Monte Carlo Methods, Fall 2007, Courant Institute, NYU
Homework 4, Due November 34

1. Verify that the "move the first of two rods" transition measure ((6) in the notes) satisfies the detailed balance condition ((14) in the notes). Part of the problem is writing an expression for $f(x_1, x_2)$, the uniform density subject to the constraints that the rods not overlap.

2. Develop the basic theory of the Metropolis Hastings algorithm for continuous random variables. Suppose that for each $x$, $T(x, y)$ is a probability density with respect to $y$. Suppose that $R(x, y) \in [0, 1]$ for all $x$ and $y$.

   (a) Write an expression for the transition "density", $p(x, y)$ that corresponds to proposing $y \sim T(x, y)$ and accepting with probability $R(x, y)$. This $p$ has a continuous density part and a $\delta-$function part.

   (b) What formula for $R$ leads to $P$ satisfying the detailed balance condition ((14) in the notes) with respect to a given density, $f$. This formula should be familiar.

3. Suppose $H(x)$ is an smooth energy function that depends on a continuous multi-component variable, $x$, and that we want to sample the Gibbs Boltzmann probability density

$$f(x) = \frac{1}{Z(T)} e^{-H(x)/kT} . \tag{1}$$

   One dynamic sampler is based on (misnamed) *Langevin* dynamics, which is the SDE

$$dX \;=\; -\nabla H(X)dt \;+\; \sqrt{2kT}dW . \tag{2}$$

   The intuition behind this is that the drift term tends to move $X$ toward smaller energy values while the noise term can increase or decrease the energy. The larger the noise coefficient, the less the system insists on keeping the energy small. The forward equation for the probability density, $f(x, t)$, of $X(t)$ is

$$\partial_t f \;=\; \nabla \cdot \left( \frac{b^2}{2} \nabla f - a(x) f \right) \;=\; L^* f , \tag{3}$$

   if $dX = a(X)dt + bdW$.

   (a) Find the operator $L$ so that

$$\langle L^* g, u \rangle_{L^2} \;=\; \langle g, Lu \rangle_{L^2}$$

   for all $g$ and $u$.

   (b) Show that the $L$ corresponding to (2) is self-adjoint with respect to the inner product $\langle u, v \rangle_f$, with $f$ the Gibbs-Boltzmann distribution (1). This shows that the dynamics (2) satisfies detailed balance with respect to the probability density (1). The noise coefficient $\sqrt{2kT}$ is needed to make this work.

(c) Verify directly that the probability density (1) is the steady state density for the process (2) because it satisfies $L^* f = 0$. Hint: Show the quantity in parentheses vanishes.

(d) The Euler approximation to (2) has a Gaussian transition probability, $T(x, y)$. Show that (1) is not the invariant probability density for the discrete time Markov chain with transition probability $T$. Hint: for smal $h$, you can calculate $(fT)(x) = f(x) + h^p g(x) + O(h^{p+1})$. The formula for $g$ involves $4^{\text{th}}$ derivatives of $f$, so it will may better not to write it in terms of $H$. All that is important here is that it is not equal to zero.

(e) Describe a Metropolis style algorithm for sampling (1) exactly that uses the Euler transition density $T$ as a trial transition probability for ejection.

4. Let $\vec{X} = (X_n, \ldots, X_n)$ with the $X_k$ distributed as in problem 7 of Homework 2, that is, independent with density $g(x) = C\sqrt{1 - x^2}$ subject to the constraint that $\sum X_k > ns$. Formally the density is (in awkward notation)

$$f(\vec{x}) \;=\; \frac{1}{Z(s,n)} \prod_k g(x_k) \mathbf{1}_{\sum x_k > ns} \;.$$

Cramer's theorem tells us that sampling $f$ by rejection from sequences of independent samples ($h(\vec{x}) = \prod_k g(x_k)$ is exponentially hopeless.

(a) Write a sampler for $f$ that works by sweeping through the components replacing each $X_k$ with an independent $X_k' \sim g$ and rejecting any component replacement that results in $\sum X_k < ns$.

(b) Explain why this simpler is correct.

(c) We are interested in the distribution of $S = \frac{1}{n}\sum X_k$ given that $S > s$. Use the sampler to make a histogram of the conditional $S$ for $s = .5$ and $n = 30$. Compare this to the behavior predicted by Cramer's theorem. Make sure to choose the horizontal axis of the histogram plot so that the plot fills the whole plot frame (i.e. does not look like a horizontal line or a right angle).

(d) Explore how the acceptance probability for your component resampler depends on $n$ and $s$. Hint: it should be more or less independent of $n$ once $n$ has any size, and it should deteriorate with increasing $s$ but not be terribly small except for extreme $s$.

(e) Make (and plot) a histogram of the conditional distribution of $X_1$ conditional on $S > s$, for $s = .5$ and $n = 30$ (and other values if interested). Compare this to the twisted distribution you used in Homework 2. They should agree. It is a simple consequence of Cramer theory that the conditional distribution is the twisted distribution. You will get a more accurate histogram if you use all the components of $\vec{X}$ to make it. This is allowed because they all have the same distribution.

(f) We want to estimate $E[S - s \mid S > s]$. Cramer's theorem (with our approximate integration technique) predicts that this is $O(1/n)$. Estimate it using the above sampler for interesting values of $s$ and $n$ (Debugging a code is hard. Playing with it is easy.)

(g) Estimate the auto-correlation time using the method of batched means and the self consistent window method ((31) in the notes). It should be fine to take $T = 10$, so just put in a check that that's OK. Keep debugging until the two estimates of $\tau$ should agree.

(h) Plot the estimated auto-corremation function, $\widehat{C}(t)$ and comment on its sign. This may be different for different values of $s$ and $n$. Look for interesting behavior. Comment on the relation of this to the results for part (g).