# Inteligencia Artificial

## Programa de Ingeniería de Sistemas

Tema: Agentes basados en la Incertidumbre - Redes Bayesianas (Regla de la Cadena)

*LuzStellaRoblesPedrozo*

Universidad Tecnológica de Bolívar
CARTAGENA DE INDIAS

# Redes Bayesianas

## Modelación

Modelos

Representación simplificada de la realidad.

Características

- Suficientemente simple para no complicar el análisis del sistema.
- Suficientemente completo para englobar la mayor parte del sistema.

# Redes Bayesianas

## Modelación

Tipos de Modelos

Existen diferentes formas de representar los sistemas a través de modelos:

- Matemáticos (analíticos)
  A través de funciones matemáticas que describan la dinámica de los sistemas.

- Diagramas de Bloques
  Síntesis de las partes que componen al sistema.

- Grafos
  Diagramas de representación simbólica sobre las relaciones existentes en un sistema.

- Híbridos
  Varios tipos de modelos que en conjunto representan un sistema.

# Redes Bayesianas

## Algoritmos de Modelado

Englobados dentro de aprendizaje máquina (***Machine Learning***) permiten llevar a cabo problemas de aproximación y clasificación.

Clasificación

- Paramétricos
- Probabilísticos
- Multivariados
- Bio-inspirados

# Redes Bayesianas

## Algoritmos de Modelado

Englobados dentro de aprendizaje máquina (*Machine Learning*) permiten llevar a cabo problemas de aproximación y clasificación.

Clasificación

- Clasificador Bayesiano Naive
- **Redes Bayesianas**

# Redes Bayesianas

## Algoritmos de Modelado

Algunos algoritmos de modelado:

- Regresión
- Redes Neuronales Artificiales
- **Redes Bayesianas**
- Lógica Difusa
- Redes Orgánicas Artificiales

# Redes Bayesianas

## Introducción

Dado un conjunto de registros en una base de datos, donde cada registro pertenece a una clase, construir un modelo que permita predecir la clase de un nuevo registro.

# Redes Bayesianas

## Introducción

Redes Bayesianas se aplican en casos de incertidumbre a la hora de hacer cálculos de probabilidades condicionales y cálculo probabilidades desconocidas, dadas las condiciones específicas.

Aplicaciones: la bioinformática y la medicina, la ingeniería, clasificación de documentos, procesamiento de imágenes, la fusión de datos, y apoyar los sistemas de decisión, etc.

Ejemplos:

- Inferencia: P (Diagnóstico | Síntoma)
- Detección de anomalías: ¿Es esta observación anómala?
- Recolección de datos: ¿Cuál es la próxima prueba de diagnóstico dado en un conjunto de observaciones?

# Redes Bayesianas

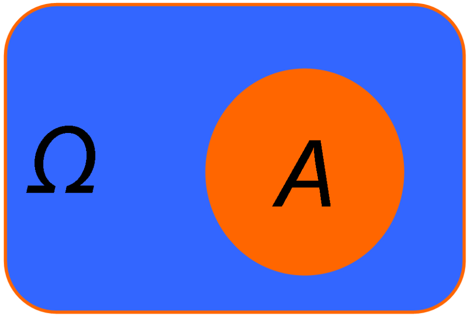
## La intuición detrás de Probabilidades

Intuitivamente, la probabilidad del evento A es igual a la porción de los resultados en donde A es cierto.

$\Omega$ es el conjunto de todos los resultados posibles. Su superficie es de $P(\Omega) = 1$

El conjunto de color naranja corresponde a los resultados en donde A es verdadera

P(A) = Área del óvalo naranja. Es evidente que $0 \leq P(A) \leq 1$



# Redes Bayesianas

## Axiomas de Probabilidad de Kolmogorov

*“La teoría de la probabilidad como una disciplina matemática, puede y debería ser desarrollada a partir de axiomas exactamente del mismo modo como la geometría y el algebra”.*

Andrey Nikolaevich Kolmogorov. Los fundamentos de la Teoría de la Probabilidad de 1933.

