and the Robbin–Monro algorithm. (The latter actually embeds the chain of interest $\theta^{(t)}$ in a larger chain $(\theta^{(t)}, \xi^{(t)})$ that also includes the parameter of the proposal distribution as well as the gradient of a performance criterion.) We will again consider adaptive algorithms in Chapter 14, with more accessible theoretical justifications.

7.7 Problems

7.1 Calculate the mean of a Gamma$(4.3, 6.2)$ random variable using
(a) Accept-Reject with a Gamma$(4, 7)$ candidate.
(b) Metropolis–Hastings with a Gamma$(4, 7)$ candidate.
(c) Metropolis–Hastings with a Gamma$(5, 6)$ candidate.
In each case monitor the convergence.

7.2 Student's $\mathcal{T}_\nu$ density with ν degrees of freedom is given by

$$f(x|\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\nu\pi}} \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2} .$$

Calculate the mean of a t distribution with 4 degrees of freedom using a Metropolis–Hastings algorithm with candidate density
(a) $\mathrm{N}(0, 1)$
(b) t with 2 degrees of freedom.
Monitor the convergence of each.

7.3 Complete some details of Theorem 7.2:
(a) To establish (7.2), show that

$$\begin{aligned} K(x, A) &= P(X_{t+1} \in A | X_t = x) \\ &= P(Y \in A \text{ and } X_{t+1} = Y | X_t = x) + P(x \in A \text{ and } X_{t+1} = x | X_t = x) \\ &= \int_A q(y|x)\varrho(x, y)dy + \int_{\mathcal{Y}} \mathbb{I}(x \in A)(1 - \varrho(x, y))q(y|x)dy, \end{aligned}$$

where $q(y|x)$ is the instrumental density and $\varrho(x, y) = P(X_{t+1} = y | X_t = x)$. Take the limiting case $A = \{y\}$ to establish (7.2).
(b) Establish (7.3). Notice that $\delta_y(x)f(y) = \delta_x(y)f(x)$.

7.4 For the transition kernel,

$$X^{(t+1)}|x^{(t)} \sim \mathcal{N}(\rho x^{(t)}, \tau^2)$$

gives sufficient conditions on ρ and τ for the stationary distribution π to exist. Show that, in this case, π is a normal distribution and that (7.4) occurs.

7.5 (Doukhan et al. 1994) The algorithm presented in this problem is used in Chapter 12 as a benchmark for slow convergence.
(a) Prove the following result:

Lemma 7.24. *Consider a probability density g on $[0, 1]$ and a function $0 < \rho < 1$ such that*

$$\int_0^1 \frac{g(x)}{1 - \rho(x)} \, dx < \infty .$$

The Markov chain with transition kernel

$$K(x,x') = \rho(x)\ \delta_x(x') + (1-\rho(x))\ g(x')\ ,$$

where δ_x is the Dirac mass at x, has stationary distribution

$$f(x) \propto g(x)/(1-\rho(x)).$$

(b) Show that an algorithm for generating the Markov chain associated with Lemma 7.24 is given by

Algorithm A.30 –Repeat or Simulate–

```
1. Take X^(t+1) = x^(t) with probability ρ(x^(t))
2. Else, generate X^(t+1) ~ g(y).                    [A.30]
```

(c) Highlight the similarity with the Accept–Reject algorithm and discuss in which sense they are complementary.

7.6 (Continuation of Problem 7.5) Implement the algorithm of Problem 7.5 when g is the density of the $\mathcal{B}e(\alpha+1,1)$ distribution and $\rho(x) = 1-x$. Give the expression of the stationary distribution f. Study the acceptance rate as α varies around 1. (*Note:* Doukhan et al. 1994 use this example to derive β-mixing chains which do not satisfy the Central Limit Theorem.)

7.7 (Continuation of Problem 7.5) Compare the algorithm $[A.30]$ with the corresponding Metropolis–Hastings algorithm; that is, the algorithm $[A.25]$ associated with the same pair (f,g). (*Hint:* Take into account the fact that $[A.30]$ simulates only the y_t's which are not discarded and compare the computing times when a recycling version as in Section 7.6.2 is implemented.)

