


Técnicas para calcular distribuciones de creencias

Estimación de habilidad en la industria del video juego

Gustavo Landfried

@GALandfried 

Licenciado en Ciencias Antropológicas
Doctorando en Ciencias de la Computación



DEPARTAMENTO
DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

Incertidumbre



Sorpresa



Sorpresa

El punto débil de las creencias



¿Cómo estimar habilidades?

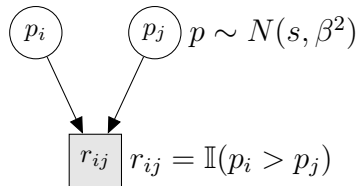


Arpad Elo

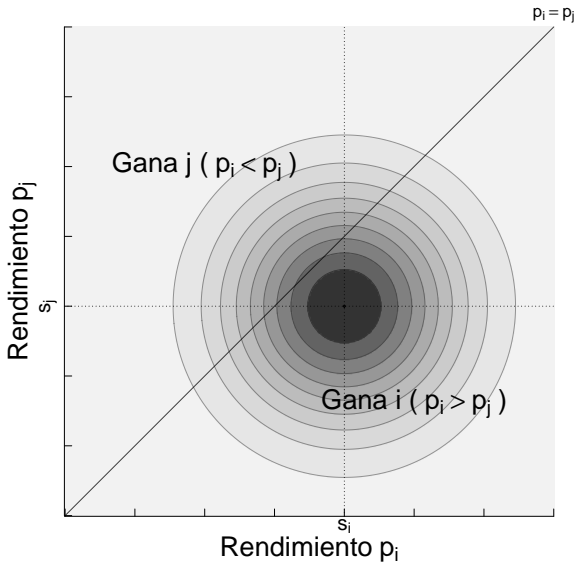
Modelo Elo

Rendimiento aleatorio oculto (p)
centrado en la habilidad estimada (s)

Resultado observado (r)



Probabilidad de ganar (gráfica)



Estimación Elo

$$s_i^{\text{new}} = s_i^{\text{old}} + K\Delta$$

Estimación Elo

$$s_i^{\text{new}} = s_i^{\text{old}} + K\Delta$$

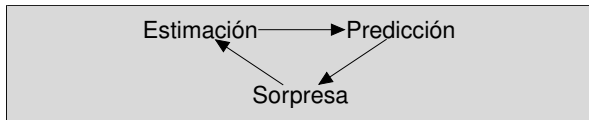
$$\Delta = \underbrace{(2r_{ij} - 1)}_{\text{Signo del resultado}} \underbrace{(1 - P(r_{ij}|s_i, s_j))}_{\text{Sorpresas del resultado}}$$

Estimación Elo

$$s_i^{\text{new}} = s_i^{\text{old}} + K\Delta$$

$$\Delta = \underbrace{(2r_{ij} - 1)}_{\text{Signo del resultado}} \underbrace{(1 - P(r_{ij}|s_i, s_j))}_{\text{Sorpresa del resultado}}$$

Modelo de solución



Hoy

Estimación de habilidad en la industria del video juego



Hoy

Estimación de habilidad en la industria del video juego



Inferencia Bayesiana

¿Por qué inferencia Bayesiana?

¿Por qué inferencia Bayesiana?

Permite computar creencias óptimas
dadas restricciones: modelos y datos

Principio de máxima incertidumbre



Principio de máxima incertidumbre

10 %



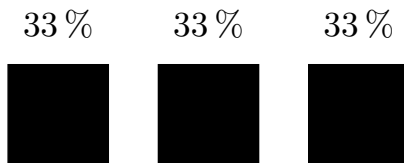
80 %



10 %



Principio de máxima incertidumbre



Principio de máxima incertidumbre Honestidad

33 %



33 %



33 %



Principio de máxima incertidumbre

Honestidad



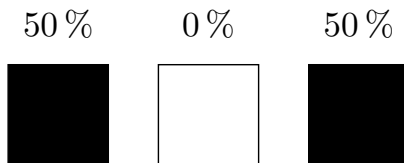
Principio de máxima incertidumbre

Honestidad

0 %



Principio de máxima incertidumbre Honestidad



Principio de máxima incertidumbre Honestidad

33 %



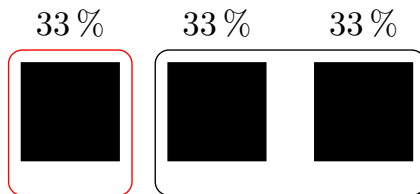
33 %



33 %

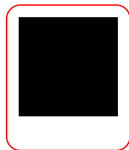


Principio de máxima incertidumbre Honestidad

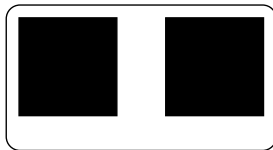


Principio de máxima incertidumbre Honestidad

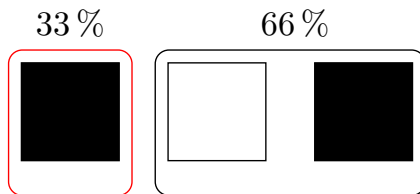
33 %



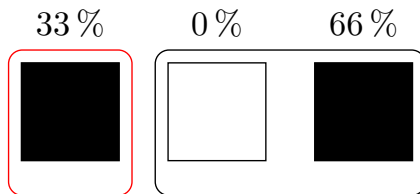
66 %



Principio de máxima incertidumbre Honestidad



Principio de máxima incertidumbre Honestidad



La sorpresa fuente de información

Entropía

La sorpresa fuente de información

Entropía

Información extraída de la sorpresa,

$$h(x) = -\log P(x)$$

La sorpresa fuente de información

Entropía

Información extraída de la sorpresa,

$$h(x) = -\log P(x)$$

Información esperada,

$$H(X) = \sum_x P(x)h(x)$$

La sorpresa fuente de información

Entropía

Información extraída de la sorpresa,

$$h(x) = -\log P(x)$$

Información esperada,

$$H(X) = \sum_x P(x)h(x)$$

Honestidad: máxima entropía

Máxima información esperada \Leftrightarrow Máxima incertidumbre

La sorpresa fuente de información

Entropía

Información extraída de la sorpresa,

$$h(x) = -\log P(x)$$

Información esperada,

$$H(X) = \sum_x P(x)h(x)$$

Honestidad: máxima entropía

Máxima información esperada \Leftrightarrow Máxima incertidumbre

Máxima entropía ante estados igualmente posibles



Máxima entropía ante estados igualmente posibles



$$\text{Creencia}(A) = \frac{\text{Cantidad de estados en que } A \text{ es verdadera}}{\text{Cantidad de estados totales}}$$

Máxima entropía ante estados no igualmente posibles

	y_j			
x_i		n_{ij}		f_i
	c_j			N

Máxima entropía ante estados no igualmente posibles

	y_j			
x_i		n_{ij}		f_i
	c_j			N

Creencia conjunta: $C(X = x_i, Y = y_j) =$

Máxima entropía ante estados no igualmente posibles

	y_j			
x_i		n_{ij}		f_i
	c_j			N

Creencia conjunta: $C(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$

Máxima entropía ante estados no igualmente posibles

	y_j			
x_i		n_{ij}		f_i
	c_j			N

Creencia conjunta: $C(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$

Creencia marginal: $C(X = x_i) =$

Máxima entropía ante estados no igualmente posibles

	y_j			
x_i		n_{ij}		f_i
	c_j			N

Creencia conjunta: $C(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$

Creencia marginal: $C(X = x_i) = \frac{f_i}{N}$

Máxima entropía ante estados no igualmente posibles

	y_j			
x_i		n_{ij}		f_i
	c_j			N

Creencia conjunta: $C(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$

Creencia marginal: $C(X = x_i) = \frac{f_i}{N} = \frac{\sum_j n_{ij}}{N}$

Máxima entropía ante estados no igualmente posibles

	y_j			
x_i		n_{ij}		f_i
	c_j			N

Creencia conjunta: $C(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$

Creencia marginal: $C(X = x_i) = \frac{f_i}{N} = \sum_j C(X = x_i, Y = y_j)$

Máxima entropía ante estados no igualmente posibles

	y_j			
x_i		n_{ij}		f_i
	c_j			N

Creencia conjunta: $C(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$

Creencia marginal: $C(X = x_i) = \frac{f_i}{N} = \sum_j C(X = x_i, Y = y_j)$

Creencia condicional: $C(Y = y_j | X = x_i) =$

Máxima entropía ante estados no igualmente posibles

	y_j			
x_i		n_{ij}		f_i
	c_j			N

Creencia conjunta: $C(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$

Creencia marginal: $C(X = x_i) = \frac{f_i}{N} = \sum_j C(X = x_i, Y = y_j)$

Creencia condicional: $C(Y = y_j | X = x_i) = \frac{n_{ij}}{f_i}$

Máxima entropía ante estados no igualmente posibles

	y_j			
x_i		n_{ij}		f_i
	c_j			N

Creencia conjunta: $C(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$

Creencia marginal: $C(X = x_i) = \frac{f_i}{N} = \sum_j C(X = x_i, Y = y_j)$

Creencia condicional: $C(Y = y_j | X = x_i) = \frac{n_{ij}}{N} \frac{N}{f_i}$

Máxima entropía ante estados no igualmente posibles

	y_j			
x_i		n_{ij}		f_i
	c_j			N

Creencia conjunta: $C(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$

Creencia marginal: $C(X = x_i) = \frac{f_i}{N} = \sum_j C(X = x_i, Y = y_j)$

Creencia condicional: $C(Y = y_j | X = x_i) = \frac{n_{ij}}{N} \frac{N}{f_i} = \frac{C(X=x_i, Y=y_j)}{C(X=x_i)}$

Las reglas de la probabilidad

$$\text{Marginal}_i = \sum_j \text{Conjunta}_{ij}$$

$$\text{Condicional}_{j|i} = \frac{\text{Conjunta}_{ij}}{\text{Marginal}_i}$$

Las reglas de la probabilidad

$$\text{Marginal}_i = \sum_j \text{Conjunta}_{ij}$$

$$\text{Condicional}_{j|i} = \frac{\text{Conjunta}_{ij}}{\text{Marginal}_i}$$

Regla de la suma

$$P(X) = \sum_Y P(X, Y)$$

Cualquier distribución marginal puede ser obtenida integrando la distribución conjunta

Regla del producto

$$P(X, Y) = P(Y|X)P(X)$$

Cualquier distribución conjunta puede ser expresada como el producto de distribuciones condicionales unidimensionales.

