

A Brief History of Bayesian Inference

Mark Andrews

Bayes's Theorem

- ▶ The origin of Bayesian inference can be traced to a posthumously published 1763 essay *Towards Solving a Problem in the Doctrine of Chances*, written some years earlier by Reverend Thomas Bayes (1701-1761).
- ▶ In this essay, Bayes focused on a problem that is mathematically identical to the problem of inferring the bias of a coin after observing the outcome of a sequence of flips of that coin.
- ▶ Bayes showed that the probability that the coin's bias is exactly θ is proportional to the prior probability of θ multiplied by the probability of the observed outcomes given θ .

Laplace and Inverse Inference

- ▶ Shortly after the publication of Bayes's essay, the French polymath Pierre-Simon Laplace (1749-1827) independently rediscovered Bayes' theorem and began to apply it to practical problems of data analysis.
- ▶ Laplace was the first to present Bayesian inference in what is now its modern form:

$$\underbrace{P(\theta|\mathcal{D})}_{\text{Posterior}} = \frac{\overbrace{P(\mathcal{D}|\theta)}^{\text{likelihood}} \overbrace{P(\theta)}^{\text{prior}}}{\underbrace{\int P(\mathcal{D}|\theta)P(\theta)d\theta}_{\text{marginal likelihood}}}$$

Frequentism

- ▶ By the early 20th century, a frequentist definition of probability became increasingly widely adopted.
- ▶ Crucially, frequentism entails that the concept of probability only applies to the outcomes of random physical processes.
- ▶ As such, a parameter such as the bias of a coin, being a fixed but unknown quantity, can not be given a probabilistic interpretation.
- ▶ From this perspective, in the words of R. A. Fisher, Bayesian methods were "...founded upon an error, and must be wholly rejected." because "(i)nferences respecting populations, from which known samples have been drawn, cannot be expressed in terms of probability". Fisher (1925).

The Return of Bayesian Methods

- ▶ Bayesian methods remained marginalized from the early to the late 20th century
- ▶ However, their potential practical advantage over classical methods was that they could be applied in principle to any problem where a probabilistic model of data could be specified.
- ▶ When computer power was minimal, any “in principle” advantages of Bayesian methods were not important.
- ▶ However, the roughly exponential growth of computational power from the 1970s onwards was the eventual catalyst for the adoption of Bayesian methods.
- ▶ By the late 1980's, Bayesian methods began returning to widespread use in science.