Fundamentos de Inferencia Bayesiana

2017

Operacionales

- Horario
- Programa
- Evaluación
- Puntos, licenciatura/doctorado
- Correlativas
- ¿Otras?

Bibliografía

- Gelman *et al.*
  Bayesian Data Analysis

- MacKay
  Information Theory, Inference, and Learning Algorithms

- Jaynes
  Probability Theory: The Logic of Science

- Barber
  Bayesian Reasoning and Machine Learning

- Hacking
  An Introduction to Probability and Inductive Logic

- Wagenmakers & Lee
  A Course in Bayesian Graphical Modeling for Cognitive Science

- Marr, Anderson, ...

Experimento mental *(gedankenexperiment)*

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

¿diríamos que la moneda está cargada?

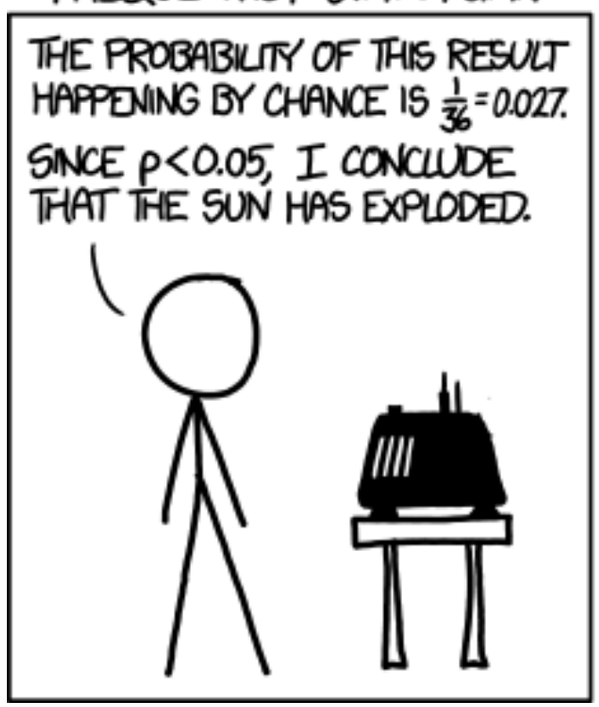
La “ciencia” hoy dice “sí”.

Más precisamente:

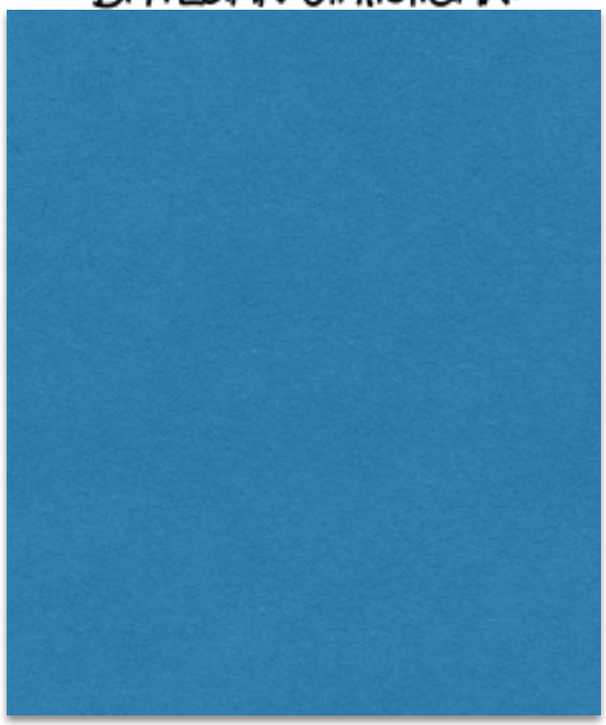
“Hay evidencia significativa de que está cargada ($p < 0.05$, test binomial)”

DID THE SUN JUST EXPLODE?
(IT'S NIGHT, SO WE'RE NOT SURE.)

FREQUENTIST STATISTICIAN:



BAYESIAN STATISTICIAN:



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

Permite (entre otras cosas) incorporar nuestro conocimiento previo

Probabilidad como *creencia*

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

inferencia bayesiana vs. estadística frecuentista

mayormente

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

subjetiva

objetiva (intersubjetiva)

interpretación de la probabilidad

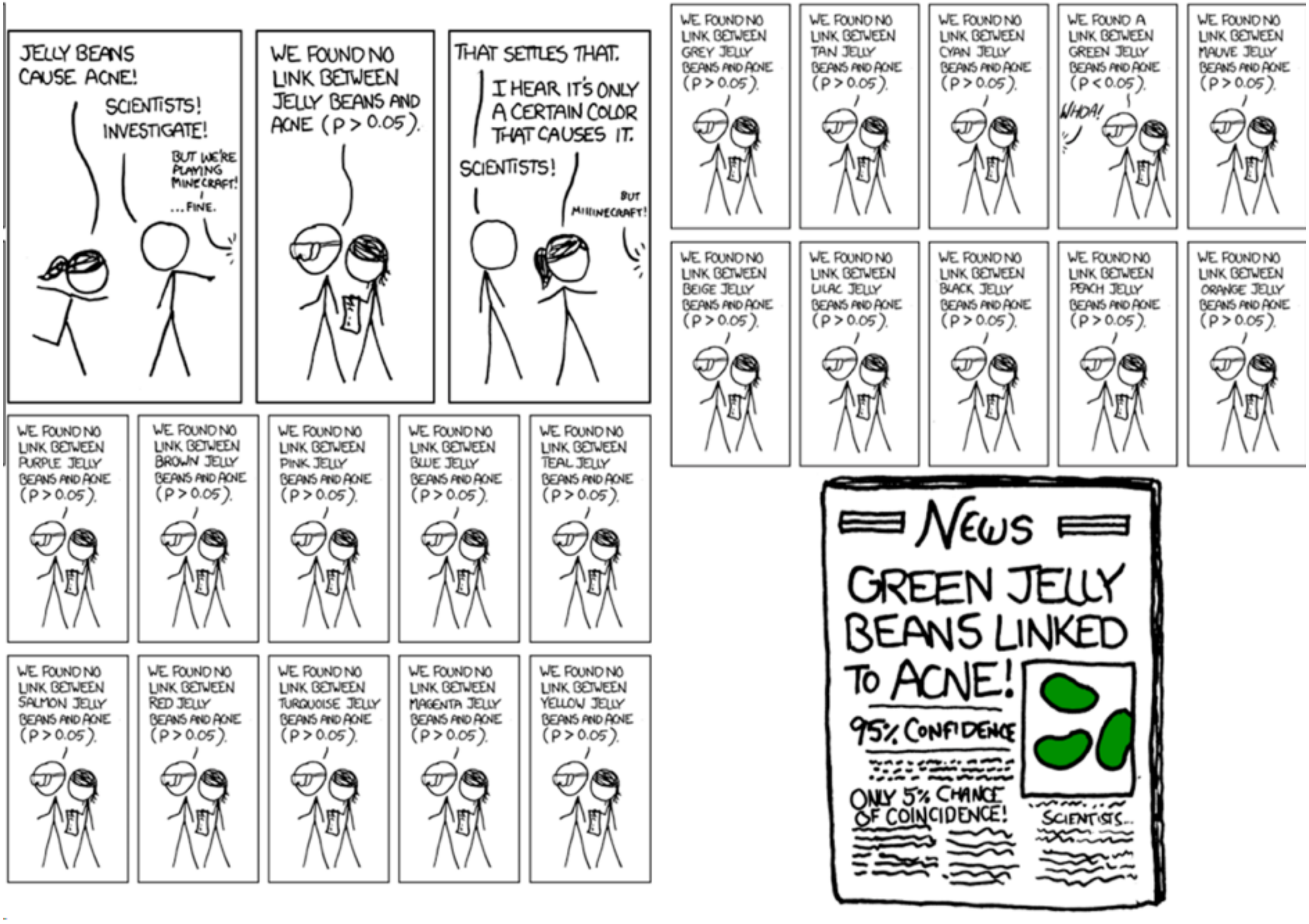
p(*lo que yo quiera*)

