O'REILLY®

Strata CONFERENCE

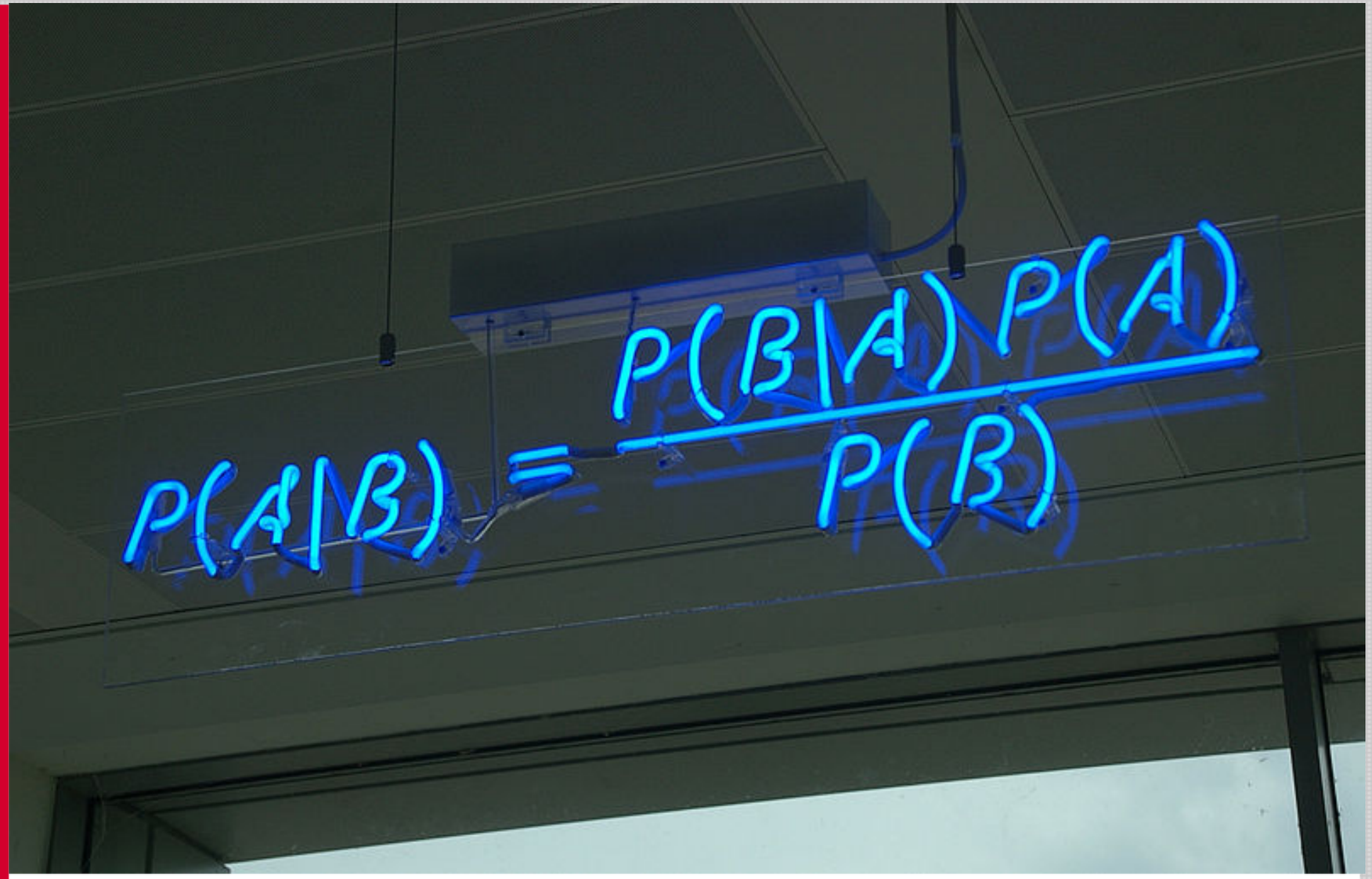
Making Data Work

11–13 Nov 2013

LONDON, ENGLAND

#strataconf

strataconf.com/london



Dealing With Uncertainty: What the reverend Bayes can teach us

Probability – Bernoulli, de Moivre

- Fair coin
  - 50% heads
  - 50% tails

What is the probability of two consecutive heads?

