Bayesian Inference

Universiteit van Amsterdam

d.huppenkothen@uva.nl

A short software interlude!

- you need PyTorch 2.5.1
- Go to PyTorch.org —> “previous version of PyTorch”
- Scroll to v2.5.1
- Find the right command for your system
- If there are multiple options, choose CPU

If conda doesn’t find `sbi`, install with

python -m pip install sbi

Anaconda versus Miniforge

In March 2024, Anaconda suddenly and surprisingly changed its licensing requirements (and soon started sending threatening letters to institutions)

… much confusion ensued!

For now, Anaconda is free to use for academic institutions, but many institutions removed the dependency on it

Conda (the package) is open-source, but the Anaconda Distribution and the defaults channel are not!

Miniforge is a fully open-source, community-maintained alternative

Bayesian Inference

UNIVERSITEIT VAN AMSTERDAM

d.huppenkothen@uva.nl

Estimating Probabilities for Rare Events

From: Sharon Bertsch McGrayne, The theory that would not die

It's the Cold War during the 1950. The US has lots of nuclear bombs on aeroplanes. How do we estimate the probability of a nuclear bomb exploding by accident?

The military analysts: no accidents have ever occurred, therefore the probability is zero and no accidents will occur in the future.

Fisher: probability is a relative frequency in an infinitely large population. Until there is a sufficiently large number of nuclear bomb accidents, we have no way of computing the future probability.

Reality: several (16) near-miss accidents had already happened.

What would you do?

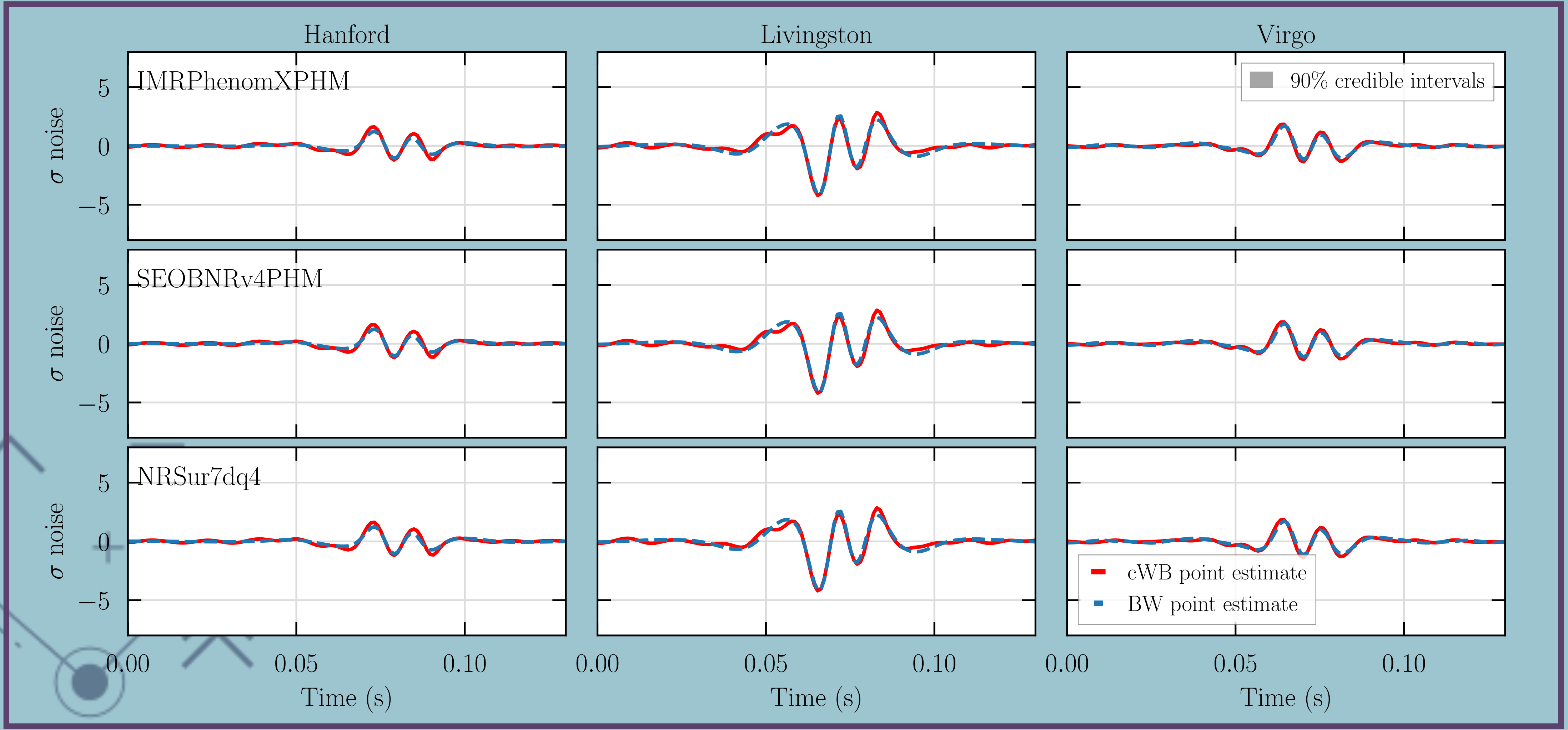
A related Astronomy problem:

estimating event rates in astronomical surveys

LIGO has found a single GW event from an intermediate-mass black hole.

What is the underlying rate of of IMBH mergers?

What is the predicted number of IMBH GW events in Advanced LIGO?



The LVK Collaboration (2021)

What is different in these questions from the ones you might have encountered so far?

- small, complicated data
- incorporate prior information from disparate sources
- update your probability based on data

Is it a random variable?

https://app.wooclap.com/BAYESSCHOOL

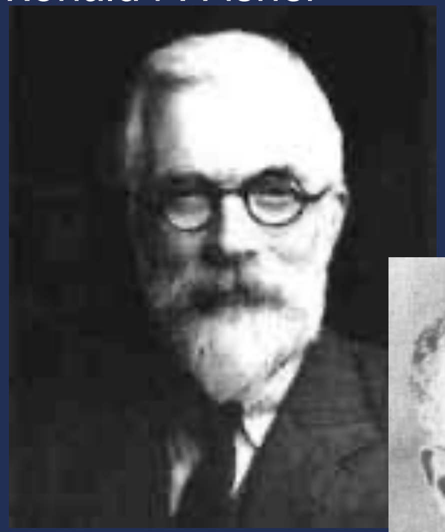
What is different in these questions from the ones you might have encountered so far?