7.8 Determine the distribution of Y_t given $y_{t-1},\ldots$ in $[A.25]$.

7.9 (Tierney 1994) Consider a version of $[A.25]$ based on a "bound" M on f/g that is not a uniform bound; that is, $f(x)/g(x) > M$ for some x.

(a) If an Accept–Reject algorithm uses the density g with acceptance probability $f(y)/Mg(y)$, show that the resulting variables are generated from

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

instead of f.

(b) Show 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

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

to produce a sample from f.

7.10 The inequality (7.8) can also be established using *Orey's inequality* (See Problem 6.42 for a slightly different formulation) For two transitions P and Q,

$$||P^n - Q^n||_{TV} \le 2P(X_n \neq Y_n), \qquad X_n \sim P^n, \quad Y_n \sim Q^n.$$

Deduce that when P is associated with the stationary distribution f and when X_n is generated by $[A.25]$, under the condition (7.7),

$$||P^n - f\|_{TV} \leq \left(1 - \frac{1}{M}\right)^n .$$

Hint: Use a coupling argument based on

$$X^n = \begin{cases} Y^n & \text{with probability } 1/M \\ Z^n \sim \dfrac{g(z) - f(z)/M}{1 - 1/M} & \text{otherwise.} \end{cases} .$$

7.11 Complete the proof of Theorem 7.8:

(a) Verify (7.10) and prove (7.11). (*Hint:* By (7.9), the inner integral is immediately bounded by $1 - \frac{1}{M}$. Then repeat the argument for the outer integral.)

(b) Verify (7.12) and prove (7.8).

7.12 In the setup of Hastings (1970) uniform-normal example (see Example 7.14):

(a) Study the convergence rate (represented by the 90% interquantile range) and the acceptance rate when δ increases.

(b) Determine the value of δ which minimizes the variance of the empirical average. (*Hint:* Use a simulation experiment.)

7.13 Show that, for an arbitrary Metropolis–Hastings algorithm, every compact set is a small set when f and q are positive and continuous everywhere.

7.14 (Mengersen and Tweedie 1996) With respect to Theorem 7.15, define

$$A_x = \{y;\ f(x) \leq f(y)\} \quad \text{and} \quad B_x = \{y;\ f(x) \geq f(y)\}.$$

(a) If f is symmetric, show that $A_x = \{|y| < |x|\}$ for $|x|$ larger than a value x_0.

(b) Define x_1 as the value after which f is log-concave and $x^* = x_0 \vee x_1$. For $V(x) = \exp s|x|$ and $s < \alpha$, show that

$$\begin{aligned} \frac{\mathbb{E}[V(X_1)|x_0 = x]}{V(x)} &\leq 1 + \int_0^x \left[e^{s(y-x)} - 1\right] g(x-y)dy \\ &\quad + \int_x^{2x} e^{-\alpha(y-x)} \left[e^{s(y-x)} - 1\right] g(x-y)dy \\ &\quad + 2\int_x^\infty g(y)dy . \end{aligned}$$

(c) Show that

$$\begin{aligned} &\int_0^x \left(e^{-sy} - 1 + e^{-(\alpha-s)y} - e^{-\alpha y}\right) g(y)dy \\ &\qquad = -\int_0^x \left[1 - e^{-sy}\right] \left[1 - e^{-(\alpha-s)y}\right] g(y)dy \end{aligned}$$

and deduce that (7.17) holds for $x > x^*$ and x^* large enough.

(d) For $x < x^*$, show that

$$\frac{\mathbb{E}[V(X_1)|x_0 = x]}{V(x)} \leq 1 + 2\int_{x^*}^\infty g(y)dy + 2e^{sx^*} \int_0^{x^*} g(z)dz$$

and thus establish the theorem.

7.15 Examine whether the following distributions are log-concave in the tails: Normal, log-normal, Gamma, Student's t, Pareto, Weibull.

7.16 The following theorem is due to Mengersen and Tweedie (1996).

Theorem 7.25. *If the support of f is not compact and if g is symmetric, the chain $(X^{(t)})$ produced by [A.29] is not uniformly ergodic.*

Assume that the chain satisfies Doeblin's condition (Theorem 6.59).

