1 Baye's rule

We have seen in lecture 1 that the conditional probability of A given B was given by:

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

given that $A \cap B = B \cap A$, it follows that:

$$\mathrm{P}(A|B)\mathrm{P}(B) = \mathrm{P}(B|A)\mathrm{P}(A)$$

or:

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

1.1 Bayes in the context of modelling

Data modelling involves choosing a model and fitting its parameters so as to best describe the data given. In this context, we denote Θ the parameters of a model H, to be identified given the data D.

Example: If we perform a linear regression on a set of data $\{x_m, y_m\}$, H is our model, here a function such as $y = ax + b$, Θ consists of the parameters a and b, and D is our set $\{x_m, y_m\}$.

In this context, Baye's rule is often written:

$$\mathrm{P}(\Theta|D,H) = \frac{\mathrm{P}(D|\Theta,H)\mathrm{P}(\Theta|H)}{\mathrm{P}(D|H)} \tag{1}$$

- $\mathrm{P}(\Theta|H)$ is called the *prior*, and represents our idea of what the parameters should be (or rather with which probability they take a given value) before we have seen the data.
- $\mathrm{P}(D|\Theta,H)$ is called the *likelihood*, which is the probability of our dataset given some values of the parameters. We will see later that we need a model to evaluate this quantity.
- $\mathrm{P}(D|H) = \int_\Theta P(D|\Theta,H)\mathrm{P}(\Theta|H)\mathrm{d}\Theta$ is called the *evidence*.

The use of Baye's rule is better illustrated through an example.

2 An example of inference

A good example to illustrate Bayes rule is that of someone drawing balls from 11 urns each containing 10 balls. Of the 10 balls in each urn, u are black, and the rest are white.

Given that the person draws N times (replacing the balls on each occasion) from a chosen urn, the problem is to guess from which urn u the person has drawn.

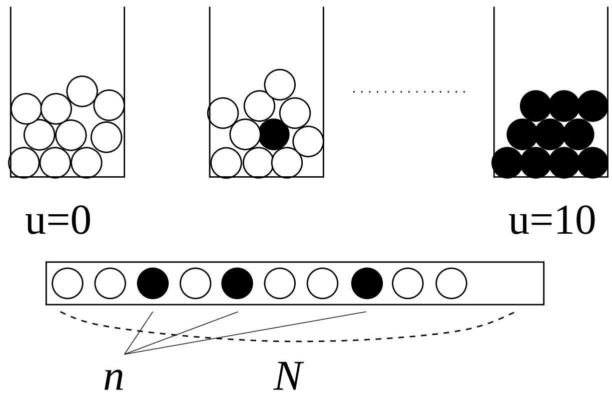

Figure 1: A person is drawing N times from a same urn u chosen among eleven and presents the result. The problem is to guess which urn the person has chosen and to use this to predict the next draw.

In this case, Bayes rules is written:

$$P\left(u|n,N\right)=\frac{P\left(n|u,N\right)P\left(u|N\right)}{P\left(n|N\right)} \tag{2}$$

that is, given that the person has drawn n black balls among N, the probability that urn u was used is the probability of drawing n black balls from u times the probability to chose u, over the overall probability of drawing n balls. This last term is also:

$$P\left(n|N\right)=\sum_u P\left(n|u,N\right)P\left(u|N\right) \tag{3}$$

To solve the problem, we can first assume (this is our hypothesis H) that the person will chose any of the urns with the same probability, so that $P\left(u|N\right)$, which is actually $P\left(u\right)$ since it does not depend on the number of draws, is 1/11; this is our *prior.* For each possible value of u, we can then compute the probability of drawing n black balls, which gives a binomial distribution, with $p=u/10$ (see Appendix for details):

$$\mathrm{P}(n|u,N)=\binom{N}{n}p^n(1-p)^{N-n} \tag{4}$$

If we are simply interested in knowing which urn the person has most probably drawn from, that is to compare the different models, there is no need to calculate the normalising constant $P(n|N)$. The result is a probability distribution on u, which, not surprisingly, is maximum for $n = u$. However, almost all other possibilities have a non zero probability, as can be seen in figure 2, apart from the two urns containing either only white or only black balls.

