Fundamentos de Probabilidad y Estadística

para Inteligencia Artificial

Estructura de la sesión

Estadística Descriptiva

- Medidas de Tendencia Central
- Medidas de Dispersión
- Correlación
- Ejemplos y aplicaciones para IA

Teoría de Probabilidad

- Conceptos Clave
- Probabilidad Conjunta, Marginal y Condicional
- Distribuciones de Probabilidad
- Inferencia Probabilística
- Ejemplos

¿Por qué...

necesitamos probabilidad y estadística en IA?

Incertidumbre

Modelar la Incertidumbre y la Toma de Decisiones

Estadística

Analizar Datos y Extraer Patrones

¿Por qué...

necesitamos probabilidad y estadística en IA?

Probabilidad

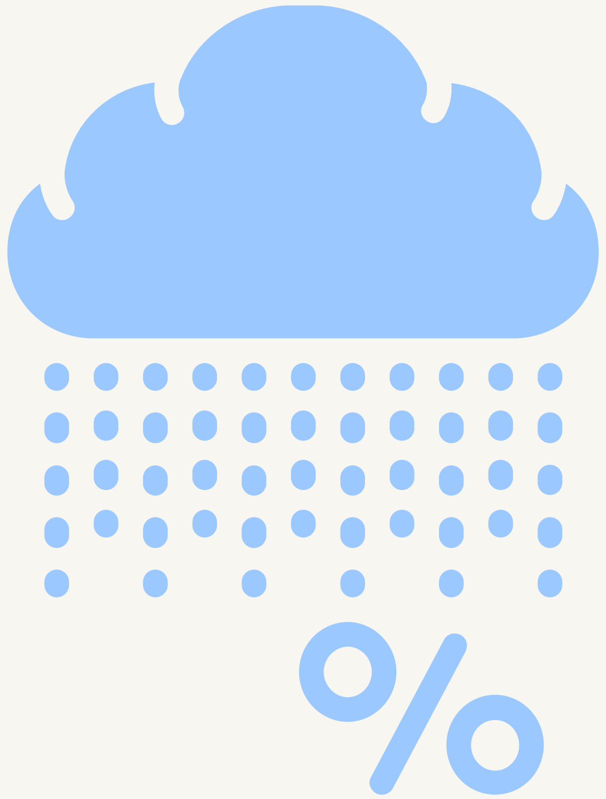
Algoritmos como clasificadores, Redes Neuronales Artificiales y Árboles de Decisión

Evaluar Modelos

¿Qué Tan Buenas Son Nuestras Predicciones?

¿Cómo la probabilidad y la estadística nos ayudan a tomar mejores decisiones con datos?

Imagina que sales de casa... No tienes certeza absoluta de si lloverá, pero puedes usar datos y probabilidades para tomar la mejor decisión.



Diferencia entre probabilidad (incertidumbre en eventos) y estadística (análisis de datos).

¿Debo llevar paraguas hoy?

Datos históricos y patrones climáticos

- ¿Cuántos días al año llueve en mi ciudad?
- ¿Llueve más en ciertas estaciones o meses?
- Si ayer llovió, ¿qué tan probable es que hoy también llueva?

Probabilidad condicional

Supongamos que:

- El pronóstico del tiempo dice que hay un 80% de probabilidad de lluvia.
- El cielo está nublado y hay alta humedad.

Decisión	¿Llovió?	Consecuencia
Llevaste paraguas	Si	No te mojas
Llevaste paraguas	No	Cargas un paraguas innecesariamente
No llevaste paraguas	Sí	Te mojas y arruinas tu ropa
No llevaste paraguas	No	No Cargas Peso Extra

Análisis de Datos

Estadística Descriptiva

es una rama de la estadística que se encarga de organizar, resumir y representar datos numéricos

Medidas de Tendencia Central

- Media, mediana y moda.

Medidas de Dispersión

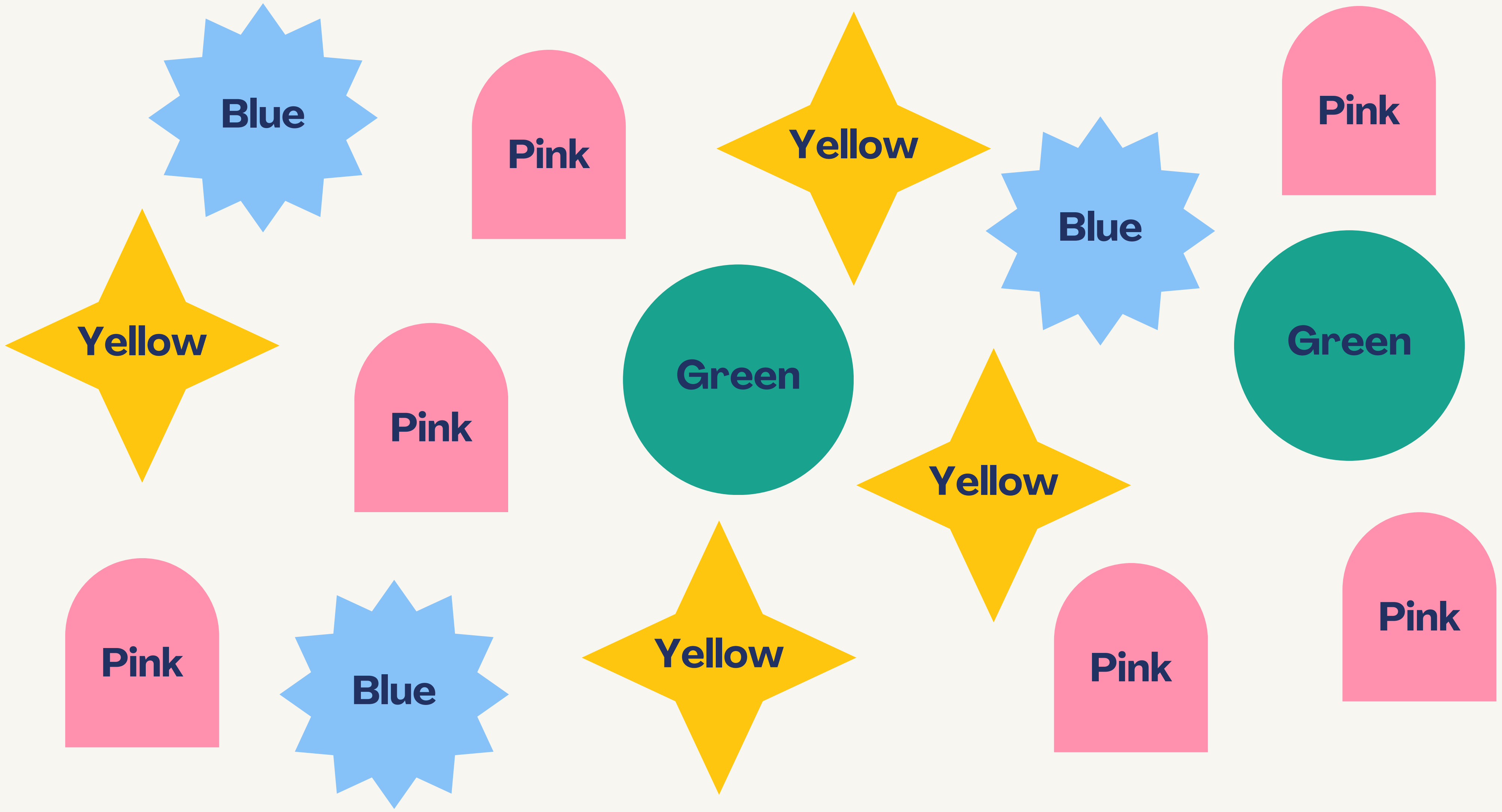
- Varianza y desviación estándar

Correlación y Covarianza

- Concepto de correlación: Relación entre variables.

Votación de Colores Favoritos

Supongamos que realizamos una encuesta a 100 personas sobre su color favorito y obtenemos los siguientes resultados:



¿Cómo analizar y organizar los resultados?

Preguntas a Resolver

¿Cuál es el color más popular?

¿Cuántos votos tiene un color en promedio?

¿Cuánta variabilidad hay entre los votos?

Estos mismos principios se aplican en Machine Learning para analizar datos antes de entrenar modelos

Análisis Exploratorio de Datos

Ahora aplicamos la estadística descriptiva para analizar estos datos:

Color	Número de personas
Azul	35
Rojo	20
Verde	15
Amarillo	10
Negro	10
Blanco	10

Which color is the mode?