(a) Take x_0 and $A_0 =]-\infty, x_0]$ such that $\nu(A_0) > \varepsilon$ and consider the *unilateral* version of the random walk, with kernel

$$K^-(x, A) = \frac{1}{2}\, \mathbb{I}_A(x) + \int_{A \cap]-\infty, x]} g(x-y)\, dy \ ;$$

that is, the random walk which only goes to the left. Show that for $y > x_0$,

$$P^m(y, A_0) \le P_y(\tau \le m) \le P_y(\tau^- \le m),$$

where τ and τ^- are the return times to A_0 for the chain $(X^{(t)})$ and for K^-, respectively,

(b) For y sufficiently large to satisfy

$$(K^-)^m(y, A_0) = (K^-)^m(0,]-\infty, x_0 - y]) < \frac{\delta}{m}\ ,$$

show that

$$P_y(\tau^- \le m) \le \sum_{j=1}^{m} (K^-)^j(y, A_0) \le m (K^-)^m(y, A_0)\ ,$$

contradicting Doeblin's condition and proving the theorem.

(c) Formulate a version of Theorem 7.25 for higher-dimensional chains.

7.17 Mengersen and Tweedie (1996) also establish the following theorem:

Theorem 7.26. *If g is continuous and satisfies*

$$\int |x|\, g(x) dx < \infty\ ,$$

the chain $(X^{(t)})$ of [A.29] is geometrically ergodic if and only if

$$\overline{\varphi} = \lim_{x \to \infty} \frac{d}{dx} \log f(x) < 0\ .$$

(a) To establish sufficiency, show that for $x < y$ large enough, we have

$$\log f(y) - \log f(x) = \int_x^y \frac{d}{dt} \log f(t) dt \le \frac{\overline{\varphi}}{2}\,(y - x).$$

Deduce that this inequality ensures the log-concavity of f and, therefore, the application of Theorem 7.15.

(b) For necessity, suppose $\overline{\varphi} = 0$. For every $\delta > 0$, show that you can choose x large enough so that $\log f(x+z) - \log f(x) \ge -\delta z$, $z > 0$ and, therefore, $f(x+z)\, \exp(\delta z) \ge f(x)$. By integrating out the z's, show that

$$\int_x^\infty f(y)\, e^{\delta y}\, dy = \infty\ ,$$

contradicting condition (7.18).

7.18 For the situation of Example 7.16, show that

$$\lim_{x\to\infty} \frac{d}{dx} \log\varphi(x) = -\infty \quad \text{and} \quad \lim_{x\to\infty} \frac{d}{dx} \log\psi(x) = 0\,,$$

showing that the chain associated with φ is geometrically ergodic and the chain associated with ψ is not.

7.19 Verify that the transition matrix associated with the geometric random walk in Example 7.17 is correct and that $\beta = \theta^{-1/2}$ minimizes λ_β.

7.20 The Institute for Child Health Policy at the University of Florida studies the effects of health policy decisions on children's health. A small portion of one of their studies follows.

The overall health of a child (metq) is rated on a 1–3 scale, with 3 being the worst. Each child is in an HMO[8] (variable np, 1=nonprofit, −1=for profit). The dependent variable of interest (y_{ij}) is the use of an emergency room (erodds, 1=used emergency room, 0=did not). The question of interest is whether the status of the HMO affects the emergency room choice.

(a) An appropriate model is the logistic regression model,

$$\text{logit}(p_{ij}) = a + bx_i + cz_{ij}, \quad i = 1, \ldots, k, \quad j = 1, \ldots, n_i\,,$$

where x_i is the HMO type, z_{ij} is the health status of the child, and p_{ij} is the probability of using an emergency room. Verify that the likelihood function is

$$\prod_{i=1}^{k}\prod_{j=1}^{n_i}\left(\frac{\exp(a+bx_i+cz_{ij})}{1+\exp(a+bx_i+cz_{ij})}\right)^{y_{ij}}\left(\frac{1}{1+\exp(a+bx_i+cz_{ij})}\right)^{1-y_{ij}}.$$

(Here we are only distinguishing between for-profit and non-profit, so $k = 2$.)

(b) Run a standard GLM on these data[9] and get the estimated mean and variance of a, b, and c.

