Fundamentos de Inferencia Bayesiana

Modelos jerárquicos

MODELO

Creencias

prior

Mecanismo

likelihood

variables ***latentes***

DATOS

Explicación

posterior

Predicción

posterior predictive

variables ***observables***

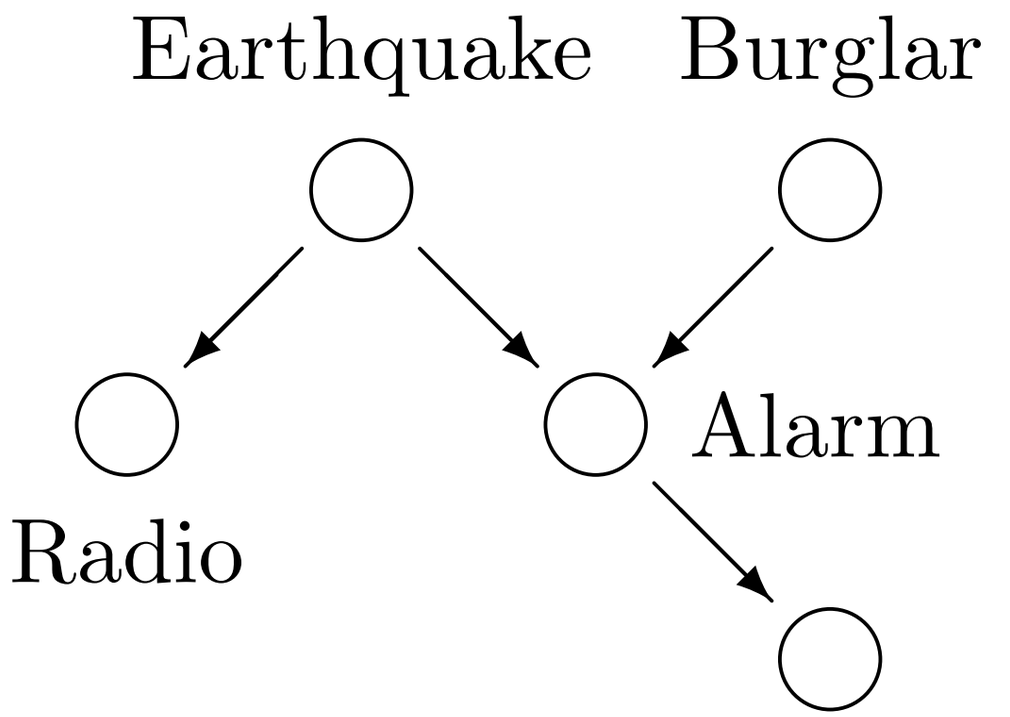
Pero.. ¿Cuál es el problema?

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

Curse of dimensionality

Blessing of abstraction

Inferencia Exacta



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

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

MODELO

Creencias

prior

Mecanismo

likelihood

imponer más y más estructura

variables *latentes*

DATOS

variables *observables*

Explicación

posterior

Predicción

posterior predictive

Modelos Jerárquicos

Limitaciones:

-Si tiramos 25 veces la moneda y sale 25 veces cara.. raro

...Pero con el prior de 100 y 100 que usamos obtenemos:

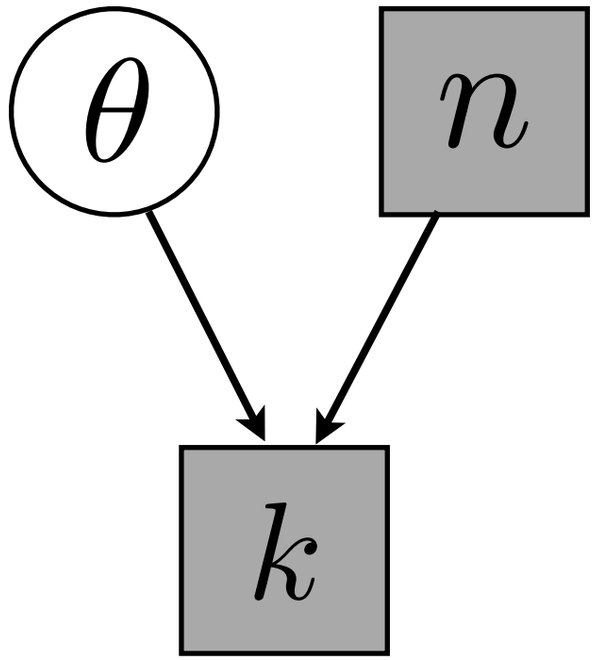
$$p(0|25,25) = 125/225 \simeq 0.56$$ ¡no tan raro!

