Chapter 3

Multiparameter Models

Virtually every practical problem in statistics involves more than one unknown or unobservable quantity. It is in dealing with such problems that the simple conceptual framework of the Bayesian approach reveals its principal advantages over other methods of inference. Although a parameter can include several parameters of interest, conclusions will often be drawn about one, or only a few, parameters at a time. In this case, the ultimate aim of a Bayesian analysis is to obtain the marginal posterior distribution of the particular parameters of the interest. In principle, the route to achieving this aim is clear: we first require the joint posterior distribution of all unknowns, and then we integrate this distribution over the unknowns that are not immediate interest to obtain the desired marginal distribution. Or equivalently, using simulation, we draw samples from the joint posterior distribution and then look at the parameters of interest and ignore the values of the other unknowns. In many problem there is no interest in making inferences about many of the unknown parameters, although they are required in order to construct a realistic model. Parameters of this kind are often called nuisance parameters. (Hoff, PD, 2009)

3.1 The normal model

A random variable $Y$ is said to be normally distributed with mean $\theta$ and variance $\sigma^2$, if the density of $Y$ is given by,

$$p(y|\theta, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{1}{2} \frac{(y - \theta)^2}{\sigma^2}\right)$$
3.1 The normal model

Properties of normal model

1. The distribution is symmetric about $\theta$, and the mode, median and mean are all equal to $\theta$.

2. About 95% of the population lies within two standard deviations of the mean.

3. If $X \sim \text{normal}(\mu, \tau^2)$ and $Y \sim \text{normal}(\theta, \sigma^2)$ and $X$ and $Y$ are independent, $aX+bY \sim \text{N}(a\mu+b\theta, a^2 \tau^2+b^2 \sigma^2)$

Features of normal distribution

Suppose our sampling model is,

$$(Y_1, \ldots, Y_n | \theta, \sigma^2) \sim \text{i.i.d normal}(\theta, \sigma^2)$$

prior,

$$\theta \sim \text{normal}(\mu_0, \sigma_0^2)$$

$$p(\theta) \propto \exp\left(\frac{-(\theta - \mu_0)^2}{2\sigma_0^2}\right)$$

then the posterior distribution of $\theta$, $\theta | y_1, \ldots, y_n \sim \text{normal}(\mu_n, \tau_n^2)$ where,

$$\tau_n^2 = \frac{1}{1/\sigma_0^2 + n/\sigma^2}$$

and

$$\mu_n = \frac{\mu_0/\sigma_0^2 + n\bar{y}/\sigma^2}{1/\sigma_0^2 + n/\sigma^2}$$

It is evident from above expression theta the posterior parameters $\tau_n^2$ and $\mu_n$ combine the prior parameters $\tau_0^2$ and $\mu_0$ with terms from the data. Moreover the formula for $\frac{1}{\tau^2}$ is,

$$\frac{1}{\tau_n^2} = \frac{1}{\tau_0^2} + \frac{n}{\sigma^2}$$

Inverse variance is referred to as the precision. For the normal model let,

$$\hat{\sigma}^2 = \frac{1}{\tau_0^2} = \text{sampling precision}$$

$$\hat{\tau}_0^2 = \frac{1}{\tau_0^2} = \text{prior precision}$$

$$\hat{\tau}_n^2 = \frac{1}{\tau_n^2} = \text{posterior precision}$$

therefore, $\hat{\tau}_n^2 = \hat{\tau}_0^2 + n\sigma^2$ and so the posterior information = prior information + data information.
3.2 Example: midge wing length

Posterior mean

\[ \mu_n = \frac{\tau_0^2 \mu_0}{\tau_0^2 + n \sigma^2} + \frac{n \sigma^2 \bar{y}}{\tau_0^2 + n \sigma^2} \]

and so the posterior mean is the weighted average of the prior mean and the sample mean.

Prediction

If \( Y \mid \theta, \sigma^2 \sim N(\theta, \sigma^2) \)

\( \Rightarrow Y = \theta + \epsilon, (\epsilon \mid \sigma^2) \sim N(0, \sigma^2) \)

then, the predictive distribution is,

\( \bar{Y} \mid \sigma^2, y_1, \ldots, y_n \sim N(\mu_n, \tau_n^2 + \sigma^2) \)

3.2 Example: midge wing length

Grogan and Wirth (1981) provide the data on the wing length in millimeters of nine members of a species of midge (small, two winged flies). From these nine measurements we wish to make inference on the population mean \( \theta \). Studies from other population suggest that wings length are typically around 1.9mm, so we set \( \mu_0 = 1.9 \)

Since for any normal distribution most of the probability is within two standard deviation of the mean,

\[ \mu_0 - 2\tau_0 > 0 \]

\[ \tau_0 < \frac{1.9}{2} = 0.95 \]

Data:

\( y = (1.64, 1.70, 1.72, 1.74, 1.82, 1.82, 1.82, 1.90, 2.08) \)

\( \bar{y} = 1.804 \)

we have \( \theta \mid y_1, \ldots, y_n, \sigma^2 \sim N(\mu_n, \tau_n^2) \)

where,

\[ \mu_n = \frac{\mu_0}{\tau_0} + \frac{n}{\sigma^2} = \frac{1.9 \times 1.804}{1.804 + 9/\sigma^2} = 1.805 \]

Hence,

\( \theta \mid y \sim N(1.805, 0.0447) \)

The posterior mean and posterior standard deviation are 1.805 and 0.0447. Therefore, 95% credible region is,

\[ 1.805 \pm 1.96 \times 0.0447 \]

\[ 1.774648, 1.893198 \]

[1] 1.774648 1.893198
3.3 Monte Carlo sampling

For many data analysis, interest primarily lies in estimating the population mean $\theta$, and so we would like to calculate quantities like $E(\theta|y_1, \ldots, y_n)$, $sd(\theta|y_1, \ldots, y_n)$, $P(\theta_1 < \theta_2|y_1, \ldots, y_n)$ and so on. These quantities are all determined by the marginal posterior distribution of $\theta$ given the data.

