CHAPTER 5

THE BABY MODEL

5.1. The model

We will start this chapter with the simplest possible model, and we will see more complicated models in chapter 6. Our model consists only in a mean plus an error term

$y_i = \mu + e_i$

along the book we will consider that the data are normally distributed, although all procedures and conclusions can be applied to other distributions. All errors have mean zero and are uncorrelated, and all data have the same mean. Thus, to describe our model we will say that

$y_i \sim N(\mu, \sigma^2)$

$\mathbf{y} \sim N(\mathbf{1}\mu, \mathbf{I}\sigma^2)$

where **1'** = [1, 1, 1, ... , 1] and **I** is the identity matrix.

$$f\left(y_i \mid \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\left(y_i - \mu\right)^2}{2\sigma^2}\right]$$

$$f\left(\mathbf{y} \mid \mu, \sigma^2\right) = f\left(y_1, y_2, \ldots, y_n \mid \mu, \sigma^2\right) = f\left(y_1 \mid \mu, \sigma^2\right) \cdot f\left(y_2 \mid \mu, \sigma^2\right) \cdots f\left(y_n \mid \mu, \sigma^2\right) =$$

$$= \prod_1^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\left(y_i - \mu\right)^2}{2\sigma^2}\right] = \frac{1}{\left(2\pi\right)^{\frac{n}{2}} \left(\sigma^2\right)^{\frac{n}{2}}} \exp\left[\sum_1^n -\frac{\left(y_i - \mu\right)^2}{2\sigma^2}\right] \quad (5.1)$$

as we saw in 3.5.3. Now we have to establish our objectives. What we want is to estimate the unknowns ‘μ’ and ‘σ^2’ that define the distribution.

5.2. Analytical solutions

5.2.1. *Marginal posterior density of the mean*

We will try to find the *marginal* posterior distributions for each unknown because this distribution takes into account the uncertainty when estimating the other parameter, as we have seen in chapters 2 and 3. Thus we should find

$$f\left(\mu \mid \mathbf{y}\right) = \int_0^\infty f\left(\mu, \sigma^2 \mid \mathbf{y}\right) d\sigma^2$$

$$f\left(\sigma^2 \mid \mathbf{y}\right) = \int_{-\infty}^\infty f\left(\mu, \sigma^2 \mid \mathbf{y}\right) d\mu$$

We have derived these distributions by calculating the integrals in chapter 3.

$$f(\mu \mid \mathbf{y}) \propto t_{n-1}\left(\bar{y}, s^2\right)$$

This is a “Student” t-distribution with parameters $\bar{y}$ and s^2, and n-1 degrees of freedom, where

$$\bar{y} = \frac{1}{n}\sum_1^n y_i \quad ; \quad s^2 = \frac{1}{n-1}\sum_1^n \left(y_i - \bar{y}\right)^2$$

The other marginal density we look for is (see Chapter 3)

$$f\left(\sigma^2 \mid \mathbf{y}\right) \propto \frac{1}{\left(\sigma^2\right)^{\frac{n-1}{2}}} \exp\left[-\frac{\sum_1^n \left(y_i - \bar{y}\right)^2}{2\,\sigma^2}\right] \propto \mathbf{IG}(\alpha, \beta)$$

which is an Inverted Gamma distribution with parameters α, β

$$\alpha = \frac{n-1}{2} - 1 \quad ; \quad \beta = \frac{1}{2}\sum_1^n \left(y_i - \bar{y}\right)^2$$

5.2.2. *Joint Posterior density of the mean and variance*

We have seen that, using flat priors for the mean and variance,

$$f\left(\mu, \sigma^2 \mid \mathbf{y}\right) = \frac{f\left(\mathbf{y} \mid \mu, \sigma^2\right) f\left(\mu, \sigma^2\right)}{f(\mathbf{y})} \propto f\left(\mathbf{y} \mid \mu, \sigma^2\right) f\left(\mu, \sigma^2\right) \propto f\left(\mathbf{y} \mid \mu, \sigma^2\right)$$

$$f\left(\mu, \sigma^2 \mid \mathbf{y}\right) \propto \frac{1}{(2\pi)^{\frac{n}{2}}\left(\sigma^2\right)^{\frac{n}{2}}} \exp\left[-\frac{\sum_1^n \left(y_i - \mu\right)^2}{2\sigma^2}\right]$$

now both parameters are in red because this is a bivariate distribution.