(c) Use normal candidate densities with mean and variance at the GLM estimates in a Metropolis–Hastings algorithm that samples from the likelihood. Get histograms of the parameter values.

7.21 The famous "braking data" of Tukey (1977) is given in Table 7.6. It is thought that a good model for this dataset is a quadratic model

$$y_{ij} = a + bx_i + cx_i^2 + \varepsilon_{ij}, \quad i = 1, \ldots, k, \quad j = 1, \ldots n_i.$$

If we assume that $\varepsilon_{ij} \sim N(0, \sigma^2)$, independent, then the likelihood function is

$$\left(\frac{1}{\sigma^2}\right)^{N/2} e^{\frac{1}{2\sigma^2}\sum_{ij}(y_{ij}-a-bx_i-cx_i^2)^2}\,,$$

where $N = \sum_i n_i$. We can view this likelihood function as a posterior distribution of a, b, c, and σ^2, and we can sample from it with a Metropolis–Hastings algorithm.

[8] A person who joins an *HMO* (for Health Maintenance Organization) obtains their medical care through physicians belonging to the HMO.

[9] Available as LogisticData.txt on the book website.

x	4	7	8	9	10	11	12
y	2,10	4,22	16	10	18,26	17,28	14,20
					34		24,28
x	13	14	15	16	17	18	19
y	26,34	26,36	20,26	32,40	32,40	42,56	36,46
	34,46	60,80	54		50	76,84	68
x	20	22	23	24	25		
y	32,48	66	54	70,92	85		
	52,56,64			93,120			

Table 7.6. Braking distances of 50 cars, x = speed (mph), y = distance to stop (feet).

(a) Get estimates of a, b, c, and σ^2 from a usual linear regression.
(b) Use the estimates to select a candidate distribution. Take normals for a, b, c and inverted gamma for σ^2.
(c) Make histograms of the posterior distributions of the parameters. Monitor convergence.
(d) Robustness considerations could lead to using an error distribution with heavier tails. If we assume that $\varepsilon_{ij} \sim t(0, \sigma^2)$, independent, then the likelihood function is

$$\left(\frac{1}{\sigma^2}\right)^{N/2} \prod_{ij} \left(1 + \frac{(y_{ij} - a - bx_i - cx_i^2)^2}{\nu}\right)^{-(\nu+1)/2},$$

where ν is the degrees of freedom. For $\nu = 4$, use Metropolis–Hastings to sample a, b, c, and σ^2 from this posterior distribution. Use either normal or t candidates for a, b, c, and either inverted gamma or half-t for σ^2.

(*Note*: See Problem 11.4 for another analysis of this dataset.)

7.22 The *traveling salesman problem* is a classic in combinatoric and operations research, where a salesman has to find the shortest route to visit each of his N customers.

(a) Show that the problem can be described by (i) a permutation σ on $\{1, \ldots, N\}$ and (ii) a distance $d(i, j)$ on $\{1, \ldots, N\}$.
(b) Deduce that the traveling salesman problem is equivalent to minimization of the function

$$H(\sigma) = \sum_i d(i, \sigma(i)).$$

(c) Propose a Metropolis–Hastings algorithm to solve the problem with a simulated annealing scheme (Section 5.2.3).
(d) Derive a simulation approach to the solution of $Ax = b$ and discuss its merits.

7.23 Check whether a negative coefficient b in the random walk $Y_t = a + b(X^{(t)} - a) + Z_t$ induces a negative correlation between the $X^{(t)}$'s. Extend to the case where the random walk has an ARCH-like structure,

$$Y_t = a + b(X^{(t)} - a) + \exp(c + d(X^{(t)} - a)^2) Z_t.$$

7.24 Implement the Metropolis–Hastings algorithm when f is the normal $\mathcal{N}(0,1)$ density and $q(\cdot|x)$ is the uniform $\mathcal{U}[-x-\delta, -x+\delta]$ density. Check for negative correlation between the $X^{(t)}$'s when δ varies.

7.25 Referring to Example 7.11

(a) Verify that $\exp\alpha$ has an exponential distribution.

