@p

Btry 694 Class Homework 6

11/6/98

1 Problem 6.6

Dennis Carlson

Consider a version of $[A.25]$ (the independent M-H algorithm), based on a "bound" M on $\frac{f}{g}$ that is not a uniform bound, that is, $\frac{f(x)}{g(x)} > M$ for some x.

(a) If an accept-reject algorithm uses the density g with acceptance probability $\frac{f(y)}{Mg(y)}$, show that the resulting variables are generated from

$$\tilde{f}(x) \propto \min\{f(x), Mg(x)\},$$

instead of f.

(b) She that this error can be corrected, for instance by using the Metropolis-Hastings algorithm:

1. Generate $Y_t \sim \tilde{f}$.

2. Accept with probability

$$P(X^{(t+1)} = y_t | x^{(t)}, y_t) = \begin{cases} \min\left\{1, \frac{f(y_t)g(x^{(t)})}{g(y_t)f(x^{(t)})}\right\} & \text{if } f(y_t) > Mg(y_t) \\ \min\left\{1, \frac{Mg(x^{(t)})}{f(x^{(t)})}\right\} & \text{otherwise} \end{cases}$$

Solutions:

(a) We essentially have two cases in the problem: $f(x) \leq MG(x)$, and $f(x) > Mg(x)$. In the first case, it follows from the proof of Lemma 2.3.1 that the Accept-Reject algorithm produces a variable distributed according to f. In the second case, since $\frac{f(X)}{Mg(X)} > 1$, and $U \sim \mathcal{U}_{[\prime,\infty]}$, it follows that

$$P(Y \le y) = P\left(X \le y | U \le \frac{f(X)}{Mg(X)}\right) = P\left(X \le y | U \le \frac{Mg(X)}{Mg(X)}\right).$$

Following the rest of the proof of Lemma 2.3.1, it follows that the A-R algorithm produces a variable distributed according to Mg. Hence, the variables are generated from $\tilde{f}(x) \propto \min\{f(x), Mg(x)\}$, where the proportionality will depend on $P(f(X) \le Mg(X))$.

(b). Again looking at the seperate cases, if $f(y_t) > Mg(y_t)$, then

$$\min\left\{1, \frac{f(y_t)g(x^{(t)})}{g(y_t)f(x^{(t)})}\right\} = \min\left\{1, \frac{f(y_t)Mg(x^{(t)})}{Mg(y_t)f(x^{(t)})}\right\} = \min\left\{1, \frac{f(y_t)\tilde{f}(x^{(t)})}{\tilde{f}(y_t)f(x^{(t)})}\right\}.$$

That is, in this case, the M-H agorithm generates f according to $[A.25]$.

Now, suppose $f(y_t) \le Mg(y_t)$. Then we have

$$\min\left\{1, \frac{Mg(x^{(t)})}{f(x^{(t)})}\right\} = \min\left\{1, \frac{f(y_t)Mg(x^{(t)})}{f(y_t)f(x^{(t)})}\right\} = \min\left\{1, \frac{f(y_t)\tilde{f}(x^{(t)})}{\tilde{f}(y_t)f(x^{(t)})}\right\}.$$

Again, in this case, the M-H agorithm generates f according to $[A.25]$.

2 Problem 6.31

Mitchell Warachka

part a

Will express the kernel $K(x,y) = p(x)q(y) + (1 - r(x))\delta_x(y)$ for independent Metropolis-Hastings Algorithm in matrix form with diagonol G and another matrix $\pi^T q$ where q is proposal distribution. The matrix G is to account for the delta function, when $x = y$.

$$\begin{aligned}
& r(x) = p(x) \\
= \;& p(x)q(x) + 1 - p(x) \text{ results from kernel} \\
& \lambda_x = \sum x \neq y \left(p_i + \frac{\pi_i}{w_x}\right) \text{ given in question, where } w_x = \frac{\pi_x}{p_x} \\
= \;& 1 - p(x) + p(x)q(x)
\end{aligned} \tag{1}$$

while for $x \neq y$, the first matrix, G can be ignored and normal transition probabilities are considered as product of two probability vectors which form an m by m matrix.