5.2.3. *Inferences*

We can draw inferences from the joint or from the marginal posterior distributions. For example, if we find the maximum from the joint posterior distribution, this is the most probable value for both parameters μ and σ^2 *simultaneously*, which is not the most probable value for the mean and the variance when all possible values of the other parameter have been weighted by their probability and summed up (i.e.: the mode of the marginal posterior densities of μ and σ^2). We will show now some inferences that have related estimators in the frequentist world.

Mode of the joint posterior density

To find the mode, as it is the maximum value of the posterior distribution, we derive and equal to zero (Appendix 5.2)

$$\text{mode } f\left(\mu,\sigma^2 \mid \mathbf{y}\right) \longrightarrow \begin{cases} \dfrac{\partial}{\partial\mu} f\left(\mu,\sigma^2 \mid \mathbf{y}\right) = 0 \longrightarrow \hat{\mu} = \dfrac{1}{n}\sum y_i = \bar{y} \quad \text{corresponding to } \hat{\mu}_{ML} \\ \dfrac{\partial}{\partial\sigma^2} f\left(\mu,\sigma^2 \mid \mathbf{y}\right) = 0 \longrightarrow \hat{\sigma}^2 = \dfrac{1}{n}\sum\left(y_i - \bar{y}\right)^2 \text{ corresponding to } \hat{\sigma}^2_{ML} \end{cases}$$

thus the mode of the joint posterior density give formulas that *look like* the maximum likelihood (ML) estimates of the variances, although here the interpretation is different because here they mean that this estimate is the most probable value of the unknowns μ and σ^2, whereas in a frequentist context this means that these values would make the sample most probable if they were the true values. The numeric value of the estimate is the same, but the interpretation is different.

Notice that we will not usually make inferences from joint posterior distributions because when estimating one of the parameters we do no take into account the uncertainty of estimating the other parameter.

Inferences from the marginal posterior density of the mean

As the marginal posterior distribution of the mean is a $t_{n-1}\left(\bar{y},s^2\right)$, the mean, median and mode are the same and they are equal to the sample mean. For credibility intervals we can consult a table of the t_{n-1} distribution.

Mode of the marginal posterior density of the variance

Deriving the marginal posterior density and equating to zero, we obtain (Appendix 5.3)

$$\text{mode } f\left(\sigma^2 \mid \mathbf{y}\right) \longrightarrow \frac{\partial}{\partial\sigma^2} f\left(\sigma^2 \mid \mathbf{y}\right) = 0 \longrightarrow \hat{\sigma}^2 = \frac{1}{n-1}\sum\left(y_i - \bar{y}\right)^2 \text{ corresponding to } \hat{\sigma}^2_{REML}$$

thus the mode of the joint posterior density give formulas that look like the maximum residual likelihood (REML) estimates of the variances, although here the interpretation is different because here this estimate is the most probable value of the unknown σ^2 when the values of the other unknown μ has been considered, weighted by its probability and integrated out (summed up), whereas in a frequentist context we mean that this value would make the sample most probable if this would be the true value when working in a subspace in which there is no μ (see Blasco, 2001 for a more detailed interpretation). The numeric value of the estimate is the same, but the interpretation is different. Here the use of this estimate is more founded that in the frequentist case, but notice that the frequentist properties are different from the Bayesian ones, thus a good Bayesian estimator is not necessarily a good frequentist estimator, and vice versa.

