Assignment 1, due September 23

Corrections: (none yet)

1. Basics of multivariate normals:

   (a) Show that if H is a symmetric positive definite $n \times n$ matrix, then

   $$Z(H) \;=\; \int_{\mathbb{R}^n} e^{-x^t Hx/2}\, dx \;=\; \frac{(2\pi)^{n/2}}{\sqrt{\det H}}$$

   First, check that this is true if $H = I$ using $\int_{-\infty}^{\infty} e^{-x^2/2} dx = \sqrt{2\pi}$. There are several ways to finish. Choose one

   i. Show that if Q is an orthogonal matrix, then $Z(H) = Z(Q^t HQ)$ by substitution.

   ii. Let $H = LL^*$ be the Cholesky factorization and change variables $y = L^t x$.

   iii. Another trick.

   (b) If $X \in \mathbb{R}^n$ is a random variable with expected value $\mu = E[X]$, the covariance matrix of X is $C = E\left[(X-\mu)(X-\mu)^t\right]$ (just a reminder). Show that if the probability density of X is

   $$f(x) \;=\; \frac{1}{Z(H)} e^{-(x-\mu)^t H(x-\mu)/2} \,, \tag{1}$$

   then its covariance satisfies $C = H^{-1}$. We write $X \sim \mathcal{N}(\mu, C)$ if X has this probability density. Hint: make the trivial reduction to the case $\mu = 0$ then use one of the hints for part (a).

   (c) Suppose A is an invertable $n \times n$ matrix and that $X = AY$, $Y = A^{-1}X$. Use the change of variables formula for probability densities to write the probability density of Y. Conclude that Y also is multivariate normal.

   (d) Suppose that $X = (Y, Z)$ with $Y \in \mathbb{R}^j$, $Z \in \mathbb{R}^k$, and $n = j + k$. Write H in block matrix form

   $$H \;=\; \begin{pmatrix} A & B \\ B^t & C \end{pmatrix} \,,$$

   where A is symmetric $j \times j$, etc. Take $\mu = 0$. Let $g(y)$ be the marginal density of Y and $h(y \mid Z)$ to be the conditional density. By direct manipulation of (1) show that both g and h are Gaussian. For g, this involves integration over z. You can simplify this by completing the square.

(e) Use parts (c) and (d) above, together with the basic methods of linear algebra, to show that if $Y \in \mathbb{R}^m$ with $m \leq n$ and $Y = LX$ with L being an onto linear mapping $\mathbb{R}^n \xrightarrow{L} \mathbb{R}^m$, then Y is a multivariate normal.

(f) Now the easy part: suppose $X \sim \mathcal{N}(\mu_X, C_X)$, and $Y = LX$. Show that $Y \sim \mathcal{N}(\mu_Y, C_Y)$ (this was part (e)) and find the formulas for μ_Y and C_Y in terms of μ_X, C_X, and L.

(g) (nothing to hand in) There are other ways to get the basic properties of multivariate normals. One uses the Fourier transform (called characteristic function in probability)

$$\widehat{f}(p) = \int_{\mathbb{R}^n} e^{-ipx} f(x)dx = E\left[e^{ipX}\right] . \tag{2}$$

with the inversion formula

$$f(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{ipx} \widehat{f}(p)dp$$

It is easy to check (for $\mu = 0$) that X has the density (1) if and only if it has the characteristic function $\widehat{f}(p) = e^{-p^t Cp/2}$. From this, the linear transformation properties follow easily by substituting $Y = LX$ into (2). A still better way to verify that a linear transformation of a gaussian is gaussian is through the multivariate central limit theorem.

2. The discrete Laplacian in one dimension (Note the change of notation between the discrete and continuous version. In part (a), x is the independent variable indexed by j or k. In part (d), u is the independent variable indexed by x or y.):

(a) Consider the quadratic form

$$H(x) = \frac{1}{2}\left(x_1^2 + x_n^2 + \sum_{k=1}^{n-1}(x_{k+1} - x_k)^2\right) . \tag{3}$$

Find the symmetric matrix, H, so that $H(x) = \frac{1}{2}x^t Hx$. The double use of H as a function and a matrix is deliberate. To check your answer, see whether your H is tridiagonal (i.e. $H_{jk} = 0$ if $|j - k| > 1$), has all diagonals equal, and has all non-zero off diagonals equal. This matrix is the "negative of the discrete Laplace operator with Dirichlet boundary conditions in one dimension".

(b) Show that the matrix H from part (a) is positive definite, which means that $x^t Hx > 0$ if $x \neq 0$. Hint: $x^t Hx \geq 0$ is obvious from part (a). Why is it impossible that $H(x) = 0$ if $x \neq 0$?