Modelo

likelihood:

$$p(k|\theta) = \binom{n}{k} \theta^k (1-\theta)^{n-k}$$

$$k \sim \text{Binomial}(\theta, n)$$



prior:

$$\theta \sim \text{Uniform}(0,1) = \text{Beta}(1,1)$$

$$\theta \sim \text{Beta}(100,100)$$

posterior:

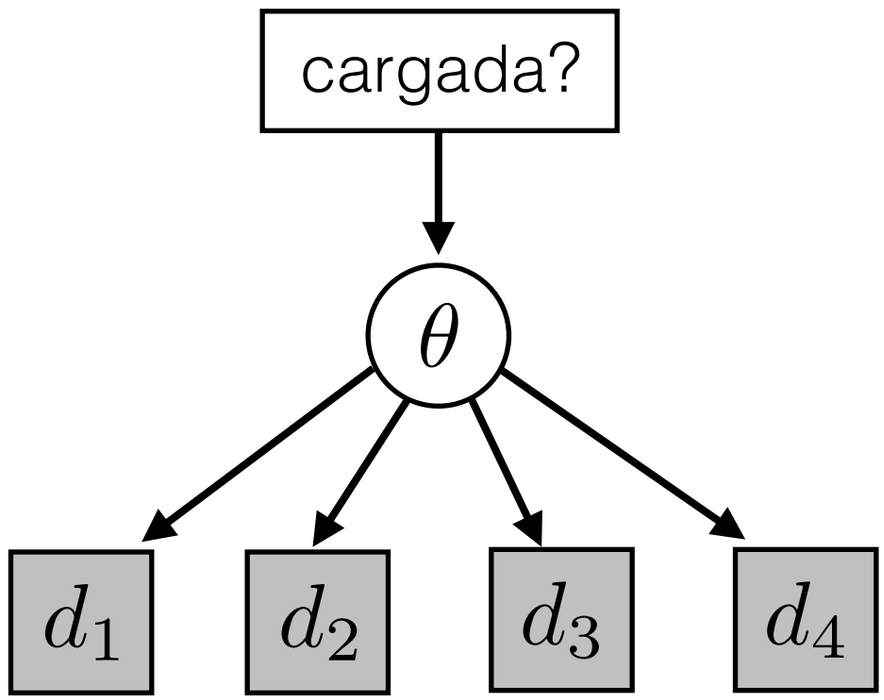
$$p(\theta|D) = \text{Beta}(k+1, n-k+1)$$

$$p(\theta|D) = \text{Beta}(k+100, n-k+100)$$

Modelos Jerárquicos

prior débil:
poco razonable

$$\text{Beta}(1, 1)$$



prior fuerte:
¡poco flexible!