5.3. Working with MCMC

5.3.1. *Using flat priors*

To work with MCMC-Gibbs sampling we need o calculate the *conditional* distributions of the parameters $f\left(\mu \mid \mathbf{y}, \sigma^2\right)$ and $f\left(\sigma^2 \mid \mathbf{y}, \mu\right)$. We do not know them, but we can calculate them using Bayes theorem. Using flat priors

$$f\left(\mu \mid \mathbf{y}, \sigma^2\right) \propto f\left(\mathbf{y} \mid \mu, \sigma^2\right) f\left(\mu\right) \propto f\left(\mathbf{y} \mid \mu, \sigma^2\right)$$

$$f\left(\sigma^2 \mid \mathbf{y}, \mu\right) \propto f\left(\mathbf{y} \mid \sigma^2, \mu\right) f\left(\sigma^2\right) \propto f\left(\mathbf{y} \mid \sigma^2, \mu\right)$$

as we know the distribution of the data, we can obtain both conditionals

$$f\left(\mu \mid \mathbf{y}, \sigma^2\right) \propto f\left(\mathbf{y} \mid \mu, \sigma^2\right) \propto \frac{1}{\left(\sigma^2\right)^{\frac{n}{2}}} \exp\left[-\frac{\sum_1^n \left(y_i - \mu\right)^2}{2\sigma^2}\right]$$

$$f\left(\sigma^2 \mid \mathbf{y}, \mu\right) \propto f\left(\mathbf{y} \mid \sigma^2, \mu\right) \propto \frac{1}{\left(\sqrt{2\pi}\right)^n \left(\sigma^2\right)^{\frac{n}{2}}} \exp\left[-\frac{\sum_1^n \left(y_i - \mu\right)^2}{2\sigma^2}\right]$$

Notice that the formulae are the same, but the variable is in red, thus the functions are completely different. As we saw in chapter 3, the first is a normal distribution and the second an inverted gamma distribution.

$$f\left(\mu \mid \mathbf{y}, \sigma^2\right) \quad \propto \quad N\left(\bar{y}, \frac{\sigma^2}{n}\right)$$

$$f\left(\sigma^2 \mid \mathbf{y}, \mu\right) \quad \propto \quad \mathbf{IG}\left(\alpha, \beta\right) \quad ; \qquad \beta = \frac{1}{2}\sum_1^n \left(y_i - \mu\right)^2 \quad ; \quad \alpha = \frac{n}{2} - 1$$

We have algorithms to sample from normal and inverted gamma functions, thus we can take random samples of them. We start, for example, with an arbitrary value for the variance and then we get a sample value of the mean. We substitute this value in the conditional of the mean and we get a random value of the variance. We substitute it in the conditional distribution of the mean and we continue the process (figure 5.1)

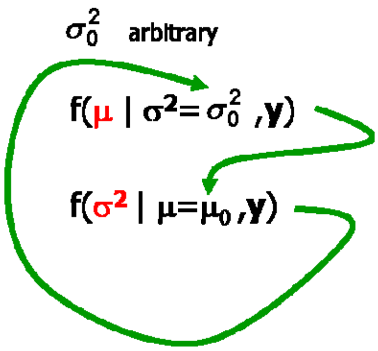

Figure 5.1. Gibbs sampling process for the mean and the variance of the baby model

We will put an example. We have a data vector with four samples

y' = [2, 4, 4, 2]

then we calculate

$$\bar{y} = 3, \quad n = 4, \quad \sum_{1}^{n} y_i^2 = 40$$

and we can prepare the first conditional distributions

$$f\left(\mu \mid \sigma^2, \mathbf{y}\right) \propto N\left(\bar{y}, \frac{\sigma^2}{n}\right) \equiv N\left(3, \frac{\sigma^2}{4}\right)$$

$$f\left(\sigma^2 \mid \mu, \mathbf{y}\right) \propto \text{Igamma}\begin{cases} \beta = \dfrac{1}{2}\left[\sum_{1}^{n} y_i^2 - \mu^2\right] \\ \alpha = \dfrac{n}{2} - 1 \end{cases} \equiv \text{Igamma}\begin{cases} \beta = \dfrac{1}{2}\left(40 - \mu^2\right) \\ \alpha = 1 \end{cases}$$

