NI-MLP-05: Applied Bayesianism

Petr Paščenko

23. 10. 2023

Covid test

PROFINIT

Suppose you go on holiday and get tested for covid and you get a positive result. What is the probability that you are covid-positive?

- the declared test accuracy (TPR and TNR) is 99%
- overall population positivity is about 1%

Answers:

a) 1%	b) 10%	c) 50%
d) 90%	e) 99%	f) can’t tell

Outline

PROFINIT

1. Probabilistic recap
2. Probabilistic subjectivity
3. Odds and log odds
4. Bayes theorem formally
5. Bayes theorem as a probabilistic rule of three
6. Evidence iteration
7. Rationalistic consequences and real life applications
8. Literature

Thomas Bayes
1701 – 1761

[ˈbeɪz]

Probabilistic recap

Definition of probability

PROFINIT

- Kolmogorov definition (second half of 20th century)
  - based on the set and measure theories
  - axiomatic (probability falls from heaven), great for mathematics
- Frequentist definition (Bernoulli 17. century, Poisson, Fischer)
  - “ratio between positive and all events in a long run”
  - empiric (probability raises from repeated experiment), great for children
- Bayesian definition (Thomas Bayes, Pierre-Simon Laplace,18. century)
  - rationalistic (probability is a measure of our limited knowledge)
  - great for decision making (AI, forecasting, every day rationality)

Kolmogorov probability space

$$\begin{pmatrix} \Omega \\ F \\ P \end{pmatrix} \sim \begin{matrix} \text{sample space} \\ \text{event space} \\ \text{prob. function} \end{matrix}$$

Frequentist probability

$$p = \lim_{N \to \infty} \frac{N_p}{N}$$

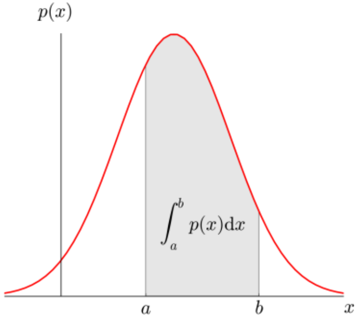
Bayesian probability

$$p = prior \; * LR$$

What is the probability of getting exactly one head by tossing two fair coins?

50%

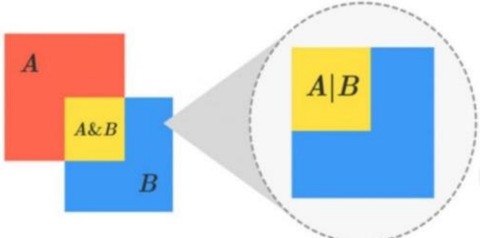
Probabilistic recap



- Probability is a distribution (non neg. function with unit integral)
  - special case $P \in \mathbb{R}$ for binary event, $0 \sim$ impossibility, $1 \sim$ cerntainty
- Conditional probability $P(A|B) = P(A \cap B)/P(B)$
  - probability of event A in case we know B is true
  - *probability of raining given the fact, we are on Sahara*
- Independence $P(A \cap B) = P(A) * P(B)$ $P(A|B) = P(A)$
  - events A and B are independent if their joint prob. is a product of their marginal probs.
    - *probability winning lottery on your birthday*
  - also, knowing B does not change the probability of A

What is more likely? $P(A \cap B) \leq P(A)$

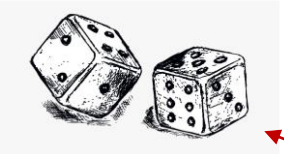
- *Mr. F. has had one or more heart attacks.*
- *Mr. F. has had one or more heart attacks and he is over 55 years old.*



Objective and subjective probability

PROFINIT

A Philosopher



A pair of dice

What is the probability of the pair of dice giving at least 10?

$^1/_6$

P is a property of the object (frequentist approach)

P is a property of the subject (bayesian approach)

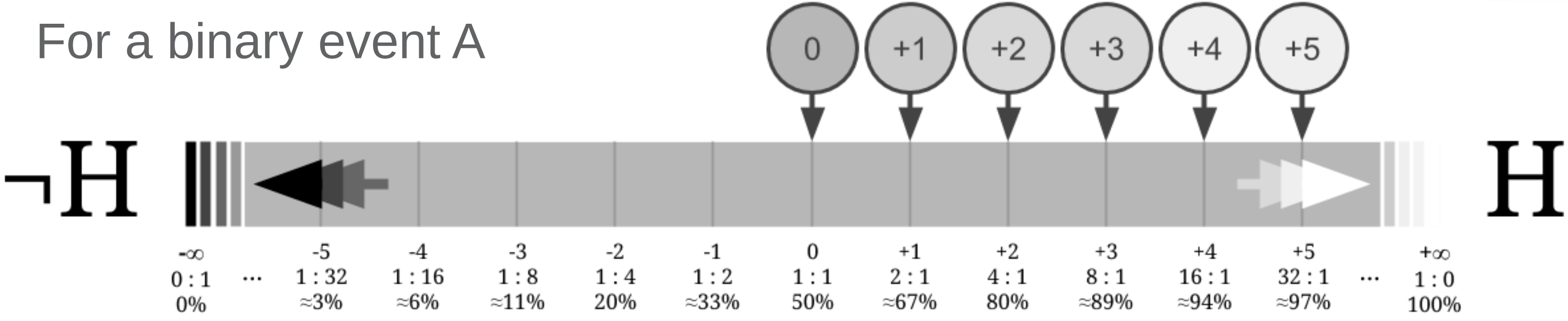
- What if you know…
  - they are the pair of lucky dice found in a famous cheat player pocket?
  - a complete stranger offers you $200 for them
  - you roll them once and you got 12