1. $P(A) \geq 0, \forall A \subset \Omega$
2. $P(\Omega) = 1$
3. $\sigma$-aditividad. Cualquier secuencia numerable de sucesos disjuntos dos a dos $A_1, A_2, \ldots P\left(\bigcup_i A_i\right) = \sum_i P(A_i)$

# Redes Bayesianas

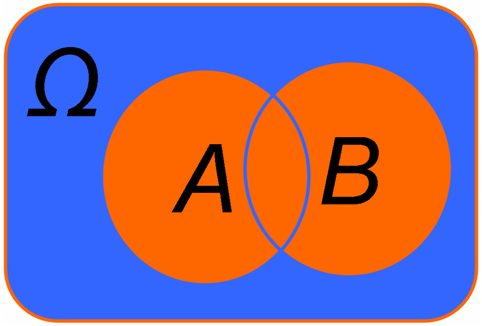
## Consecuencias de los Axiomas

$P(\overline{A}) = 1 - P(A)$. Donde $\overline{A} = \Omega - A$

$P(\varnothing) = 0$

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

$$P(A) = P(A \cap B) + P\left(A \cap \overline{B}\right)$$



# Redes Bayesianas

## Probabilidad Conjunta

Es la probabilidad de ocurrencia de dos o más eventos.

De la expresión

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

Al despejar *$P(A \cap B)$*, se obtiente: *$P(B|A)P(A)$*

Esta expresión es llamada Ley de multiplicación de probabilidades.

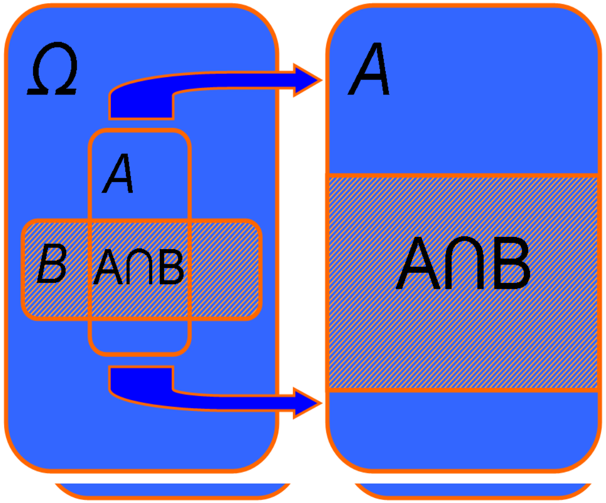
$P(A \cap B)$ recibe el nombre de Probabilidad Conjunta y corresponde a la probabilidad de que se presenten resultados comunes a los eventos A y B.

## Redes Bayesianas

### Probabilidad Condicional

P(B|A) = Proporción del espacio en el que B es verdad, dado que A también es cierto.

Definición formal de Probabilidad Condicional: $$P(B|A) = \frac{P(A \cap B)}{P(A)}$$



# Redes Bayesianas

## Independencia de Eventos

Definición

Dos eventos A y B son **independientes** si y solo si

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

Notación: I(A, B).

La independencia de A y B implica

$$\begin{aligned} P(A|B) &= P(A), \quad si \quad P(B) \neq 0 \\ P(B|A) &= P(B), \quad si \quad P(A) \neq 0 \end{aligned}$$

Universidad Tecnológica de Bolívar
CARTAGENA DE INDIAS

Acreditación Institucional de Alta Calidad

# Redes Bayesianas

## Independencia de Eventos

Tenemos: Suponiendo A y B son **independientes**:

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

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

Prueba:

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

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

La independencia de A y B implica

$$\begin{aligned} P(A|B) &= P(A), \quad si \quad P(B) \neq 0 \\ P(B|A) &= P(B), \quad si \quad P(A) \neq 0 \end{aligned}$$

# Redes Bayesianas

## Independencia Condicional

Definición

Dos eventos A y B son condicionalmente independientes dado un evento C, si y solo si

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

Notación: I(A, B|C).