(b) Show that the posterior distribution is proper, that is

$$\int L(\alpha,\beta|\mathbf{y})\pi(\alpha,\beta)d\alpha d\beta < \infty\,.$$

(c) Show that $\mathbb{E}\alpha = \log b - \gamma$, where γ is Euler's constant.

7.26 Referring to the situation of Example 7.12:

(a) Use Taylor series to establish the approximations

$$\begin{aligned} K_X(t) &\approx K_X(0) + K_X'(0)t + K_X''(0)t^2/2 \\ tK_X'(t) &\approx t\left[K_X'(0) + K_X''(0)t\right] \end{aligned}$$

and hence (7.14).

(b) Write out the Metropolis–Hastings algorithm that will produce random variables from the saddlepoint distribution.

(c) Apply the Metropolis saddlepoint approximation to the noncentral chi squared distribution and reproduce the tail probabilities in Table 7.1.

7.27 Given a Cauchy $\mathcal{C}(0,\sigma)$ instrumental distribution:

(a) Experimentally select σ to maximize (i) the acceptance rate when simulating a $\mathcal{N}(0,1)$ distribution and (ii) the squared error when estimating the mean (equal to 0).

(b) Same as (a), but when the instrumental distribution is $\mathcal{C}(x^{(t)},\sigma)$.

7.28 Show that the Rao–Blackwellized estimator δ^{RB} does not depend on the normalizing factors in f and g.

7.29 Reproduce the experiment of Example 7.20 in the case of a Student's $\mathcal{T}_7$ distribution.

7.30 In the setup of the Metropolis–Hastings algorithm [$A.24$], the Y_t's are generated from the distributions $q(y|x^{(t)})$. Assume that $Y_1 = X^{(1)} \sim f$.

(a) Show that the estimator

$$\delta_0 = \frac{1}{T}\sum_{t=1}^{T} \frac{f(y_t)}{q(y_t|x^{(t)})} h(y_t)$$

is an unbiased estimator of $\mathbb{E}_f[h(X)]$.

(b) Derive, from the developments of Section 7.6.2, that the Rao–Blackwellized version of δ_0 is

$$\begin{aligned} \delta_1 = \frac{1}{n}\Bigg\{ & h(x_1) + \frac{f(y_2)}{q(y_2|x_1)} h(y_2) \\ & + \frac{\sum_{i=3}^{T}\sum_{j=1}^{i-1} f(y_i)\delta_j\zeta_{j(i-2)}(1-\rho_{j(i-1)})\omega_i^j}{\sum_{i=1}^{T-1}\delta_i\zeta_{i(T-1)}} h(y_i) \Bigg\}\,. \end{aligned}$$

(c) Compare δ_1 with the Rao–Blackwellized estimator of Theorem 7.21 in the case of a $\mathcal{T}_3$ distribution for the estimation of $h(x) = \mathbb{I}_{x>2}$.

7.31 Prove Theorem 7.19 as follows:

(a) Use the properties of the Markov chain to show that the conditional probability τ_i can be written as

$$\tau_i = \sum_{j=0}^{i-1} P(X_i = y_i | X_{i-1} = y_j)\, P(X_{i-1} = y_j) = \sum_{j=0}^{i-1} \rho_{ji}\, P(X_{i-1} = y_j).$$

(b) Show that

$$\begin{aligned} P(X_{i-1} = y_j) &= P(X_j = y_j, X_{j+1} = y_j, \ldots, X_{i-1} = y_j) \\ &= (1 - \rho_{j(j+1)}) \cdots (1 - \rho_{j(i-1)}) \end{aligned}$$

and, hence, establish the expression for the weight φ_i.

7.32 Prove Theorem 7.21 as follows:

(a) As in the independent case, the first step is to compute $P(X_j = y_i | y_0, y_1, \ldots, y_n)$. The event $\{X_j = y_i\}$ can be written as the set of all the i-tuples $(u_1, \ldots, u_i)$ leading to $\{X_i = y_i\}$, of all the $(j-i)$-tuples $(u_{i+1}, \ldots, u_j)$ corresponding to the rejection of $(y_{i+1}, \ldots, y_j)$ and of all the $(n-j)$-tuples $u_{j+1}, \ldots, u_n$ following after $X_j = y_i$.
Define $B_0^1 = \{u_1 > \rho_{01}\}$ and $B_1^1 = \{u_1 < \rho_{01}\}$, and let $B_k^t(u_1, \ldots, u_t)$ denote the event $\{X_t = y_k\}$. Establish the relation