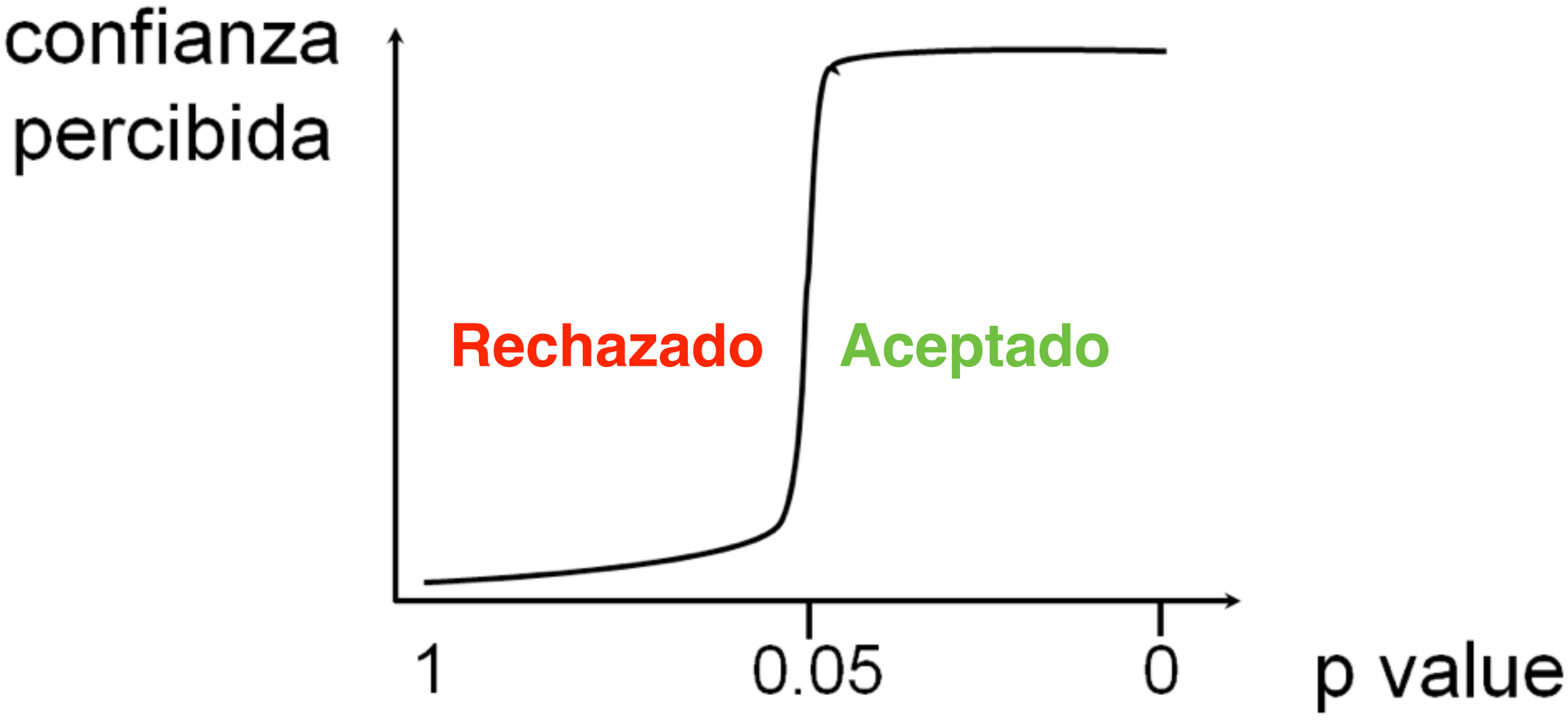
p sólo para muestreo

Críticas al frecuentismo: 2 niveles

- críticas de fondo:
  - ignora conocimiento previo
  - dilema (mezcla) Fisher (significance testing) vs. Neyman/Pearson (hypothesis testing)
  - ¿inconsistente? (discusión todavía abierta)

- críticas al (ab)uso:
  - *p hacking*
  - uso ciego en general
  - no específico del frecuentismo


JELLY BEANS CAUSE ACNE!
SCIENTISTS! INVESTIGATE!
BUT WE'RE PLAYING MINECRAFT!
... FINE.
WE FOUND NO LINK BETWEEN JELLY BEANS AND ACNE (P > 0.05).
THAT SETTLES THAT.
I HEAR IT'S ONLY A CERTAIN COLOR THAT CAUSES IT.
SCIENTISTS!
BUT MINECRAFT!
WE FOUND NO LINK BETWEEN GREY JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN CYAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND A LINK BETWEEN GREEN JELLY BEANS AND ACNE (P < 0.05).
WHOA!
WE FOUND NO LINK BETWEEN MAUVE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BEIGE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN LILAC JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLACK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PEACH JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN ORANGE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PINK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLUE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN SALMON JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN RED JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TURQUOISE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN MAGENTA JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN YELLOW JELLY BEANS AND ACNE (P > 0.05).
NEWS
GREEN JELLY BEANS LINKED TO ACNE!
95% CONFIDENCE
ONLY 5% CHANCE OF COINCIDENCE!
SCIENTISTS...


confianza
percibida
Rechazado
Aceptado
1
0.05
0
p value

Inferencia Bayesiana

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

$$\Downarrow$$

posterior *likelihood* *prior*

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

Teorema de Bayes

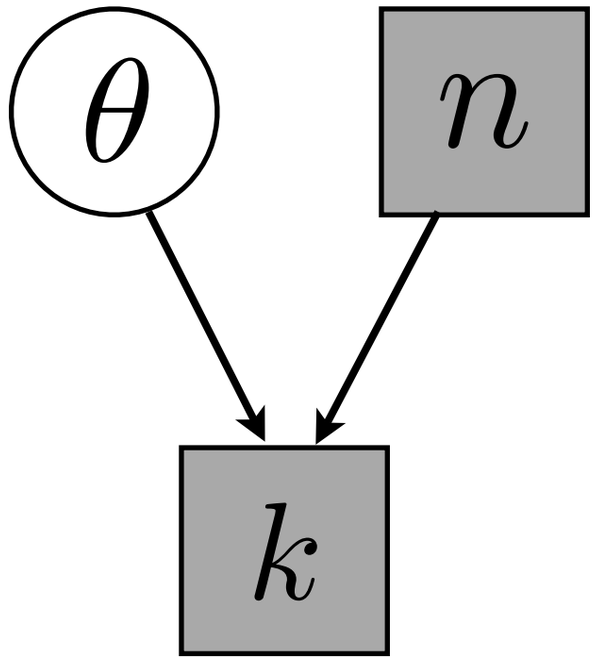
Inferencia Bayesiana

Modelo

likelihood:

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

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

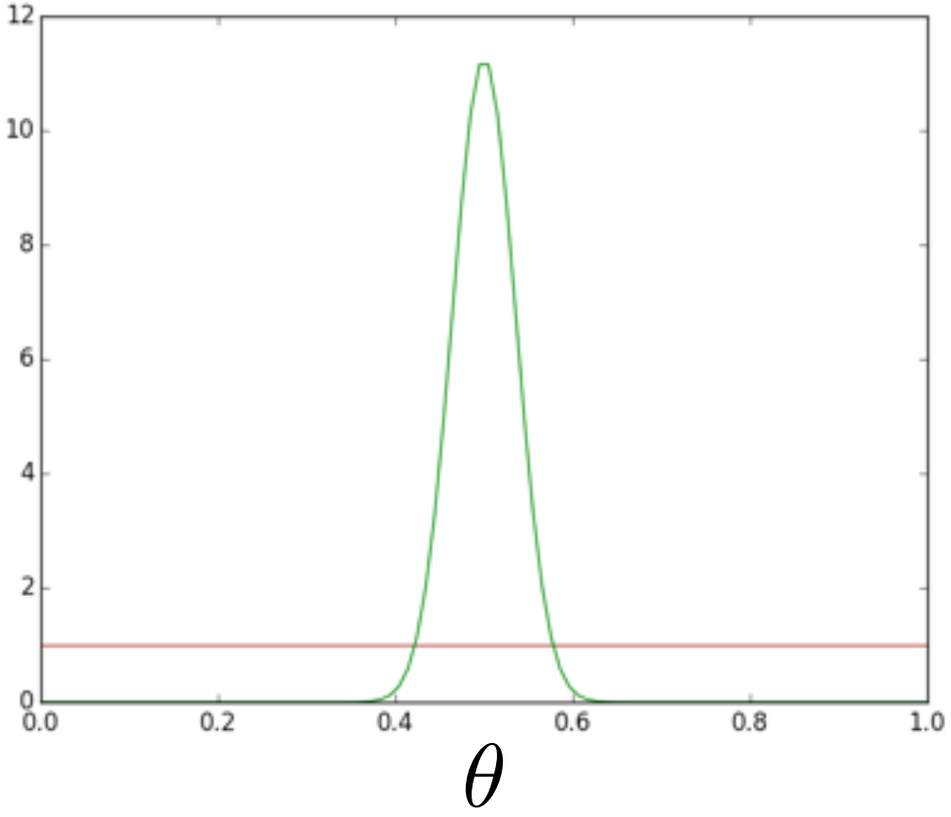


prior:

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

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

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

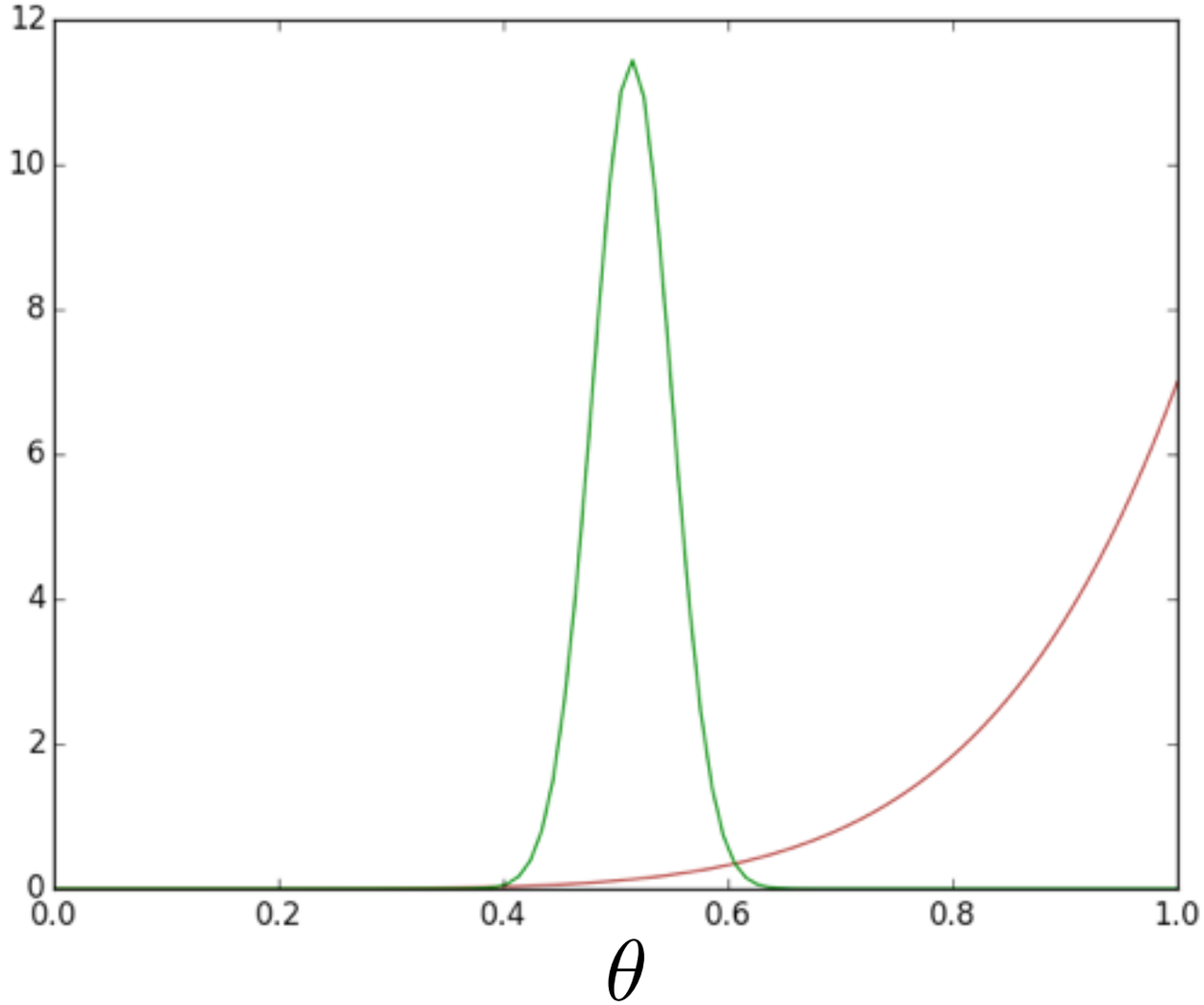


prior

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

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

posterior
(luego de 6 caras)



Acumulación de Evidencia

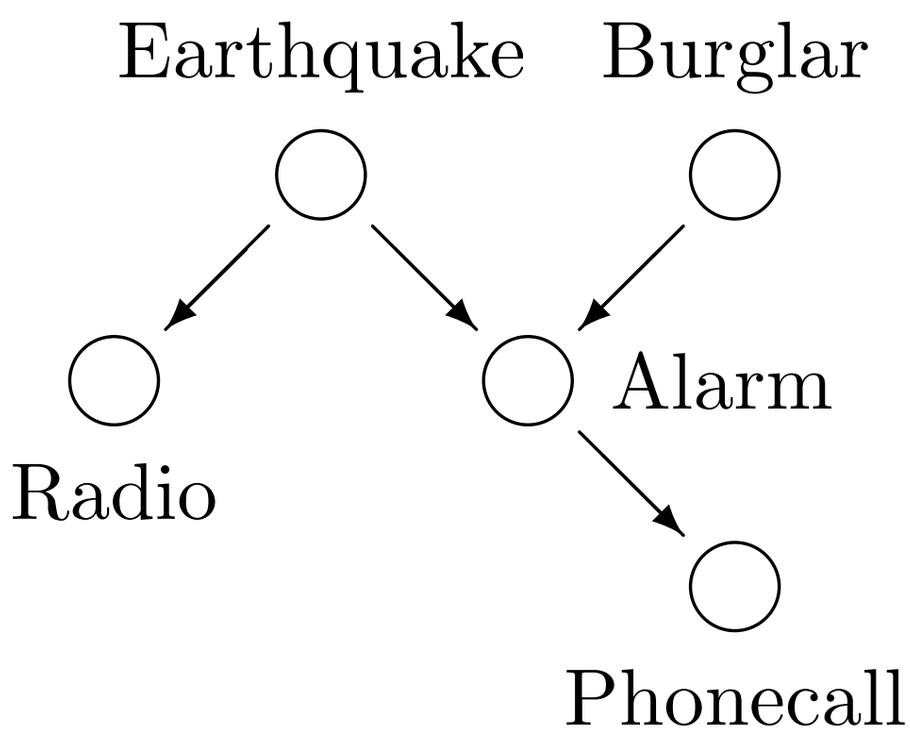
$$p(H|D_1) = \frac{p(D_1|H)P(H)}{p(D_1)}$$

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

"The posterior is the new prior"