Medidas de Tendencia Central:

- Media
- Mediana
- Moda

Medidas de Dispersión:

- Rango
- Varianza
- Desviación

Color	Número de personas
Azul	35
Rojo	20
Verde	15
Amarillo	10
Negro	10
Blanco	10

Ejemplo Varianza

Tiempos de entrega en dos sucursales de una empresa

Imagina que dos sucursales de una empresa entregan paquetes, y queremos analizar la rapidez en la entrega.

Sucursal 1:

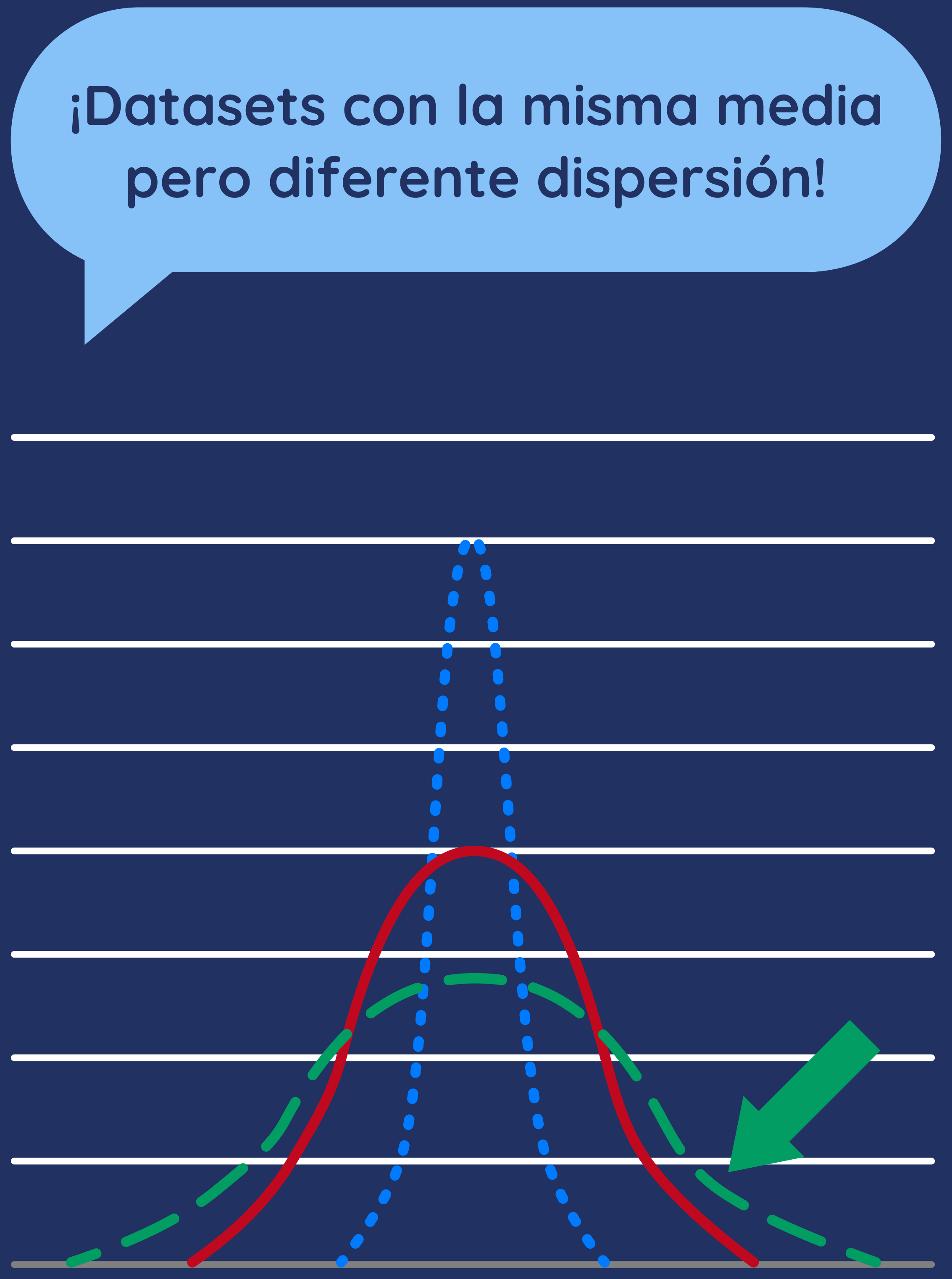
Tiempos en días: 5, 5, 5, 5, 5

- Media = 5 días
- Varianza = 0 (todas las entregas son en el mismo tiempo)

Sucursal 2:

Tiempos en días: 1, 3, 5, 7, 9

- Media = 5 días
- Varianza > 0 (algunas entregas son muy rápidas y otras tardan más)



Supongamos que, además de preguntar el color favorito, también preguntamos la edad de los encuestados y obtenemos la siguiente tabla:

Color Favorito	Número de Votos (X)	Edad Promedio (Y)
Azul	35	25
Rojo	20	30
Verde	15	22
Amarillo	10	35
Negro	10	28
Blanco	10	40

Relación entre variables

¿Existe una relación entre los votos y la edad?

La covarianza nos indica si existe una relación entre los votos y la edad.

Correlación

¿Qué tan fuerte es la relación?

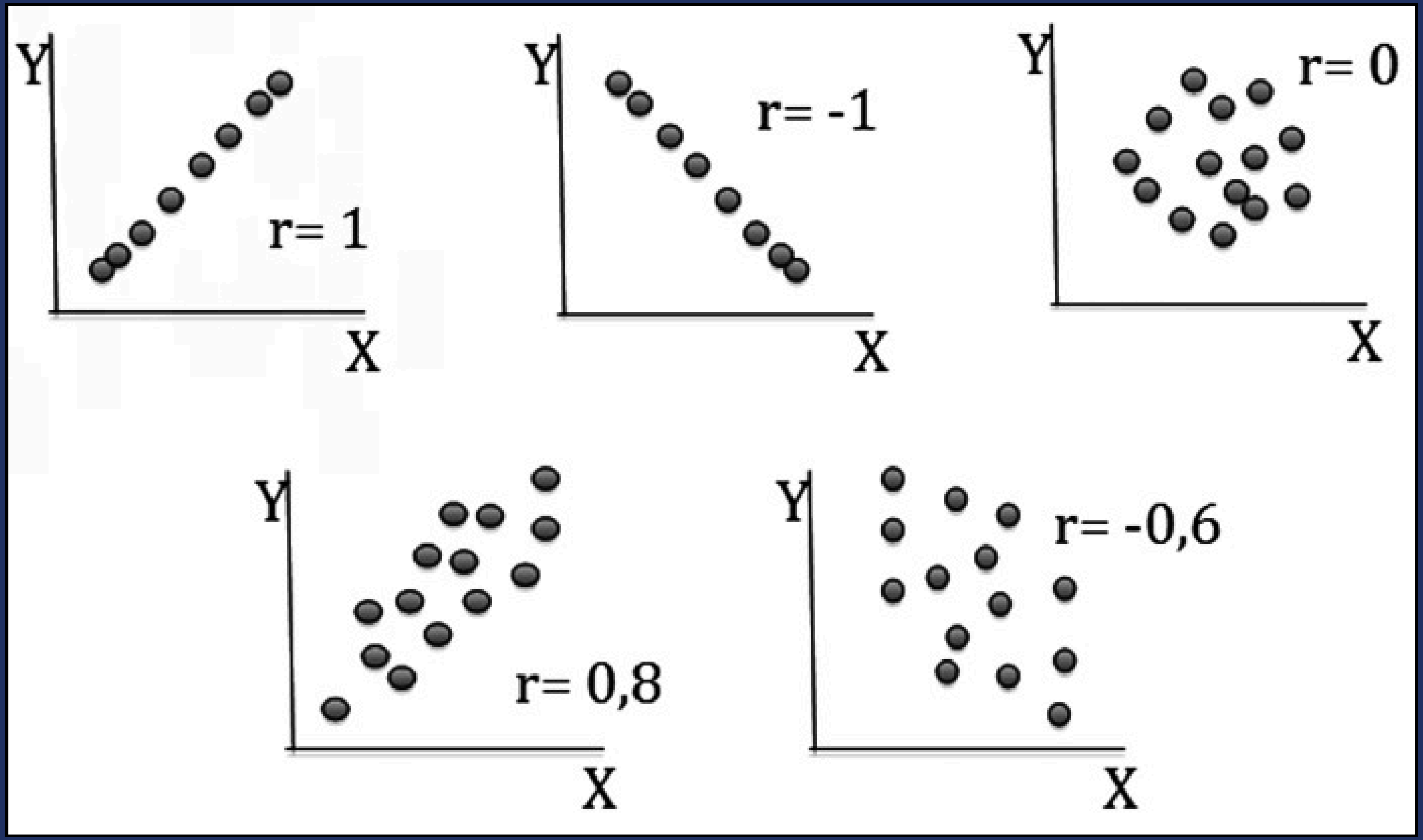
Estos cálculos ayudan a entender qué variables están relacionadas en un dataset, lo cual es útil para seleccionar características

Covarianza

Si Cov(X, Y) > 0, significa que a mayor número de votos, mayor edad promedio.
Si Cov(X, Y) < 0, significa que a mayor número de votos, menor edad promedio.
Si Cov(X, Y) ≈ 0, no hay relación entre los votos y la edad.