$$\begin{aligned} B_k^t(u_1, \ldots, u_t) = \bigcup_{m=0}^{k-1} \Big[& B_m^{k-1}(u_1, \ldots, u_{k-1}) \\ & \cap \{u_k < \rho_{mk}, u_{k+1} > \rho_{k(t+1)}, \ldots, u_t > \rho_{kt}\} \Big], \end{aligned}$$

and show that

$$\begin{aligned} \{x_j = y_i\} = \bigcup_{k=0}^{i-1} \Big[& B_k^{i-1}(u_1, \ldots, u_{i-1}) \\ & \cap \{u_i < \rho_{ki}, u_{i+1} > \rho_{i(i+1)}, \ldots, u_j > \rho_{ij}\} \Big]. \end{aligned}$$

(b) Let $p(u_1, \ldots, u_T, y_1, \ldots, y_T) = p(\mathbf{u}, \mathbf{y})$ denote the joint density of the $U_i's$ and the $Y_i's$. Show that $\tau_i = \int_A p(\mathbf{u}, \mathbf{y}) du_1 \cdots du_i$, where $A = \bigcup_{k=0}^{i-1} B_k^{i-1}(u_1, \ldots, u_{i-1}) \cap \{u_i < \rho_{ki}\}$.

(c) Show that $\omega_{j+1}^i = \int_{\{x_j = y_i\}} p(\mathbf{u}, \mathbf{y}) du_{j+1} \cdots du_T$ and, using part (b), establish the identity

$$\tau_i \prod_{t=i+1}^{j} (1 - \rho_{it}) q(y_{t+1} | y_i) \omega_{j+1}^i = \tau_i \prod_{t=i+1}^{j} \underline{\rho}_{it} \omega_{j+1}^i = \tau_i\, \zeta_{ij}\, \omega_{j+1}^i.$$

(d) Verify the relations

$$\tau_i = \sum_{t=0}^{i-1} \tau_i\, \zeta_{t(i-1)j} \overline{\rho}_{ti} \quad \text{and} \quad \omega_{j+1}^i = \omega_{j+2}^{j+1}\, \underline{\rho}_{i(j+1)} + \omega_{j+2}^i\, \overline{\rho}_{i(j+1)},$$

which provide a recursion relation on the ω_j^i's depending on acceptance or rejection of y_{j+1}. The case $j = T$ must be dealt with separately, since there is no generation of y_{T+1} based on $q(y|x_T)$. Show that ω_T^i is equal to $\rho_{iT} + (1 - \rho_{iT}) = 1$

(e) The probability $P(X_j = y_i)$ can be deduced from part (d) by computing the marginal distribution of $(Y_1, \ldots, Y_T)$. Show that $1 = \sum_{i=0}^{T} P(X_T = y_i) = \sum_{i=0}^{T-1} \tau_i \zeta_{i(T-1)}$, and, hence, the normalizing constant for part (d) is $(\sum_{i=0}^{T-1} \tau_i \zeta_{i(T-1)})^{-1}$, which leads to the expression for φ.

7.33 (Liu 1996b) Consider a finite state-space $\mathcal{X} = \{1, \ldots, m\}$ and a Metropolis–Hastings algorithm on $\mathcal{X}$ associated with the stationary distribution $\pi = (\pi_1, \ldots, \pi_m)$ and the proposal distribution $p = (p_1, \ldots, p_m)$.

(a) For $\omega_i = \pi_i/p_i$ and $\lambda_k = \sum_{i=k}^{m}(p_i - \pi_i/\omega_k)$, express the transition matrix of the Metropolis–Hastings algorithm as $K = G + \mathbf{e}p^T$, where $\mathbf{e} = (1, \ldots, 1)^T$. Show that G is upper triangular with diagonal elements the λ_k's.

(b) Deduce that the eigenvalues of G and K are the λ_k's.