La independencia condicional de A y B, dado un evento C, implica:

$$P(A|C, B) = P(A|C), \text{ si } P(B|C) \neq 0$$
$$P(B|C, A) = P(B|C), \text{ si } P(A|C) \neq 0$$

# Redes Bayesianas

## Independencia Condicional

Tenemos:

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

La independencia condicional de A y B, dado un evento C, implica

Prueba:

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

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

La independencia condicional de A y B, dado un evento C, implica:

$P(A|C, B) = P(A|C)$, si $P(B|C) \neq 0$

$P(B|C, A) = P(B|C)$, si $P(A|C) \neq 0$

# Redes Bayesianas

## Regla de Bayes

La definición de Probabilidad Condicional

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

Implica la Regla de la Cadena

$$\begin{aligned} P(A \cap B) &= P(A|B)P(B) \\ P(A \cap B) &= P(B|A)P(A) \end{aligned}$$

Regla de Bayes

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

# Redes Bayesianas

## Regla de Bayes

Desde el punto de vista bayesiano, todas las probabilidades son condicionales porque casi siempre existe algún conocimiento previo o experiencia acerca de los sucesos.

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

# Redes Bayesianas

## Regla de Bayes

### Ley de la Probabilidad Total

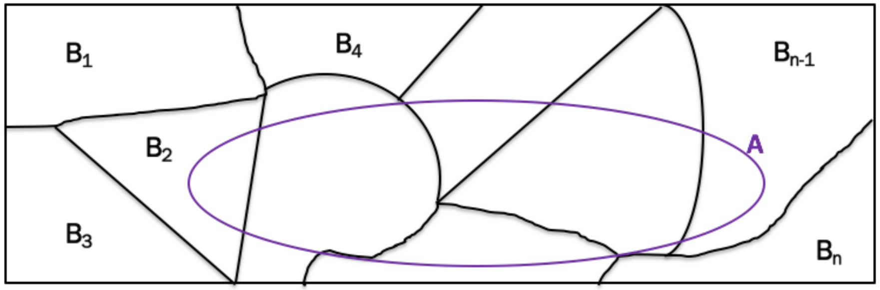
Para un suceso A y unas particiones $B_1,\ldots, B_k$, mutuamente excluyentes y exhaustivos entre si: $B_i \cap B_j = \varnothing$, con $i \neq j$.

Excluyentes: Si no hay dos con resultados comunes.
Exhaustivos: Si debe ocurrir un $B_i$, entonces $B_1 \cup B_2 \cup B_3 \cup \ldots \cup B_k = S$

Si tomamos dos eventos $B_i$ diferentes, su intersección da como resultado el conjunto vacío.

Adicionalmente, la unión de todos los eventos $B_i$, cubre el espacio de eventos, así:

$$\bigcup_{i=1}^{n} B_i = S$$



# Redes Bayesianas

## Regla de Bayes

### Ley de la Probabilidad Total

Para conocer el evento A a través de los eventos Bi, se tiene:

$A = (A \cap B_1) \cup (A \cap B_2) \cup (A \cap B_3) \cup \ldots \cup (A \cap B_{n-1}) \cup (A \cap B_n)$

A se define como la unión de las intersecciones del evento A con los eventos $B_i$, así:

$$A = \bigcup_{i=1}^{n} (A \cap B_i)$$

Por Probabilidad Conjunta, se tiene que: $P(A \cap B) = P(A|B)P(B)$

Y la probabilidad de A, como: $$P(A) = \sum_{i=1}^{n} P(A \cap B_i)$$

Con $k = n$

$$P(A) = \sum_{i=1}^{k} P(A|B_i)P(B_i)$$

# Redes Bayesianas

## Regla de Bayes

Permite obtener inferencias con mayor probabilidad asumiendo que se conoce información inicial. En términos de Bayes e inteligencia artificial, se define como:

probabilidad de verosimilitud

probabilidad *a priori*

probabilidad *a posterior* $\longrightarrow$

$$P(X|E) = \frac{P(E|X)P(X)}{P(E)}$$

probabilidad conjunta de E

Donde:

*X*: es el conjunto de eventos que se desean conocer
*E*: es el conjunto de evidencias

# Redes Bayesianas

## Regla de Bayes

Suponiendo que las particiones son sobre A:

Sea $A_1,\ldots, A_n$ una colección de k eventos mutuamente excluyentes y exhaustivos, con $P(A_i) > 0$ para $i = 1, 2, \ldots, k$. Entonces para cualquier otro evento B para el cual $P(B) > 0$,

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{k} P(B|A_i)P(A_i)} \quad j = 1, 2, \ldots, k$$

P(B) es una constante que puede ser calculada como: $P(B) = \sum_{i=1}^{k} P(B|A_i)P(A_i)$

$P(A_i)$ se puede estimar a partir del conjunto de entrenamiento (fracción de registros de entrenamiento que pertenecen a cada clase)

$P(B|A_i)$ es una tarea más difícil. Métodos:

- Clasificador Naive Bayes
- **Red Bayesiana**

# Redes Bayesianas

## Regla de Bayes

$P(B|A_i)$ Métodos para estimarlos:

- Clasificador Naive Bayes
- **Red Bayesiana**

Clasificador Naive Bayes, se asume que los atributos son condicionalmente independientes, dada una variable-clase *y*.

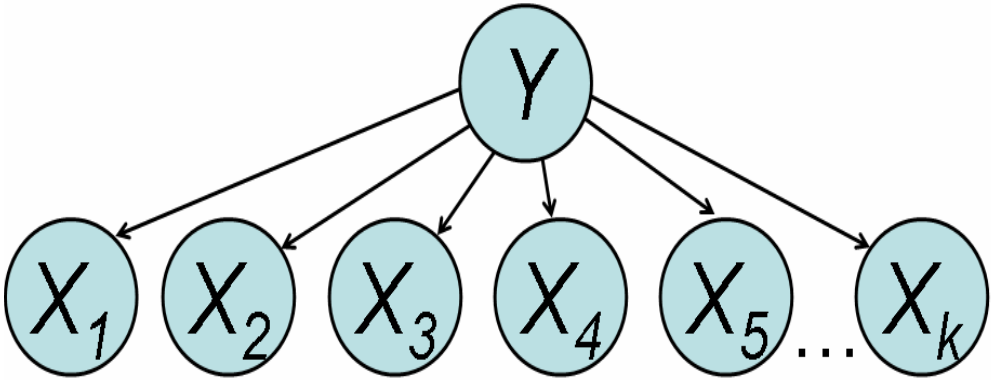
Redes Bayesianas, se tiene en cuenta la dependencia entre los atributos.

# Redes Bayesianas

## Regla de Bayes

Clasificador Naive Bayes, se asume que los atributos son condicionalmente independientes, dada una variable-clase *y*.

Esto puede ser representado usando una Red Bayesiana, como sigue:



# Redes Bayesianas

## Regla de Bayes

Reglas de la Probabilidad Total

Probabilidad Total Conjunta para una variable

$$P(C) = P(B \cap C) + P(B^c \cap C)$$
$$P(C) = P(B, C) + P(B^c, C)$$

Probabilidad Total Condicional para una variable

$$P(C) = P(C|B)\, P(B) + P(C|B^c)\, P(B^c)$$

Probabilidad Total Conjunta para dos variables

$$P(B, C) = P(A, B, C) + P(A^c, B, C)$$
$$P(B, C) = P(A \cap B \cap C) + P(A^c \cap B \cap C)$$

→ Regla de la Cadena

Probabilidad Total Condicional para dos variables

$$P(C|B) = P(C, B)/P(B) = P(C, A, B)/P(B) + P(C, A^c, B)/P(B)$$
$$= P(C, A|B) + P(C, A^c|B)$$

→ Definición de Probabilidad Condicional

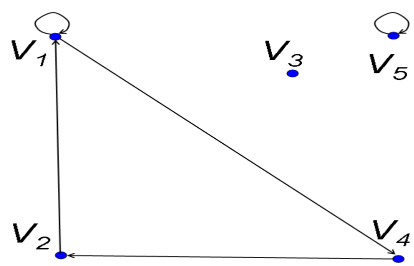
# Redes Bayesianas

## Grafo Dirigido

Definición

Un grafo dirigido o digrafo G es un par ordenado

$G := (V, A)$, donde $V$ es un conjunto cuyos elementos son llamados vértices. $A \subset V \times V$ es un conjunto de pares ordenados de vértices llamdos bordes dirigidos, arcos o flechas




