Ex. 7.2.15. Location Normal model.

$(X_1, \ldots, X_n) \sim N(\mu, \sigma_0^2)$; $\mu \sim N(\mu_0, \tau_0^2)$

One can show that the posterior predictive distribution is

$$X_{n+1} \sim N\left(\bar{x}, \left(\frac{1}{\tau_0^2} + \frac{n}{\sigma_0^2}\right)^{-1} \sigma_0^2\right) \Rightarrow \hat{t} = \bar{x}, \text{ this is both}$$

the posterior mode and the posterior mean. Note that the future obs. is not independent of the observed data.

7.3 Bayesian Computation.

7.3.1 Asymptotic Normality of the Posterior

Under some regularity conditions, if $(X_1, \ldots, X_n) \sim f_\theta$,

$$\Pi\left(\frac{\theta - \hat{\theta}(x_1, \ldots, x_n)}{\hat{\sigma}(x_1, \ldots, x_n)} \le z \,\middle|\, x_1, \ldots, x_n\right) \to \Phi(z) \text{ as } n \to \infty$$

where $\hat{\theta}(x_1, \ldots, x_n)$ is the posterior mode, and

$$\hat{\sigma}^2(x_1, \ldots, x_n) = \left(- \frac{\partial^2 \ln L(\theta | x_1, \ldots, x_n)\pi(\theta)}{\partial \theta^2}\bigg|_{\theta = \hat{\theta}}\right)^{-1}$$

7.3.2 Sampling from the posterior

May want to compute the posterior probability of $A \subset \Omega$ given s, i.e. $\Pi(A|s) = E(I_A(\theta)|s)$

More generally, for some $w(\theta)$, we may want to compute $E(w(\theta)|s)$ (*). Thus, many things can be reduced to estimating expectations. But note, that, if due to sampling error, the value of a parameter may only be known to within $\pm$ 0.1 units, there is no point in computing an estimate to more digits of accuracy. Thus, the exact evaluation of (*) may not be necessary. Monte Carlo: if I can generate an i.i.d sequence from the posterior distribution of θ, can estimate (*) by

$$\bar{w} = \frac{1}{N} \sum_{i=1}^{N} w(\theta_i).$$

Then, by SLLN, $\bar{w} \xrightarrow{a.s.} \mathbb{E}(w(\theta)/X)$ as $N \to \infty$.

How large should N be? If $\mathbb{E}(w^2(\theta)/S) < \infty$,

by CLT $\dfrac{\bar{w} - \mathbb{E}(w(\theta)/S)}{\sigma_w/\sqrt{N}} \xrightarrow{D} N(0,1)$ as $N \to \infty$ where

$\sigma_w^2 = \cancel{\mathrm{Var}\left(\frac{w(\theta)/S}{N}\right)}\ \mathrm{Var}\left(w(\theta)/S\right)$. Typically,

$\hat{\sigma}_w^2 = \dfrac{1}{N-1} \sum_{i=1} \left(w(\theta_i) - \bar{w}\right)^2$ when $w(\theta)$ is a quantitative variable;

if $w = I_A$ for $A \subset \Omega$, $S_w^2 = \bar{w}(1-\bar{w})$. In either case

S_w^2 is a consistent estimate of σ_w^2. $\Rightarrow$

$\Rightarrow \dfrac{\bar{w} - \mathbb{E}(w(\theta)/S)}{S_w/\sqrt{N}} \xrightarrow{D} N(0,1)$ as $N \to \infty$.

Thus, $\bar{w} \pm 3\dfrac{S_w}{\sqrt{N}} \simeq 100\%$ CI for $\mathbb{E}(w(\theta)/S)$ ....

this would suggest how large N should be selected.

Note that $\dfrac{S_w}{\sqrt{N}}$ is itself subject to error. Usually, one

monitors $3\dfrac{S_w}{\sqrt{N}}$ for successively larger values of N

and stops sampling only when it is clear that $3\dfrac{S_w}{\sqrt{N}}$

is small enough for accuracy desired and is declining

appropriately.

