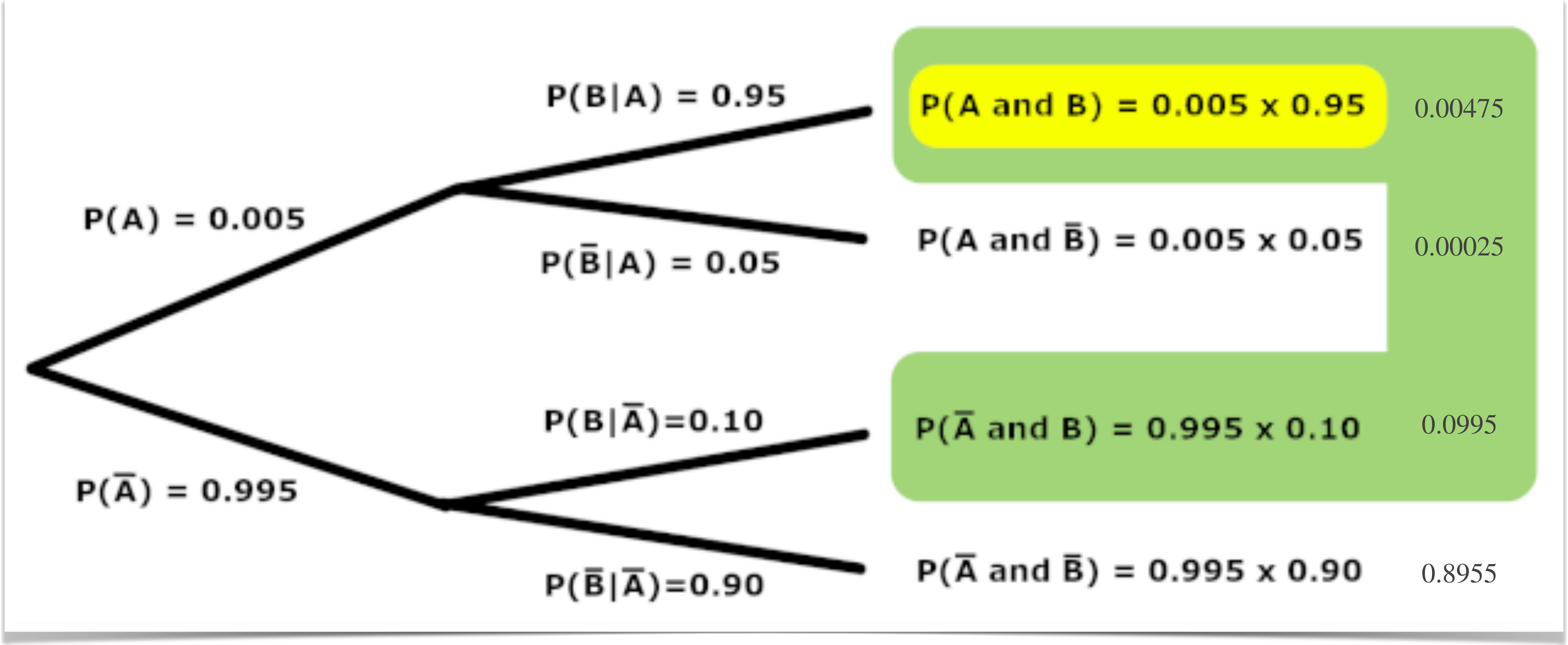
$$P(A\,|\,B) = \frac{P(A)P(B\,|\,A)}{P(B)} = \frac{P(A)P(B\,|\,A)}{P(A \cap B) + P(\bar{A} \cap B)}$$

$$P(B) = P(A \cap B) + P(\bar{A} \cap B) = = P(A) \cdot P(B\,|\,A) + P(\bar{A}) \cdot P(B\,|\,\bar{A})$$

Teorema de Bayes

Diagramas de árbol

https://towardsdatascience.com/bayes-theorem-clearly-explained-with-visualization-5083ea5e9b14



$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} = \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid \bar{A})P(\bar{A})}$$

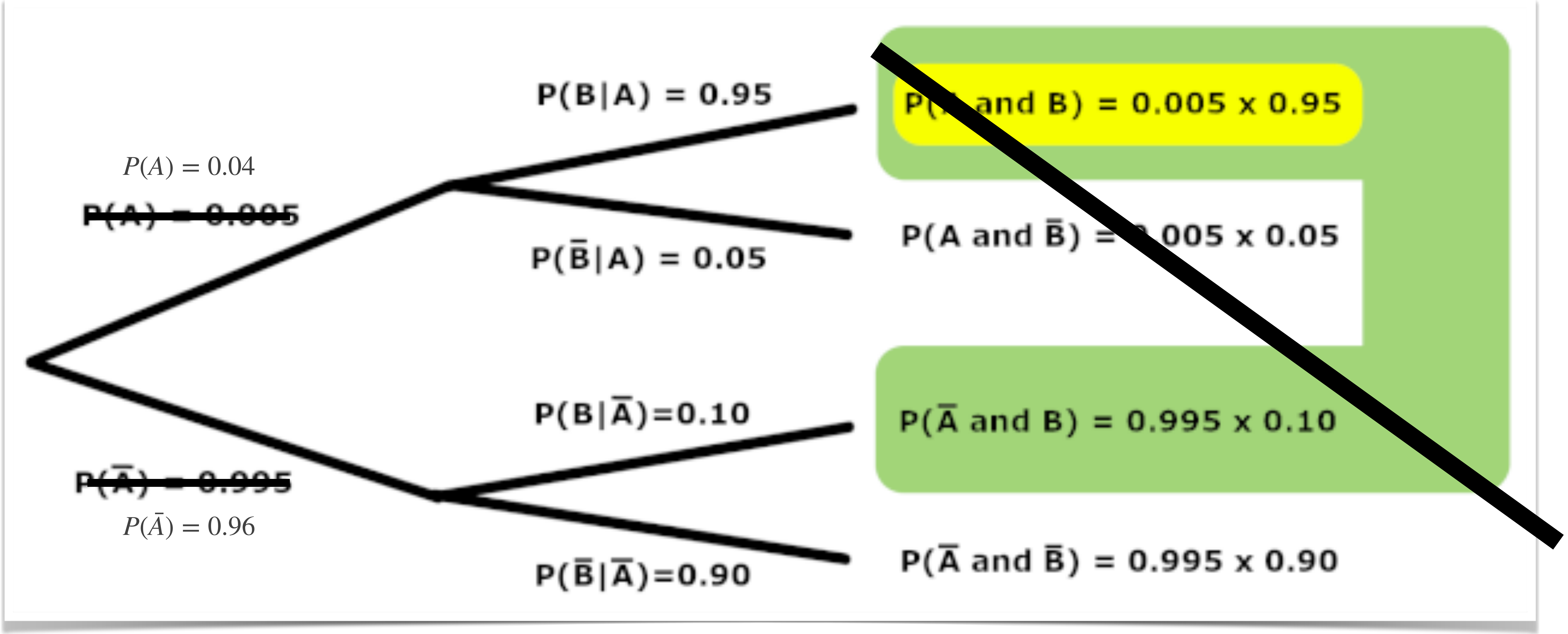
Teorema de Bayes

Diagramas de árbol

- Para nuestro caso particular:
- $$P(enfermedad\,|\,test) = \frac{(0.005)(0.95)}{(0.005)(0.95) + (0.995)(0.10)} \sim 0.04$$
- Es es contra intuitivo, demasiado bajo, pero viene del hecho de que la enfermedad es de por sí rara en la población. 99.5% del set de personas que sufren de la enfermedad es en comparación aún más pequeño que el 0.5% más grande de personas que no la tienen.

Teorema de Bayes

Diagramas de árbol (Segunda ronda). Actualización de información.

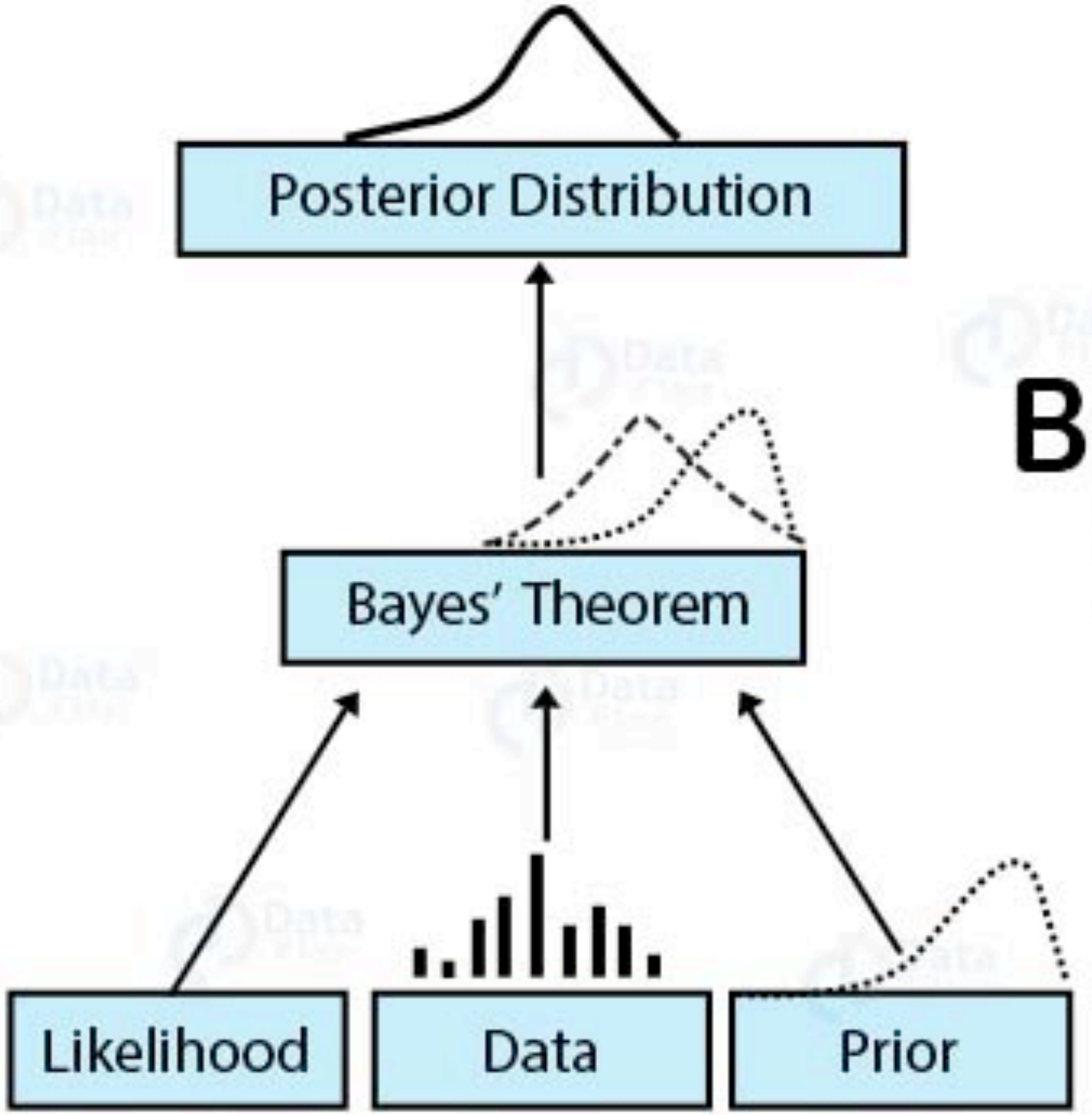


- Para nuestro caso particular:
- $$P(enfermedad\,|\,test) = \frac{(0.04)(0.95)}{(0.04)(0.95) + (0.96)(0.10)} \sim 0.28$$
- En la segunda instancia, valores de entrada para el evento A son actualizados.