$V = \{V_1, V_2, V_3, V_4, V_5\}$

$E = \{(V_1, V_1), (V_1, V_4),$
$(V_2, V_1), (V_4, V_2),$
$(V_5, V_5)\}$

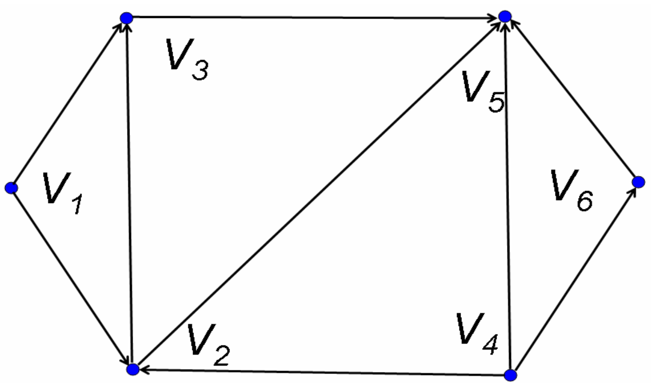
Cycle: $V_1 \rightarrow V_4 \rightarrow V_2$

# Redes Bayesianas

## Grafo Acíclico Dirigido

Definición

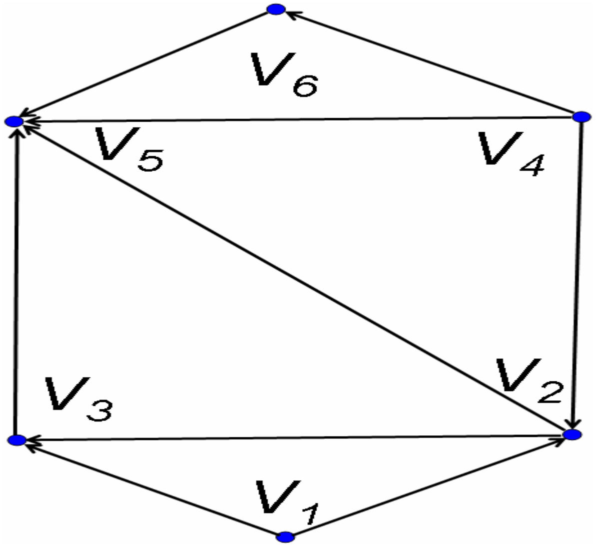
Un grafo dirigido acíclico digrafo (DAG) es un grafo dirigido sin ciclos dirigidos, es decir, para cualquier vértice $V_j$, no hay un camino dirigido que empieza en $V_j$ y termina en $V_j$.



# Redes Bayesianas

## Algunas descripciones en la Teoría de Grafos

- $V_1$ y $V_4$ son padres de $V_2$, pues $(V_1, V_2) \in E$, $(V_4, V_2) \in E$
- $V_5$, $V_3$ y $V_2$ son descencientes de $V_1$
- $V_4$ y $V_2$ son antepasados de $V_3$ existen caminos dirigidos de estos vertices a $V_3$
- $V_4$ y $V_6$ no son descendientes de $V_1$ no existen caminos dirigidos a estos vertices desde $V_1$



# Redes Bayesianas

## Definición de una Red Bayesiana

Una Red Bayesiana es:

- Es un Grafo Acíclico Dirigido (DAG) G, que registra la relación de dependencia entre un conjunto de variables.
- Una tabla de probabilidades que asocia cada nodo a sus nodos padres inmediatos.
- Cada nodo del grafo representa una variable (El grafo está formado por un conjunto de variables proposicionales, V).
- Un conjunto de relaciones binarias definidas sobre las variables de V, E.
- Cada arco confirma la relación de dependencia entre el par de variables.
- Una distribución de probabilidad conjunta sobre las variables de V.
- DAG satisface la Condición de Markov.