  - you roll them 100 times and you got 70% times result over 10

Probability axes

PROFINIT

- For a binary event A



probability axis

$$P_A = \frac{N_A}{N_A + N_{\neg A}}$$

odds axis

$$O_A = N_A : N_{\neg A}$$

log odds axis

$$L_A = \log_2\left(\frac{N_A}{N_{\neg A}}\right)$$

33%	1:2	−1 bit
20%	1:4	−2 bit
11%	1:8	−3 bit

50%	1:1	0 bit

67%	2:1	+1 bit
80%	4:1	+2 bit
89%	8:1	+3 bit

Bayes Theorem

Bayes theorem

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

PROFINIT

- $P(H|E)$ – probability of hypothesis H given observation / evidence E
- $P(E|H)$ – probability of observing E given H aka likelihood of H given E
- $P(H)$ – prior probability of hypothesis H
- $P(E)$ – overall probability of observing evidence E

$$P(E) = P(E|H) * P(H) + P(E|\neg H) * P(\neg H)$$

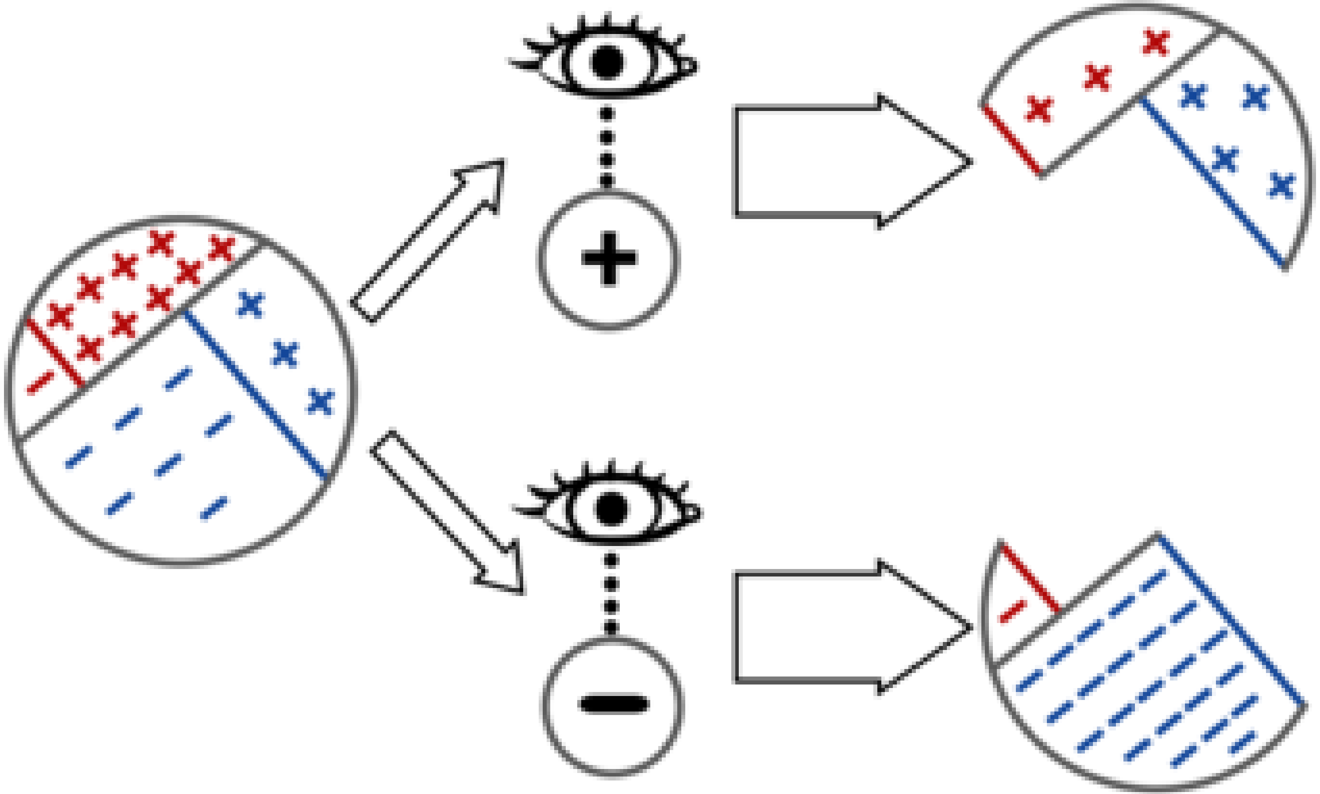
Suppose you are going to holiday and you get tested for covid and you get positive result. What is the probability of you being covid-positive?

