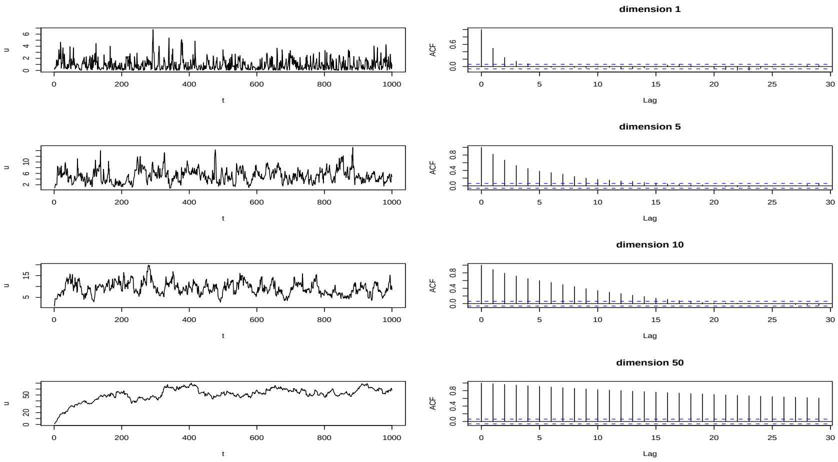

Fig. 8.5. Rawplots and autocorrelation functions for the series $(z^{(t)})$ generated by the slice samplers associated with $f_1(z) = z^{d-1}\, e^{-z}$ for $d = 1, 5, 10, 50$.

8.4 Problems

8.1 Referring to Algorithm [A.31] and equations (8.1) and (8.2):

(a) Show that the stationary distribution of the Markov chain [A.31] is the uniform distribution on the set $\{(x, u) : 0 < u < f(x)\}$.

(b) Show that the conclusion of part (a) remains the same if we use f_1 in [A.31], where $f(x) = C f_1(x)$.

8.2 In the setup of Example 8.1, show that the cdf associated with the density $\exp(-\sqrt{x})$ on $\mathbb{R}_+$ can be computed in closed form. (*Hint:* Make the change of variable $z = \sqrt{x}$ and do an integration by parts on $z\exp(-z)$.)

8.3 Consider two unnormalized versions of a density f, f_1 and f_2. By implementing the slice sampling algorithm [A.31] on both f_1 and f_2, show that the chains $(x^{(t)})_t$ produced by both versions are exactly the same if they both start from the same value $x^{(0)}$ and use the same uniform $u^{(t)} \sim \mathcal{U}([0, 1])$.

8.4 As a generalization of the density in Example 8.1, consider the density $f(x) \propto \exp\{-x^d\}$, for $d < 1$. Write down a slice sampler algorithm for this density, and evaluate its performance for $d = .1, .25, .4$.

8.5 Show that a possible slice sampler associated with the standard normal density, $f(x) \propto \exp(-x^2/2)$, is associated with the two conditional distributions

$$\omega | x \sim \mathcal{U}_{[0, \exp(-x^2/2)]}\,, \qquad X|\omega \sim \mathcal{U}\left([-\sqrt{-2\log(\omega)}, \sqrt{-2\log(\omega)}]\right)\,. \tag{8.6}$$

Compare the performances of this slice sampler with those of an iid sampler from $\mathcal{N}(0, 1)$ by computing the empirical cdf at 0, .67, .84, 1.28, 1.64, 1.96, 2.33, 2.58, 3.09, and 3.72 for two samples of same size produced under both approaches. (Those figures correspond to the .5, .75, .8, .9, .95, .99, .995, .999 and .9999 quantiles, respectively.)

8.6 Reproduce the comparison of Problem 8.5 in the case of (a) the gamma distribution and (b) the Poisson distribution.

8.7 Consider the mixture distribution (1.3), where the p_j's are unknown and the f_j's are known densities. Given a sample $x_1, \ldots, x_n$, examine whether or not a slice sampler can be constructed for the associated posterior distribution.

8.8 (Neal 2003) Consider the following hierarchical model ($i = 1, \ldots, 10$)

$$x_i|v \sim \mathcal{N}(0, e^v)\,, \qquad v \sim \mathcal{N}(0,3)\,.$$

(a) When simulating this distribution from a independent Metropolis–Hastings algorithm with $\mathcal{N}(0,1)$ proposals on all variables, show that the algorithm fails to recover the proper distribution for v by looking at the smaller values of v.
(b) Explain why this poor behavior occurs by considering the acceptance probability when v is smaller than -5.
(c) Evaluate the performance of the corresponding random Metropolis–Hastings algorithm which updates one component at a time with a Gaussian proposal centered at the current value and variance equal to one. Explain why the problem now occurs for the larger values of v.
(d) Compare with a single-variable slice sampling, that is, slice sampling applied to the eleven full conditional distributions of v given the x_i's and of the x_i's given the x_j's and v.

8.9 (Roberts and Rosenthal 1998) Show that, for the slice sampler [$A.31$], $(f_1(X^{(t)}))$ is a Markov chain with the same convergence properties as the original chain $(X^{(t)})$.

8.10 (Roberts and Rosenthal 1998) Using the notation of Section 8.3, show that, if f_1 and $\tilde{f}_1$ are two densities on two spaces $\mathcal{X}$ and $\tilde{\mathcal{X}}$ such that $\mu(w) = \tilde{\mu}(aw)$ for all w's, the kernels of the chains $(f_1(X^{(t)}))$ and $(\tilde{f}_1(\tilde{X}^{(t)}))$ are the same, even when the dimensions of $\mathcal{X}$ and $\tilde{\mathcal{X}}$ are different.