Correlación de Pearson

El resultado estará en el rango [-1, 1]:
✅ r > 0 → Correlación positiva (cuando sube X, sube Y)
❌ r < 0 → Correlación negativa (cuando sube X, baja Y)
⚠ r ≈ 0 → No hay relación



Con el ejemplo, ¿Cómo interpretar los valores?

- Si encontramos una correlación fuerte entre edad y popularidad de un color, podríamos predecir qué colores son más atractivos para diferentes grupos de edad.
- Si la correlación es débil, significa que los gustos por los colores son independientes de la edad.

Por lo tanto...

- La covarianza indica la dirección de la relación, pero no la intensidad.
- La correlación de Pearson nos da tanto la dirección como la fuerza de la relación de forma estandarizada.

¿Cómo analizar y organizar los resultados?

Presenting Results

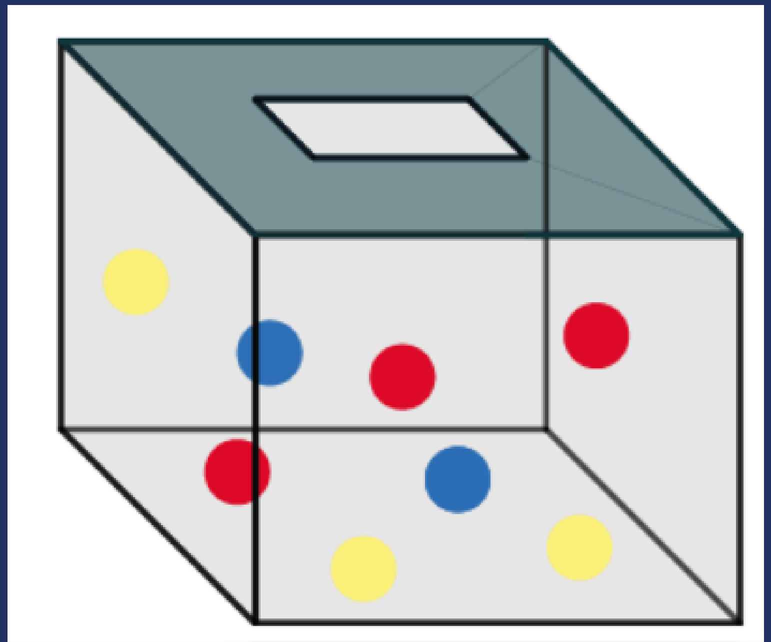
Fundamentos de Probabilidad

La probabilidad compuesta o conjunta es el cálculo de la probabilidad cuando un experimento de probabilidad simple se repite varias veces o se relaciona un experimento con otro.

Sucesos Independientes vs Sucesos Dependientes

Sucesos independientes:
Dos sucesos son independientes si la ocurrencia de uno no afecta la probabilidad de que ocurra el otro. En otras palabras, el hecho de que uno suceda no influye en la probabilidad del otro.

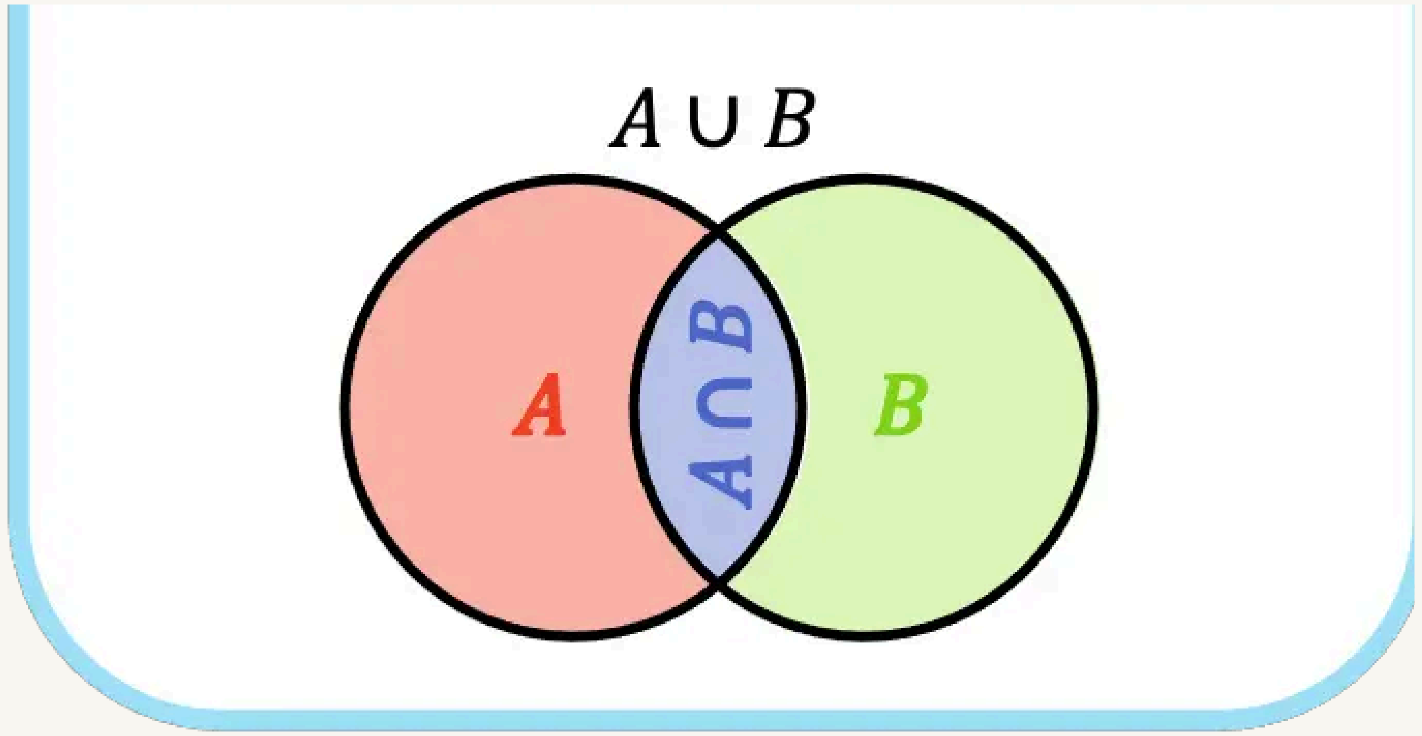
Sucesos dependientes:
Dos sucesos son dependientes si la ocurrencia de uno afecta la probabilidad de que ocurra el otro. En este caso, la probabilidad de un suceso cambia dependiendo de si el otro suceso ha ocurrido o no.