Las reglas de la probabilidad

Teorema de Cox

Son las reglas del razonamiento con incertidumbre

Las únicas que garantizan:

- Representación de las creencias con valores reales
- Consistencia, cualquier camino lleva a la misma conclusión
- Actualización de las creencias en la dirección de la evidencia

Las reglas de la probabilidad

Teorema de Cox

Son las reglas del razonamiento con incertidumbre

Las únicas que garantizan:

- Representación de las creencias con valores reales
- Consistencia, cualquier camino lleva a la misma conclusión
- Actualización de las creencias en la dirección de la evidencia

Teorema de Bayes

$$P(X, Y) = P(Y, X)$$

Teorema de Bayes

$$P(Y|X)P(X) = P(X, Y) = P(Y, X) = P(X|Y)P(Y)$$

Teorema de Bayes

$$P(Y|X)P(X) = P(X|Y)P(Y)$$

Teorema de Bayes

$$P(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}$$

Teorema de Bayes

$$P(\text{Hipótesis} \mid \text{Datos}) = \frac{P(\text{Datos} \mid \text{Hipótesis})P(\text{Hipótesis})}{P(\text{Datos})}$$

Teorema de Bayes

$$\underbrace{P(\text{Hipótesis} \mid \text{Datos})}_{\text{Posteriori}} = \frac{\overbrace{P(\text{Datos} \mid \text{Hipótesis})}^{\text{Verosimilitud}} \overbrace{P(\text{Hipótesis})}^{\text{Priori}}}{\underbrace{P(\text{Datos})}_{\text{Evidencia}}}$$

Teorema de Bayes

$$\underbrace{P(\text{Hipótesis} \mid \text{Datos, Modelo})}_{\text{Posteriori}} = \frac{\overbrace{P(\text{Datos} \mid \text{Hipótesis, Modelo})}^{\text{Verosimilitud}} \overbrace{P(\text{Hipótesis} \mid \text{Modelo})}^{\text{Priori}}}{\underbrace{P(\text{Datos} \mid \text{Modelo})}_{\text{Evidencia}}}$$

Teorema de Bayes

$$\underbrace{P(\text{Hipótesis} \mid \text{Datos, Modelo})}_{\text{Posteriori}} = \frac{\overbrace{P(\text{Datos} \mid \text{Hipótesis, Modelo})}^{\text{Verosimilitud}} \overbrace{P(\text{Hipótesis} \mid \text{Modelo})}^{\text{Priori}}}{\underbrace{P(\text{Datos} \mid \text{Modelo})}_{\text{Evidencia}}}$$

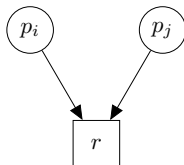
El **modelo** es lo que permite relacionar los **datos** con nuestras **hipótesis**!

