lecture 30

Probability Functions

When dealing with discrete random variables, define a **Probability Function** as probability for i[th] possibility

$$P(x_i) = p_i$$

Defined as limit of long term frequency

- probability of rolling a 3 := limit $_{\#\text{trials}\to\infty}$ (# rolls with 3 / # trials)
  - you don't need an infinite sample for definition to be useful

Normalization

$$\sum_i P(x_i) = 1$$

Probability Density Functions

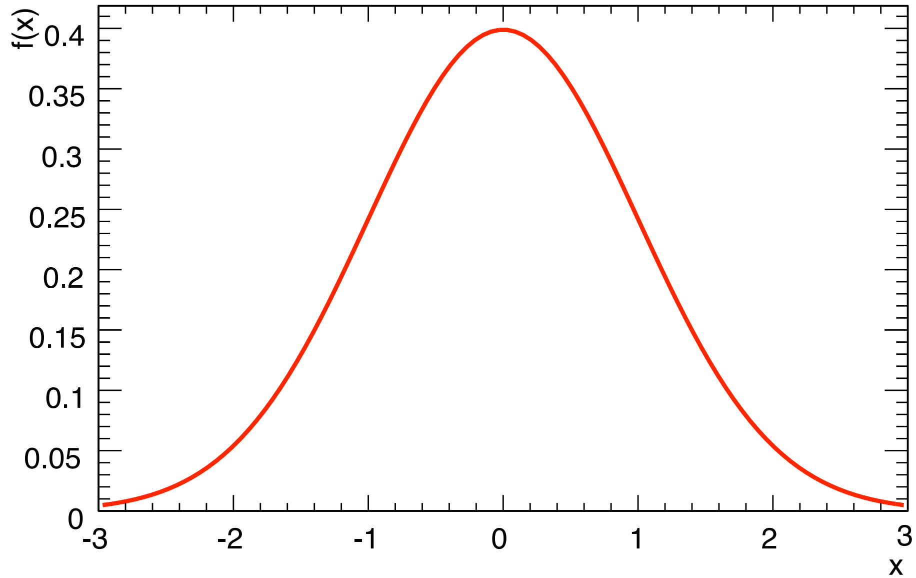
When dealing with continuous random variables, need to introduce the notion of a **Probability Density Function**

$$P(x\ E\ [x, x + dx]) = f(x)dx$$

Note, $f(x)$ is NOT a probability

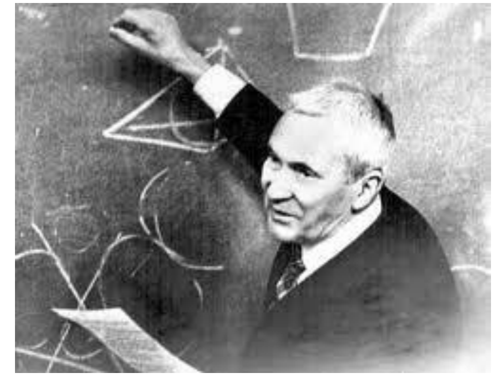
PDFs are always normalized to unity:

$$\int_{-\infty}^{\infty} f(x)dx = 1$$



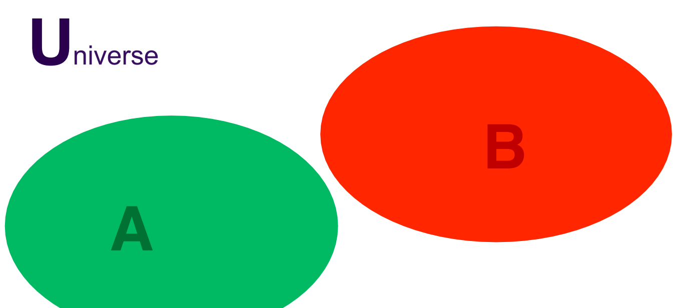
What is Probability

- **Axioms of probability: Kolmogorov (1933)**
  - $P(A) \geq 0$
  - $\int_U P(A)dU = 1$
  - **if:** $(A \text{ and } B) \equiv (A \cap B) = 0$
    (i.e disjoint/independent/exclusive)
    $P(A \text{ or } B) \equiv (A \cup B) = P(A) + P(B)$

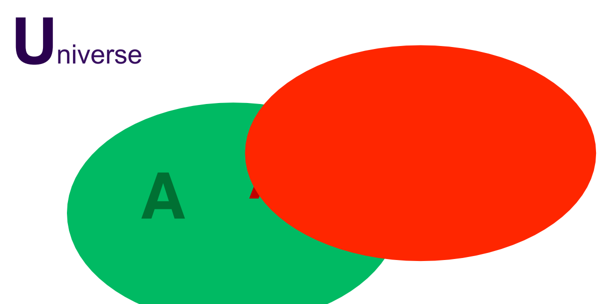


define e.g.: conditional probability

$$P(A|B) \equiv P(A \text{ given } B \text{ is true}) = \frac{P(A \cap B)}{P(B)}$$