Teorema de Bayes

General

- En general, el teorema de Bayes es definido por:
- $$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2)}$$
- Donde se asume que los eventos A1 y A2 son independientes y exhaustivos.
- Aunque claro, puede haber más eventos independientes
- $$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_n)P(B|A_n)}$$



Bayesian Inference for Data Science

Estadística y probabilidad

Ciencia de datos

- Machine learning is statistical learning.
- Métodos en ML realizan predicciones basadas en:
  - Datos de una muestra (datos de entrenamiento)
  - Estadísticos (parámetros de un modelo)
  - Modelos de probabilidad (Información previa, condicionales, distribuciones, etc.
- Un uso robusto de probabilidad y estadística es un prerequisito para ML.

Distribuciones de Probabilidad

Distribuciones de probabilidad

- Una distribución de probabilidad muestra los posibles resultados de un experimentos y la probabilidad de cada uno de estos resultados.

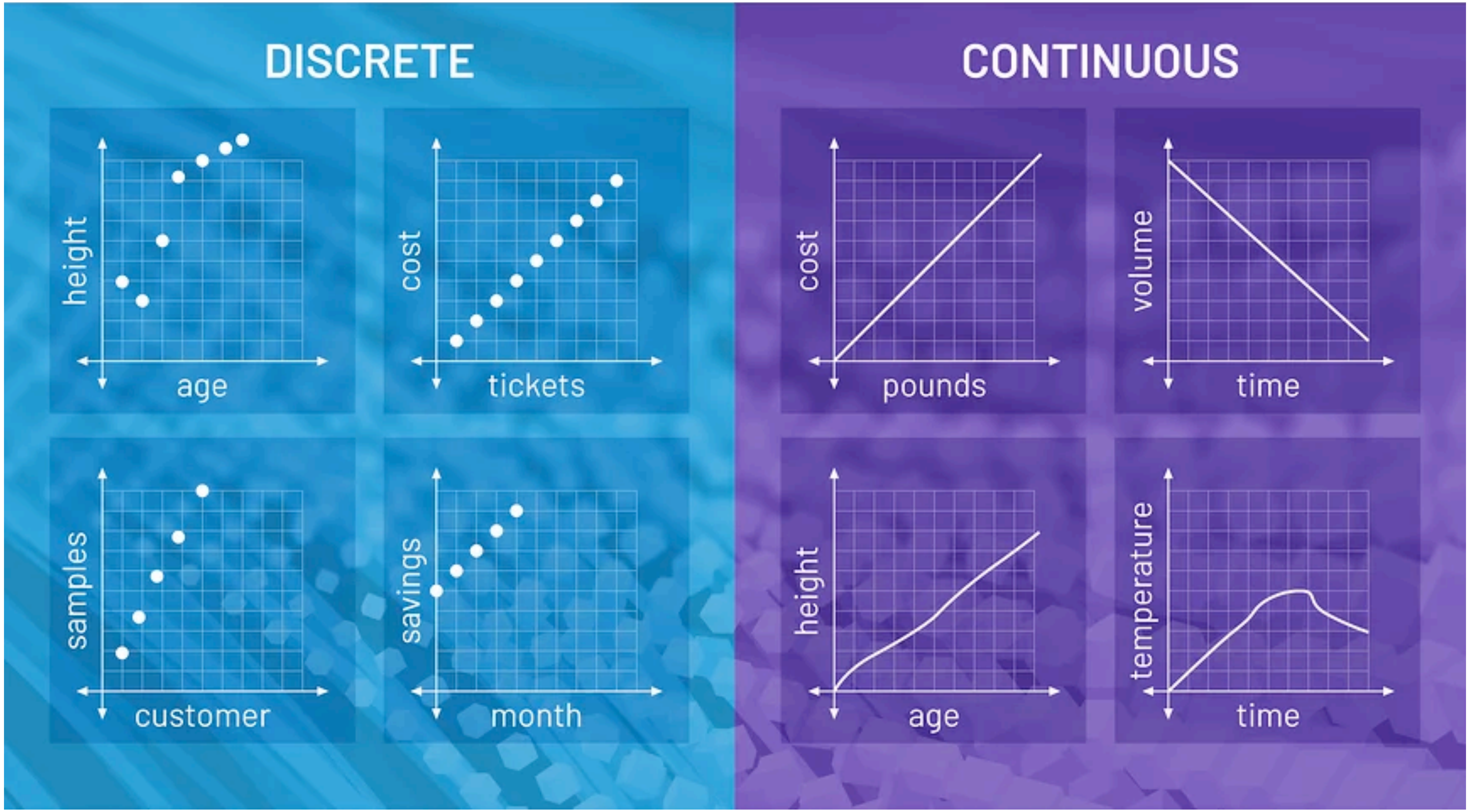
- Recordar que tenemos dos tipos de variables aleatorias:

  - **Discretas:** toman valores aislados (e.g., el número de caras al lanzar tres monedas).

  - **Continuas:** pueden tomar cualquier valor dentro de un intervalo (e.g., la altura de una persona)

