Assignment 3

Corrections: (none yet).

1. (*NGAR(1) processes.*) Suppose $\phi(z)$ is a smooth PDF with $E_\phi[Z] = 0$, and that $\{Z_k\}$ is a collection of independent samples $Z_k \sim \phi(z)$. Suppose $|a| < 1$ and consider the *non-Gaussian auto-regressive* process

$$X_{n+1} = aX_k + Z_k \ .$$

Assume that this process has a steady state PDF $f(x)$, which means that if $X_n \sim f$, then $X_{n+1} \sim f$. If X_n is not in the steady state distribution, write $X_n \sim f_n(x)$.

(a) The forward equation for this process has the form

$$f_{n+1}(x) = \int_{-\infty}^{\infty} f_n(y) L^*(y,x)\, dy \ .$$

Write a formula for $L^*(y,x)$ in terms of ϕ and a.

(b) The *generator* of this process (or any discrete time Markov process like this) is defined by the backward equation satisfied by expected values. If $V(x)$ is a suitable function, and

$$W(x) = \mathrm{E}[V(X_{n+1}) \mid X_n = x] \ ,$$

then

$$W(x) = \int_{-\infty}^{\infty} L(x,y) V(y)\, dy \ .$$

Use this to derive a formula for L in terms of ϕ and a. What is the relationship between L^* and L?

(c) Find the condition on ϕ that is equivalent to the process satisfying detailed balance, which means that L is self adjoint in the inner product:

$$\langle u, v\rangle = \int_{-\infty}^{\infty} u(x) v(x) f(x)\, dx \ .$$

Show that some NGAR(1) processes satisfy detailed balance and some do not.

(d) Find the eigenvalues and eigenfunctions of L^*. An eigenfunction/eigenvalue pair is g and λ with

$$\int g(y) L^*(y,x)\, dy = \lambda g(x) \ .$$

Hint: Use the Gaussian case as a model.

(e) Find the eigenvalues and eigenfunctions of L. An eigenfunction/eigenvalue pair is v and λ with

$$\int L(x,y)v(y)\,dy = \lambda v(x) \; .$$

Hint: The eigenvalues of L and L^* are related here as they are in the finite dimensional case. Show that the eigenfunctions of L and L^* satisfy a *biorthogonality* relation: if g is an eigenfunction of L^* with eigenvalue λ and v is an eigenfunction of L with eigenvalue μ, and if $\lambda \neq \mu$, then

$$\int g(x)v(x)\,dx = 0 \; .$$

2. The family of Student t distributions (5) in the Week 5 notes depends on three parameters. Study the Fisher information matrix for this family enough to determine that it is positive definite. Determine which entries of I are zero and which are non-zero.

3. Suppose $X \in \mathbb{R}^d$, with $X = (X_1, \ldots, X_d)$. Suppose there is a single variable probability density $g(y)$ and that the probability density of X is

$$f(x) = \begin{cases} \frac{1}{Z}\prod_{k=1}^d g(x_k) & \text{if } \sum_{k=1}^d X_k \geq A \\ 0 & \text{otherwise} \end{cases}$$

This is the density of d independent samples of g conditioned on the sum being larger than A. Suppose you have a direct sampler for g. Show that you can make this into an MCMC sampler for f by resampling the components X_k and rejecting when the constraint is violated. Does this satisfy detailed balance? *Extra credit, may be hard*: Let $C(t)$ be the auto-covariance function for $V_n = \sum_k X_k$. Show that it is possible to have $C(1) < 0$.

4. Write a program in our combination of C++ and Python that applies the isotropic Gaussian proposal (length scale r) Metropolis algorithm to a generic probability density $f(x)$ with $x \in \mathbb{R}^d$. The code should allow you to change f and d easily so you can explore the behavior of the method. The actual sampling should be done in C++, while the analysis can be done in Python. Use your judgement and initiative to decide precisely what to do, but it should include the following elements:

   (a) Verification: show that your sampler gets the right answer for a simple $d = 1$ problem (make a histogram, do a very long run to get rid of visible statistical scatter). Choose the target density f. Verify for $d > 1$ by making histograms of some scalar functions of the random variable x (for example, x_1 and $x_1^2 + x_2^2$).

   (b) Auto-correlations: compute (estimate) $C(t)$ for some functionals, particularly for some of the hard cases below.

(c) Apply your method to the highly anisotropic problem in 2D $f(x) = 0$ if $|x_1 - 3x_2| > \varepsilon$, and $f(x) = 0$ if $|x_1| > 1$, and $f(x)$ constant in the allowed region. Explore what relation between ε and r gives the best decay of $C(t)$.

(d) Try a "double well" problem such as $f(x) = \frac{1}{Z}e^{-\beta\phi(x)}$, where $\phi(x) = (x^2 - 1)^2$. The method will work poorly when β is large. Make some plots of the time series of X_n to see it switch from one well to the other. Choose β so that this happens only once, say, in about 1000 steps.