Consider simulating parameter values using the following Monte Carlo procedure.

$$
\sigma^{2(1)} \sim \text{inverse-gamma}(\frac{\nu_n}{2}, \frac{\nu_n \sigma_n^2}{2}) \\
\vdots \\
\sigma^{2(s)} \sim \text{inverse-gamma}(\frac{\nu_n}{2}, \frac{\nu_n \sigma_n^2}{2}) \\
\theta^{(1)} \sim \text{normal}(\mu_n, \frac{\sigma^{2(1)}}{k_n}) \\
\vdots \\
\theta^{(s)} \sim \text{normal}(\mu_n, \frac{\sigma^{2(s)}}{k_n})
$$

Note that each $\theta^{(s)}$ is sampled from its conditional distribution given the data and $\sigma^2 = \sigma^{2(s)}$.
Figure 3.2: Monte Carlo samples from and estimates of the joint and marginal distributions of the population mean and variance. The vertical lines in the third plot give a 95 percent quantile based posterior interval for $\theta$ (gray), as well as the 95% confidence interval based on t-statistics (black).
3.3 Monte Carlo sampling

Figure 3.3: The first panel shows 1000 samples from the Gibbs sampler, plotted over the contours of the discrete approximation. The second and third panels give kernel density estimates to the distribution of Gibbs samples of \( \theta \) and \( \sigma^2 \). Vertical gray bars on the second plot indicate 2.5 and 97.5 percent quantiles of the Gibbs samples of \( \theta \), while nearly identical black vertical bars indicates the 95 percent confidence interval based on the t-test.

Joint inference for the mean and variance

For any joint prior distribution \( p(\theta, \sigma^2) \) for \( \theta \) and \( \sigma^2 \), posterior inference proceeds using Bayes’ rule: Joint posterior of \( \theta, \sigma^2 \)

\[
p(\theta, \sigma^2 | y_1 \ldots y_n) = p(y_1 \ldots y_n | \theta, \sigma^2) p(\theta, \sigma^2) / p(y_1 \ldots y_n)
\]  

\[
\text{variance} = \sigma^2 \sim \text{inverse gamma}(a, b) \\
\text{precision} = \frac{1}{\sigma^2} \sim \text{gamma}(a, b)
\]

Instead of using \( a \) and \( b \) we will parameterized this prior distribution as:

\[
\frac{1}{\sigma^2} \sim \text{gamma}(\nu_0, \frac{\nu_0 \sigma_0^2}{2})
\]
\[ \frac{1}{\sigma^2} | y_1 \ldots y_n | \sim \text{gamma}(\nu_n, \nu_n \sigma_n^2) \]

\[ (\theta | y_1 \ldots y_n, \sigma^2) \sim \text{normal}(\mu_n, \frac{\sigma^2}{k_n}) \]

where, \( k_n = k_0 + n \)

\[ \mu_n = \frac{k_0 \mu_0 + n \bar{y}}{k_n} \]

\[ \nu_n = \nu_0 + n \]

\[ \sigma_n^2 = \frac{1}{\nu_n} [\nu_0 \sigma_0^2 + (n - 1) \sigma^2 + \frac{k_0}{k_n}(\bar{y} - \mu_0)^2] \]

### 3.4 Discussion and further references

The normal sampling model can be justified in many different ways. For example, Lukcas (1942) shows that characterizing feature of the normal distribution is that the sample mean and the sample variance are independent. From a subjective probability perspective, this suggests that if our beliefs about the sample mean are independent from those about the sample variance, then a normal model is appropriate.


Using the posterior predictive distribution to assess model fit was suggested by Guttman (1967) and Rubin (1984), and is now common practice. In some problems, it is useful to evaluate goodness of fit using functions that depend on parameters as well as predicted data. This is discussed in Gelman et al. (1996) and more recently in Johnson (2007).

The term Gibbs sampling was coined by Geman and Geman (1984) in their paper of image analysis, but the algorithm appears earlier in the context of spatial statistics, for example, Besag (1974) or Ripley (1979). However, the general utility of the Gibbs sampler for Bayesian data analysis was not fully realized the late 1980s (Gelfand and Smith, 1990) and Upadhayay for his paper Full posterior analysis of three parameter lognormal distribution using Gibbs sampler.

Assessing the convergence of the Gibbs sampler and the accuracy of the
MCMC approximation is difficult. Several authors have come up with convergence diagnostics (Gelman and Rubin, 1992; Geweke, 1992; Raftery and Lewis, 1992), although these can only highlight problems and not guarantee a good approximation (Geyer, 1992).