part b

Since G is diagonol, the diagonol elements are the eigenvalues and the diagonol elements of G are the $\lambda_k' s$.

part c

Using the total variation norm,

$$||K^n(x,\cdot) - \pi(\cdot)|| = \frac{1}{2}\sum_y |K^n(x,y) - \pi(y)| \tag{2}$$

an upper bound on the total variation can be computed where λ_1 is assumed to be the largest eigenvalue

$$\begin{aligned}
\left(\sum_y |K^n(x,y) - \pi(y)|\right)^2 &= \left(\sum_y \frac{|K^n(x,y) - \pi(y)|}{\sqrt{\pi(y)}}\sqrt{\pi(y)}\right)^2 \\
&\leq \sum_y \frac{|K^n(x,y) - \pi(y)|^2}{\pi(y)} \text{ by Cauchy - Schwartz} \\
&\leq \sum_y \frac{K^{2n}(x,y)}{\pi(y)} - 1 \\
&\leq \frac{K^{2n}(x,x)}{\pi(x)} \\
&\leq \frac{\lambda_1^{2n}}{\pi(x)}
\end{aligned}$$

3 Problem 6.37

Bárbara Patricia González

Show that

$$\pi(\theta_1, \theta_2 | y) = f(y) \exp\left(-\frac{1}{2}\left(\theta_1^2 \frac{5}{4} + \theta_2^2 \frac{5}{4} + \frac{\theta_1\theta_2}{2} - (\theta_1 + \theta_2)y\right)\right)$$

can be directly simulated as

$$\theta_2 \sim N\left(\frac{y}{3}, \frac{5}{6}\right), \theta_1|\theta_2 \sim N\left(\frac{2y-\theta_2}{5}, \frac{4}{5}\right).$$

3.1 Solution

We can write

$$\pi(\theta_1, \theta_2|y) = f_{\Theta_1|\Theta_2,Y}(\theta_1|\theta_2, y) f_{\Theta_2|Y}(\theta_2|y) =$$

$$= \frac{1}{\sqrt{2\pi\frac{4}{5}}} \exp\left(-\frac{5}{8}\left(\theta_1 - \left(\frac{2y-\theta_2}{5}\right)\right)^2\right) \frac{1}{\sqrt{2\pi\frac{5}{6}}} \exp\left(-\frac{3}{5}\left(\theta_2 - \frac{y}{3}\right)^2\right) =$$

$$= \frac{1}{2\pi\sqrt{\frac{2}{3}}} \exp\left(-\frac{5}{8}\left(\theta_1 - \left(\frac{2y-\theta_2}{5}\right)\right)^2 - \frac{3}{5}\left(\theta_2 - \frac{y}{3}\right)^2\right) =$$

$$= \frac{1}{2\pi\sqrt{\frac{2}{3}}} \exp\left(-\frac{5}{8}\left(\theta_1^2 - 2\theta_1\left(\frac{2y-\theta_2}{5}\right) + \left(\frac{2y-\theta_2}{5}\right)^2\right) -\right.$$

$$\left. -\frac{3}{5}\left(\theta_2^2 - \frac{2\theta_2 y}{3} + \frac{y^2}{3}\right)\right) =$$

$$= \frac{1}{2\pi\sqrt{\frac{2}{3}}} \exp\left(-\frac{3y^2}{10}\right) \exp\left(-\frac{1}{2}\left(\theta_1^2\frac{5}{4} + \theta_2^2\frac{5}{4} + \frac{\theta_1\theta_2}{2} - (\theta_1+\theta_2)y\right)\right) =$$

$$= f(y) \exp\left(-\frac{1}{2}\left(\theta_1^2\frac{5}{4} + \theta_2^2\frac{5}{4} + \frac{\theta_1\theta_2}{2} - (\theta_1+\theta_2)y\right)\right).$$

4 Problem 6.44

Krishanu Maulik

[Peskun 1973] For $\mathbb{P}_1$ and $\mathbb{P}_2$, irreducible, reversible stochastic matrices with the same stationary distribution π on $\{1,\ldots,m\}$, show that, if $\mathbb{P}_1 \le \mathbb{P}_2$ (meaning that the off-diagonal elements are smaller in the first case), for every function h,

$$\lim_{N\to\infty} \mathrm{v}ar\left[\sum h(X_1^{(t)})\right]/N \ge \lim_{N\to\infty} \mathrm{v}ar\left[\sum h(X_1^{(t)})\right]/N,$$

where $(X_i^{(t)})$ is a Markov chain with transition matrix $\mathbb{P}_i$ $(i=1,2)$.