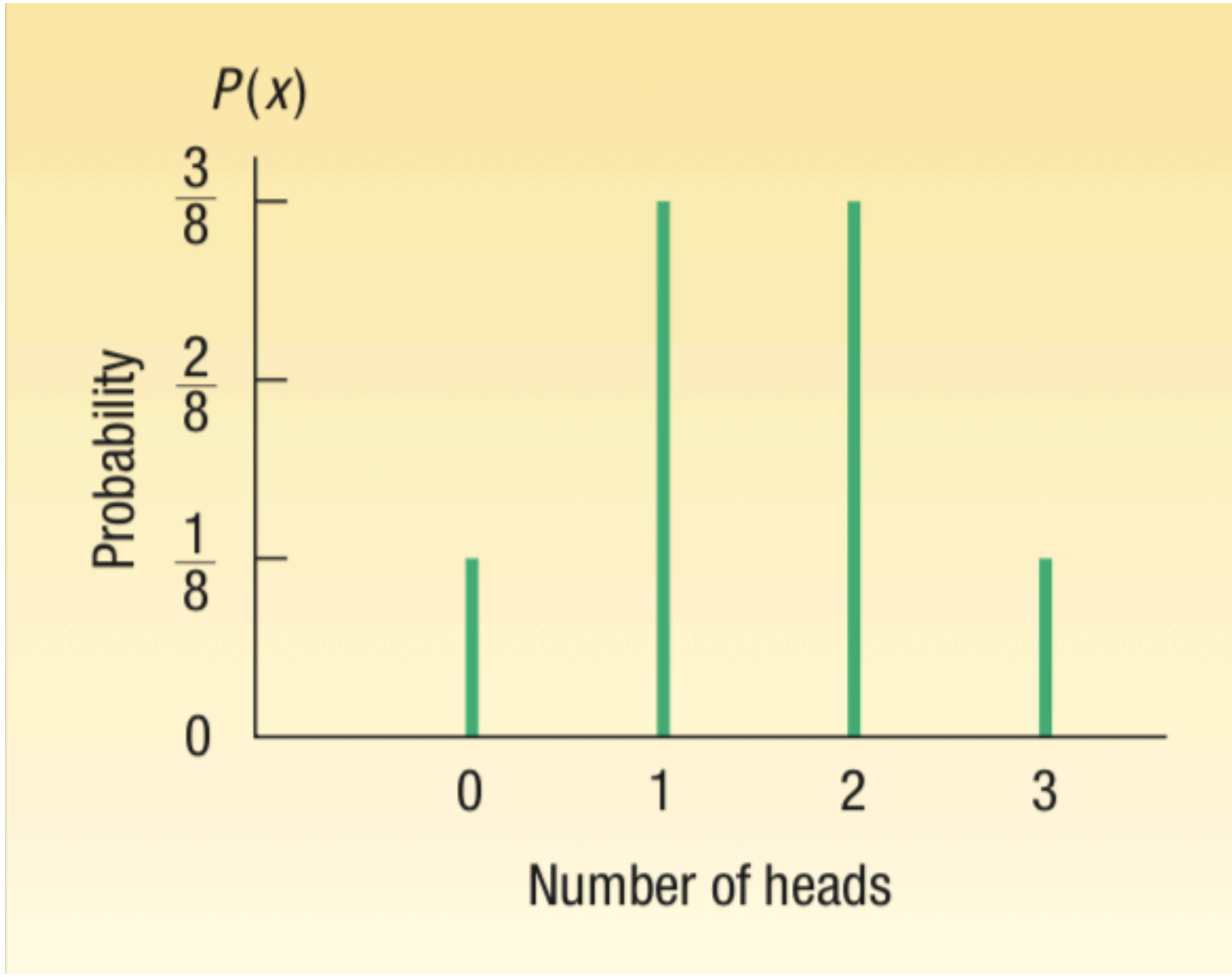
Distribuciones de probabilidad

Ejemplos de variables



Distribuciones discretas

Ejemplo



$$\sum_{i=0}^{n} P(x_i) = 1$$

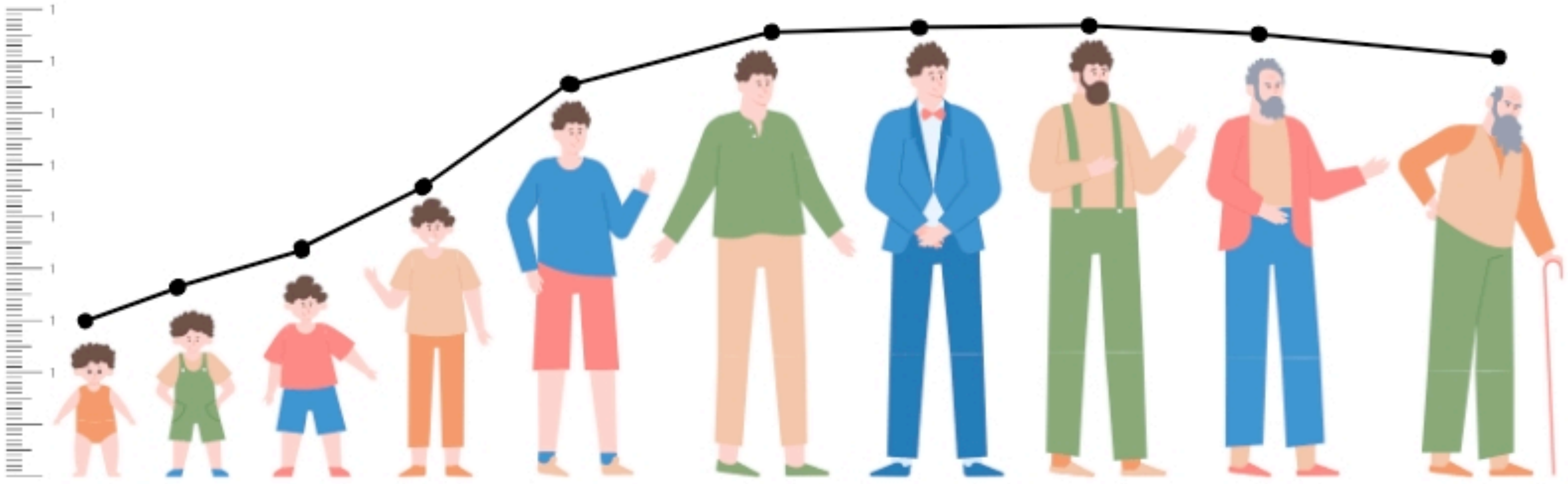
Number of Heads, x	Probability of Outcome, $P(x)$
0	$\frac{1}{8} = .125$
1	$\frac{3}{8} = .375$
2	$\frac{3}{8} = .375$
3	$\frac{1}{8} = .125$
Total	$\frac{8}{8} = 1.000$

Possible Result	Coin Toss First	Second	Third	Number of Heads
1	T	T	T	0
2	T	T	H	1
3	T	H	T	1
4	T	H	H	2
5	H	T	T	1
6	H	T	H	2
7	H	H	T	2
8	H	H	H	3