$$\text{Beta}(100, 100)$$

Hipótesis de más alto nivel: *overhypothesis*, o teoría

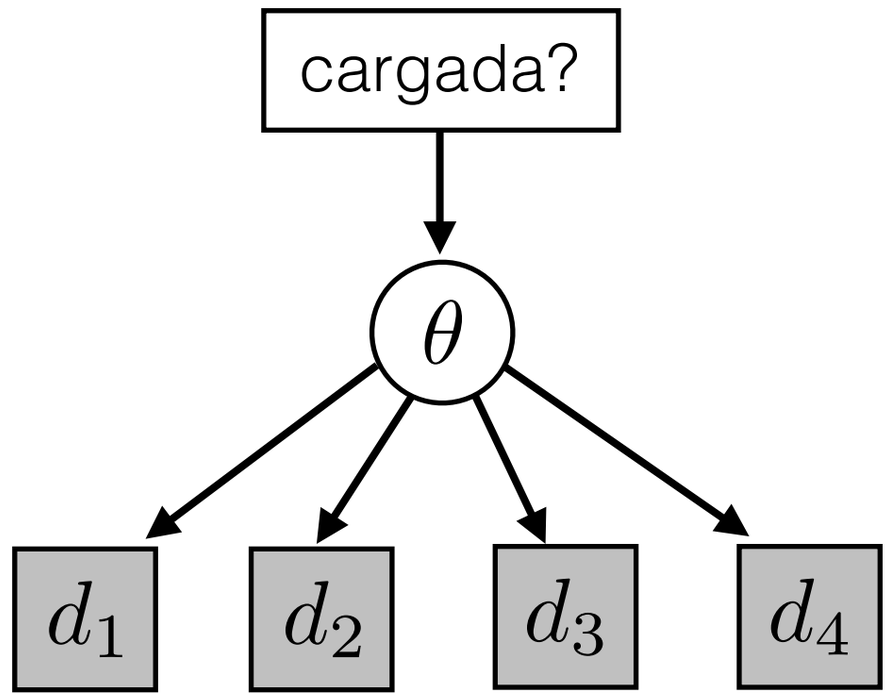
Modelo jerárquico

$$k \sim \text{Binomial}(\theta, n)$$

$$p(\theta|\text{no cargada}) = \text{Beta}(100, 100)$$

$$p(\theta|\text{cargada}) = \text{Beta}(1, 1)$$

$$p(\text{cargada}) = 0.01$$



La información de los datos se “propaga” hacia arriba

Hacemos inferencia sobre la teoría... $p(\text{cargada}|25 \text{ caras})$

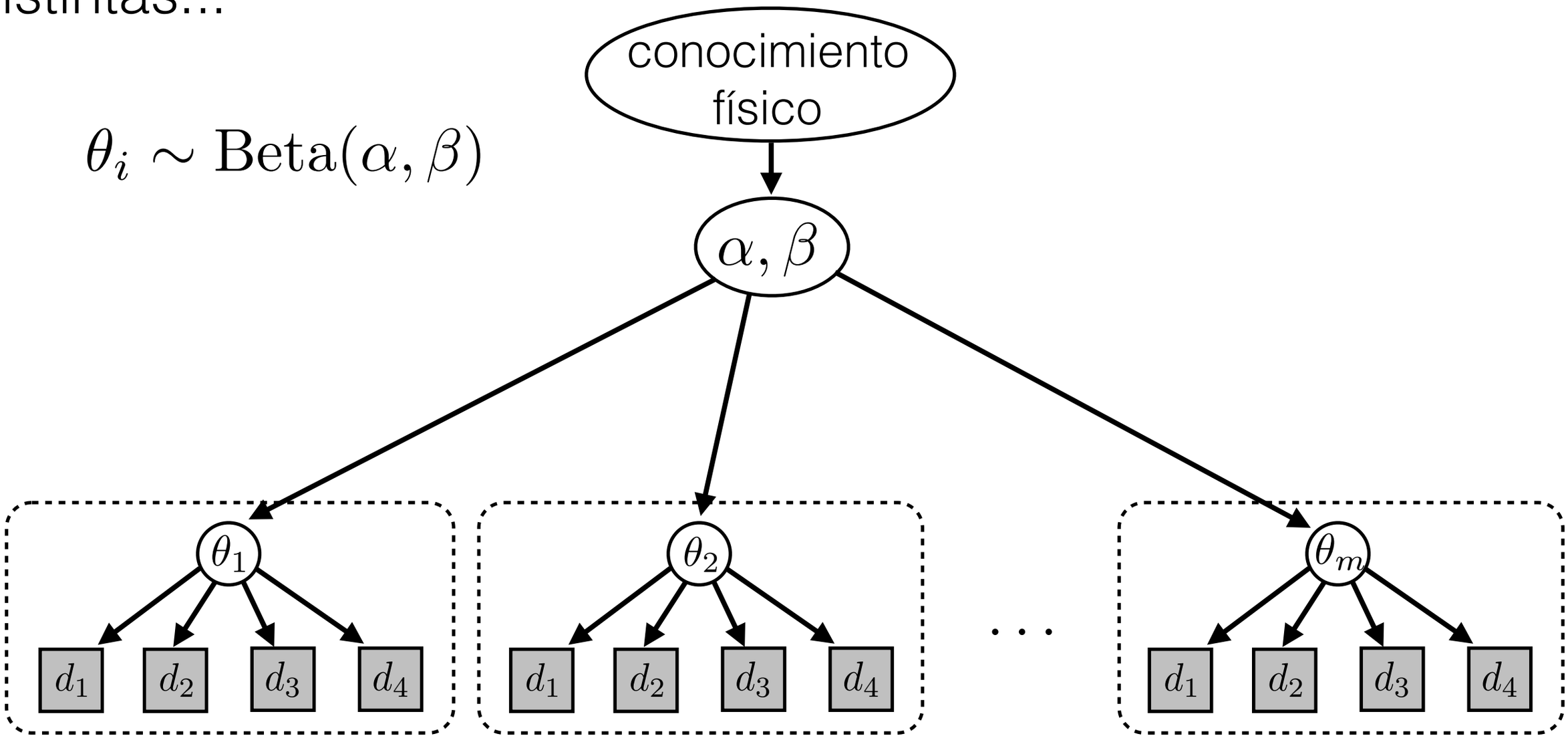
$$\frac{p(\text{cargada}|25\text{ caras})}{p(\text{no cargada}|25\text{ caras})} = \frac{p(25\text{ caras}|\text{cargada})}{p(25\text{ caras}|\text{no cargada})}\frac{p(\text{cargada})}{p(\text{no cargada})}$$