(c) Show that for $||p|| = \sum |p_i|$,

$$||K^n(x, \cdot) - \pi||^2 \leq \frac{\lambda_1^{2n}}{\pi(x)},$$

following a result by Diaconis and Hanlon (1992).

7.34 (Ó Ruanaidh and Fitzgerald 1996) Given the model

$$y_i = A_1 e^{\lambda_1 t_i} + A_2 e^{\lambda_2 t_i} + c + \epsilon_i,$$

where $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ and the t_i's are known observation times, study the estimation of $(A_1, A_2, \lambda_1, \lambda_2, \sigma)$ by recursive integration (see Section 5.2.4) with particular attention to the Metropolis–Hastings implementation.

7.35 (Roberts 1998) Take f to be the density of the $\mathcal{E}xp(1)$ distribution and g_1 and g_2 the densities of the $\mathcal{E}xp(0.1)$ and $\mathcal{E}xp(5)$ distributions, respectively. The aim of this problem is to compare the performances of the independent Metropolis–Hastings algorithms based on the pairs (f, g_1) and (f, g_2).

(a) Compare the convergences of the empirical averages for both pairs, based on 500 replications of the Markov chains.

(b) Show that the pair (f, g_1) leads to a geometrically ergodic Markov chain and (f, g_2) does not.

7.36 For the Markov chain of Example 7.28, define

$$\xi^{(t)} = \mathbb{I}_{[0,13]}(\theta^{(t)}) + 2\mathbb{I}_{[13,\infty)}(\theta^{(t)}).$$

(a) Show that $(\xi^{(t)})$ is not a Markov chain.

(b) Construct an estimator of the pseudo-transition matrix of $(\xi^{(t)})$.

7.37 Show that the transition associated with the acceptance probability (7.20) also leads to f as invariant distribution, for every symmetric function s. (*Hint:* Use the reversibility equation.)

7.38 Show that the Metropolis–Hastings algorithm is, indeed, a special case of the transition associated with the acceptance probability (7.20) by providing the corresponding $s(x, y)$.

7.39 (Peskun 1973) Let $\mathbb{P}_1$ and $\mathbb{P}_2$ be regular (see Problems 6.9 and 6.10), reversible stochastic matrices with the same stationary distribution π on $\{1, \ldots, m\}$. Show that if $\mathbb{P}_1 \leq \mathbb{P}_2$ (meaning that the off-diagonal elements are smaller in the first case) for every function h,

$$\lim_{N\to\infty} \text{var}\left[\sum h(X_1^{(t)})\right]/N \geq \lim_{N\to\infty} \text{var}\left[\sum h(X_2^{(t)})\right]/N\ ,$$

where $(X_i^{(t)})$ is a Markov chain with transition matrix $\mathbb{P}_i$ $(i = 1, 2)$. (*Hint:* Use Kemeny and Snell 1960 result on the asymptotic variance in Problem 6.50.)

7.40 (Continuation of Problem 7.39) Deduce from Problem 7.39 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

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

where $s_{ij} = s_{ji}$ and $0 \leq \alpha_{ij} \leq 1$. (*Hint:* Give the corresponding α_{ij} for the Metropolis–Hastings algorithm and show that it is maximal for $i \neq j$. Tierney 1998, Mira and Geyer 1998, and Tierney and Mira 1998 propose extensions to the continuous case.)

7.41 Show that f is the stationary density associated with the acceptance probability (7.20).

7.42 In the setting of Example 7.28, implement the simulated annealing algorithm to find the maximum of the likelihood $\mathrm{L}(\theta|x_1, x_2, x_3)$. Compare with the performances based on $\log \mathrm{L}(\theta|x_1, x_2, x_3)$.

7.43 (Winkler 1995) A *Potts model* is defined on a set S of “sites” and a finite set G of “colors” by its energy

$$H(x) = -\sum_{(s,t)} \alpha_{st} \mathbb{I}_{x_s = x_t}, \qquad x \in G^S,$$

where $\alpha_{st} = \alpha_{ts}$, the corresponding distribution being $\pi(x) \propto \exp(H(x))$. An additional structure is introduced as follows: “Bonds” b are associated with each pair (s,t) such that $\alpha_{st} > 0$. These bonds are either active ($b = 1$) or inactive ($b = 0$).