Reglas fundamentales de probabilidad

- **Regla de la suma:**

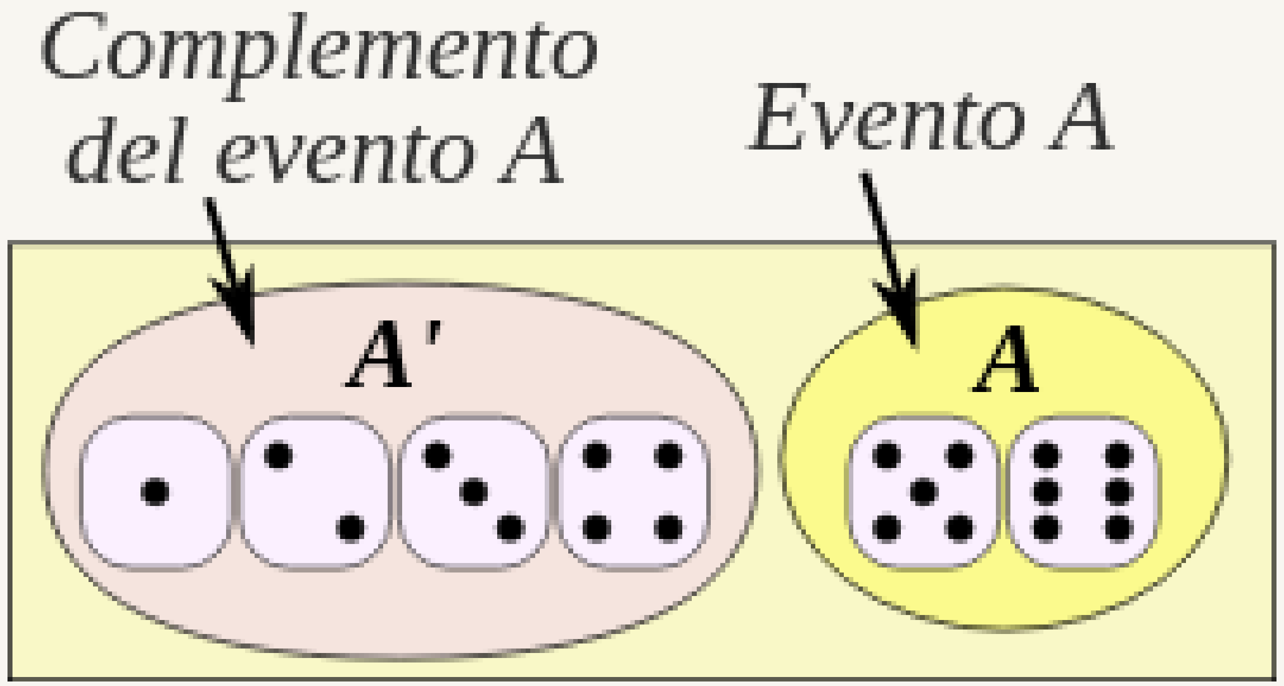
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$



Reglas fundamentales de probabilidad

- Regla del complemento:

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

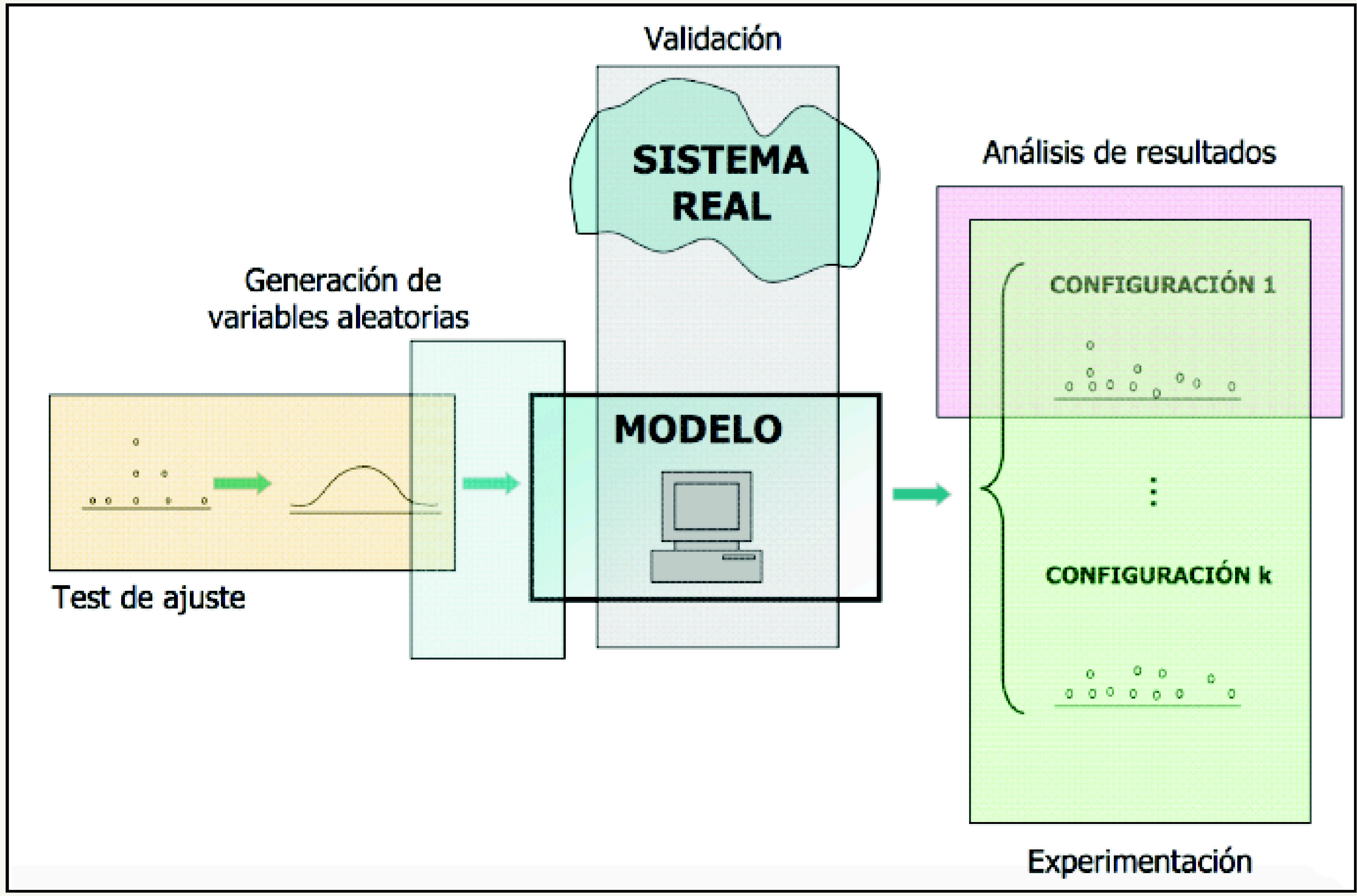


Distribución de Probabilidad

Describe cómo se reparten las probabilidades de los posibles valores de una variable aleatoria

En un contexto de simulación, cada elemento incierto del sistema, se modela mediante una variable aleatoria con una cierta distribución.

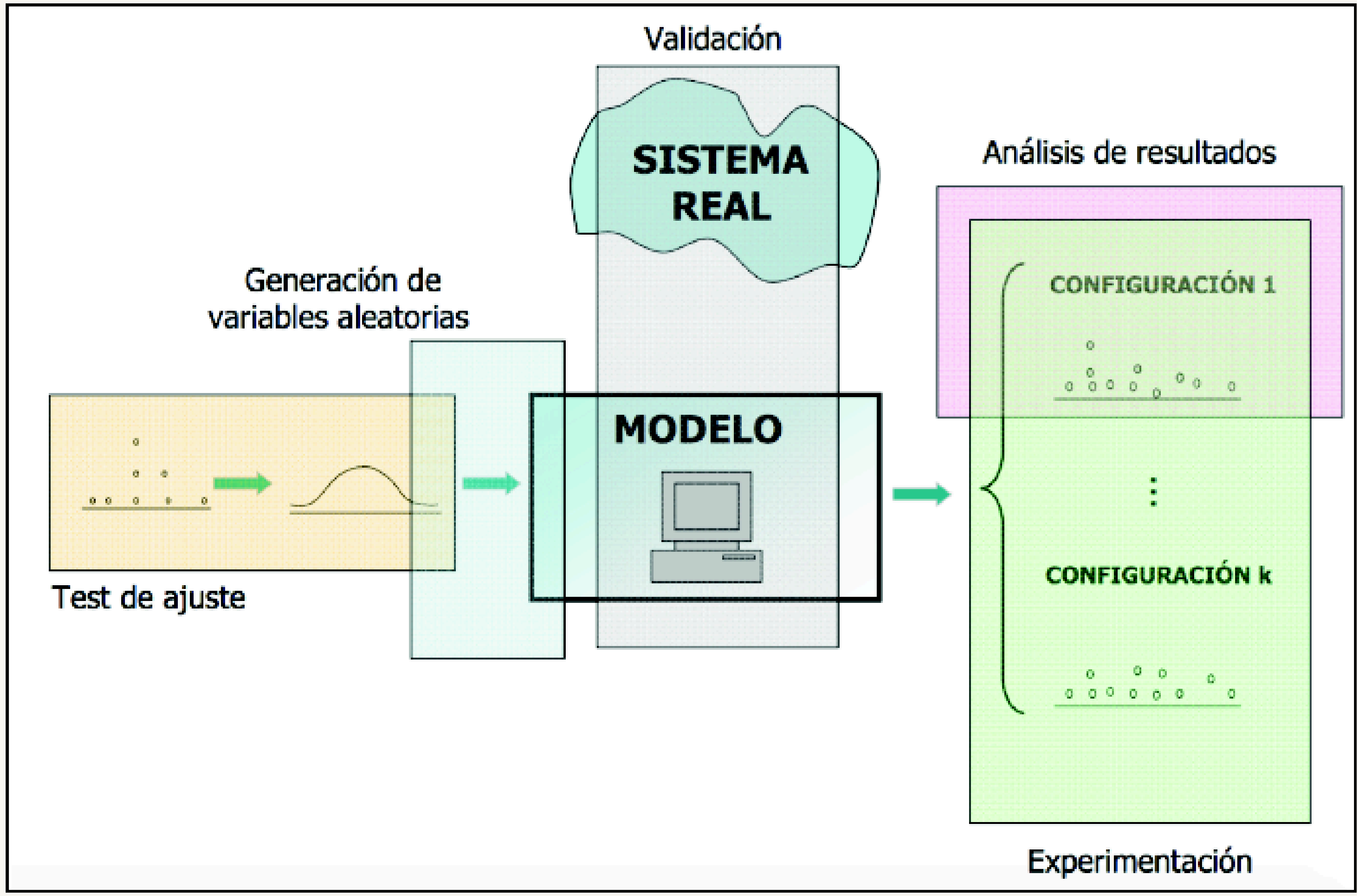
El uso de distribuciones en el modelo permite incorporar el comportamiento estocástico del sistema, imitando su variabilidad natural.