50%

the declared test accuracy (tpr and tnr) is 99%, overal population postivity is 1%

Bayes theorem visual intuition

PROFINIT



Probability vs Likelihood and Bayes Factor

PROFINIT

Probability
probability of Hypothesis H being true given I see the Evidence E

$$P(H|E)$$

probability of you being the shooter given your fingerprints found on the smoking gun

Likelihood
probability of seeing Evidence E given Hypothesis H being true

$$P(E|H)$$

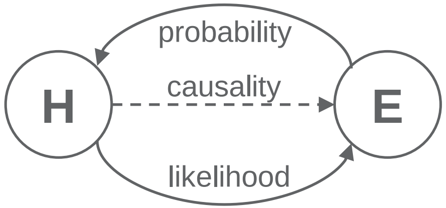
$$L(H|E)$$

probability of your fingerprints found on the smoking gun given you are the shooter

Likelihood Ratio ~ Bayes Factor
How much more likely is the Evidence E given H compared to not H

$$\frac{P(E|H)}{P(E|\neg H)}$$

how much more likely you are the shooter if we found your fingerprints on the smoking gun



Bayes theorem for odds

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

PROFINIT

$$\frac{P(H|E)}{P(\neg H|E)} = \frac{\frac{P(E|H) * P(H)}{P(E)}}{\frac{P(E|\neg H) * P(\neg H)}{P(E)}} = \frac{P(H)}{P(\neg H)} * \frac{P(E|H)}{P(E|\neg H)}$$

odds = prior odds times likelihood ratio

$$\begin{array}{lll} prior\ odds & P(C): P(\neg C) & 1:99 \\ likelihood\ ratio & P(T|C): P(T|\neg C) & 99:1 \\ \hline posterior\ odds & P(C|T): P(\neg C:T) & 1:1 \end{array}$$

Log odds version

$$\begin{array}{c} \log_2(1/99) = -6.6\ bit \\ \log_2(99/1) = +6.6\ bit \\ \hline \log_2(1) = 0\ bit \end{array}$$

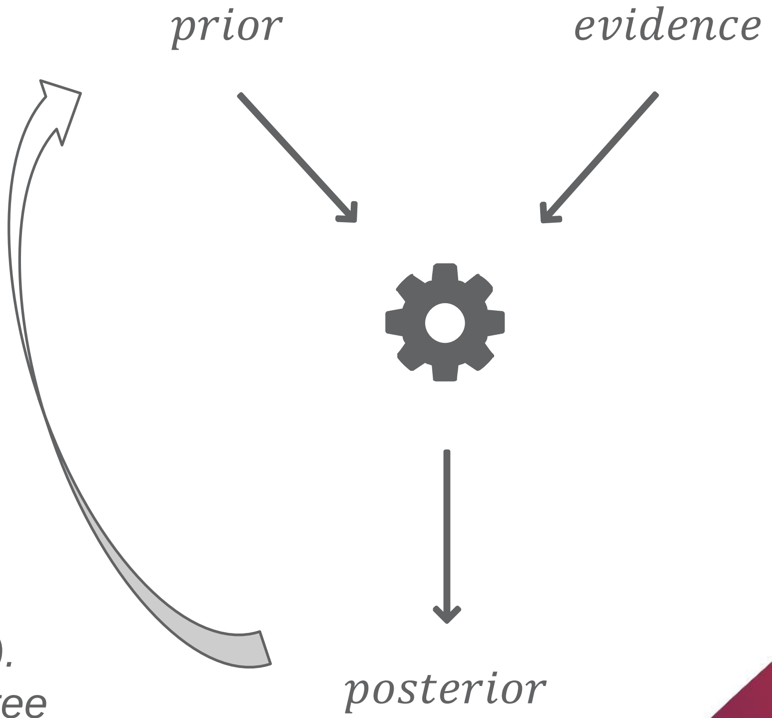
Suppose you live in Scotland (rainy 80% of days). What are the odds of being sunny tomorrow if weather forecast (accurate 2/3 of time) say so?

$$1:2 \sim 33\%$$

Evidence iteration

PROFINIT

- Bayes theorem works iteratively
- Posterior reflect our best knowledge after observing the evidence
- When considering next evidence the posterior becomes next prior
- Expects independent evidence
  - rarely happens in real world
- Continual belief improvement

