Inferencia Bayesiana

Módulo 1: Estadística

Tirada	Resultado
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
…	…

¿Está cargada la moneda?

Hipótesis/Modelo

Estadística: nos permite sacar conclusiones del mundo exterior

Test Binomial

`scipy.stats.binom_test`

Hipótesis: resultados distribuidos binomialmente

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

Hipótesis/Modelo

¿Cuál es la probabilidad de obtener *m* o más caras de *n* tiradas?

$$\sum_{k \geq m}^{n} p\left(k \middle| \theta = \frac{1}{2}\right)$$

¿Cuál es la probabilidad de obtener un resultado *tan o más extremo*?

$$\sum_{k=m}^{n} p\left(k \middle| \theta = \frac{1}{2}\right) + \sum_{k=0}^{n-m} p\left(k \middle| \theta = \frac{1}{2}\right)$$

¡El *p-valor*!

Test Binomial

`scipy.stats.binom_test`

$$p = \sum_{k=m}^{n} p\left(k \middle| \theta = \frac{1}{2}\right) + \sum_{k=0}^{n-m} p\left(k \middle| \theta = \frac{1}{2}\right)$$

p-valor para 12 caras de 12 tiradas…

$$p = \sum_{k=12}^{12} p\left(k \middle| \theta = \frac{1}{2}\right) + \sum_{k=0}^{0} p\left(k \middle| \theta = \frac{1}{2}\right) = 0.00049$$

Si tiramos 6 veces la moneda,
y sale 6 veces cara...

¿diríamos que la moneda está cargada?

La “ciencia” hoy dice “sí”.

Más precisamente:

“Hay evidencia significativa de que está cargada ($p < 0.05$, test binomial)”

$$p = \sum_{k=6}^{6} p\left(k \middle| \theta = \frac{1}{2}\right) + \sum_{k=0}^{0} p\left(k \middle| \theta = \frac{1}{2}\right) = 0.03125$$

¡¿Es razonable esto?!

Estadística frecuentista…

$$p(6 \text{ caras}|\text{moneda honesta})$$

$$p(\text{datos}|\text{hipótesis})$$

Lo que queremos realmente..

$$p(\text{moneda honesta}|6 \text{ caras})$$

$$p(\text{hipótesis}|\text{datos})$$

¡Inferencia bayesiana!

Estadística (o inferencia, o análisis de datos) bayesiana
como alternativa a la estadística frecuentista o “clásica”

Permite (¡exige!) incorporar nuestro
conocimiento previo

Podemos hacer afirmaciones sobre cuán
probable es que la moneda esté cargada
(prohibido en frecuentismo)

Probabilidad como *creencia*

inferencia bayesiana vs. estadística frecuentista

mayormente

probabilidad como *creencia* vs. probabilidad como *frecuencia*

interpretación de la probabilidad

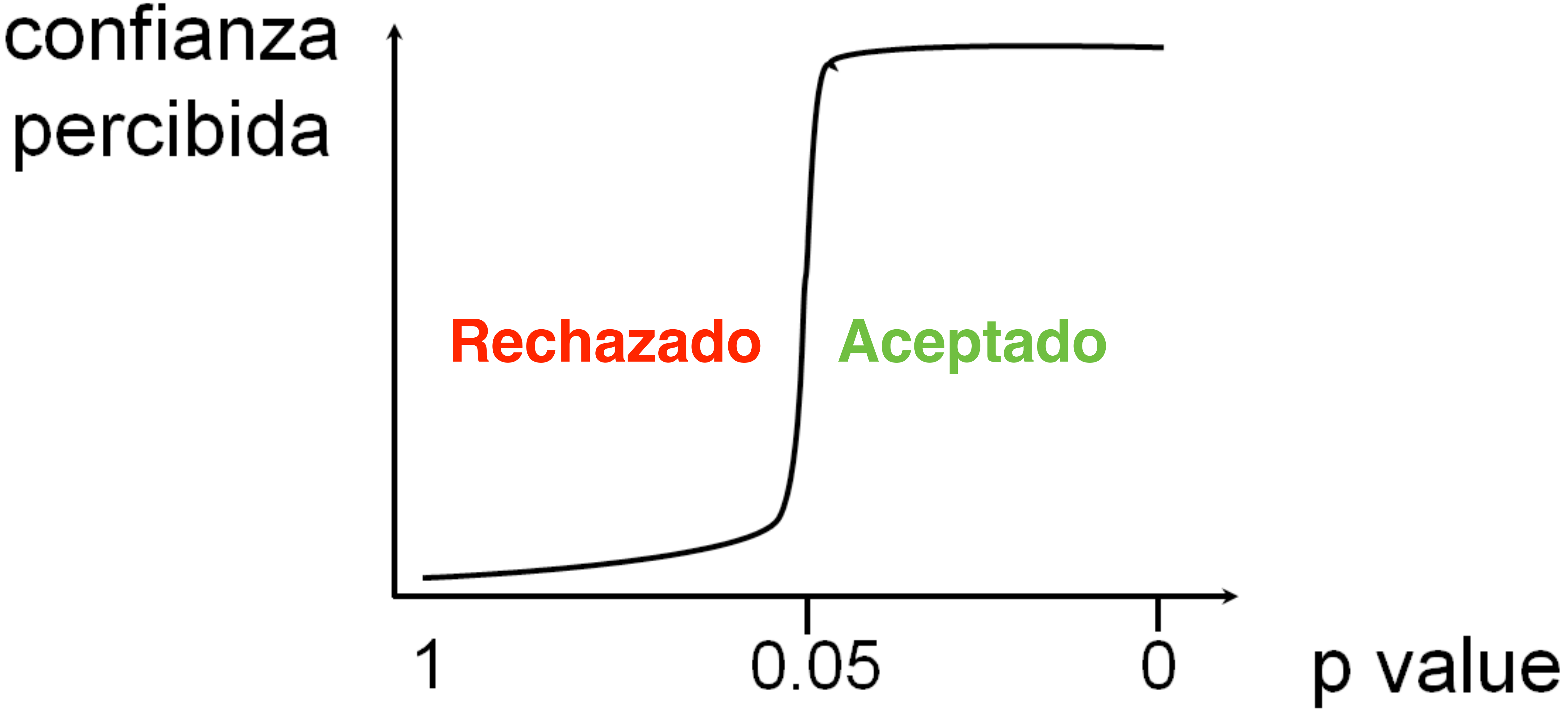
p(*lo que yo quiera*)

p sólo para muestreo ("datos")

