Chapter 6

Normal Model with Mean and Variance Unknown

6.1 Multiparameter problems

"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." (Gelman et al. 1995, p. 65).

Although a problem 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 this particular parameter of interest. In principle, the route to achieving this aim is clear: we first require the joint posterior distribution of all unknowns,

$$p(\theta_1, \theta_2, ..., \theta_q \mid y)$$

and then we integrate this distribution over the unknowns that are not of immediate interest to obtain the desired marginal distribution. E.g.,

$$p(\theta_1) = \int \int ...... \int p(\theta_1, \theta_2, ..., \theta_q \mid y) d\theta_2 d\theta_3 ....d\theta_q$$

Or equivalently, using simulation, we draw samples from the joint posterior distribution, $p(\theta_1, \theta_2, ..., \theta_q \mid y)$, and then look at the parameters of interest and ignore the values of the other unknowns. In many problems 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 called *nuisance parameters*.

To draw samples from the joint posterior distribution, we may use grids on which we calculate the joint posterior distribution, or we may be lucky that the posterior factors into known densities and we can simulate directly from them to obtain the joint posterior distribution through numerical integration.

We will be working with joint distributions of only two parameters in this lesson. At this point, you may want to review course outline 2, section 2.8.3, and review the corresponding section in chapter 2 of Hoff's book.

Look also at the table of distributions that I gave you in class to see what a normal distribution looks like, since we are going to be illustrating these computations with the normal distribution of unknown mean and variance.

6.2 Averaging over nuisance parameters

To express the ideas of joint and marginal posterior distributions mathematically, suppose θ has two parts, each of which can be a vector $\theta = (\theta_1, \theta_2)$, and further suppose that we are only interested (at least for the moment) in inference for θ_1, so θ_2 may be considered a 'nuisance' parameter. For instance, in the simple example,

$$y \mid \mu, \sigma^2 \sim N(\mu, \sigma^2),$$

in which both $\mu(\theta_1)$ and $\sigma^2(\theta_2)$ are unknown, interest commonly centers on μ. We seek the conditional distribution of the parameter of interest given the observed data; in this case, $p(\theta_1 \mid y)$. This is derived from the joint posterior density,

$$p(\theta_1, \theta_2 \mid y) \propto p(y \mid \theta_1, \theta_2) p(\theta_1, \theta_2),$$

by averaging over θ_2.

$$p(\theta_1 \mid y) = \int p(\theta_1, \theta_2 \mid y) d\theta_2$$

Alternatively, the joint posterior density can be factored to yield

$$p(\theta_1 \mid y) = \int p(\theta_1 \mid \theta_2, y) p(\theta_2 \mid y) d\theta_2,$$

which shows that the posterior distribution of interest, $p(\theta_1 \mid y)$, is a *mixture* of the conditional posterior distributions given the nuisance parameter, θ_2, where $p(\theta_2 \mid y)$ is a weighting function for the different possible values of θ_2. The weights depend on the posterior density of θ_2 and thus on a combination of evidence from data and prior model. The averaging over nuisance parameters θ_2 can be interpreted very generally; for example, θ_2 can include a discrete component representing different possible submodels.

We rarely evaluate this last integral explicitly, but it suggests an important practical strategy for both constructing and computing with multiparameter models. Posterior distributions can be computed by marginal and conditional simulation, first drawing θ_2 from its marginal posterior distribution and then θ_1 from its conditional posterior distribution, given the drawn value of θ_2. That is,

(a) draw $\theta_2^{(m)} \sim p(\theta_2 \mid y)$

(b) draw $\theta_1^{(m)} \sim p(\theta_1 \mid \theta_2^{(m)}, y)$

(c) find $\theta_1^{(m)}, \theta_2^{(m)}$ for $m = 1, ..., M$

In this way the integration embodied above is performed indirectly via numerical integration A canonical example of this form of analysis is provided by the normal model with unknown mean and variance, to which we now turn. μ is the parameter of interest. σ^2 is the nuisance parameter

6.3 Normal model, μ and σ^2 unknown

As the prototype example of estimating the mean of a population from a sample, we consider a vector y of n iid observations from a univariate normal distribution, $N(\mu, \sigma^2)$. We begin by analyzing the model under a noninformative prior distribution, with the understanding that this is no more than a convenient assumption for the purposed of exposition and is easily extended to informative prior distributions.

