- **Short history of Bayes' theorem**

reference:

– S B McGrayne: The theory that would not die: how Bayes' rule cracked the enigma code, hunted down Russian submarines, and emerged triumphant from two centuries of controversy.

Yale University Press, 2011.

- 1740: reverent Bayes.

  - Studied theology in Edinburgh, but was also ’amateur’ mathematician.
  - 1748: David Hume (Edinburgh): ”we can rely only on what we learn from experience”
    - Dilemma in those times: we cannot be sure that a specific cause will lead to a specific effect
→ only probable causes with probable effects.
    - **Newtonian mechanics had promised something exact**!

The question: probabilities of causes?

- Probability calculus could solve: P(effect | cause).
- But not: P(cause | effect)
  - This was called ”inverse probability”
  - ”What is the probability that a dice is weighted if we get 5 times six in 5 trials?” → ”then what is the probability to get a six in the next trial?”

– *Cause, effect, uncertainty…*

→Bayes , sometime between 1746-1749 **Heureka!**

Solution by using a specific example.

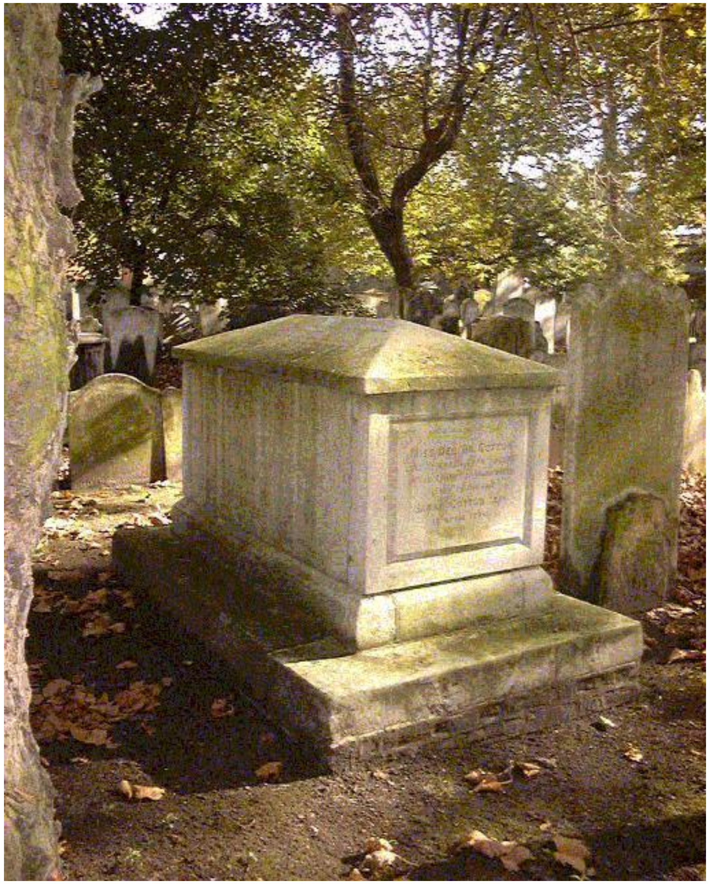
- The example:
  - Imagine a square, flat table.
  - An assistant throws ”randomly” a ball on the table and takes note of where it stops.
  - The assistant throws new balls and tells whether they stop to the left or right from the first ball.
  - If all balls stop to the right , what can we say about the position of the first ball?

- Bayes figured out:
  - The more balls are thrown, the better we should know the position of the first ball.
  - *This is a learning process*.
  - Before observations, any position is as possible as any other → Uniform(0,1) distribution.
    - Understandable if the first ball is thrown ”randomly” → can this be generalized?

