Stat 461-561: Exercises 6

Exercise 2 *Changepoint detection.*
(a) Derive the Gibbs sampler to sample from the posterior distribution

$$\pi\left(k,\mu,\lambda,b_1,b_2\right|x_{1:m}).$$

We have

$$\begin{aligned}&\pi\left(k,\mu,\lambda,b_1,b_2\right|x_{1:m})\\ \propto\ &\prod_{i=1}^{k}\exp\left(-\mu\right)\frac{\mu^{x_i}}{x_i!}\prod_{i=k+1}^{m}\exp\left(-\lambda\right)\frac{\lambda^{x_i}}{x_i!}\\ &\times\frac{1}{b_1^{a_1}}\mu^{a_1-1}\exp\left(-\mu/b_1\right)\times\frac{1}{b_2^{a_2}}\lambda^{a_2-1}\exp\left(-\lambda/b_2\right)\\ &\times b_1^{-c_1-1}\exp\left(-1/\left(d_1b_1\right)\right)\times b_2^{-c_2-1}\exp\left(-1/\left(d_2b_2\right)\right)\end{aligned}$$

and we use the convention $\prod_{i=m+1}^{m}\exp\left(-\lambda\right)\frac{\lambda^{x_i}}{x_i!}=1.$

We have

$$\begin{aligned}&\pi\left(\mu\right|x_{1:m},k,\lambda,b_1,b_2)=\pi\left(\mu\right|x_{1:k},b_1)\\ &\propto\mu^{a_1-1}\exp\left(-\mu/b_1\right)\prod_{i=1}^{k}\exp\left(-\mu\right)\tfrac{\mu^{x_i}}{x_i!}\\ &\propto\mu^{a_1+\sum_{i=1}^{k}x_i-1}\exp\left(-\mu/b_1-k\mu\right)\\ &=\mathit{Gamma}\left(\mu;a_1+\textstyle\sum_{i=1}^{k}x_i,\left(1/b_1+k\right)^{-1}\right).\end{aligned}$$

Similarly, we have for $k<m$

$$\begin{aligned}&\pi\left(\lambda\right|x_{1:m},k,\mu,b_1,b_2)=\pi\left(\lambda\right|x_{k+1:m},b_2)\\ &=\mathit{Gamma}\left(\lambda;a_2+\textstyle\sum_{i=k+1}^{m}x_i,\left(1/b_2+m-k\right)^{-1}\right)\end{aligned}$$

and if $k=m$

$$\begin{aligned}\pi\left(\lambda\right|x_{1:m},k,\mu,b_1,b_2)&=\pi\left(\lambda\right|b_2)\\ &=\mathit{Gamma}\left(\lambda;a_2,b_2\right).\end{aligned}$$

We have

$$\begin{aligned}\pi\left(b_1\right|x_{1:m},k,\lambda,\mu,b_2)&=\pi\left(b_1\right|\mu)\\ &\propto\frac{1}{b_1^{a_1}}\exp\left(-\mu/b_1\right)b_1^{-c_1-1}\exp\left(-1/\left(b_1d_1\right)\right)\\ &=\mathit{InvGamma}(b_1;a_1+c_1,\left(\mu+d_1^{-1}\right)^{-1}),\\ \pi\left(b_2\right|x_{1:m},k,\lambda,\mu,b_1)&=\pi\left(b_2\right|\lambda)\\ &=\mathit{InvGamma}(b_2;a_2+c_2,\left(\lambda+d_2^{-1}\right)^{-1}).\end{aligned}$$

We also have for $k=1,...,m$

$$\pi\left(k|\, x_{1:m},\lambda,\mu,b_1,b_2\right)=\frac{f\left(x_{1:m}|\, k,\mu,\lambda,b_1,b_2\right)}{\sum_{j=1}^{m} f\left(x_{1:m}|\, k=j,\mu,\lambda,b_1,b_2\right)}.$$

(b) Is it not possible to select an improper prior for λ as

$$\begin{aligned}\pi\left(\lambda|\, x_{1:m}\right) &= \sum_{i=1}^{m}\pi\left(\lambda,k=i|\, x_{1:m}\right)\\ &= \sum_{i=1}^{m}\pi\left(\lambda|\, x_{1:m},k=i\right)\pi\left(k=i|\, x_{1:m}\right).\end{aligned}$$

But if there is no changepoint (i.e. $k=m$) then $\pi\left(\lambda|\, x_{1:m},k=i\right)=\pi\left(\lambda\right)$ which is improper.