6.3.1 A Noninformative Prior Distribution

A sensible vague prior density for μ and σ^2, assuming prior independence of location and scale parameters, is uniform on $(\mu, \log\sigma)$ or, equivalently,

$$p(\mu, \sigma^2) \propto (\sigma^2)^{-1}.$$

6.3.2 The joint posterior distribution, $p(\mu, \sigma^2 \mid y)$

Under this conventional improper prior density, the joint posterior distribution is proportional to the likelihood function multiplied by the factor $\frac{1}{\sigma^2}$. At this point, you may want to read chapter 5 of Hoff's book, to see the details of derivations. You will see there that:

(a) The joint posterior is

$$p(\mu, \sigma^2 \mid y) \propto \sigma^{-(n+2)} \exp\left(-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{y}-\mu)^2]\right)$$

(b) The conditional posterior distribution of μ is

$$\mu \mid \sigma^2 \sim N\left(\mu \mid \bar{y}, \frac{\sigma^2}{n}\right)$$

(c) The marginal posterior distribution of σ^2 is

$$\sigma^2 \mid y \sim IG\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)$$

where
$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$
Notice that you may express this differently. That is, $IG\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right) = IChisquare(n-1, s^2)$.

6.4 Obtaining marginal posterior distribution for μ via simulation from the posterior with numerical integration

$$p(\mu, \sigma^2) = p(\mu \mid \sigma^2, y)p(\sigma^2 \mid y)$$

(a) draw $\sigma^{2(m)} \sim IG\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)$

(b) draw $\mu^{(m)} \sim N(\mu \mid \bar{y}, \sigma^{2(m)}/n)$

(c) $\mu^{(m)}, \sigma^{2(m)}, m = 1, ..., M$ is a sample from the joint posterior distribution $p(\mu, \sigma^2 \mid y)$.

```
####################################
# Numerical integration
#  Example of generating a sample of size NN
# from the joint posterior distribution of a normal model
#  with uninformative prior.
# Our data is going to be generated artificially, to
# convince ourselves that this works
####################################

yy=rnorm(1000,0,sqrt(4))   # this is our data

  nn=length(yy)   # gives us sample size of the data
 ss=var(yy)       # gives us variance of the data
 ybar=mean(yy)    # gives us mean of the data
  AA=(nn-1)/2     # prepare the arguments of the inverse gamma
```

```
 BB=0.5*(nn-1)*ss

########## Open space to put the generated mu and sigma squared ######

mu=rep(NA, 5000)           #   we will generate 5000 simulated mus
 sigmasqr=rep(NA,5000)   #

####### We have two ways of generating the inverted gamma ############

library(MCMCpack)   # if you have installed MCMCpack you can use rinvgamma

 for(m in 1:5000){    # this loop will give the simulations from the joint posterior
 sigmasqr[m]=1/(rgamma(1,AA)/BB)    # This generates the same as next command
# sigmasqr[m]= rinvgamma(1, AA, BB)  # use this if you downloaded MCMCpack

#### For each sigma squared generate a mu  ################

 mu[m]=rnorm(1,ybar,sqrt(sigmasqr[m]/nn))   # plug in sigmasqr and simulate mu given it
 }

 ############## Plot marginals for mu and sigma squared ############

  par(mfrow=c(2,1))  # create space to put the plots.
 hist(mu,density=-1,yaxt="n",yalb="",xlab="mu",cex=2,nclass=50)
 hist(sqrt(sigmasqr),density=-1,yaxt="n",ylab="",xlab="sigma",cex=2,nclass=50)

 ############ Things to do on your own####################
 #
 # (a) Copy paste the plots into your file.
#
# (b) Summarize the marginal posterior distribution for mu. Find marginal mean of mu,
#      marginal variance, intervals,  some probabilities.
#####################################################
```

6.5 Obtaining the marginal posterior for μ by simulating directly from the known distribution of the marginal for μ

The approach we used above to find the posterior distribution of μ is very convenient if we know how to factor the posterior the way we did. Sometimes, in other cases, that is possible also.

In rare cases we know the mathematical form of the posterior distribution of μ but in this case we are studying we do. μ is a t distribution (see Hoff's chapter 5). So an alternative to generate the marginal posterior distribution for μ is to use this result and simulate directly from it (at least in this case we are studying).