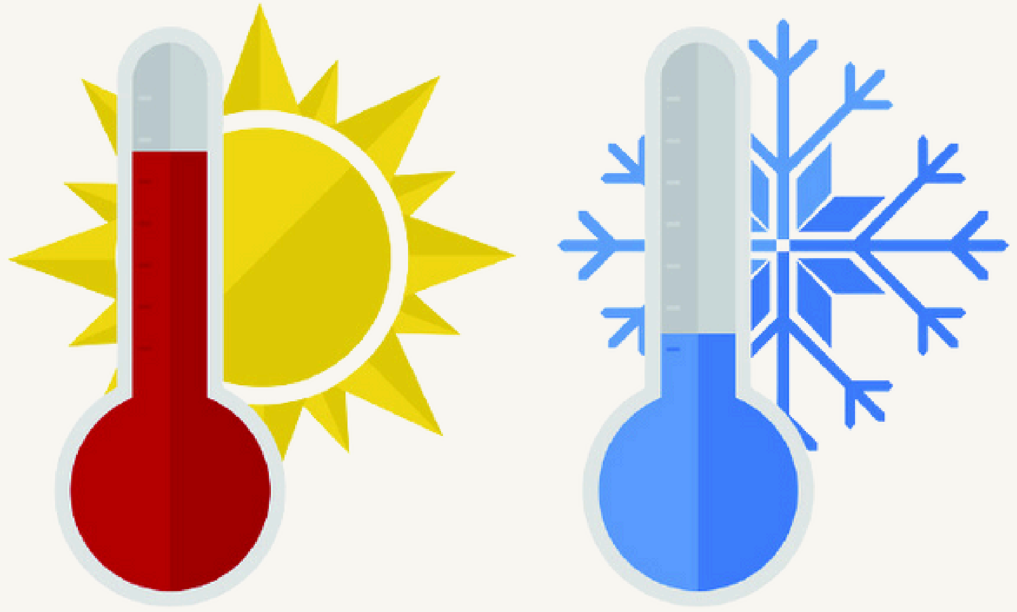
...es un modelo matemático que describe la incertidumbre sobre una variable aleatoria. Se usa para representar y predecir el comportamiento de datos en algoritmos de aprendizaje automático, modelos probabilísticos y redes neuronales.

Distribución de Probabilidad

1. Se observa el sistema real y se recolectan datos.
2. Se ajustan distribuciones para modelar las variables aleatorias.
3. Se crea un modelo computacional que integra esas distribuciones y la lógica del sistema.
4. Se realizan experimentos cambiando configuraciones.
5. Se analizan los resultados para tomar decisiones o proponer mejoras en el sistema real.



Distribuciones

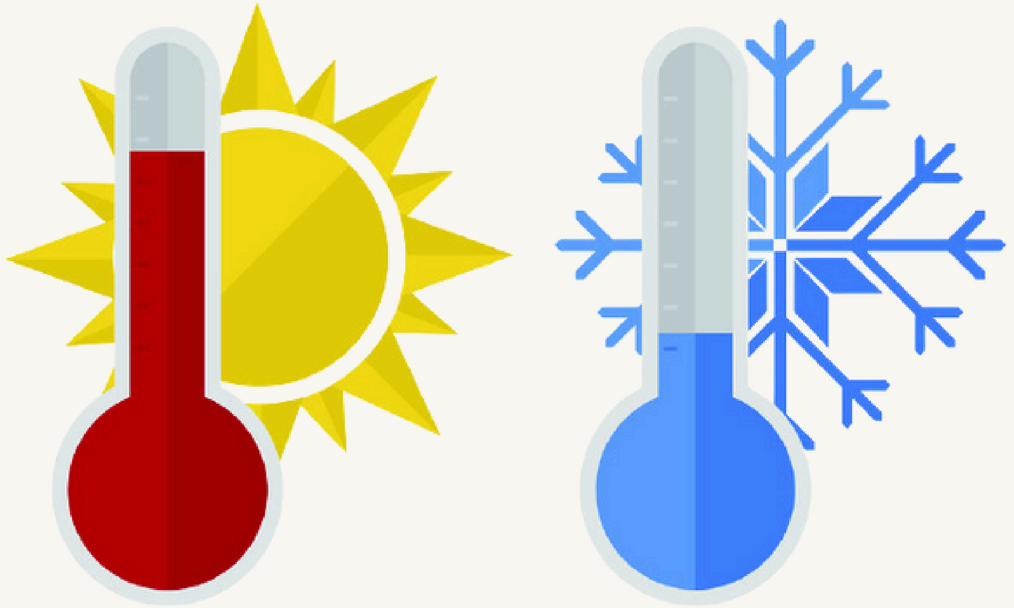


Probabilidad conjunta y marginal

Estado	**Eventos**
Llueve	**9**
No Llueve	**12**

Eventos de lluvia

Distribuciones

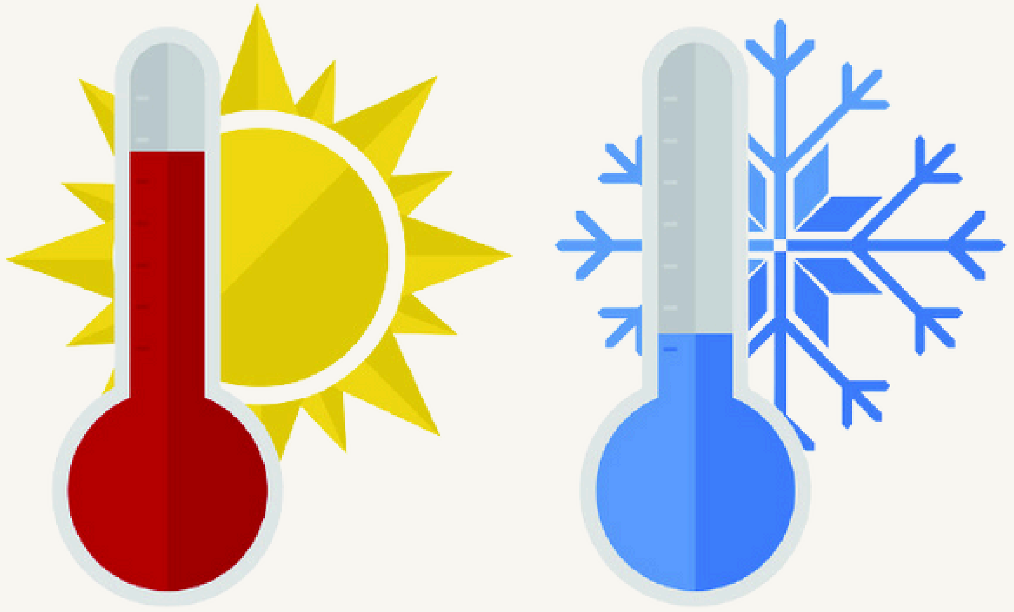


Estado	Eventos
Soleado	10
Nublado	11

Eventos del tiempo

Probabilidad conjunta y marginal

Tabla de Eventos



	Soleado	Nublado	
Llueve	2	7	9
No llueve	8	4	12
	10	11	

Probabilidad conjunta y marginal

Probabilidad Conjunta: Especifica la probabilidad de que dos eventos sucedan (NO necesariamente dependientes).