Solution

Since we have a finite state space and the transition matrix is irreducible, the matrix is regular too. Thus using the result in Kemeny and Snell (1968) pp. 84-85, as given in the problem 4.52, we have, for $i=1,2$,

$$\lim_{N\to\infty} \mathrm{v}ar\left[\sum h(X_i^{(t)})\right]/N = h^T C_i h,$$

where

$$C_i = DZ_i + (DZ_i)^T - D - \pi\pi^T \tag{3}$$

with

$$D = \mathrm{d}iag\{\pi_1,\cdots,\pi_m\},$$

$$Z_i = (\mathbf{I} - (\mathbb{P}_i - A))^{-1}, \text{ the fundamental matrix}$$

and

$$A = \mathbf{1}\pi^T,$$

where 1 is the m-vector of 1's.

Let $\mathbb{P}$ be an irreducible, reversible stochastic matrix with the stationary distribution π and $(X^{(t)})$ be a Markov chain with the transition matrix $\mathbb{P}$. We define C and Z analougously.

Then D is invertible and is $\mathrm{d}iag\{\pi_1^{-1},\cdots,\pi_m^{-1}\}$. Also $DZ = (D^{-1} - (\mathbb{P}D^{-1} - AD^{-1}))^{-1}$. Thus DZ is symmetric if and only if $\mathbb{P}D^{-1} - AD^{-1}$ is so. But as $A = \mathbf{1}\pi^T$, we have $a_{ij} = \pi_j$. Hence, for $i \ne j$,

$$(\mathbb{P}D^{-1} - AD^{-1})_{ij} = p_{ij}\pi_j^{-1} - a_{ij}\pi_j^{-1} = p_{ij}/\pi_j - 1$$

and, similarly,

$$(\mathbb{P}D^{-1} - AD^{-1})_{ji} = p_{ji}/\pi_i - 1.$$

But, by reversibility, $p_{ij}/\pi_j = p_{ji}/\pi_i$. Hence $\mathbb{P}D^{-1} - AD^{-1}$ and hence DZ is symmetric. So, from equation (1),

$$C = 2DZ - D - \pi\pi^T.$$

Now, for $k \neq l$, we have $\frac{dC}{dp_{kl}} = 2D\frac{dZ}{dp_{kl}}$. But,

$$\frac{dZ}{dp_{kl}} = -Z\frac{dZ^{-1}}{dp_{kl}}Z.$$

$$\therefore \frac{dh^T Ch}{dp_{kl}} = -2h^T DZ\frac{dZ^{-1}}{dp_{kl}}Zh = -2(Zh)^T D^T\frac{dZ^{-1}}{dp_{kl}}(Zh) = -2(Zh)^T D\frac{dZ^{-1}}{dp_{kl}}(Zh),$$

since D is symmetric. Also $\mathbb{P}$ is a reversible transition matrix, and hence satisfies the conditions

$$p_{ii} = -\sum_{j\neq i} p_{ij},$$

and

$$p_{ji} = p_{ij}\pi_i/\pi_j.$$

So $D\frac{dZ^{-1}}{dp_{kl}}(Zh)$ has all entries zero, except for $(k,k),(k,l),(l,k),(l,l)$th entries which are $\pi_l, -\pi_l, -\pi_k, \pi_k$ and hence is non-negative definite. Thus,

$$\frac{dh^T Ch}{dp_{kl}} \leq 0.$$

This implies as the off-diagonal entries are reduced, $h^T Ch$ is increased. So in our problem, as $\mathbb{P}_1 \leq \mathbb{P}_2$, we have,

$$\lim_{N\to\infty} \mathrm{var}\left[\sum h(X_1^{(t)})\right]/N \geq \lim_{N\to\infty} \mathrm{var}\left[\sum h(X_1^{(t)})\right]/N.$$