<u>Ex. 7.3.1</u> Location - scale normal

$$(x_1, \ldots, x_n) \sim N(\mu, \sigma^2); \text{ prior}$$

$$\mu/\sigma^2 \sim N(\mu_0, \tau_0\sigma^2),$$

$$\frac{1}{\sigma^2} \sim \mathrm{Gamma}(\alpha_0, \beta_0)$$

Recall that $\mu/\sigma^2, x_1, \ldots, x_n \sim N\left(\mu_x, \left(n + \frac{1}{\tau_0^2}\right)^{-1}\sigma^2\right)$

$$\frac{1}{\sigma^2}\Big/ x_1, \ldots, x_n \sim \mathrm{Gamma}\left(\alpha_0 + \frac{n}{2}, \beta_x\right)$$

In practice, one usually generates $\dfrac{1}{\sigma^2}$ from the gamma distr;

then μ/σ^2 from the normal distribution.

Thus, we can have a sample $(\mu_1, \sigma_1^2), \dots, (\mu_N, \sigma_N^2)$

Suppose we want to obtain the posterior distribution of $\psi = \frac{\sigma}{\mu}$ – clear how to get a posterior sample for ψ.

Consider the following sample $(x_1, \dots, x_{15})$:

11.6714 1.8957 2.1228 2.1286 1.0751
8.1631 1.8236 4.0362 6.8513 7.6461
1.9020 7.4899 4.9233 8.3223 7.9486

Here, $\bar{x} = 5.2$; $s = 3.3$. Let the prior be: $\mu_0 = 4$, $\tau_0^2 = 2$, $\alpha_0 = 2$, $\beta_0 = 1$. By (7.1.7), $\mu_x = \left(15 + \frac{1}{2}\right)^{-1}\left(\frac{4}{2} + 15 \cdot 5.2\right) = 5.161$

By (7.1.8) $\beta_x = 1 + \frac{15}{2}(5.2)^2 + \frac{4^2}{2 \cdot 2} + \frac{14}{2}(3.3)^2 - \frac{1}{2}\left(15 + \frac{1}{2}\right)^{-1} \times$

$\times \left(\frac{4}{2} + 15 \cdot 5.2\right)^2 = 77.578$

$\Rightarrow$ generate $\frac{1}{\sigma^2} \mid x_1, \dots, x_n \sim \text{Gamma}(9.5, 77.578)$

& $\mu \mid \sigma^2, x_1, \dots, x_n \sim N\left(5.161, (15.5)^{-1}\sigma^2\right)$

Code – Appendix B; illustrate the histogram of $\psi = \frac{\sigma}{\mu}$ using 10^4 observations; note the long right tail.

Now, suppose we want to estimate $\Pi(\psi \le 0.5 \mid x_1, \dots, x_n) = \mathbb{E}\left(I_{(-\infty, 0.5)}(\psi) \mid x_1, \dots, x_n\right)$. Note that $I_{(-\infty, 0.5)}(\psi)$ is finite, thus there exists the finite variance and CLT can be used. Illustrate the estimated $\Pi(\psi \le 0.5 \mid x_1, \dots, x_n)$ and its standard error for e.g. $N = 10^4$

Example 9.3.1

```
x <- read.csv(file = "datavector731.csv")$X
post <- post_normal(10^4, x, 2, 1, 4, 2)
z <- sqrt(post$sigmasq)/post$mu
hist(z)
```

OR (after post):

```
psi <- sqrt(post$sigmasq)/post$mu
psi_hat <- mean(psi <= .5)
psi_se <- sqrt(psi_hat*(1-psi_hat))/sqrt(length(psi))
cq <- 3
```

cat("The estimate interval is", "[", psi_hat - cq*psi_se, ",", psi_hat + cq*psi_se, "]" \n")

7.3.3 Gibbs sampling.

1. Need to generate samples from the joint distribution of $(Y_1, \ldots, Y_k) \in \mathbb{R}^k$

Moreover, suppose we can generate from each of the full conditional distributions

$$Y_i \mid Y_{-i} = y_{-i},$$

where $Y_{-i} = (Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_k)$

The following steps define Gibbs sampler

1. Specify an initial value $(Y_{1(0)}, \ldots, Y_{k(0)})$ for $(Y_1, \ldots, Y_k)$

2. For $N > 0$, generate $Y_{i(N)}$ from its conditional distribution given $(Y_{1(N)}, \ldots, Y_{i-1(N)}, Y_{i+1(N-1)}, \ldots, Y_{k(N-1)})$ for each $i = 1, \ldots, k$

For example, if $k=3$, we specify first $(Y_{1(0)}, Y_{2(0)}, Y_{3(0)})$.