6.5.1 Example

Estimation from two independent experiments: an experiment was performed on the effects of magnetic fields on the flow of calcium out of chicken brains. The experiment involved two groups of chickens: a control group of 32 chickens and an exposed group of 36 chickens. One measurement was taken on each chicken, and the purpose of the experiment was to measure the average flow μ_c in untreated (control) chickens and the average flow μ_t in treated chickens. The 32 measurements on the control group had a sample mean of 1.013 and a sample standard deviation of 0.24. The 36 measurements on the treatment group had a sample mean of 1.173 and a sample standard deviation of 0.20.

(a) Assuming the control measurements were taken at random from a normal distribution with mean μ_x and variance σ_c^2, what is the posterior distribution of μ_c? Similarly, use the treatment group measurements to determine the marginal posterior distribution of μ_t. In both cases, assume a uniform prior distribution on $(\mu, \log\sigma)$.

The likelihood: $x_i \sim N(\mu_c, \sigma_c^2), \quad i = 1, ..., n_c = 32; \quad y_i \sim N(\mu_t, \sigma_t^2), \quad i = 1, ..., n_t = 36.$

Priors: $p(\mu_c, \sigma_c^2) = \frac{1}{\sigma_c^2} \quad p(\mu_t, \sigma_t^2) = \frac{1}{\sigma_t^2}$

Posterior: $p(\mu_c \mid x) = t_{n_c-1}(\bar{x}, \frac{s_x^2}{n_c}) = t_{31}(1.013, 0.24^2/32) \quad p(\mu_t \mid y) = t_{n_t-1}(\bar{y}, \frac{s_y^2}{n_t}) = t_{35}(1.173, 0.2^2/36)$

Notice that $\frac{\mu-\bar{x}}{s/sqrtn} = t$ which is centered at 0. We generate the latter in R and use that equality to convert to the original t distribution.

(b) What is the posterior distribution for the difference, $\mu_t - \mu_c$? To get this, you may sample from the independent Student-t distributions you obtained in part (a) above. Plot a histogram of your samples and give an approximate 95% posterior interval for $\mu_t - \mu_c$. Also, provide other summary statistics, and calculate the probability that the treatment group has higher flow than the control group.

Simulate n samples $\mu_{t,j} \sim p(\mu_t \mid y)$ and $\mu_{c,j} \sim p(\mu_c \mid y), \quad j = 1, ..., n$ Compute $d_j = \mu_{c,j} - \mu_{t,j}$ and plot a histogram and compute some summary statistics.

```
######################################################
# We generate the posterior distribution of the difference of means #
# ####################################################


muc= rt(1000,31)*(0.24/sqrt(32))+1.103   # this generates from the posterior for muc
mut= rt(1000,35)*(0.24/sqrt(36))+1.173   # this generates from the posterior for muc
mudif=muc-mut    # the difference between means

summary(mudif)     # will give us some summary statistics
sd(mudif)   # gives the standard deviation

layout(rbind(c(1,3), c(0,0), c(2,2)), heights=c(2,lcm(0.5),1),respect=TRUE)
hist(muc, xlab="muc", main="Posterior for mean of control group")
hist(mudif, xlab="muc-mut")
post.interval=quantile(mudif, c(0.025, 0.975))
abline(v=post.interval)
hist(mut, xlab="mut", main="Posterior for mean of treatment group")

prob.zero=(sum(mudif<0))/1000
prob.zero

####### To do on your own ################
# Copy paste the histograms in your file. Summarize the posterior difference
# between the means of the two groups (mean, variance). In addition to
# Interpret the result obtained by prob.zero... what is the conclusion you
# reach in this study?
#######################################
```

6.6 Obtaining the marginal distribution for μ via grids

This is an approach that will work with any bivariate posterior distribution. It has the advantage that it does not rely on factoring the joint posterior or knowing a closed for for the marginal posterior.

There are several steps to follow to use this approach in any example.

(a) Create a grid. A sequence of values for parameter θ_1 and a sequence for parameter θ_2.

(b) Create a function containing the functional form of your joint posterior distribution $p(\theta_1, \theta_2)$. That is, just type the equation.

(c) Find the value of the $p(\theta_1, \theta_2)$ for all values of those parameters on the grid.

(d) Sample values of θ_1, θ_2 from this grid-created posterior.

(e)