5 Problem 6.45

Krishanu Maulik

Dedude from Problem 6.44 that, for a given instrumental matrix Q in a Metropolis-Hastings algorithm, the choice

$$p_{ij} = q_{ij}\left(\frac{\pi_j q_{ji}}{\pi_i q_{ij}} \wedge 1\right)$$

is optimal among the transitions such that

$$p_{ij} = \frac{q_{ij}s_{ij}}{1+\frac{\pi_i q_{ij}}{\pi_j q_{ji}}} = q_{ij}\alpha_{ij},$$

where $0 \le \alpha_{ij} \le 1$ **and** s_{ij} **is a symmetric function of** i **and** j**.**

Solution

Let

$$t_{ij} = \frac{\pi_i q_{ij}}{\pi_j q_{ji}}.$$

Since $0 \le \alpha_{ij} \le 1,$ **we have** $s_{ij} \le 1 + t_{ij}$**. Also since** s_{ij} **is a symmetric function of** i **and** j**,** $s_{ij} = s_{ji} \le 1 + t_{ji}$**. Combining,** $s_{ij} \le 1 + (t_{ij} \wedge t_{ji})$**. Thus,**

$$p_{ij} \le q_{ij}\frac{1 + (t_{ij} \wedge t_{ji})}{1+t_{ij}} = q_{ij}\left(1 \wedge \frac{1+tji}{1+t_{ij}}\right) = p^*_{ij},$$

where $\mathbb{P}^*$ **is the transition matrix of Metropolis-Hastings algorithm.**

So for any other transition kernel $\mathbb{P}$**, we have** $\mathbb{P} \le \mathbb{P}^*$**, and hence the asymptotic variance will be more than that of Metropolis-Hastings and Thus it is the optimal one.**

References

Kemeny, J.G. and Snell, J.L. (1968) *Finite Markov Chains.* **Springer-Verlag, New York.**

Peskun, P.H. (1973) Optimum Monte-Carlo sampling using Markov chains. *Biometrika* **60, pp. 607-612.**

6 Problem 6.46

Bárbara Patricia González

Show that f **is the stationary density associated with the acceptance probability**

$$\rho(x,y) = \frac{s(x,y)}{1+\frac{f(x)q(y|x)}{f(y)q(x|y)}},$$

where s **is an arbitrary positive symmetric function such that** $\rho(x,y) \le 1$**.**

6.1 Solution

It is enough to prove that $K(x,y)f(x) = K(y,x)f(y)$, where

$$K = \rho(x,y)q(y|x) + (1 - r(x))\delta_x(y)$$

is the kernel of the Markov chain produced by the Metropolis-Hastings algorithm. So we have that

$$K(x,y)f(x) = \frac{s(x,y)}{1 + \frac{f(x)q(y|x)}{f(y)q(x|y)}} q(y|x)f(x) + (1 - r(x))\delta_x(y).$$

Since $(1 - r(x))\delta_x(y) \neq 0$ only when $x = y$ we have that

$$K(x,y)f(x) =$$

$$= \frac{s(x,y)q(x|y)f(x)}{f(y)q(x|y) + f(x)q(y|x)} f(y)q(x|y) + (1 - r(y))\delta_y(x) =$$

$$= \frac{s(x,y)}{\frac{f(y)q(x|y)}{f(x)q(y|x)} + 1} f(y)q(x|y) + (1 - r(y))\delta_y(x).$$

Since $s(x,y) = s(y,x)$ we have that

$$K(x,y)f(x) = \frac{s(y,x)}{\frac{f(y)q(x|y)}{f(x)q(y|x)} + 1} f(y)q(x|y) + (1 - r(y))\delta_y(x).$$

So finally we have that

$$K(x,y)f(x) = K(y,x)f(y).$$