Bayesian Inference and Computational Methods 2009

Homework 6

1. Suppose that the bivariate density $\pi(x, y)$ on $\mathbb{R}^2_+$ (i.e. the *positive* quadrant $\{(x, y)\colon x \geq 0, y \geq 0\}$) is given by, for some parameters $\lambda > 0$, $\mu > 0$,

$$\pi(x, y) \propto \exp(-[\lambda x + \lambda y + \mu xy]), \qquad x \geq 0, \quad y \geq 0.$$

   (a) Show that the conditional distribution of y, given x, is $\mathrm{Exp}(\lambda + \mu x)$, and similarly that the conditional distribution of x, given y, is $\mathrm{Exp}(\lambda + \mu y)$.

   (b) Write a programme using the Gibbs sampler to sample from π, using these conditional distributions. (One possible such programme is the **R** function `gscexp` available at http://www.ma.hw.ac.uk/~stan/f73bi/R/) For the case $\lambda = \mu = 1$, investigate both theoretically and practically the mixing properties of the sampler.

   (c) Again for the case $\lambda = \mu = 1$, generate a histogram estimate of the marginal density of x. Estimate also the mean and the standard deviation of the marginal distribution of x.

   (d) Estimate the probability, under the *joint distribution* of x and y, that *both* x and y are less than 1.

2. Let $\pi(x, y)$ denote a bivariate density which is uniform over the region bounded by the lines $x = 0$, $x = 1$, $x + y = 1$, and $x + y = 2$.

   (a) Identify the conditional densities $\pi(y \,|\, x)$, $\pi(x \,|\, y)$, for $0 < x < 1$ and $0 < y < 2$. (A sketch of the region will help).

   (b) Write a programme using the Gibbs sampler to sample from π, using these conditional densities.

   (c) Generate histogram estimates of the marginal densities $\pi(x)$ and $\pi(y)$, and verify that your algorithm produces estimates that agree with the theoretical marginal densities (which you will have to work out!).

   (d) Suppose that you used a similar algorithm to sample from a density that was uniform on the region bounded by $x = 0$, $x = 1$, $x + y = 1$, and $x + y = 1.1$. By considering the shape of this region, suggest why your Gibbs sampler would not explore the target density efficiently.

3. Suppose that the bivariate density $\pi(x, y)$ on $\mathbb{R}^2$ is given by

$$\pi(x, y) = \begin{cases} kxy, & x \geq 0, \quad y \geq 0, \quad x + y \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

   for some normalising constant $k > 0$.

   (a) Show that the conditional density $\pi(y \,|\, x)$ of y, given x, is given by

$$\pi(y \,|\, x) = \begin{cases} \dfrac{2y}{(1 - x)^2}, & 0 \leq y \leq 1 - x, \\ 0 & \text{otherwise.} \end{cases}$$

   Clearly a similar result holds for the conditional density $\pi(x \,|\, y)$ of x, given y.

   (b) Write a programme using the Gibbs sampler to sample from π, using these conditional distributions. (You will find it easiest to simulate from these conditional distributions by using the inverse transform method. One possible such programme is the **R** function `gsbv` available at
   http://www.ma.hw.ac.uk/~stan/f73bi/R/
   but you will learn much more if you write your own.) Investigate both theoretically and practically the mixing properties of the sampler.

(c) Generate a histogram estimate of the marginal density of x. Estimate also the unconditional (marginal) probability that $x \leq 1/4$, and compare with the theoretical value.

(d) Estimate the probability, under the *joint distribution* of x and y, that *both* x and y are less than $1/2$.

4. Suppose that $\boldsymbol{y} = (y_1, \ldots, y_n)$ is a random sample of size n from a normal distribution with unknown mean μ and unknown variance $\sigma^2 = \frac{1}{\psi}$. The prior distributions of μ and ψ are taken to be independent, with μ having an improper prior distribution which is uniform on $\mathbb{R}$ (i.e. constant density on $\mathbb{R}$), and ψ having an improper prior distribution whose density on $\mathbb{R}_+$ is proportional to ψ^{-1} (note that the latter is equivalent to a constant, improper, density for $\log \psi$).