Redes Bayesianas

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



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

Modelos Jerárquicos

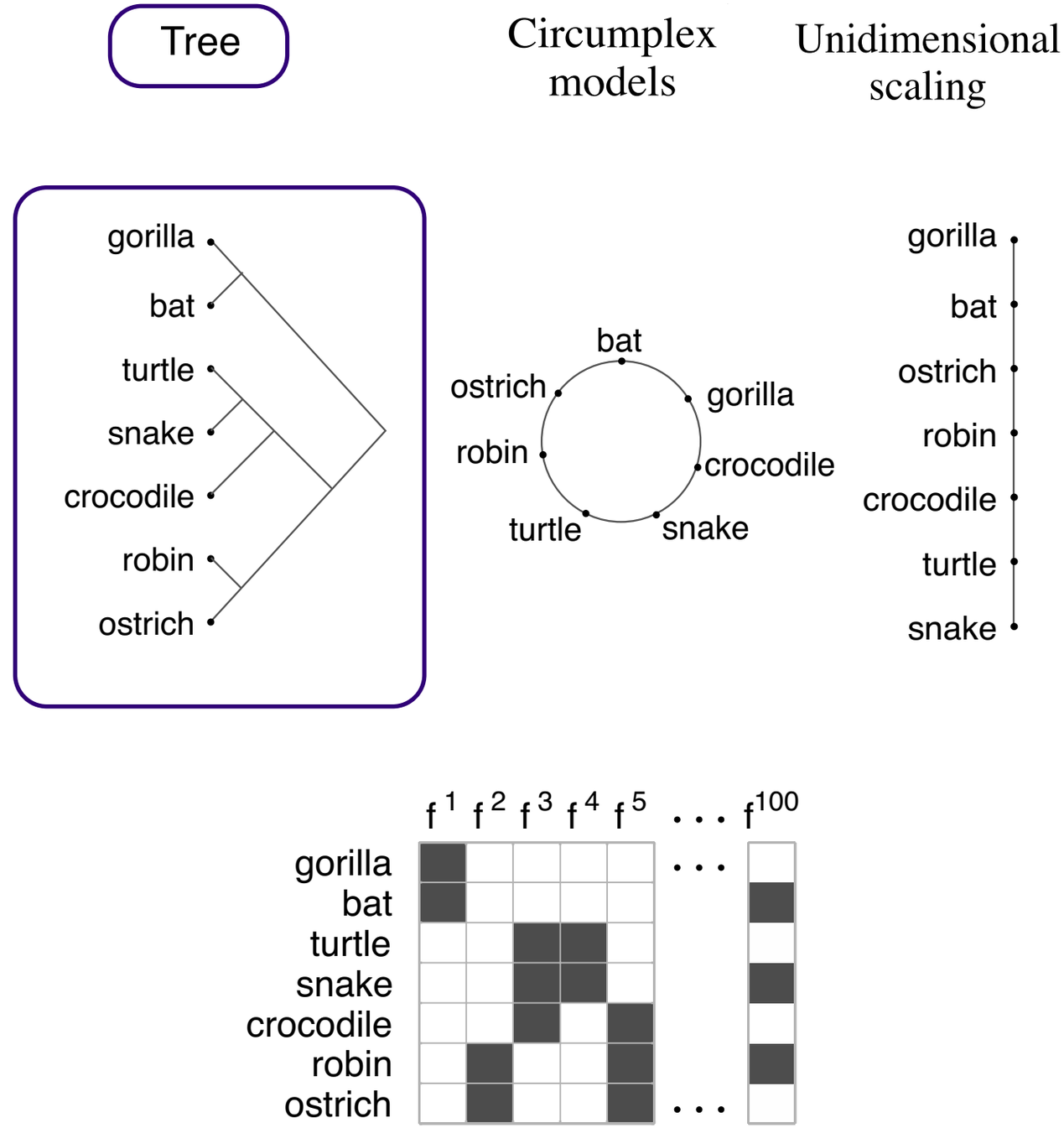
Teorías

↓

Hipótesis

↓

Datos



Inferencia Bayesiana

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

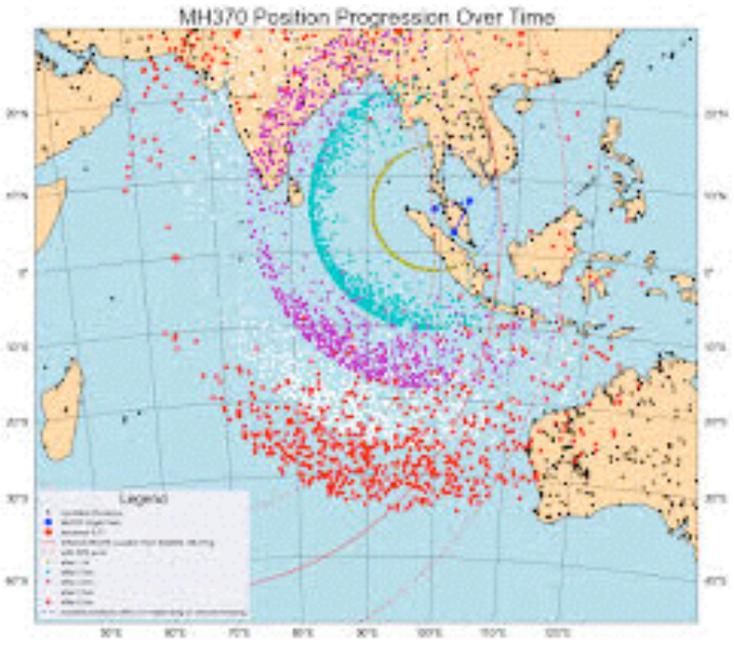
- Elude los *p values*

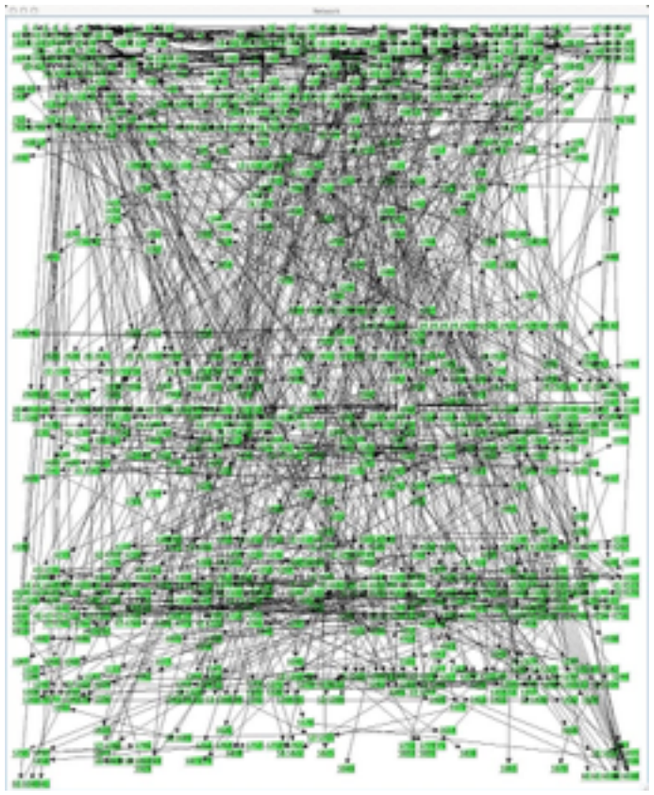
- Responde en términos de distribuciones de probabilidad (nuestro *grado de creencia*)

- Datos secuenciales: modelo de aprendizaje

- Redes, modelos jerárquicos, estructura x estadística

¿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… ¿por qué probabilidades? ¿por qué Bayes?

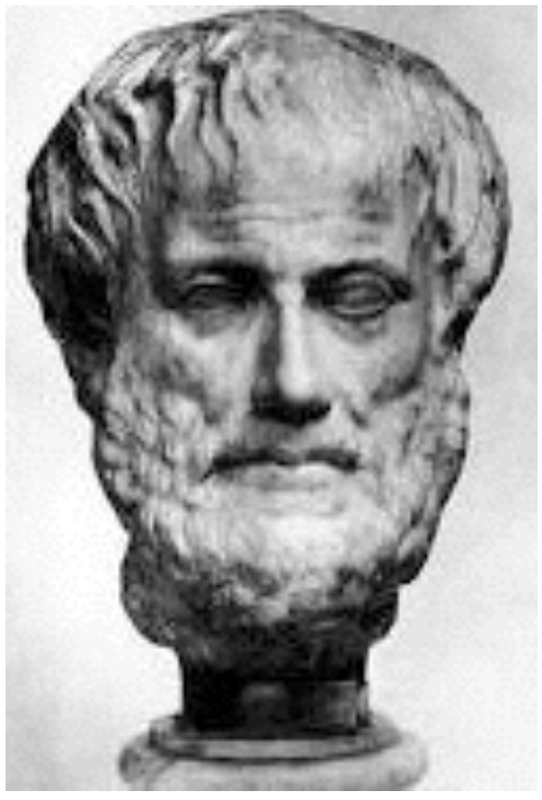
Dutch book arguments -coherencia

Axiomas de Cox para los niveles de creencia:
1) Representados por un número real
2) “Sentido común” (inc. lógica Aristotélica)
3) Consistencia

Alternativamente:
estadística en términos de toma de decisiones,
minimizar la péridida/maximizar la utilidad

Racionalidad en contextos de incertidumbre...
¡Modelo del pensamiento humano!