- small, complicated data
- incorporate prior information from
- update your probability based on

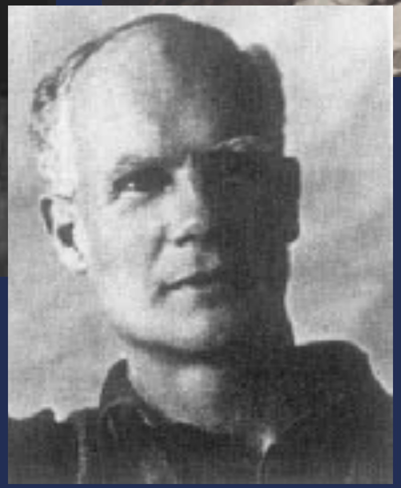
IS A RANDOM VARIABLE!

Statistics Battle of the Century

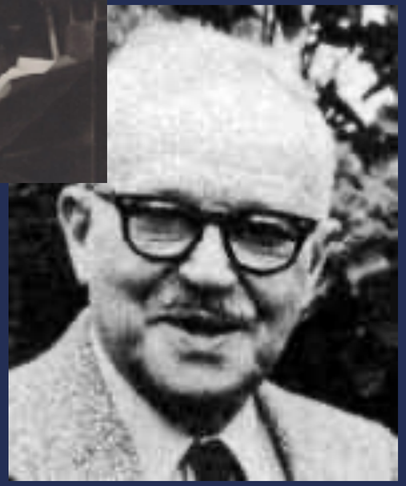
Ronald A Fisher



Karl Pearson



Egon Pearson



Jerzy Neyman

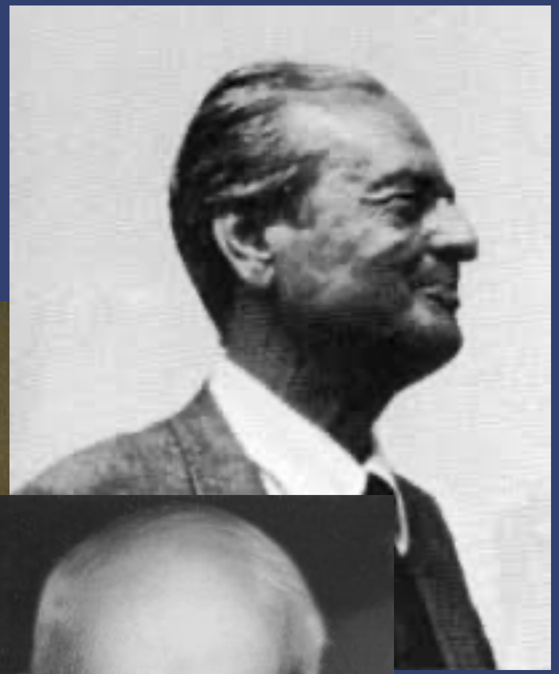
+ most of 20th century academic statisticians

Frequentist

VS

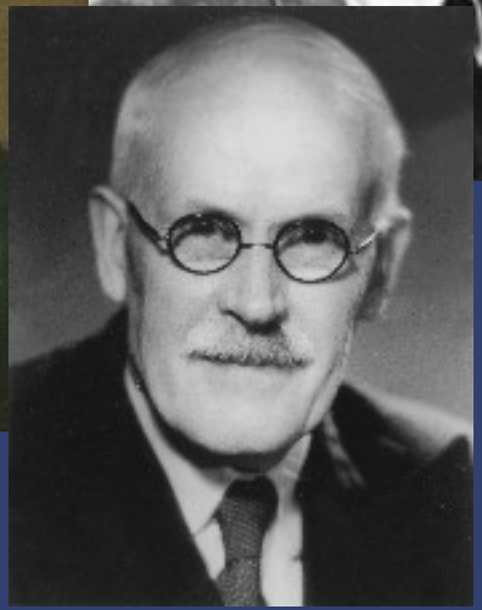
Thomas Bayes

Bruno de Finetti



Alan Turing

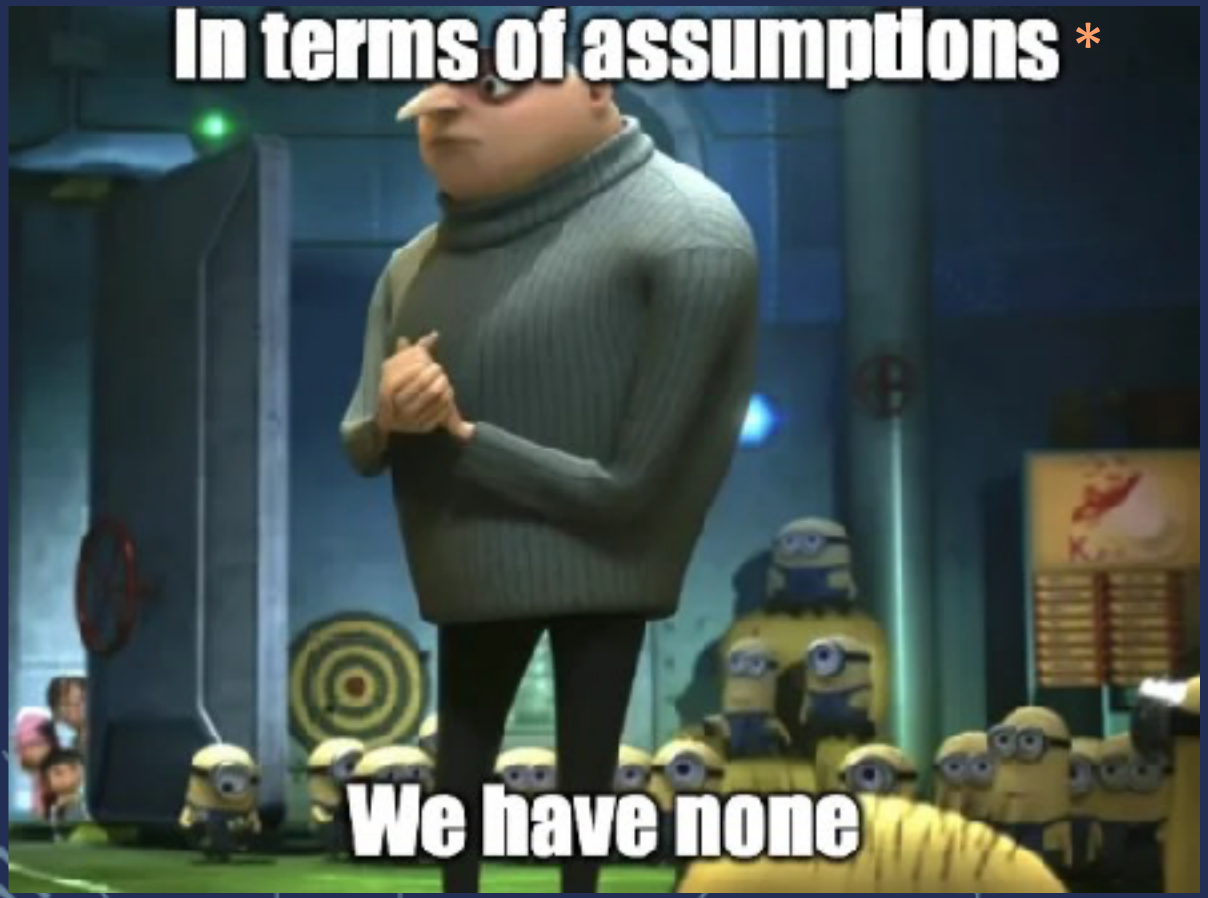
Pierre Simon
Maquis de Laplace



Harold Jeffreys

Bayesian

Frequentist



*except the mathematical assumptions of your statistical tests

Bayesian

FREQUENTIST

- probability defined as a long-running frequency derived from many repeated, identical trials (“proportion”)
- Only stochastic quantities can be random variables
- Only the data counts
- Relies mostly on hypothesis tests