25%

25%

25%

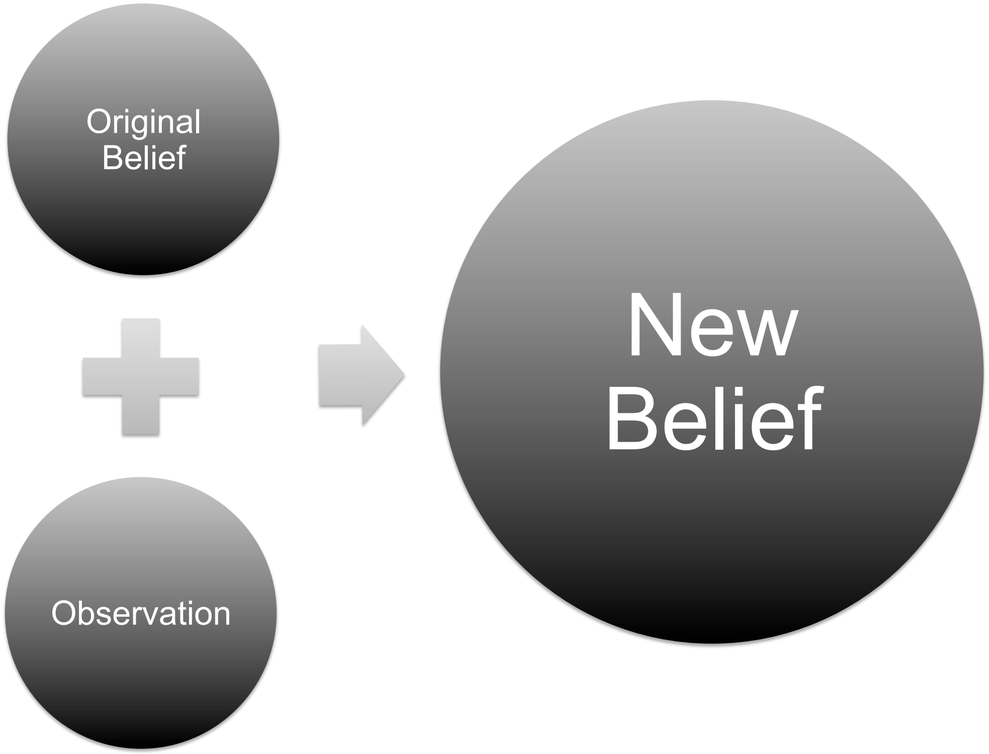
25%

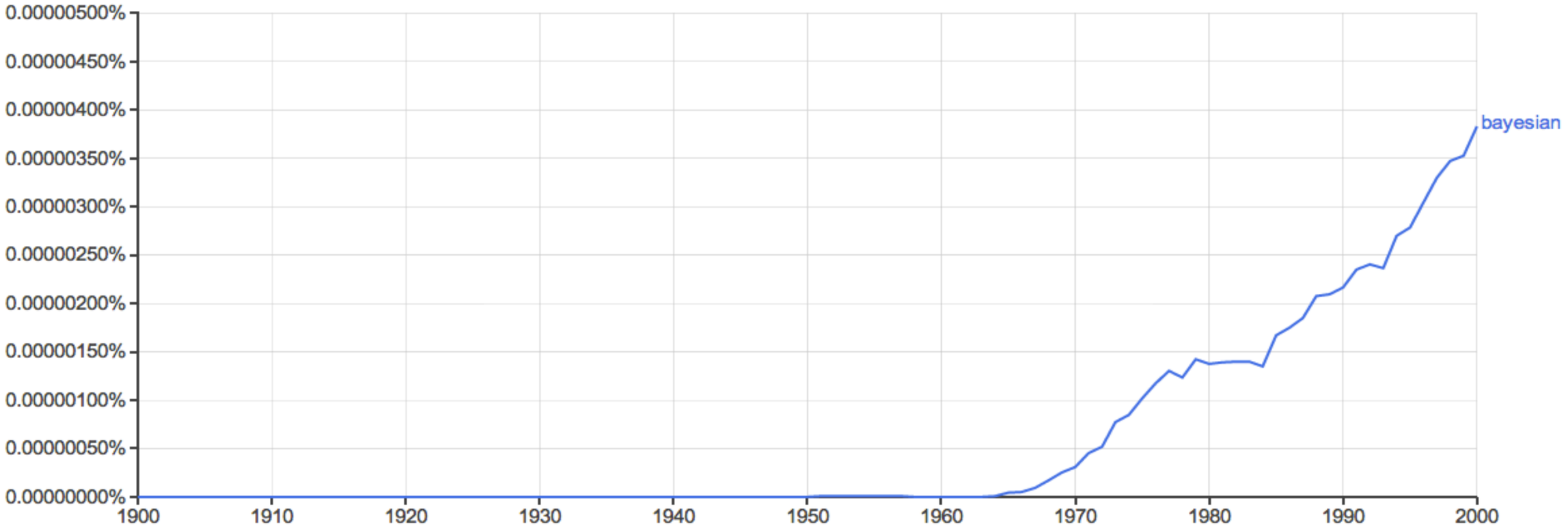
1 7 0 1

Strata
CONFERENCE

Inverse Probability (Bayes)

- Given a coin, not sure whether biased or not?
- If two rolls turn up heads, is the coin biased or not?




0.00000500%
0.00000450%
0.00000400%
0.00000350%
0.00000300%
0.00000250%
0.00000200%
0.00000150%
0.00000100%
0.00000050%
0.00000000%
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
bayesian

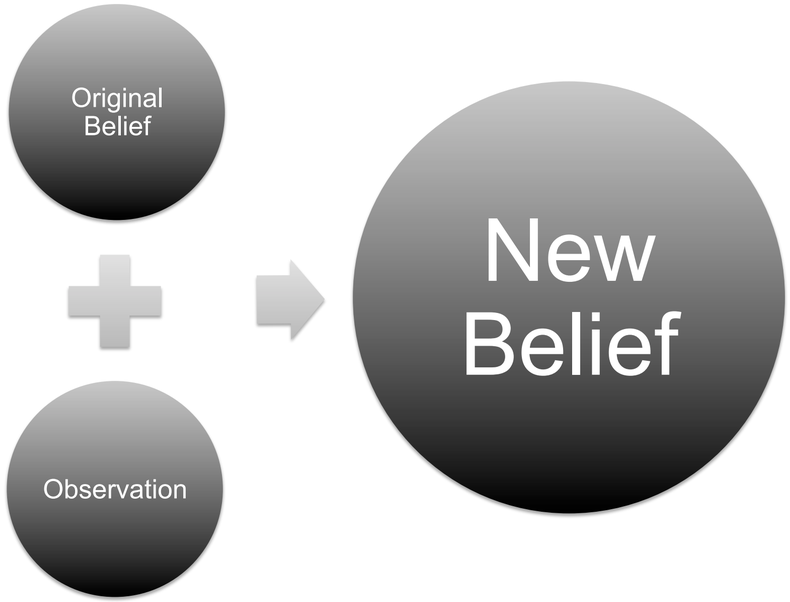
BAYESIAN PROBABILITY

Cox Axioms

- The plausibility of a statement is a real number and is dependent on information we have related to the statement.
- Plausibilities should vary sensibly with the assessment of plausibilities in the model.
- If the plausibility of a statement can be derived in many ways, all the results must be equal.

