#### Fundamentos de Inferencia Bayesiana

#### Modelos Gráficos e Inferencia Exacta

# Modelos Gráficos Probabilísticos (*PGMs, Pearl* 1998)

- Los nodos son variables, las aristas codifican dependencias. Cuidado: ¡No es *causalidad*!
- En particular: Directed Acyclic Graphs (DAGs)
- Nodos *observables* y *latentes*
- Notación de *placas*: variables repetidas
- Convenciones frecuentes: gris: observable, blanco: latente. circular: continuo, cuadrado: discreto.





"Inversa": No todas las independencias pueden ser capturadas en un DAG

Semántica: independencia condicional... Las flechas no necesariamente indican causalidad



#### Inferencia Exacta



Llama el vecino diciendo que sonó la alarma... ¿Hubo un ladrón?

En la radio reportan que hay un terremoto... "Ah, era eso"

#### Inferencia Exacta



P(b, e, a, p, r) = P(r|p, a, e, b)P(p|a, e, b)P(a|e, b)P(e|b)P(b)  $\downarrow$  P(b, e, a, p, r) = P(r|e)P(p|a)P(a|e, b)P(e)P(b)  $\downarrow$ Pero por qué?  $\downarrow$ Cuál es el problema?

#### Midiendo el espacio...

P(b, e, a, p, r) = P(r|p, a, e, b)P(p|a, e, b)P(a|e, b)P(e|b)P(b)

31 = 16 + 8 + 4 + 2 + 1

#### P(b, e, a, p, r) = P(r|e)P(p|a)P(a|e, b)P(e)P(b)2 + 2 + 4 + 1 + 1 = 10

Redujimos el tamaño del problema estableciendo las (in)dependencias

## Curse of dimensionality

El espacio de hipótesis crece exponencialmente con el número de variables.. Imposible explorarlo

# Blessing of abstraction

Si imponemos estructura, reducimos la dimensionalidad efectiva

**Structure x Statistics** 

#### Poniendo números

P(b, e, a, p, r) = P(r|e)P(p|a)P(a|e, b)P(e)P(b)

 $\begin{array}{ll} P(b=1)=\beta, & P(b=0)=1-\beta & \beta=0.001 & (\text{un robo cada} \\ P(e=1)=\epsilon, & P(e=0)=1-\epsilon & \epsilon=0.001 & \text{tres años}) \end{array}$ 

 $P(a=0 | b=0, e=0) = (1-f), \qquad P(a=1 | b=0, e=0) = f$   $P(a=0 | b=1, e=0) = (1-f)(1-\alpha_b), \qquad P(a=1 | b=1, e=0) = 1-(1-f)(1-\alpha_b)$   $P(a=0 | b=1, e=1) = (1-f)(1-\alpha_e), \qquad P(a=1 | b=0, e=1) = 1-(1-f)(1-\alpha_e)$   $P(a=0 | b=1, e=1) = (1-f)(1-\alpha_b)(1-\alpha_e), \qquad P(a=1 | b=1, e=1) = 1-(1-f)(1-\alpha_b)(1-\alpha_e)$   $f = 0.001 \qquad \alpha_b = 0.99 \qquad \alpha_e = 0.01 \qquad \bigvee$  P(p=1 | a=0) = 0 P(a=0 | b=0, e=1) = 0

P(p=1 | a=1) no hace falta, ya que p=1 => a=1

$$P(r=1 | e=0) = 0$$
  
 $P(r=1 | e=1)$  no hace falta, ya que  $e=1 => r=1$ 

Luego del llamado...

$$P(b, e \mid a = 1) = \frac{P(a = 1 \mid b, e)P(b)P(e)}{P(a = 1)} \qquad P(b = 1, e = 0 \mid a = 1) = 0.4947$$
$$P(b = 0, e = 1 \mid a = 1) = 0.0055$$
$$P(b = 1, e = 1 \mid a = 1) = 0.0005$$

¿Cuál es la probabilidad de que haya entrado un ladrón? Marginalizamos...

$$\begin{array}{rcl} P(b=0 \mid a=1) &=& P(b=0, \ e=0 \mid a=1) + P(b=0, \ e=1 \mid a=1) &=& 0.505 \\ P(b=1 \mid a=1) &=& P(b=1, \ e=0 \mid a=1) + P(b=1, \ e=1 \mid a=1) &=& 0.495 \end{array}$$

Luego del reporte en la radio...

$$P(b \mid e = 1, a = 1) = P(b, e = 1 \mid a = 1) / P(e = 1 \mid a = 1)$$

$$P(b=0 | e=1, a=1) = 0.92$$
  

$$P(b=1 | e=1, a=1) = 0.08$$

"Explaining away" Una de las posibles causas de un efecto se vuelve *menos* probable cuando otra causa se vuelve *más* probable

P(b=0, e=0 | a=1) =

0.4993

### Patrones de inferencia

- Explaining away: una de las posibles causas de un efecto se vuelve menos probable cuando otra causa se vuelve más probable
- Screening off: dos variables dependientes sólo a través de una tercera variable se vuelven independientes condicionadas a la observación de esa variable (en causa común como en cadenas)
- Non-monotonic reasoning: agregar una hipótesis puede reducir el número de conclusiones a las que se puede llegar (imposible en primer orden). Ejemplo: Tweety ave, Tweety avestruz

#### De vuelta a la binomial.. ahora con n desconocido

 $k \sim \text{Binomial}(\theta, n)$  $\theta \sim \text{Uniforme}(0, 1)$  $n \sim \text{Uniforme}(1, N_{max})$ 



Ejemplo con varias observaciones (encuesta)

 $k_i \sim \text{Binomial}(\theta, n)$  $\theta \sim \text{Uniforme}(0, 1)$  $n \sim \text{Uniforme}(1, N_{max})$ 



 $k = \{16, 18, 22, 25, 27\}$ 



¡Atención a la conjunta! (problemático con muchas variables..)

# Práctica



Muestre cómo se simplifica la probabilidad conjunta usando las independencias condicionales codificadas en el DAG.

Escriba primero una expresión general sin usar las independencias, y luego otra utilizándolas.

Cuente el número de entradas en cada expresión asumiendo variables *ternarias.* 

Wagenmakers & Lee: 3.4, 3.5