

Methods 3, Week 2: Introduction to Bayesian Thinking

Maxime Sainte-Marie

Development



All ravens are black



Rara avis in terris nigroque simillima cyg

Ravens and swans are troublemakers

- The problem of induction (1739)
 - *There can be no demonstrative arguments to prove, that those instances, of which we have had no experience, resemble those, of which we have had experience* (T 1.3.6.5)
 - * *Custom, then, is the great guide of human life. It is that principle alone, which renders our experience useful to us, and makes us expect, for the future, a similar train of events with those which have appeared in the past.* (E 5.6)
 - Belief in causality cannot be rationally justified
 - ... God?

An improbable savior

Of miracles (1748)

- Testimonies in favor of any given miracle should be ruled irrelevant as evidence in support of any religion
 - Miracles are definition a violation of natural law
 - * Dead men never return to life
 - * Established through observation over a very long time period
 - P(deception) » P(resurrection)

An improbable savior, probabilistically saved

- Thomas Bayes, then Richard Price, started working on a probabilistic approach to inductive reasoning
 - *An Essay towards solving a Problem in the Doctrine of Chances* (1764)
 - *A Demonstration of the Second Rule in the Essay toward the Solution of a Problem in the Doctrine of Chances* (1765)

– *Four Dissertations* (1767)

- Claims

- Hume underestimated the impact of there being a number of independent witnesses to a miracle
- The multiplication of even fallible evidence can overwhelm the great improbability of an event and establish it as fact

Laplace and inverse probabilities

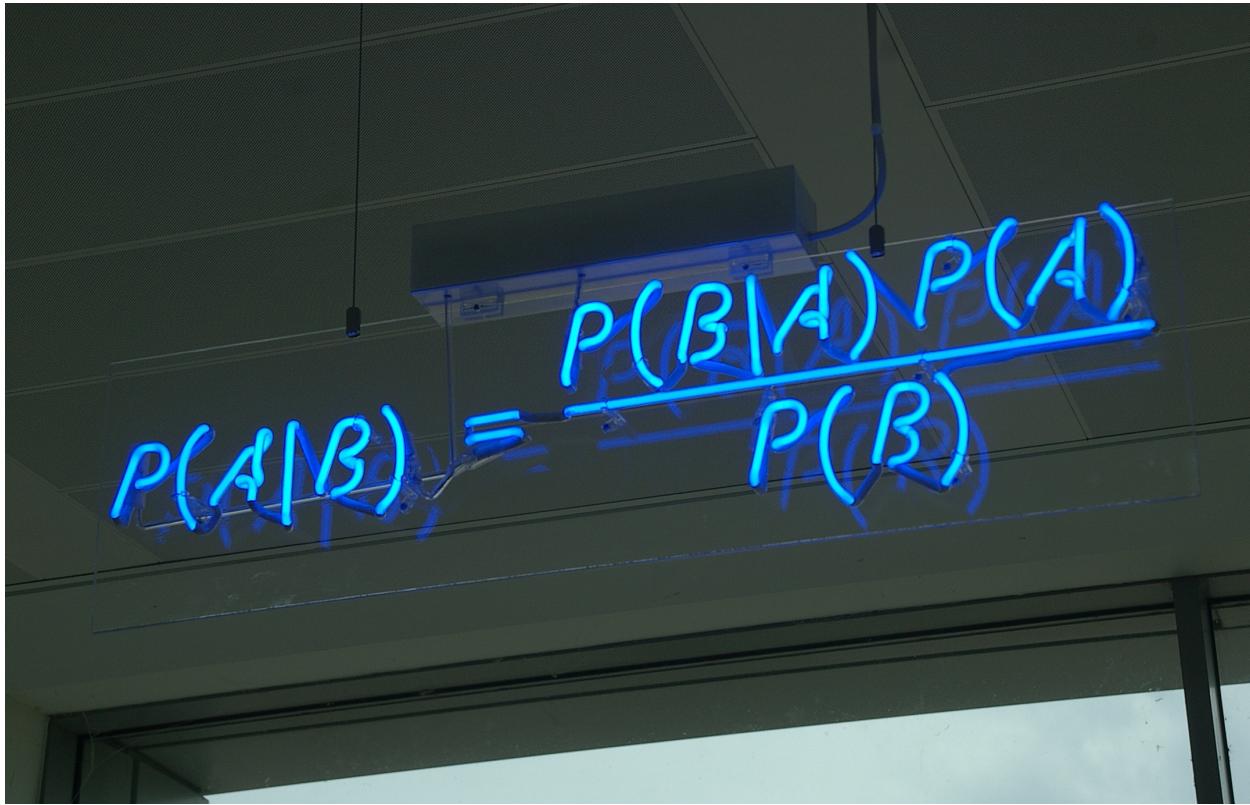
Mémoire sur la probabilité des causes par les événements (1774)

