Assignment #1 (Due April 16)

1. (Portfolio rebalancing) The goal is to compare two investment strategies for a volatile stock (in an idealized setting in which, among other things, money is infinitely divisible). At the beginning of Day 1, Investor A buys \$1 worth of a stock and holds onto it until the beginning of Day 20. By the end of each day, the stock's value has either doubled or halved in value, with each possibility equally likely. Investor B, in contrast to the "buy and hold" strategy of Investor A, uses a "rebalancing" strategy that buys or sells stock each day to maintain a 50/50 balance between stock and cash. For example, at the beginning of Day 1, Investor B buys 50¢ worth of stock and keeps the remaining 50¢ in cash. If the stock doubles during the day, Investor B, who now has \$1.00 in stock and 50¢ in cash, sells 25¢ of stock so that, at the beginning of Day 2, she has 75¢ in stock and 75¢ in cash. (Thus, if the stock does well, she sells some of it and, if the stock does poorly, she buys more of it.) Denote by U and V the value of the total holdings of Investors A and B, respectively, at the beginning of Day 20.

   a) Write a simulation program to compute a point estimate and 95% confidence interval for E[U], based on 3000 simulation replications. Repeat this exercise for $E[V]$, using different random numbers so that your point and interval estimates for $E[V]$ are independent of those for $E[U]$.

   b) Combine your two confidence intervals from Part (a) to obtain an approximate 95% confidence interval for $E[V]-E[U]=E[V-U]$. [Hint: if X is approximately $\mathrm{Normal}(\mu_X,\sigma_X^2)$ and Y is independent of X and approximately $\mathrm{Normal}(\mu_Y,\sigma_Y^2)$, then $aX+bY$ is approximately $\mathrm{Normal}(a\mu_X+b\mu_Y,a^2\sigma_X^2+b^2\sigma_Y^2)$ for any real constants a and b.]

   c) Now compute a 95% confidence interval for $E[V-U]$ based on 3000 simulation replications. That is, for each replication, compute both V and U based on the *same* sequence of stock prices, and then subtract U from V. Is the resulting confidence interval wider or narrower than that of Part (b)? Give an intuitive explanation of the difference in widths.

   d) Using the simulation strategy of Part (c), estimate both $E[V-U]$ and $\Pr(V>U)$ to within ±10% with 95% probability. [To understand any seeming inconsistencies between these results, it is helpful to take a detailed look at the outcomes of 20 or so replications.]

   e) Now suppose that the degree of satisfaction from having \$$x$ is $\log_{10}(x)$—such a "logarithmic utility function" models the fact that, as an investor becomes richer, the marginal satisfaction from each additional dollar decreases (\$10 is worth much more to a beggar than to a billionaire). Similarly to Part (c), compute a point estimate and 95% confidence interval for $E[\log_{10}(V)-\log_{10}(U)]$ and for $E[\log_{10}(V)]$, each based on 3000 replications. Compare the first of these two results to Part (d). Using such utility functions is considered by some to yield more satisfying comparisons of investment alternatives.

   f) Using the results of Parts (a) and (e), compare $E[\log_{10}(V)]$ to $\log_{10}(E[V])$; which is larger? What well known probability inequality would lead you to expect this result? [Hint: It is mentioned in the *Probability and Statistics Refresher* document on the class website.]

2. (Monte Carlo integration) As mentioned in class, we can use simulation to solve deterministic numerical computation problems. Suppose in the following problems that we have available sequences $U_1, U_2, \ldots$ and $V_1, V_2, \ldots$ of i.i.d. uniform(0,1) random numbers.
   a) Suppose that we wish to evaluate the integral $I = \int_0^1 h(x)\,dx$, where the nonnegative function h is so complicated that analytical or numerical evaluation of the integral is impossible. Explain why we can estimate I by $\hat{I}_n = (1/n)\sum_{i=1}^n Z_i$, where $Z_i = h(U_i)$ for $i \ge 1$ and n is sufficiently large. How can we choose n?
   b) Now describe an estimation procedure as above, but for the integral $I = \int_a^b h(x)\,dx$, where a and b are specified numbers. [Hint: consider the transformation $y = (x-a)/(b-a)$.]
   c) Same as part (b), but for the integral $I = \int_0^\infty h(x)\,dx$. [Hint: consider the transformation $y = 1/(x+1)$.]
   d) Same as part (b), but for the integral $I = \int_0^1\int_0^1 e^{(x+y)^2}\,dx\,dy$ and using both sequences $U_1, U_2, \ldots$ and $V_1, V_2, \ldots$.
   e) Same as part (b), but for the integral $I = \int_0^\infty \int_0^x e^{-(x+y)^2}\,dy\,dx$. [Hint: using the function $g(x,y)$ which equals 1 if $y \le x$ and 0 otherwise, rewrite I so that the upper limit of the inner integral doesn't depend on x.]

3. (A probability refresher) Let X and Y be independent uniform(0,1) random variables. Compute and graph the probability density function (pdf) of $Z = X + Y$. [Hint: Use the law of total probability to write
$$P(Z \le z) = P(X+Y \le z) = \int_0^1 P(X+Y \le z \mid Y = y) f_Y(y)\,dy = \int_0^1 P(X \le z-y) f_Y(y)\,dy$$
and consider the cases $z \in [0,1]$ and $z \in [1,2]$ separately.]