Exercise 3. This was covered during the course (Baseball data).

Exercise 4. *Brook's formula.* Consider the following conditional distributions $\pi\left(\theta_2|\,\theta_1\right)$ and $\pi\left(\theta_1|\,\theta_2\right)$ which are assumed strictly positive for any θ_1,θ_2.

(a) We have

$$\begin{aligned}&\pi\left(\theta_1^0,\theta_2^0\right)\frac{\pi\left(\theta_1|\,\theta_2\right)}{\pi\left(\theta_1^0\middle|\,\theta_2\right)}\frac{\pi\left(\theta_2\middle|\,\theta_1^0\right)}{\pi\left(\theta_2^0\middle|\,\theta_1^0\right)}\\
&= \pi\left(\theta_2^0\middle|\,\theta_1^0\right)\pi\left(\theta_1^0\right)\frac{\pi\left(\theta_1|\,\theta_2\right)}{\pi\left(\theta_1^0\middle|\,\theta_2\right)}\frac{\pi\left(\theta_2\middle|\,\theta_1^0\right)}{\pi\left(\theta_2^0\middle|\,\theta_1^0\right)}\\
&= \pi\left(\theta_1^0\right)\frac{\pi\left(\theta_1|\,\theta_2\right)\pi\left(\theta_2\middle|\,\theta_1^0\right)}{\pi\left(\theta_1^0\middle|\,\theta_2\right)}\\
&= \frac{\pi\left(\theta_1|\,\theta_2\right)\pi\left(\theta_1^0,\theta_2\right)}{\pi\left(\theta_1^0\middle|\,\theta_2\right)}\\
&= \frac{\pi\left(\theta_1|\,\theta_2\right)\pi\left(\theta_1^0\middle|\,\theta_2\right)\pi\left(\theta_2\right)}{\pi\left(\theta_1^0\middle|\,\theta_2\right)}\\
&= \pi\left(\theta_1,\theta_2\right)\end{aligned}$$

(b) Given $\pi\left(\theta_2|\,\theta_1\right)=\mathcal{N}\left(\theta_2;0,\frac{1}{1+\theta_1^2}\right)$ and $\pi\left(\theta_1|\,\theta_2\right)=\mathcal{N}\left(\theta_1;0,\frac{1}{1+\theta_2^2}\right)$, establish the expression of $\pi\left(\theta_1,\theta_2\right)$ up to a normalizing constant.

We select $\left(\theta_1^0,\theta_2^0\right)=(0,0)$ and

$$\begin{aligned}\pi\left(\theta_1,\theta_2\right) &\propto \frac{\pi\left(\theta_1|\,\theta_2\right)}{\pi\left(\theta_1^0\middle|\,\theta_2\right)}\pi\left(\theta_2\middle|\,\theta_1^0\right)\\
&\propto \frac{\sqrt{1+\theta_2^2}\exp\left(-\frac{\left(1+\theta_2^2\right)\theta_1^2}{2}\right)}{\sqrt{1+\theta_2^2}}\times\sqrt{1+\left(\theta_1^0\right)^2}\exp\left(-\frac{\left(1+\left(\theta_1^0\right)^2\right)\theta_2^2}{2}\right)\\
&\propto \exp\left(-\frac{\left(1+\theta_2^2\right)\theta_1^2+\theta_2^2}{2}\right)\\
&\propto \exp\left(-\frac{\theta_1^2+\theta_1^2\theta_2^2+\theta_2^2}{2}\right).\end{aligned}$$

(c) Are the marginal distributions $\pi(\theta_1)$ and $\pi(\theta_2)$ Gaussians?

The marginals are clearly not Gaussian due to the term $\theta_1^2\theta_2^2$.

Exercise 5. *Two-stage Gibbs sampler.* In the case of the two-stage Gibbs sampler, the relationship between the Gibbs sampler and the Metropolis-Hastings algorithm becomes particularly clear. If we have the bivariate Gibbs sampler $X \sim \pi(x|\, y)$ and $Y \sim \pi(y|\, x)$, consider the X chain alone and show:

(a) its transition kernel is given by $K(x, x') = \int \pi(y|\, x)\, \pi(x'|\, y)\, dy$.