Now we start the Gibbs sampling process by taking an arbitrary value for σ^2, for example

$$\sigma_0^2 = 1$$

then we substitute this arbitrary value in the first conditional distribution and we have

$$f\left(\mu \mid \sigma^2, \mathbf{y}\right) \propto N\left(3, \frac{1}{4}\right)$$

we sample from this distribution using an appropriate algorithm and we find

$\mu_0 = 4$

then we substitute this sampled value in the second conditional distribution,

$$f\left(\sigma^2 \mid \mu, \mathbf{y}\right) \propto \text{Igamma}\left[12,\ 1\right]$$

now we sample from this distribution using an appropriate algorithm and we find

$$\sigma_1^2 = 5$$

then we substitute this sampled value in the first conditional distribution,

$$f\left(\mu \mid \sigma^2, \mathbf{y}\right) \propto N\left(3, \frac{5}{4}\right)$$

now we sample from this distribution using an appropriate algorithm and we find

$\mu_1 = 3$

then we substitute this sampled value in the second conditional distribution, and continue the process. Notice that we sample each time from a different conditional distribution. The first conditional distribution of μ was a normal with mean equal to 3 and variance equal to 1, but the second time we sampled it was a normal with the same mean but with variance equal to 5. The same happened with the different Inverted Gamma distributions from which we were sampling.

We obtain two chains. All the samples belong to different conditional distributions, but after a while they are also samples from the respective marginal posterior distributions (figure 5.2). After rejecting the samples of the "burning period", we can use the rest of the samples for inferences as we did in chapter 4 with the MCMC chains.

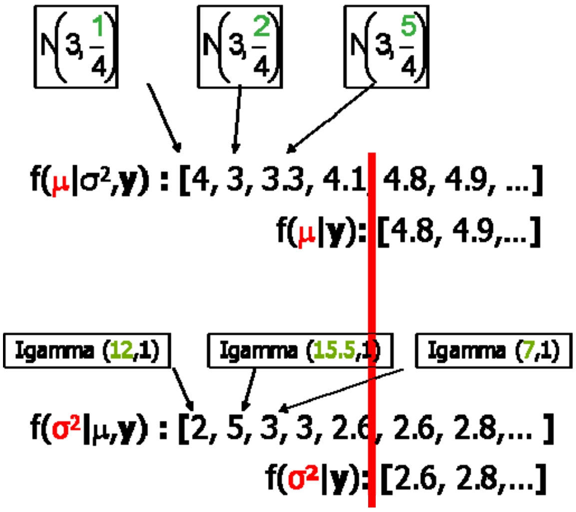

Figure 5.2. A Gibbs sampling process. Samples are obtained from different conditional distributions, but after a "burning" in which samples are rejected, the rest of them are also samples of the marginal posterior distributions.

5.2.3. Common misinterpretations

The parameters of the inverted gamma distribution are degrees of freedom: The concept of degrees of freedom was developed by Fisher (1922), who represented the sample in a space of n-dimensions (see Blasco 2001 for an intuitive representation of degrees of freedom). This has no relationship with what we want. We manipulate the parameters in order to change the shape of the function, and it is irrelevant whether these "hyper parameters" are natural numbers or fractions. For example, Blasco et al. (1998) use fractions for these parameters.

One of the parameters of the inverted gamma distribution represents credibility and the other represents the variance of the function: Both parameters modify the shape of the function in both senses: dispersion and sharpness showing more credibility, thus it is incorrect to name one of them as the parameter of credibility and to use standard deviations or variances coming from other experiments as a parameter of the inverted gamma distribution. Both parameters should be manipulated in order to obtain a shape that will show our beliefs, and it is irrelevant which values they have as far as the shape of the function represents something similar to our state of beliefs.