Outcome:

- If A is true then p(A) = 1
- p(A) + p(not A) = 1
- p(A and B) = p(A|B) x p(B)



$$p(\text{“cause”}\,|\,\text{“effect”}) = \frac{p(\text{“effect”}\,|\,\text{“cause”})p(\text{“cause”})}{p(\text{“effect”})}$$

What is the probability that the person behind the screen is a girl?

50%

What is the probability that the person called Charlie behind the screen is a girl?

Something about probability of Charlie

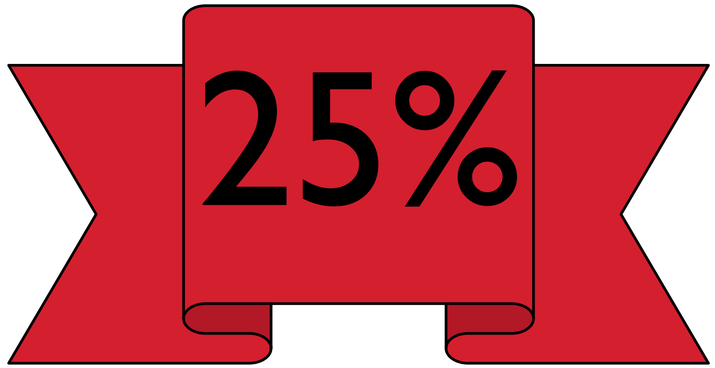
- Girls: 32 / 22989 = 0.13%
- Buys: 89 / 22070 = 0.4%

What is the probability that the person called Charlie behind the screen is a girl?

32 / 22989 = 0.13%

50%

$$p(\text{Girl}|\text{“Charlie”}) = \frac{p(\text{“Charlie”}|\text{Girl})p(\text{Girl})}{p(\text{“Charlie”})}$$



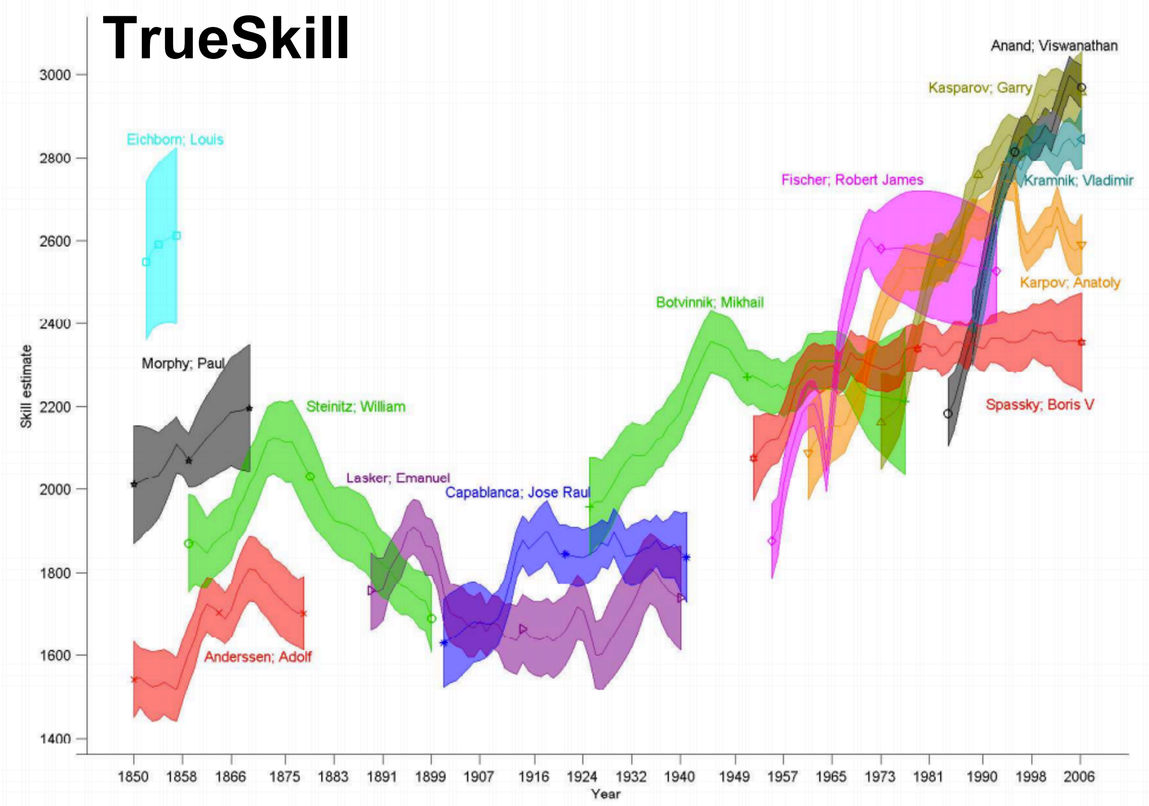
$$p(\text{“Charlie”}|\text{Girl})p(\text{Girl}) + p(\text{“Charlie”}|\text{Boy})p(\text{Boy})$$

