STAT 591 NOTES – Gibbs sampler examples

Ryan Martin (rgmartin@uic.edu)

November 11, 2013

Example 1: Bivariate normal

Let $Y = (Y_1, Y_2)^\top$ be a bivariate normal sample (of size 1), where the mean $\theta = (\theta_1, \theta_2)^\top$ is unknown, but the variances are known and equal to 1, and the correlation ρ is also known. If we take a flat prior for θ, then the posterior is also bivariate normal, with the same correlation matrix, but with mean Y. We can do a Gibbs sampler from this bivariate normal distribution as follows:

1. Start with, say, $\theta_1^{(0)} = Y_1$.
2. For $m = 1, \ldots, M$, do

$$\begin{aligned} \theta_2^{(m)} \mid (\theta_1^{(m-1)}, Y) &\sim \mathsf{N}(Y_2 + \rho(\theta_1^{(m-1)} - Y_1), 1 - \rho^2) \\ \theta_1^{(m)} \mid (\theta_2^{(m)}, Y) &\sim \mathsf{N}(Y_1 + \rho(\theta_2^{(m)} - Y_2), 1 - \rho^2). \end{aligned}$$

   Can throw away some of the first few iterations as a burn-in.

Then for sufficiently large M, the sample $\{\theta^{(m)} = (\theta_1^{(m)}, \theta_2^{(m)})^\top : m = 1, \ldots, M\}$ ought to be close to a sample from the bivariate normal posterior. In particular, if R is the correlation matrix defined above, then

$$Z_m = (\theta^{(m)} - Y)^\top R^{-1} (\theta^{(m)} - Y), \quad m = 1, \ldots, M$$

should be close to an independent $\mathsf{ChiSq}(2)$ sample. We can do a Kolmogorov–Smirnov test (or some other goodness-of-fit test) to check this.

Example 2: Normal–inverse gamma sampling

Consider a distribution known as the normal–inverse gamma,[1] with parameters $m \in (-\infty, \infty)$, $r > 0$, $a > 0$, and $b > 0$; this will be denoted by $\mathsf{NiGam}(m, r, a, b)$. The density function for $(U, V) \sim \mathsf{NiGam}(m, r, a, b)$ is given by

$$f(u, v) \propto v^{-(2a+3)/2} \exp\Bigl\{-\frac{r(u-m)^2 + 2b}{2v}\Bigr\}, \quad (u, v) \in (-\infty, \infty) \times (0, \infty);$$

the proportionality constant can be evaluated, but it's not important. This distribution is just the joint distribution of (U, V) when $\frac{1}{V} \sim \mathsf{Gam}(a, b)$ and $U \mid V \sim \mathsf{N}(m, V/r)$; note that b is a *rate parameter* in this version. To implement a Gibbs sampler, in addition to the $U \mid V$ conditional distribution, we need $V \mid U$. The easiest way to see this is to note that the conditional density of $V \mid U$, as a function of v for given u, is proportional to the joint density $f(u, v)$. With a little reflection, one will see that $f(u, v)$ is proportional to an inverse gamma density in v with shape parameter $(2a + 1)/2$ and rate parameter $(r/2)(u - m)^2 + b$. Therefore,

$$U \mid (V = v) \sim \mathsf{InvGam}(a + \tfrac{1}{2}, \tfrac{r}{2}(u - m)^2).$$

So, a Gibbs sampler for $\mathsf{NiGam}(m, r, a, b)$ goes as follows:

[1] http://en.wikipedia.org/wiki/Normal-inverse-gamma_distribution

1. Start with, say, $U_0 = m$.
2. For $t = 1, \ldots, T$, do:

$$\begin{aligned} V_t \mid U_{t-1} &\sim \mathsf{InvGam}(a + \tfrac{1}{2}, \tfrac{r}{2}(U_{t-1} - m)^2) \\ U_t \mid V_t &\sim \mathsf{N}(m, V_t/r). \end{aligned}$$

The reason why this distribution is important is that NiGam is a conjugate prior for the $\mathsf{N}(\mu, \sigma^2)$ model. In particular, if the Bayes model is

$$\begin{aligned} (X_1, \ldots, X_n) \mid (\mu, \sigma^2) &\overset{\text{iid}}{\sim} \mathsf{N}(\mu, \sigma^2) \\ (\mu, \sigma^2) &\sim \mathsf{NiGam}(m, r, a, b), \end{aligned}$$

then $(\mu, \sigma^2) \mid (X_1, \ldots, X_n) \sim \mathsf{NiGam}(m', r', a', b')$, where

$$\begin{aligned} m' &= \frac{rm + n\hat{\mu}}{r + n} \\ r' &= r + n \\ a' &= a + \frac{n}{2} \\ b' &= b + \frac{n}{2}\Big\{\hat{\sigma}^2 + \frac{r}{r+n}(\hat{\mu} - m)^2\Big\}, \end{aligned}$$

and $\hat{\mu} = \bar{X}$, $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \hat{\mu})^2$ are the MLEs of μ and σ^2, respectively. See, for example, Gelman et al. (2004), Section 3.3. For this model, posterior inference can be obtained by running the Gibbs sampler described above.

Example 3: Capture–recapture study

(Taken from Lange 2010, Example 26.3.3.) Consider a population of unknown size N. For concreteness, consider a lake that contains N fish. To estimate N, a capture–recapture study is conducted, i.e., at n different times, fish are caught, marked, and returned to the lake. At each time point $i = 1, \ldots, n$, the data consists of a pair (C_i, R_i), where C_i denotes the total number of fish caught at time i and R_i is the number of "recaptures" at time i. Obviously, $R_1 \equiv 0$. Let $U_i = \sum_{j=1}^i (C_j - R_j)$ denote the total number of distinct fish caught by time i, where $U_0 \equiv 0$. Introduce a new set of parameters $\omega_1, \ldots, \omega_n)$, all taking values in $(0, 1)$. The model assumes independent binomial sampling, so the likelihood function looks like

$$\begin{aligned} L(N, \omega) &= \prod_{i=1}^n \binom{U_{i-1}}{R_i} \omega_i^{R_i}(1 - \omega_i)^{U_{i-1} - R_i} \binom{N - U_{i-1}}{C_i - R_i} \omega_i^{C_i - R_i}(1 - \omega_i)^{N - U_{i-1} - C_i + R_i} \\ &= \prod_{i=1}^n \binom{U_{i-1}}{R_i}\binom{N - U_{i-1}}{C_i - R_i} \omega_i^{C_i}(1 - \omega_i)^{N - C_i} \\ &= \frac{N!}{(N - U_n)!} \prod_{i=1}^n \binom{U_{i-1}}{R_i} \omega_i^{C_i}(1 - \omega_i)^{N - C_i}. \end{aligned}$$

For a prior, let's take $N \sim \mathsf{Pois}(m)$ and $\omega_i \sim \mathsf{Beta}(a, b)$, all independent. So, then, the posterior distribution looks like

$$\pi(N, \omega \mid \text{data}) \propto \frac{N!}{(N-U_n)!} \frac{m^N}{N!} \prod_{i=1}^{n} \binom{U_{i-1}}{R_i} \omega_i^{C_i+a-1}(1-\omega_i)^{N-C_i+b-1}.$$

For the Gibbs sampler, we need all the conditional distributions. It is clear that

$$(\omega_1, \ldots, \omega_n) \mid (\text{data}, N) \overset{\text{ind}}{\sim} \mathsf{Beta}(a + C_i, b + N - C_i), \quad i = 1, \ldots, n.$$

For N, given data and all the ω_i's, a little reflection will reveal that

$$(N - U_n) \mid (\text{data}, \omega_1, \ldots, \omega_n) \sim \mathsf{Pois}\Big(m \prod_{i=1}^{n} (1-\omega_i)\Big).$$

So, a Gibbs sampler goes as follows:

1. Set $N^{(0)} = U_n + 1$, say.
2. For $t = 1, \ldots, T$, do

$$\omega_i^{(t)} \mid (\text{data}, N^{(t-1)}) \overset{\text{ind}}{\sim} \mathsf{Beta}(a + C_i, b + N^{(t-1)} - C_i), \quad i = 1, \ldots, n,$$
$$N^{(t)} \mid (\text{data}, \omega_1^{(t)}, \ldots, \omega_n^{(t)}) \sim U_n + \mathsf{Pois}\Big(m \prod_{i=1}^{n} (1-\omega_i^{(t)})\Big).$$

For illustration, consider the data on Gordy lake sunfish presented in Example 26.3.3 in Lange (2010). Here fishing was done at $n = 14$ different times; the data is given with the R code in Figure 2. As a starting point, we use a data-based choice $N^{(0)} = 457$, which is based on the MLE. For hyper parameters, we take $a = b = 1$ and $m = 457$. A sample of size 10,000 is taken from the posterior distribution of N, and this is displayed in the histogram in Figure 1. In this case, the posterior mean is roughly 443.

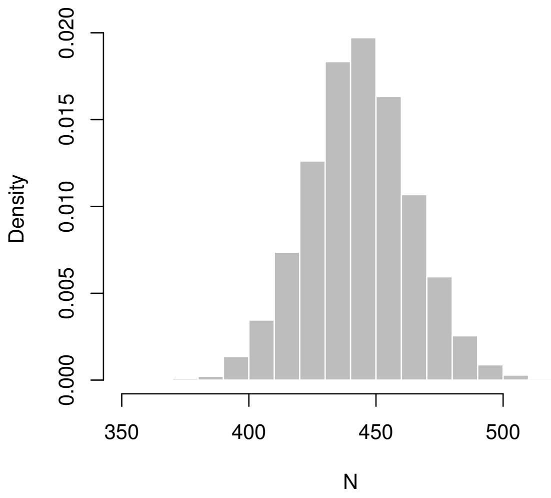

Figure 1: Posterior distribution of N, the total number of sunfish in Gordy lake.

```
cc <- c(10, 27, 17, 7, 1, 5, 6, 15, 9, 18, 16, 5, 7, 19)
rr <- c(0, 0, 0, 0, 0, 0, 2, 1, 5, 5, 4, 2, 2, 3)
uu <- cumsum(cc - rr)
n <- length(cc)
M <- 10000
B <- 1000
N <- numeric(B + M)
w <- matrix(0, nrow=B + M, ncol=n)
a <- b <- 1
m <- 457
N[1] <- m
w[1,] <- 0.02
for(r in 2:(B + M)) {

  w[r,] <- rbeta(n, a + cc, b + N[r-1] - cc)
  N[r] <- uu[n] + rpois(1, m * prod(1 - w[r,]))

}
N.gibbs <- N[-(1:B)]
hist(N.gibbs, freq=FALSE, xlab="N", col="gray", border="white", main="")
```

Figure 2: R code for capture–recapture example.