The joint distribution of (y, x') given x is

$$\pi(y|\, x)\, \pi\left(x'\middle|\, y\right).$$

So integrating out the variable y, we have

$$K\left(x, x'\right) = \int \pi(y|\, x)\, \pi\left(x'\middle|\, y\right) dy.$$

(b) show that $\frac{\pi(x')K(x',x)}{\pi(x)K(x,x')} = 1$ so a Metropolis-Hastings with proposal $K(x', x)$ is always accepted.

We have

$$\begin{aligned}
& \pi(x)\, K\left(x, x'\right) \\
= \;& \pi(x) \int \pi(y|\, x)\, \pi\left(x'\middle|\, y\right) dy \\
= \;& \pi(x) \int \frac{\pi(x|\, y)\, \pi(y)}{\pi(x)} \frac{\pi(y|\, x')\, \pi(x')}{\pi(y)} dy \\
= \;& \pi\left(x'\right) \int \pi(x|\, y)\, \pi\left(y\middle|\, x'\right) dy \\
= \;& \pi\left(x'\right) K\left(x', x\right).
\end{aligned}$$

Exercise 6. *Binomial model.* For $i = 1, 2, 3$, consider $Y_i = X_{1i} + X_{2i}$ with

$$X_{1i} \sim \mathcal{B}(n_{1i}, \theta_1),\; X_{2i} \sim \mathcal{B}(n_{2i}, \theta_2).$$

(a) Give the likelihood for $n_{1i} = 5, 6, 3$, $n_{2i} = 5, 4, 6$ and $y_i = 7, 5, 6$.

We can compute the distribution of the observations as follows

$$\prod_{i=1}^{3} \left[\sum_{j_i} \binom{n_{1i}}{j_i} \binom{n_{2i}}{y_i - j_i} \theta_1^{j_i} (1-\theta_1)^{n_{1i}-j_i} \theta_2^{y_i - j_i} (1-\theta_2)^{n_{2i}-y_i+j_i} \right]$$

where $\max\{0, y_i - n_{2i}\} \le j_i \le \min\{n_{1i}, y_i\}$.

(b) For a uniform prior on (θ_1, θ_2), derive the Gibbs sampler.

To implement the Gibbs sampler, we introduce missing data and sample from

$$\pi(\theta_1, \theta_2, x_{1,1:3}, x_{2,1:3}|\, y_{1:3})$$

where $x_{i,1:3} = (x_{i1}, x_{i2}, x_{i3})$. We have

$$\begin{aligned}
\pi(\theta_1, \theta_2|\, y_{1:3}, x_{1,1:3}, x_{2,1:3}) &= \pi(\theta_1, \theta_2|\, x_{1,1:3}, x_{2,1:3}) \\
&= \pi(\theta_1|\, x_{1,1:3})\, \pi(\theta_2|\, x_{2,1:3})
\end{aligned}$$

where

$$\begin{aligned}\pi\left(\theta_i \mid x_{i,1:3}\right) &\propto \prod_{i=1}^{3} \theta_1^{x_{1i}}\left(1-\theta_1\right)^{n_{i1}-x_{1i}} \\ &\propto \theta_1^{\sum_{i=1}^{3} x_{1i}}\left(1-\theta_1\right)^{3n_{i1}-\sum_{i=1}^{3} x_{1i}} \\ &= Beta\left(\theta_1; 1+\sum_{i=1}^{3} x_{1i}, 1+3n_{i1}-\sum_{i=1}^{3} x_{1i}\right).\end{aligned}$$

Now we have

$$\begin{aligned}&\pi\left(x_{1,1:3}, x_{2,1:3} \mid y_{1:3}, \theta_1, \theta_2\right) \\ &= \prod_{i=1}^{3} \pi\left(x_{1,i}, x_{2,i} \mid y_i, \theta_1, \theta_2\right)\end{aligned}$$

and

$$\pi\left(x_{1,i}, x_{2,i} \mid y_i, \theta_1, \theta_2\right) = \mathcal{M}\left(y_i, \frac{n_{i1}\theta_1}{n_{i1}\theta_1+n_{i2}\theta_2}, \frac{n_{i2}\theta_1}{n_{i1}\theta_1+n_{i2}\theta_2}\right)$$

or equivalently

$$\pi\left(x_{1,i} \mid y_i, \theta_1, \theta_2\right) = \mathcal{B}\text{in}\left(y_i, \frac{n_{i1}\theta_1}{n_{i1}\theta_1+n_{i2}\theta_2}\right)$$

(c) Examine whether an alternative parameterization or a MH algorithm may speed up convergence.