Probabilidad Marginal: Es la probabilidad de que una sola variable tome un valor específico, ignorando las demás variables.

Probabilidad Marginal

Var 1	Var 2	Probabilidad
Soleado	Llueve	2/21
Soleado	No Llueve	8/21
Nublado	Llueve	7/21
Nublado	No Llueve	4/21

Marginalizamos variable 2 ignorando la variable 1

Estado 2	Probabilidad
Llueve	9/21
No Llueve	12/21

Distribucion Marginal: Es la distribución de una de las variables (o un subconjunto) cuando no nos interesa el resto.

Se obtiene “integrando” o “sumando” la distribución conjunta sobre las variables que se ignoran.

Ejemplo: calcular la **probabilidad que llueva o no llueva**

NOTA: Una distribución es el conjunto completo de probabilidades para todos los valores/eventos, mientras que la probabilidad es la medida para un evento en particular.

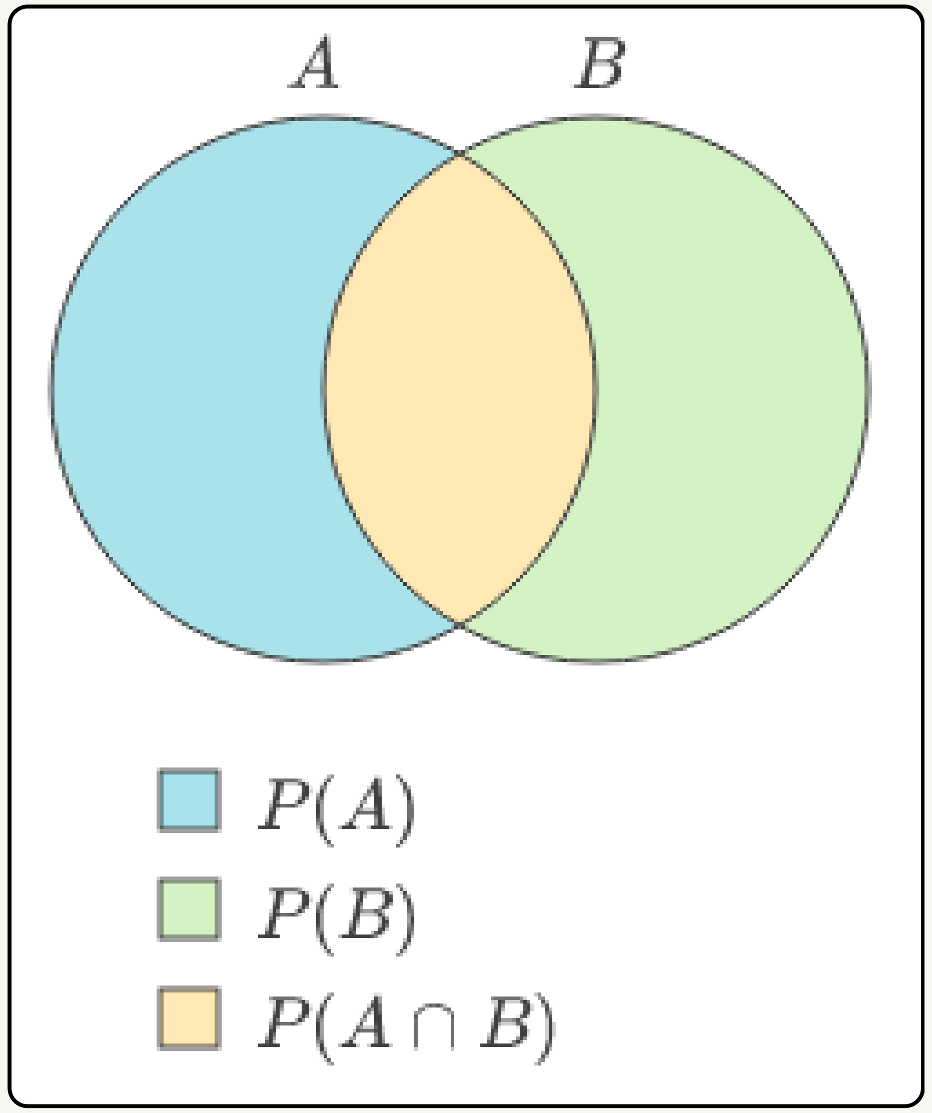
Probabilidad Condicional

Es la probabilidad de que ocurra un evento dado que otro ya ha ocurrido.

Se denota como:

P(A | B) <- Cual es la probabilidad que ocurra A dado que ocurrio B

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



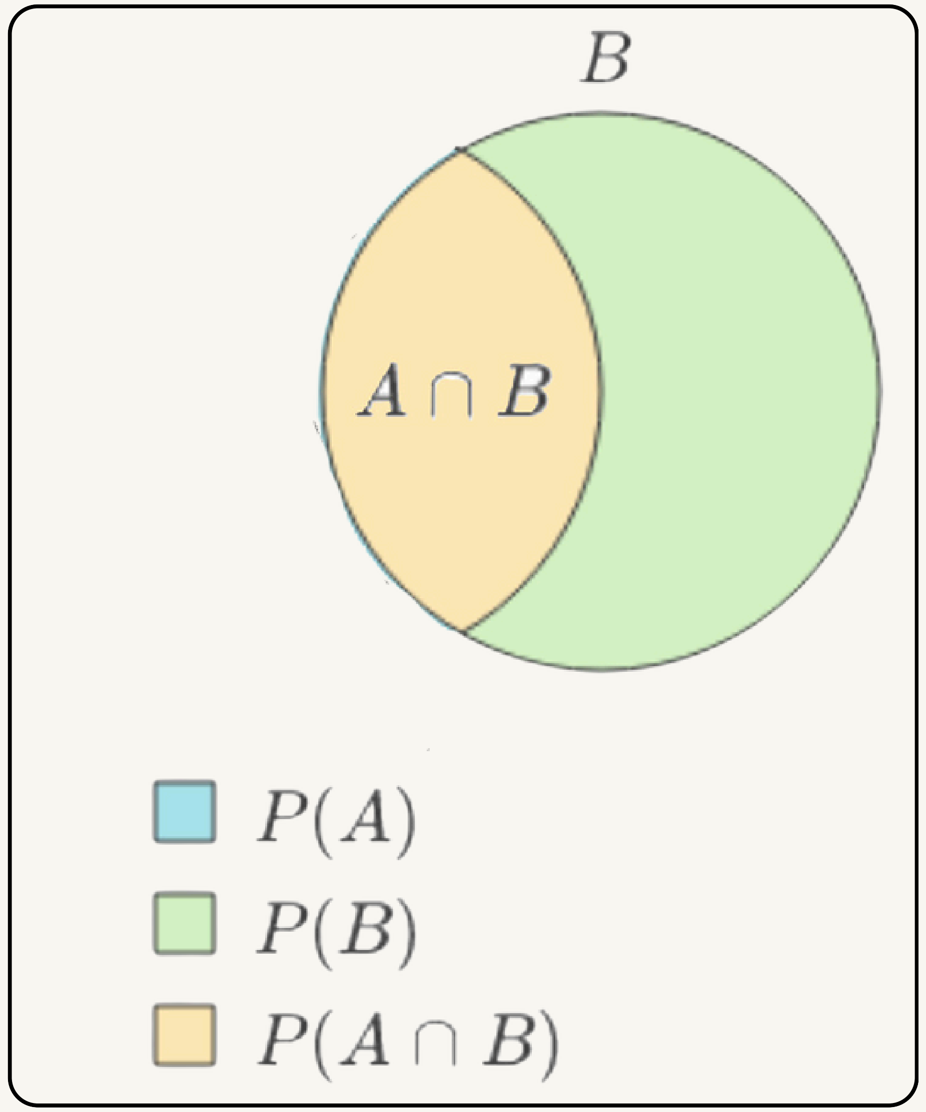
Probabilidad Condicional

Es la probabilidad de que ocurra un evento dado que otro ya ha ocurrido.

Se denota como:

P(A | B) <- Cual es la probabilidad que ocurra A dado que ocurrio B

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



Probabilidad Condicional

Dada la tabla de probabilidad conjunta

X	Y	P
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

Calcule:

1. $P(+x|+y)$
2. $P(-x|+y)$
3. $P(-y|+x)$

Para: $P(+x|+y)$

$$P(+x|+y) = \frac{P(+x,+y)}{P(+y)}$$

NO TENEMOS P(+y) !!!!!