BAYESIAN

- Probability defined as a degree of belief or degree of uncertainty in an outcome (sometimes also quantified as “betting odds”)
- Anything you’re uncertain about can be a random variable
- Our belief in an outcome is a combination of our prior beliefs in that outcome, and the information supplied by the data.
- Enables direct computation of probabilities of hypotheses

Conditional Probabilities

DANIELA HUPPENKOTHEN

Universiteit van Amsterdam

d.huppenkothen@uva.nl

Example: Monty Hall Problem

Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?

What will we eat at the dinner tomorrow?

$$F = (\text{tapas}, \text{pizza})$$

$$P(F=\text{tapas}) \qquad P(F=\text{pizza})$$

Basic rule of probability

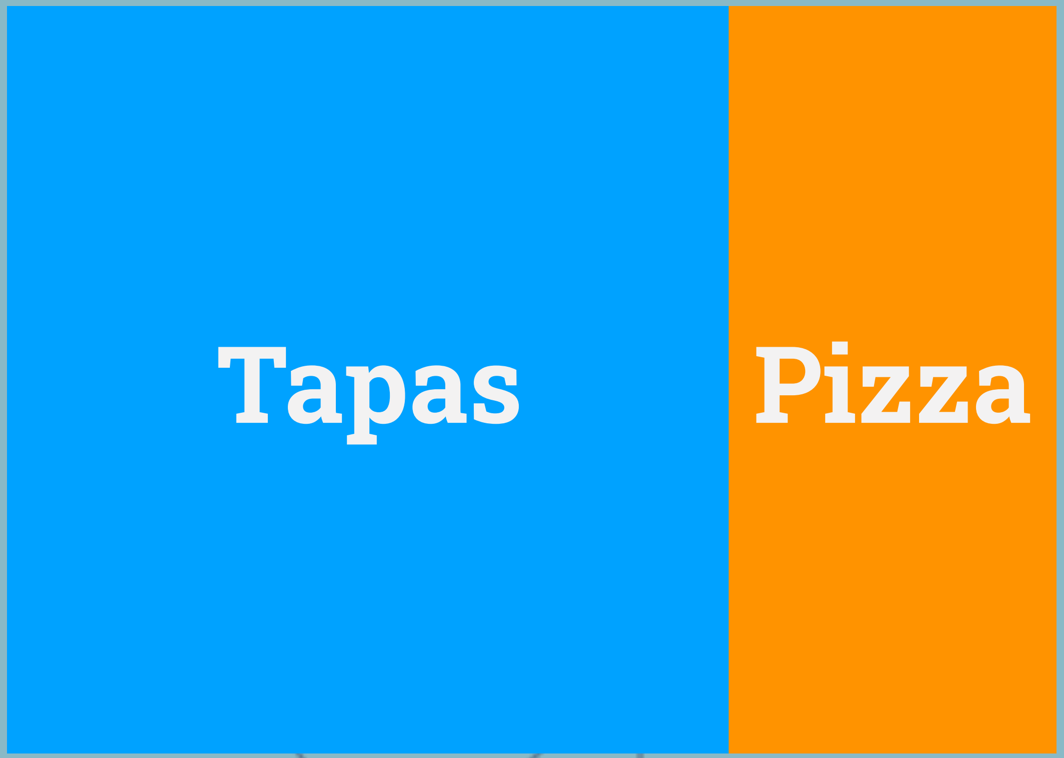
(First and second axioms of probability)

$$0 \leq \mathbf{P(F)} \leq 1$$

If p(tapas)=0.7, what is p(pizza)?

a) 0.3
b) 0.5
c) 0.01
d) 0.7

If P(tapas)=0.7,
what is P(pizza)?



P(tapas) + P(pizza) = 1

$$\sum_i P_i(x) = 1$$

Second axiom of probability

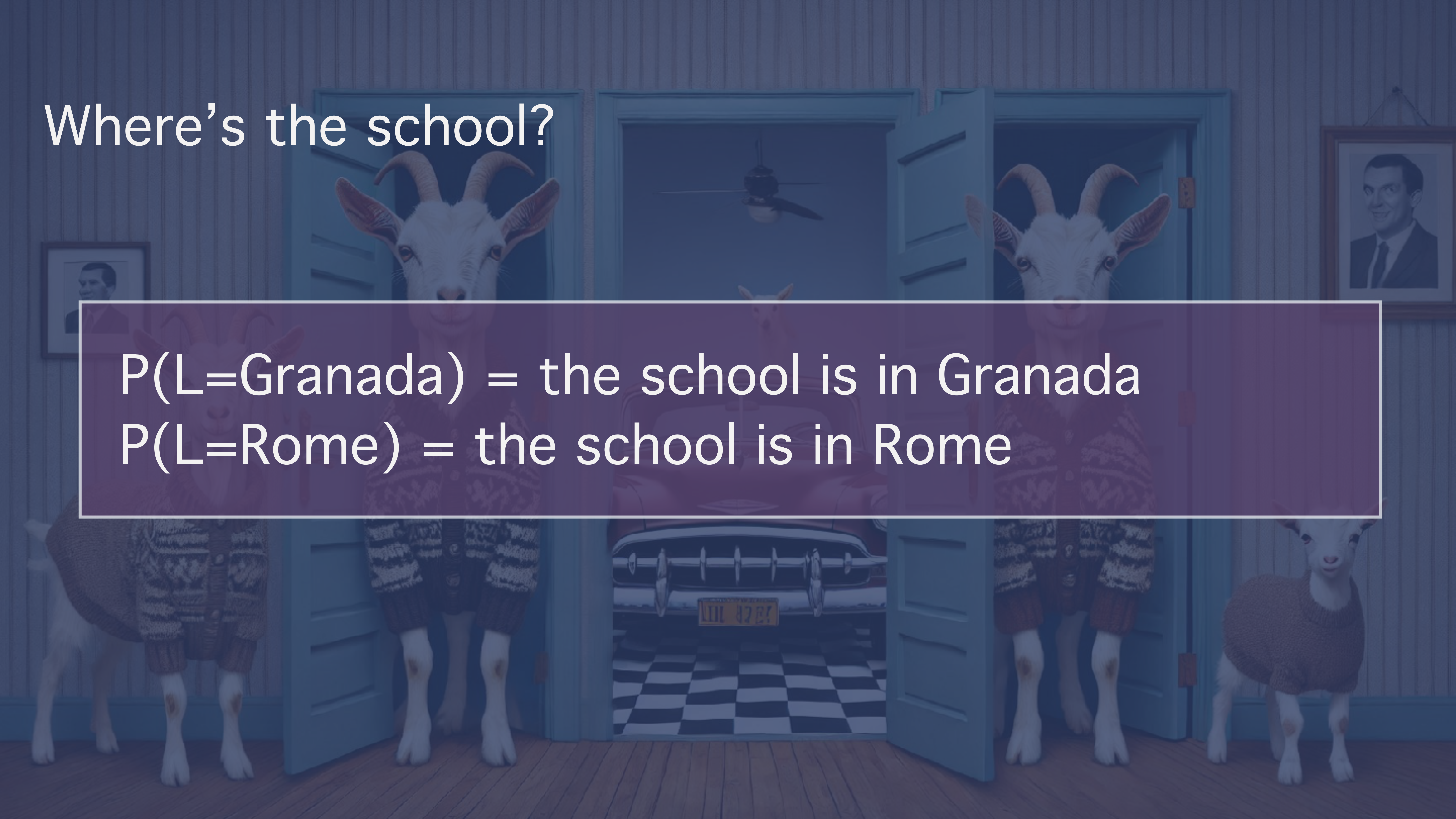
$$P(\text{tapas}) + P(\text{pizza}) + P(\text{curry}) = 1$$