There is no clear cut answer to this question. It might be interesting to use the logit (log-odds) reparameterization, that is

$$\mu_i = \log\left(\frac{\theta_i}{1-\theta_i}\right).$$

Also it is not necessary to introduce missing data and it could be possible alternatively to use the MH algorithm to sample directly from $\pi\left(\theta_1, \theta_2 \mid y_{1:3}\right)$.

Exercise 7. *More two-stage Gibbs sampler.* For the Gibbs sampler

$$X \mid y \sim \mathcal{N}\left(\rho y, 1-\rho^2\right), \quad Y \mid x \sim \mathcal{N}\left(\rho x, 1-\rho^2\right).$$

(a) Show that for the X chain, the transition kernel is

$$K\left(x', x\right) = \frac{1}{2\pi\left(1-\rho^2\right)} \int \exp\left(-\frac{\left(x-\rho y\right)^2}{2\left(1-\rho^2\right)}\right) \exp\left(-\frac{\left(y-\rho x'\right)^2}{2\left(1-\rho^2\right)}\right) dy.$$

This follows straightforwardly from the fact that

$$K\left(x', x\right) = \int \pi\left(y \mid x'\right) \pi\left(x \mid y\right) dy.$$

(b) Show that $\mathcal{N}(0,1)$ is the invariant distribution of the X chain.

Assume that $X' \sim \mathcal{N}(0,1)$ then as

$$Y|\, x' \sim \mathcal{N}\left(\rho x', 1-\rho^2\right)$$

we have

$$\begin{aligned}\mathbb{E}\left[Y|\, x'\right] &= \rho x', \\ \mathbb{V}\left[Y|\, x'\right] &= 1-\rho^2\end{aligned}$$

so

$$\begin{aligned}\mathbb{E}[Y] &= \mathbb{E}\left[\mathbb{E}\left[Y|\, X'\right]\right] = \rho\mathbb{E}\left[X'\right] = 0, \\ \mathbb{V}[Y] &= \mathbb{E}\left[\mathbb{V}\left[Y|\, X'\right]\right] + \mathbb{V}\left[\mathbb{E}\left[Y|\, X'\right]\right] \\ &= 1-\rho^2+\rho^2\mathbb{V}\left[X'\right] \\ &= 1.\end{aligned}$$

Similarly we show that $X|\, y \sim \mathcal{N}\left(\rho y, 1-\rho^2\right)$ is such that $\mathbb{E}[X]=0$ and $\mathbb{V}[X]=1$. As X is Gaussian then this proves that $\mathcal{N}(0,1)$ is the invariant distribution of $K(x', x)$. Alternatively, you could just check it by showing that

$$\int \mathcal{N}\left(x';0,1\right) K\left(x',x\right) dx = \mathcal{N}\left(x';0,1\right).$$

(c) It follows directly by completing the square in y and integrating with respect to y.

(d) We have $X_k = \rho^2 X_{k-1} + U_k$ $(k=1,2,...)$ where $U_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0,1-\rho^4\right)$ directly because of (c). Now we have

$$\begin{aligned}\mathbb{E}[X_1] &= \mathbb{E}\left[\rho^2 X_0 + U_1\right] \\ &= \rho^2\mathbb{E}[X_0] + \mathbb{E}[U_1] \\ &= 0\end{aligned}$$

if $\mathbb{E}[X_0]=0$ and by induction $\mathbb{E}[X_k]=0$.

We have

$$\begin{aligned}cov(X_0,X_1) &= \mathbb{E}[X_0X_1] - \mathbb{E}[X_0]\,\mathbb{E}[X_1] \\ &= \rho^2\mathbb{E}\left[X_0^2\right] = \rho^2.\end{aligned}$$

Assume we have $cov(X_0,X_k)=\rho^{2k}$ then

$$\begin{aligned}cov(X_0,X_{k+1}) &= \mathbb{E}[X_0X_{k+1}] \\ &= \rho^2\mathbb{E}[X_0X_k] + \mathbb{E}[X_0U_k] \\ &= \rho^2\mathbb{E}[X_0X_k] + \mathbb{E}[X_0]\,\mathbb{E}[U_k] \\ &= \rho^{2(k+1)}.\end{aligned}$$