Understanding MCMC: Exercise Solutions

1. (a) We compute that $\sum_{x\in\mathcal{X}}(1/3)\,P(x,\{y\}) = 1/3$ for all $y \in \mathcal{X}$.

   (b) It is not reversible since e.g. $\pi\{1\}P(1,\{2\}) = (1/3)(3/4) \neq (1/3)(1/4) = \pi\{2\}P(2,\{1\})$.

2. (a) Let $h(t) = \frac{1}{\sqrt{2\pi}}e^{-t^2/2}$ be the standard normal density. Then if X_n has density $h(x)$, then X_{n+1} has density given by $\int_{\mathbf{R}} h(t)\,h\big((x-t/2)/\sqrt{3/4}\big)\,dt$ which we compute is equal to $h(x)$ for all $x \in \mathbf{R}$, so that X_{n+1} also has density $h(x)$.

   (b) If X_n and Z_{n+1} are i.i.d. standard normal, then $X_n/2 + \sqrt{3/4}\,Z_{n+1}$ is also standard normal.

3. We compute the function $q(x,y)$ as follows. Let $g(z) = xe^z$ (for fixed x). Then $Y_{n+1} = g(Z_{n+1})$, where $Z_{n+1} \sim N(0,\sigma^2)$ with density f_Z (say). Now, $g'(z) = xe^z$ and $g^{-1}(y) = \log(y/x)$, so $g'(g^{-1}(y)) = xe^{\log(y/x)} = x(y/x) = y$. Hence, by the change-of-variable formula, the density of Y_{n+1} is given by $f_Y(y) = f_Z(g^{-1}(y))\,|g'(g^{-1}(y)| = f_Z(\log(y/x))\,y$. We conclude that $q(x,y) = f_Z(\log(y/x))\,y$.

   Now, if $C(x,y) = f_Z(\log(y/x))$, then $C(x,y) = f_Z(\log y - \log x)$, so $C(x,y) = C(y,x)$. Hence,

$$\alpha(x,y) = \min\left[1,\ \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}\right] = \min\left[1,\ \frac{\pi(y)C(y,x)\,x}{\pi(x)C(x,y)\,y}\right] = \min\left[1,\ \frac{x\,\pi(y)}{y\pi(x)}\right].$$

4. (a) Given X_n, propose $Y_{n+1} \sim \text{Uniform}[X_n - 1,\ X_n + 1]$, then accept (and set $X_{n+1} = Y_{n+1}$) with probability $\min[1,\ \pi(Y_{n+1})/\pi(X_n)]$, otherwise reject (and set $X_{n+1} = X_n$).

   (b) Let λ be Lebesgue measure on $\mathbf{R}$. Then if $\lambda(A) > 0$, we can find $r \in \mathbf{R}$ with $\lambda(A \cap [r, r+1]) > 0$. Then from $X_0 = x$, we have positive probability of being inside $[r, r+1]$ after $\geq |x - r| + 1$ iterations. From there, we have positive probability of entering A on the next iteration. Hence, the chain is λ-irreducible.

   (c) Assume to the contrary that the chain has periodic decomposition $\mathcal{X} = \mathcal{X}_1 \cup \ldots \cup \mathcal{X}_d$ for some $d \geq 2$. Find $r \in \mathbf{R}$ and $A \subseteq \mathcal{X}_1 \cap [r, r+1]$ with $\lambda(A) > 0$. Then for $x \in [r, r+1]$, we have $P(x, A) > 0$, contradicting the fact that $P(x, \mathcal{X}_1) = 0$ for all $x \in \mathcal{X}_1$.

(d) We conclude that $\lim_{n\to\infty} \|P^n(x,\cdot) - \pi(\cdot)\| = 0$ for π-a.e. $x \in \mathcal{X}$.

5. (a) Given x_2, the 1-component update (P_1) replaces x_1 by a draw from the density on $[0,1]$ given by $f(x_1) = \pi((x_1,x_2)) \Big/ \int_0^1 \pi((x_1,z))\,dz$. Similarly, the 2-component update (P_2) replaces x_2 by a draw from the density on $[0,1]$ given by $h(x_2) = \pi((x_1,x_2)) \Big/ \int_0^1 \pi((z,x_2))\,dz$. The deterministic-scan Gibbs sampler then alternately applies P_1 and P_2, while the random-scan Gibbs sampler repeated chooses one of P_1 and P_2 uniformly at random.