# Redes Bayesianas

## La Condición de Markov

Definición:

Suponga que se tiene una distribución de probabilidad P de variables aleatorias en un conjunto V, y un DAG G = (V, E). Se dice que (G, P) satisface la Condición de Markov si por cada variable $X \in V$, {X} (el conjunto de los padres directos de X), es condicionalmente independiente del conjunto de todos los otros nodos de la red -NDx- (o, lo que es lo mismo, separa condicionalmente a X de todo otro nodo Y de la red, que no sea X), (salvo sus descendientes), dados sus padres - PAx.

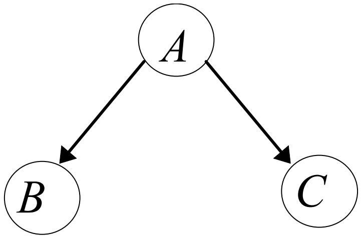
$$I(\{X\}, NDx|PAx)$$

La definición implica que el nodo raíz X, el cual no tiene padres, es incondicionalmente independiente de los otros nodos de la red.

# Redes Bayesianas

## Ejemplo 1

Consideremos la red dada en la siguiente figura:




$P(a_1) = 0.3$;  $P(b_1/a_1) = 0.4$;  $P(b_1/a_2) = 0.6$;  $P(c_1/a_1) = 0.7$;  $P(c_1/a_2) = 0.2$

$a_1 = A$
$a_2 = A^c$

En el que las variables que aparecen son binarias, junto con la siguiente distribución de probabilidad conjunta:

# Redes Bayesianas

## Ejemplo 1

En el que las variables que aparecen son binarias, junto con la siguiente distribución de probabilidad conjunta:

$P(a_1,b_1,c_1) = 0.084$

$P(a_1,b_2,c_1) = 0.126$

$P(a_2,b_1,c_1) = 0.084$

$P(a_2,b_1,c_2) = 0.336$

$P(a_1,b_1,c_2) = 0.036$

$P(a_1,b_2,c_2) = 0.054$

$P(a_2,b_2,c_1) = 0.056$

$P(a_2,b_2,c_2) = 0.224$

Es esto una Red Bayesiana?

# Redes Bayesianas

**Ejemplo 1**

Es esto una Red Bayesiana?

Solución:

En este caso, la condición de independencia condicional se satisface si C y B son independientes dado su padre común A.

Se debe probar entonces, que: $P(B|A, C) = P(B|A)$.

Para lo anterior, se debe probar que:

$P(b_1/a_1, c_1) = P(b_1/a_1)$ $P(b_2/a_1, c_1) = P(b_2/a_1)$

$P(b_1/a_1, c_2) = P(b_1/a_1)$ $P(b_2/a_1, c_2) = P(b_2/a_1)$

$P(b_1/a_2, c_1) = P(b_1/a_2)$ $P(b_2/a_2, c_1) = P(b_2/a_2)$

$P(b_1/a_2, c_2) = P(b_1/a_2)$ $P(b_2/a_2, c_2) = P(b_2/a_2)$

# Redes Bayesianas

**Ejemplo 1**

Es esto una Red Bayesiana?

$$P(B) = \sum_{i=1}^{k} P(B|A_i)P(A_i)$$

Solución:

A modo de ejemplo, comprobaríamos:

$P(b_1/a_1, c_2) = P(b_1/a_1)$

1. $P(b_1/a_1, c_2)$
2. $P(b_1/a_1)$

1. $$\boldsymbol{P(b_1|a_1, c_2)} = \frac{P(b_1, a_1, c_2)}{P(a_1, c_2)} = \frac{P(a_1, b_1, c_2)}{P(a_1, c_2)} =$$

$$\frac{0{,}036}{P(a_1, b_1, c_2) + P(a_1, b_2, c_2)} = \frac{0{,}036}{0{,}036 + 0{,}054} = \mathbf{0{,}4}$$