“complement”

$$P(\text{pizza}) + P(\text{curry}) = P(\text{not tapas})$$

$$\boxed{P(\text{tapas}) = 1 - P(\text{not tapas})}$$

Let’s add another category



Where’s the school?

P(L=Granada) = the school is in Granada
P(L=Rome) = the school is in Rome

The joint distribution

P(tapas ∩ Granada) = "P(tapas and Granada)"
= P(tapas, Granada)

F, L are independent:

$$\boxed{P(\text{tapas}, \text{Granada}) = P(\text{tapas})P(\text{Granada})}$$

The joint distribution

$$P(\text{tapas} \cap \text{Granada}) = \text{“}P(\text{tapas and Granada})\text{”}$$
$$= P(\text{tapas, Granada})$$

F, L are not independent:

$$\boxed{P(\text{tapas, Granada})}$$

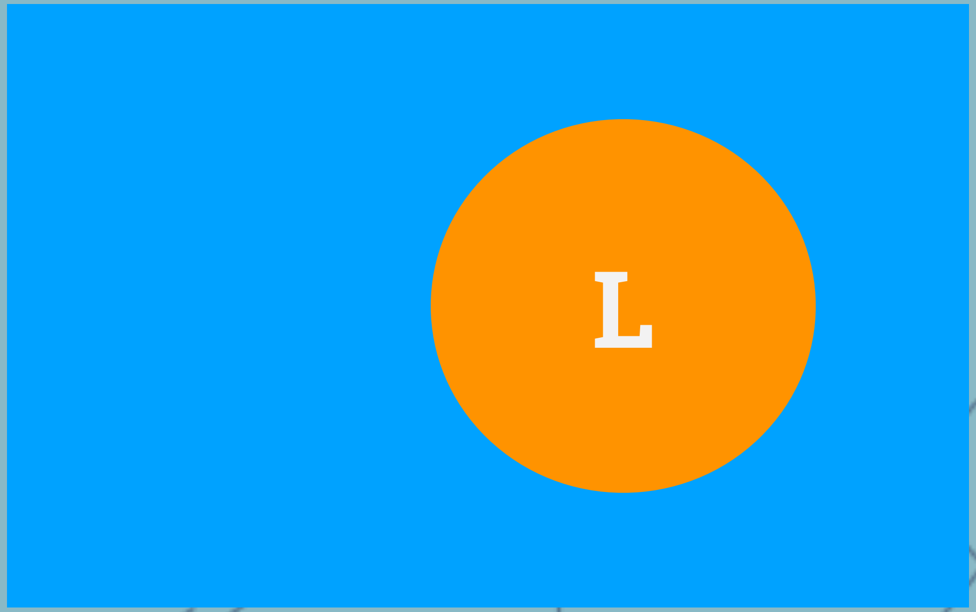
The joint distribution

P(tapas ∪ Granada) = "P(tapas or Granada)"

P(tapas or Granada) = P(tapas) + P(Granada) - P(tapas, Granada)

Unless F and L are mutually exclusive

<u>Marginalisation</u>:
What is p(F)?



$$P(F) = \sum_{L=\{\text{Granada, Rome}\}} P(F,L)$$

Conditional probabilities: What is the probability of having tapas at dinner, given that we're in Granada?

P(tapas | Granada) = "probability of eating tapas given that we're in Granada "

Independent variables:

P(tapas | Granada) = p(tapas)

Multiplication rule

Dependent variables

P(tapas | Granada) = P(tapas , Granada) / P(Granada)

Which dinner will we eat given our location?

	Tapas	Pizza	Curry
Granada	0.2	0.15	0.08
Rome	0.5	0.05	0.02

- P(Curry) = ?
- P(Granada | tapas) = ?
- P(curry | Granada) = ?
- Is p(tapas | Granada) = p(Granada | tapas)?
- $\sum_{L} p(\text{tapas}, L) = ?$