Probabilidad Condicional

Dada la tabla de probabilidad conjunta

X	Y	P
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

Calcule:

1. $P(+x|+y)$
2. $P(-x|+y)$
3. $P(-y|+x)$

Para: $P(+x|+y)$

$$P(+x|+y) = \frac{P(+x,+y)}{P(+y)}$$

NO TENEMOS P(+y) !!!!!
entonces:

Probabilidad marginal P(+y)

$$P(+y) = P(+x,+y)+P(-x,+y)=0.2+0.4=0.6$$

$$P(+x|+y) = \frac{P(+x,+y) = 0.2}{P(+y) = 0.6}$$

$$P(+x|+y) \approx 0.333$$

Probabilidad Condicional

Dada la tabla de probabilidad conjunta

X	Y	P
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

Calcule:

1. $P(+x|+y)$
2. $P(-x|+y)$
3. $P(-y|+x)$

Para: $P(-x|+y)$

$$P(-x|+y) = \frac{P(-x,+y)}{P(+y)}$$

$$P(-x|+y) = \frac{P(-x,+y) = 0.4}{P(+y) = 0.6}$$

$$P(-x|+y) \approx 0.667$$

Probabilidad Condicional

Dada la tabla de probabilidad conjunta

X	Y	P
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

Calcule:

1. $P(+x|+y)$
2. $P(-x|+y)$
3. $P(-y|+x)$

Para: $P(-y|+x)$

$$P(-y|+x) = \frac{P(-y,+x)}{P(+x)}$$

NO TENEMOS P(+x) !!!!!
entonces:

Probabilidad marginal P(+x)

$$P(+x) = P(+x,-y)+P(+x,-y)=0.2+0.3=0.5$$

$$P(-y|+x) = \frac{P(-y,+x) = 0.3}{P(+x) = 0.5}$$

$$P(-y|+x) = 0.6$$

y si nos piden calcular ¿P(X|Y = -y)?

Normalizacion

Es el proceso de asegurar que la suma de todas las probabilidades en un espacio dado sea igual a 1.

Supongamos la siguiente tabla de probabilidades

Var 1	Var 2	Probabilidad
Soleado	Llueve	0.3
Soleado	No Llueve	0.5
Nublado	Llueve	0.4
Nublado	No Llueve	0.6

Suma de probabilidades = 1.8

Normalizar dividiendo entre la suma total

Normalizacion

Esto se logra ajustando las probabilidades para que puedan interpretarse como proporciones relativas.

Se parte de una función que no está garantizado que sume 1, y se fuerza a que lo haga dividiéndola entre la suma total de sus valores, obteniendo así una distribución válida.

Supongamos la siguiente tabla de probabilidades

Var 1	Var 2	Probabilidad	Normalizacion	Nueva Probabilidad
Soleado	LLueve	0.3	0.3/1.8	0.1667
Soleado	No Llueve	0.5	0.5/1.8	0.2778
Nublado	Llueve	0.4	0.4/1.8	0.2222
Nublado	No Llueve	0.6	0.6/1.8	0.3333

la suma de las nuevas probabilidades es 1, por lo que la distribución está normalizada

INFERENCIA

Inferencia Probabilistica

Es un enfoque utilizado en inteligencia artificial, estadística y ciencias de la computación para modelar y razonar bajo incertidumbre.

Consiste en utilizar probabilidades para representar creencias sobre eventos desconocidos y actualizar esas creencias a medida que se obtiene nueva información (evidencia).

Su objetivo es calcular la probabilidad de ciertas hipótesis o eventos, dados los datos observados y un modelo probabilístico.

INFERENCIA

Inferencia Probabilistica

Incertidumbre:

- Modela situaciones donde no hay certeza absoluta, como: diagnóstico médico, predicción del clima, detección de spam.
- Ejemplo: ¿Cuál es la probabilidad de que un paciente tenga una enfermedad, dado un resultado positivo en una prueba?

INFERENCIA

Inferencia Probabilistica

Modelos Probabilísticos:

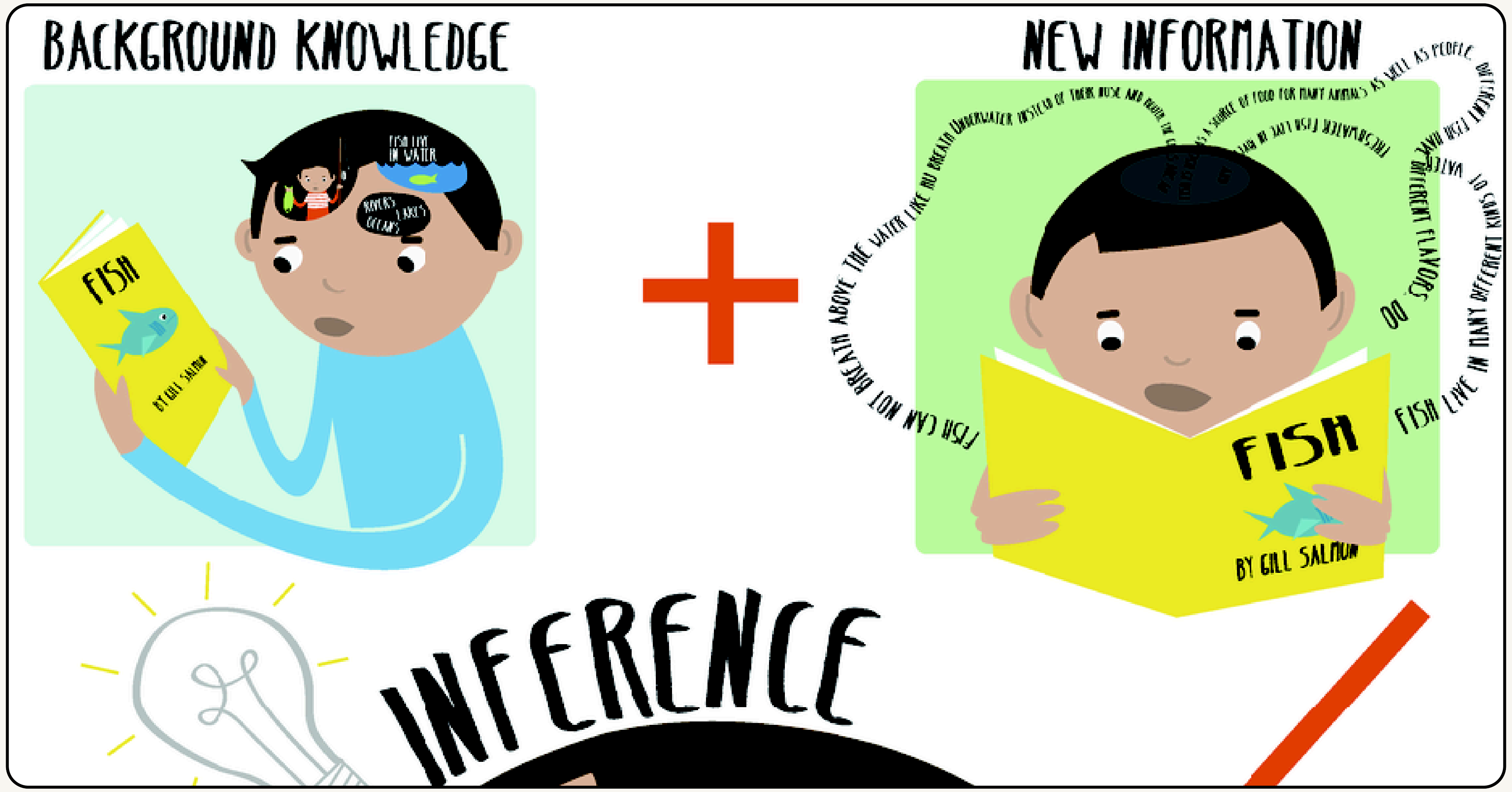
- Representan relaciones entre variables mediante distribuciones de probabilidad.
- Ejemplos: Redes Bayesianas, Cadenas de Markov, Modelos de Mezclas Gaussianas.

INFERENCIA

Inferencia Probabilistica

Actualización de Creencias:

- Usa reglas como el Teorema de Bayes para ajustar probabilidades ante nueva evidencia.



INFERENCIA

Inferencia por Enumeracion

Es un método exacto utilizado en probabilidad y redes bayesianas para calcular probabilidades marginales o condicionales enumerando y sumando todas las posibles combinaciones de variables ocultas (no observadas).

Consiste en:

- Identificar variables:
  - Separar variables en: consulta (X), evidencia (E), y ocultas (Y).