$$p(25\text{ caras}|\text{cargada}) = \int p(25\text{ caras}|\theta)p(\theta|\text{cargada})d\theta$$

$$\frac{p(\text{cargada}|25\text{ caras})}{p(\text{no cargada}|25\text{ caras})} \simeq \frac{0.038}{1.14\times 10^{-7}}\frac{0.01}{0.99} \simeq 3367$$

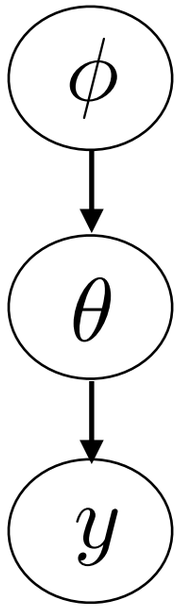
...No es lo mismo ver 200 tiradas de una moneda que 20 de 10 distintas...

$$\theta_i \sim \mathrm{Beta}(\alpha, \beta)$$



Inferencia sobre los *hiperparámetros..*
caracterizan la población de monedas

Modelos jerárquicos



$p(\phi)$

$p(\theta|\phi)$

$p(y|\theta)$

$\phi \rightarrow$ hiperparámetros

$\theta \rightarrow$ parámetros

$y \rightarrow$ datos

$p(\phi, \theta) \rightarrow prior$ conjunto

$p(\phi, \theta|y) \rightarrow posterior$ conjunto

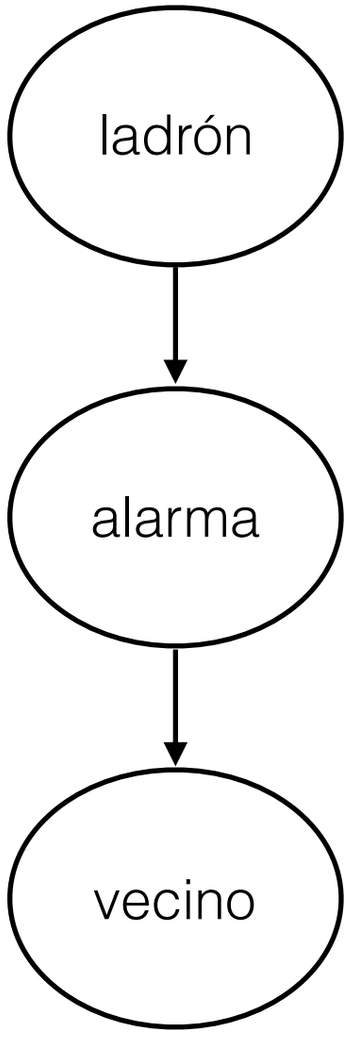
Bayes:

$$p(\phi, \theta|y) \propto p(y|\theta)p(\theta|\phi)p(\phi)$$

Nos pueden interesar tanto los parámetros como los hiperparámetros para la *posterior,* y tanto los parámetros como los datos para la *predicción*