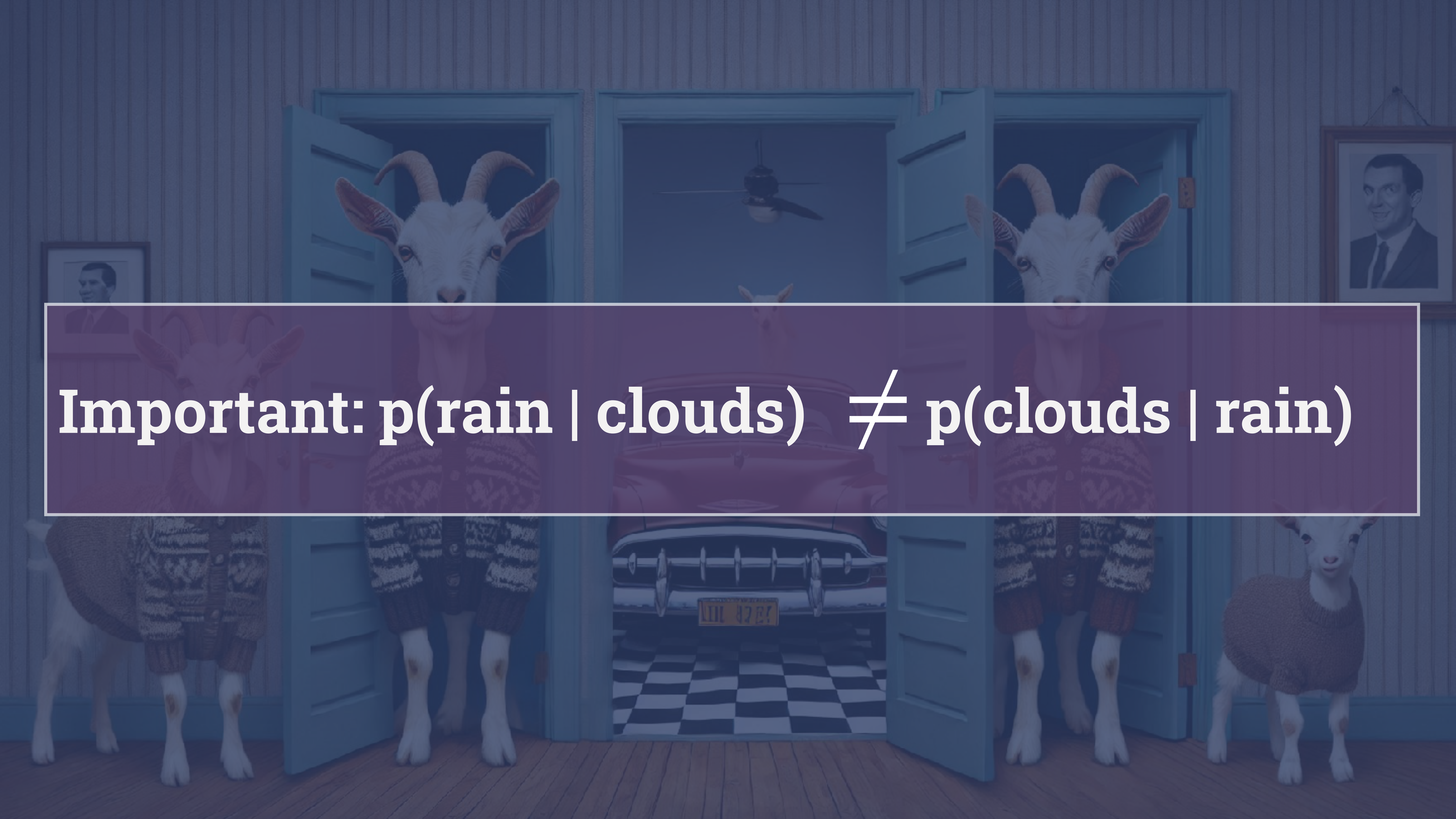
Which dinner will we eat given our location?

	Tapas	Pizza	Curry
Granada	0.2	0.15	0.08
Rome	0.5	0.05	0.02

- P(Curry) = ? **0.1**
- P(Granada | tapas) = ? **0.2/0.7**
- P(curry | Granada) = ? **0.08/0.43**
- Is p(tapas | Granada) = p(Granada | tapas)?

NO! Emphatically not!

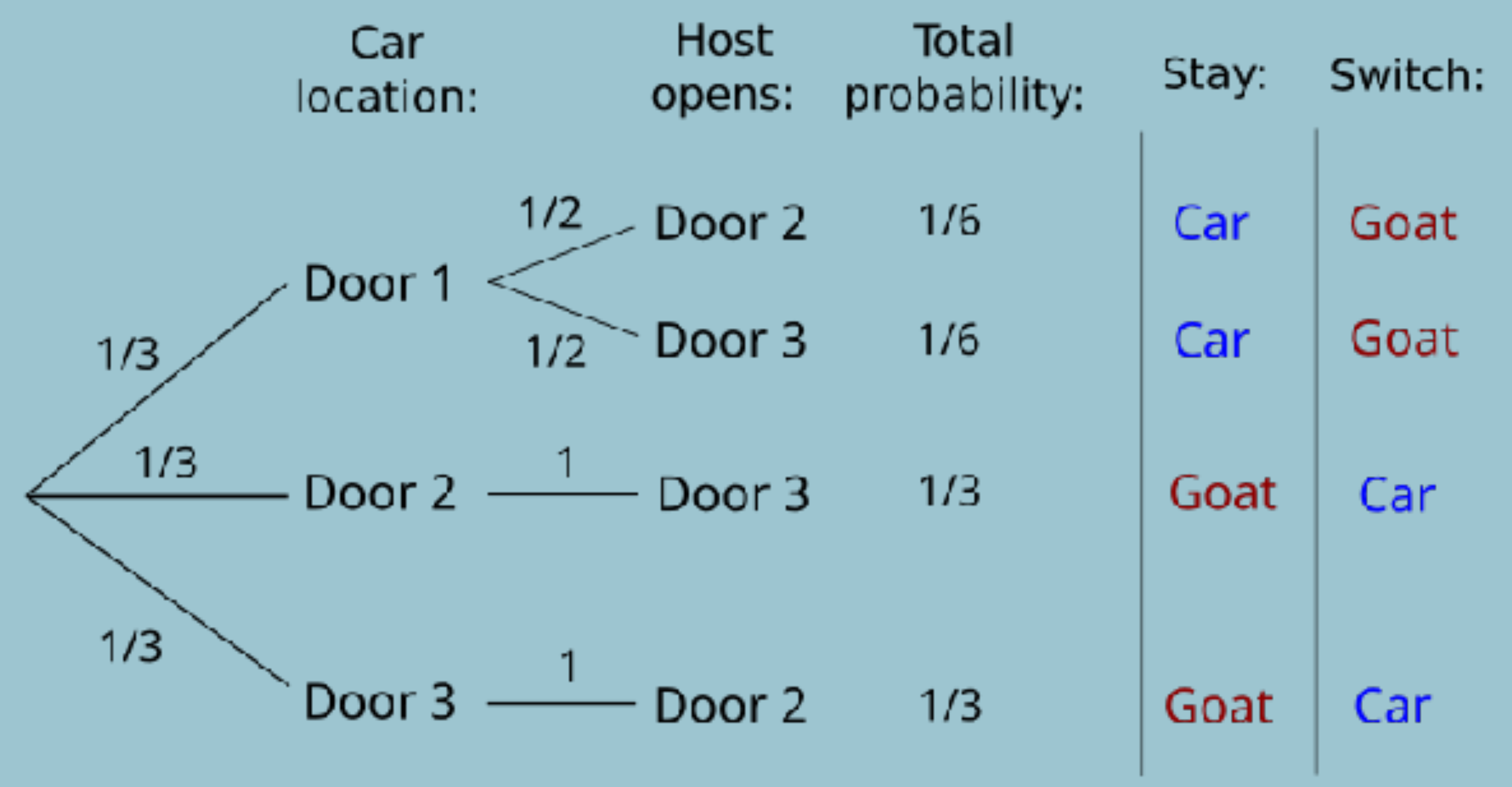
- $\sum_{L}$ p(tapas, L) = ? **0.7**



Important: $p(\text{rain} \mid \text{clouds}) \neq p(\text{clouds} \mid \text{rain})$