Desde Aristóteles…



Lógica

Todo **hombre** es **mortal**
Sócrates es **hombre**
Ergo **Sócrates** es **mortal**

Planteada como modelo del pensamiento

pero…

La Paradoja de Linda

$$p(cajera) \not\geqslant p(cajera \,\&\, feminista)$$

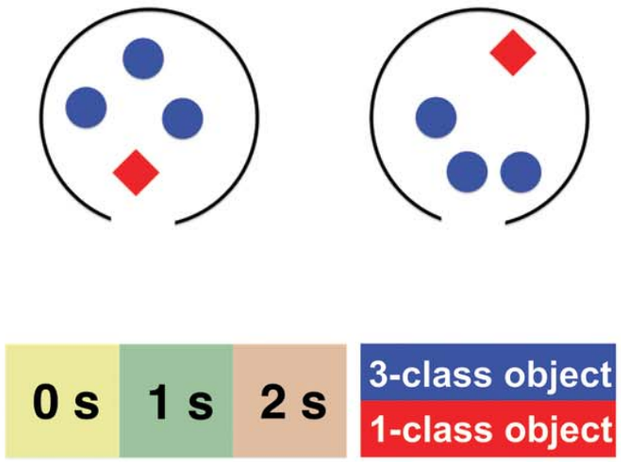
(racional)

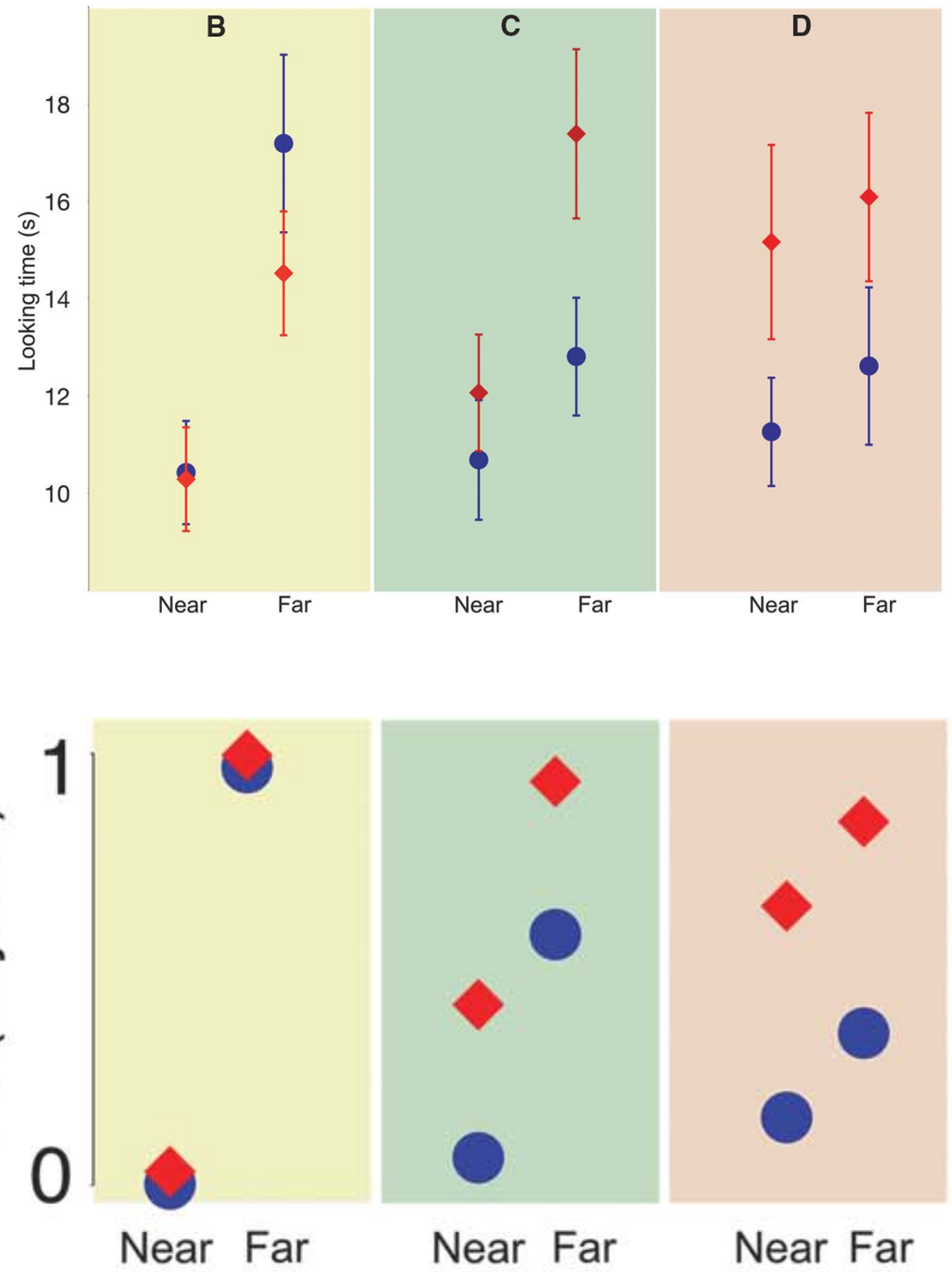
Sesgos y Heurísticas, hombre irracional
(Tversky & Kahneman, 1980s)

Programa de la Cognición Bayesiana

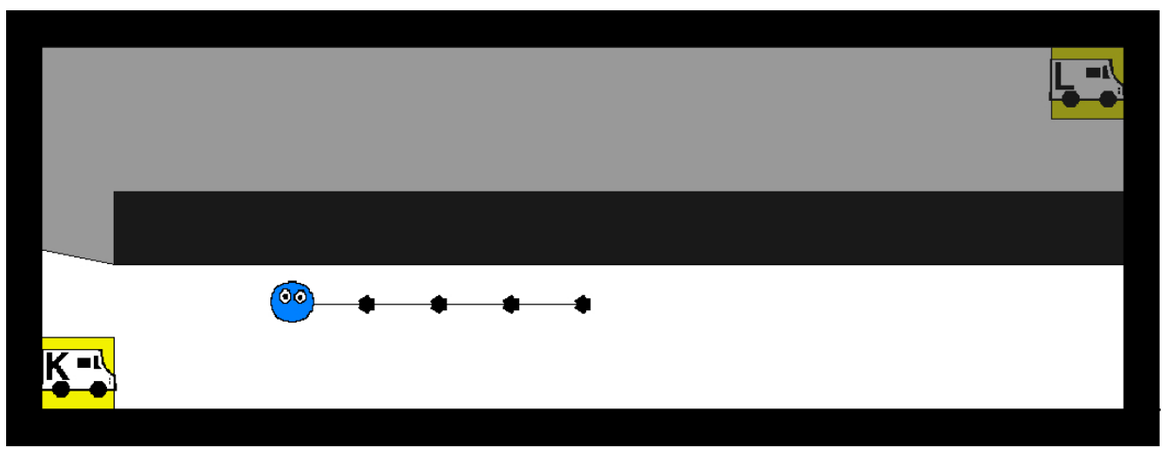
- Lógica es adecuada en contextos de **certidumbre**
- Cuando hay **incertidumbre**, el lenguaje racional es la teoría de probabilidad
- **Racionalidad acotada** por nuestros recursos de cómputo: ilusión de irracionalidad
- Programa **general** y **cuantitativo**