1. Calcular la distribución conjunta
2. Marginalizar variables ocultas
3. Normalizar

Ejemplo:

Supongamos que estamos modelando si una persona tiene gripe X, dado que tiene fiebre E, pero no sabemos si estuvo expuesta a un virus H

- X = ¿El paciente tiene gripe?
- E ?
- Y ?

INFERENCIA

Inferencia por Enumeracion

Supongamos los siguientes datos probabilisticos:

Estacion	Temp.	Clima	Prob.
Verano	calido	Soleado	0.30
Verano	calido	Lluvioso	0.05
Verano	frio	Soleado	0.10
Verano	frio	Lluvioso	0.05
Invierno	calido	Soleado	0.10
Invierno	calido	Lluvioso	0.05
Invierno	frio	Soleado	0.15
Invierno	frio	Lluvioso	0.20

Cual es la probabilidad de:

- P(Clima)
- P(Clima | Invierno)
- P(Clima | Invierno, Calido)

Estacion	Temp.	Clima	Prob.
Verano	calido	Soleado	0.30
Verano	calido	Lluvioso	0.05
Verano	frio	Soleado	0.10
Verano	frio	Lluvioso	0.05
Invierno	calido	Soleado	0.10
Invierno	calido	Lluvioso	0.05
Invierno	frio	Soleado	0.15
Invierno	frio	Lluvioso	0.20

Para P(Clima):

Identificacion de variables:

- Variables de consulta: Clima - C
- Variables de evidencia: Ninguna
- Variables de ocultas: Temperatura - T, Estacion E

P(C) = P(C=Soleado) , P(C=Lluvioso)

Marginalizar P(C)
P(C=Soleado) = 0.30 + 0.10 + 0.10 + 0.15 = 0.65

P(C=Lluvioso) = 0.05 + 0.05 + 0.05 + 0.20 = 0.35

Entonces P(C) es:

Clima	Prob.
Soleado	0.65
Lluvioso	0.35

Estacion	Temp.	Clima	Prob.
Verano	calido	Soleado	0.30
Verano	calido	Lluvioso	0.05
Verano	frio	Soleado	0.10
Verano	frio	Lluvioso	0.05
Invierno	calido	Soleado	0.10
Invierno	calido	Lluvioso	0.05
Invierno	frio	Soleado	0.15
Invierno	frio	Lluvioso	0.20

Para P(Clima | Invierno):

Identificacion de variables:

- Variables de consulta: Clima - C
- Variables de evidencia: Estacion - E = Invierno
- Variables de ocultas: Temperatura - T

P(C | E = Invierno) = P(Soleado | E = Invierno) , P(Lluvioso | E = Invierno)

$$P(C \mid E = \text{Invierno}) = \frac{P(C = \text{Soleado}, E = \text{Invierno})}{P(E = \text{Invierno})}, \frac{P(C = \text{Lluvioso}, E = \text{Invierno})}{P(E = \text{Invierno})}$$

Probabilidad total de Invierno (Marginalizacion):

P(Invierno) = 0.10 + 0.05 + 0.15 + 0.20 = 0.50

Probabilidades de cada clima en Invierno:

P(Soleado,Invierno) = 0.10 + 0.15 = 0.25

P(Lluvioso,Invierno) = 0.05 + 0.20 = 0.25

Estacion	Temp.	Clima	Prob.
Verano	calido	Soleado	0.30
Verano	calido	Lluvioso	0.05
Verano	frio	Soleado	0.10
Verano	frio	Lluvioso	0.05
Invierno	calido	Soleado	0.10
Invierno	calido	Lluvioso	0.05
Invierno	frio	Soleado	0.15
Invierno	frio	Lluvioso	0.20

Para P(Clima | Invierno):

$$P(C \mid E = \text{Invierno}) = \frac{P(C = \text{Soleado}, E = \text{Invierno})}{P(E = \text{Invierno})}, \frac{P(C = \text{Lluvioso}, E = \text{Invierno})}{P(E = \text{Invierno})}$$

Probabilidad total de Invierno (Marginalizacion):

P(Invierno) = 0.50

Probabilidades de cada clima en Invierno (Marginalizacion):

P(Soleado,Invierno) = 0.25
P(Lluvioso,Invierno) = 0.25

$$P(C \mid E = \text{Invierno}) = \frac{0.25}{0.5}, \frac{0.25}{0.5}$$

Estacion	Clima	Prob.
Invierno	Soleado	0.50
Invierno	Lluvioso	0.50

Estacion	Temp.	Clima	Prob.
Verano	calido	Soleado	0.30
Verano	calido	Lluvioso	0.05
Verano	frio	Soleado	0.10
Verano	frio	Lluvioso	0.05
Invierno	calido	Soleado	0.10
Invierno	calido	Lluvioso	0.05
Invierno	frio	Soleado	0.15
Invierno	frio	Lluvioso	0.20

Para P(Clima | Invierno, Calido):

Identificacion de variables:

- Variables de consulta: Clima - C
- Variables de evidencia: Estacion - E = Invierno, T = Calido
- Variables de ocultas: Ninguna

P(C | Calido, Invierno) = P(Soleado | Calido,Invierno) , P(Lluvioso | Calido, Invierno)

$$P(C \mid Calido, Invierno) = \frac{P(C= Soleado, Cal\text{-}Invierno)}{P(Calido\text{-}Invierno)} , \frac{P(C= Lluvioso, Cal\text{-}Invierno)}{P(Calido\text{-}Invierno)}$$

Probabilidad total de Invierno (Marginalizacion):

P(Calido-Invierno) = 0.10 + 0.05 = 0.15

Probabilidades de cada clima en Invierno:

P(Soleado,Calido-Invierno) = 0.10 = 0.10

P(Lluvioso,Calido-Invierno) = 0.05 = 0.05

Estacion	Temp.	Clima	Prob.
Verano	calido	Soleado	0.30
Verano	calido	Lluvioso	0.05
Verano	frio	Soleado	0.10
Verano	frio	Lluvioso	0.05
Invierno	calido	Soleado	0.10
Invierno	calido	Lluvioso	0.05
Invierno	frio	Soleado	0.15
Invierno	frio	Lluvioso	0.20

Para P(Clima | E=Invierno, T=Calido):

$$P(C \mid E = \text{Invierno}) = \frac{P(C = \text{Soleado}, \text{Cal-Invierno})}{P(\text{Calido-Invierno})}, \frac{P(C = \text{Lluvioso}, \text{Cal-Invierno})}{P(\text{Calido-Invierno})}$$

Probabilidad total de Invierno (Marginalizacion):

P(Calido-Invierno) = 0.15

Probabilidades de cada clima en Invierno (Marginalizacion):

P(Soleado,Calido-Invierno) = 0.10
P(Lluvioso,Calido-Invierno) = 0.05

$$P(C \mid E = \text{Invierno}) = \frac{0.10}{0.15}, \frac{0.05}{0.15}$$

Estacion	Temp.	Clima	Prob.
Invierno	calido	Soleado	0.6667
Invierno	calido	Lluvioso	0.3333

Regla del Producto

Si sabemos que:

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

$$\downarrow$$

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

Regla del Producto

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

P(C)

Clima	Prob.
Soleado	0.8
Lluvioso	0.2

X

P(D | C)

Humedad	Clima	Prob.
Humedo	Soleado	0.10
Seco	Soleado	0.90
Humedo	Lluvioso	0.70
Seco	Lluvioso	0.30

P(D , C)

Humedad	Clima	Prob.
Humedo	Soleado	0.08
Seco	Soleado	0.72
Humedo	Lluvioso	0.14
Seco	Lluvioso	0.06

INFERENCIA

Regla de la Cadena

Es una herramienta en probabilidad y cálculo que permite descomponer probabilidades conjuntas complejas en multiplicaciones de probabilidades condicionales más simples.

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

Es decir, la probabilidad de que ocurran **A, B** y **C** juntos es igual a la probabilidad de A, multiplicada por la probabilidad de B dado A, multiplicada por la probabilidad de C dado A y B.

INFERENCIA

Regla de la Cadena

Ejemplo: Calcular la probabilidad de que usando la regla de la cadena:

1. Hoy esté nublado (N),
2. Luego llueva (L), y
3. Tu paraguas se moje (M).

$$P(N,L,M)=P(N)\cdot P(L \mid N)\cdot P(M \mid N,L)$$

- $P(N)$: Probabilidad de que esté nublado.
- $P(L|N)$: Probabilidad de lluvia si está nublado.
- $P(M|N, L)$: Probabilidad de que el paraguas se moje si está nublado y llueve.