Modelos gráficos

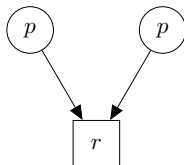
Modelos gráficos

Rendimiento aleatorio
centrado en la habilidad

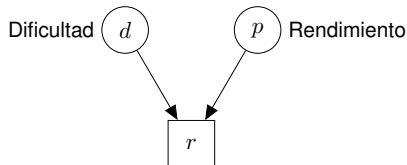
Resultado



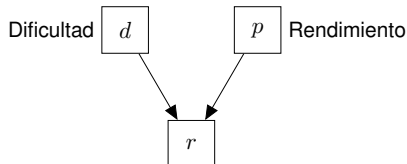
Modelos gráficos



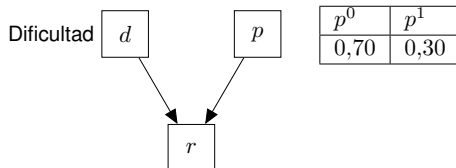
Modelos gráficos



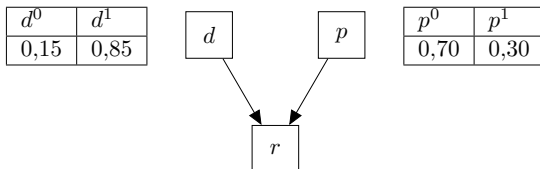
Modelos gráficos



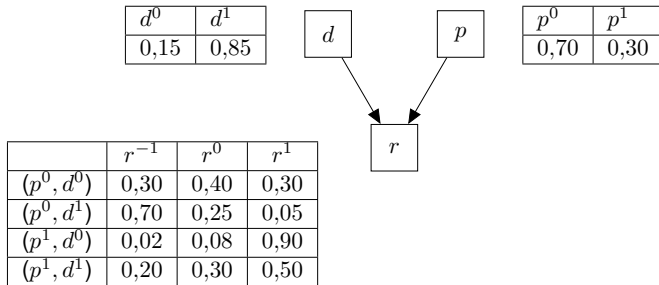
Modelos gráficos



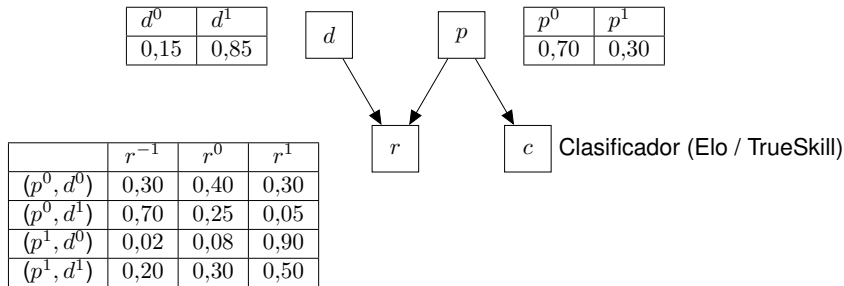
Modelos gráficos



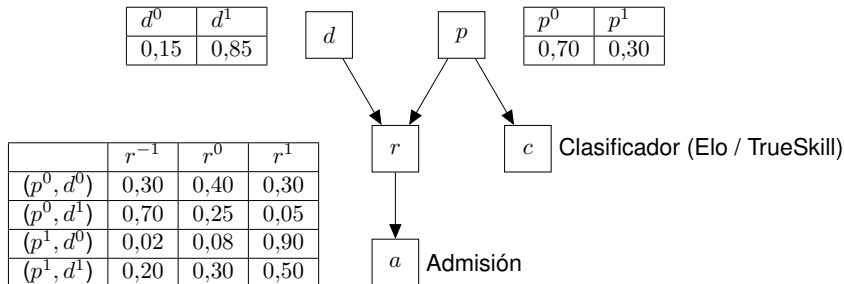
Modelos gráficos



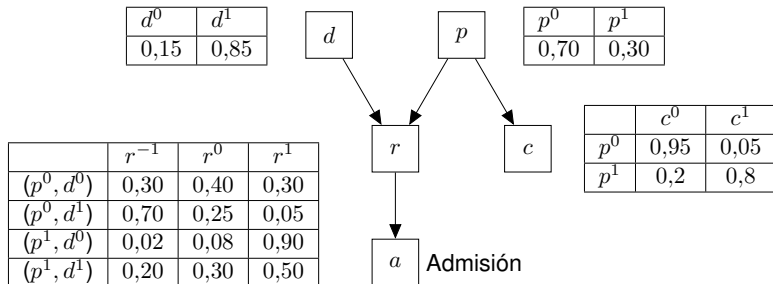
Modelos gráficos



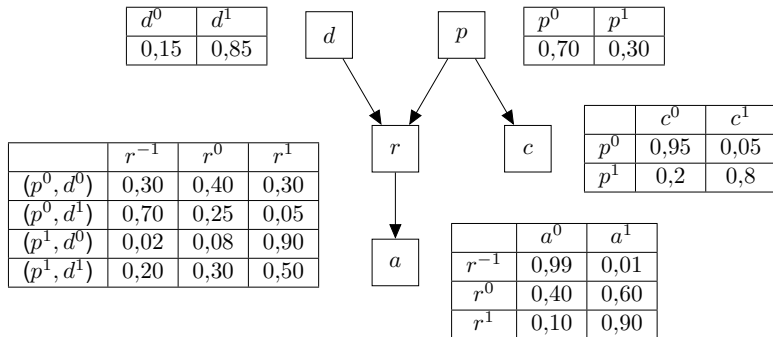
Modelos gráficos



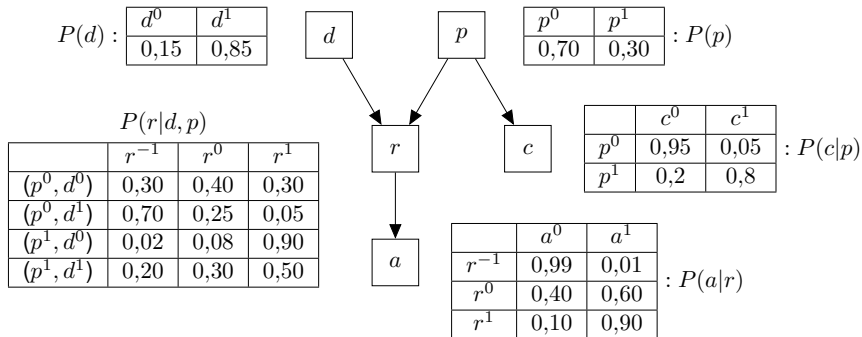
Modelos gráficos



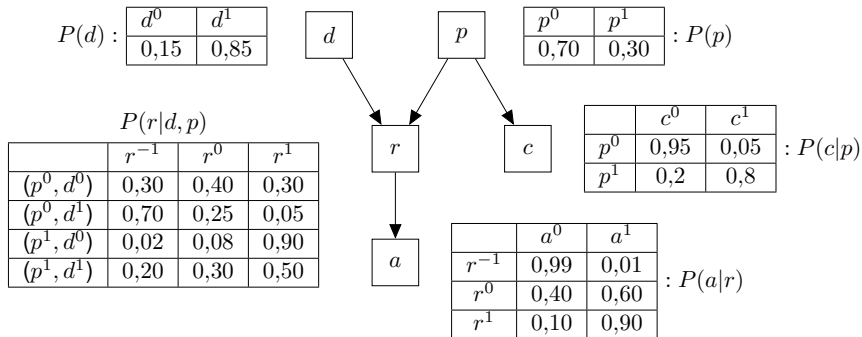
Modelos gráficos



Modelos gráficos

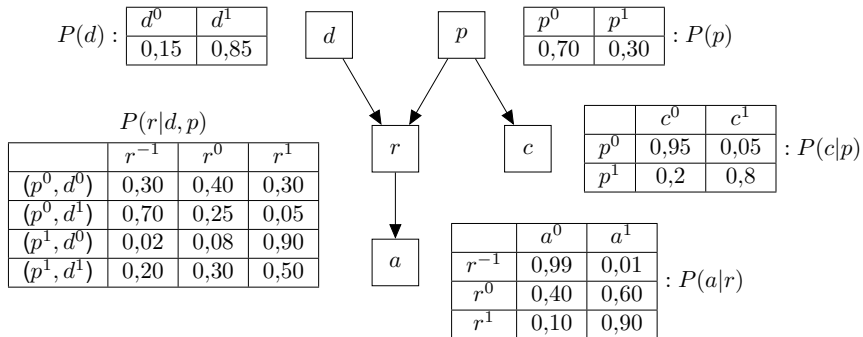


Modelos gráficos



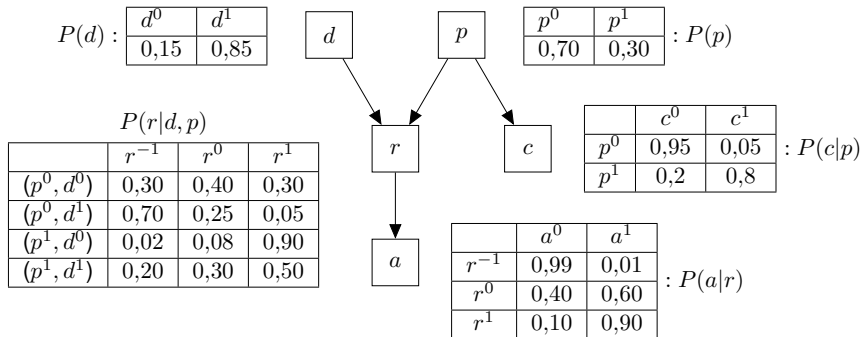
Los factores son útiles para definir distribuciones de probabilidad en espacios de alta dimensión.

Modelos gráficos



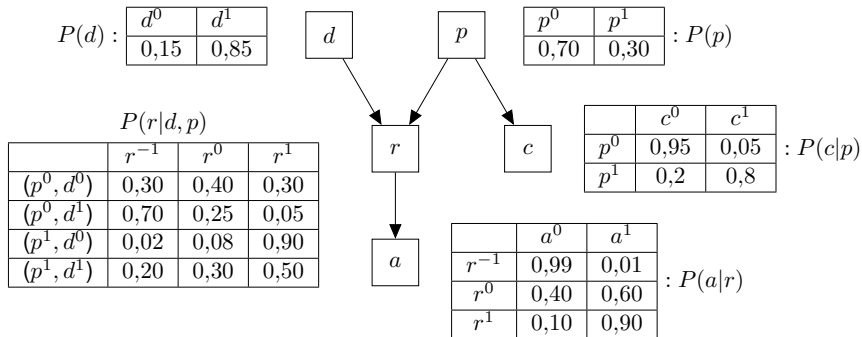
$$P(d, p, r, c, a) =$$

Modelos gráficos



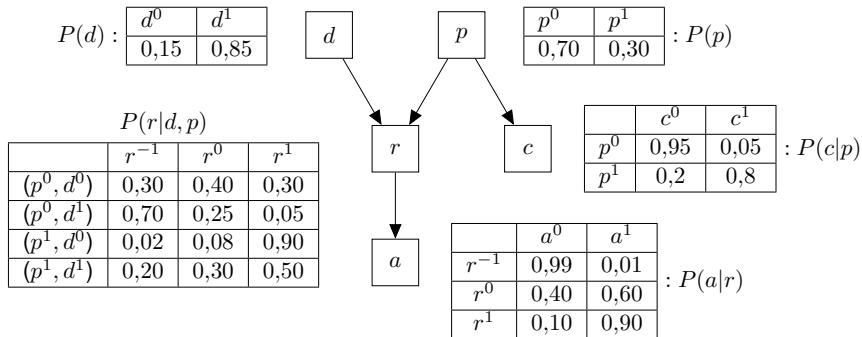
$$P(d, p, r, c, a) = P(d)P(p|d)P(r|d, p)P(c|d, p, r)P(a|d, p, r, c)$$

Modelos gráficos



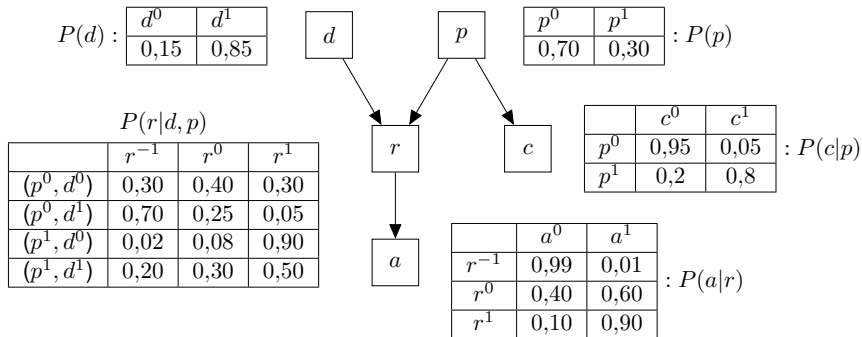
$$P(d, p, r, c, a) = P(d)P(p|\emptyset)P(r|d, p)P(c|\emptyset, p, r)P(a|\emptyset, p, r, c)$$

Modelos gráficos



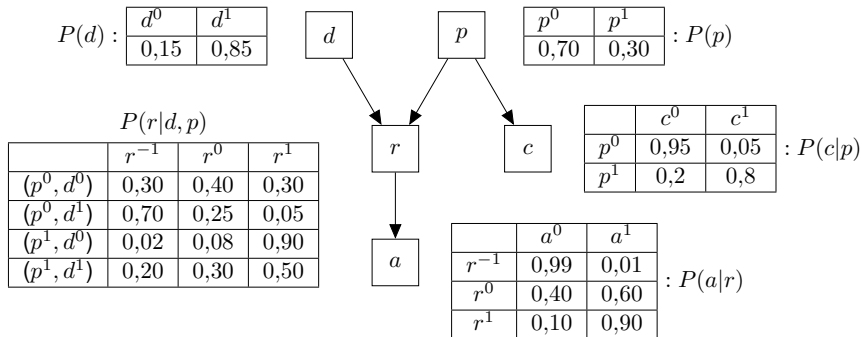
$$P(d, p, r, c, a) = P(d)P(p)P(r|d, p)P(a|r)P(c|p)$$

Modelos gráficos



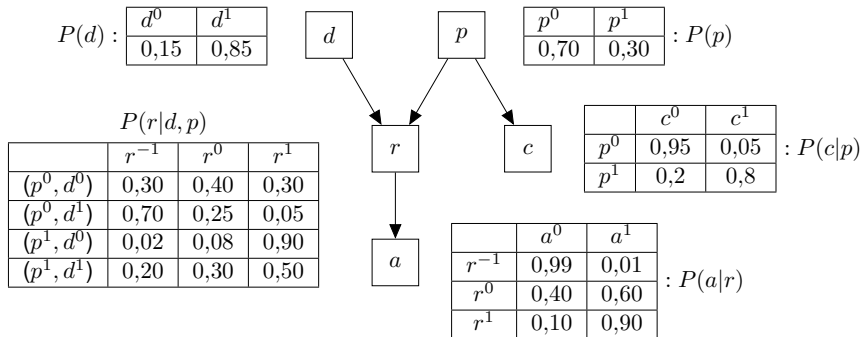
$$P(d^1, p^1, r^1, c^1, a^1) = P(d^1)P(p^1)P(r^1|d^1, p^1)P(a^1|r^1)P(c^1|p^1)$$