Venn-Diagram



Venn-Diagram

What is Probability

- Axioms of probability: - pure “set-theory”

1) a measure of how likely an event will occur, expressed as a the ratio of favourable—to—all possible cases in repeatable trials

- Frequentist (classical) probability

$$P(\text{“Event”}) = \lim_{n\to\infty}\left(\frac{\#\text{outcome is "Event"}}{n_{\text{trials}}}\right)$$

2) the “degree of belief” that an event is going to happen

- Bayesian probability:
  - $P(\text{“Event”})$: degree of belief that “Event” is going to happen -> no need for “repeatable trials”
  - degree of belief (in view of the data AND previous knowledge(belief) about the parameter) that a parameter has a certain “true” value

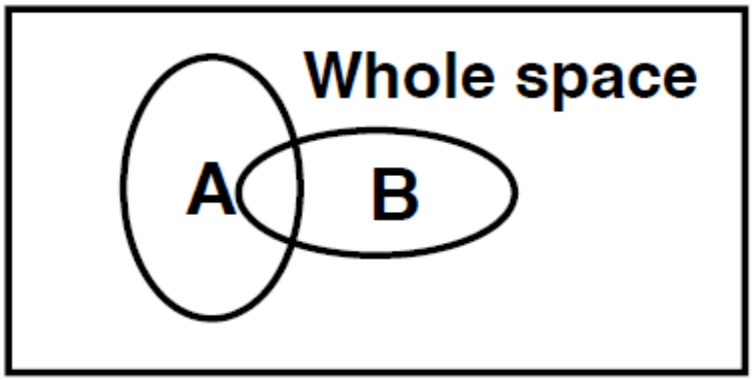
Frequentist vs. Bayesian

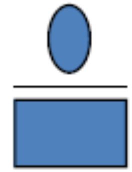
Bayes' Theorem $$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = P(B|A)\,\frac{P(A)}{P(B)}$$

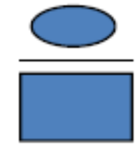
- **This follows simply from the "conditional probabilities":**

Derivation of Bayes' Theorem

… in picture …taken from Bob Cousins



P(A) = 

P(B) = 

Bob Cousins, CMS, 2008

Frequentist vs. Bayesian

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

- **This follows simply from the "conditional probabilities":**

$$P(A|B)P(B) = P(A \cap B) = P(B \cap A) = P(B|A)P(A)$$

$$P(A|B)P(B) = P(B|A)P(A)$$

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

Frequentist vs. Bayesian

Bayes' Theorem $$P(\mu|n) = \frac{P(n|\mu)P(\mu)}{P(n)}$$

- $P(n|\mu)$: Likelihood function
- $P(\mu|n)$:posterior probability of μ
- $P(\mu)$: the "prior"
- $P(n)$: just some normalisation

.: Nobody doubts Bayes' Theorem:
discussion starts ONLY if it is used to turn

frequentist statements:

- probability of the observed data given a certain model: $P(Data|Model)$

into Bayesian probability statements:

- probability of a the model being correct (given data): $P(Model|Data)$

- **… there can be heated debates about 'pro' and 'cons' of either….**

P (Data|Theory) ≠ P (Theory|Data)

- Higgs search at LEP: the statement
  - the probability that the data is in agreement with the Standard Model background is less than 1% (i.e. P(data| SMbkg) < 1%)

went out to the press and got turned round to:

~~P(data|SMbkg) = P(SMbkg|data)~~ < 1% P(Higgs|data) > 99% !

WRONG!

An easy example:

Theory = fish (hypothesis) .. mammal (alternative)
Data = swimming or not swimming

P(swimming | fish) ~ 100% but P(fish | swimming) = ??

... OK... but what does it SAY?

The correct frequentist interpretation

we know: P (Data | Theory) ≠ P (Theory | Data)

Bayes Theorem: $$P(\text{Data}|\text{Theory}) = P(\text{Theory}|\text{Data}) \frac{P(\text{Theory})}{P(\text{Data})}$$

Frequentists answer ONLY: P (Data | Theory)

in reality - we are all interested in P(Theory…)

We only learn about the "probability" to observe certain data under a given theory. Without knowledge of how likely the theory (or a possible "alternative" theory) is .. we cannot say anything about how unlikely our current theory is !

We can define "confidence levels" … e.g., if P(data) < 5%, discard theory.

- can accept/discard theory and state how often/likely we will be wrong in doing so. But again: It does not say how "likely" the theory itself (or the alternative) is true
- note the subtle difference !!

Frequentist vs. Bayesian

- Certainly: both have their “right-to-exist”
  - Some “probably” reasonable and interesting questions cannot even be ASKED in a frequentist framework :
    - “How much do I trust the simulation”
    - “How likely is it that it will raining tomorrow?”
    - “How likely is it that climate change is going to…
  - after all.. the “Bayesian” answer sounds much more like what you really want to know: i.e.
    “How likely is the “parameter value” to be correct/true ?”
- <u>BUT:</u>
  - NO Bayesian interpretation exist w/o “prior probability” of the parameter
    - where do we get that from?
    - all the actual measurement can provide is “frequentist”!

Bayesian Prior Probabilties

- **“flat” prior $\pi(\theta)$ to state “no previous” knowledge (assumptions) about the theory?**
  - **often done, BUT WRONG:**
    - e.g. flat prior in M_{Higgs} -> not flat in M^2_{Higgs}
  - **Choose a prior that is invariant under parameter transformations**

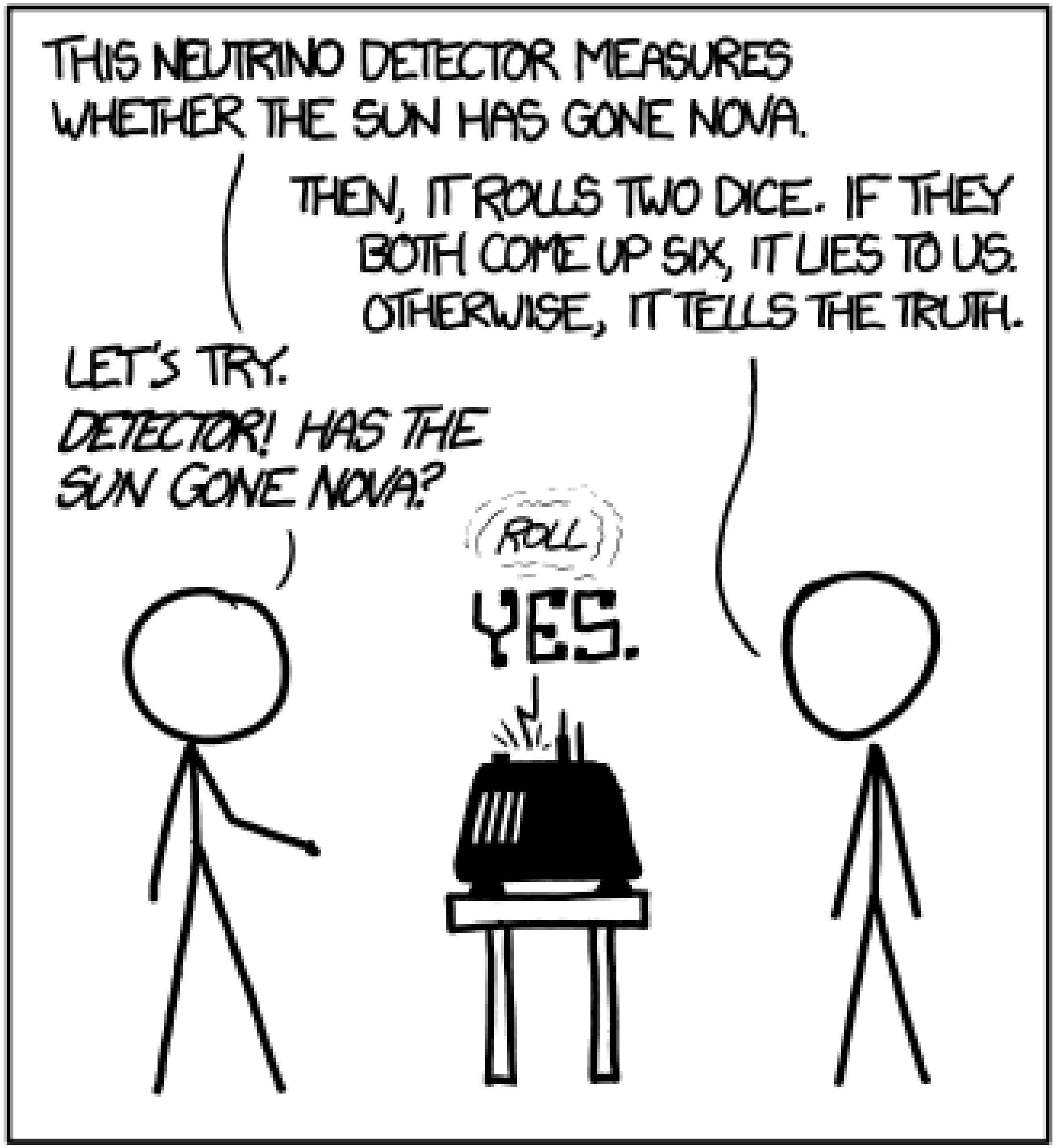
  Jeffrey’s Prior → objective Bayesian”:
    - “flat” prior in Fisher’s information space
    - $\pi(\theta) \propto \sqrt{\mathrm{I}(\theta)}$ ($\pi(\theta) \propto \sqrt{\det \mathrm{I}(\theta)}$ if several parameters)

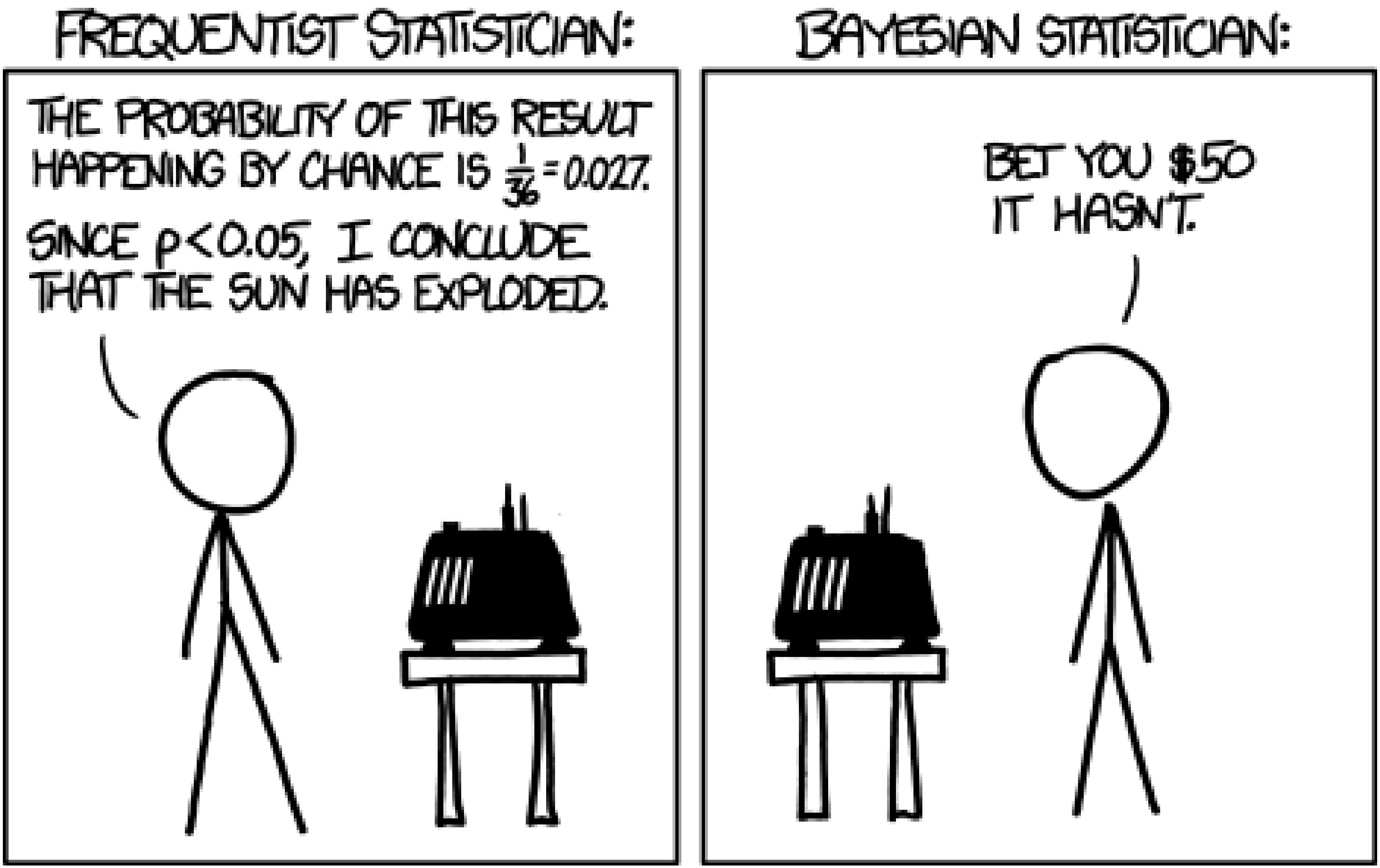
$I(\theta) = -E_x\left[\frac{\partial^2}{\partial\theta^2} log(f(x;\theta)\right]$:

- $f(x;\theta)$: Likelihood function of θ, probability to observe x for a give parameter θ
- amount of “information” that data x is ‘expected’ to contain about the parameter θ

- **personal remark: nice idea, but “WHY” would you want to do that?**
  - **still use a “arbitrary” prior, only make sure everyone does the same way**
  - **loose all “advantages” of using a “reasonable” prior if you choose already to**

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




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:
BET YOU $50 IT HASN'T.