(a) Defining the joint distribution

$$\mu(x, b) \propto \prod_{b_{st}=0} q_{st} \prod_{b_{st}=1} (1 - q_{st}) \mathbb{I}_{x_s = x_t},$$

with $q_{st} = \exp(\alpha_{st})$, show that the marginal of μ in x is π. Show that the marginal of μ in b is

$$\mu(b) \propto |G|^{c(b)} \prod_{b_{st}=0} q_{st} \prod_{b_{st}=1} (1 - q_{st}),$$

where $c(b)$ denotes the number of *clusters* (the number of sites connected by active bonds).

(b) Show that the *Swendson–Wang* (1987) algorithm

```
1. Take b_st = 0 if x_s ≠ x_t and, for x_s = x_t,
```

$$b_{st} = \begin{cases} 1 & \texttt{with probability } 1 - q_{st} \\ 0 & \texttt{otherwise.} \end{cases}$$

```
2. For every cluster, choose a color at random on G.
```

leads to simulations from π (*Note:* This algorithm is acknowledged as accelerating convergence in image processing.)

7.44 (McCulloch 1997) In a generalized linear mixed model, assume the link function is $h(\xi_i) = x_i'\beta + z_i'\mathbf{b}$, and further assume that $\mathbf{b} = (b_1, \ldots, b_I)$ where $\mathbf{b} \sim f_{\mathbf{b}}(\mathbf{b}|D)$. (Here, we assume φ to be unknown.)

(a) Show that the usual (incomplete-data) likelihood is

$$L(\theta, \varphi, D|y) = \int \prod_{i=1}^{n} f(y_i|\theta_i) f_{\mathbf{b}}(\mathbf{b}|D) d\mathbf{b} \, .$$

(b) Denote the complete data by $\mathbf{w} = (\mathbf{y}, \mathbf{b})$, and show that

$$\log L_W = \sum_{i=1}^{n} \log f(y_i|\theta_i) + \log f_{\mathbf{b}}(b_i|D) \, .$$

(c) Show that the EM algorithm, given by

1. Choose starting values $\beta^{(0)}$, $\varphi^{(0)}$, and $D^{(0)}$.
2. Calculate (expectations evaluated under $\beta^{(m)}$, $\varphi^{(m)}$, and $D^{(m)}$) $\beta^{(m+1)}$ and $\varphi^{(m+1)}$, which maximize $\mathbb{E}[\log f(y_i|\theta_i, u, \beta, \varphi)|y]$.
3. $D^{(m+1)}$ maximizes $\mathbb{E}[f_{\mathbf{b}}(\mathbf{b}|D)|y]$.
4. Set m to $m+1$.

converges to the MLE.

The next problems (7.45–7.49) deal with Langevin diffusions, as introduced in Section 7.8.5.

7.45 Show that the naïve discretization of (7.22) as $dt = \sigma^2$, $dL_t = X^{(t+\sigma^2)} - X^{(t)}$, and $dB_t = B_{t+\sigma^2} - B_t$ does lead to the representation (7.23).

7.46 Consider f to be the density of $\mathcal{N}(0,1)$. Show that when $\sigma = 2$ in (7.23), the limiting distribution of the chain is $\mathcal{N}(0,2)$.

7.47 Show that (7.27) can be directly simulated as

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

7.48 Show that when (7.24) exists and is larger (smaller) than 1 (-1) at $-\infty$ (∞), the random walk (7.23) is transient.

7.49 (Stramer and Tweedie 1999a) Show that the following stochastic differential equation still produces f as the stationary distribution of the associated process:

$$dL_t = \sigma(L_t)\nabla \log f(L_t) + b(L_t)dt,$$

when

$$b(x) = \frac{1}{2}\nabla \log f(x)\sigma^2(x) + \sigma(x)\nabla\sigma(x).$$

Give a discretized version of this differential equation to derive a Metropolis–Hastings algorithm and apply to the case $\sigma(x) = \exp(\omega|x|)$.

7.8 Notes

7.8.1 Background of the Metropolis Algorithm

The original Metropolis algorithm was introduced by Metropolis et al. (1953) in a setup of optimization on a discrete state-space, in connection with particle physics: