# Matemática para Ciencia de Datos

## Probabilidades

Dr. Daniel Alexis Gutierrez Pachas (dgutierrezp@ucsp.edu.pe)

25 de noviembre de 2023

Departamento de Ciencia de la Computación, Universidad Católica San Pablo, Arequipa, Perú.

# Experimento Aleatorio

Es un experimento bajo el mismo conjunto aparente de condiciones iniciales, que pueda presentar resultados diferentes. Entonces, no se puede predecir o reproducir el resultado exacto de cada experiencia particular.

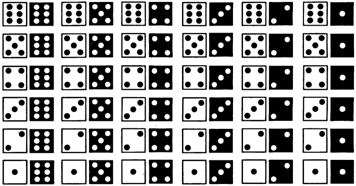
# Espacio Muestral

El **espacio muestral** es el conjunto de todos los resultados posibles de un experimento aleatorio, y que denotamos por $\Omega$.

- Si lanzamos una moneda $\Omega = \{C, S\}$, $|\Omega| = 2$.
- Si lanzamos un dado $\Omega = \{1, 2, 3, 4, 5, 6\}$, $|\Omega| = 6$.
- Si lanzamos un dado y una moneda,

$$\Omega = \{(1, C), (1, S), (2, C), (2, S), \ldots, (6, C), (6, S)\}, |\Omega| = 12.$$

- Si lanzamos dos dados, tenemos $|\Omega| = 36$.



# Evento

Un **evento** es un subconjunto del espacio muestral.

- Si $A$ es el evento de obtener un número mayor a 3 al lanzar un dado, entonces $A = \{4, 5, 6\}$.

- Si $B$ es el evento de obtener al menos un sello al lanzar dos monedas, entonces, $B = \{(C, S), (S, C), (S, S)\}$.

- Si $C$ es el evento de obtener suma igual a 7 al lanzar dos dados, entonces, $C = \{(1, 6), (6, 1), (5, 2), (2, 5), (4, 3), (3, 4)\}$.

- Si $D$ es el evento de obtener el primer sello al lanzar una moneda infinitas veces, entonces,

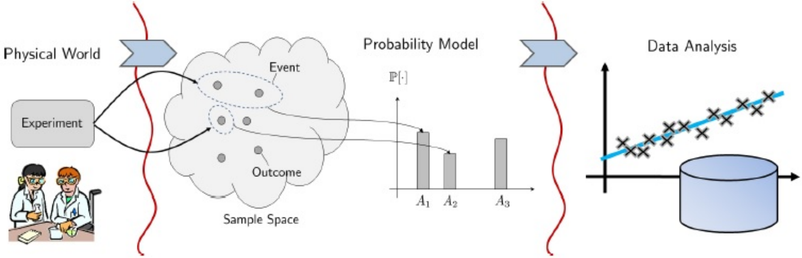
$$D = \{(S), (C, S), (C, C, S), (C, C, C, S), (C, C, C, C, S), (C, C, C, C, C, S), \ldots\}.$$

# Definición de Probabilidad

Dado un evento $A$, decimos que la probabilidad del evento $A$, y que denotamos por $p(A)$, es:

$$P(A) = \frac{\text{N° de ocurrencias del evento } A}{\text{N° de posibilidades del espacio muestral}} = \frac{|A|}{|\Omega|}.$$

Denotamos $(\Omega, \mathcal{F}, P)$ al espacio de probabilidad.



## Ejemplo 1

De los 39 alumnos de una clase, 16 escogieron francés y 27 inglés. 9 alumnos eligieron ambos, y el resto no escogió ninguno de ellos. Si se elige al azar un alumno de dicha clase, determinar la probabilidad que el alumno

(a) elija francés.
(b) elija inglés.
(c) elija ambos idiomas.
(d) elija francés o inglés.
(e) elija o francés o inglés.
(f) elija francés, pero no inglés.
(g) elija inglés, pero no francés.
(h) no elija francés ni inglés.

# Ejemplo 2

En un almacén hay 10 motores, 3 de los cuales están defectuosos. Los otros siete están en buenas condiciones de funcionamiento. Un primer inspector entra al almacén y selecciona uno de los motores. Al día siguiente, un segundo inspector entra en este mismo almacén y selecciona un motor. La respuesta depende de si el primer motor seleccionado permaneció en el almacén o fue retirado. Elaborar un diagrama de árbol para representar las probabilidades en cada caso.

## Ejemplo 3

Calcula la probabilidad de:

(a) Obtener aleatoriamente un número entero positivo de dos cifras múltiplo de 3.

(b) Obtener aleatoriamente un número entero positivo de tres cifras múltiplo de 6.