Then, $Y_{1(1)} \mid Y_{2(0)} = y_{2(0)}, Y_{3(0)} = y_{3(0)}$

$$Y_{2(1)} \mid Y_{1(1)} = y_{1(1)}, Y_{3(0)} = y_{3(0)}$$

$$Y_{3(1)} \mid Y_{1(1)} = y_{1(1)}, Y_{2(1)} = y_{2(1)}$$

to obtain $(Y_{1(1)}, Y_{2(1)}, Y_{3(1)})$. Next,

$$Y_{1(2)} \mid Y_{2(1)} = y_{2(1)}, Y_{3(1)} = y_{3(1)}$$

$$Y_{2(2)} \mid Y_{1(2)} = y_{1(2)}, Y_{3(1)} = y_{3(1)}$$

$$Y_{3(2)} \mid Y_{1(2)} = y_{1(2)}, Y_{2(2)} = y_{2(2)}$$

to obtain $(Y_{1(2)}, Y_{2(2)}, Y_{3(2)})$ etc. Note we didn't have to specify $Y_{1(0)}$ as it is never used.

Under some technical conditions, $(Y_{1(N)}, \ldots, Y_{k(N)}) \xrightarrow{D} (Y_1, \ldots, Y_k)$ as $N \to \infty$. Furthermore (even more relevant for application) $\bar{w} = \frac{1}{N}\sum_{i=1}^{N} w(Y_{1(i)}, \ldots, Y_{k(i)}) \xrightarrow{a.s.} \mathbb{E}(w(Y_1, \ldots, Y_k))$

However, estimation of the variance of $\bar{w}$ is a bit more difficult because terms $w(Y_{1(i)}, \ldots, Y_{k(i)})$ are not independent for different i

The most commonly used approach is branching.
First, the sequence $w(Y_{1(0)},\ldots,Y_{k(0)}),\ldots,w(Y_{1(N)},\ldots,Y_{k(N)})$ is divided into N/m nonoverlapping sequential batches of size m, calculate the mean in each batch obtaining $\bar{w}_1,\ldots,\bar{w}_{N/m}$, then estimate $Var(\bar{w}) \approx \frac{S_b^2}{N/m}$
where $S_b^2 = \frac{1}{N/m-1}\sum_{i=1}^{N/m}(\bar{w}_i-\bar{w})^2$. This works because, for sufficiently large m, $(Y_{1(i)},\ldots,Y_{k(i)})$ and $(Y_{1(i+m)},\ldots,Y_{k(i+m)})$ are approximately independent. m has to be chosen sufficiently large but still allowing a sufficient number of df.

Sometimes, even this approach cannot be used directly.
Then, Gibbs sampling via latent variables is typically used.
For this, we search for some RVs $(V_1,\ldots,V_l)$ such that Y_i is a function of $(V_1,\ldots,V_l)$ for each i and such that Gibbs sampling can be applied to the joint distr. of $(V_1,\ldots,V_l)$.

Ex. Location-scale Student.
$(X_1,\ldots,X_n)$ is a sample from $X=\mu+\sigma Z$, $Z\sim t(\lambda)$.
If $\lambda>2$, μ is the mean, $\sigma\sqrt{\frac{\lambda}{\lambda-2}}$ is the SD.
Fix λ at some specified value.
Prior: $\mu|\sigma^2 \sim N(\mu_0,\tau_0^2\sigma^2)$,
$1/\sigma^2 \sim Gamma(\alpha_0,\beta_0)$.
The likelihood function is $\left(\frac{1}{\sigma^2}\right)^{n/2}\prod_{i=1}^{n}\left[1+\frac{1}{\lambda}\left(\frac{x_i-\mu}{\sigma}\right)^2\right]^{-\frac{\lambda+1}{2}}$ (*)

Hence the posterior density of $(\mu,\frac{1}{\sigma^2})$ is

$$\propto \left(\frac{1}{\sigma^2}\right)^{n/2}\prod_{i=1}^{n}\left[1+\frac{1}{\lambda}\left(\frac{x_i-\mu}{\sigma}\right)^2\right]^{-\frac{\lambda+1}{2}}\times\left(\frac{1}{\sigma^2}\right)^{1/2}e^{-\frac{1}{2\tau_0^2\sigma^2}(\mu-\mu_0)^2}\times$$

$$\times\left(\frac{1}{\sigma^2}\right)^{\alpha_0-1}\exp\left(-\frac{\beta_0}{\sigma^2}\right)$$

How to generate from it — unclear!!