- The example in modern notations:
  - Observations X have conditional distribution $P(X \mid p)$:
    - Binomial(N,p) where p = unknown position in [0,1]
  - Want to calculate $P(p \mid X)$
  - Note that: $P(X,p) = P(X|p)P(p) = P(p|X)P(X)$
  - Solve: $\boxed{P(p|X)} = P(X,p)/P(X) = \boxed{P(X|p)P(p)/P(X)}$
  - Nowadays known as Bayes' formula!
    - $P(X|p)$ is easy to write and calculate: binomial probability.
    - $P(p)$ is **uniform density function** (prior)
    - $P(X)$ is normalizing constant $= \int P(X|p)P(p)\, dp$ = const.

Bayes solved the inverse problem for binomial model

- Bayes' solution:
  - We obtain P(p |X), posterior probability density of p.
  - This is Beta(X+1,N-X+1)
  - Bayes left it forgotten in the drawer…
  - After Bayes had died, 1761, Richard Price studied the papers and published them.



- **But Price first edited and corrected the manuscript for 2 years.**
  - ”an imperfect solution of one of the most difficult problems in the doctrine of chances”
  - It gave a response to Hume’s critique of causes and effects.
  - Royal Society’s Philosophical Transactions: *”An Essay toward solving a Problem in the Doctrine of Chances”. 1763.*