## Axiomas de Probabilidad

Dados los eventos $A$ y $B$ en $\Omega$ del espacio de probabilidad $(\Omega, \mathcal{F}, P)$, se cumplen los axiomas:

(a) $0 \leq P(A) \leq 1$.

(b) $P(\emptyset) = 0$.

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

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

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

(f) Si $P(A \cap B) = 0$ entonces $P(A \cup B) = P(A) + P(B)$, y decimos que $A$ y $B$ son eventos mutuamente excluyente.

# Ejemplo 4

Sean A y B son eventos en $\Omega$ con probabilidades 0.8 y 0.7, respectivamente. Supongamos también que $P(A \cap B) = 0{,}6$. Calcular $P(A \cup B)$.

# Ejemplo 5

Sean A y B son eventos en $\Omega$. Si $P(A) = 0{,}6$, $P(B^c) = 0{,}4$ y $P(A \cap B) = 0{,}2$.

# Ejemplo 6

Los números telefónicos en una país tienen 9 cifras, y debe cumplir solo dos restricciones: inicie en 6, y sus últimas 5 cifras no tengan la cadena 091. Por ejemplo, 675091822 es válido, mientras que 675809122 no lo es. Si generamos el número telefónico aleatoriamente, ¿Cuál es la probabilidad de obtener un número válido?

## Ejemplo 7: El problema de Monty Hall

Suponga que está jugando un juego con tres puertas y detrás de una de ellas hay un automóvil como premio. El presentador del juego te pide que elijas una puerta. Si eliges la puerta que tiene el automóvil, ganas el automóvil. Sin embargo, después de elegir su puerta, el presentador abre otra puerta en el juego y muestra que contiene una cabra. El presentador pregunta ¿Quieres cambiar la puerta?.

# Ejemplo 8: El Problema del Apostador

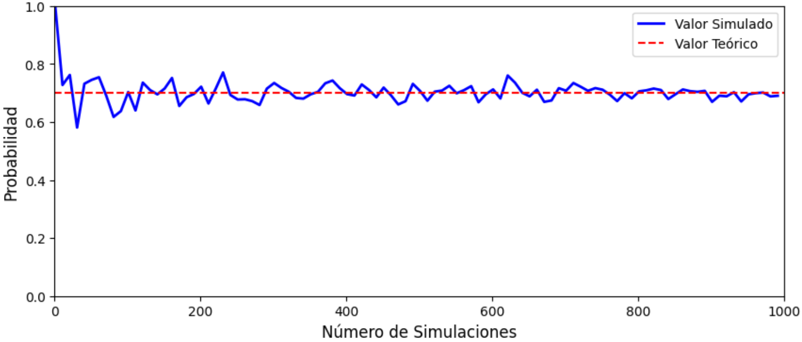
Un apostador paga 5 soles para participar en un juego de dados, disputando con un oponente haber quien tiene el punto más alto. El apostador y el oponente tiran sus dados, con las reglas:

- Si el punto del apostador es más alto, gana 2 veces la diferencia entre su punto y el obtenido por el oponente.
- Si el punto del apostador es menor o igual que el del oponente, no gana nada.

¿Vale la pena para el apostador? ¿Mejorará la situación si aumentamos la apuesta?

# Ley de los grandes números

El matemático italiano Gerolamo Cardano (1501–1576) afirmó sin pruebas que la precisión de las estadísticas empíricas tienden a mejorar con el número de intentos. Luego, esta observación fue formalizado como la ley de los grandes números.



# Simulación Montecarlo

Fue inventado por John von Neumann y Stanislaw Ulam durante la Segunda Guerra Mundial para mejorar la toma de decisiones en condiciones de incertidumbre. Tiene el nombre de un conocido barrio de Mónaco célebre por su casino, ya que el elemento de la suerte es la base del método.

Se usa una distribución de probabilidad para cualquier variable que tenga una incertidumbre inherente. Luego, calculamos los resultados repetidamente, utilizando cada vez un conjunto diferente de números aleatorios. Este proceso se repite miles de veces y así generar un gran número de resultados probables.

Las aplicaciones son diversas y busca evaluar el impacto del riesgo en muchas situaciones de la vida real como, por ejemplo, precios de las acciones, previsión de ventas, gestión de proyectos, fijación de precios y en modelos de Inteligencia Artificial.

## Estimación de $\pi$ usando Simulación Montecarlo

Considerando 1000 simulaciones, tenemos 775 puntos internos y el valor estimado de $\hat{\pi} \approx 4 \cdot \frac{775}{1000} = 3{,}1$, y a medida que aumenta la cantidad de simulaciones, $\hat{\pi}$ se aproxima a $\pi$.