4. (An alternative Monte Carlo integration method) **Note: do this problem individually**. Suppose that we want to evaluate the two-dimensional integral $I = \int_0^1\int_0^1 e^{-(x_1-0.5)^2-(x_2-0.5)^2}\,dx_1\,dx_2$ using Monte Carlo integration. A standard way to do this would be to generate N uniform number pairs $\mathbf{U}_1 = (U_{1,1}, U_{1,2}), \ldots, \mathbf{U}_N = (U_{N,1}, U_{N,2})$ and estimate I by $\hat{I}_N = (1/N)\sum_{i=1}^N h(\mathbf{U}_i)$. Here all $U_{i,j}$'s are mutually independent and $h(\mathbf{x}) = e^{-(x_1-0.5)^2-(x_2-0.5)^2}$ for $\mathbf{x} = (x_1, x_2)$. Now consider the following alternative algorithm. Choose K such that K divides N and set $n = N/K$. Then execute the following unbiased estimation procedure:
   A. For $i = 1$ to n
      1. Generate K uniform pairs $\mathbf{U}_1^{(i)}, \ldots, \mathbf{U}_K^{(i)}$.
      2. Generate two independent random permutations $\pi_1^{(i)}$ and $\pi_2^{(i)}$ of $(1, 2, \ldots, K)$. E.g., one possible realization of $\pi_1^{(i)}$ is given by $\pi_1^{(i)}(1) = K,\ \pi_1^{(i)}(2) = K-1,\ \ldots,\ \pi_1^{(i)}(K) = 1$.
      3. Define $\mathbf{V}_k^{(i)} = (V_{k,1}^{(i)}, V_{k,2}^{(i)})$ for $k = 1, 2, \ldots, K$ by setting $V_{k,1}^{(i)} = \left(\pi_1^{(i)}(k) - U_{k,1}^{(i)}\right)/K$ and $V_{k,2}^{(i)} = \left(\pi_2^{(i)}(k) - U_{k,2}^{(i)}\right)/K$.
      4. Set $Y_i = (1/K)\sum_{k=1}^K h(\mathbf{V}_k^{(i)})$.

B. Estimate I by $\hat{I} = (1/n)\sum_{i=1}^{n} Y_i = (1/N)\sum_{i=1}^{n}\sum_{k=1}^{K} h(\mathbf{V}_k^{(i)})$.

C. Compute a $100(1-\delta)\%$ confidence interval as $\hat{I} \pm z s_n / \sqrt{n}$, where z is the usual $(1-\delta/2)$-quantile of the standard normal distribution and s_n is the sample standard deviation of $Y_1, Y_2, \ldots, Y_n$.

a) Using the R, Matlab, or Octave software packages, estimate the standard error (i.e., sample standard deviation divided by square root of sample size) for the standard and alternative estimation methods described above, with $K = 5$, and $N = 150$. Which method is superior? [The file hw1code.txt on the course web site contains R code and Matlab/Octave code for doing the calculations, along with download instructions for the open source R and Octave packages. One point of this exercise is to introduce you to these useful packages for Monte Carlo simulation. If you already know them, you are encouraged to try writing the code yourself.]

b) Draw a sketch in which the unit square is divided into $K \times K$ square subregions (for $K = 5$). On this sketch, indicate a typical realization of the K two-dimensional points $\mathbf{V}_1^{(i)}, \mathbf{V}_2^{(i)}, \ldots, \mathbf{V}_K^{(i)}$. Using your drawing, give an intuitive explanation of why one of the estimation methods is better than the other.

c) For the alternative method, could we have estimated the standard deviation simply as the sample standard deviation of the N numbers $\{h(\mathbf{V}_k^{(i)}) : i = 1, 2, \ldots n \text{ and } k = 1, 2, \ldots, K\}$?

5. (Extra credit) Let $x_1, x_2, \ldots, x_n$ be a sequence of real numbers. Let $S_n = x_1 + x_2 + \cdots + x_n$ denote the sum, $\bar{X}_n = S_n / n$ the average, and $V_n = 1/(n-1)\sum_{i=1}^{n}(x_i - \bar{X}_n)^2$ the sample variance of the numbers. Establish the following numerically stable recursion formula for the sample variance:

$$(k-1)V_k = (k-2)V_{k-1} + \left(\frac{S_{k-1} - (k-1)x_k}{k}\right)\left(\frac{S_{k-1} - (k-1)x_k}{k-1}\right), \qquad k \ge 2.$$

The computer program for the gambling simulation sets $V_1 = 0$ and then uses this formula to compute the sample variance during a single pass through the data, i.e., without having to store the results of each simulation repetition. [Hint: First show that

$$(k-1)V_k = \sum_{i=1}^{k} x_i^2 - k\bar{X}_k^2 = \sum_{i=1}^{k-1} x_i^2 - k\left(\frac{S_{k-1}}{k} + \frac{x_k}{k}\right)^2 + x_k^2.$$

Then subtract and add $(k-1)\bar{X}_{k-1}^2 = S_{k-1}^2/(k-1)$ in the rightmost expression, and expand the 2nd term in this expression (i.e., the one that multiplies k).]