A brief introduction to (simulation based) Bayesian inference

The basic idea of Bayesian inference is to setup a full probability model for both observed and unobserved quantities. Inference is then based on the so-called posterior density — that is the conditional density of the unobserved quantity conditional on the observed quantity.

Let y denote the observed quantity (the data) which we assume is a realisation of a random variable Y. Assume futher that the dsitribution of Y depends on an unobserved quantity θ which we assume is a realisation of another random variable Θ. More precisely we assume that Θ is distributed according to the so-called *prior density* $\pi(\theta)$. Given $\Theta = \theta$ we assume that Y is distributed according to the so-called *sanpling/data density* $\pi(y|\theta)$ — sometimes also referred to as the *likelihood*. By the definition of conditional densities, these assumption imply that the joint distribution of Y and Θ has density

$$\pi(y, \theta) = \pi(\theta)\pi(y|\theta).$$

The prior density should reflect our prior knowledge (or our prior uncertainty) regarding Θ, i.e. our knowledge about Θ *before* we observe Y. The data density should be chosen so that it is consistent with our knowledge about the problem of interest.

From the definition of conditional densities we obtain the posterior density of Θ:

$$\pi(\theta|y) = \frac{\pi(y, \theta)}{\pi(y)} = \frac{\pi(\theta)\pi(y|\theta)}{\pi(y)}. \tag{1}$$

Notice that given the data $Y = y$ the term $\pi(y)$ is a constant and hence

$$\pi(\theta|y) \propto \pi(\theta)\pi(y|\theta)$$

is an unnormalised posterior density. The posterior density can be interpreted as our updated knowledge about Θ after having observed Y. Inference is typically based on reproducing all or parts of the posterior density graphically (as graphs or contour plots). Another option is to report e.g. posterior mean, mode, and quantiles. Notice that a *central 95% postrior interval* (e.g. the interval between the 2.5% and 97.5% quantiles) can directly be interpreted as containing θ with high probability unlike the classical confidence intervals. It is however not always trivial to obtain the posterior density — or even an approximation of it.

Classical Bayesian inference has been limited by the fact that to make a posterior analysis feasible the prior should be chosen so that the resulting posterior density can be recognised as the density of a known distribution. Such prior distribution are called conjugated priors.

This limitation has been drastically reduced in last 15 years by a combination of Markov chain Monte Carlo (MCMC) methods and an increase in available computing power.

For a more detailed introduction to Bayesian inference, including many example of applications, see

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2004). *Bayesian Data Analysis*, 2nd ed. Chapman & Hall/CRC.

Example: Placenta Previa data (cont.)

Recall Exercise 5 in the text "Basic methods for simulation of random variables: 1. Inversion" regarding estimating the probability for a female birth given a special condition called placenta previa. The number of female births is the observed quantity and the probability of a female birth is the unobserved quantity θ which we assume is the realisation of a random variable Θ. This leads us to assume that the number of observed female births given $\Theta = \theta$ is binomially distributed with parameter θ, so Y has density

$$\pi(y|\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y},$$

where we assume that the total number of births $n = 980$ is known. In Exercise 5 the prior distribution has a non-standard density (a witch hat). Here we assume instead that the prior density is beta with parameters α and β:

$$\pi(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}.$$

This is an example of a conjugated prior, since the resulting posterior density:

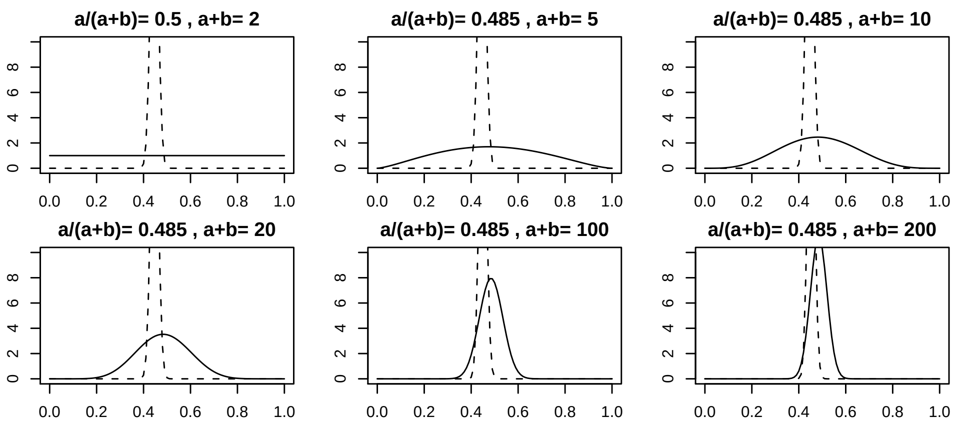
$$\pi(\theta|y=437) \propto \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\theta^{437}(1-\theta)^{543} \propto \theta^{437+\alpha-1}(1-\theta)^{543+\beta-1}$$

can be recognised as the unnormalised density of a beta distributed random variable Θ with parameters $437+\alpha$ and $543+\beta$.

In Bayesian statistics it is good practice to perform a so-called *sensitivity analysis* to assess how sensitive the posterior distribution is to the choice of prior. Table 1 contains the 2.5%, 50% and 97.5% quantiles for the posterior distribution for a range of α and β values reparameterised as $\alpha/(\alpha+\beta)$ (the prior mean) and $\alpha+\beta$. Further, Figure 1 shows the prior and posterior densities of Θ for the same values of α and β. Table 1 shows that except for the last row the prior has little influence on the posterior distribution. Note that the prior and posterior densities are quite different, again except the last case, and even here the 95% posterior interval does not contain the prior mean.

		Quantiles		
$\frac{\alpha}{\alpha+\beta}$	$\alpha+\beta$	2.5%	50%	97.5%
0.5	2	0.415	0.446	0.477
0.485	5	0.415	0.446	0.477
0.485	10	0.415	0.446	0.477
0.485	20	0.416	0.447	0.478
0.485	100	0.420	0.450	0.479
0.485	200	0.424	0.453	0.481

Tabel 1: Prior parameters and corresponding posterior quantiles.



Figur 1: Prior (solid line) and posterior (dashed line) densities.