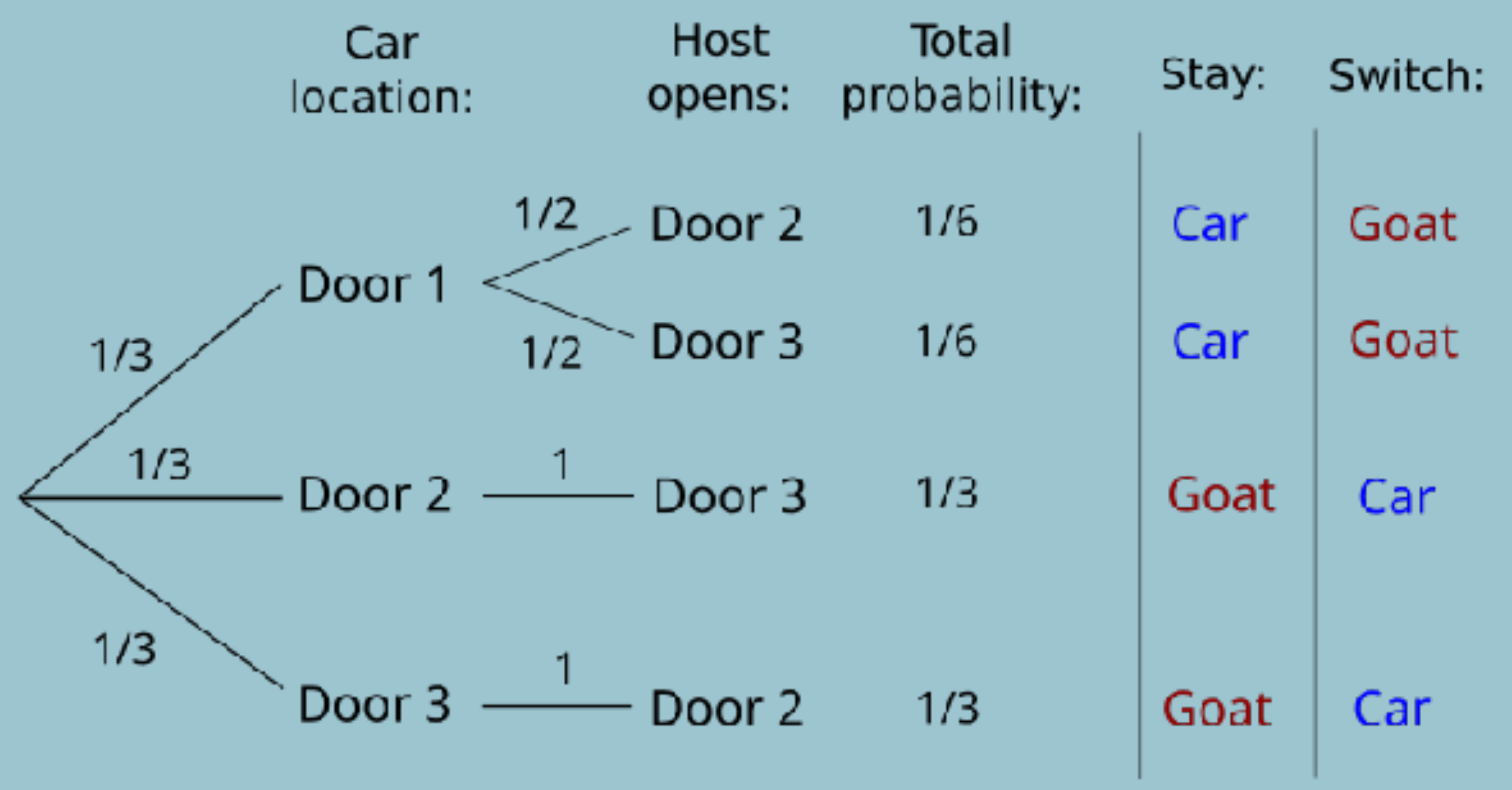
Example: Monty Hall Problem

Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?



Example: Monty Hall Problem

Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?



the conditional probability of winning by switching given the contestant initially picks door 1 and the host opens door 3 is the probability for the event "car is behind door 2 and host opens door 3" divided by the probability for "host opens door 3".