Modelos gráficos



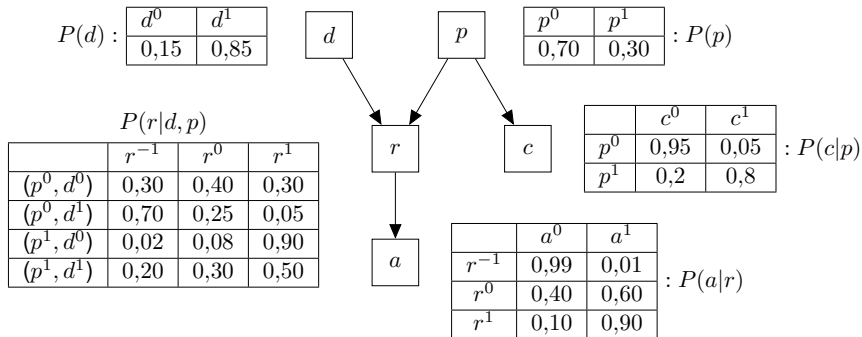
$$P(d^1, p^1, r^1, c^1, a^1) = 0.85 \cdot P(p^1)P(r^1|d^1, p^1)P(a^1|r^1)P(c^1|p^1)$$

Modelos gráficos



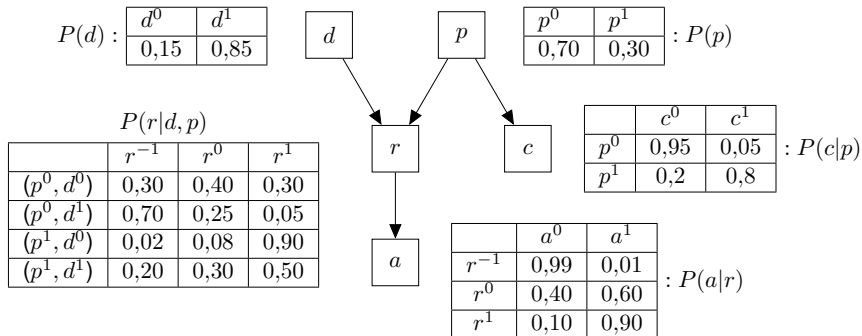
$$P(d^1, p^1, r^1, c^1, a^1) = 0.85 \cdot 0.30 \cdot 0.50 \cdot 0.80 \cdot 0.90$$

Modelos gráficos



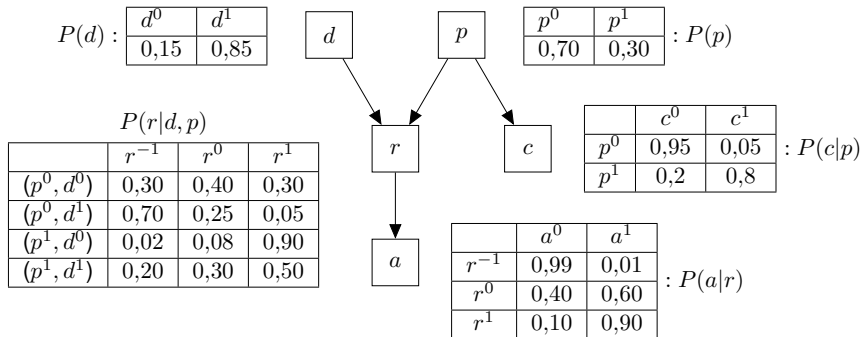
$$P(d^1, p^1, r^1, c^1, a^1) = 0.85 \cdot 0.30 \cdot 0.50 \cdot 0.80 \cdot 0.90 \approx 0.09$$

Modelos gráficos



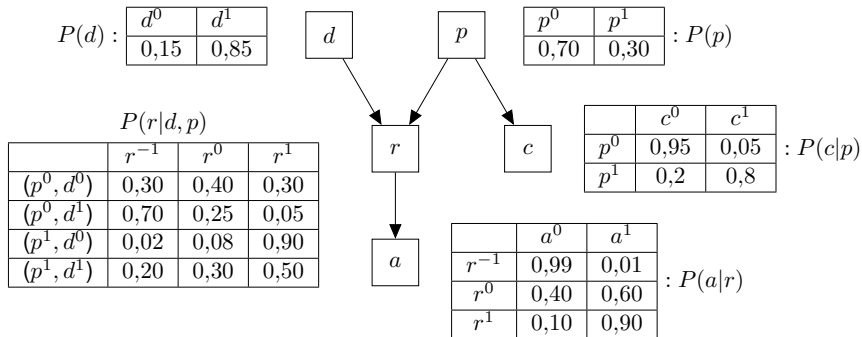
$$P(r^1) =$$

Modelos gráficos



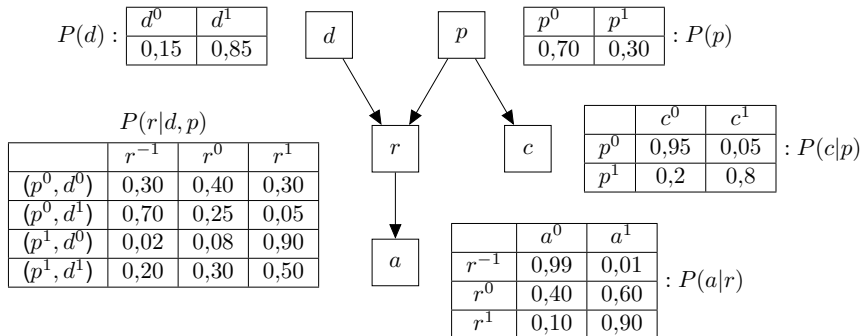
$$P(r^1) = \sum_i \sum_j \sum_l \sum_m P(d^i, p^j, r^1, c^l, a^m)$$

Modelos gráficos



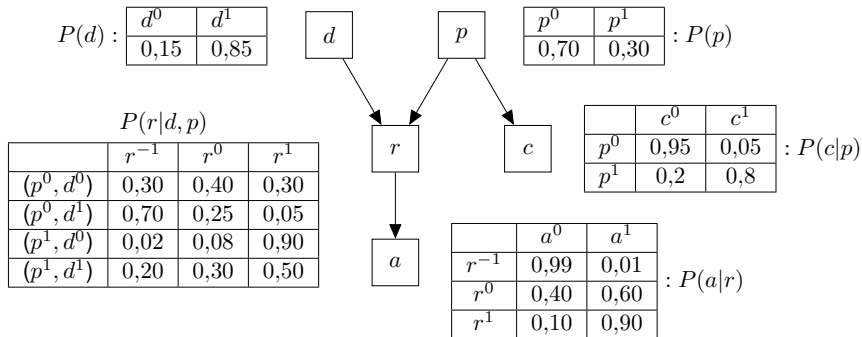
$$P(r^1) = \sum_{i,j,l,m} P(d^i, p^j, r^1, c^l, a^m)$$

Modelos gráficos



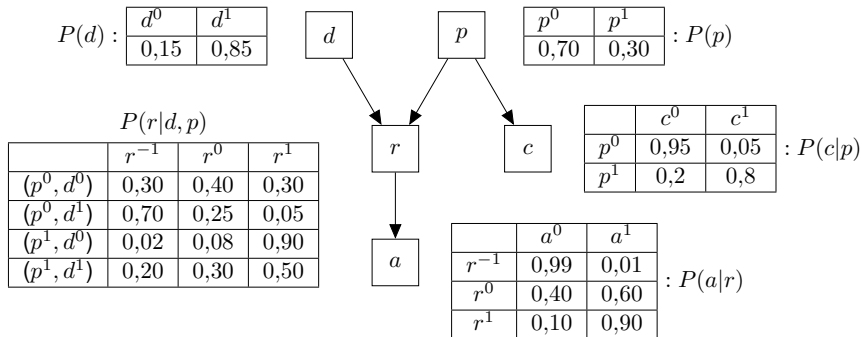
$$P(r^1) = \sum_{i,j,l,m} P(d^i)P(p^j)P(r^1|d^i, p^j)P(c^l|p^j)P(a^m|r^1)$$

Modelos gráficos



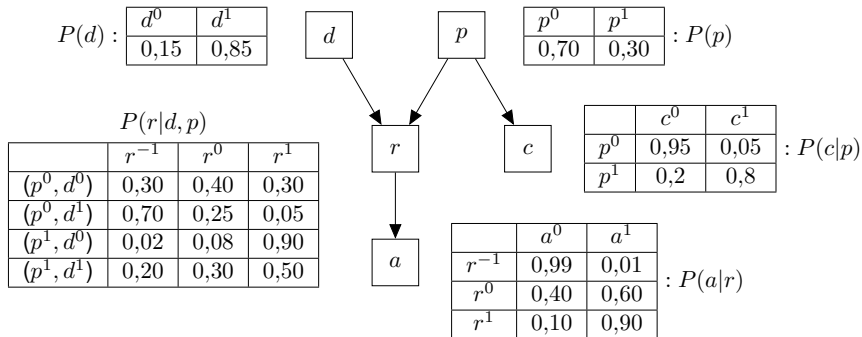
$$P(r^1) = \left(\sum_m P(a^m | r^1) \right) \left(\sum_{i,j,l} P(d^i) P(p^j) P(r^1 | d^i, p^j) P(c^l | p^j) \right)$$