Inferencia Causal

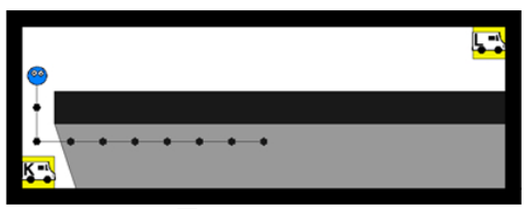




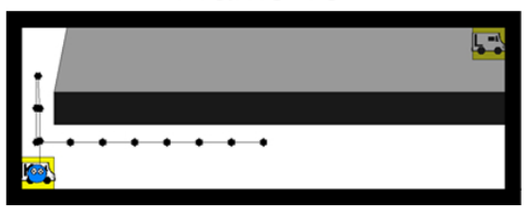
Teoria de la Mente (*joint Belief-Desire*)



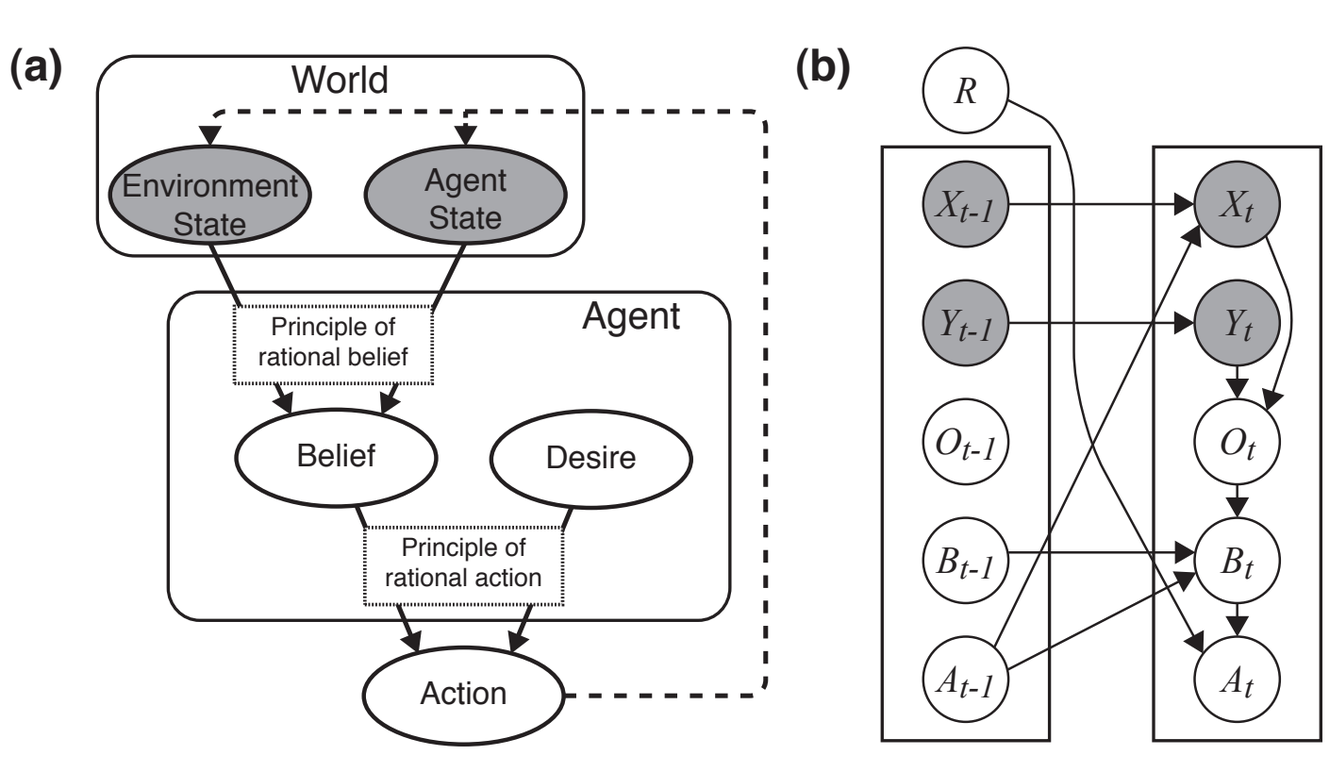
Frame 5

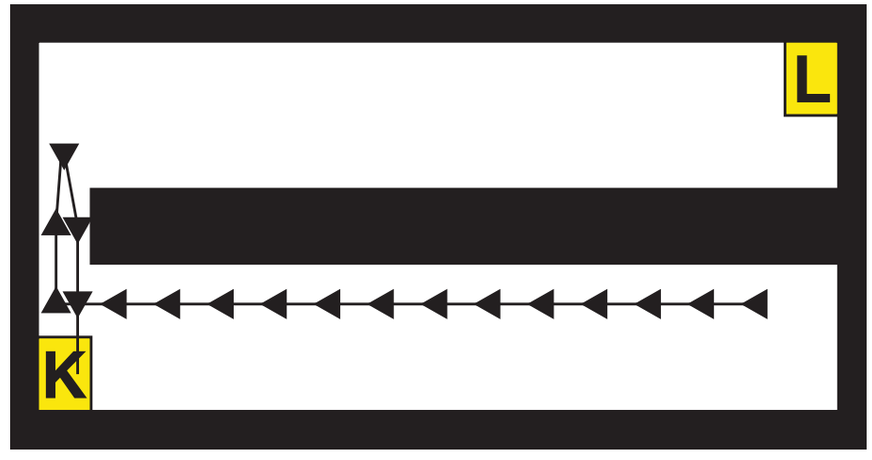


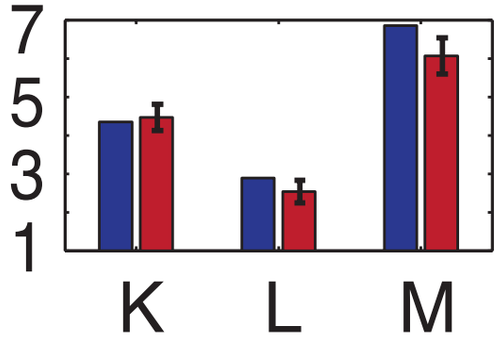
Frame 10



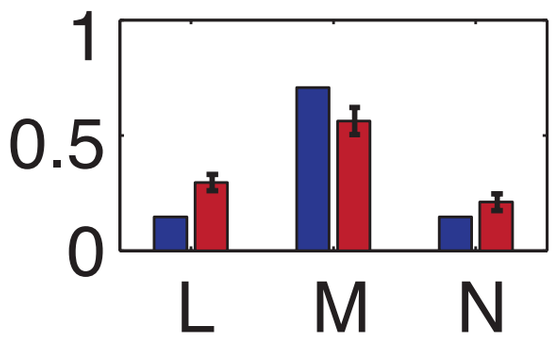
Frame 15







Desires

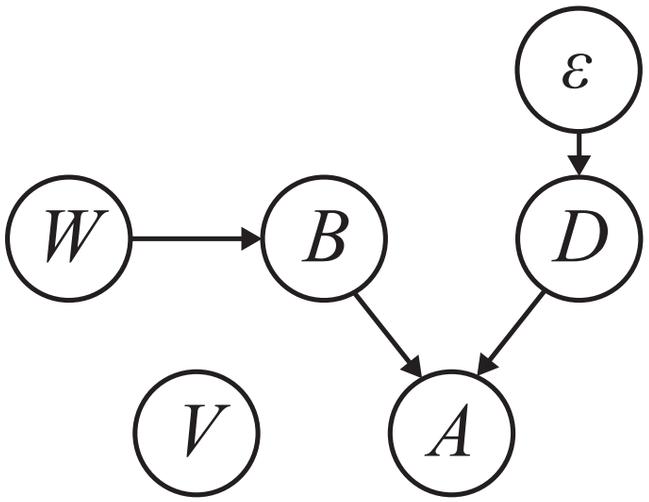


Beliefs

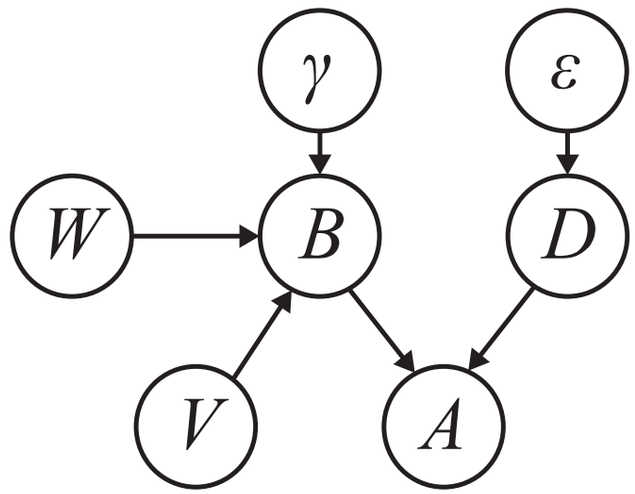
False Belief

Variable	Description	States
World (W)	Location of the toy.	0: Original location, 1: New location.
Access (V)	Could Sally see the toy moved?	0: No, 1: Yes.
Action (A)	Where Sally looks for her toy.	0: Original location, 1: New location.
Belief (B)	Where Sally thinks the toy is.	0: Original location, 1: New location.
Desire (D)	Sally's primary desire.	1: To find the toy, 0: Anything else.