# Redes Bayesianas

**Ejemplo 1**

Solución:

$$P(B) = \sum_{i=1}^{k} P(B|A_i)P(A_i)$$

$P(b_1/a_1, c_2) = P(b_1/a_1)$

1. $P(b_1/a_1, c_2)$
2. $P(b_1/a_1)$

2. $$\boldsymbol{P(b_1|a_1)} = \frac{P(b_1, a_1)}{P(a_1)} = \frac{P(a_1, b_1, c_1) + P(a_1, b_1, c_2)}{P(a_1, b_1) + P(a_1, b_2)}$$

$$= \frac{P(a_1, b_1, c_1) + P(a_1, b_1, c_2)}{P(a_1, b_1, c_1) + P(a_1, b_1, c_2) + P(a_1, b_2, c_1) + P(a_1, b_2, c_2)}$$

$$= \frac{0{,}084 + 0{,}036}{0{,}084 + 0{,}036 + 0{,}126 + 0{,}054} = \frac{0{,}12}{0{,}3} = \mathbf{0,4}$$

Realizando las comprobaciones restantes, se vería que en este caso sí tenemos una Red Bayesiana!

# Redes Bayesianas

## Definición

Si se cumplen las condiciones de independencia condicional, a partir de las probabilidades condicionales, es posible calcular la distribución conjunta, así, por ejemplo:

Para un grafo G $\quad A \longrightarrow B \longrightarrow C$

$P(B, C) = P(B).P(C|B)$

$P(B, C) = P(A, B, C) + P(A^c, B, C) =$

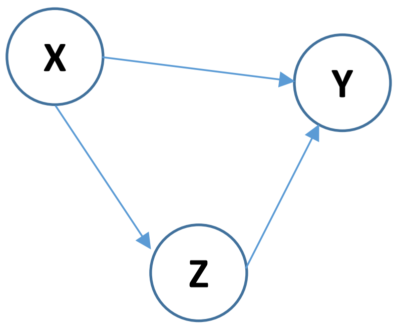
$P(A, B, C) = P(A).P(B|A).P(C|A, B)$

$P(A^c, B, C) = P(A^c).P(B|A^c).P(C|A^c, B)$

# Redes Bayesianas

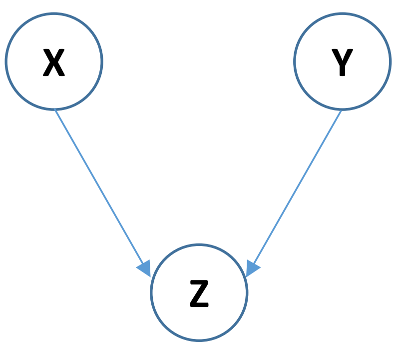
## Ejemplo 2

Representación de Grafos




P(X, Y, Z) = ?

P(X, Y, Z) = P(X).P(Z|X).P(Y|X, Z)

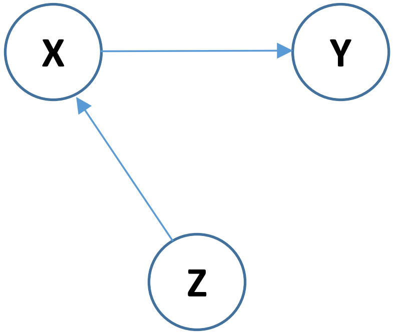



P(X, Y, Z) = ?

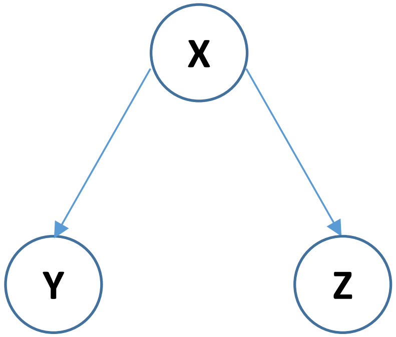
# Redes Bayesianas

## Ejemplo 2

Representación de Grafos




P(X, Y, Z) = ?




P(X, Y, Z) = ?

# Referencias

Material de apoyo de la semana registrados en SAVIO (Diapositivas):


Bayesian networks without tears: making Bayesian networks more accessible to the probabilistically unsophisticated. Eugene Charniak

AI Magazine Volume 12 , Issue 4 (Winter 1991) Pages: 50 - 63 Year of Publication: 1991

Learning Bayesian Networks. Richard E. Neapolitan. Prentice Hall

# Gracias!