Modelos gráficos



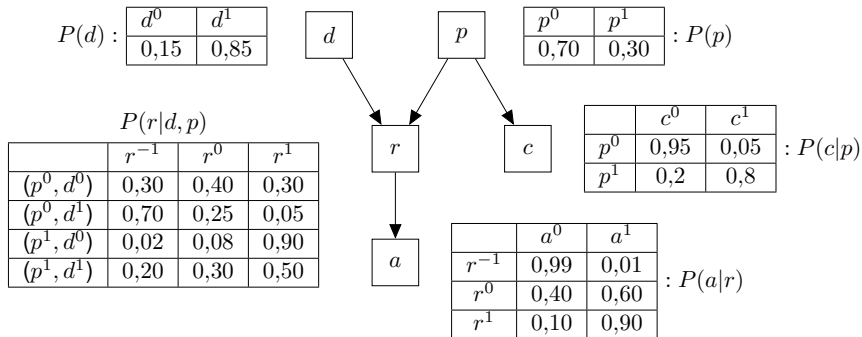
$$P(r^1) = \left(\sum_m P(a^m | r^1) \right) \left(\sum_{i,j} P(d^i) P(p^j) P(r^1 | d^i, p^j) \left(\sum_l P(c^l | p^j) \right) \right)$$

Modelos gráficos



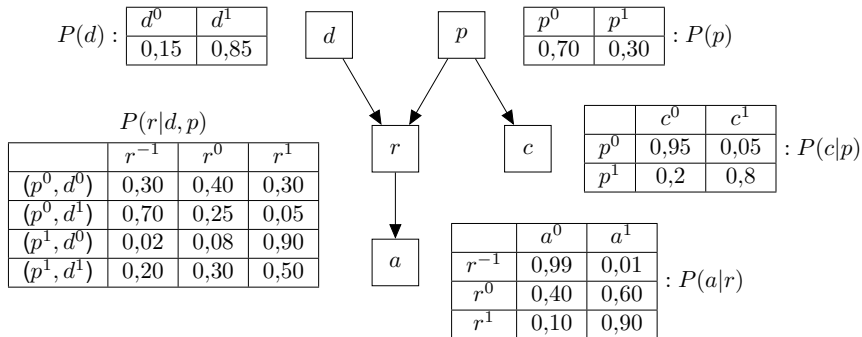
$$P(r^1) = \left(\sum_m P(a^m | r^1) \right) \left(\sum_{i,j} P(d^i) P(p^j) P(r^1 | d^i, p^j) \left(\sum_l P(c^l | p^j) \right) \right)$$

Modelos gráficos



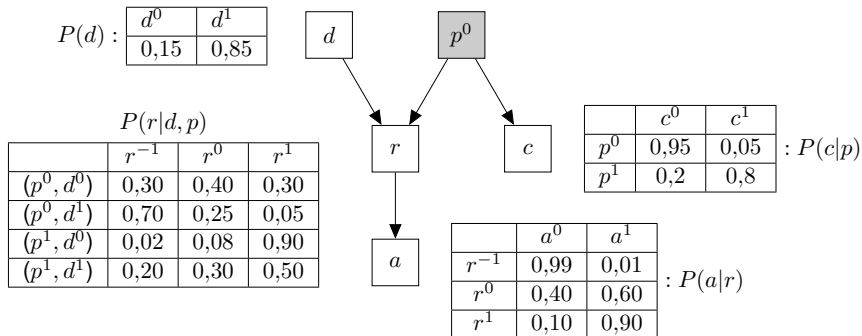
$$P(r^1) = \left(\sum_m \cancel{P(a^m|r^1)} \right) \left(\sum_{i,j} P(d^i)P(p^j)P(r^1|d^i,p^j) \left(\sum_l \cancel{P(c^l|p^j)} \right) \right)$$

Modelos gráficos



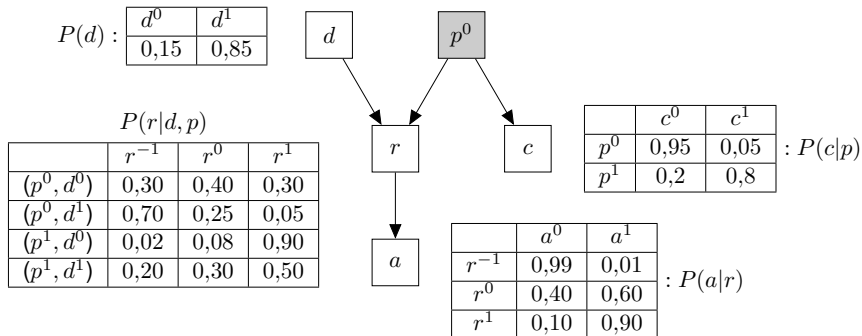
$$P(r^1) = \sum_{i,j} P(d^i)P(p^j)P(r^1|d^i, p^j)$$

Modelos gráficos



$$P(r^1) = \sum_{i,j} P(d^i)P(p^j)P(r^1|d^i, p^j) \approx 0.23$$

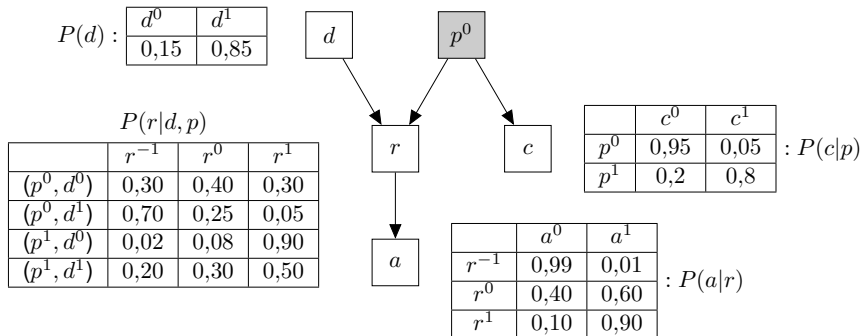
Modelos gráficos



$$P(r^1) = \sum_{i,j} P(d^i)P(p^j)P(r^1|d^i, p^j) \approx 0.23$$

$$P(r^1|p^0) =$$

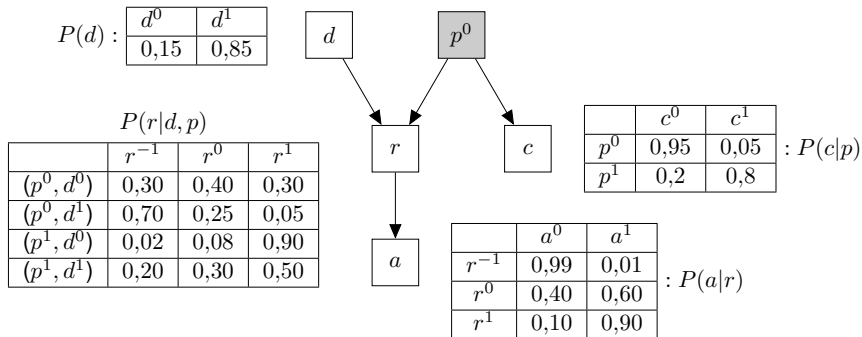
Modelos gráficos



$$P(r^1) = \sum_{i,j} P(d^i)P(p^j)P(r^1|d^i, p^j) \approx 0.23$$

$$P(r^1|p^0) = \frac{P(r^1, p^0)}{P(p^0)}$$

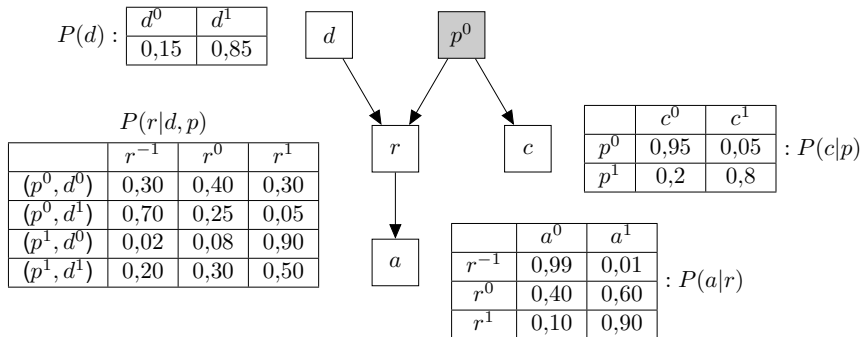
Modelos gráficos



$$P(r^1) = \sum_{i,j} P(d^i)P(p^j)P(r^1|d^i, p^j) \approx 0.23$$

$$P(r^1|p^0) = \frac{1}{P(p^0)} \sum_{i,l,m} P(d^i)P(p^0)P(r^1|d^i, p^0)P(c^l|p^0)P(a^m|r^k)$$

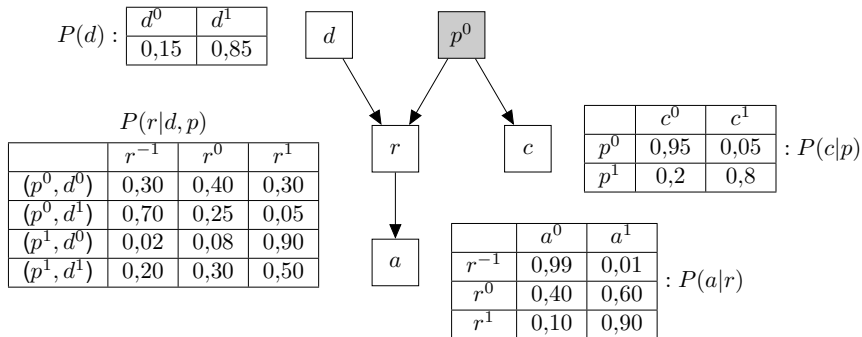
Modelos gráficos



$$P(r^1) = \sum_{i,j} P(d^i)P(p^j)P(r^1|d^i, p^j) \approx 0.23$$

$$P(r^1|p^0) = \frac{1}{P(p^0)} \sum_i P(d^i)P(p^0)P(r^1|d^i, p^0)$$

Modelos gráficos



$$P(r^1) = \sum_{i,j} P(d^i)P(p^j)P(r^1|d^i, p^j) \approx 0.23$$

$$P(r^1|p^0) = \frac{1}{P(p^0)} \sum_i P(d^i)P(p^0)P(r^1|d^i, p^0) \approx 0.09$$

Flujos de inferencia

Flujos de inferencia

$V \notin \text{Observable}$	$V \in \text{Observable}$
$X \rightarrow V \rightarrow Y$	
$X \leftarrow V \leftarrow Y$	
$X \leftarrow V \rightarrow Y$	
$X \rightarrow V \leftarrow Y$	

Flujos de inferencia

	$V \notin \text{Observable}$ Sí	$V \in \text{Observable}$
$X \rightarrow V \rightarrow Y$		
$X \leftarrow V \leftarrow Y$		
$X \leftarrow V \rightarrow Y$		
$X \rightarrow V \leftarrow Y$		

Flujos de inferencia

	$V \notin \text{Observable}$	$V \in \text{Observable}$
$X \rightarrow V \rightarrow Y$	Sí	
$X \leftarrow V \leftarrow Y$	Sí	
$X \leftarrow V \rightarrow Y$		
$X \rightarrow V \leftarrow Y$		

Flujos de inferencia

	$V \notin \text{Observable}$	$V \in \text{Observable}$
$X \rightarrow V \rightarrow Y$	Sí	
$X \leftarrow V \leftarrow Y$	Sí	
$X \leftarrow V \rightarrow Y$	Sí	
$X \rightarrow V \leftarrow Y$		

Flujos de inferencia

	$V \notin \text{Observable}$	$V \in \text{Observable}$
$X \rightarrow V \rightarrow Y$	Sí	No
$X \leftarrow V \leftarrow Y$	Sí	
$X \leftarrow V \rightarrow Y$	Sí	
$X \rightarrow V \leftarrow Y$		

Flujos de inferencia

	$V \notin \text{Observable}$	$V \in \text{Observable}$
$X \rightarrow V \rightarrow Y$	Sí	No
$X \leftarrow V \leftarrow Y$	Sí	No
$X \leftarrow V \rightarrow Y$	Sí	
$X \rightarrow V \leftarrow Y$		

Flujos de inferencia

	$V \notin \text{Observable}$	$V \in \text{Observable}$
$X \rightarrow V \rightarrow Y$	Sí	No
$X \leftarrow V \leftarrow Y$	Sí	No
$X \leftarrow V \rightarrow Y$	Sí	No
$X \rightarrow V \leftarrow Y$		

Flujos de inferencia

	$V \notin \text{Observable}$	$V \in \text{Observable}$
$X \rightarrow V \rightarrow Y$	Sí	No
$X \leftarrow V \leftarrow Y$	Sí	No
$X \leftarrow V \rightarrow Y$	Sí	No
$X \rightarrow V \leftarrow Y$	No	

Flujos de inferencia

	$V \notin \text{Observable}$	$V \in \text{Observable}$
$X \rightarrow V \rightarrow Y$	Sí	No
$X \leftarrow V \leftarrow Y$	Sí	No
$X \leftarrow V \rightarrow Y$	Sí	No
$X \rightarrow V \leftarrow Y$	No	Sí

Flujos de inferencia

	$V \notin \text{Observable}$	$V \in \text{Observable}$
$X \rightarrow V \rightarrow Y$	Sí	No
$X \leftarrow V \leftarrow Y$	Sí	No
$X \leftarrow V \rightarrow Y$	Sí	No
$X \rightarrow V \leftarrow Y$	No	<div>Sí</div> <div>O algún descendiente observable</div>

Flujos de inferencia

	$V \notin \text{Observable}$	$V \in \text{Observable}$
$X \rightarrow V \rightarrow Y$	Sí	No
$X \leftarrow V \leftarrow Y$	Sí	No
$X \leftarrow V \rightarrow Y$	Sí	No
$X \rightarrow V \leftarrow Y$	No	Sí
	Y ningún descendiente observable	O algún descendiente observable

Flujos de inferencia

	$V \notin \text{Observable}$	$V \in \text{Observable}$
$X \rightarrow V \rightarrow Y$	Sí	No
$X \leftarrow V \leftarrow Y$	Sí	No
$X \leftarrow V \rightarrow Y$	Sí	No
$X \rightarrow V \leftarrow Y$	No	Sí
	Y ningún descendiente observable	O algún descendiente observable

Un flujo de inferencia permanece abierto si:

- Toda consecuencia común (o alguno de sus descendientes) es observable
- Ninguna otra variable es observable

Flujos de inferencia

	$V \notin \text{Observable}$	$V \in \text{Observable}$
$X \rightarrow V \rightarrow Y$	Sí	No
$X \leftarrow V \leftarrow Y$	Sí	No
$X \leftarrow V \rightarrow Y$	Sí	No
$X \rightarrow V \leftarrow Y$	No	Sí
	Y ningún descendiente observable	O algún descendiente observable

Un flujo de inferencia permanece abierto si:

- Toda consecuencia común (o alguno de sus descendientes) es observable
- Ninguna otra variable es observable

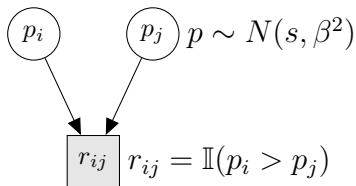
TrueSkill



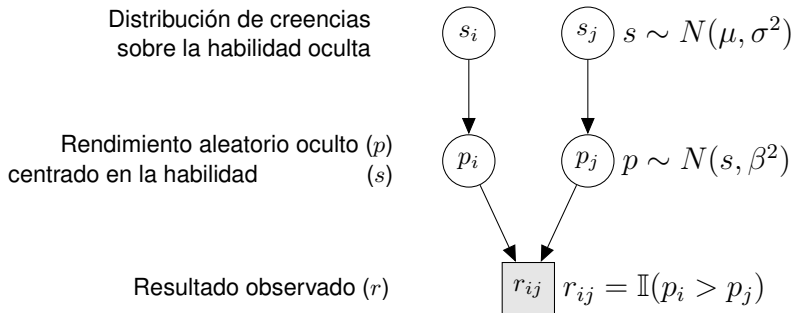
TrueSkill

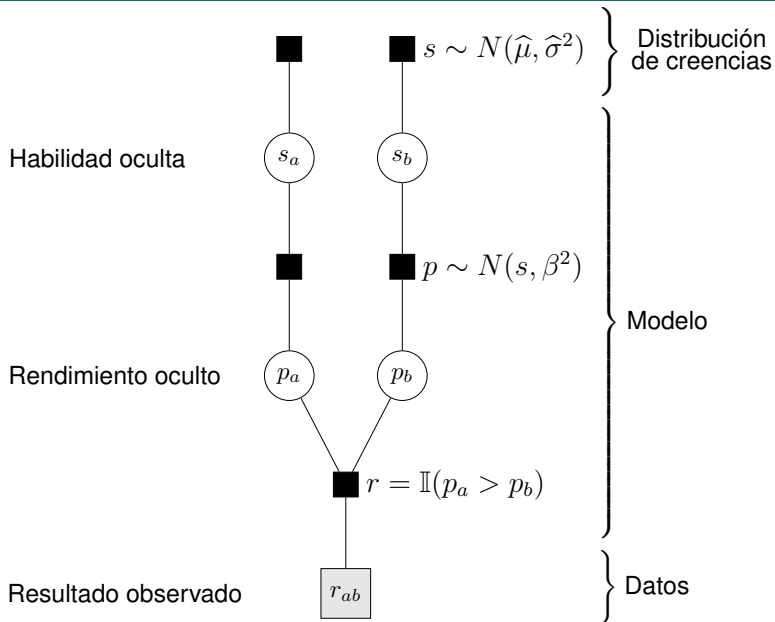
Rendimiento aleatorio oculto (p)
centrado en la habilidad estimada (s)

Resultado observado (r)

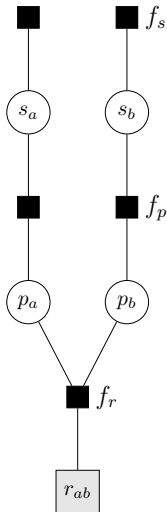


TrueSkill





Sum-product algorithm



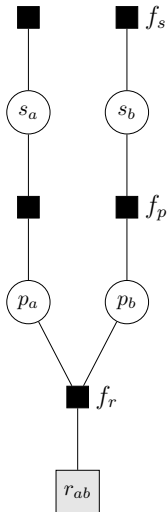
$$P(x) = \prod_{h \in n(x)} m_{h \rightarrow x}$$

$m_{x \rightarrow f}(x)$: Mensaje de variable x a factor f

$m_{f \rightarrow x}(x)$: Mensaje factor f a variable x

$n(v)$: Conjunto de nodos vecinos del nodo v

Sum-product algorithm



$$P(x) = \prod_{h \in n(x)} m_{h \rightarrow x}$$

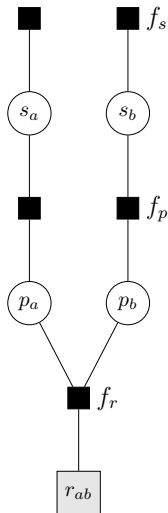
$m_{x \rightarrow f}(x)$: Mensaje de variable x a factor f

$m_{f \rightarrow x}(x)$: Mensaje factor f a variable x

$n(v)$: Conjunto de nodos vecinos del nodo v

$$m_{x \rightarrow f}(x) = \prod_{h \in n(x) \setminus \{f\}} m_{h \rightarrow x}(x)$$

Sum-product algorithm



$$P(x) = \prod_{h \in n(x)} m_{h \rightarrow x}$$

$m_{x \rightarrow f}(x)$: Mensaje de variable x a factor f

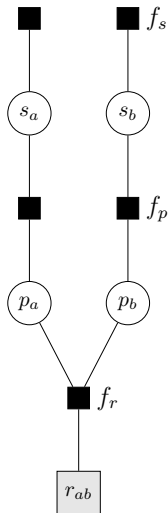
$m_{f \rightarrow x}(x)$: Mensaje factor f a variable x

$n(v)$: Conjunto de nodos vecinos del nodo v

$$m_{x \rightarrow f}(x) = \prod_{h \in n(x) \setminus \{f\}} m_{h \rightarrow x}(x)$$

$$m_{f \rightarrow x}(x) = \sum_{X \setminus \{x\}} \left(f(X) \prod_{h \in n(f) \setminus \{x\}} m_{h \rightarrow f}(h) \right)$$

Sum-product algorithm



$$P(x) = \prod_{h \in n(x)} m_{h \rightarrow x}$$

$m_{x \rightarrow f}(x)$: Mensaje de variable x a factor f

$m_{f \rightarrow x}(x)$: Mensaje factor f a variable x

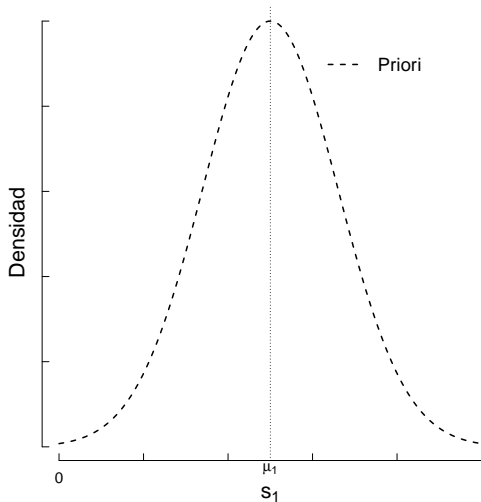
$n(v)$: Conjunto de nodos vecinos del nodo v

$$m_{x \rightarrow f}(x) = \prod_{h \in n(x) \setminus \{f\}} m_{h \rightarrow x}(x)$$

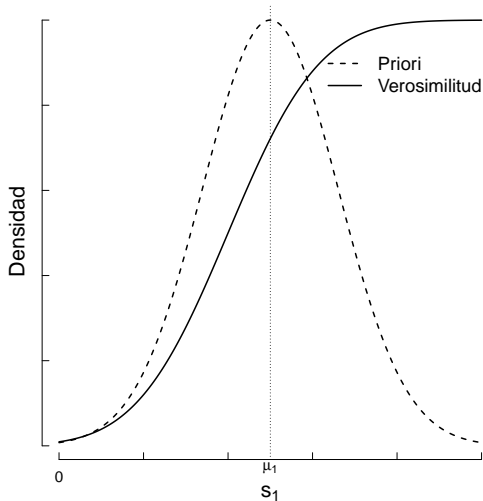
$$m_{f \rightarrow x}(x) = \sum_{X \setminus \{x\}} \left(f(X) \prod_{h \in n(f) \setminus \{x\}} m_{h \rightarrow f}(h) \right)$$

$$\overbrace{P(s_1 \mid r, \text{Modelo})}^{\text{Posteriori}} \propto \overbrace{N(s_1 \mid \mu_1, \sigma_1^2)}^{\text{Priori}} \overbrace{1 - \Phi(0 \mid s_1 - \mu_2, \vartheta^2 - \sigma_1^2)}^{\text{Verosimilitud}} \quad \text{Caso ganador}$$

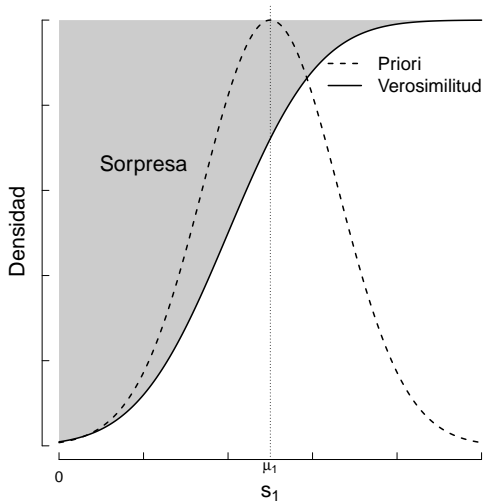
$$\overbrace{P(s_1 | r, \text{Modelo})}^{\text{Posteriori}} \propto \overbrace{N(s_1 | \mu_1, \sigma_1^2)}^{\text{Priori}} \overbrace{1 - \Phi(0 | s_1 - \mu_2, \vartheta^2 - \sigma_1^2)}^{\text{Verosimilitud}} \quad \text{Caso ganador}$$



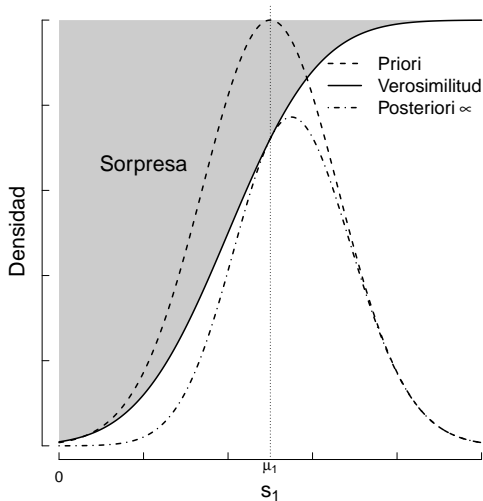
$$\overbrace{P(s_1 | r, \text{Modelo})}^{\text{Posteriori}} \propto \overbrace{N(s_1 | \mu_1, \sigma_1^2)}^{\text{Priori}} \overbrace{1 - \Phi(0 | s_1 - \mu_2, \vartheta^2 - \sigma_1^2)}^{\text{Verosimilitud}} \quad \text{Caso ganador}$$



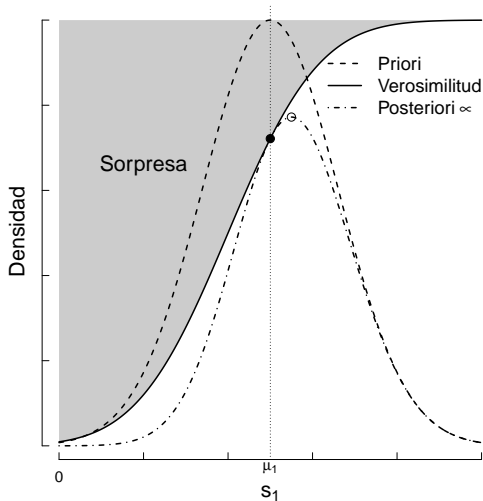
$$\overbrace{P(s_1 | r, \text{Modelo})}^{\text{Posteriori}} \propto \overbrace{N(s_1 | \mu_1, \sigma_1^2)}^{\text{Priori}} \overbrace{1 - \Phi(0 | s_1 - \mu_2, \vartheta^2 - \sigma_1^2)}^{\text{Verosimilitud}} \quad \text{Caso ganador}$$



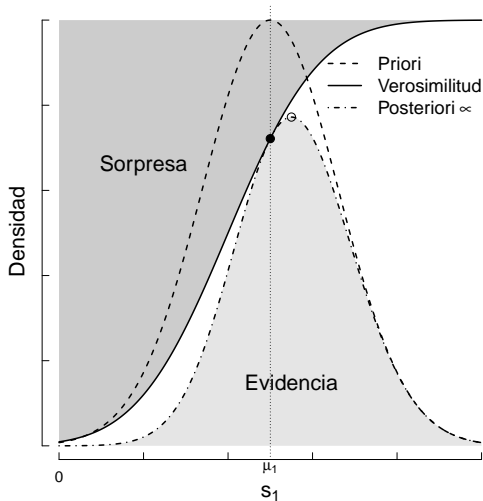
$$\overbrace{P(s_1 | r, \text{Modelo})}^{\text{Posteriori}} \propto \overbrace{N(s_1 | \mu_1, \sigma_1^2)}^{\text{Priori}} \overbrace{1 - \Phi(0 | s_1 - \mu_2, \vartheta^2 - \sigma_1^2)}^{\text{Verosimilitud}} \quad \text{Caso ganador}$$



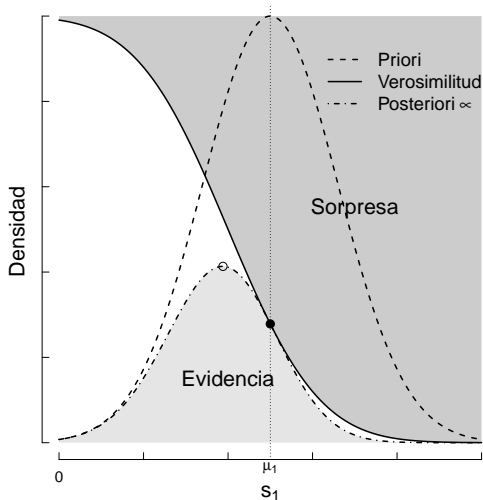
$$\overbrace{P(s_1 | r, \text{Modelo})}^{\text{Posteriori}} \propto \overbrace{N(s_1 | \mu_1, \sigma_1^2)}^{\text{Priori}} \overbrace{1 - \Phi(0 | s_1 - \mu_2, \vartheta^2 - \sigma_1^2)}^{\text{Verosimilitud}} \quad \text{Caso ganador}$$



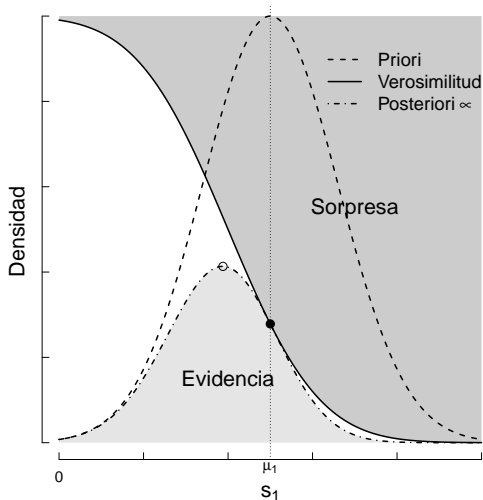
$$\overbrace{P(s_1 | r, \text{Modelo})}^{\text{Posteriori}} \propto \overbrace{N(s_1 | \mu_1, \sigma_1^2)}^{\text{Priori}} \overbrace{1 - \Phi(0 | s_1 - \mu_2, \vartheta^2 - \sigma_1^2)}^{\text{Verosimilitud}} \quad \text{Caso ganador}$$



$$\overbrace{P(s_2 \mid r, \text{Modelo})}^{\text{Posteriori}} \propto \overbrace{N(s_2 \mid \mu_2, \sigma_2^2)}^{\text{Priori}} \overbrace{\Phi(0 \mid \mu_1 - s_2, \vartheta^2 - \sigma_1^2)}^{\text{Verosimilitud}} \quad \text{Caso perdedor}$$



$$\overbrace{P(s_2 \mid r, \text{Modelo})}^{\text{Posteriori}} \propto \overbrace{N(s_2 \mid \mu_2, \sigma_2^2)}^{\text{Priori}} \overbrace{\Phi(0 \mid \mu_1 - s_2, \vartheta^2 - \sigma_1^2)}^{\text{Verosimilitud}} \quad \text{Caso perdedor}$$



Selección de modelo

¿Y nuestras creencias respecto de modelos alternativos?

Selección de modelo

$$P(M|D) = \frac{P(D|M)P(M)}{P(D)}$$

Selección de modelo

$$P(M|D) = \frac{P(D|M)P(M)}{P(D)}$$

$$\frac{P(M_i|D)}{P(M_j|D)} = \frac{P(D|M_i) P(M_i)}{P(D|M_j) P(M_j)}$$

Selección de modelo

$$P(M|D) = \frac{P(D|M)P(M)}{P(D)}$$

$$\frac{P(M_i|D)}{P(M_j|D)} = \frac{P(D|M_i) P(M_i)}{\underbrace{P(D|M_j) P(M_j)}_{\text{Evidencia!}}}$$

Selección de modelo

$$P(M|D) = \frac{P(D|M)P(M)}{P(D)}$$

$$\frac{P(M_i|D)}{P(M_j|D)} = \frac{P(D|M_i) \cancel{P(M_i)}}{\underbrace{P(D|M_j)}_{\text{Evidencia!}} \cancel{P(M_j)}}$$

Selección de modelo

$$P(M|D) = \frac{P(D|M)P(M)}{P(D)}$$

$$\frac{P(M_i|D)}{P(M_j|D)} = \frac{P(D|M_i) \cancel{P(M_i)}}{\underbrace{P(D|M_j)}_{\text{Evidencia!}} \cancel{P(M_j)}}$$

Lo único que necesitamos es la **evidencia!**

Selección de modelo

$$P(M|D) = \frac{P(D|M)P(M)}{P(D)}$$

$$\frac{P(M_i|D)}{P(M_j|D)} = \frac{P(D|M_i) \cancel{P(M_i)}}{\underbrace{P(D|M_j)}_{\text{Evidencia!}} \cancel{P(M_j)}}$$

Lo único que necesitamos es la **evidencia!**

Evidencia

$$P(C|D, M) = \frac{P(D|C, M)P(C|M)}{\underbrace{P(D|M)}_{\text{Evidencia}}}$$

Evidencia

$$P(C|D, M) = \frac{P(D|C, M)P(C|M)}{\underbrace{P(D|M)}_{\text{Evidencia}}}$$

$$P(D|M) = \sum_C P(D|C, M) P(C|M)$$

Evidencia

$$P(C|D, M) = \frac{P(D|C, M)P(C|M)}{\underbrace{P(D|M)}_{\text{Evidencia}}}$$

$$P(D|M) = \sum_C \underbrace{P(D|C, M)}_{\text{Predicción de D dado C}} \underbrace{P(C|M)}_{\text{Creencia de C a Priori}}$$

Evidencia

$$P(C|D, M) = \frac{P(D|C, M)P(C|M)}{\underbrace{P(D|M)}_{\text{Evidencia}}}$$

$$P(D|M) = \sum_C \underbrace{P(D|C, M)}_{\substack{\text{Predicción} \\ \text{de D dado C}}} \underbrace{P(C|M)}_{\substack{\text{Creencia de} \\ \text{C a Priori}}}$$

Predicción de la datos observados
pesando todas las creencias a priori

Evidencia

$$P(C|D, M) = \frac{P(D|C, M)P(C|M)}{\underbrace{P(D|M)}_{\text{Evidencia}}}$$

$$P(D|M) = \sum_C \underbrace{P(D|C, M)}_{\substack{\text{Predicción} \\ \text{de D dado C}}} \underbrace{P(C|M)}_{\substack{\text{Creencia de} \\ \text{C a Priori}}}$$

Predicción de la datos observados
pesando todas las creencias a priori

Preferimos modelos con la menor **sorpresa** en la evidencia!

Evidencia

$$P(C|D, M) = \frac{P(D|C, M)P(C|M)}{\underbrace{P(D|M)}_{\text{Evidencia}}}$$

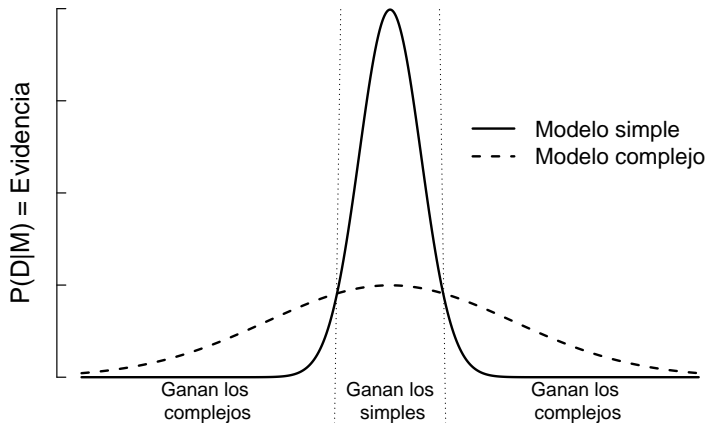
$$P(D|M) = \sum_C \underbrace{P(D|C, M)}_{\substack{\text{Predicción} \\ \text{de D dado C}}} \underbrace{P(C|M)}_{\substack{\text{Creencia de} \\ \text{C a Priori}}}$$

Predicción de la datos observados
pesando todas las creencias a priori

Preferimos modelos con la menor **sorpres**a en la evidencia!

$$P(M_q|D) > P(M_r|D) \iff P(D|M_q) > P(D|M_r)$$

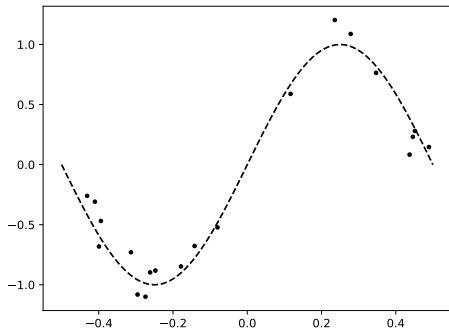
Evidencia



Balance natural entre complejidad y predicción

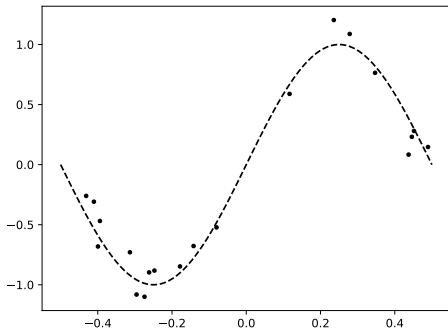
Regresión lineal Bayesiana

Función objetivo

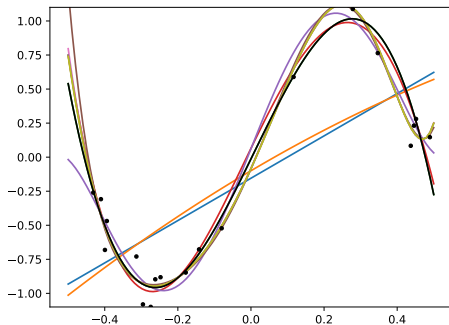


Regresión lineal Bayesiana

Función objetivo

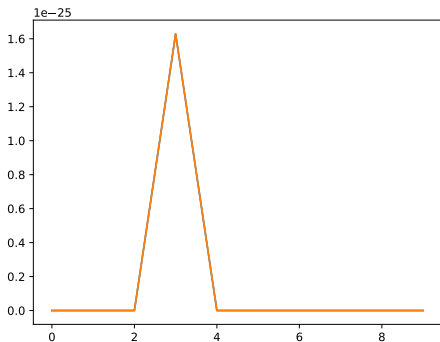


Modelos polinomiales



Evidencia vs Verosimilitud

Evidencia conjunta



Máxima verosimilitud