Copy Theorist

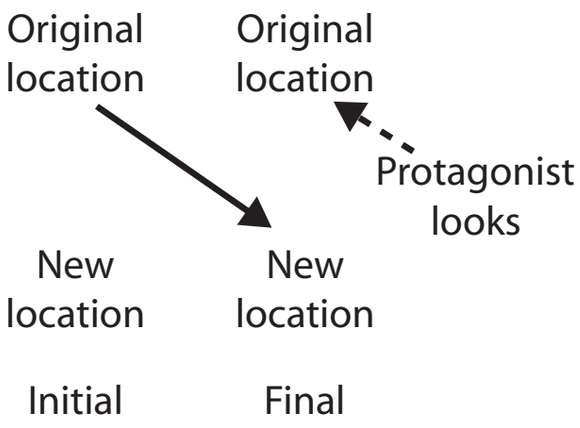


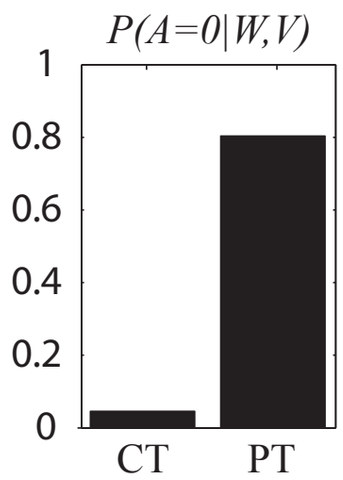
Perspective Theorist

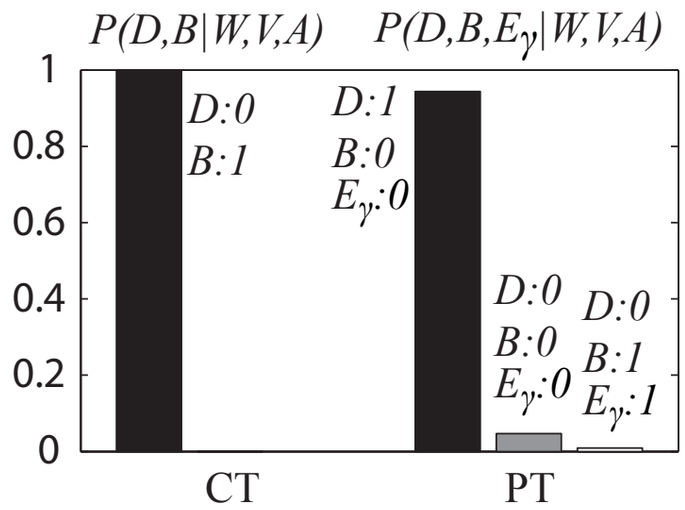


$P(A=1\|B,D)$	B	D
0	0	1
1	1	1
0.5	0	0
0.5	1	0

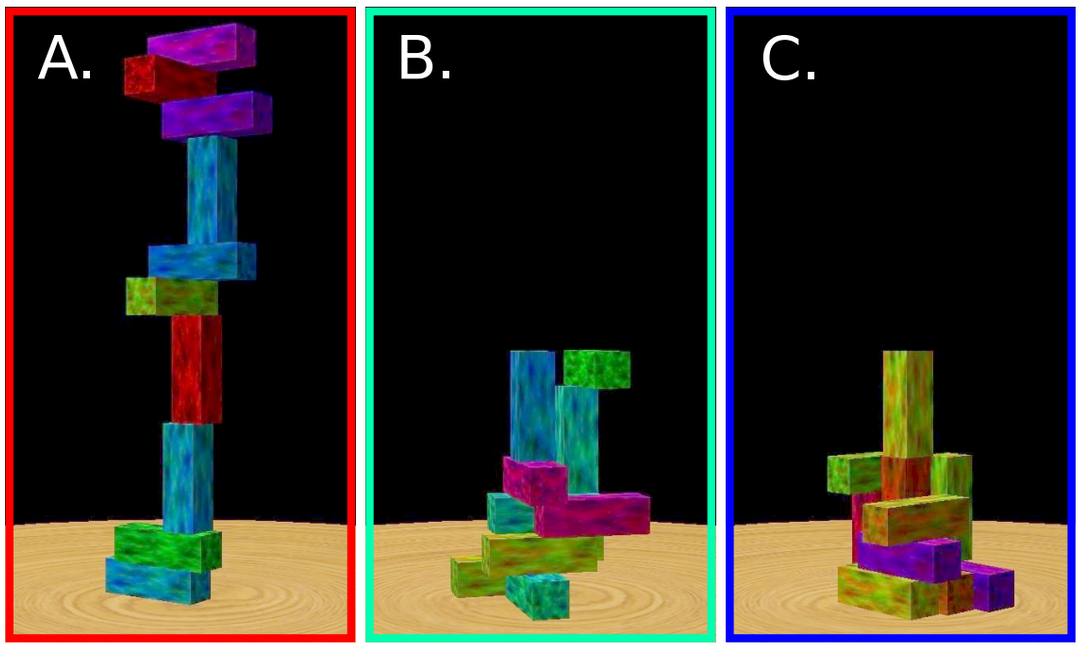
$P_{\mathrm{PT}}(B=1\|W,V)$	W	V
0	0	1
1	1	1
γ	0	0
γ	1	0