32 / 22989 = 0.13% 50% 89 / 22070 = 0.4% 50%

BAYESIAN MACHINE LEARNING

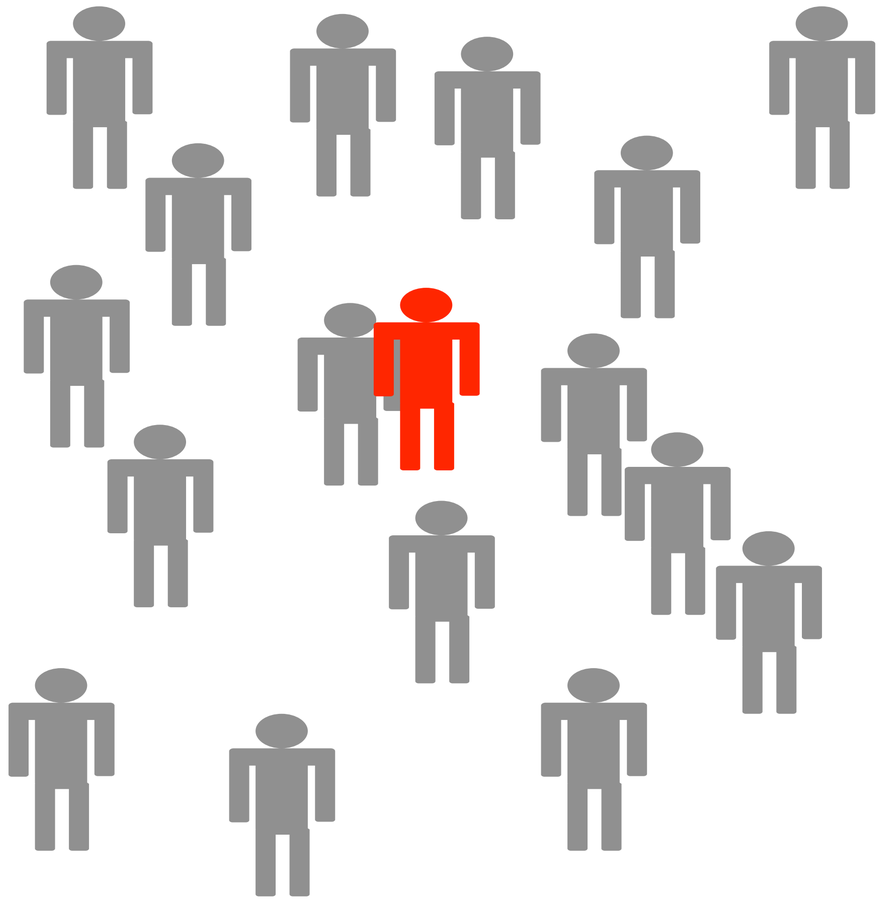
$$p(Spam|Content) = \frac{p(Content|Spam) \times p(Spam)}{p(Content)}$$

$$p(Skill|\text{Match Outcomes}) = \frac{p(\text{Match Outcomes}|Skill) \times p(Skill)}{p(\text{Match Outcomes})}$$



$$p(Road_{t+1}|\text{Image}_t) = \frac{p(\text{Image}_t|Road_t) \times p(Road_t)}{p(\text{Image}_t)}$$

Bayesian Sick People Experiment



- 1 in 100 has health issue.
- Test is 90% accurate.
- You test positive, what are the odds that you need a treatment?

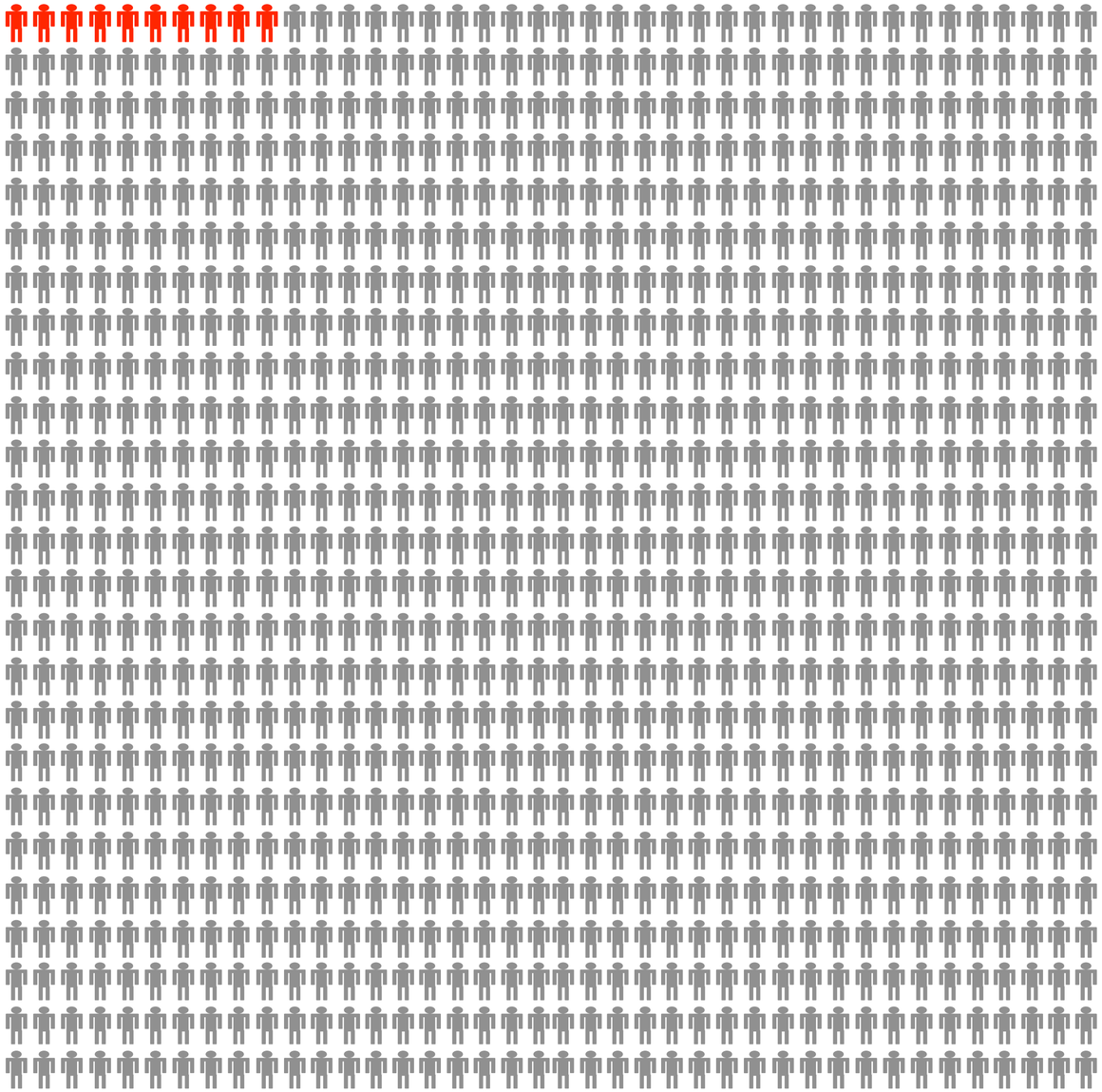
What is the probability of being sick?