(a) Show that the posterior joint distribution of (μ, ψ) is given by the density

$$
\begin{aligned}
\pi(\mu, \psi \,|\, \boldsymbol{y}) &\propto \psi^{\frac{n-2}{2}} \exp\left(-\frac{\psi}{2}\sum_{i=1}^{n}(y_i - \mu)^2\right) \\
&\propto \psi^{\frac{n-2}{2}} \exp\left(-\frac{\psi}{2}\left(\sum_{i=1}^{n}(y_i - \bar{y})^2 + n(\bar{y} - \mu)^2\right)\right) \\
&\propto \psi^{\frac{n-2}{2}} \exp\left(-\frac{\psi}{2}\Big((n-1)s^2 + n(\bar{y} - \mu)^2\Big)\right),
\end{aligned}
$$

where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ is the sample mean and $s^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2/(n-1)$ is the sample variance. Deduce that the conditional distribution of ψ given μ is $\Gamma\left(\frac{n}{2}, \frac{1}{2}\Big((n-1)s^2 + n(\bar{y} - \mu)^2\Big)\right)$, while the conditional distribution of μ given ψ is $\mathrm{N}\left(\bar{y}, (n\psi)^{-1}\right)$.

(b) Write a programme (taking as its input the vector $\boldsymbol{y}$ of observations and computing initially $\bar{y}$ and $(n-1)s^2$ as sufficient statistics, given this model) to use the Gibbs sampler to simulate from the above posterior distribution of (μ, ψ). (Recall also that the **R** function `rnorm` uses the *standard deviation*—not the variance—as its scale parameter.)

(c) Generate a random sample of size 20 from an $\mathrm{N}(1, 0.5)$ distribution (standard deviation $(0.5)^{1/2}$). Use the Gibbs sampler to investigate the above posterior distribution of (μ, ψ) given your data. As always verify the mixing properties of the Gibbs sampler. By generating a suitably long sequence of iterates from the chain, estimate the posterior mean of each of the parameters, and the posterior probabilities that $\mu > 1.5$ and that $\sigma^2 > 0.75$. Estimate also 95% equal-tailed credible intervals for μ and for ψ.

(d) Repeat these calculations for a sample of size 60 from an $\mathrm{N}(1, 0.5)$ distribution.

(e) From the results of Section 1.5.1 of the lecture notes it follows that (i) the marginal posterior distribution of μ is such that

$$
\frac{\mu - \bar{y}}{s/n^{1/2}} \sim t_{n-1},
$$

and (ii) the marginal posterior distribution of ψ is such that $(n-1)s^2\psi \sim \chi^2_{n-1}$. Use these results to check whether your Gibbs sampler is working correctly.

5. Let n_1 and n_2 denote the number of calls being served at any one time by each of two adjacent mobile phone masts. The joint distribution of n_1 and n_2 is given by the probability function

$$\pi(n_1, n_2) = a \frac{\lambda_1^{n_1}}{n_1!} \frac{\lambda_2^{n_2}}{n_2!} \phi(n_1, n_2)$$

where $\lambda_1 > 0$, $\lambda_2 > 0$, $\phi(n_1, n_2) \leq 1$ for all n_1, n_2 and $\phi(0,0) = 1$, and a is the appropriate normalising constant. Thus $\phi(n_1, n_2)$ represents the reduction in the probability of the configuration (n_1, n_2), relative to $(0,0)$, due to interference between the two masts.

   (a) Suppose that

   $$\phi(n_1, n_2) = \begin{cases} 1, & n_1 + n_2 \leq N \\ 0, & n_1 + n_2 > N, \end{cases}$$

   for some N which denotes the maximum number of calls which can be in progress at any time. Show that the conditional distributions $\pi(n_2|n_1)$ and $\pi(n_1|n_2)$ are both truncated Poisson distributions. Write a program using the Gibbs sampler to simulate the joint distribution $\pi(n_1, n_2)$ of n_1 and n_2 and to estimate (via histograms) the marginal distributions of n_1 and n_2, their means, and also the constant $a = \pi(0,0)$. Carry out the estimations in the case $\lambda_1 = \lambda_2 = 20$ and $N = 30$. (You may wish to compare your results with the theoretical values, which are here computable.)

   (b) Suppose instead that $\phi(n_1, n_2) = \alpha^{n_1 n_2}$ for some $\alpha < 1$. Find the conditional distributions $\pi(n_2|n_1)$ and $\pi(n_1|n_2)$ in this case. Modify the Gibbs sampler of the first part of the question to estimate (again via histograms) the marginal distributions of n_1 and n_2, together with their means, in the case $\lambda_1 = \lambda_2 = 20$ and $\alpha = 1/2$.

6. *(Challenging.)*

   (a) Write a program using the Gibbs sampler for the spatial process considered in the lecture notes. In the case $N = 11$, $\alpha = 1$, $\beta = 0.9$, estimate the probability that the vertex $(6, 6)$ is in the state 1.

   (b) Adapt the program to study also the cases where (a) all states on the boundary are constrained to be 0, (b) all states on the boundary are constrained to be 1. For $N = 11$, $\alpha = 1$, $\beta = 0.9$, investigate how the probability that the vertex $(6, 6)$ is in the state 1 varies between these two cases.

   (c) Investigate the difficulties caused by using values of β far from 1.

Bayesian Inference and Computational Methods 2009

Homework 6 – Solutions

1. (a) The conditional density of y, given x, is given by

$$\begin{aligned}\pi_{y\,|\,x}(y\,|\,x) &\propto \exp(-[\lambda x + \lambda y + \mu xy]) \\ &\propto \exp(-(\lambda + \mu x)y),\end{aligned}$$

which corresponds to an $\mathrm{Exp}(\lambda + \mu x)$ distribution. The second result follows similarly.

(b) The supplied **R** function `gscexp` can be used, for example, as follows.

```
gs = gscexp(100000,1,1)
plot(gs[1:1000,])
plot(gs[1:100,],type='l')
```

The mixing is clearly very good, as can also be argued theoretically by considering the effects of simulating alternately from the two conditional distributions.

(c) Noting that `gs[,1]` corresponds to the sequence of simulations of the x coordinate of the chain, we can treat this sequence as a sample from the marginal distribution of x. A long simulation as above gives estimates of the mean and standard deviation of this marginal distribution of 0.68 and 0.74 respectively.

(d) **R** code to estimate the probability, under the *joint distribution* of x and y, that *both* x and y are less than 1 is given, for example, by

```
sum(gs[,1]<1 & gs[,2]<1)/100000
```

and gives an estimate of 0.57.

2. (a) For $0 < x < 1$, the conditional distribution of y, given x, is $\mathrm{U}(1-x, 2-x)$. For $0 < y \leq 1$, the conditional distribution of x, given y, is $\mathrm{U}(1-y, 1)$, while for $1 < y < 2$, the conditional distribution of x, given y, is $\mathrm{U}(0, 2-y)$.

(b) A programme using the Gibbs sampler to sample from π may be constructed by suitably modifying the **R** function `gscexp` used earlier. Not the need to distinguish between values of y between 0 and 1 and between 1 and 2.

(c) Again we may proceed as in the earlier question. Note that the theoretical marginal distribution of x is $\mathrm{U}(0, 1)$, while that of y is given by the density

$$\pi(y) = \begin{cases} y, & 0 < y \leq 1, \\ 2 - y, & 1 < y < 2, \\ 0, & \text{otherwise.} \end{cases}$$

(d) A sketch of the region would show clearly that in this case the Gibbs sampler (which, recall, samples alternately from the conditional distribution of y given x and from the conditional distribution of x given y) would mix rather slowly.

3. (a) Clearly the conditional density $\pi(y \mid x)$ of y, given x, is of the form

$$\pi(y \mid x) = \begin{cases} k'y, & 0 \leq y \leq 1-x, \\ 0 & \text{otherwise.} \end{cases}$$

for an appropriate normalising constant k'. The requirement that the integral of the density is 1 shows $k' = 2/(1-x)^2$.

(b) The **R** code

```
> gs = gsbv(100000)
> plot(gs[1:100,],type='l')
> plot(gs[1:1000,])
```

shows the mixing properties of the Gibbs sampler to be very good in this case. This is also not difficult to see by considering the theoretical behaviour of the sampler.

(c) The unconditional (marginal) probability that $x \leq 1/4$ may be estimated by

```
sum(gs[,1]<=1/4)/100000
```

and compare with the theoretical value of 67/256.

(d) Appropriate **R** code is

```
sum(gs[,1]<1/2 & gs[,2]<1/2)/100000
```

Again the result may be compared with the theoretical value (*exercise*).

4. (a) The prior joint density is given by $\pi(\mu, \psi) \propto \psi^{-1}$, and so the posterior joint density is immediate on writing down the likelihood function associated with the data. The required conditional distributions now follow immediately.

(b) The required **R** function is entirely analogous to previous **R** functions for the Gibbs sampler.

(c) This is a *practical exercise*, the results to be confirmed using the theory in part (e).

(d) This is again a *practical exercise*.

(e) The mean of the posterior distribution of μ is $\bar{y}$, while the posterior probability that $\mu > 1.5$ is

$$\mathbf{P}\left(T > \frac{1.5 - \bar{y}}{s/n^{1/2}}\right)$$

where the random variable $T \sim t_{n-1}$. A 95% equal-tailed credible interval for μ is given by

$$\left(\bar{y} - \frac{s}{n^{1/2}} t_{n-1,0.975},\ \bar{y} - \frac{s}{n^{1/2}} t_{n-1,0.975}\right)$$

where $t_{n-1,0.975}$ is the 0.975-quantile of the t_{n-1} distribution—exactly as the 95% confidence interval of classical statistics in this case. The results for ψ (*exercise*) similarly correspond to those obtained using classical statistics.

5. (a) For the given function ϕ, the conditional distribution of n_2 given n_1 has probability function

$$\phi(n_2 \,|\, n_1) = \begin{cases} a' \dfrac{\lambda_2^{n_2}}{n_2!}, & 0 \le n_2 \le N - n_1, \\ 0, & \text{otherwise,} \end{cases}$$

for some appropriate normalising constant a'. This is a truncated Poisson distribution. Similarly the conditional distribution of n_1 given n_2 is a truncated Poisson distribution.

The **R** function gsmp at http://www.ma.hw.ac.uk/~stan/f73bi/R/) uses the Gibbs sampler to simulates the joint distribution $\pi(n_1, n_2)$ of n_1 and n_2. It uses rejection sampling to simulate the conditional distributions. Suitable code (with sample output) to estimate the marginal distribution of n_1 and its mean, for the given parameters, is given by

```
> gs = gsmp(100000,20,20,30)
> plot(gs[1:1000,], type='l')  # show distribution and check mixing
> histogram(gs[,1])
> table(gs[,1])/100000
      0       3       4       5       6       7       8
0.00001 0.00005 0.00020 0.00077 0.00267 0.00705 0.01539
      9      10      11      12      13      14      15
0.03004 0.05111 0.08189 0.10967 0.12971 0.13773 0.13273
     16      17      18      19      20      21      22
0.10851 0.08238 0.05246 0.03100 0.01584 0.00686 0.00264
     23      24      25
0.00094 0.00029 0.00006
> mean(gs[,1])
[1] 14.0159
```

Note also the general appearance of the joint distribution, and the very good mixing properties of the Gibbs sampler in this case.

The joint distribution can also be computed exactly via, for example, the following **R** code

```
> tp = dpois(0:30,20)                                  # Poisson probs
> pp = tp%o%tp                                         # outer product
> for(i in 1:31) for(j in 1:31) if(i+j>32) pp[i,j]=0  # truncate
> pp = pp/sum(pp)                                      # normalise
> ppmarg = apply(pp,1,'sum')                           # marginal probs
> sum(ppmarg)                                          # check
[1] 1
> sum(0:30*ppmarg)                                     # calculate mean
[1] 14.01387
> pp[1,1]
[1] 6.886154e-17                                       # calculate a
```

(b) The conditional distribution of n_2 given n_1 is now $\text{Pois}(\lambda_2 \alpha^{n_1})$, while that of n_1 given n_2 is $\text{Pois}(\lambda_1 \alpha^{n_2})$.

The Gibbs sampler of the first part of the question is easily modified (see the **R** function gsmp2 at http://www.ma.hw.ac.uk/~stan/f73bi/R/). However, in this case its mixing properties are *extremely* poor. This is closely related to the pronounced bimodality of the marginal distributions of n_1 and n_2.

6. This is again a *practical exercise*!