$$p(\theta|\phi, y) \neq p(\theta|\phi)$$

Implementando inferencia exacta



$$p(l = 1) = 0.001$$

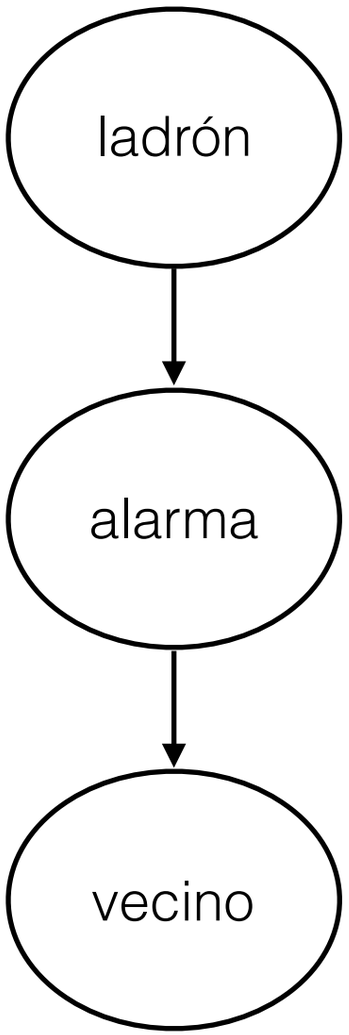
$$p(a = 1|l = 1) = 0.99$$
$$p(a = 1|l = 0) = 0.05$$

$$p(v = 1|a = 1) = 0.85$$
$$p(v = 1|a = 0) = 0.02$$

$$p(l = 1|v = 1) = ?$$

$$p(l = 1|v = 1) = \frac{p(v = 1|l = 1)p(l = 1)}{p(v = 1)} = \frac{\sum_a p(v = 1|a)p(a|l = 1)p(l = 1)}{\sum_l [\sum_a p(v = 1|a)p(a|l)p(l)]}$$

Implementando inferencia exacta



$$p(a = 1|l = 1) = 0.99$$

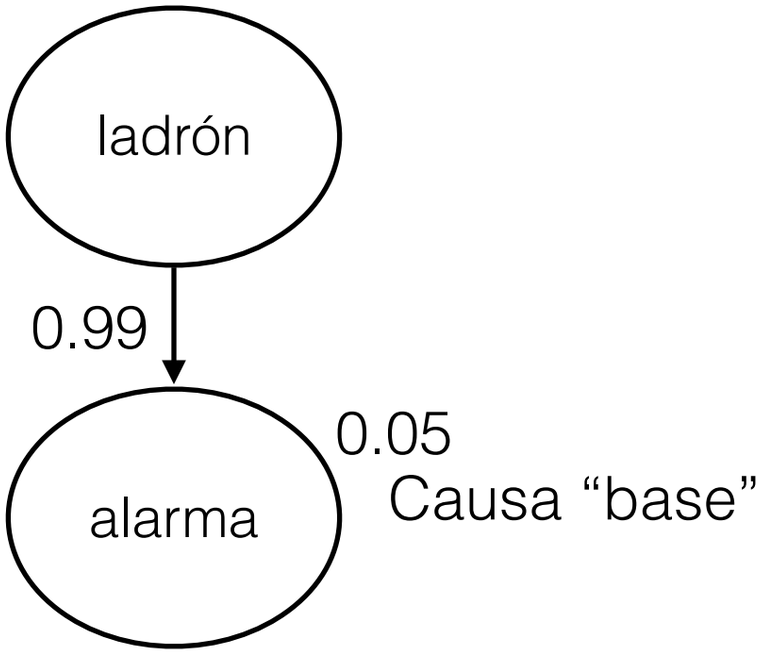
```
def p_vecino_ladron(v,l):
    if l==1:
        if v==1:
            …
```

$$p(l = 1|v = 1) = \frac{\sum_a p(v = 1|a)p(a|l = 1)p(l = 1)}{\sum_l [\sum_a p(v = 1|a)p(a|l)p(l)]}$$

```
def p_ladron_vecino(v,l):
    numerador = sum(…)
    denominador = sum(…)
    …
```

$$p(l = 1|v = 1) = 0.019$$

Noisy-OR



$$p(a = 1|l = 0) = 0.05$$

~~$p(a = 1|l = 1) = 0.99$~~

¡Pueden suceder las dos!