A. ≈ 95%

B. ≈ 90%

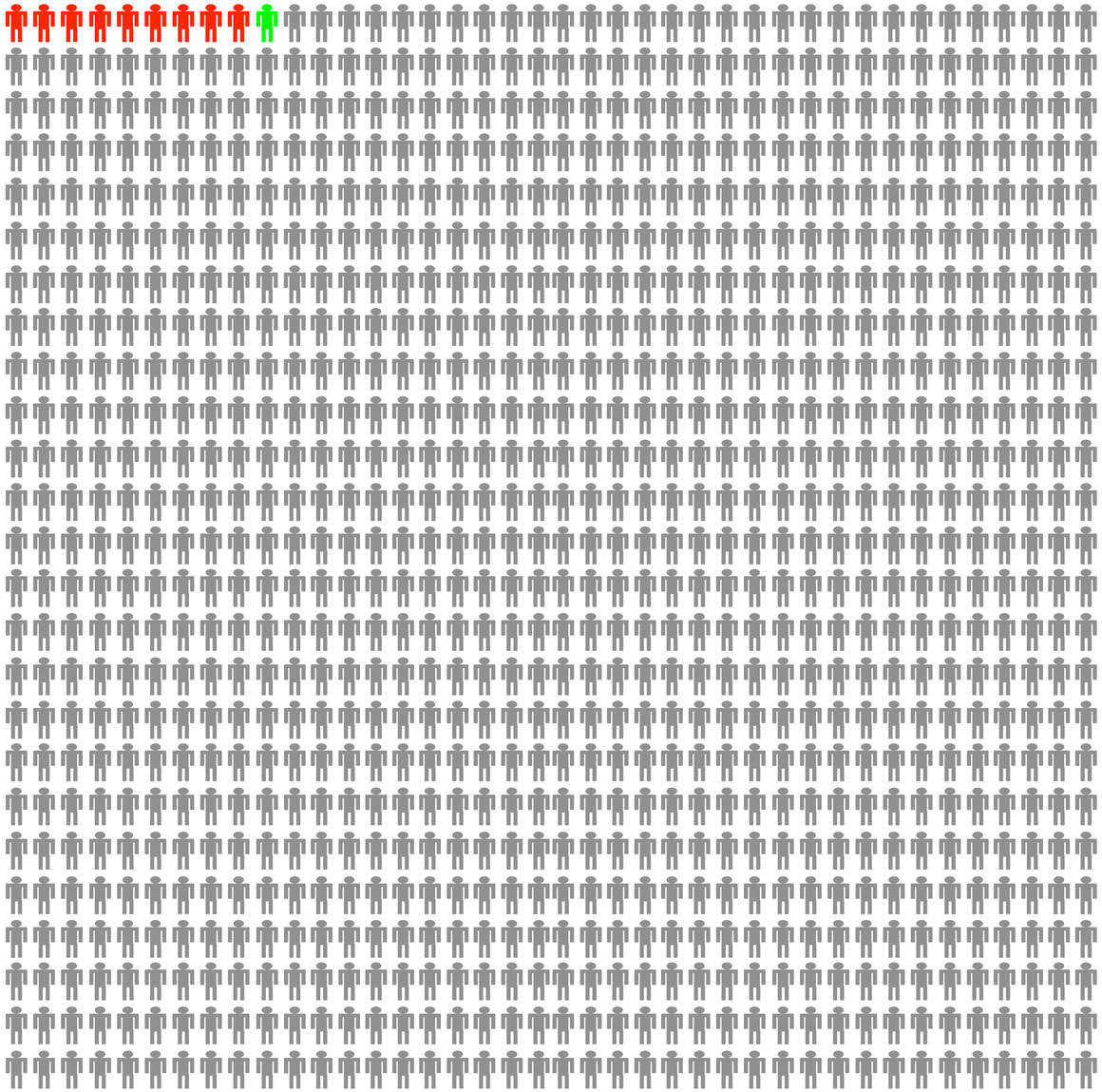
C. ≈ 50%

D. ≈ 10%



- 1000 people in our sample.
- We expect 10 people to be sick (give or take).

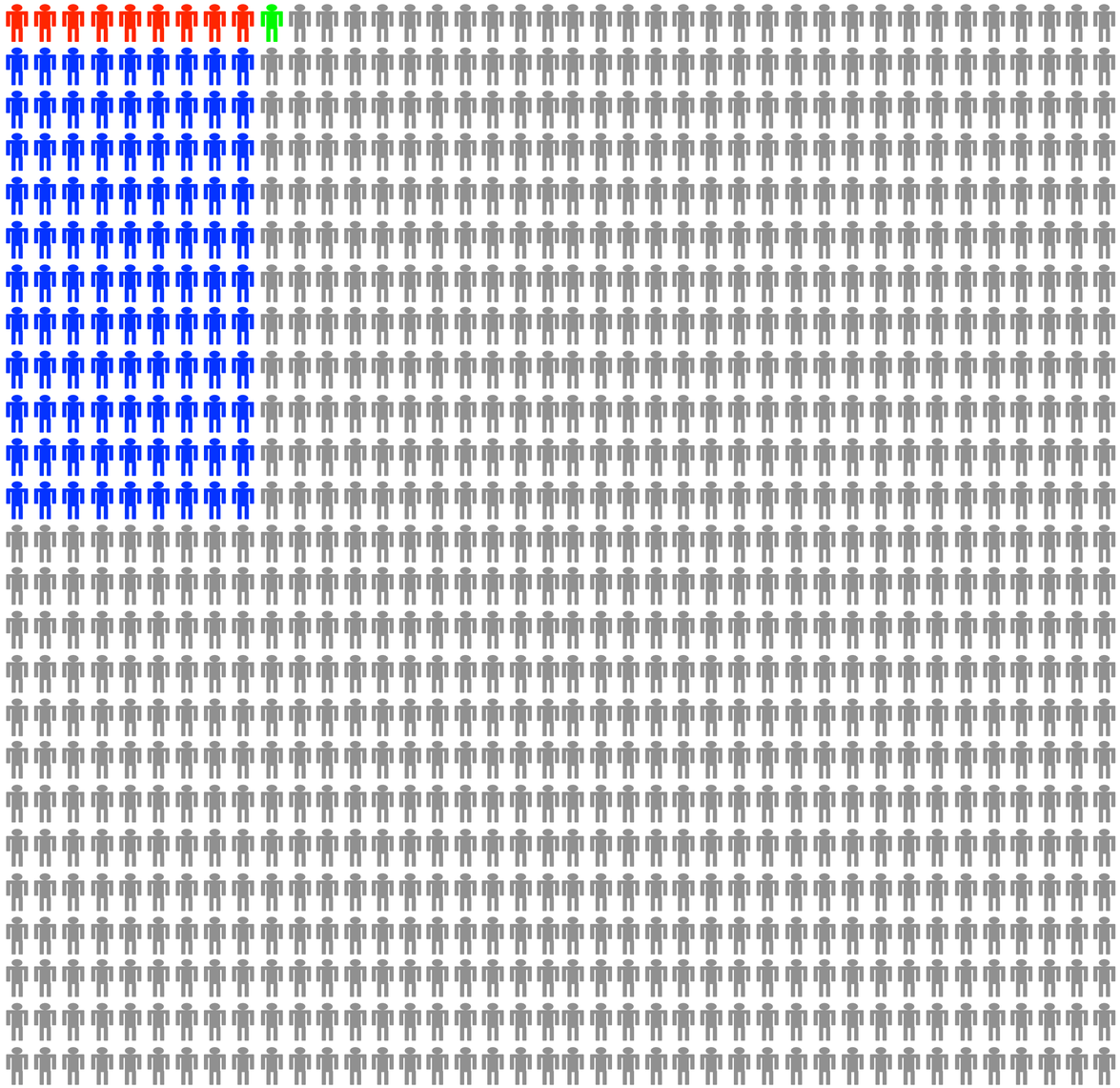
- Imagine testing all individuals?



- 1000 people in our sample.
- We expect 10 people to be sick (give or take).

- Imagine testing all individuals?

→ 9 out of 10 sick people test positive.



- 1000 people in our sample.
- We expect 10 people to be sick (give or take).

- Imagine testing all individuals?

→ 9 out of 10 sick people test positive.

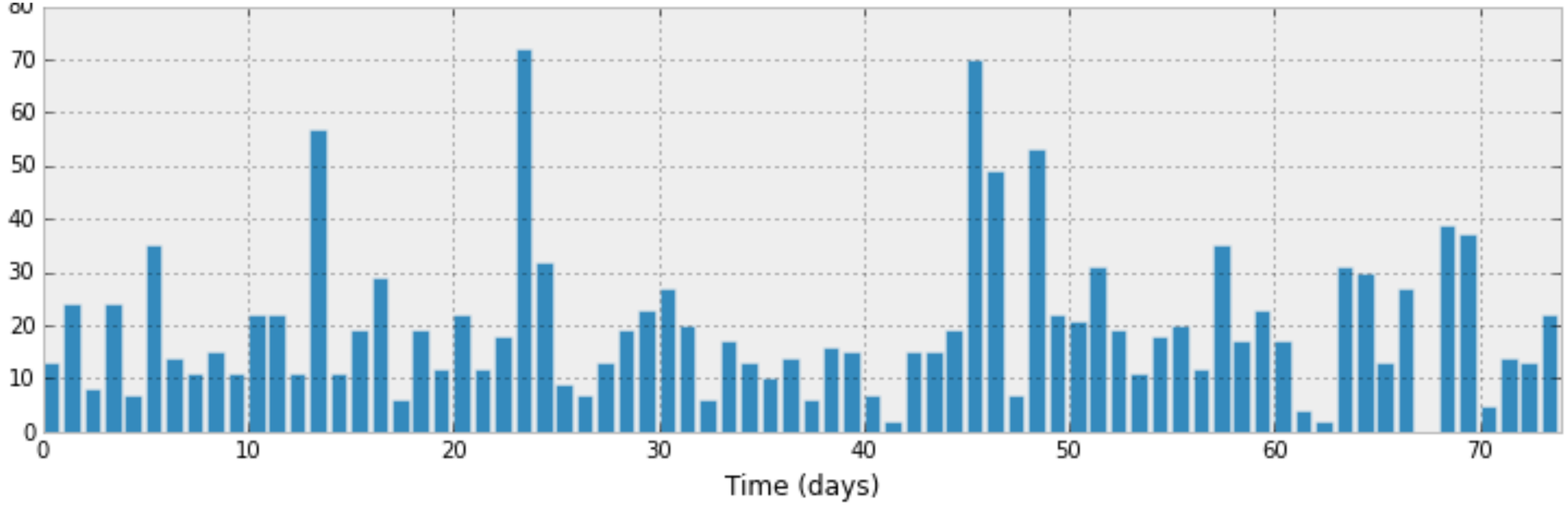
→ 99 out of 990 healthy people test positive!

- I.o.w. if you test positive, it is actually not very likely that you are sick.

PROBABILISTIC PROGRAMMING

Cause → Effect

Effect → Cause



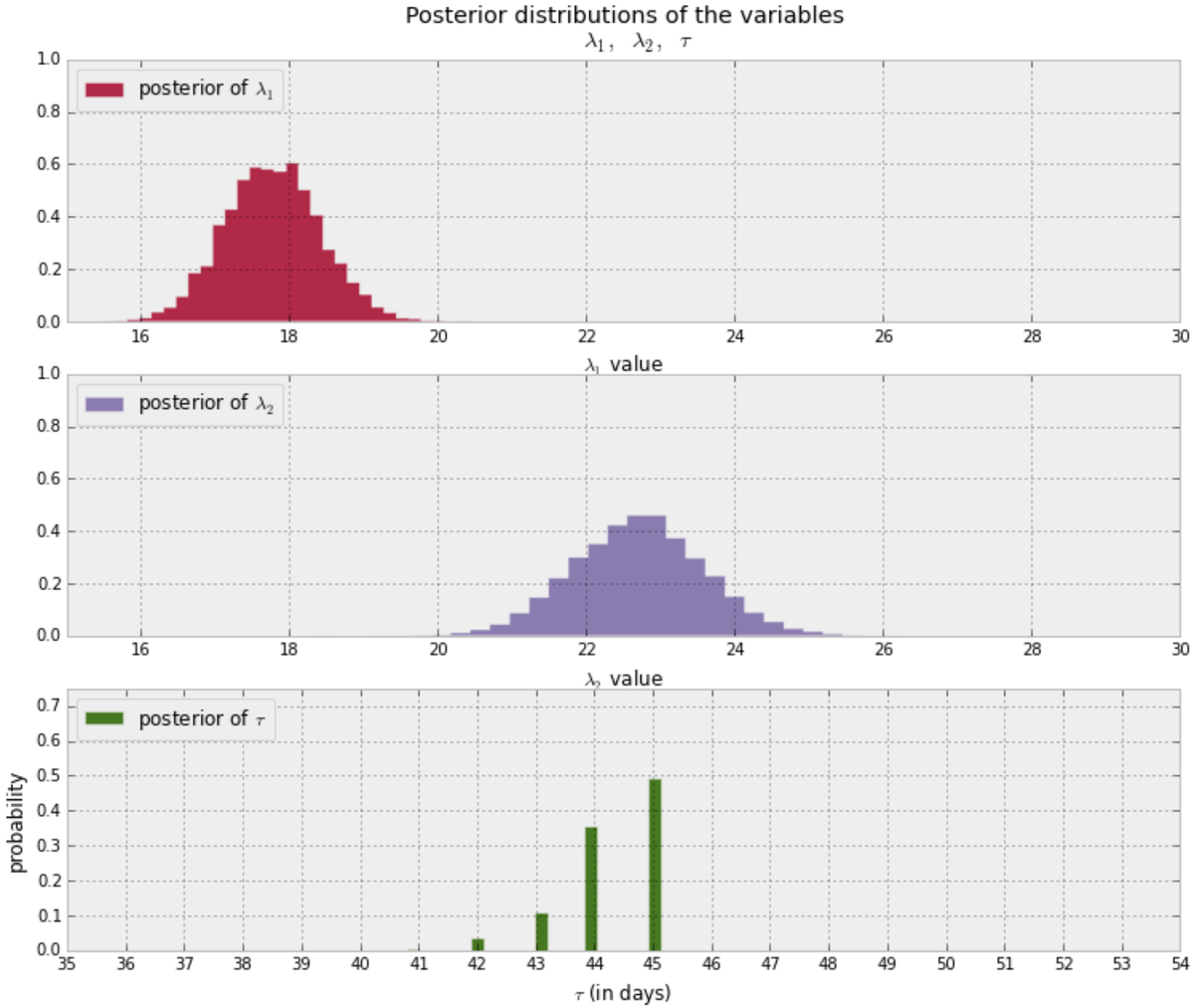
- Imagine a timeline of sales per day for a particular product.
- Did the sales rate for this product change over time?

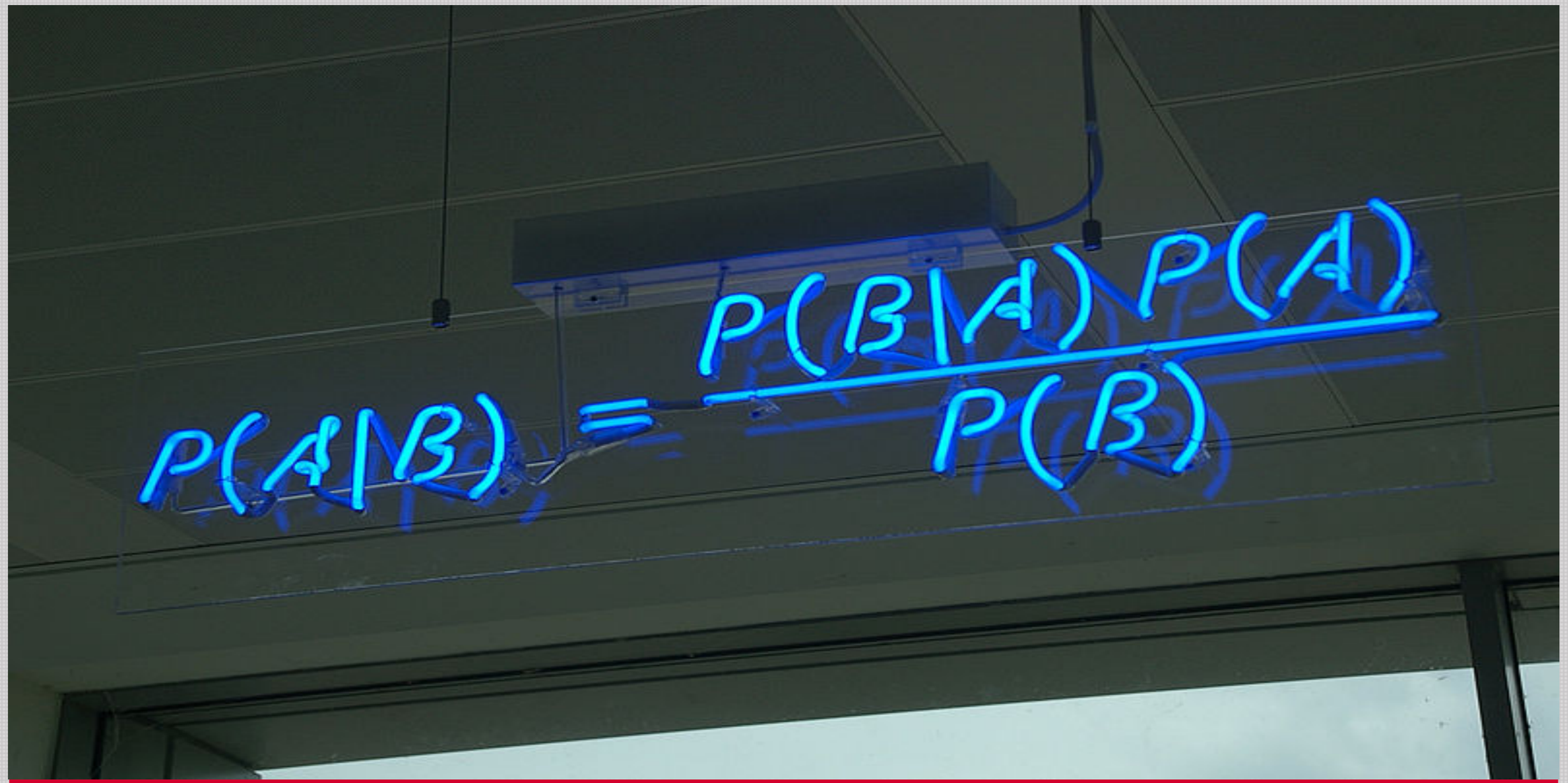
Thinking From Cause to Effect

- In:
  - Sales rate for period 1.
  - Sales rate for period 2.
  - Switchover point between period 1 and 2.
- Output:
  - Unit sales over period 1 and 2.

```
model = pymc.Model()
with model:
    switch = pymc.DiscreteUniform(lower=0, lower=70)
    rate_1 = pymc.Exponential(1.0)
    rate_2 = pymc.Exponential(1.0)

    rates = pymc.switch(switch >= arange(70), rate_1, rate_2)
    unit_sales = pymc.Poisson(rates, observed=data)
```


Posterior distributions of the variables
λ_1, λ_2, τ
posterior of λ_1
λ_1 value
posterior of λ_2
λ_2 value
posterior of τ
probability
τ (in days)


$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$

References

- Bayesian vs. Frequentist Statistics
  - http://www.stat.ufl.edu/~casella/Talks/BayesRefresher.pdf
- Probabilistic Programming & Bayesian Methods for Hackers
  - https://github.com/CamDavidsonPilon/Probabilistic-Programming-and-Bayesian-Methods-for-Hackers
- Bayesian Methods
  - http://www.gatsby.ucl.ac.uk/~zoubin/tmp/tutorial.pdf
- “The Theory That Would not Die”, Sharon Bertsch Mcgrayne
  - http://www.amazon.co.uk/dp/0300188226

Medical Example using PyMC

```
model = pymc.Model()
with model:
    sick = pymc.Bernoulli(p=0.01)
    test_result = pymc.Bernoulli(sick * 0.9 + (1-sick) * (1.0-0.9), observed=[1])

    algorithm = pymc.Metropolis()
    print “Pr(Sick | Test) = %f” % pymc.sample(1000, algorithm)[sick].mean()
```