(b) Let λ be Lebesgue measure on $\mathcal{X} = [0,1] \times [0,1]$. Then if $\lambda(A) > 0$, then (since $\pi(\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathcal{X}$) the chain can reach A with positive probability in one step of deterministic scan, or two steps of random scan. Hence, the chain is λ-irreducible.

(c) Random-scan Gibbs sampler is always aperiodic (since it might repeat the same update twice). For deterministic-scan, if $\pi(A) > 0$, then the chain has positive probability of reaching A in one iteration from anywhere, so it cannot be periodic.

(d) The deterministic-scan Gibbs sampler has transitions which are absolutely continuous (i.e. have density), so it must be Harris recurrent. For random-scan the chain is absolutely continuous as soon as it has updated *both* components at least once, which must happen eventually with probability 1.

(e) We conclude that $\lim_{n\to\infty} \|P^n(x,\cdot) - \pi(\cdot)\| = 0$ for <u>all</u> $x \in \mathcal{X}$.

6. Let $\rho(A) = \epsilon^{-1} \int_A \big(\inf_{x\in C} p(x,y)\big)\nu(dy)$, for $A \subseteq \mathcal{X}$. Then we claim that $P(x,\cdot) \geq \epsilon\,\rho(\cdot)$, where $\epsilon = \int_{y\in\mathcal{X}} \big(\inf_{x\in C} p(x,y)\big)\nu(dy)$. The proof is that for $x \in \mathcal{X}$ and any $A \subseteq \mathcal{X}$,

$$P(x,A) = \int_A p(x,y)\,\nu(dy) \geq \int_A \big(\inf_{x\in C} p(x,y)\big)\nu(dy) = \epsilon\,\rho(A)\,.$$

7. (a) $PV(x) \equiv \mathbf{E}[V(X_{n+1} \,|\, X_n = x] = 1 + (x/2)^2 + (3/4) = x^2/4 + 7/4$.

(b) We verify that $PV(x) \leq (5/8)\,V(x) + (9/8)\mathbf{1}_C(x)$, i.e. we may take $\lambda = 5/8$ and $b = 9/8$.

(c) Here $\inf_{x\in C} p(x,y) = p(\sqrt{3},y) = h((y-\sqrt{3}/2)/\sqrt{3/4})$ for $y < 0$, and $\inf_{x\in C} p(x,y) = p(-\sqrt{3},y) = h((y-\sqrt{3}/2)/\sqrt{3/4})$ for $y > 0$, where again $h(t) = \frac{1}{\sqrt{2\pi}}e^{-t^2/2}$ is the standard normal density. Then $\epsilon = \int_{y\in\mathbf{R}} \big(\inf_{x\in C} p(x,y)\big)dy = \int_{-\infty}^{\infty} h((y-\sqrt{3}/2)/\sqrt{3/4}) +$

$\int_0^\infty h((y+\sqrt{3}/2)/\sqrt{3/4}) = 2\,\Phi(-1) > 0.31$ (where $\Phi(x) = \int_{-\infty}^x h(t)\,dt$ is the cdf of a standard normal).

(d) To obtain a quantitative bound, apply the above values of $\epsilon = 0.31$, $\lambda = 5/8$, $b = 9/8$, and $d = \sqrt{3}$ to the results on slides 83 and 84.

8. (a) If $x \geq \delta$, then we can reject only to the right, and

$$P[X_{n+1} = X_n \,|\, X_n = x] = (2\delta)^{-1} \int_x^{x+\delta} (1 - e^{x-y}) dy = (2\delta)^{-1}(\delta - 1 + e^{-\delta})\,.$$

If $x < \delta$, then we can also reject to the far left, and

$$P[X_{n+1} = X_n \,|\, X_n = x] = (2\delta)^{-1}(\delta - 1 + e^{-\delta} + (\delta - x))\,.$$

(b) The stationary rejection probability is then given by

$$R_\delta = (2\delta)^{-1}\Big(\delta - 1 + e^{-\delta} + \int_0^\delta (\delta - x)e^{-x}dx\Big) = 1 - (1 - e^{-\delta})\,/\,\delta\,.$$

We should then choose δ so that $1 - R_\delta \approx 0.234$, which is achieved at $\delta \doteq 4.2$ (though any value close to this is fine too).