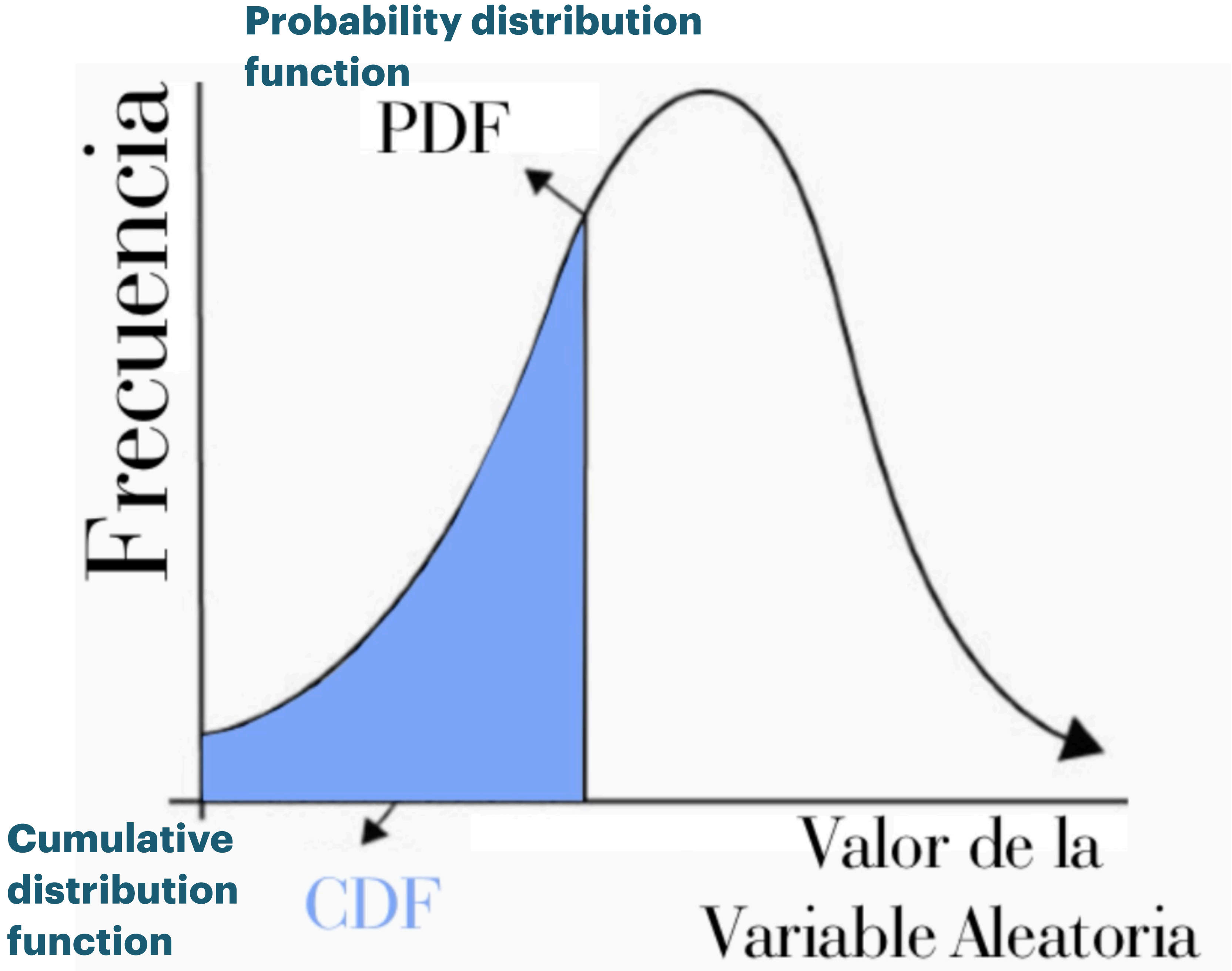
Distribuciones continuas

Ejemplo

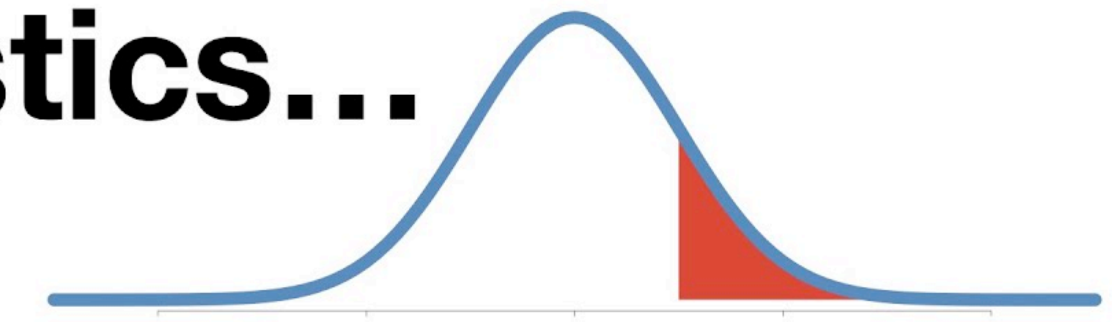
$$\sum \rightarrow \int$$



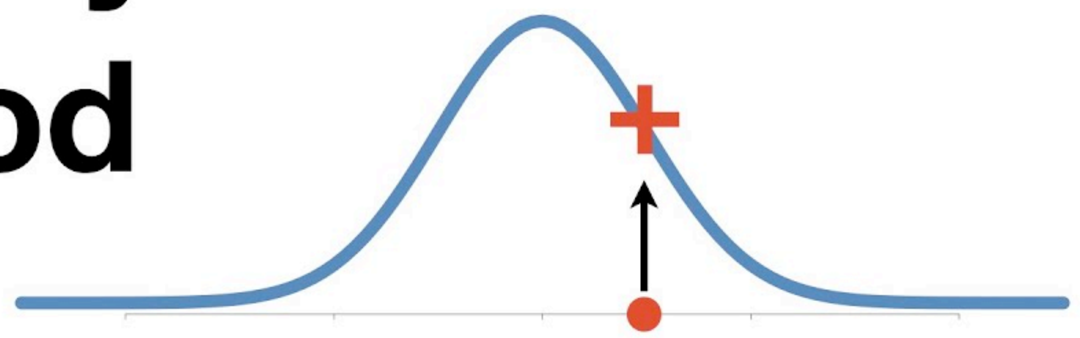
$$\int P(x)dx = 1$$



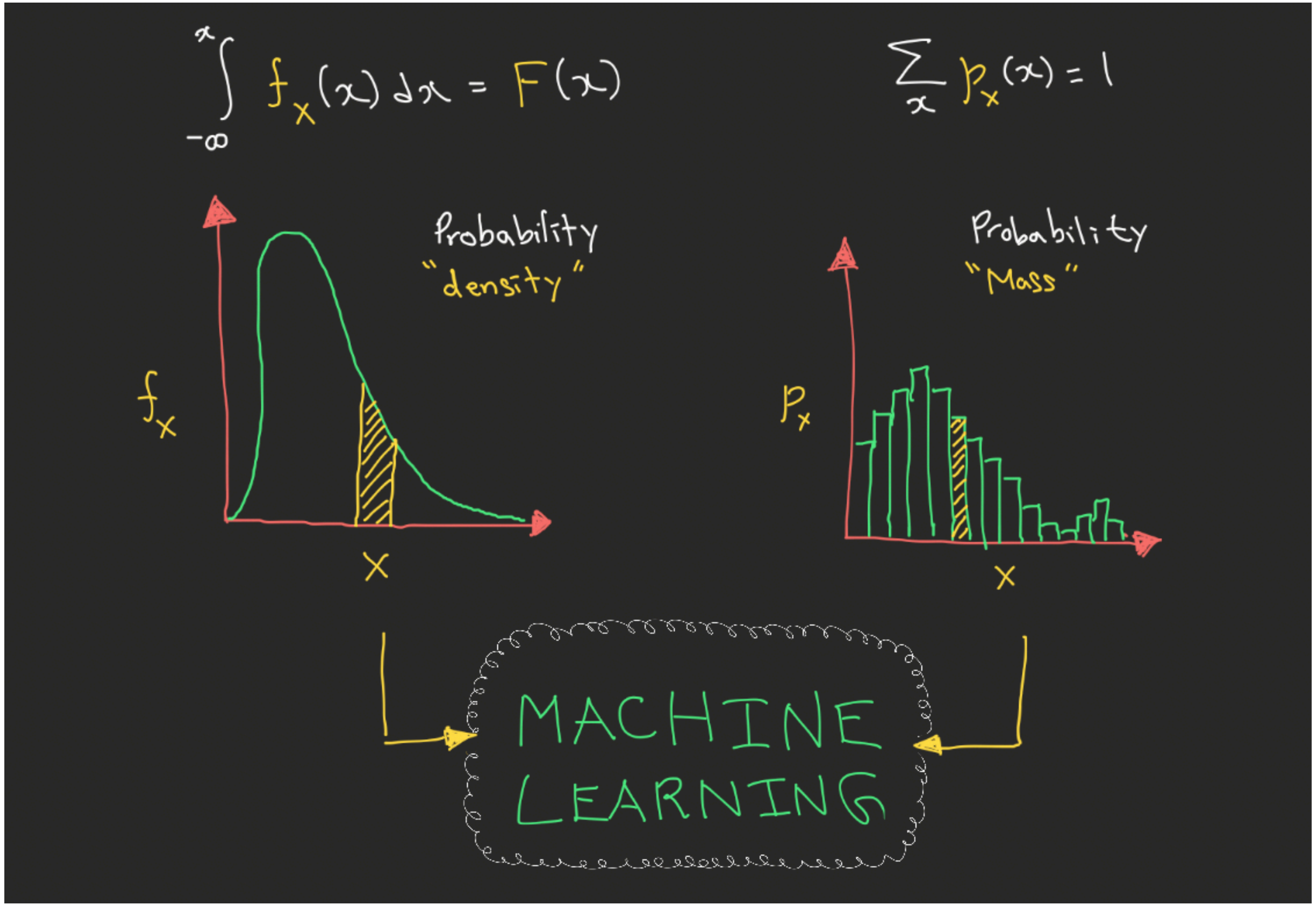
In Statistics…



Probability is not Likelihood



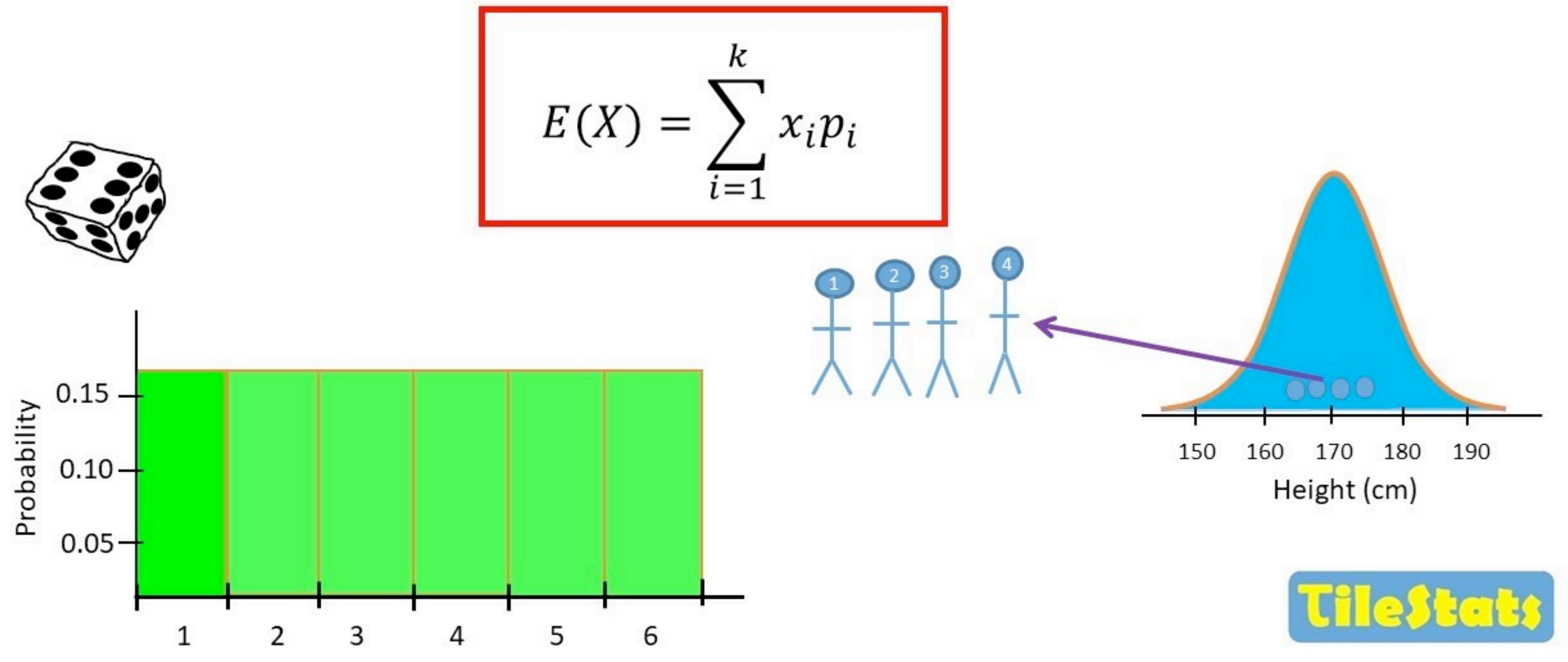
Funciones de masa y densidad de probabilidad



Estadísticos

Esperanza matemática

- La media de una distribución de probabilidad es llamada valor esperado. Y se expresa como
- $$E(x) = \mu = \sum_{i=0} [x_i P(x_i)]; \quad \mu = \int xP(x)dx$$
- donde P(x) es la probabilidad de un valor particular x. Sería un promedio ponderado con todos los posibles valores de la variable aleatoria y pesados por sus correspondientes probabilidades de ocurrencia.



Estadísticos

Varianza

- Algo similar pasa con la varianza, donde se tiene una media de las desviaciones al cuadrado $(x-\mu)^2$ pesadas con la distribución de probabilidad, es decir,
- $Var(x) = \sigma^2 = \sum \left[(x-\mu)^2 P(x)\right]; \quad \sigma^2 = \int (x-\mu)^2 P(x) dx$
- Utilizando algunas propiedades matemáticas de los valores esperados, podemos incluso reescribirla de otra forma,
- $Var(x) = E(X^2) - \mu^2$
- Y la desviación estándar sería $SD(X) = \sigma = \sqrt{\sum_{x \in S} [x^2 P(x)] - \mu^2}$

Distribuciones discretas

Ensayos de Bernoulli

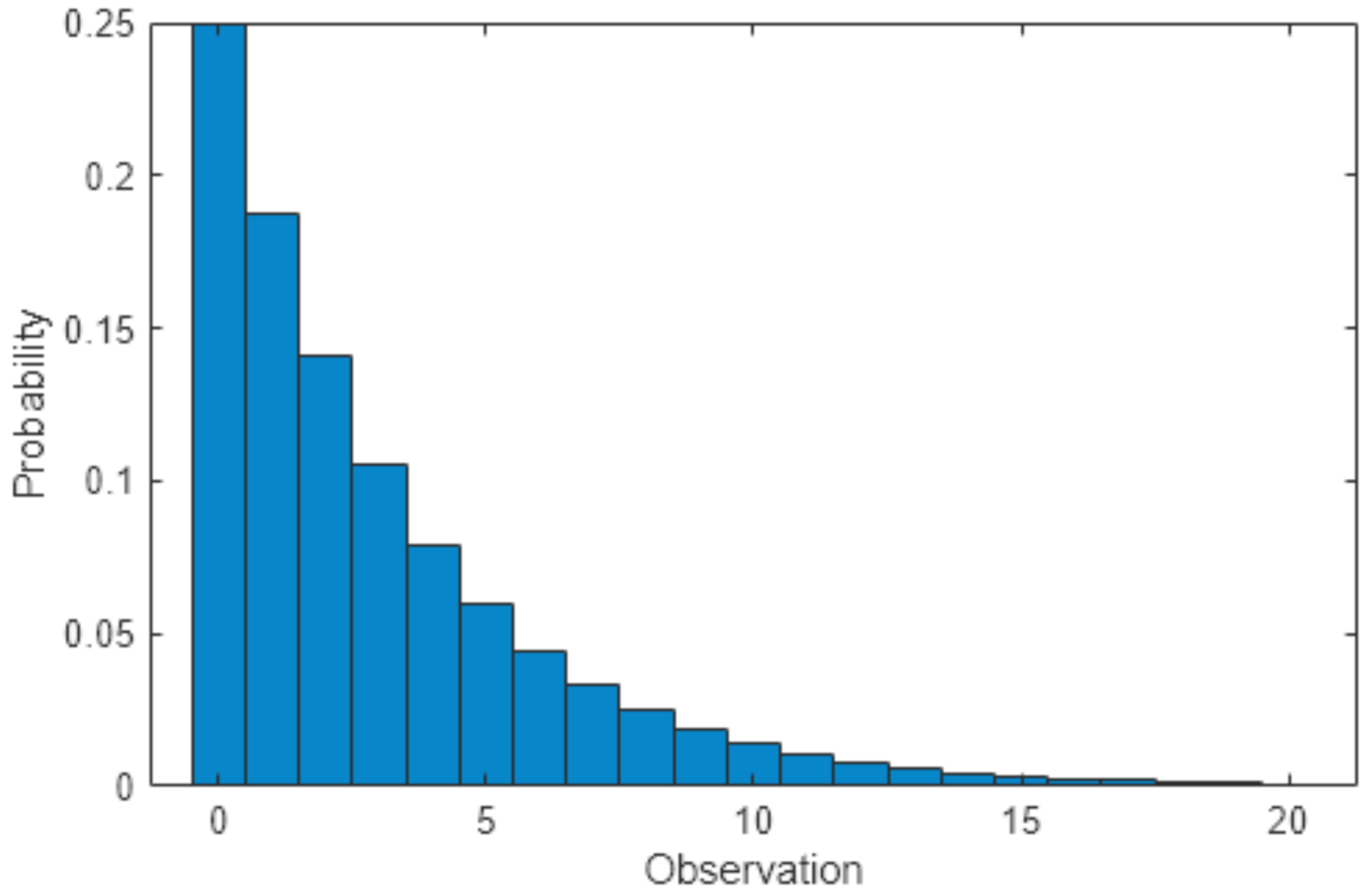
- Considera un ejemplo de “tirar una moneda”. Vas a tener “Cara” o “Sello”. Si obtener “Cara” es un éxito, entonces tener “Sello” es un fallo.
- Estos ensayos son llamados Ensayos de Bernoulli , si se tiene un escenario donde se cumple lo siguiente
- Sólo dos posibles resultados se dan en cada ensayo: éxito o fracaso. (Aplicados en contexto con el problema)
- La probabilidad de éxito es denotada por p, que es constante en cada ensayo.
- Cada ensayo es independiente. Obtener un sello en un ensayo no afecta el siguiente.