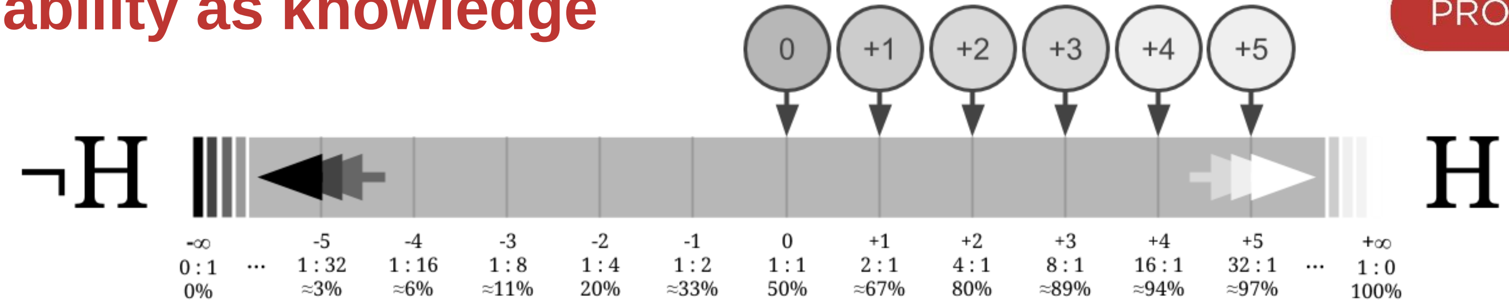


Suppose you live in Scotland (rainy 80% of days). What are the odds of being sunny tomorrow if three independent weather forecasts (accurate 2/3 of time) say so?

$2:1 \sim 67\%$

Probability as knowledge

PROFINIT



- Seeking the truth we move our position on the probability axis
- We start in the middle, we know nothing, having 0 bits of knowledge
- Observation 2 times more likely if H=true moves us 1 bit right and vice versa
- The axis is linear to log-odds but shrinking to percentages
  - Distance between 98% and 99% is much greater than distance between 50% and 51%
- The majority of human senses are logarithmic, the sense of probability is logarithmic as well.

Rationalistic consequences

Rationalistic consequences

Do I have a dragon in my basement?

- **ECREE**: **extraordinary claims require extraordinary evidence**
- Extraordinary claim = very unexpected
  - low prior (-10 bits) req. strong evidence (10 bits)
- Extraordinary evidence
  - $LR = \frac{\text{Probability of seeing the evidence if claim is true.}}{\text{Probability of seeing the evidence if claim is false.}}$
  - The key is in a very small denominator

Evidence I provide:

- a tape with the dragon roaring
- a scale of a dragon skin
- to burn the city flying on a dragon

- Examples of extraordinary claims supported by weak evidence
  - Conspiracy theories (minor inconsistencies in the facts, noisy observations)
  - Paranormal physics (irreproducible experiments, random coincidencies)
  - Religions (third hand testimony, some old books, single source of wisdom)
- **OCROOE**: **ordinary claims require only ordinary evidence**

Rationalistic consequences

$$P(H|E) = \frac{P(E|H) * 0}{P(E)} = 0$$

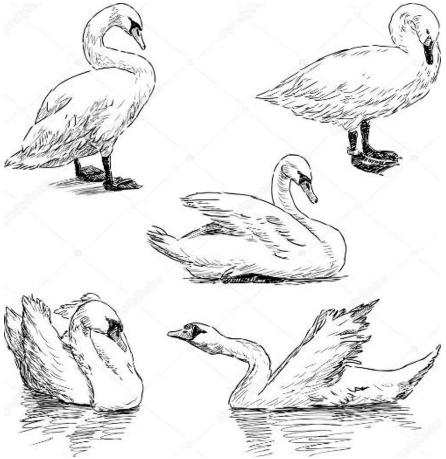
- **Prior P = 1 and P = 0 are taboo**
  - No matter how strong evidence you observe,
    your belief does not change – i.e. your mind is broken
  - The Bayesian definition of fanaticism is an infinite prior
- **Evidence is double sided**
  - Every piece of evidence should move
    us in opposite direction than its absence.
  - However of different size

Inquisition logic
- to confess proves the guilt
- to refuse the confession proves it even more

- **The absence of evidence actually is an evidence of absence**
  - Suppose you search car keys in your house
  - Bayesian argument for P is not NP
  - Very weak evidences are often neglected.
  - Generalization of Popper's Falsification Principle

Counting the evidence

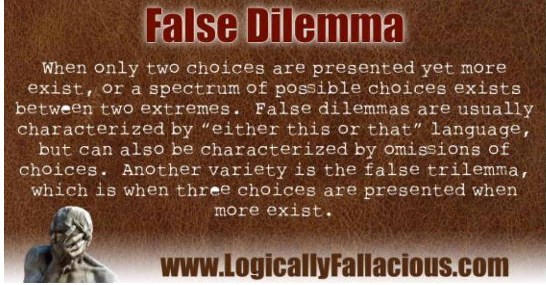
› **Hypothesis: all swans are white**



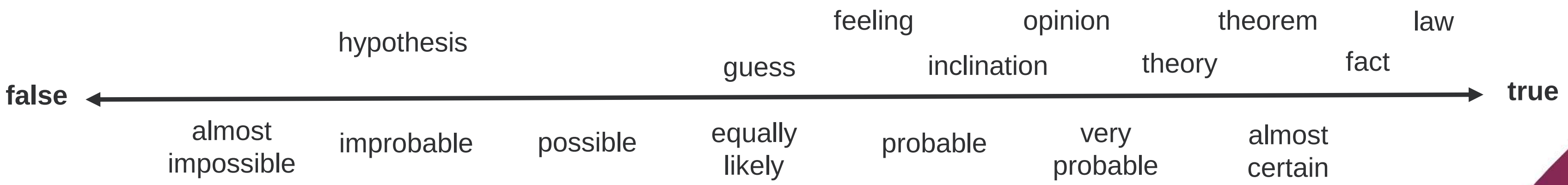
prior	$P(H):P(\neg H)$	$1:1$	w – to our best knowledge it is 1/2
I see 1 white swan	$P(W\|H):P(W\|\neg H)$	$1:w$	
posterior	$P(H\|W):P(\neg H\|W)$	$2:1$	
prior	$P(H):P(\neg H)$	$2:1$	$w \sim 2/3$
I see 2nd white swan	$P(W\|H):P(W\|\neg H)$	$1:w$	i.e. 1:w = 1:2/3 = 3:2
posterior	$P(H\|W):P(\neg H\|W)$	$3:1$	
prior	$P(H):P(\neg H)$	$3:1$	$w \sim 3/4$
I see 3rd white swan	$P(W\|H):P(W\|\neg H)$	$1:w$	i.e. 1:w = 1:3/4 = 4:3
posterior	$P(H\|W):P(\neg H\|W)$	$4:1$	
prior	$P(H):P(\neg H)$	$4:1$	
I see a black swan	$P(B\|H):P(B\|\neg H)$	$0:1-w$	$w \sim 4/5$ i.e. $1-w = 1/5$
posterior	$P(H\|B):P(\neg H\|B)$	$0:1$	**FALSIFIED!**

w – ratio of white swans

Rationalistic consequences



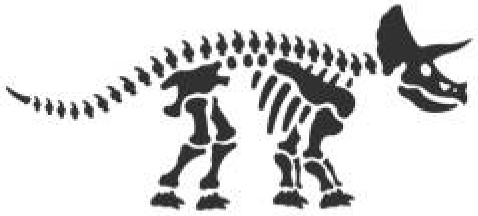
- **Facts vs. opinions is mostly a simplification**
  - There is no fundamental difference between fact and opinion
  - Every statement has just different prior
  - The difference among priors however can be huge
    - Pythagoras theorem is wrong: $1:10^{10}$
    - ČR will win a medal in Hockey World Cham.: $1:3$
  - Still it is useful to have different names for different prior classes
    - law, fact, theorem, theory, hypothesis, opinion, inclination, feeling

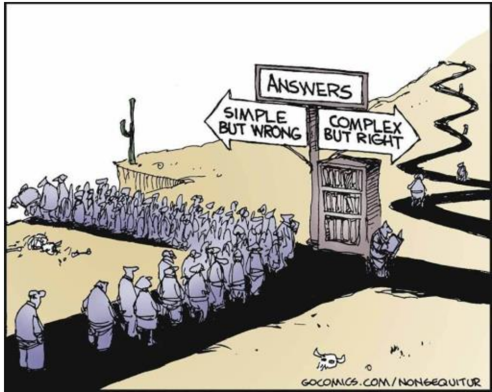


Rationalistic consequences

PROFINIT

- **Making the prior is a generalization of Occam's razor**
  - *All things being equal, the simplest solution tends to be the best one*
  - simple is not easy
    - simple means fewer unobserved assumptions
    - not easily to comprehend!
      - Genesis is much more easy to comprehend than Big Bang Theory
  - Do I have dragon in my basement or do I just lie (or got mad,…)?
    - people lie all the time
    - new fantastic creatures are discovered rather rarely

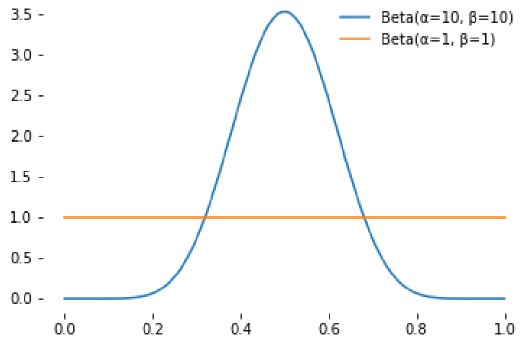




How to make the prior?

PROFINIT

- **Uninformative prior**
  - uniform distribution of probability
  - We pretend to be objective and know nothing
  - Hardly often rational
  - *Either I win the lottery or no, so 50:50*
- **Informative prior**
  - We accept we can know something apriori
  - Cognitive burden for the agent
  - Fits real world situations
  - *Lottery has 10M tickets with just one winning*



Techniques for probability estimates
1. Introspection

- **Measure your own surprise**
  - What answer do you really expect from oracle?
- **Bet on it**
  - At which ratio are you betting on it
  - Here and now, actual medium size (lunch) money
- **Imagine Hypothetical Evidence**
  - What (random) evidence would make you switch your belief?
  - How likely is that evidence?

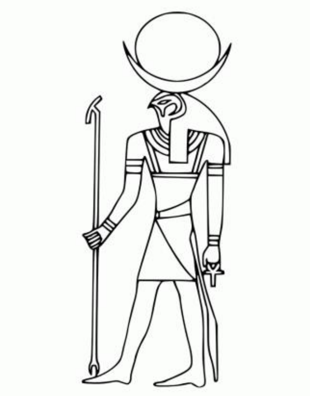
What is the probability you would get the Hogwards letter?

Scot Alexander: Codex – Techniques for probability estimates

Techniques for probability estimates
2. Enumerative

- **Convert to a Frequency**
  - How often do you see a red car going through your street
  - What is the probability the Sun will not rise tomorrow?
  - **Fermization** – rough numerical estimates based on variable decomposition
- **Find a Reference Class**
  - How often a well established scientific truth turned out to be false before?
- **Make Multiple Statements**
  - What is the probability Allah, Zeus, Baal, Ra, Jesus, Jupiter, exists?

What is the probability you will meet a friend in metro this afternoon?



Scot Alexander: Codex – Techniques for probability estimates

Remarks to prior

PROFINIT

- **Too strict prior**
  - Pythagoras theorem is wrong: $1:10^{10}$ $10\,000\ \times$
    - imagine a book with so many lines only one of them to be false.
  - For any statement worth considering anything more than $1:1000$ is too strong

- **Prior does not really matter after all**
  - With enough evidence, every reasonable prior can be overturned
- **What if you don’t like making up the prior**
  - **Usually pulling numbers out of your arse and using them to make a decision is better than pulling a decision out of your arse.**

Remarks to posterior

PROFINIT

- **Too strict posterior**
  - Mr X will win the president elections in CR: $1:100\,000$
  - What is the real probability?
    - probability given by model times probability the model is not significantly flowed
    - it is much higher probability the model is significantly flowed than $1:100\,000$
- There are limits of certainty the bayes theorem can deliver in practice.
- **Internal vs External confidence**
  - internal – inside the debate
  - external – meta level confidence about the debate as such
  - every debate needs fixed and moving parts, sometimes fixed parts are not really fixed and moving parts are not fully moving…
    - Einstein: time is relative, space is curved, weight changes with speed etc.

Remarks to prior

PROFINIT

- **Too strict prior**
  - Pythagoras theorem is wrong: $1:10^{10}$ $10\,000\ \times$
    - imagine a book with so many lines only one of them to be false.
  - For any statement worth considering anything more than $1:1000$ is too strong

- **Prior does not really matter after all**
  - With enough evidence, every reasonable prior can be overturned
- **What if you don't like making up the prior**
  - **Usually pulling numbers out of your arse and using them to make a decision is better than pulling a decision out of your arse.**

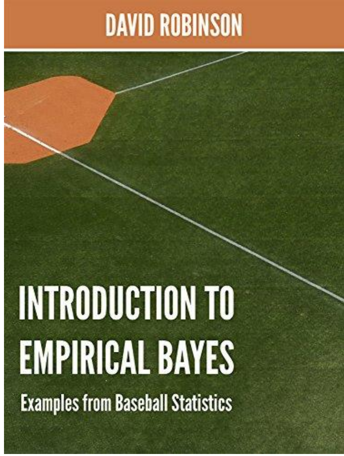
Literature

PROFINIT

- Eliezer S. Yudkowsky: An Intuitive Explanation of Bayes' Theorem
  - A bit more comprehensive introduction to Bayes Rule
- Allen B. Downey: Think Bayes 2
- Cameron Davidson-Pilon: Probabilistic Programming and Bayesian Methods for Hackers
- Eliezer S. Yudkowsky: RATIONALITY: A-Z
  - WHAT DO WE MEAN BY "RATIONALITY"?
- Scott Alexander: The Codex (Probability and Predictions)
- Phillip Tetlock: Superforecasters
- Nate Silver: Signal and Noise
- David Robinson: Introduction to empirical Bayes

Real life applications

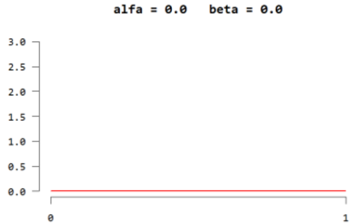
Empirical bayes



- **Restaurant rating app**
  - 0 – 5 stars for worst and best possible restaurant
- What restaurant is better?
  - 1 rating (5 stars), 10 ratings (avg. 4.5), 100 ratings (avg. 4.2), 1000 ratings (avg. 3.9)
- **Baseball player statistics**
  - BA – batting average (hits/at bat)
- Which hitter is better?
  - 1 hit of 1 at bat, 30 hits of 90 at bats, 270 hits of 1000 at bats?
- **Decision based on imperfect information**
  - Because of small and varying sample size – very typical real life situatuation

