Assignment 8.

Given May 6, due May 15.

Objective: To explore Gaussian random variables and variance reduction.
Warning: Monte–Carlo burns computer time. This assignment has been scaled to fit into a recent model Pentium computer or better. If it takes too long on your computer, scale it down a bit.

(1) The Box–Muller algorithm takes two uniform random variables and produces two independent standard normals, X and Y. We will test whether (X, Y) produced have the correct density function,

$$f(x, y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2} \quad , \tag{1}$$

using bins in two dimensions. Choose $h = \Delta x = \Delta y$, take $x_j = j \cdot j$ and $y_k = k \cdot h$, and define bins in the plane by:

$$B_{jk} = \left\{ (x, y) \mid |x - x_j| \leq \frac{h}{2} \text{ and } |y - y_k| \leq \frac{h}{2} \right\} \quad .$$

If we take n pairs, (X_t, Y_t), $t = 1, \ldots, n$, then the bin counts are

$$N_{jk} = \# \left\{ (X_t, Y_t) \in B_{jk} \mid 1 \leq t \leq n \right\} \quad .$$

The expected counts are

$$< N_{jk} >= n \iint_{B_{jk}} f(x, y) dx dy \quad , \tag{2}$$

where f is given by (1). The midpoint rule for the integral in (2) is second order accurate. In terms of local truncation error, this means that

$$\iint_{B_{jk}} f(x, y) dx dy = h^2 f(x_j, y_k) + O(h^4) \quad .$$

(This is different from the one dimensional case.) In two dimensions we do not want h so small because that would create too many bins. Therefore we want a more accurate integration formula. Show that

$$\iint_{B_{jk}} f(x, y) dx dy = h^2 f(x_j, y_k) + C h^4 \left(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \right) (x_j, y_k) + O(h^6) \quad . \tag{3}$$

Find the constant, C, to make this true. Choose h so that for $n = 10^6$, the central bin, B_{00}, has on the order of a thousand points in it. Do the bin counts agree with the theoretical prediction from (2) and (3)?

(2) Now call the independent Gaussians given by Box–Muller X_t. Box–Moller gives X_t, X_{t+1} from T_t, T_{t+1}. This a change of notation from problem 1, where X_{2t-1} and X_{2t} were called X_t and Y_t. Use the standard Monte–Carlo estimates

$$< X^{2p} >= \frac{1}{\sqrt{2\pi}} \int_{\infty}^{\infty} x^{2p} e^{-x^2/2} dx \approx \frac{1}{n} \sum_{t=1}^{n} X_t^{2p} \quad .$$

For $p = 1$, 2, and 5, how many samples are required to get the answer to within 1%? Why are more samples required for large p? It is not simply because $< X^{10} >= 945$ is larger than $< X^2 >$; we are looking for relative accuracy.

(3) A standard Brownian motion can be simulated by taking

$$X_{k+1} = X_k + \sqrt{\Delta t} Z_k \quad ,$$

where the Z_k are i.i.d. standard normals, $X_k \approx X(k\Delta t)$, and $X_0 = 0$. Here, take $\Delta t = .1$. We want to estimate

$$\mathcal{P} = \Pr\left(X_k \geq 3 \;\text{ for some }\; t_k \leq 1 \right) \quad .$$

(a) What accuracy to you get with 10^6 samples? Note that this uses 10^7 i.i.d. standard normals.

(b) Express $\mathcal{P}$ as

$$\mathcal{P} = \int \cdots \int \phi(z_1, \ldots, z_{10}) f(z_1, \ldots, z_{10}) dz_1 \cdots dz_{10} \quad ,$$

where

$$\phi = \begin{cases} 1 & \text{if } X_k \geq 3 \text{ for some } t_k \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

and

$$f(z_1, \ldots, z_{10}) = \frac{1}{(2\pi)^{10/2}} \exp\left\{ -(z_1^2 + \cdots + z_{10}^2)/2 \right\} \; ; .$$

Suppose instead we take Z_k to be i.i.d. Gaussians with mean a and variance 1. Show that this is equivalent to writing $\mathcal{P}$ in the form

$$\mathcal{P} = \int \cdots \int \phi(z_1, \ldots, z_{10}) \frac{f(z)}{g(z)} \cdot g(z_1, \ldots, z_{10}) dz_1 \cdots dz_{10} \quad ,$$

where g represents the joint density for ten independent biased gaussian random variables. Show that

$$\mathcal{P} \approx \frac{1}{n} \sum_{t=1}^{n} \phi(Z^{(t)}) \frac{f(Z^{(t)})}{g(Z^{(t)})} \quad , \tag{4}$$

where the $Z^{(t)}$ are each ten dimensional vectors.

(c) On the computer, show that (4) can estimate $\mathcal{P}$ to 1% accuracy with far smaller n if a is chosen well.