$$P(X = x) = \begin{cases} p \text{ for } x = 1 \\ 1-p \text{ for } x = 0 \end{cases}$$

Distribuciones discretas

Distribución geométrica

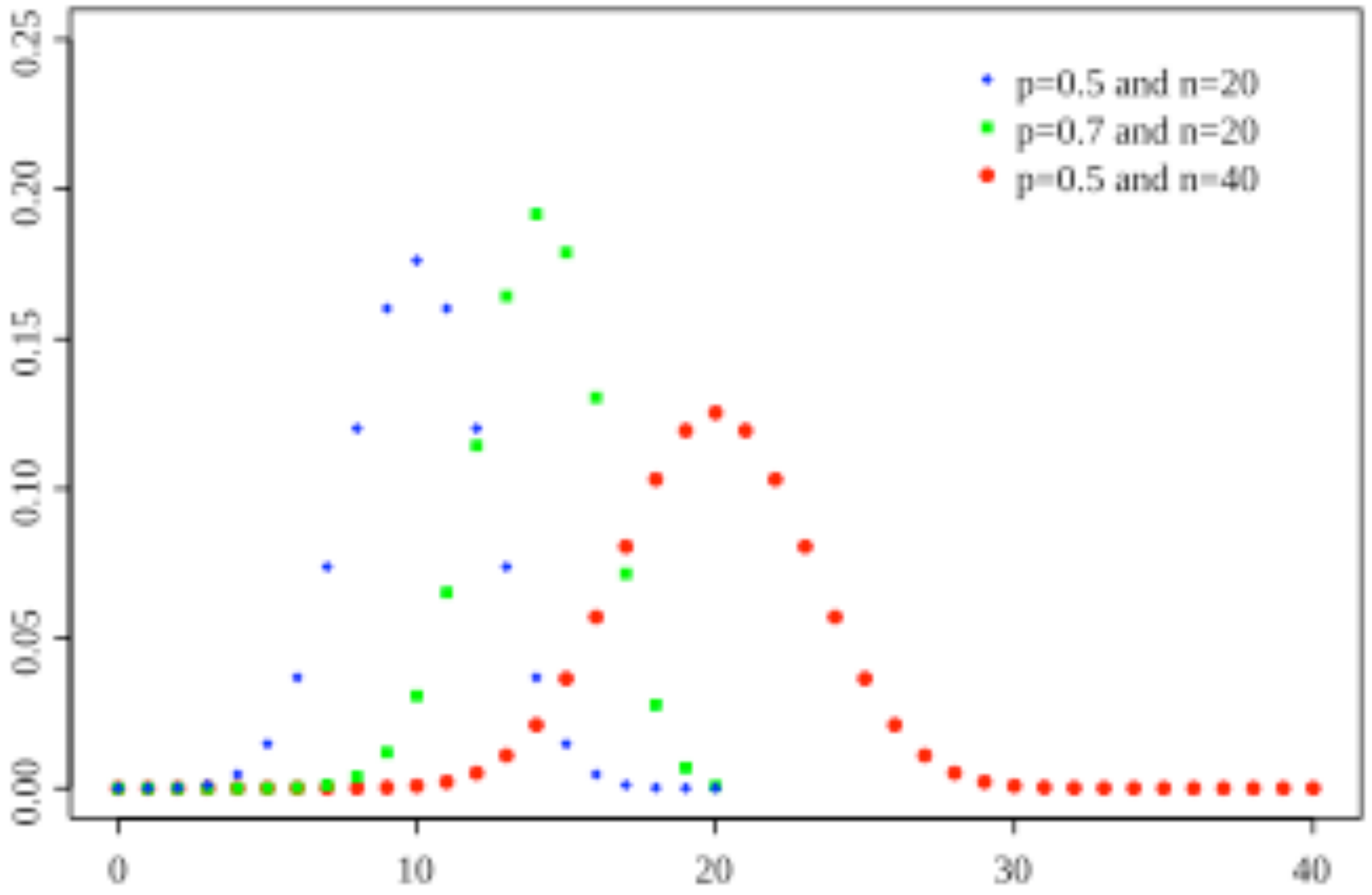
- Así que, cuál es la probabilidad de obtener un éxito en el tercer intento.?
- Intento 1- fallo con prob q=0.95, Intento 2-fallo con prob q=0.95, Intento 3- éxito con probabilidad p =0.05.
- $P(X=3) = 0.95 \cdot 0.95 \cdot 0.05 = 0.95^2 \cdot 0.05 = 0.045$
- Este modelo es llamado Modelo Geométrico que modela que tanto toma obtener el primer éxito en una serie de ensayos de Bernoulli. Denotado por
- $P(X=x) = q^{x-1}p; \quad q = 1-p$
- Esto tendría un valor esperado $E(X) = \mu = \dfrac{1}{p}$, Desviación Estándar $\sigma = \sqrt{\dfrac{q}{p^2}}$
- $\mu = E(X) = \sum xP(x) = \sum x(q^{x-1}p)$
- $\sigma^2 = Var(X) = \sum (x-\mu)^2 P(x) = \sum (x-\mu)^2 (q^{x-1}p)$



Distribuciones discretas

Distribución binomial

- Cuando la pregunta es ¿Cual es la probabilidad de escoger exactamente 2 puertas con un coche detrás de ellas?
- Utilizamos la distribución binomial, con probabilidad
- $P(X=x) = {}_nC_x p^x q^{n-x}$
- Esta distribución tiene una media de $\mu = np$ y una desviación estándar dada por $\sigma = \sqrt{npq}$.
- La distribución binomial nos dice la probabilidad de tener x éxitos en n ensayos.

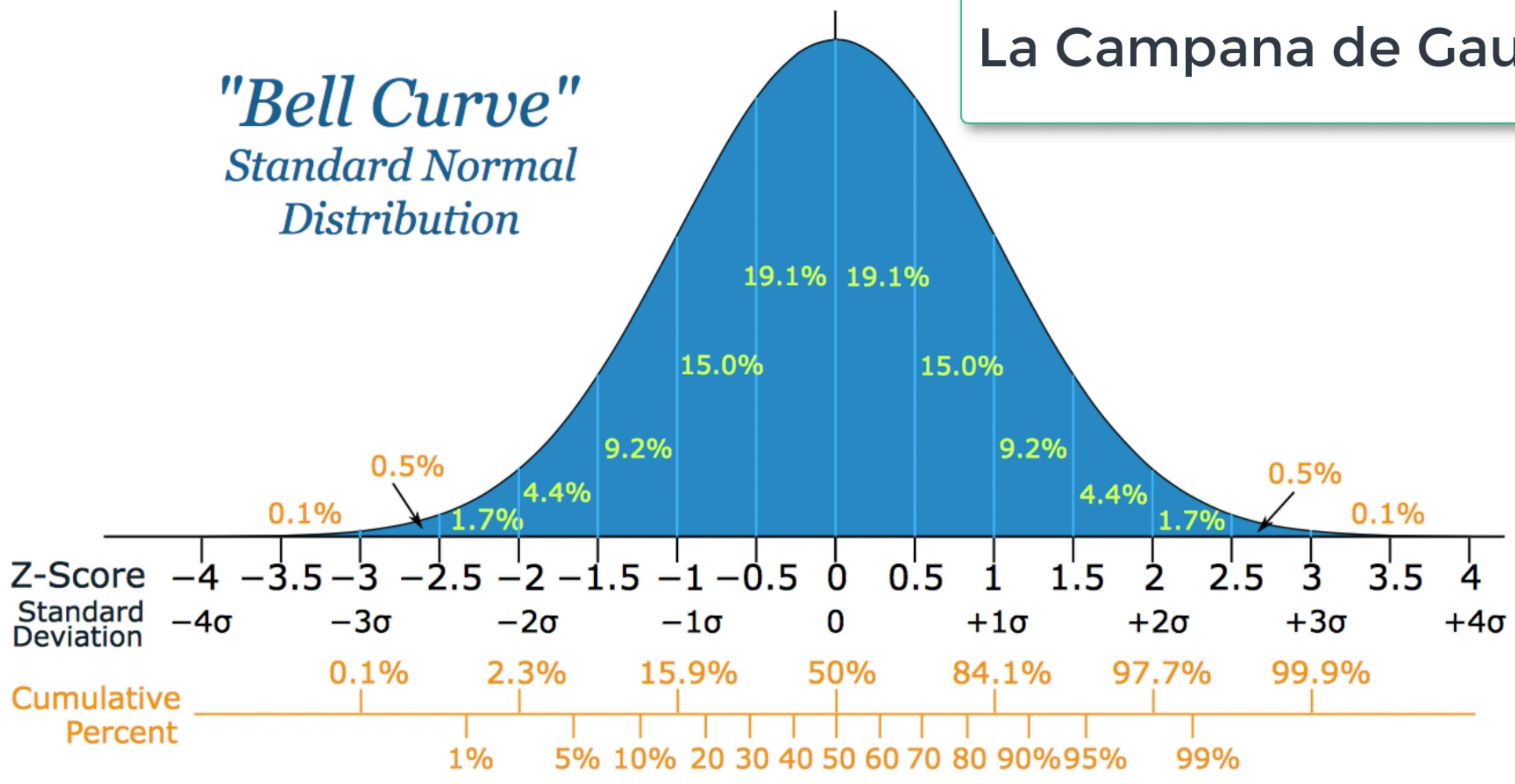


Distribuciones continuas

Distribución normal

- Muchas cosas de la vida normal siguen una distribución normal:
- Ejemplos: peso, edad, calificaciones, medidas de error.
- Este tipo de distribución tiene la peculiaridad de que los valores de los datos tienden a concentrarse alrededor de un valor central sin sesgo hacia la derecha o izquierda.