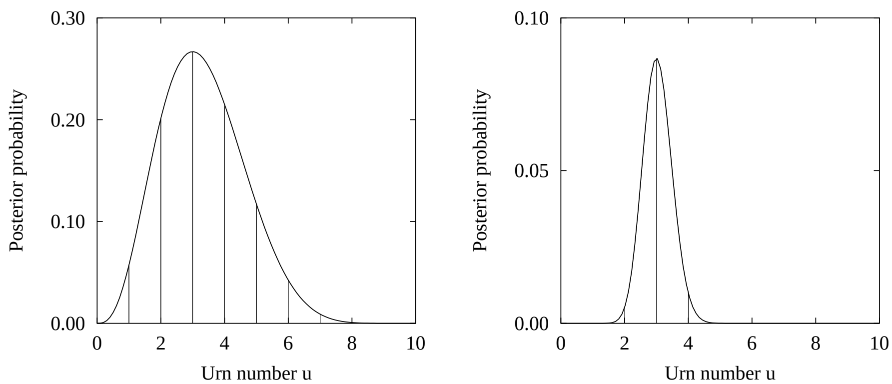

Figure 2: The posterior probability for the chosen urn u, when $n = 3$ ($N = 10$). The person has most probably drawn from the urn containing 3 black balls, but the probability that he/she has drawn from other urns is not negligible. The right figure shows that when more data are presented, ($n = 30$, $N = 100$) the probability of $u = 3$ is much greater than $u = 2$ and $u = 4$ and that any other value of u is virtually impossible

2.1 Marginalising

An interesting problem is to now use these results to predict how likely the person is to obtain a black ball at the next draw. For this, we consider the different possible values of u with their respective probability; for each of them, the probability of obtaining a black ball is $u/10$, so the overall probability is

$$\sum_{u=1}^{u=10} P(u|n,N)\, u/10 \tag{5}$$

This process is called *marginalising* over all the possible values of u.

A traditional statistic approach would have used the most probable hypothesis ($u = 3$) to make this prediction, and would have led to a value of 1/3. In the Bayesian approach, all the possibilities are taken into account, but weighted according to how probable they are.

2.2 Quantifying the uncertainty

One of the most interesting features of the Bayesian approach to inference problems, is that it provides a quantitative estimation of the uncertainty in fitting a model to the data. Figure 2 shows the posterior distributions for $P(u)$ when $n = 3,\ N = 10$ and $n = 30,\ N = 100$. In the first case, the distribution is spread, while it is sharply peaked in the second one. Consequently, the predictions made on the basis of the second model ($n = 30,\ N = 100$) is associated with a much smaller uncertainty than those made using the first set of data. We will go into further details when studying the linear regression in Bayesian framework.

Appendix

To obtain the probability of of drawing n black balls in N, first we estimate the probability of a particular sequence:

- The probability to draw a black ball is $u/10$, and that to draw a white one is $1-(u/10)$, further denoted p and $1-p$ respectively. A particular sequence of n black and $N-n$ white balls has therefore a probability:

$$p^n(1-p)^{N-n}$$

- Count the number of ways to draw n black balls out of N: imagine a series of N boxes which needs to be filled with our results, for the first black ball, there are N possibilities, for the second, $N-1$, etc. For the n^{th} black ball, we only have $N-(n-1)$ choices. This is a total of:

$$N \times (N-1) \times \ldots \times (N-(n-1)) = \frac{N!}{(N-n)!}$$

  However, this assumes that the case where, for example, the first black ball is placed in position i and the second in j is distinct from the one where the first is placed in j and the second in i. The above number, given that the black balls are not distinct, is therefore to be divided by $n!$, the number of identical sequences.

- This leads to:

$$\frac{N!}{n!(N-n)!} = \binom{N}{n}$$