- Re-discovery & algebraic proof of the Bayes-Price Rule
 - Applications in astronomy, geodesy, meteorology, population statistics, jurisprudence...
 - * Using data on the mutual perturbations of Saturn and Jupiter, Laplace estimated the mass of Saturn at 1/3512 of the solar mass
 - Probability of 0.99991 true mass within 1%
 - Modern value: 0.63% higher
- Inverse probability problem (De Morgan, 1837)
 - Inference of an unobserved cause from observed effects

The Birth of Bayesian Analysis

- 1925: “The theory of inverse probability is founded upon an error, and must be wholly rejected” (Fischer)
 - First to use the term “Bayesian” in his critics
 - * 1939: The theory of probability (H. Jeffreys)
 - Axiomatization of Laplace’s formulations
- WWII and the post-war era marked a turning point for statistical science
 - US/UK: Bayesian turn
 - * Bletchley Park (Turing, Good, Barnard)
 - * Columbia (Statistical Research Group), Chicago (Savage), Harvard (Raiffa)
- 1950s-1960s: Creations of (frequentist) statistics departments, bayesian developments in other disciplines

Conceptualisation



Commuting from $A|B$ to $B|A$

Direct and inverse (conditional) probabilities

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

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

Procedure:

$$P(B|A)P(A) = P(B \cap A) \quad (1)$$

$$P(B|A)P(A) = P(A \cap B) \quad (2)$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (3)$$

$$P(A|B) = P(A) \frac{P(B|A)}{P(B)} \quad (4)$$

Formal Interpretation

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

posterior = *prior* × *normalized likelihood*

Update Factor = $P(B|A)/P(B)$

- Captures how more likely B being true becomes when A is
 - If A being true results in B being more likely, then...

$$P(B|A)/P(B) > 1 \quad (5)$$

$$P(A|B) > P(A) \quad (6)$$

Experimental interpretation

- The Bayes Rule describes how to determine the probability of a hypothesis, given evidence

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

where:

$$P(H|E) = \text{Probability of the hypothesis, given the evidence (Posterior)} \quad (7)$$

$$P(H) = \text{Probability of the hypothesis, before getting the evidence (Prior)} \quad (8)$$

$$P(E|H) = \text{Probability of the evidence, given the hypothesis (Likelihood)} \quad (9)$$

$$P(E) = \text{Probability of the evidence (Marginal)} \quad (10)$$

$P(A)$ = Prior

- Initial plausibility assignment for each possible value of each parameter to estimate
- Represents what was thought or known before seeing the data
 - Subjective priors
 - * Represent personal beliefs, “educated guesses”
 - Rare in the sciences
 - Objective priors
 - * Represent knowledge about a parameter, before any data is observed
- Useful for constraining parameters to reasonable ranges
 - If the prior is a bad one, the resulting inference will be misleading.

$P(B|A)/P(B)$ = Normalized likelihood

- Specifies the plausibility of the data.
- Maps each conjecture onto the relative number of ways the data could occur, given that possibility.

- Derived by enumerating all the possible data sequences that could have happened and then eliminating those sequences inconsistent with the data.
- Common to both frequentist and bayesian statistics
 - Influence of likelihood proportional to sample size
 - * Similarity of Bayesian and non-Bayesian inferences
- Normalized or standardized (with marginal)
 - Integrates posterior probability to a probability density function with total probability of one.