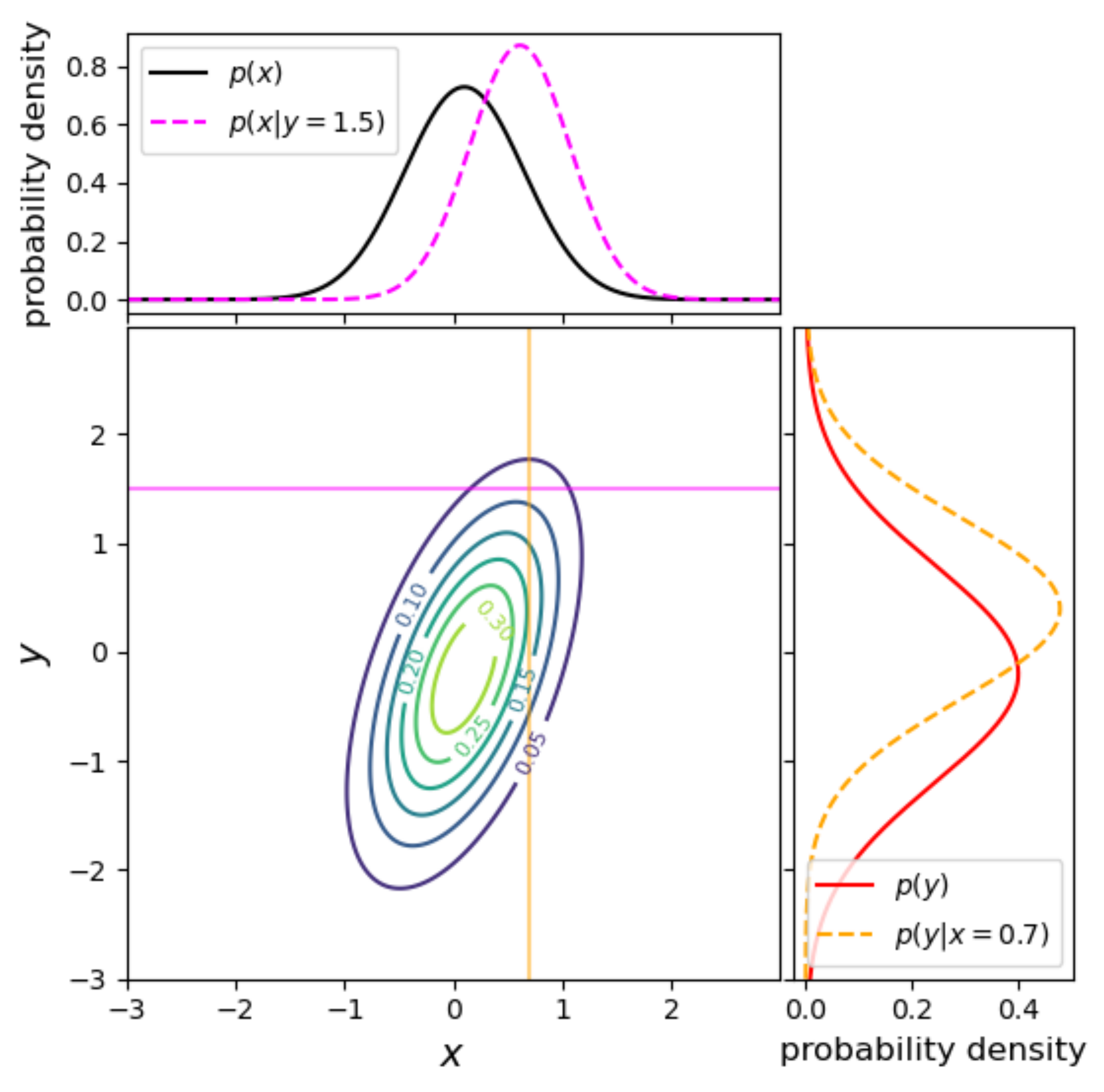
$$P = 1/3 / (1/3 + 1/6) = 2/3$$

Joint Probability Distributions

Questions

$$\int p(x \mid y)\,dx = ?$$

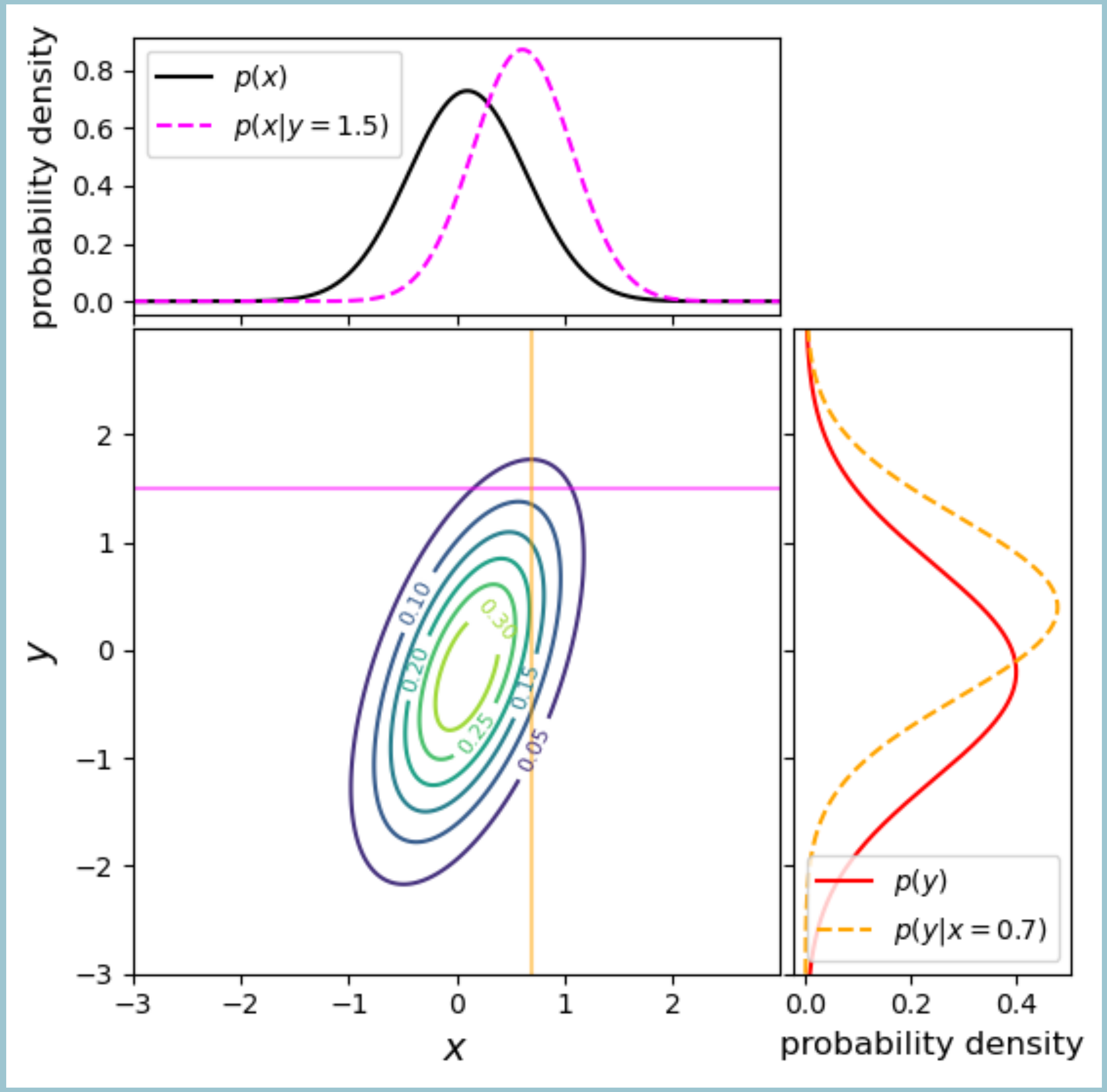
$$\int p(x \mid y)\,dy = ?$$



Questions

$$\int p(x \mid y)dx = 1$$

$$\int p(x \mid y)dy = $$ 👎



Marginalisation

$$p(x \mid y) = p(x, y)/p(y)$$

$$p(x, y) = p(x \mid y)p(y)$$

$$p(x) = \int_{-\infty}^{\infty} p(x, y)dy = \int_{-\infty}^{\infty} p(x \mid y)p(y)dy$$

Bayes Theorem

Multiplication Rule

$$p(x \mid y) = p(x, y)/p(y)$$

$$p(x, y) = p(x \mid y)p(y)$$

$$p(x) = \int_{-\infty}^{\infty} p(x, y)dy = \int_{-\infty}^{\infty} p(x \mid y)p(y)dy$$

