A primer on Bayesian Hierarchical modelling

Gianluca Rossi

January 18, 2017

Let's start with some art

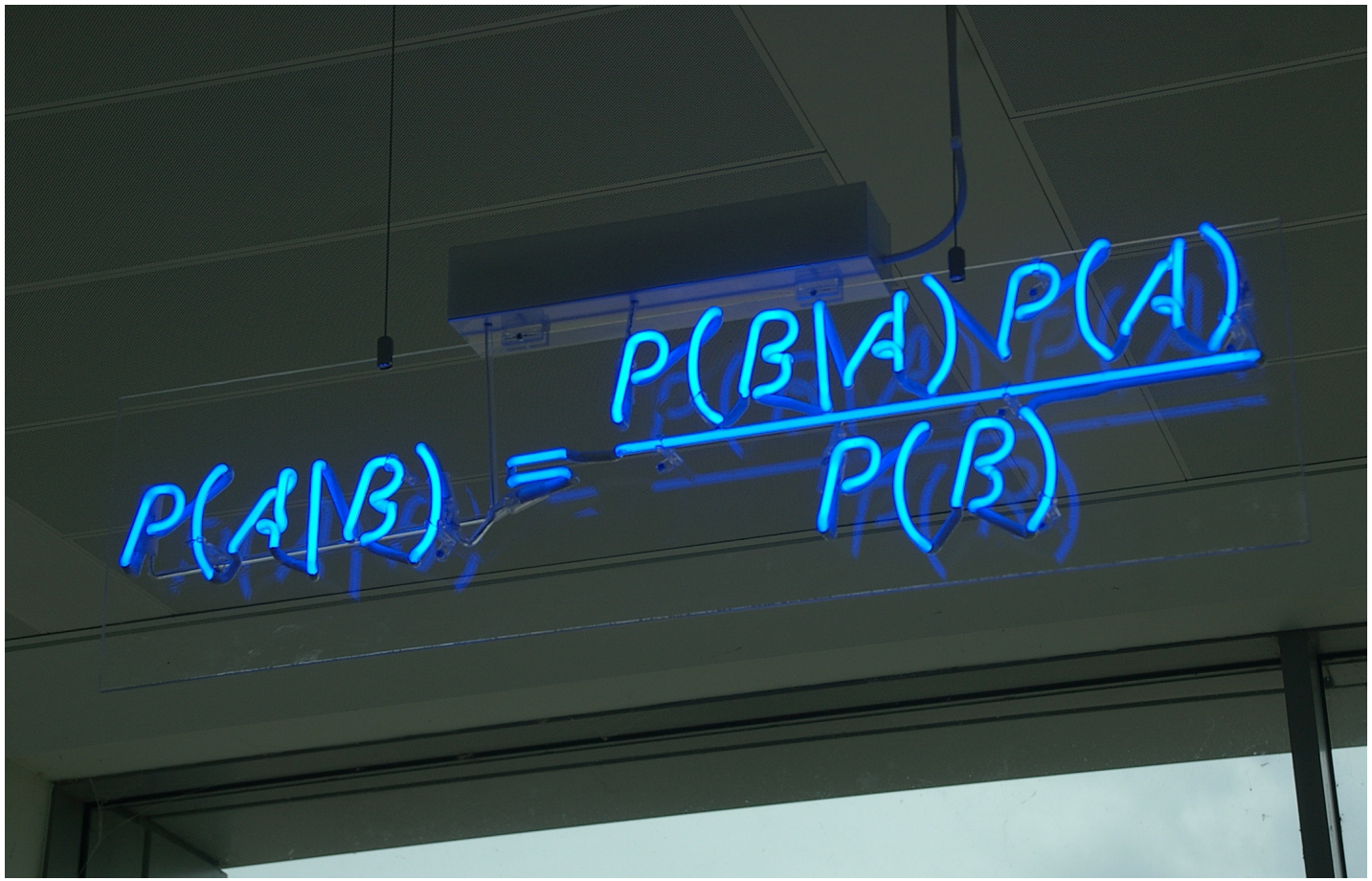

Figure: Bayes' theorem spelt out in blue neon at the offices of Autonomy in Cambridge. (credit: Mattbuck, Wikipedia)

The exact formulation

Assuming we are working in a continuous space, the Bayes' Theorem can be written as:

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} \tag{1}$$

$$= \frac{P(D|\theta)P(\theta)}{\sum P(D|\theta)P(\theta)} \tag{2}$$

$$\propto P(D|\theta)P(\theta) \tag{3}$$

The three components of the equation are:

- $P(\theta|D)$, the posterior
- $P(D|\theta)$, the likelihood function
- $P(D)$, the normalising function

The likelihood function

The likelihood is a function of the parameters of a statistical model given data. In other words, the likelihood describes the probability of observing the data given the parameters of the statistical model.

The prior distribution

The prior is a probabilistic formulation of our beliefs before collecting new data.

- Priors can be weak, moderate or strong, depending on how concentrated the probability density is compared to the problem space
- In hierarchical models, priors can be implicit, thus determined by higher level's parameters
- It's important to investigate priors when working with complex model because these could strongly impact the final results

The posterior distribution

The posterior distribution is the updated belief after collecting new data.

- Could have the same form of the prior (*conjungancy)*

The main reasons behind the success of Bayesian Inference are:

- It allows to account for prior knowledge, when this is relevant
  - Speed-up convergence
  - Attribute non-zero probability to yet un-observed events (contrary to frequentist approach)
- Moving beyond single point estimation
- Estimating probability in a joint parameter space

Empty frame

Empty frame

Empty frame

Hierarchical modelling means expressing the Bayes' Theorem as depencencies between parameters instead of an expression about the joint parameter space.

$$P(\theta, \omega|D) = \frac{P(D|\theta, \omega)P(\theta, \omega)}{P(D)} \tag{4}$$

$$= \frac{P(D|\theta)P(\theta, \omega)}{P(D)} \tag{5}$$

$$= \frac{P(D|\theta)P(\theta|\omega)P(\omega)}{P(D)} \tag{6}$$

We made the following steps:

- (5) because likelihood doesn't depend on ω
- (6) because $P(\theta|\omega) = \frac{P(\theta,\omega)}{P(\omega)}$ thus $P(\theta, \omega) = P(\theta|\omega)P(\omega)$

When dealing with a continuous problem space (6) becomes:

$$P(\theta_1, \dots m|D) = \frac{P(D|\theta_1, \dots m)P(\theta_1, \dots m)}{P(D)} \tag{7}$$

$$= \frac{P(D|\theta_1, \dots m)P(\theta_1, \dots m)}{\sum_m \int \mathrm{d}\theta_m P(D|\theta_1, \dots m)P(\theta_1, \dots m)} \tag{8}$$

$$= \frac{\prod P_m(D|\theta_m, m)P_m(\theta_m|m)P(m)}{\sum_m \int \mathrm{d}\theta_m \prod_m P_m(D|\theta_m, m)P_m(\theta_m|m)P(m)} \tag{9}$$

Empty frame

Empty frame

Empty frame

Empty frame

Empty frame