(c) Define the difference operator D by $(Dx)_j = x_{j+1} - x_j$ (if $1 \leq j < n$). The second difference is $D^2 x = D(Dx)$, which is defined if $1 \leq x < n-2$. (a slight misuse of notation: D^2 is not exactly the square of D) Show that if $D^2 x_j = 0$ for a range of j values, then x_j is a linear function of j in that range. Find the relationship between H and D^2.

(d) The true Laplace operator in one dimension with Dirichlet boundary conditions at $x = 0$ and $x = L$ is $u \to Hu = -\partial_x^2 u$, (yet another use of H) if $u(x)$ is twice differentiable on $(0, L)$, and continuous on $[0, L]$ with $u(0) = u(L) = 0$. The "Green's function" for this H is a function $G(x, y)$ so that

   i. $G(0, y) = G(L, y) = 0$ for all $y \in (0, L)$,
   ii. $\partial_x^2 G(x, y) = 0$ if $x \neq y$,
   iii. $G(x, y)$ is a continuous function of $(x, y) \in [0, L] \times [0, L]$,
   iv. $[\partial_x G(y, y)] = \partial_x G(y + 0, y) - \partial_x G(y - 0, 0) = -1$. Here $\partial_x G(y + 0, y) = \lim_{x \downarrow y} \partial_x G(x, y)$, which exists because of (ii).

   Show that the unique function with these properties is $G(x, y) = \frac{1}{L} x(y - L)$ if $x \leq y$ and $G(x, y) = \frac{1}{L}(x - L)y$ if $x \geq y$.

(e) Show that if $Hu = f$ and f is continuous, then

$$u(x) \;=\; \int_{y=0}^{L} G(x, y) f(y) dy \; .$$

(f) Return to the discrete case and let G be the $n \times n$ symmetric matrix $G = H^{-1}$. Find a formula for G by imitating the continuous case described in part (d).

(g) If $X \in \mathbb{R}^n$ has the Gibbs Boltzmann probability density distribution with the quadratic energy function given by part (a):

$$f(x) \;=\; \frac{1}{Z} e^{-H(x)} \; , \tag{4}$$

   find a formula for the variances of the components $\sigma_j^2 = \mathrm{var}(X_j)$. Which X_j has the maximum variance? How does this maximum variance scale with n?

(h) (nothing to hand in) The discrete and continuous Laplace operator have much in common. In both cases there is a linear operator with boundary conditions and a Green's function that gives the covariance matrix. It is common informally to write the probability density in the continuous case as

$$f(u) \;=\; \frac{1}{Z} \exp\left(\frac{-1}{2} \int_0^L \left(\partial_x u(x)\right)^2 dx \right) \; .$$

This formula has problems, such as the fact that in this probability density, the exponent seems to be infinite almost surely. What is true is that the Dirichlet integral

$$H(u) \;=\; \frac{1}{2}\int_0^L \left(\partial_x u(x)\right)^2 dx$$

can be expressed in the form $H(u) = \frac{1}{2}\langle < u, Hu >$, where $\langle \cdot, \cdot \rangle$ is the L^2 inner product, and $Hu = -\partial_x^2 u$, assuming that $u(0) = u(L) = 0$.

3. Write a program to sample the density (4) using the heat bath/Gibbs sampler method. Use the following as a guide.
   (a) Suppose all the X_k are fixed except $k = j$, use (4) to figure out the conditional density of X_j.
   (b) Write a sampler for this density using standard normal random variables.
   (c) Make a histogram of the values of $X_{j,k}/\sigma_j$ for $j = n/2$ and compare it to the histogram of a standard normal. Scale the histograms properly so that when you put them on the same plot they agree, except for statistical noise in the empirical one.
   (d) Make a plot of $X_{j,k}/\sigma_j$ (component j of sample k) as a function of k for $j = n/2$. Do this for several values of n and see how the time to reach equilibrium increases with increasing n.
   (e) Compute the estimated autocovariance function for the time series $X_{j,k}$ with $j = n/2$. Do this for several n values to see that the autocorrelation time is large when n is large.

4. Repeat the steps of problem 3. but using a local Metropolis step instead of heat bath/Gibbs sampler. Use a uniform proposal $X_{j,k} \to X_{j,k} + U - \frac{1}{2}$, where U is a standard uniform random variable in the interval $[0, 1]$. This is partly a test of software. If the code for problem 3. is general enough, it should be very easy and quick to swap the samplers and let the computer do the runs over. The behavior should be similar, though the Metropolis will be a little slower.

5. (Not an action item unless you have extra time on your hands) Of course, there are easier ways to generate tied discrete Gaussian random walks, such as Cholesky. The purpose of problems 3. and 4. is to try out the general MCMC sampling methods in a hard case where you can check that it was correct. The methods of problems 3. and 4. apply to many other problems, such as the *Ising model*. This has the same energy function (3), but now the x_k may take only the values $+1$ or -1, corresponding to *spin up* or *spin down*.