Marginalisation

$$p(x \mid y) = p(x, y)/p(y)$$

$$p(x, y) = p(x \mid y)p(y)$$

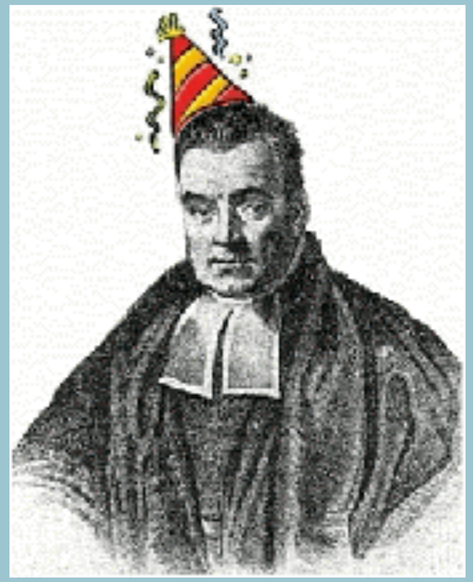
$$p(y \mid x) = ?$$

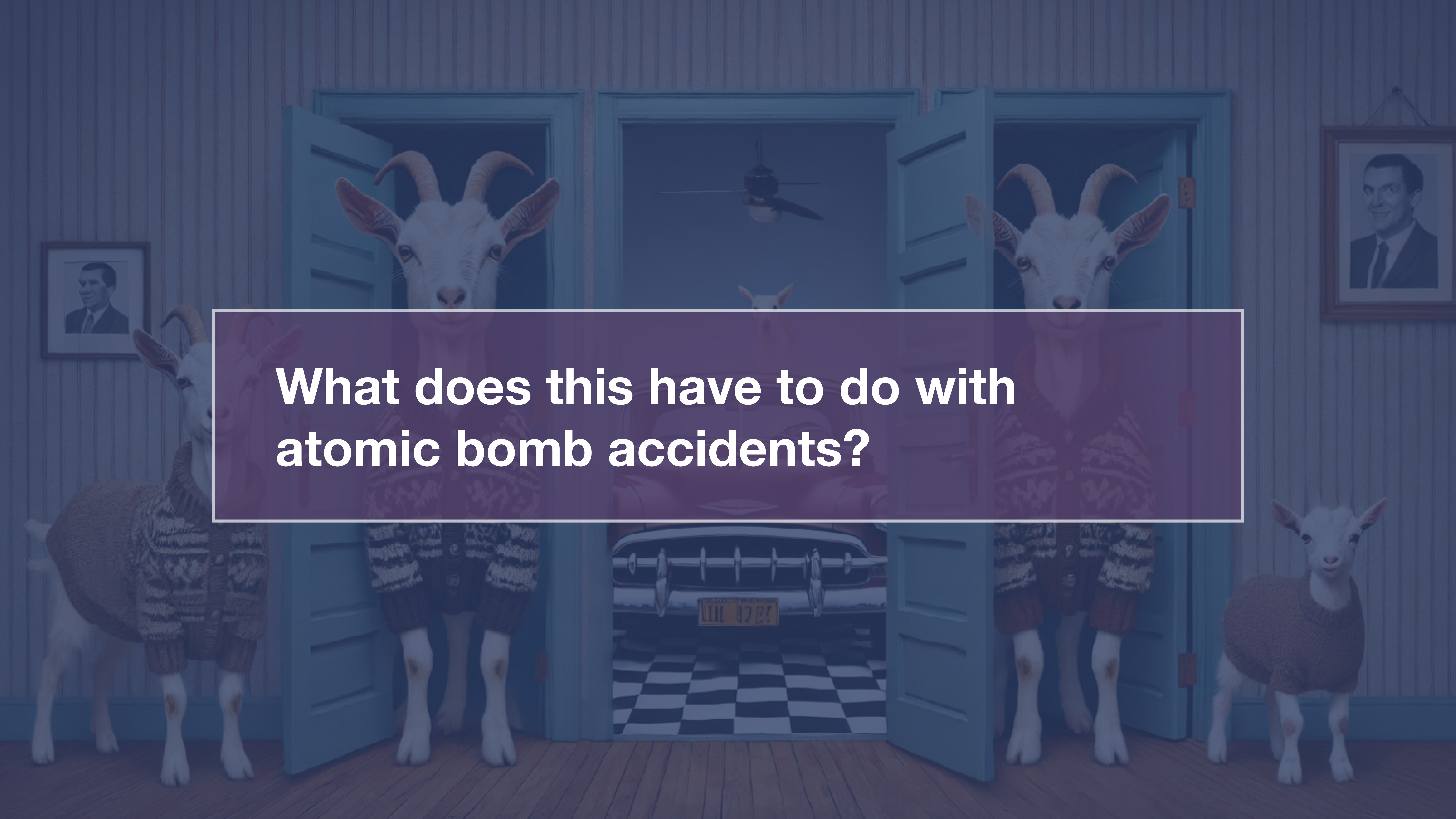
Bayes' Theorem

$$p(x, y) = p(x \mid y)p(y)$$

$$p(x, y) = p(y \mid x)p(x)$$

$$\boxed{p(x \mid y)p(y) = p(y \mid x)p(x)}$$





What does this have to do with atomic bomb accidents?

What is the probability of a nuclear bomb accident?

From: Sharon Bertsch McGrayne, The theory that would not die

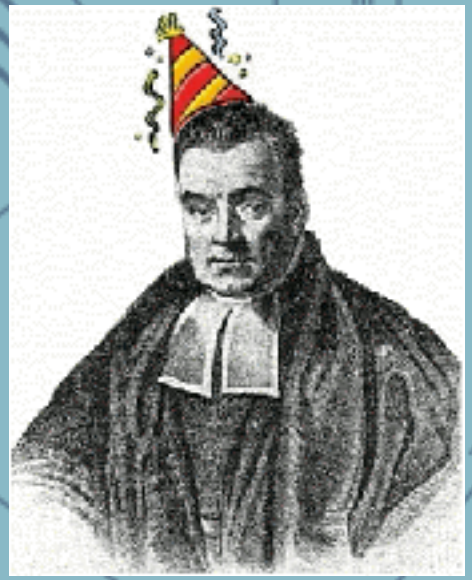
It's the Cold War during the 1950. The US has lots of nuclear bombs on aeroplanes. How do we estimate the probability of a nuclear bomb exploding by accident?

Albert Madansky:

- start with a priori assumption that the probability of an accident is very small, but not zero
- Probability of accident-free future depends on the length of the accident-free past and the number of future accident opportunities
- Include plans for future number of planes carrying nuclear weapons to calculate the number of "accident opportunities"
- Use plane accidents involving non-nuclear bombs as data

Result: expect 19 "conspicuous weapon accidents" per year. Led to implementation of major safety features.

Bayes’ Theorem



$$\boxed{p(x\,|\,y)p(y) = p(y\,|\,x)p(x)}$$

$x =$ future accidents with nuclear bombs

$y =$ recorded plane accidents without nuclear bombs

$p(x) =$ prior assumptions about the probability of accidents, based on projections of future plane flights with nuclear bombs

$p(y\,|\,x) =$ a model for observing the data, given the prior assumptions

$p(x\,|\,y) =$ posterior probability of observing future accidents with nuclear bombs, given our data on plane accidents and our prior assumptions

$$\boxed{P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}}$$

Bayes rule

$\theta =$ model parameters

$D =$ data

Bayesian

$$P(\theta \mid D) = \frac{P(D \mid \theta) P(\theta)}{P(D)}$$

- Probability defined as a degree of belief or degree of uncertainty in an outcome (sometimes also quantified as “betting odds”)
- Anything you’re uncertain about can be a random variable
- Our belief in an outcome is a combination of our prior beliefs in that outcome, and the information supplied by the data.
- Enables direct computation of probabilities of hypotheses