  - Bayes theorem → Bayes-Price theorem ?

- **Bayes** did not create modern concepts such as Bayesian statistics or Bayesian inference. These were introduced in 1950’s.
- **Bayes** did not provide any other examples, or more general interpretations.

Laplace

- After Bayes and Price, hardly anyone touched the problem, *Until*:

- ”The man who did everything”

Pierre Simon Laplace

1749-1827

- D’Alembert urged Laplace to study astronomy.
- Dilemma of the times: was the universe stable?
- Newton’s **theory vs observations**.

→ theory could be validated by exact observations.

… *exact*?

- Laplace noted: **big problem was the data**!
- Empirical planetary data was from ancient studies from China 1100 BC, Mesopotamy 600 BC, Greece 200 BC, Rome 100 AD, Arabia 1000 AD.
- Lots of errors, missing data, imperfections, uncertainty.

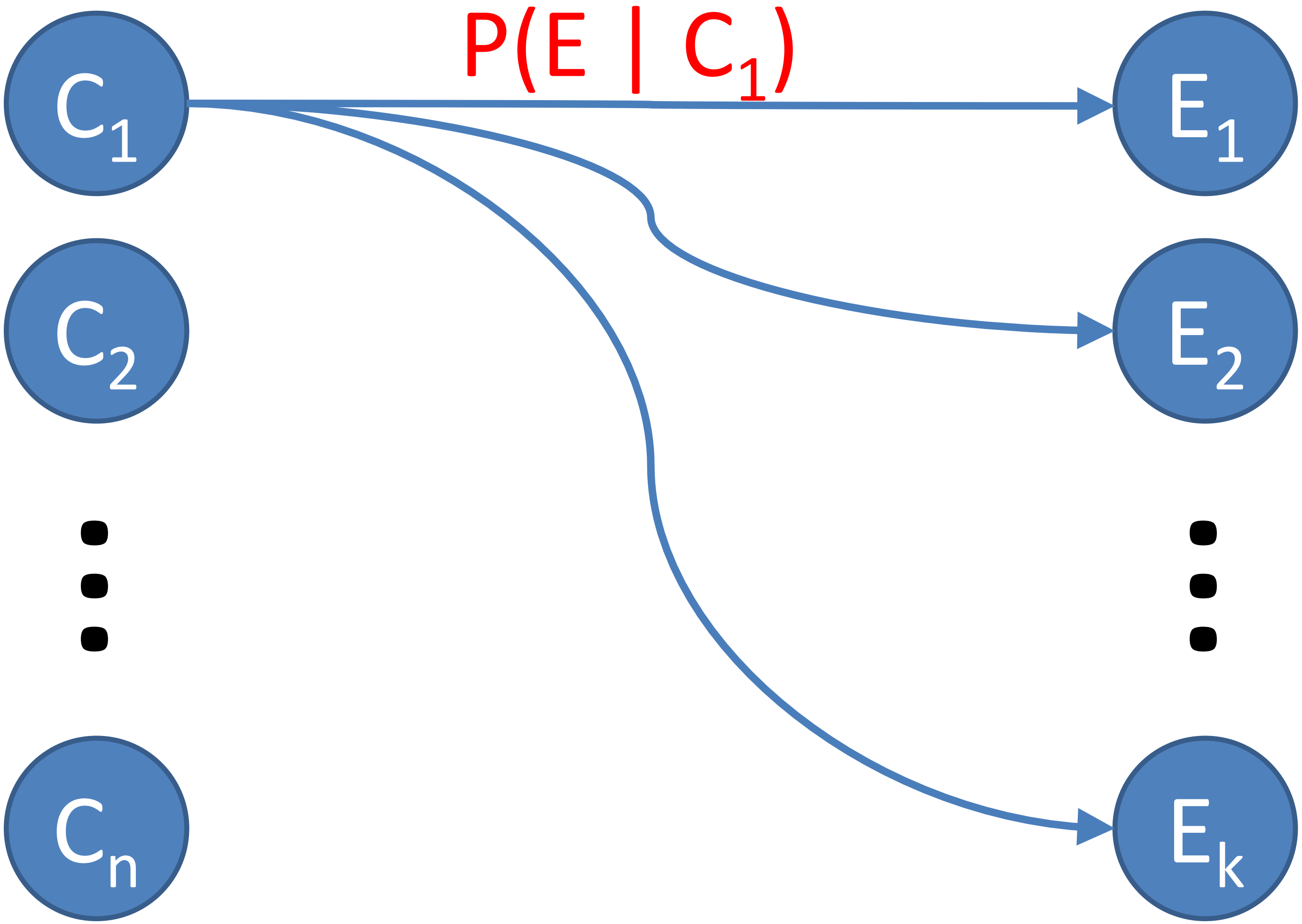
- Also new observations were gathered.
  - Transit of Venus, observed at 120 locations on Earth.
  - By comparing these French mathematicians estimated the distance of Earth from the Sun.

→Increased need to analyse complicated empirical data.

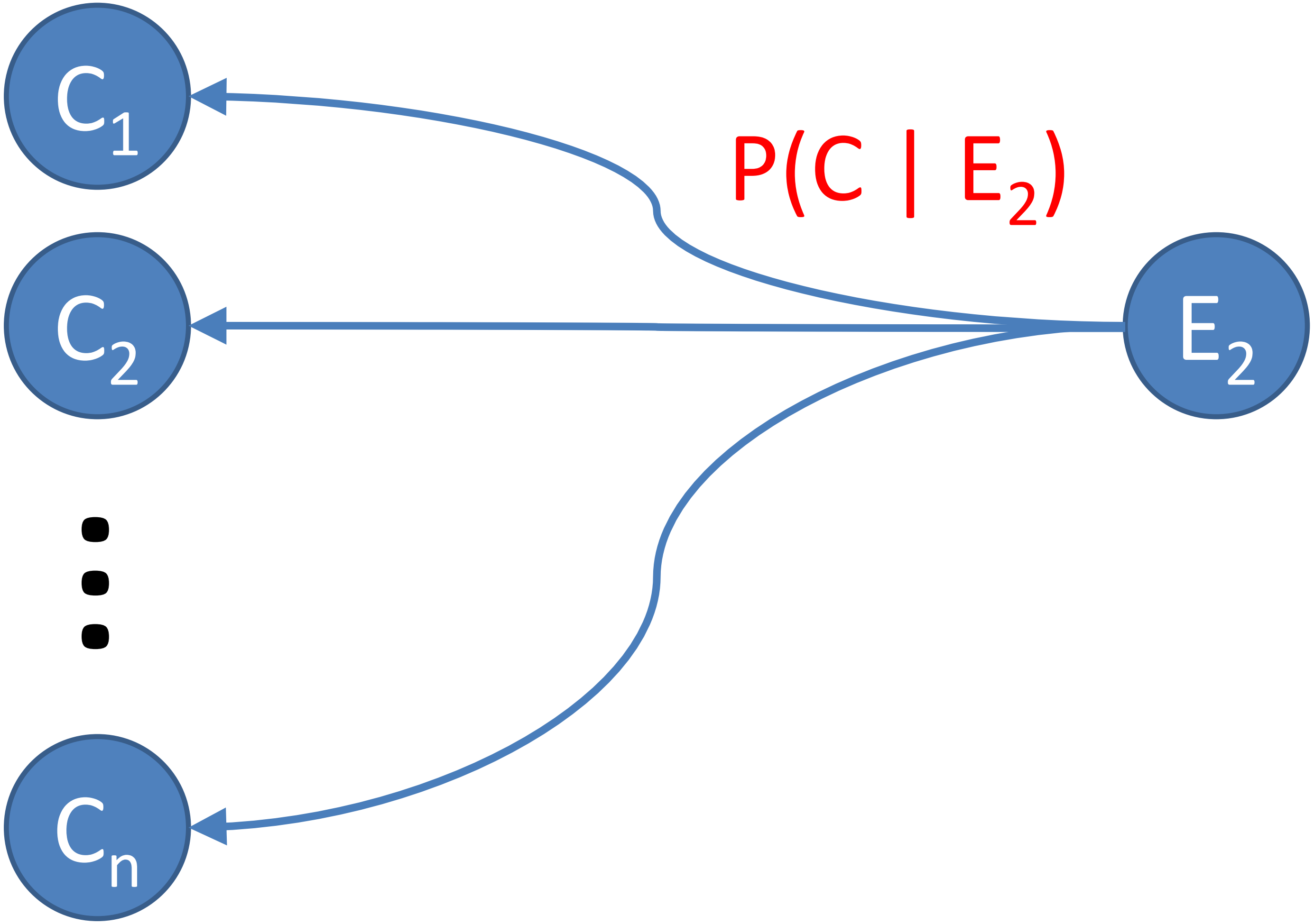
- Laplace thought probability could be the tool for dealing with **uncertainties**.
  - Found a book about probabilities in games of chances: de Moivre: The Doctrine of Chances.
  - Bayes had read the earlier edition of the same book.
  - Laplace: *Mémoire on the Probability of the Causes Given Events.*
    - Contains the first version of what we now call Bayes theorem !

- Even so, Laplace did not write formally Bayes theorem, but described it in words.
  - Idea: enumerate all possible reasons C, and compare them after we observe E.
  - Formally expressed:

$$P(C_i \mid E) = P(E \mid C_i) \;/\; [P(E \mid C_1)+\ldots+P(E \mid C_n)]$$


P(E | C1)
C1
C2
Cn
E1
E2
Ek

After observing E_2



- With this principle Laplace was able to do everything that Bayes could have done.
  - As long as one assumes all reasons C are equally possible before observing E.
  - **Voilà !** → general method for any empirical research!
- BUT: mathematical solutions in real problems proved to be difficult even for Laplace.
  - Even today the computational burden shadows applications of Bayesian methods!

- 1781 Price visits Paris and tells about Bayes' original theorem.
- Later, more challenges:
  - Equal prior probabilities criticised.
  - **Serious** technical computational problems in practice.

- New applications: 1771 French provinces begin reporting birth and death statistics to Paris.
- Apparently, more boys were born than girls, **but X % ?**
- Binomial model, lots of data (big N).

→ Laplace tries to estimate X.

- But assuming X=52%, and observing 58000 boys, need to evaluate 0.52^58000, and similarly for girls.
  - Difficult even for Laplace.
  - Need to approximate this somehow.

- Laplace collected birth and death statistics from many places and combined with previous data.
  - First real Bayesian analysis, in which new evidence was used to update earlier probabilities.
  - Mathematical model for scientific inference.
  - Conclusion in 1812: ”X>50% seems to be a general law for all humans”.
  - Laplace also estimated the size of French population.

- 1810-1814 Laplace writes more general formula:

$P(C_i \mid E) =$

$P(E \mid C_i)P(C_i) / [P(E \mid C_1)P(C_1)+\ldots+P(E \mid C_n)P(C_n)]$

”It was the formula he had been dreaming about”

→ From Bayes-Price example to Laplace’s general result.

- Laplace and intuition: ”essentially, the theory of probability is nothing but good common sense reduced to mathematics”.
- *What kind of problems Laplace had?*
  - *Data from several sources.*
    - *Many imperfections and uncertainties.*
  - *Nothing like straightforward repeatable experiments.*
  - ***In the end of his career, also developed frequentist approach.***

- *Mechanique Celeste*
- *Exposition du Systeme du Monde*
- *Theorie Analytique des Probabilités*

(St. Julien-de-Mailloc)

Silence after Laplace

- After Laplace 1827-
  - Bayes theorem unpopular: subjective = bad.
  - **More official statistical data collected**: list of objective facts, mathematical analysis not thought important.
    - →**”Facts, pure facts”, ”objective frequency”**
    - →**”Statistician has nothing to do with causation”**
  - Theoreticians buried Bayes, uniform prior attacked *(!uniformity is not required by Bayes!)*

Bayes remained in applications

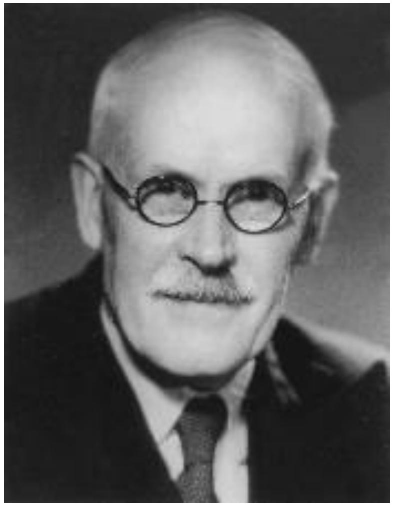
- Astronomy: objective frequency difficult to apply.
- Artillery: Joseph Louis Francois Bertrand (1822-1900)
  - How to aim cannons?
  - Uniform prior only if all causes are known to be equally probably or if nothing at all is known.
- Telecommunication, Bell Telephone Systems: Edward Molina: ”***Methods for utilizing both statistical and nonstatistical types of evidence were needed***”.
- Insurance mathematics: Isaac Rubinow: ***”every scrap of information must be used!”***, Albert Whitney: simplified Bayes formula, ’credibility theory’.

Frequentist foundation of statistics

- Karl Pearson, Ronald Fisher
- **Statistical Methods for Research Workers**. Fisher 1925.
  - ”Cook book” of statistics for non-statisticians.
  - Seven editions.
- Egon Pearson, Jerzy Neyman 1933: Neyman-Pearson theory for hypothesis testing.
  - Type I & type II errors.
- **Data was the only and sufficient source of knowledge.**
  - **Frequencies in repeatable, controllable experiments.**
  - **”Subjective priors banned”. But ok, if ’a real prior’ known (=frequency).**
  - ***No need for supplementary information.***

- **Italy**:
  - 1937, Bruno de Finetti: *”Bayes’ subjectivity on a firm mathematical foundation”*
    - Representation theorem.
    - Exchangeability → ”as if” prior. (inevitable consequence).

- **Harold Jeffreys**:
  - Almost the only Bayesian 1930-1940.
  - Geologist: earthquakes, tsunamis,…
  - Center of an earthquake?
    - Classical inverse problem → Bayes.
    - Wanted: probability of a hypothesis.
  - *”Perhaps in no other field were as many remarkable inferences drawn from so ambiguous and indirect data”.*
  - *Jeffreys’ prior: an objective prior.*



- **Jeffreys:**
  - Book: ”Theory of probability”
  - Bayes still leads to difficult calculations in practical applications.
- **Jeffreys & de Finetti:**
  - **Objective Bayes & subjective Bayes !**

WW2 and after

- **Encrypted messages**, Enigma
  - Decryption: inference under insufficient data → Bayes.
- Far too many possible combinations.
  - Impossible to try all of them.
  - Some are more probable than others.
  - Clues from different sources→ evidence builds up → probability can be updated → Bayes.
  - Example: word ”ein” was found in 90% of Enigma messages. This could still be coded in (only) 17,000 different ways.

  - Turing: measure of information: ”ban” → ”bit”
    - Birth of computer science.

Still not widely applicable

- First publication of Bayesian methods aimed for applied scientists not until 1963.
- RAND: a question for a visiting statistician: how to estimate the probability of a breaking war in the next five years?
  - ”Oh, that question just doesn’t make sense. Probability applies to a long sequence of repeatable events, and this is clearly a unique situation. The probability is either 0 or 1, but we won’t know for five years”.

    - ”I was afraid you were going to say that. I have spoken to several other statisticians and they all told me the same thing”.

Foundations

- **Savage, Lindley**: aimed for axiomatic foundation of statistics.
  - Leads to Bayesian theory ’almost accidentally’.
- Problem: if priors different, also posterior will be different. Objectivity?
- Savage: ”When they have little data, scientists disagree and are subjectivists; when they have piles of data, they agree and become objectivists”.
- Lindley agreed: ”That’s the way science is done”.

Foundations ok, but

- **Posterior probabilities still too difficult to compute.**
  - Approximation methods developed.
  - Practical examples far too artificial.
  - Lindley: ”*Bayesian statistics is not a branch of statistics, it is a way of looking at the whole of statistics*”.

→ Bayes = science of uncertainty, but how to apply it?

Breaking the wall: MCMC

- **Hierarchical Models (Lindley and Smith, *Journal of the Royal Statistical Society*, Series B, 1972) and Markov chain Monte Carlo (Gelfand and Smith, *Journal of the American Statistical Society*, 1990).**
- **1990**: MCMC and WinBUGS
  - Easier practical computation.
  - Enables bigger, more realistic models.
  - Examples from many fields of application.
  - Finally a working tool to apply Bayesian methods!

Arguments…

→"Bayes *added* a distribution for a parameter, a distribution that was not part of the binomial example under consideration and then used that distribution for probability analysis"

Fraser: Is Bayes Posterior just Quick and Dirty Confidence. Statistical Science 2011, Vol 26, no 3, 299-316.

Is this part of the problem or part of the solution?

*(Frequentists have also added **other** subjective things)*

Bayesian statistics / frequentist statistics ?

This way or that way?

Which one is of interest?

$P(X \mid \theta)$ or $P(\theta \mid X)$?

“**Thus conditioning on the data we have, rather than the data we might have had makes eminently more sense to me”.**

S.E. Fienberg. Statistical Science, 2011, Vol 26, no 2, 238-239.

But if we predict repeatedly, the predictions should be more often right than wrong. Bayesian updating should lead to better predictions , in the long run, in terms of frequency?

Bayesian methods in health technology assessment: a review

Spiegelhalter, Myles, Jones, Abrams. Health Technology Assessment 2000; Vol 4. No. 38.

- Key points
  - Claims of advantages and disadvantages of Bayesian methods are now largely based on pragmatic reasons rather than blanket ideological positions.
  - A Bayesian approach can lead to flexible modelling of evidence from diverse sources.
  - Bayesian methods are best seen as a transformation from initial to final opinion, rather than providing a single ’correct’ inference.

http://bayesian.org/

ISBA Lectures on Bayesian Foundations are now available!

The videos from the four inaugural ISBA Lectures on Bayesian Foundations presented at the 2012 ISBA World Meeting in Kyoto, Japan are now publically available. Click on any title to view the video and slides at videolectures.net exchange ideas & share knowledge

Confidence in nonparametric credible sets?
Aad van der Vaart

Bayesian dynamic modelling
Mike West

Approximate Bayesian computation (ABC): advances and questions
Christian P. Robert

Slowly but surely, Bayesian ideas revolutionize medical research
Donald A. Berry

Be sure to check out other videos in the ISBA Video and Slide Gallery!