La Campana de Gauss



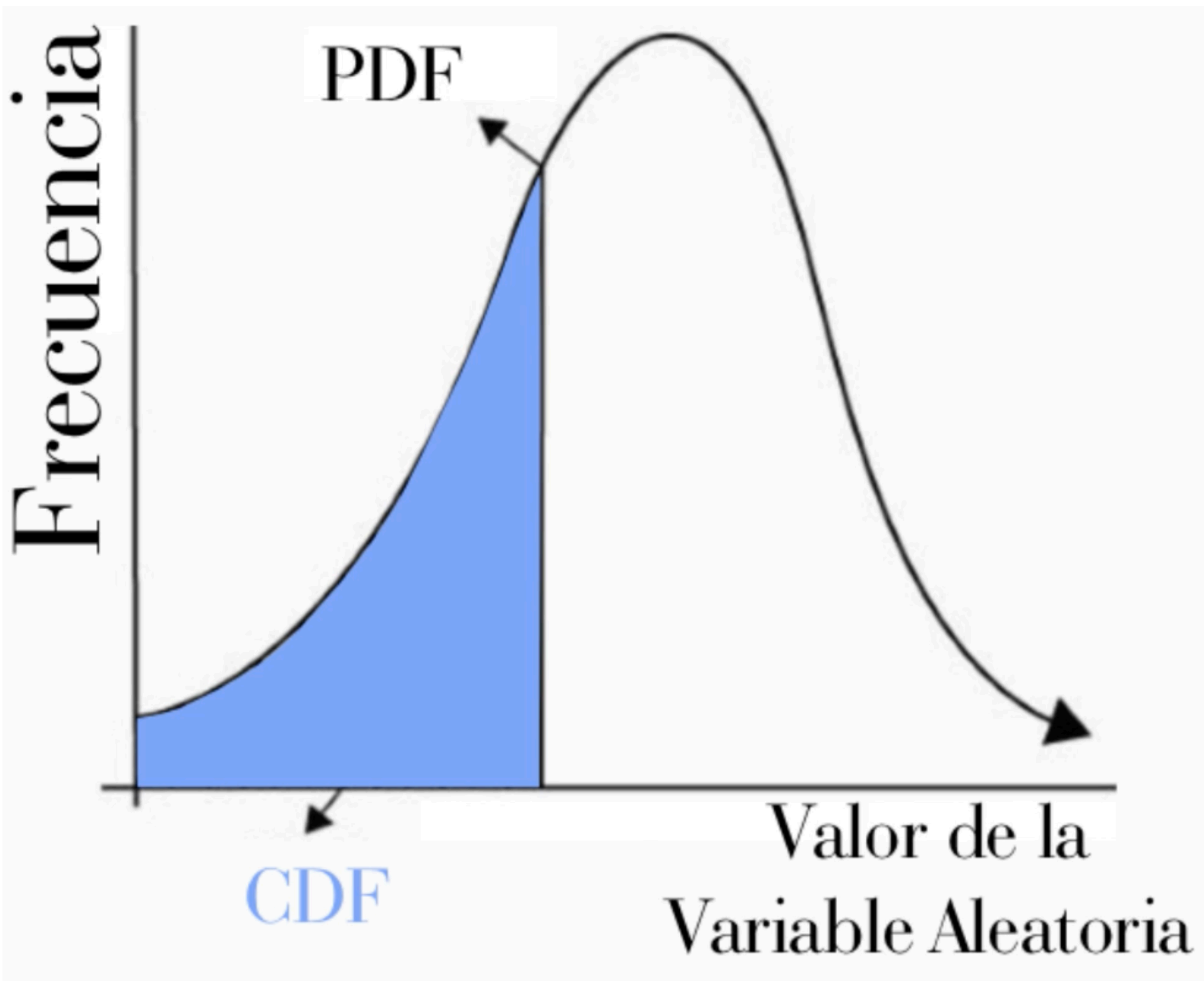
Distribución normal

Media y desviación estándar

- Por ser una distribución continua, los valores esperados se calculan mediante una integral sobre un rango (de $-\infty$ a ∞) en lugar de una sumatoria.
- $E(X) = \int_{-\infty}^{\infty} xP(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} xe^{\frac{-(x-\mu)^2}{2\sigma^2}}dx = \mu$
- $Var(X) = E(X^2) - \mu^2 = \int_{-\infty}^{\infty} [x^2P(x)]dx - \mu^2 = \left(\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2e^{\frac{-(x-\mu)^2}{2\sigma^2}}dx\right) - \mu^2 = \sigma^2$

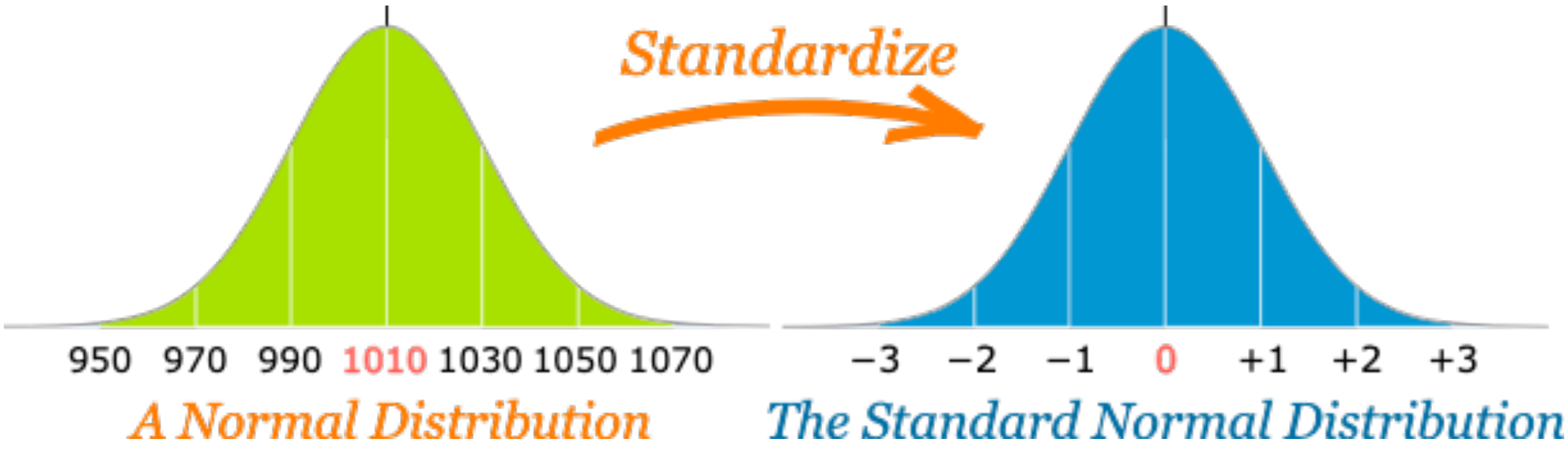
Distribución normal
Media y desviación estándar

- Para calcular la probabilidad dentro de un cierto intervalo (como ejemplo la función cumulativa) se tiene que evaluar la integral en dicho intervalo.
- $$P(X : x < c) = \int_{-\infty}^{c} P(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{c} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx$$
- Esta integral se aproxima de manera numérica.



Distribución normal

Estandarización



$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

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

Distribución normal

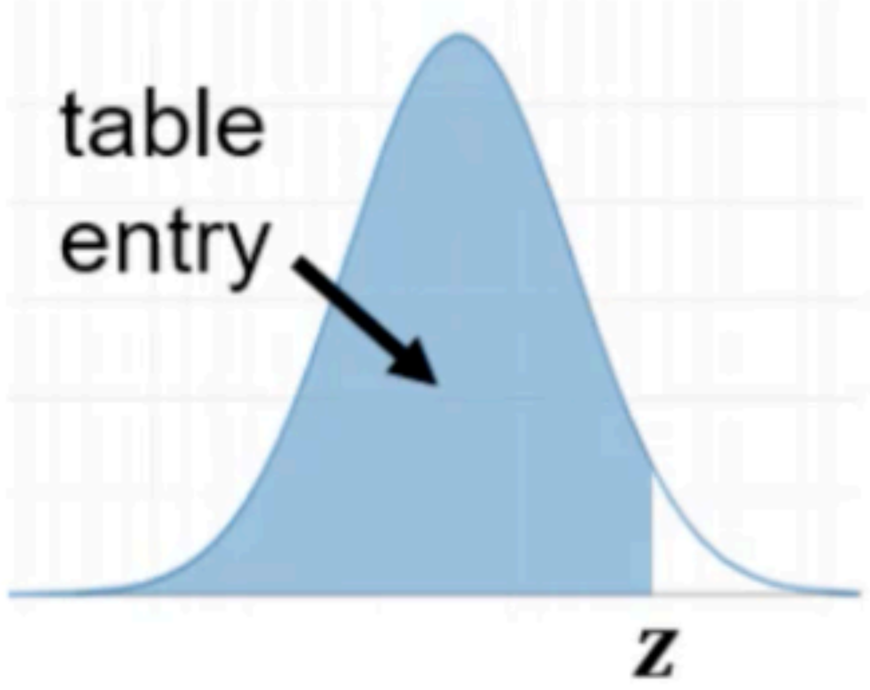
Estandarización

- Una distribución normal arbitraria puede ser convertida en una distribución normal estándar (media 0, desviación estándar 1) mediante un cambio de variable $z = \frac{x - \mu}{\sigma}$, de tal forma que la derivada implícita es $dz = \frac{dx}{\sigma}$ y la distribución de probabilidad puede reescribirse como:
- $P(x)dx = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} dz$
- Como se puede ver, esto ya tiene la forma de una distribución normal estándar, con la diferencia que ahora la variable de la función es z.
- De manera inversa $\sigma z = x - \mu \rightarrow x = \mu + \sigma z$

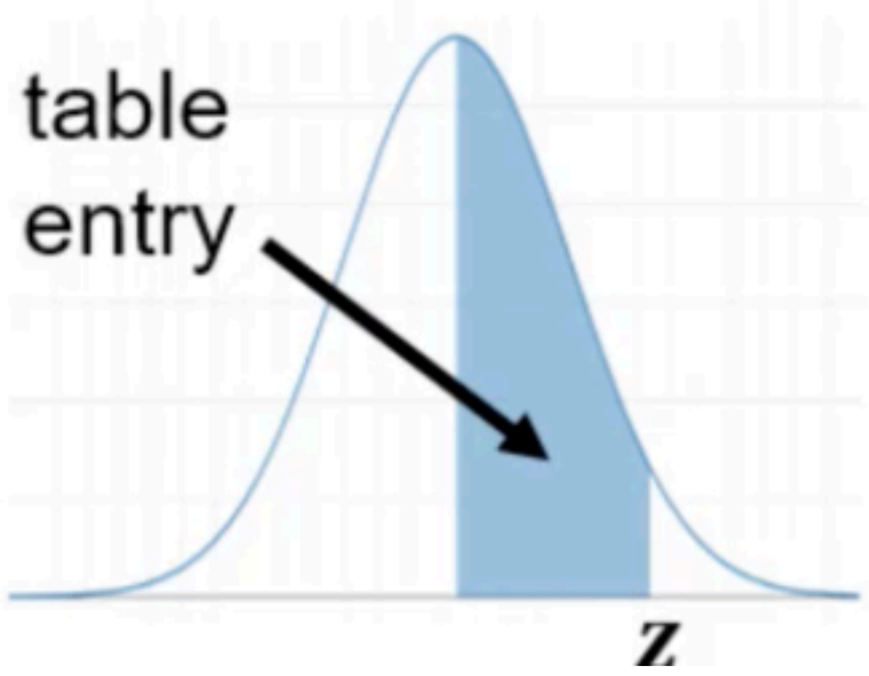
Distribución normal

Estandarización

- Con esto uno puede expresar la función cumulativa, o calcular la probabilidad dentro de un intervalo mediante valores de la variable z (en unidades de desviación estándar) y no de x.
- $$P(X : x < z) = \int_{-\infty}^{c} P(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{c} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{\frac{-z^2}{2}} dz$$
- Mediante esto también evitamos la dependencia en las unidades de x (peso, edad, etc.), pues vemos que z es adimensional, y es válido para cualquier distribución normal.
- Con los valores z, uno puede calcular un percentil de la distribución, es decir, que tanto porcentaje (área bajo la curva = probabilidad) cae debajo de este valor.



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141

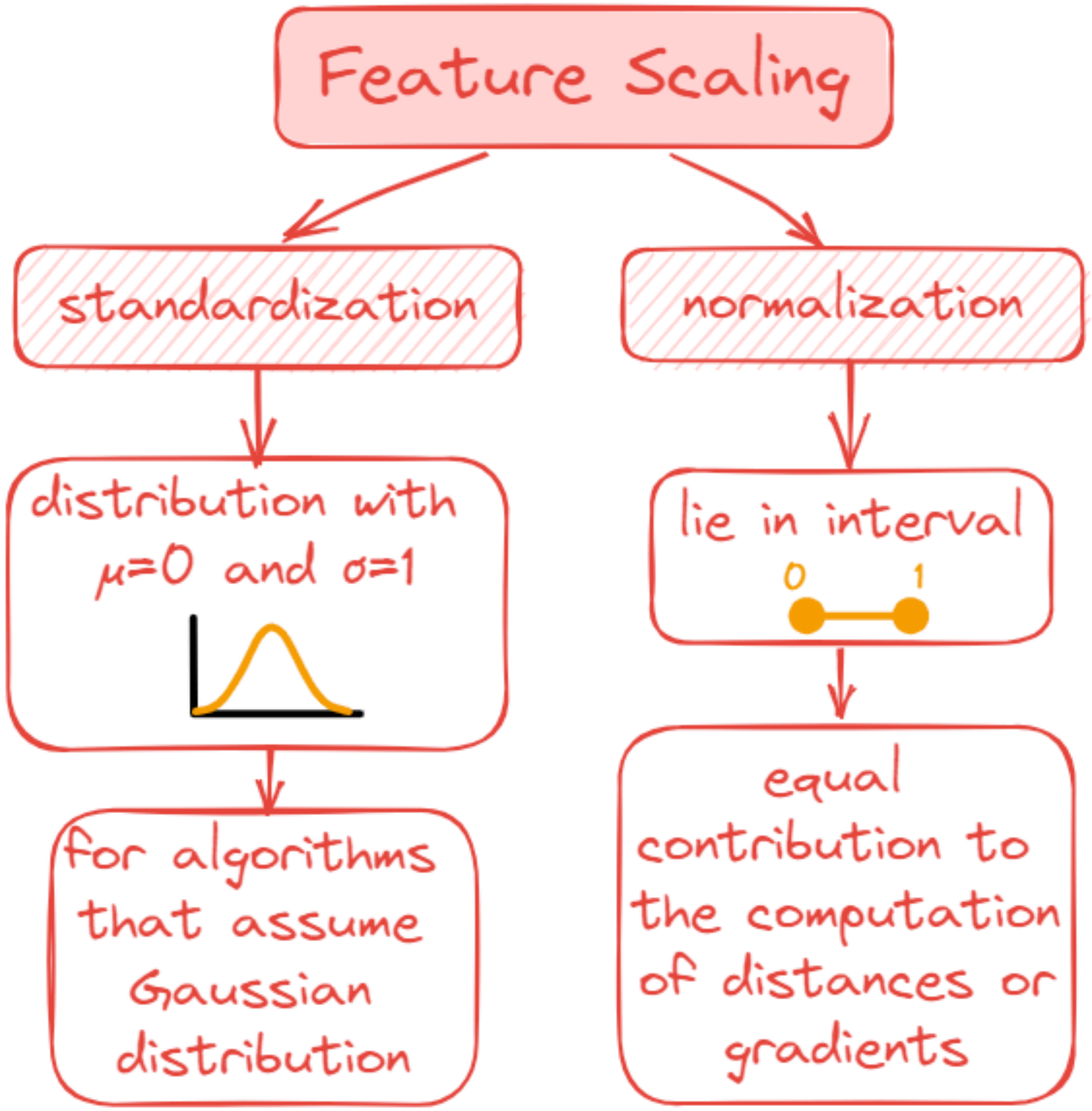
Distribución normal

Estandarización en Machine Learning

- En Machine Learning, muchos algoritmos son sensibles a la escala de los datos. Por eso, es común aplicar un proceso de **estandarización**, que transforma cada variable para que tenga media cero y desviación estándar uno.

- Esto no cambia la forma de la distribución, pero sí asegura que todas las características contribuyan de manera comparable al modelo.

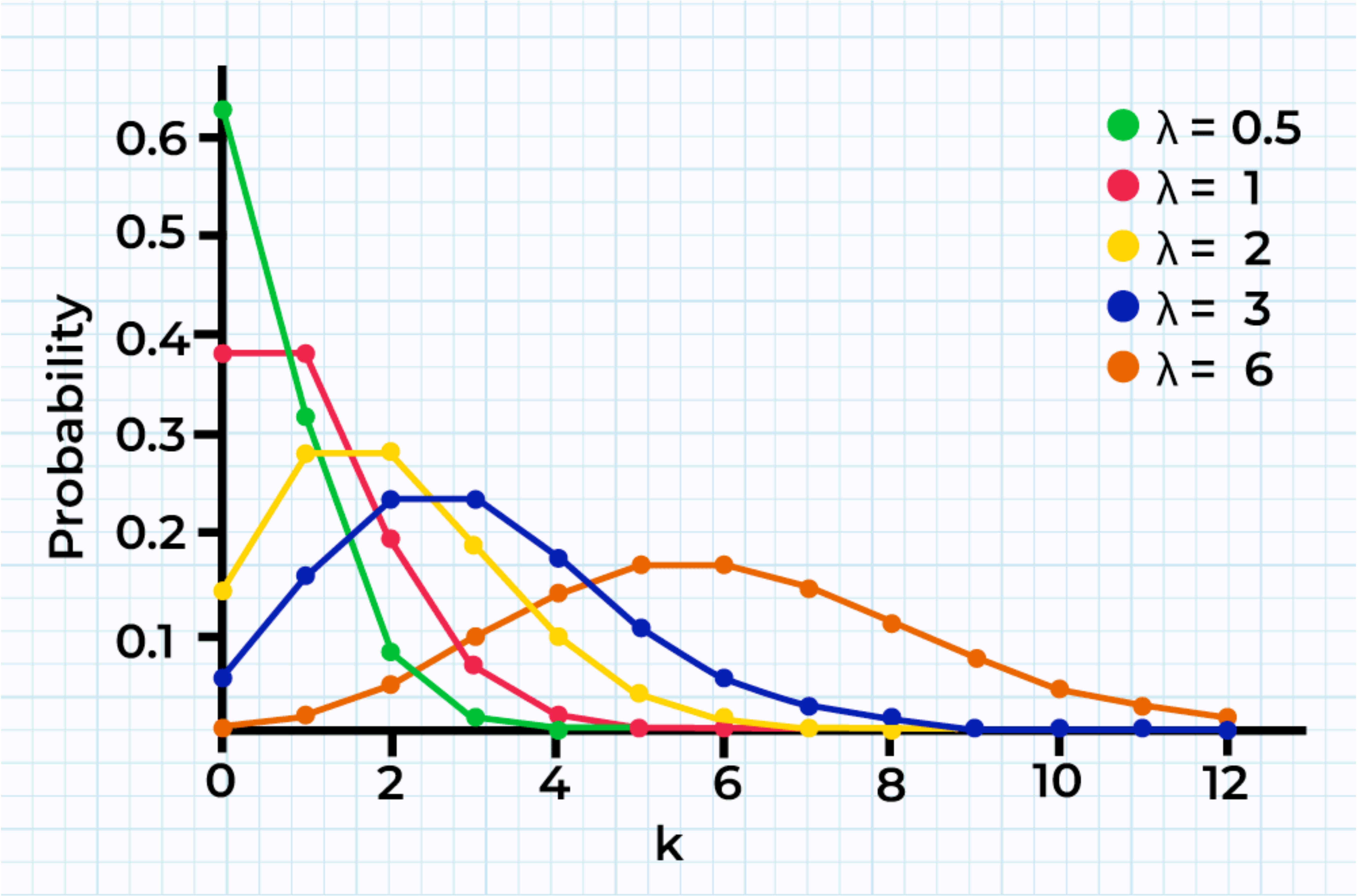
- Modelos como **regresión logística**, **SVM** o **redes neuronales** suelen beneficiarse mucho de esta práctica.



Distribuciones discretas

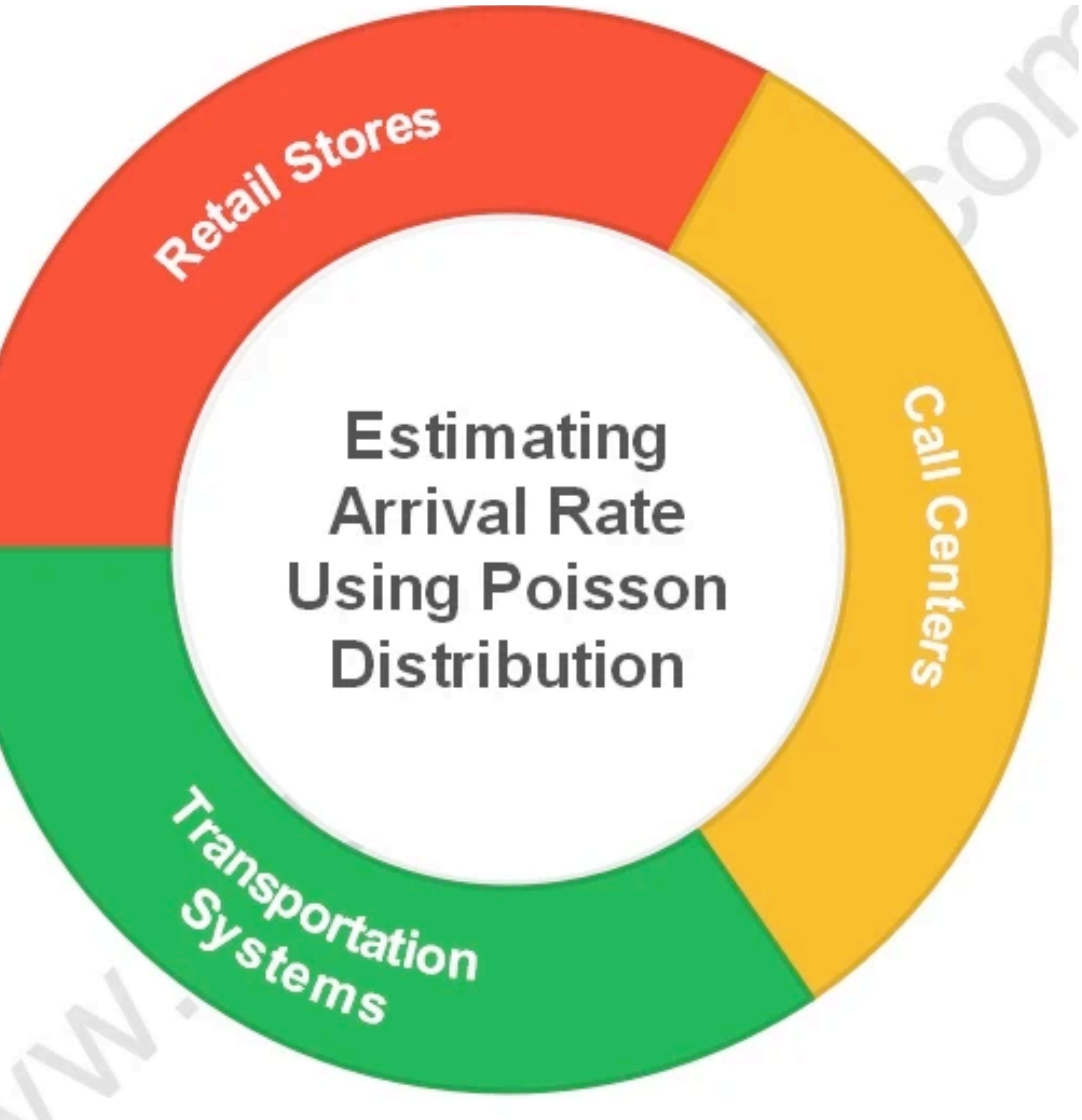
Distribución de Poisson

- Se utiliza una distribución de Poisson cuando uno espera encontrar una media de algún número de objetos aleatorios y uniformemente distribuidos o las ocurrencias (independientes entre sí) de algo en un cierto área, volumen, tiempo, etc. y uno quiere encontrar la probabilidad de ver exactamente X objetos/ocurrencias (que pueden ser diferentes del número esperado).



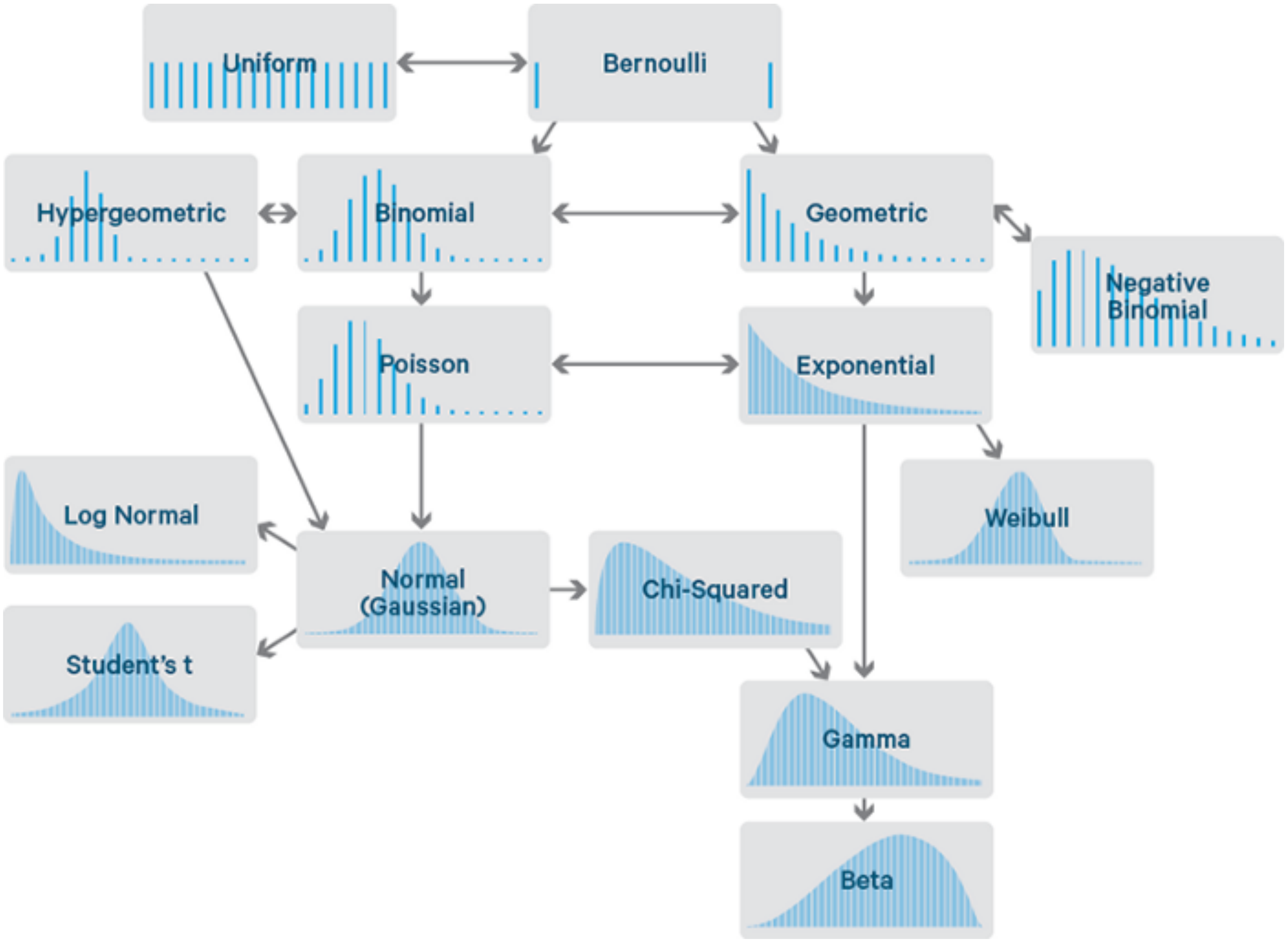
Distribuciones discretas

Distribución de Poisson

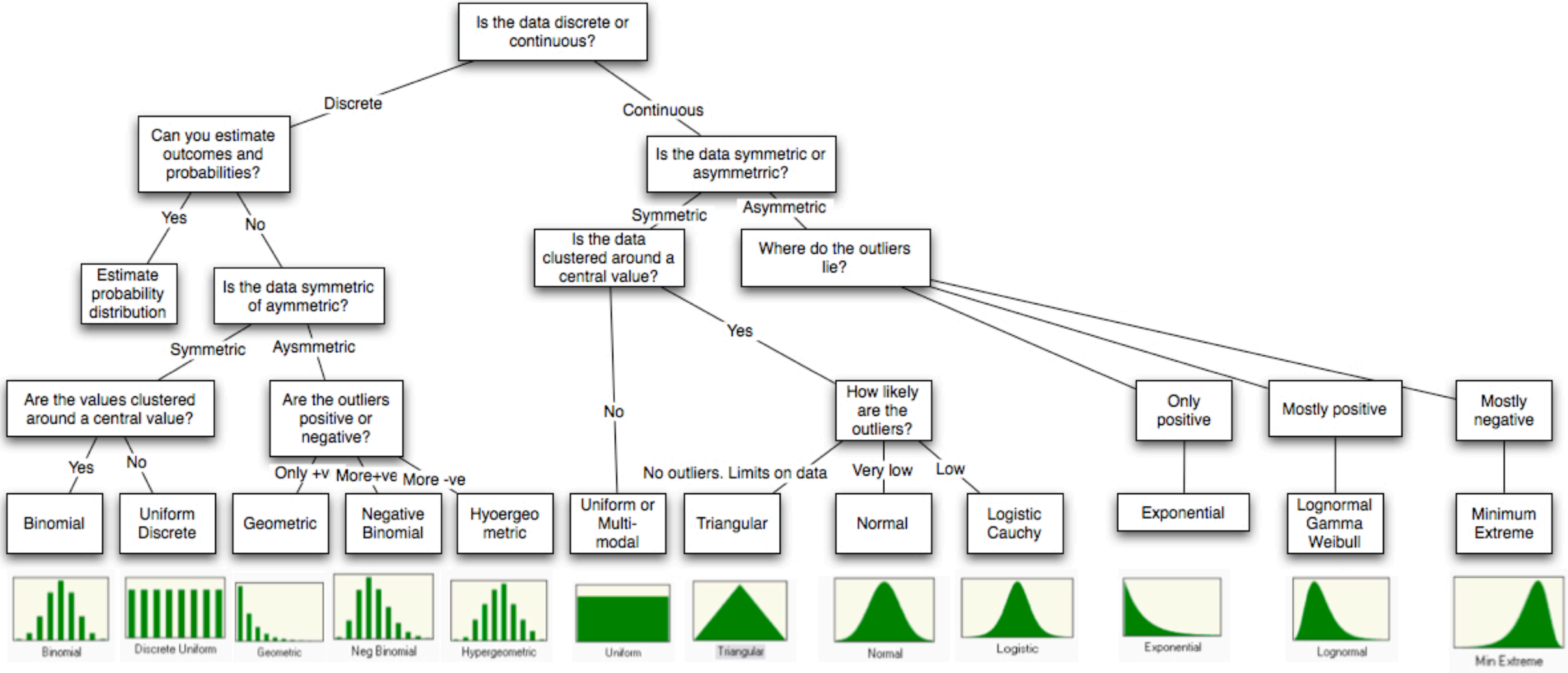


- $P(X) = \dfrac{e^{-\lambda}\lambda^{X}}{X!}$
- Donde λ es el número esperado (o sea la media) de objetos/ocurrencias, y X son los objetos/ocurrencias que queremos.
- Tiene una media de $\mu = np = \lambda = Var(X)$

Otras distribuciones de probabilidad



Otras distribuciones de probabilidad



Lecturas sugeridas

- Blitzstein, J. K., & Hwang, J. (2019). Introduction to probability (2nd ed.). CRC Press.

- Bruce, P., Bruce, A., & Gedeck, P. (2020). Practical statistics for data scientists: 50+ essential concepts using R and Python (2nd ed.). O'Reilly Media.

- Videos sugeridos relacionados con el tema:

1. https://www.youtube.com/watch?v=oI3hZJqXJuc (Introducción a distribuciones de probabilidad)
2. https://www.youtube.com/watch?v=pYxNSUDSFH4 (Probabildad y verosimilitud)
3. https://www.youtube.com/watch?v=rzFX5NWojp0 (Distribución Normal)
4. https://www.youtube.com/watch?v=7S7j75d3GM4&t=73s (Distribucion muestral de la media)