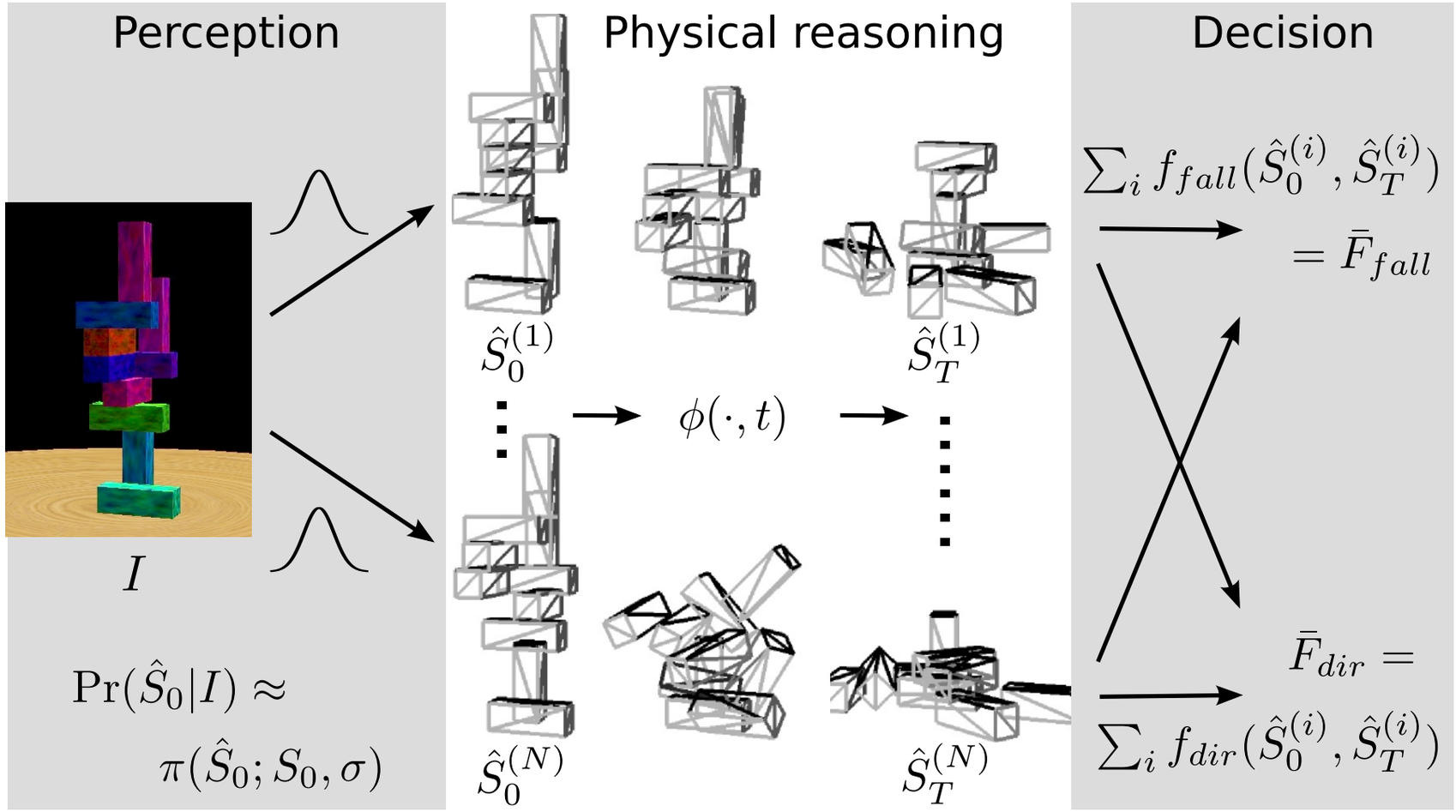






Física Intuitiva





Fundamentos de Inferencia Bayesiana

La práctica

Operacionales

- Trabajo en grupos de N < 3 (N<=2)
- Modalidad "*Hágalo Usted Mismo*"
- Mayormente en computadora
- Seguimos apunte de *Wagenmakers & Lee*
- Herramienta a elección

Herramientas: Programación Probabilística

```
INFER age, has_heart_disease FROM patients
WHERE age > 30 WITH CONFIDENCE 0.8 LIMIT 3;
```

age	has_heart_disease
66	???
44	yes
31	???

age	*has_heart_disease*
66	yes
44	*yes*
31	*???*

Lenguajes Probabilísticos

Expresivos

- Figaro (Scala)
- WebPPL (Javascript, probmods.org)

Rápidos

- STAN
- WinBUGS, JAGS (R)
- PyMC3 (Python)

- Edward (TensorFlow, edwardlib.org)
  Deep Probabilistic Programming

Venture, BayesDB, Anglican…

WinBUGS/jags

- Enlatado/open source
- Interfaz gráfica, pero se puede usar sin
- Sintaxis cercana a R
- Scripts de MATLAB y R para correr
- Usado en el apunte
- Herramienta más estándar, no muy flexible
- http://mcmc-jags.sourceforge.net/

WinBUGS/jags

```
# Difference Between Two Rates

model {

# Prior on Rates
theta1 ~ dbeta(1,1)
theta2 ~ dbeta(1,1)

# Observed Counts
k1 ~ dbin(theta1,n1)
k2 ~ dbin(theta2,n2)

# Difference between Rates
delta <- theta1-theta2

}
```

Calling matjags.m

```
%% Defining observed data
k1=0;  % number of observed successes
n1=10; % number of observations total
k2=10;
n2=10;
% Create a single structure that has the data for all observed JAGS nodes
%datastruct = struct('k',k,'n',n);
datastruct = struct('k1',k1, 'k2',k2, 'n1',n1, 'n2',n2);

%% Set initial values for latent variable in each chain
for i=1:nchains
    S.theta1 = 0.5; % An Initial Value for the Success Rate
    S.theta2 = 0.5; % An Initial Value for the Success Rate
    init0(i) = S; % init0 is a structure array that has the initial values for all latent va
end

%% Calling JAGS to sample
doparallel = 0; % do not use parallelization
fprintf( 'Running JAGS...\n' );
tic

[samples, stats, structArray] = matjags( ...
    datastruct, ...                     % Observed data   %fullfile('Rate_1.txt'),
    fullfile(pwd, 'Rate_2.txt'), ...    % File that contains model definition
    init0, ...                          % Initial values for latent variables
    'doparallel' , doparallel, ...      % Parallelization flag
    'nchains', nchains,...              % Number of MCMC chains
    'nburnin', nburnin,...              % Number of burnin steps
    'nsamples', nsamples, ...           % Number of samples to extract
    'thin', 1, ...                      % Thinning parameter
    'dic', 1, ...                       % Do the DIC?
    'monitorparams', {'delta'},     %{'theta'},         % List of latent variables to monit
```

pymc

- Librería de python
- Más flexible, v2 menos robusta, v3?

- python
- https://pypi.python.org/pypi/pymc

pymc

```python

import numpy as np
import pymc

def make_model(data, subject_trials):

    n_blocks=data.shape[1]
    n_trials_0=subject_trials/2
    n_trials_1=subject_trials/2

    theta0=np.empty(n_blocks, dtype=object)
    theta1=np.empty(n_blocks, dtype=object)
    k0=np.empty(n_blocks, dtype=object)
    k1=np.empty(n_blocks, dtype=object)

    for i in range(n_blocks):
        theta0[i]=pymc.Uniform('theta0_{0}'.format(i), 0, 1)
        theta1[i]=pymc.Uniform('theta1_{0}'.format(i), 0, 1)

        k0[i]=pymc.Binomial('k0_{0}'.format(i), p=theta0[i], n=n_trials_0, value=data[0
        k1[i]=pymc.Binomial('k1_{0}'.format(i), p=theta1[i], n=n_trials_1, value=data[1

    @pymc.deterministic
    def delta(x=theta0, y=theta1):
        return [xi-yi for xi,yi in zip(x,y)]

    return locals()


##########################################
model=pymc.MCMC(testBernoulliModel)
model.sample(iter=1000, burn=100, thin=5)

```

WebPPL

- Lenguaje funcional probabilístico, *subset* de JavaScript
- Muestras <-> Ejecuciones
- Diseñado para ciencias cognitivas
- “Lenguaje de la mente”
- Tutorial Bayes: https://probmods.org
- Lenguaje en profundidad: http://dippl.org/

WebPPL

```
var options = {method: 'enumerate'}
var model = function () {
  var A = flip()
  var B = flip()
  var C = flip()
  var D = A + B + C
  //add the desired assumption:
  condition(D >= 2)
  return A
};
var dist = Infer(options, model)
viz(dist)
```

Edward

Models

Most Edward random variables are light wrappers around distributions in TensorFlow. Therefore the list of available random variables depends on the TensorFlow version installed. For TensorFlow 1.1.0, the following are available:

- `edward.models.RandomVariable`
- `edward.models.DirichletProcess`
- `edward.models.Empirical`
- `edward.models.PointMass`
- `edward.models.Bernoulli`
- `edward.models.BernoulliWithSigmoidProbs`
- `edward.models.Beta`

Running inference is as simple as running one method.

```
inference = ed.Inference({z: qz, beta: qbeta}, {x: x_train})
inference.run()
```

Edward

```python
import tensorflow as tf
import edward as ed
from edward.models import Bernoulli, Beta, Binomial
import matplotlib.pyplot as plt
import seaborn as sns

##Single coin weight inference

##Model:
theta = Beta(1.0, 1.0)
x = Bernoulli(probs=theta)

##Sampling:
with tf.Session() as sess:
    for i in range(10):
        print(x.eval())

##Observations:
data=1

##Infer:
qtheta = Beta(tf.Variable(1.0), tf.Variable(1.0))
inference = ed.KLqp({theta: qtheta}, {x: data})
inference.run()

##Results:
qtheta_samples = qtheta.sample(1000).eval()
print(qtheta_samples.mean())
plt.hist(qtheta_samples)
plt.show()
```