? True ratio approximation
? Measure the uncertainty

Beta distribution



$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(x,y)}$$

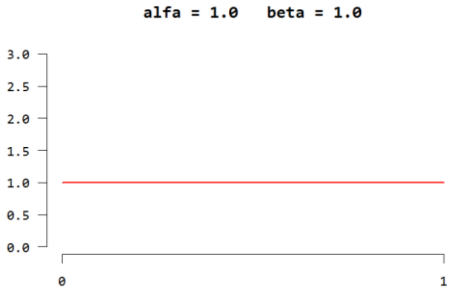
– for 2 parameters $\alpha, \beta \in [1, \infty)$

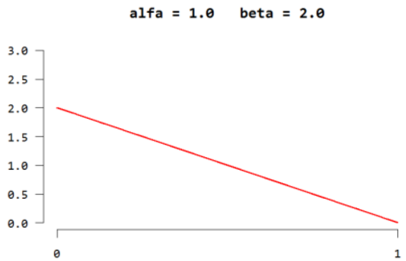
› $E(X) = \frac{\alpha}{\alpha+\beta}$ generalized ratio $\alpha : \beta$

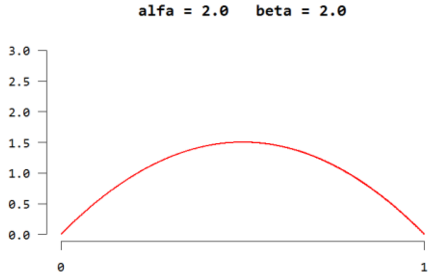
$$Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

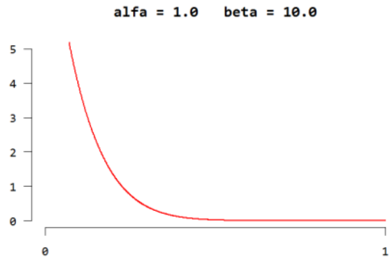
- std. dev. $\sim {}^{1}/\sqrt{(\alpha+\beta)}$

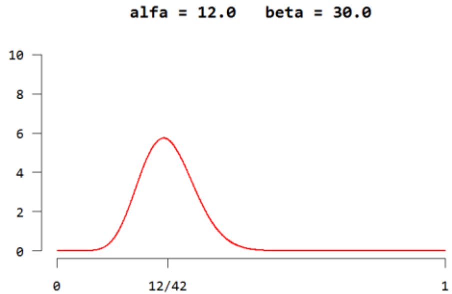
› approximation of probability

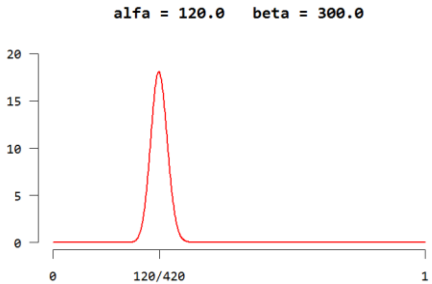










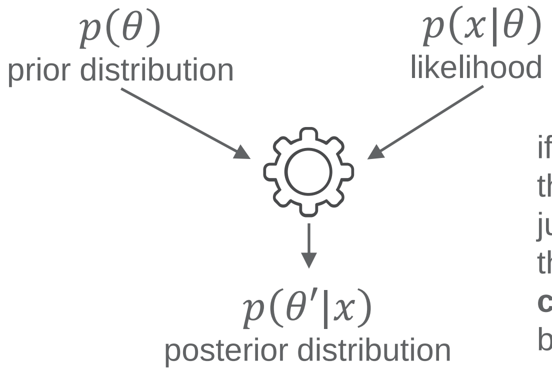


Where is the Bayes?

› Prior
  – $\text{Beta}(\alpha = 4, \beta = 7)$
    • $\text{EX} = 4/11$
    • $sd = 0.139$

› Evidence
  – hit $(\alpha, \beta) \to (\alpha + 1,\ \beta)$
  – miss $(\alpha, \beta) \to (\alpha,\ \beta + 1)$

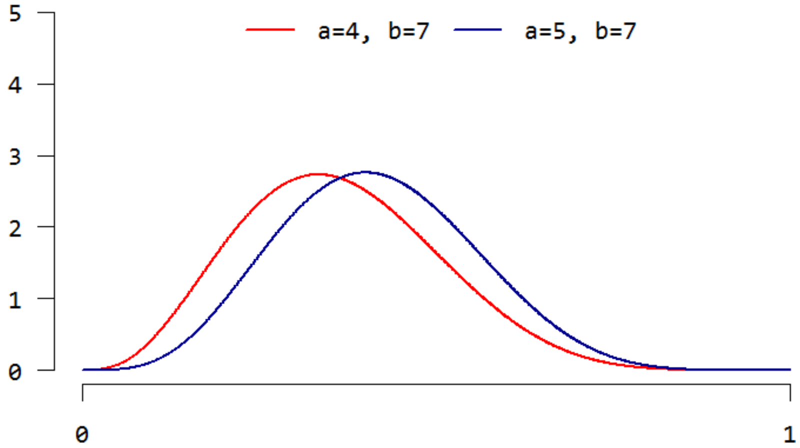
› Posterior
  – $\text{Beta}(\alpha = 5, \beta = 7)$
    • $\text{EX} = 5/11$
    • $sd = 0.137$



if prior and posterior are the same function with just different parameters, the function is called the **conjugate prior** and bayes update reduces to

$$\theta \to \theta'$$

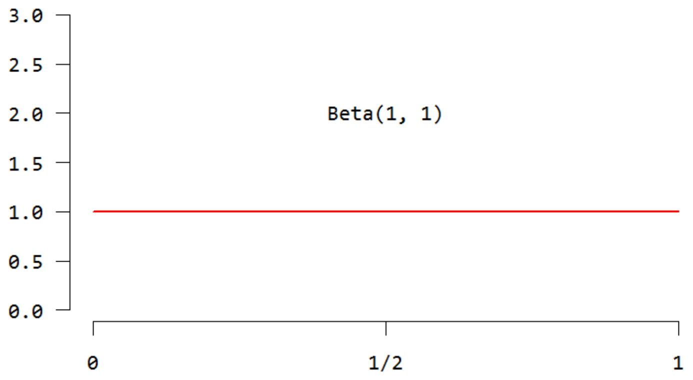
The player hits the ball



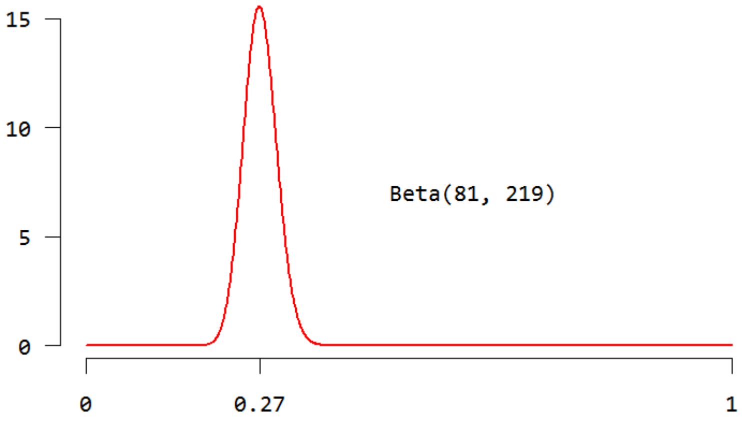
Where is the empirical?

- Uninformative prior distribution
  - $\text{Beta}(\alpha = 1, \beta = 1) \sim U(0,1)$
    - $\text{EX} = 1/2$
    - $sd = \sqrt{1/12} \sim 0.29$
- What if we know
  - most of players BA is between 0.21 – 0.35
- Empirical prior distribution
  - $\text{Beta}(\alpha = 81, \beta = 219)$
    - $\text{EX} = 0.27$
    - $sd = 0.143$

Uninformative prior distribution



Empirical – hit rate – prior distribution



Multi-Armed bandit

Multi-Armed bandit

- You enter a casino with n coins. You can not keep any of those but you can throw them into m machines. Every machine returns the coin with unknown probability p_i. You can take all returned coins with you. Maximize your return.
- Mathematical abstraction of set of real world problems
  - buying coffee/wine/whisky of various brands
  - hiring employees from various schools
  - watching movies from various directors
  - treating patients with different medications
- Inevitable tradeoff between exploration and exploitation
  - both extremes are bad, the optimum is somewhere in between

Cameron Davidson-Pilon: Bayesian Methods for Hackers

Multi-Armed bandit – the strategy

PROFINIT

- Bet on Luck!
  - randomly pick a machine and throw it all in
- Don’t put all eggs…
  - regularly distribute coins among machines
- Hire and Fire!
  - switch machine if it haven't returned the coin
- Explore first, then exploit!
  - to spend some of the coins to approximate the return rates
  - then throw all remaining coins to the machine with maximal expected return
    - ? where to put the threshold

Multi-Armed bandit – Bayes sampling strategy

PROFINIT

- Approximate return rate $p_i \sim \mathrm{Beta}(a_i, b_i)$
  - with a_i, b_i being counts of returned/lost coins of machine i
  - initiate $a_i, b_i = (1, 1)$ for every machine
- Strategy
  1. randomly sample x_i from $\mathrm{Beta}(a_i, b_i)$ for every machine
  2. find $k = \mathrm{argmax}_i\ (x_i)$
  3. pick machine k, throw coin and update a_i, b_i
  4. repeat until you have coins
- At beginning, we are sampling randomly, as soon as we get some information, we slightly incline towards higher expected returns.
- If single machine achieve statistically significant dominance, we continue sampling from this machine only.

Multi-Armed bandit – modifications

PROFINIT

- Multilevel bandit
  - Two casinos each with its set of bandits. One of them possibly with more generous return rates.
- Forgetting
  - if a performance drift is expected, we can apply forgetting rate.
- Different distribution of reward
  - instead of simple binary return we can model normal returns or any other probability distribution

PROFINIT

Questions?