8.11 (Roberts and Rosenthal 1998) Consider the distribution on $\mathbb{R}^d$ with density proportional to

$$f_0(x)\,\mathbb{I}_{\{y\in\mathcal{X};\, f_1(y)\geq\omega\}}(x)\,.$$

Let T be a differentiable injective transformation, with Jacobian J.
(a) Show that sampling x from this density is equivalent to sampling $z = T(x)$ from

$$f_0^T(z) = f_0(T^{-1}(z))/J(T^{-1}(z))\,\mathbb{I}_{T(\{y\in\mathcal{X};\, f_1(y)\geq\omega\})}(x)\,,$$

and deduce that $P_{f_0}(x, A) = P_{f_0^T}(T(x), T(A))$.
(b) Show that the transformation $T(x) = (T_1(x), x_2, \ldots, x_d)$, where $T_1(x) = \int_0^{x_1} f_0(t, x_2, \ldots, x_d)\,dt$, is such that $f_0(T^{-1}(z))/J(T^{-1}(z))$ is constant over the range of T.

8.12 (Roberts and Rosenthal 1998) In the proof of Lemma 8.7, when $f_1(x) \leq \epsilon^\star$:
(a) Show that the first equality follows from the definition of V.
(b) Show that the second equality follows from

$$\mu(\omega) = \int_\omega^\infty (-\mu'(z))\,dz\,.$$

(c) Show that the first inequality follows from the fact that $z^{-\beta}$ is decreasing and the ratio can be expressed as an expectation.
(d) Show that, if g_1 and g_2 are two densities such that $g_1(x)/g_2(x)$ is increasing, and if h is also increasing, then $\int h(x)g_1(x)dx \geq \int h(x)g_2(x)dx$. Deduce the second inequality from this general result.

(e) Establish the third equality.
(f) Show that

$$\frac{z^{-\beta} - \epsilon^{\star -\beta}}{z^{-1/\alpha} - \epsilon^{\star -1/\alpha}}$$

is an increasing function of $z < \epsilon^\star$ when $\beta\alpha < 1$. Derive the last inequality.

8.13 (Roberts and Rosenthal 1998) Show that $\mu(y)' y^{1+1/\alpha}$ is non-increasing in the following two cases:
(a) $\mathcal{X} = \mathbb{R}^+$ and $f_1(x) \propto \exp -\gamma x$ for x large enough;
(b) $\mathcal{X} = \mathbb{R}^+$ and $f_1(x) \propto x^{-\delta}$ for x large enough.

8.14 (Roberts and Rosenthal 1998) In Lemma 8.7, show that, if $\epsilon^\star = 1$, then the bound b on $KV(x)$ when $x \in \mathscr{S}(\epsilon^\star)$ simplifies into $b = \frac{\alpha\beta(1-\epsilon_\star^\beta)}{(1+\alpha\beta)}$.

8.15 (Roberts and Rosenthal 1998) Show that, if the density f on $\mathbb{R}$ is log-concave, then $\omega\mu(\omega)'$ is non-increasing. (*Hint:* Show that, if $\mu^{-1}(\omega)$ is log-concave, then $\omega\mu(\omega)'$ is non-increasing.)

8.16 (Roberts and Rosenthal 1998) Show that, if the density f in $\mathbb{R}^d$ is such that

$$y\, D(y;\theta)^{d-1} \frac{\partial}{\partial y} D(y;\theta)$$

is non-increasing for all θ's and $D(y;\theta) = \sup\{t > 0; f(t\theta) \le y\}$, then $\omega\mu(\omega)'$ is non-increasing.

8.17 Examine whether or not the density on $\mathbb{R}_+$ defined as $f(u) \propto \exp -u^{1/d}$ is log-concave. Show that $f(x) \propto \exp -||x||$ is log-concave in any dimension d.

8.18 In the general strategy proposed by Neal (2003) and discussed in Note 8.5.1, $\mathscr{I}_t$ can be shrunk during a rejection sampling scheme as follows: if ξ is rejected as not belonging to $\mathscr{A}_t$, change L_t to ξ if $\xi < x^{(t)}$ and R_t to ξ if $\xi > x^{(t)}$. Show that this scheme produces an acceptable interval $\mathscr{I}_t$ at the end of the rejection scheme.

8.19 In the stepping-out procedure described in Note 8.5.1, show that choosing ω too small may result in an non-irreducible Markov chain when $A^{(t)}$ is not connected. Show that the doubling procedure avoids this difficulty due to the random choice of the side of doubling.

8.20 Again referring to Note 8.5.1, when using the same scale ω, compare the expansion rate of the intervals in the stepping-out and the doubling procedures, and show that the doubling procedure is faster.

8.5 Notes

8.5.1 Dealing with Difficult Slices

After introducing the term "slice sampling" in Neal (1997), Neal (2003) proposed improvements to the standard slice sampling algorithm [$A.31$]. In particular, given the frequent difficulty in generating exactly a uniform $\mathcal{U}_{A^{(t+1)}}$, he suggests to implement this slice sampler in a univariate setting by replacing the "slice" $A^{(t)}$ with an interval $\mathscr{I}_t = (L_t, R_t)$ that contains most of the slice. He imposes the condition on $\mathscr{I}_t$ that the set $\mathscr{A}_t$ of x's in $A^{(t)} \cap \mathscr{I}_t$ where the probability of constructing $\mathscr{I}_t$ starting from x is the same as the probability of constructing $\mathscr{I}_t$ starting from $x^{(t-1)}$ can be constructed easily. This property is essential to ensure that the stationary distribution is the uniform distribution on the slice $A^{(t)}$. (We refer the reader to Neal