Críticas al frecuentismo: 2 niveles

- críticas de fondo:
  - ignora conocimiento previo
  - dilema (mezcla) Fisher (significance testing) vs. Neyman/Pearson (hypothesis testing)

- críticas al (ab)uso:
  - convención de p: ¡1 de 20 estudios están mal!
  - uso ciego en general (“imposible hacer ciencia”)
  - *p hacking*
  - crisis de reproducibilidad: 36 % papers replicados en psicología



Propuesta real: cambiar a *p<0.005*

Inferencia Bayesiana

$$p(H,D) = p(D|H)p(H) = p(H|D)p(D)$$

$$\Downarrow$$

likelihood

prior

posterior

$$\boxed{p(H|D)} = \frac{\boxed{p(D|H)}\,p(H)}{p(D)}$$

Teorema de Bayes

Acumulación de Evidencia

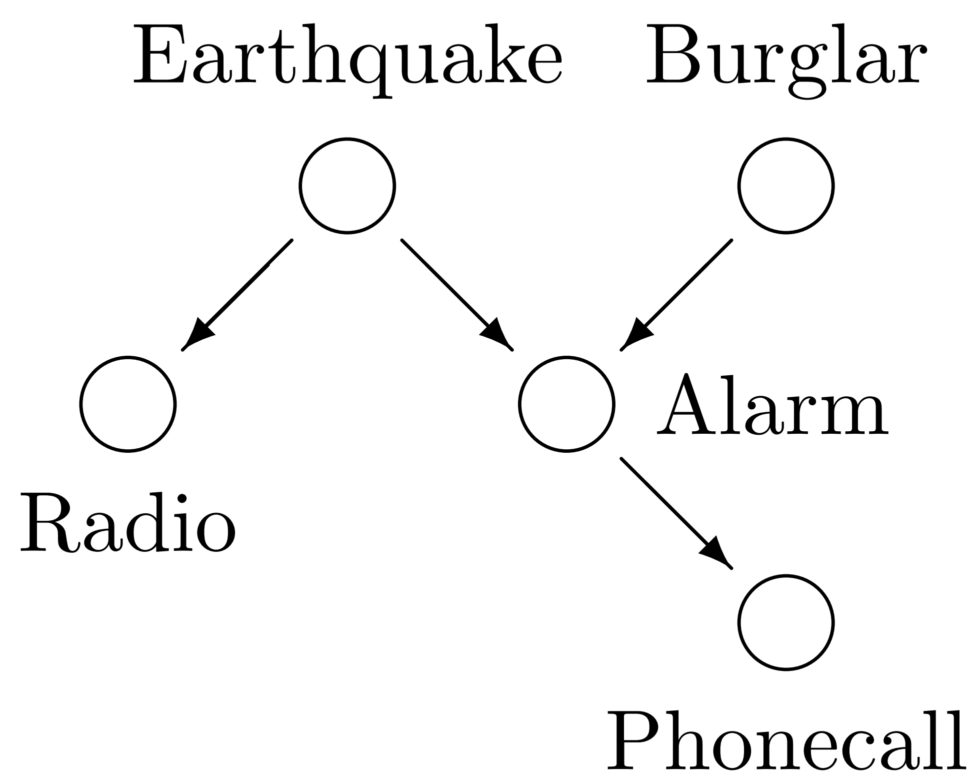
$$p(H|D_1) = \frac{p(D_1|H)P(H)}{p(D_1)}$$

$$p(H|D_2D_1) = \frac{p(D_2|HD_1)P(H|D_1)}{p(D_2|D_1)}$$

"The posterior is the new prior"

Redes Bayesianas/Modelos Jerárquicos

$$p(B,E,A,P,R) = p(B)p(E|B)p(A|B,E)p(P|A,B,E)p(R|P,A,B,E)$$



$$p(B,E,A,P,R) = p(B)p(E)p(A|B,E)p(P|A)p(R|E)$$

Inferencia Bayesiana

- Incorpora conocimiento *a priori* en forma natural (¡y obligatoria!)

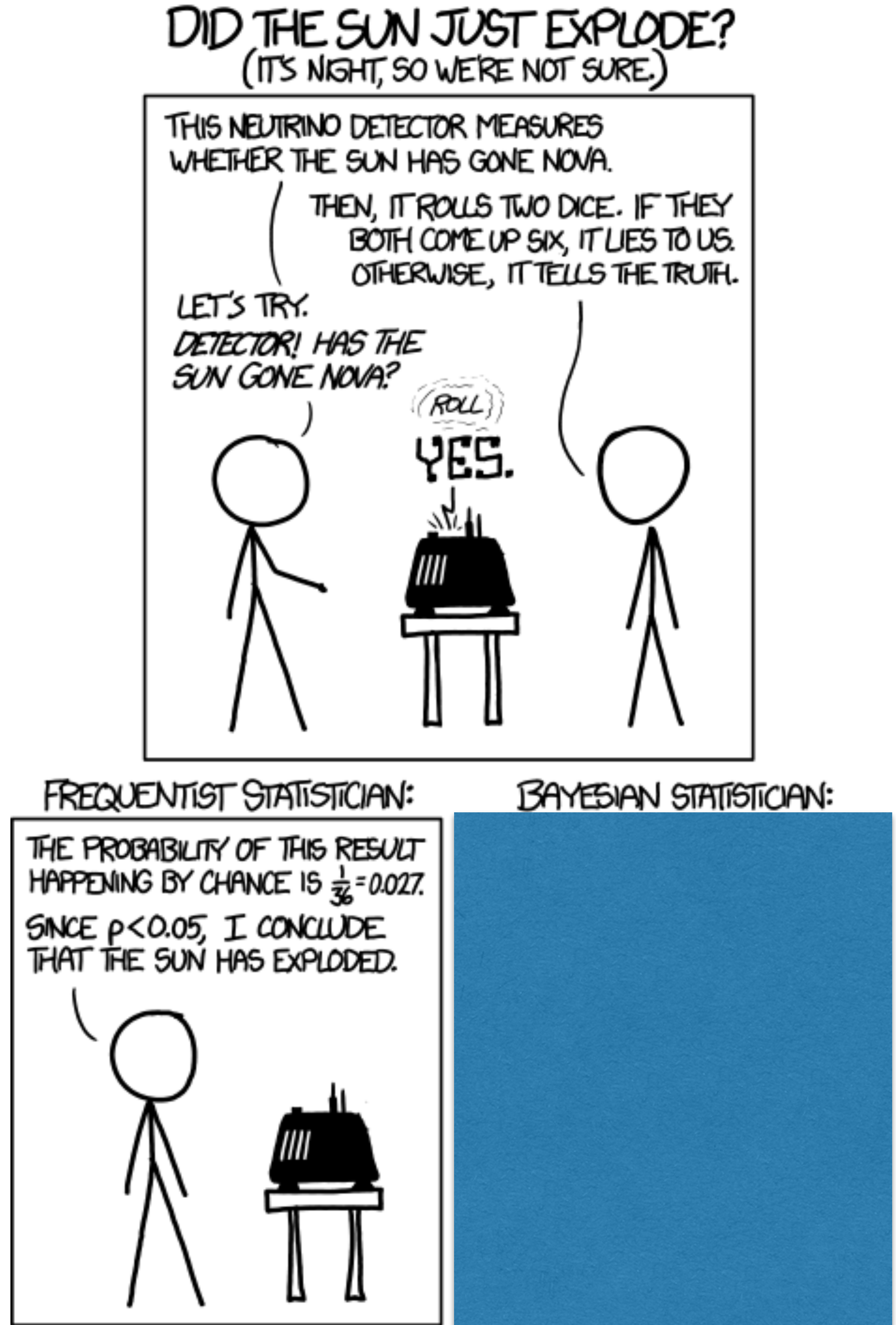
- Elude los *p values*

- Responde en términos de distribuciones de probabilidad (nuestro *grado de creencia*)

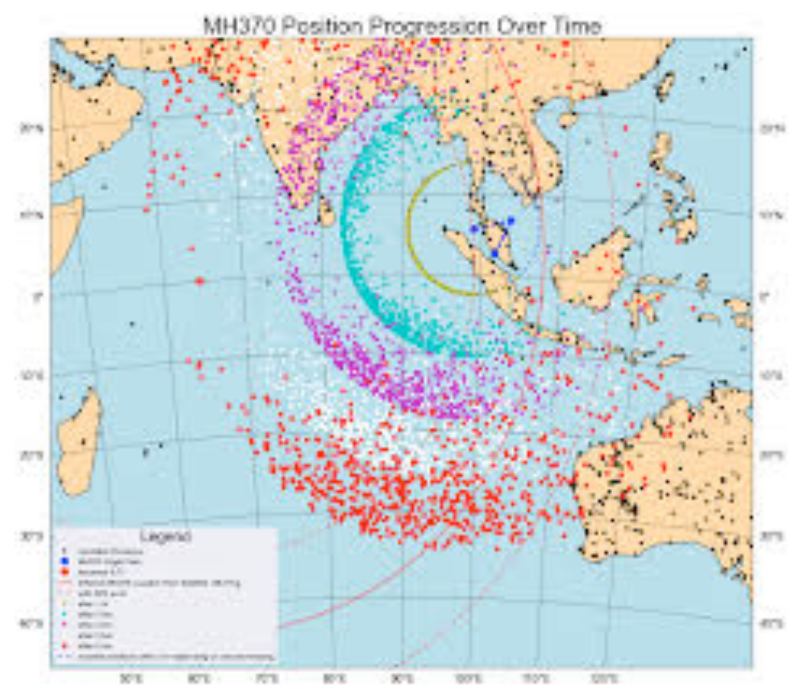
- Datos secuenciales: modelo de aprendizaje

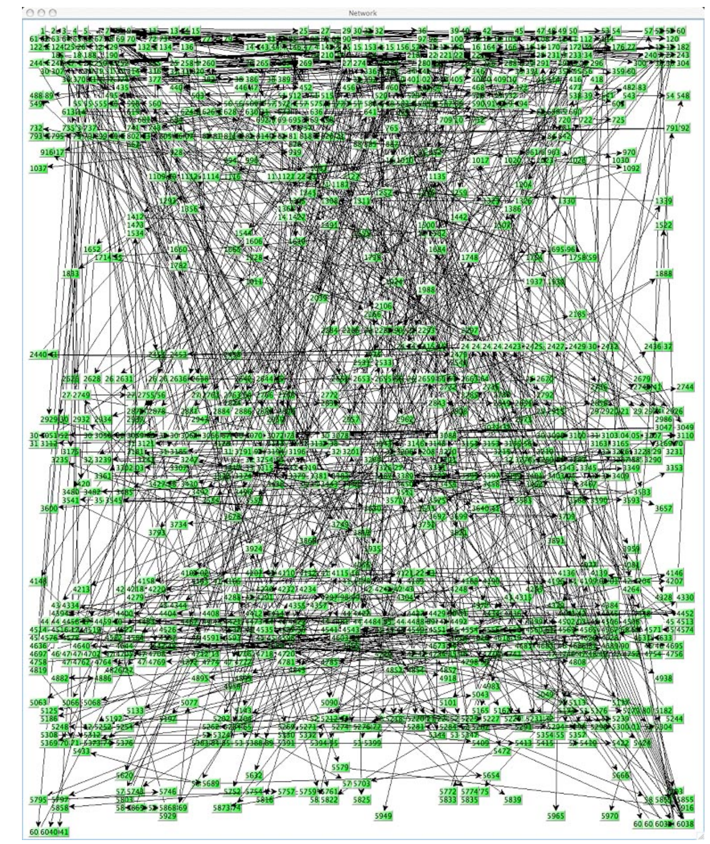
- Redes, modelos jerárquicos, estructura x estadística

"Formalización del sentido común"


DID THE SUN JUST EXPLODE?
(IT'S NIGHT, SO WE'RE NOT SURE.)
THIS NEUTRINO DETECTOR MEASURES WHETHER THE SUN HAS GONE NOVA.
THEN, IT ROLLS TWO DICE. IF THEY BOTH COME UP SIX, IT LIES TO US. OTHERWISE, IT TELLS THE TRUTH.
LET'S TRY.
DETECTOR! HAS THE SUN GONE NOVA?
ROLL
YES.
FREQUENTIST STATISTICIAN:
THE PROBABILITY OF THIS RESULT HAPPENING BY CHANCE IS 1/36 = 0.027.
SINCE p<0.05, I CONCLUDE THAT THE SUN HAS EXPLODED.
BAYESIAN STATISTICIAN:

¿Por qué ahora?
Bayes/Laplace siglo XVIII
Keynes 1920s
de Finetti 1930, *Probabilismo*





computadoras + algoritmos

↓

uso práctico

inferencia bayesiana vs. estadística frecuentista

abordaje moderno: convivencia

- proyectos de largo plazo, con un especialista en el campo
- fuerte en modelado
- ¿más 'honesta'?

- *software bundles*
- uso repetido
- modelado mínimo

Pero entonces…

¿diríamos que la moneda está cargada?

¡Usemos Bayes!

$$p(\text{moneda honesta}|6 \text{ caras})$$

Repaso de probabilidad

probabilidad conjunta, condicional

$$p(\theta, y) = p(\theta|y)p(y)$$

simétrica

$$p(\theta, y) = p(y|\theta)p(\theta)$$

marginalización

$$p(y) = \int p(y, \theta)d\theta$$

simétrica

$$p(\theta) = \int p(y, \theta)dy$$

θ discreta

$$p(y) = \sum_{\theta} p(y, \theta)$$

Independencia

$$p(A,B) = p(A)p(B)$$

Independencia Condicional

$$p(A,B|C) = p(A|C)p(B|C)$$

$$\not\Rightarrow p(A,B) = p(A)p(B) \qquad \Rightarrow p(A|B,C) = p(A|C)$$

“Desarmando” la conjunta…

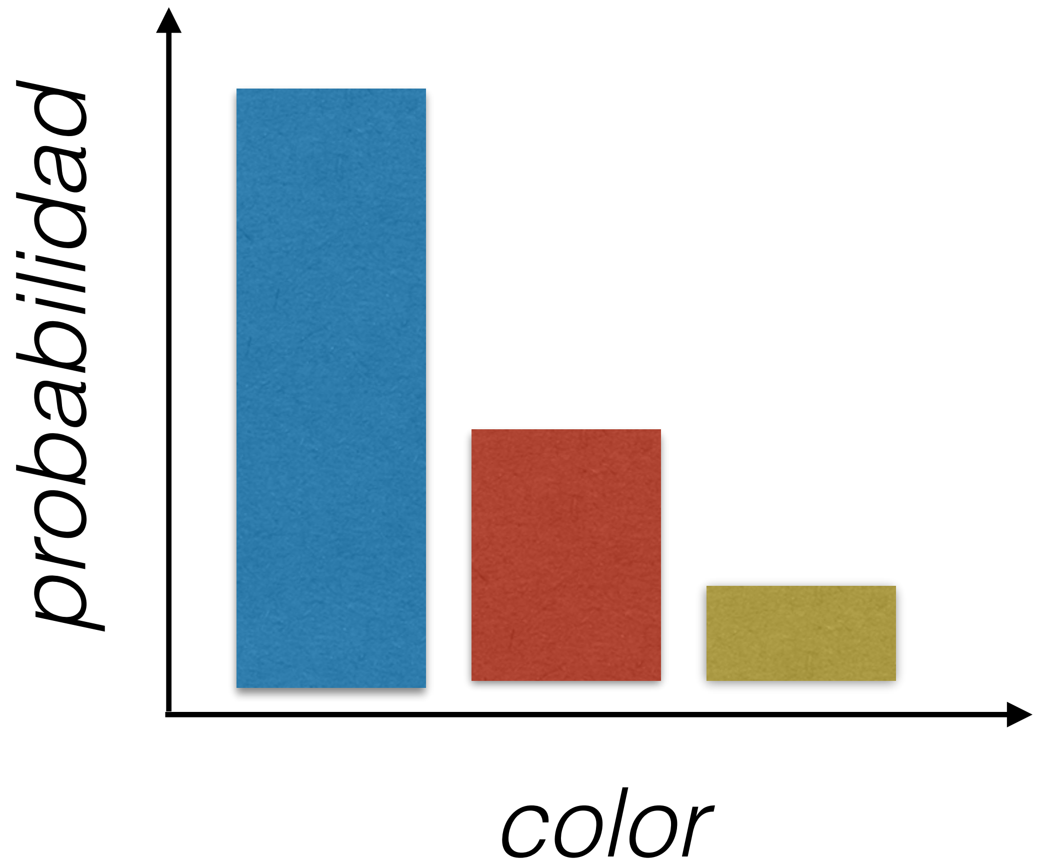
$$p(A,B,C) = p(A|B,C)p(B|C)p(C)$$

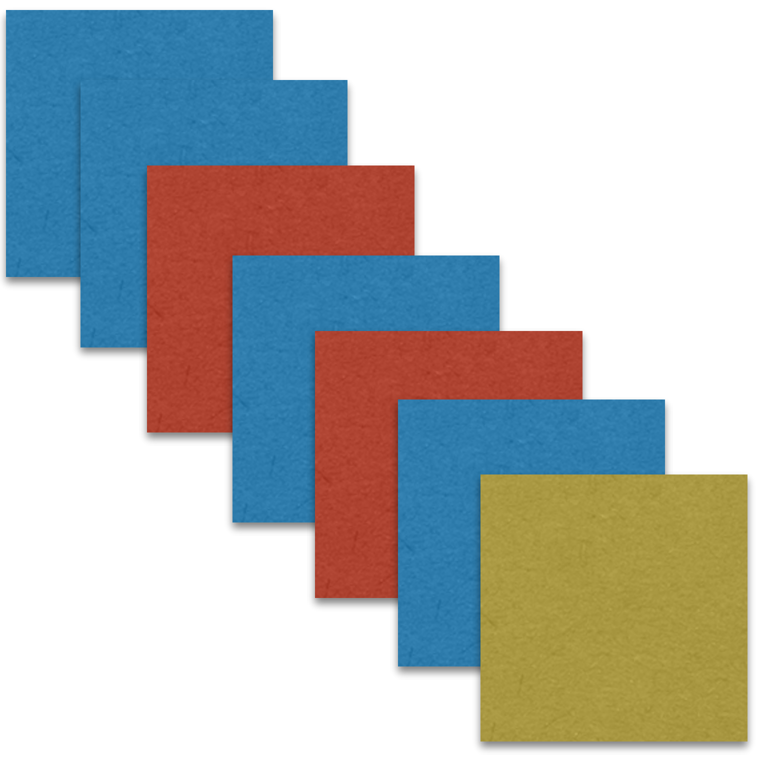
$$p(A,B,C) = p(B|A,C)p(A|C)p(C)$$

$$p(A,B,C) = p(C|A,B)p(B|A)p(A)$$

$$\cdots$$

Muestras *vs.* distribuciones





$x \sim \mathrm{Distrib}$ (“x” distribuida como …)

$p(x) = \mathrm{pdf}(\mathrm{Distrib})$

Con muchas muestras, puedo *aproximar* la pdf/pmf…

Teorema de Bayes

$$p(H|D) = \frac{p(D|H)p(H)}{p(D)} = \frac{p(D|H)p(H)}{\int p(D|H)p(H)dH}$$

$$p(H|D) \propto p(D|H)p(H)$$

posterior

likelihood

prior

$p(\text{moneda honesta}|6 \text{ caras})$

?

?

Likelihood: distribución Binomial

$$\overset{\textstyle D \mid H}{\uparrow}$$

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

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

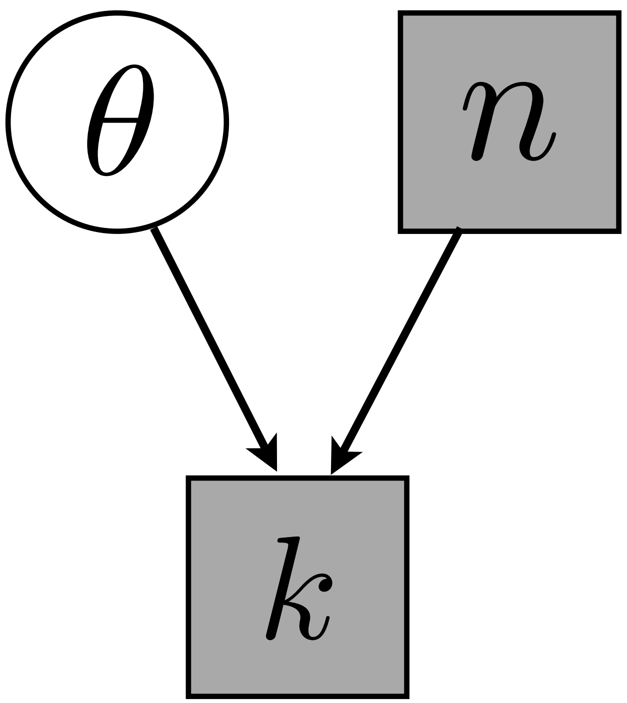
k sucesos de *n* “tiradas”, con probabilidad de suceso theta

Modelo Beta-Binomial

likelihood:

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

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

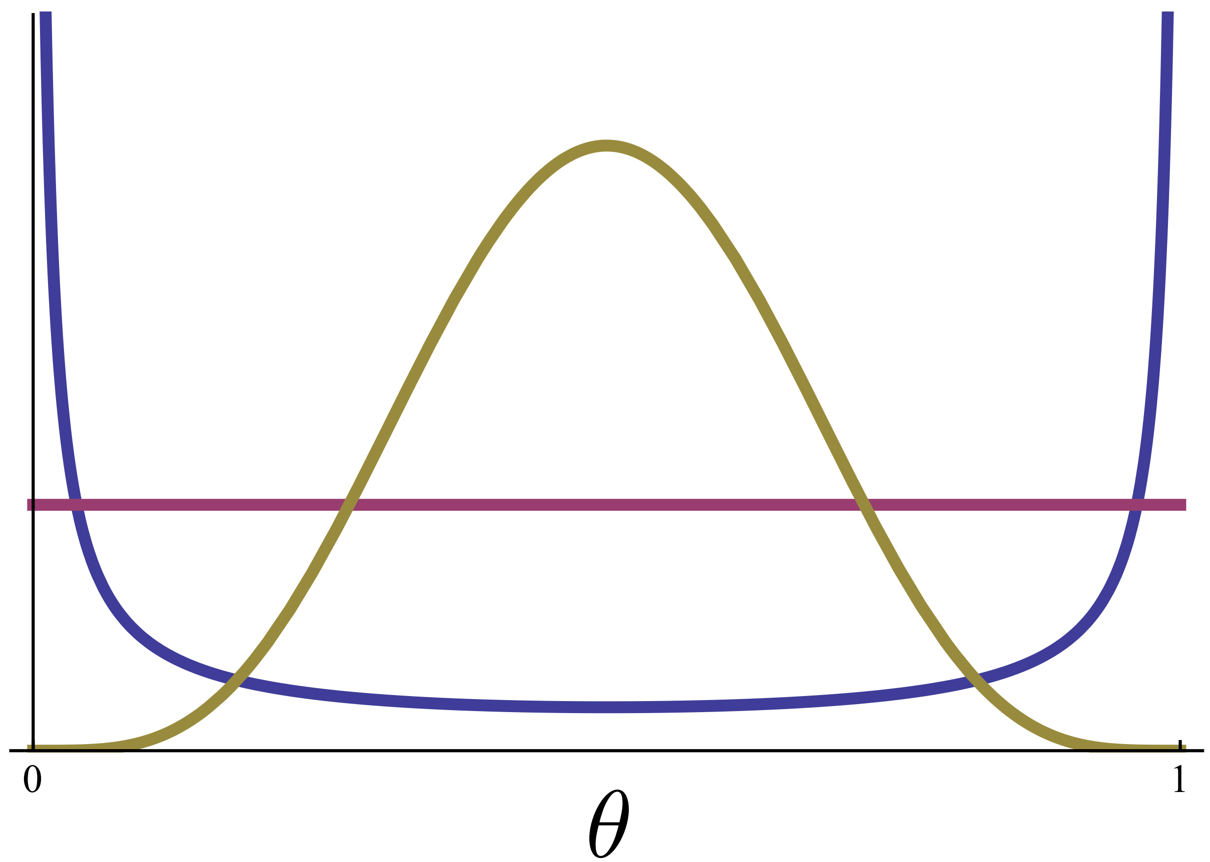


prior:

$p(\theta)$ ...

Prior: distribución Beta

$$p(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{\mathrm{B}(\alpha,\beta)} \qquad \alpha, \beta > 0 \qquad \theta \in [0,1]$$



$$\mathrm{Beta}(\alpha, \alpha)$$

α

0.01

1

5

$$\mathrm{E}[\theta] = \frac{\alpha}{\alpha+\beta}$$

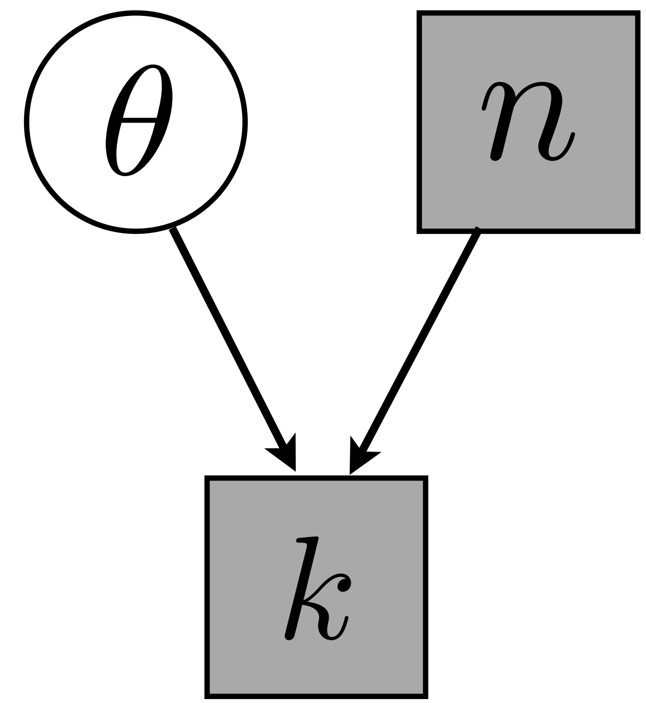
¿Cómo cambia si los parámetros difieren?

Modelo Beta-Binomial

likelihood:

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

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

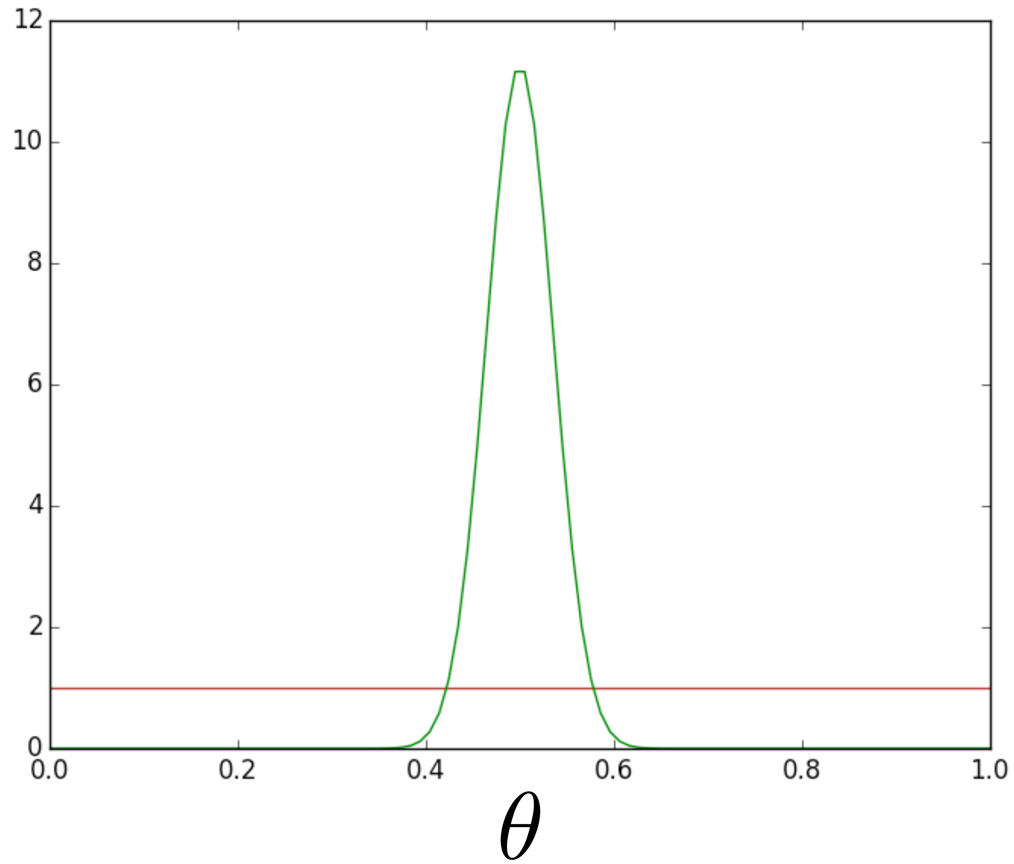


prior:

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

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

posterior con Bayes: $$p(\theta|k) = \frac{p(k|\theta)p(\theta)}{p(k)}$$



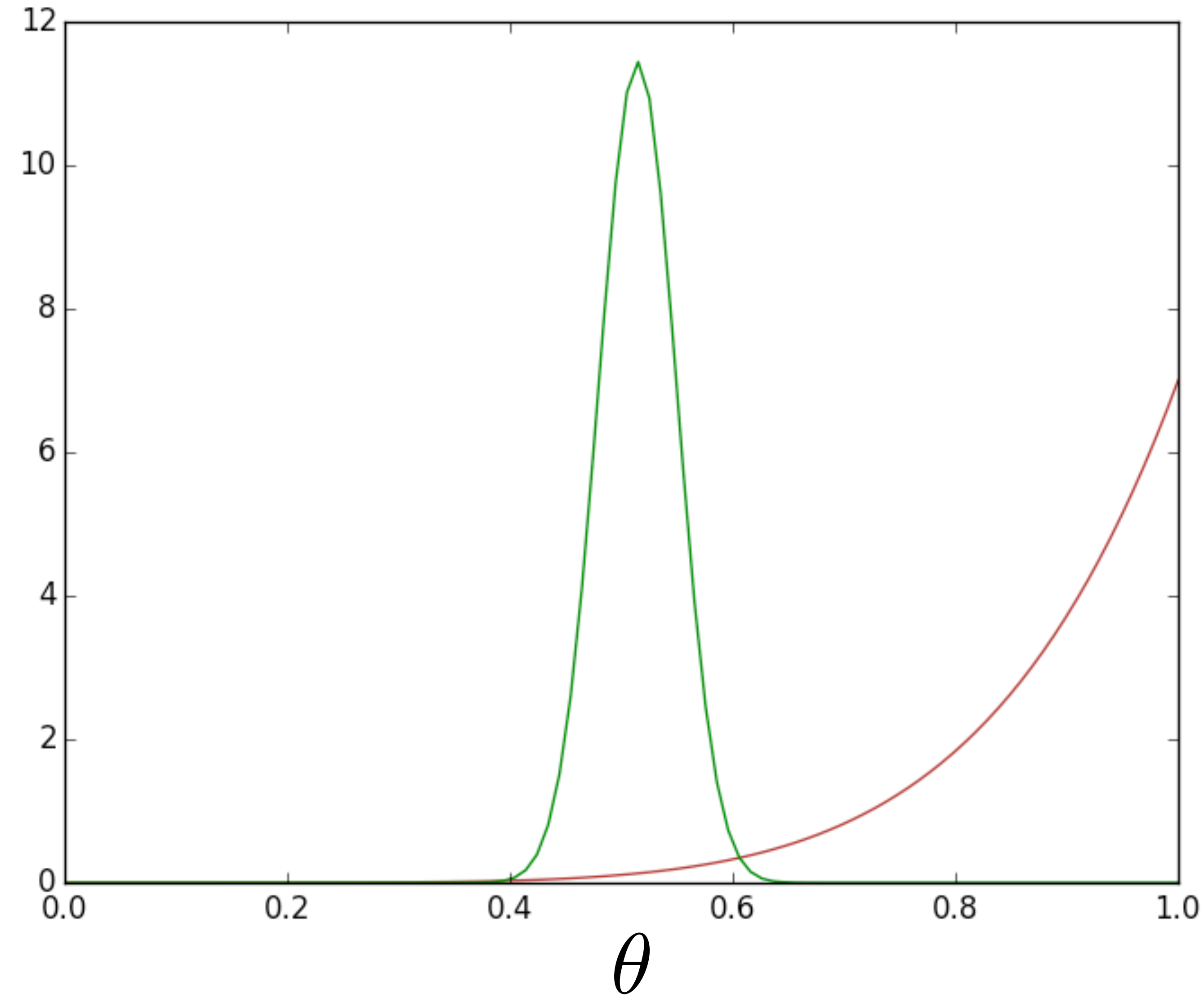
prior

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

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

posterior

(luego de 6 caras)



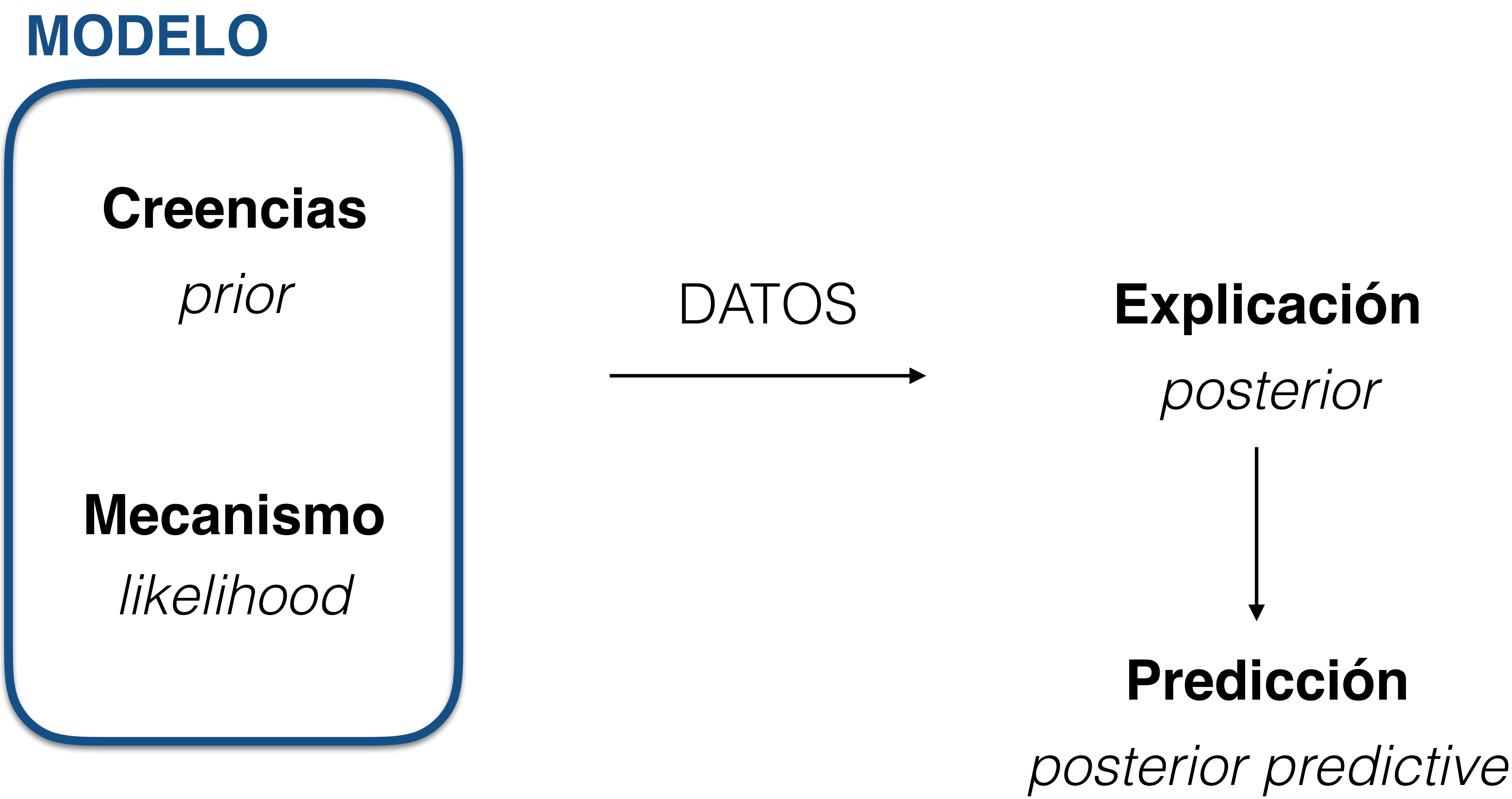
¡¿Y ENTONCES?!

¿diríamos que la moneda está cargada?

¡Depende del *prior*!

¿Y está bien esto? ¡Sí! Va a cambiar con quién me de la moneda, etc.. Es más:

No hay inferencia sin hipótesis


MODELO
Creencias
prior
Mecanismo
likelihood
DATOS
Explicación
posterior
Predicción
posterior predictive

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

posterior *likelihood* *prior*

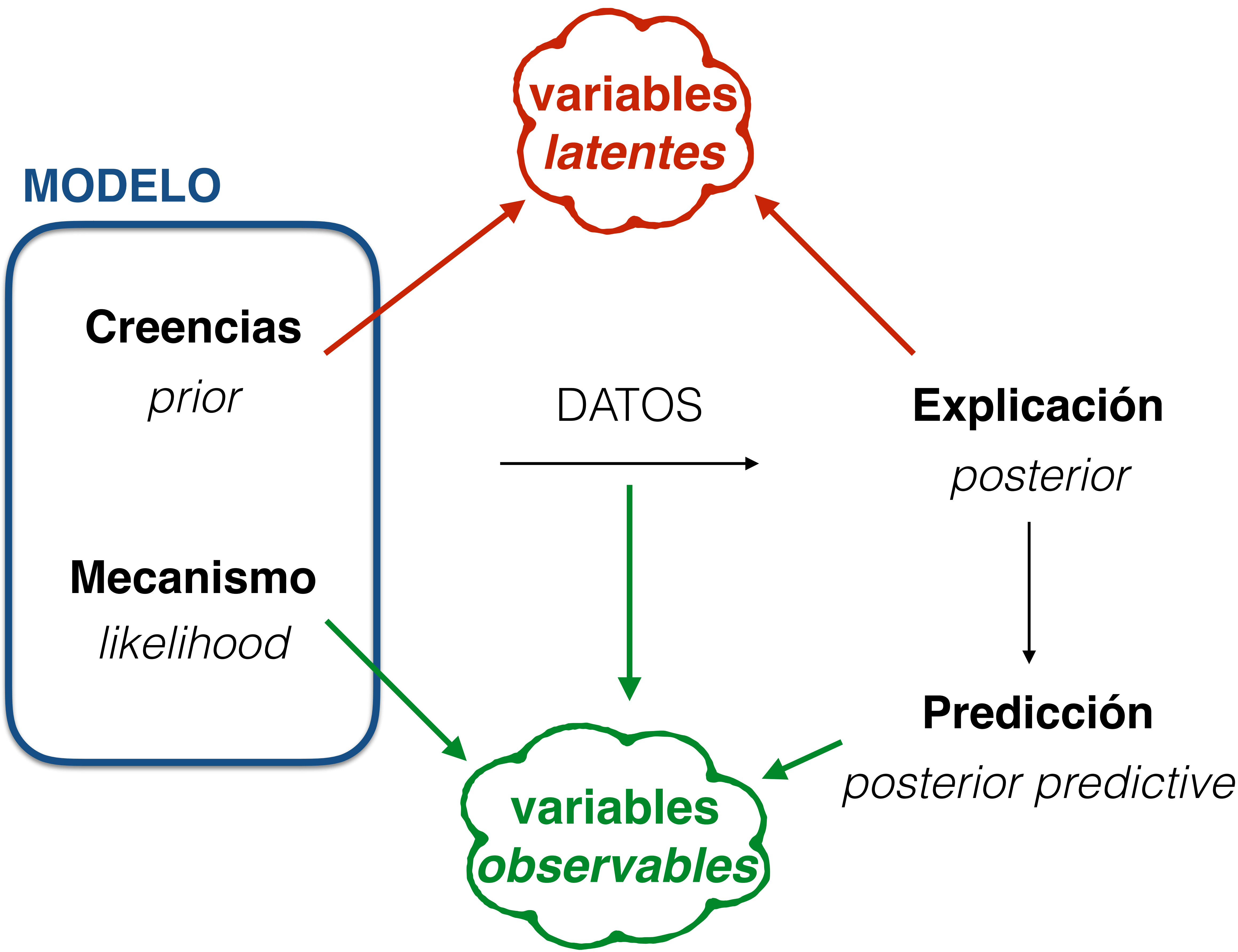
prior predictive

$$p(y) = \int p(y, \theta) d\theta = \int p(y|\theta)p(\theta) d\theta$$

posterior predictive

$$\begin{aligned} p(\tilde{y}|y) &= \int p(\tilde{y}, \theta|y) d\theta \\ &= \int p(\tilde{y}|\theta, y)p(\theta|y) d\theta \\ &= \int p(\tilde{y}|\theta)p(\theta|y) d\theta \end{aligned}$$

promedio con mi *posterior!*


variables
latentes
MODELO
Creencias
prior
DATOS
Explicación
posterior
Mecanismo
likelihood
Predicción
posterior predictive
variables
observables

pymc3

- http://docs.pymc.io/
- Modelado Bayesiano y Machine Learning Estadístico
-
```
import pymc3 as pm
my_model = pm.Model()
with my_model:
    theta = pm.Beta(…)
    …
    trace = pm.sample()
```