9. Here

$$\mathcal{X}(1) = \{(0,0,0,0,0,0),\ (0,0,0,0,0,1),\ (0,0,0,0,1,1),\ (0,0,0,1,1,1),$$
$$(0,0,1,1,1,1),\ (0,1,1,1,1,1),\ (1,1,1,1,1,1),\ (1,1,1,1,1,0),$$
$$(1,1,1,1,0,0),\ (1,1,1,0,0,0),\ (1,1,0,0,0,0),\ (1,0,0,0,0,0)\}\,,$$

while

$$\mathcal{X}(2) = \{(0,1,0,1,0,1),\ (0,1,0,1,0,0),\ (0,1,0,1,1,0),\ (0,1,0,0,1,0),$$
$$(0,1,1,0,1,0),\ (0,0,1,0,1,0),\ (1,0,1,0,1,0),\ (1,0,1,0,1,1),$$
$$(1,0,1,0,0,1),\ (1,0,1,1,0,1),\ (1,0,0,1,0,1),\ (1,1,0,1,0,1)\}\,,$$

and the chain never moves between $\mathcal{X}(1)$ and $\mathcal{X}(2)$. Hence, the chain is <u>not</u> ϕ-irreducible. On the other hand, for the uniform distribution on either $\mathcal{X}(1)$ or $\mathcal{X}(2)$, we have $\mathbf{P}[a_i = 0] = \mathbf{P}[a_i = 1] = 1/2$ for each i. Hence, if you used any one-dimensional convergence diagnostic, you would conclude that the chain had converged, even though it was actually stuck in either $\mathcal{X}(1)$ or $\mathcal{X}(2)$.

10. We compute that

$$p(x,y) = q(y)\Big(1 \wedge \frac{\pi(y)q(x)}{\pi(x)q(y)}\Big) = \Big(q(y) \wedge \frac{\pi(y)q(x)}{\pi(x)}\Big) \geq \Big(\beta\pi(y) \wedge \pi(y)\beta\Big) = \beta\pi(y)\,.$$

It then follows from Theorem 5.7 on slide 57 that $\|P^n(x,\cdot) - \pi(\cdot)\| \leq 2(1-\beta)^n$, for all $x \in \mathcal{X}$.

11. We have $\alpha(x,y) = \min[1,\ \pi(y)/\pi(x)]$ which equals 1 for $y \leq x$ and equals a if $y = x+1$. The computation for $PV(x)$ is then the sum of three terms from either proposing a move right and accepting, proposing a move right and rejecting, and proposing a move left (and therefore accepting). The formula for λ follows by factoring out $V(x) = e^{\beta x}$. For large enough β, we have $a - e^{-\beta} > 0$, and then $\lambda < 1$.

12. The conditional densities are a standard result about the bivariate normal distribution. Hence the (deterministic-scan) Gibbs sampler sets $Y_{n+1} = \rho X_n + \sqrt{1-\rho^2}\, Z_{n+1}$ and then $X_{n+1} = \rho Y_{n+1} + \sqrt{1-\rho^2}\, W_{n+1} = \rho^2 X_n + \rho\sqrt{1-\rho^2}\, Z_{n+1} + \sqrt{1-\rho^2}\, W_{n+1}$ (where $\{Z_n\}$ and $\{W_n\}$ are i.i.d. standard normal). Hence, conditional on X_n, the conditional distribution of X_{n+1} is normal with mean $\rho^2 X_n$ and variance $\Big(\sqrt{1-\rho^2}\Big)^2 + \Big(\rho\sqrt{1-\rho^2}\Big)^2 = (1-\rho^2) + (\rho^2(1-\rho^2)) = 1-\rho^4$, as stated. The final statement about X_n in terms of X_0 then follows by induction.