$$p(a = 0|l = 1) = (1 - 0.05) * (1 - 0.99)$$

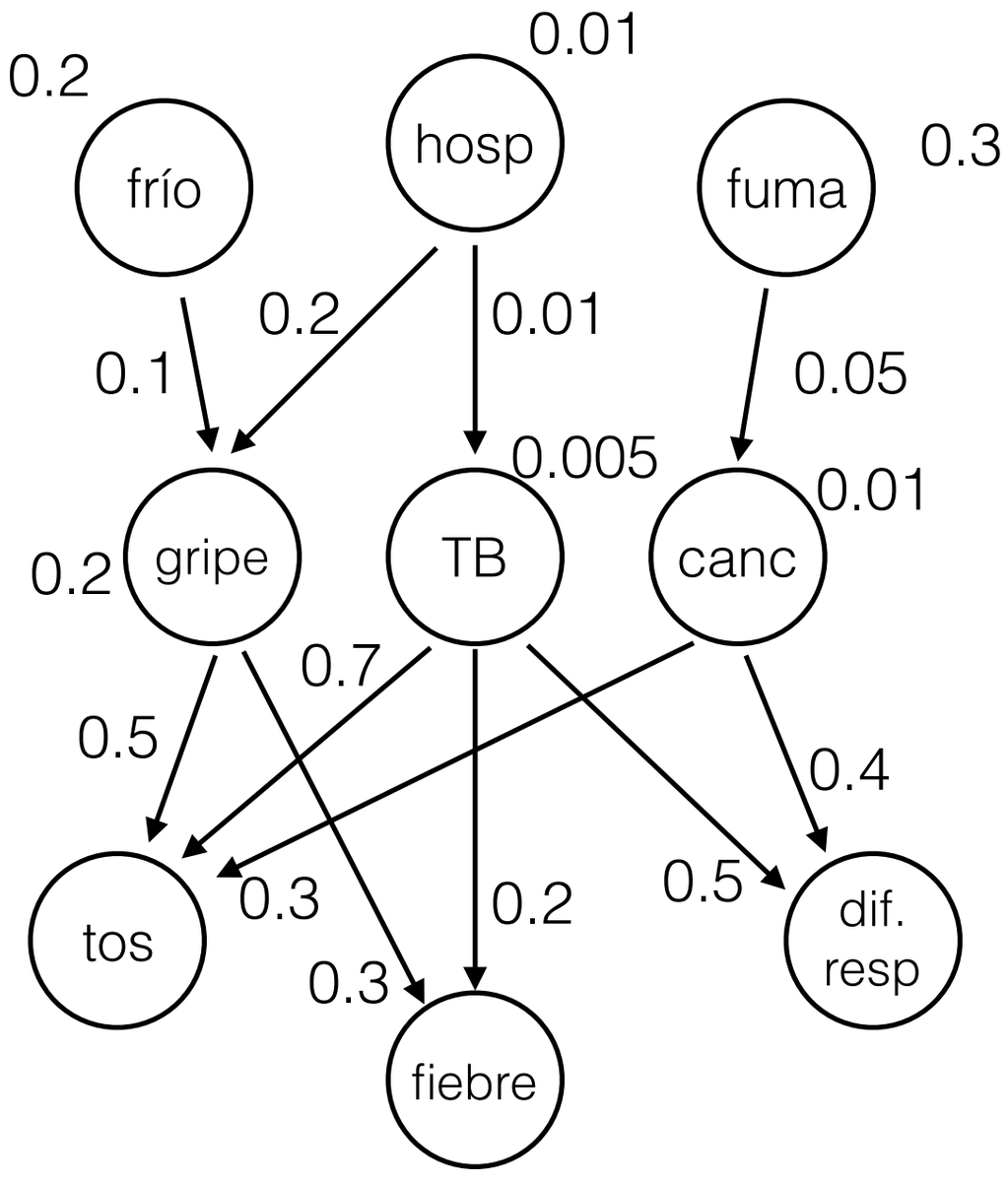
$$p(a = 1|l = 1) = 1 - (1 - 0.05) * (1 - 0.99)$$

$$p(a = 1|l = 1) = 0.9905$$

$$p(X = 1|Padre_1, \ldots, Padre_n) = 1 - \prod_i falla_i$$

Práctica - ¡para entregar!

Implementar inferencia exacta



Implementar el modelo usando Noisy-ORs para codificar las probabilidades.

Escribir cuidadosamente las probabilidades de cada nivel en función del nivel anterior.

Inferir las probabilidades de TB y cáncer dado que los tres síntomas están presentes. ¿Dónde entran los priors?

Usar un lenguaje “común”!!!.. y, *luego,* WebPPL