Mathematical modeling.

Assignment 7, due March 22.

1. Stirling's formula is the approximation

$$n! \approx \sqrt{2\pi n} \cdot n^n e^{-n} \ .$$

Use this to derive the approximation from problem 3 of homework 6.

2. The Poisson (pronounced "pwasson" with an "a" as in "father", an accent on the "a", and a French "on" if possible) process, $N(t)$, is the number of "arrivals" up to time t if the "interarrival times" are independent exponential random variables with rate λ. To be more precise, let S_1, S_2, …, S_k, …, be independent samples of the same exponential with rate λ. The probability density function for each of the S_k is $f(s) = \lambda e^{-\lambda s}$. Now the arrival times are $T_1 = S_1$, $T_2 = S_1 + S_2$, …, $T_k = S_1 + \cdots + S_k$. You can think of the Poisson process like this: first we wait for the first exponential, S_1. As soon as that happens, we start waiting for the second one, S_2, and so on. With this, we define $N(t)$ as the number of exponentials we have seen by time t. That is the same thing as

$$N(t) = \max \{ k \mid T_k < t \ \} \ .$$

For any particular t, $N(t)$ is a discrete random variable taking values 0, 1, and so on.

a. Find a formula for $f_0 = \Pr(N(t) = 0)$. This is not too hard because it only involves S_1.

b. In order to have $N(t) = 1$, we must have S_1 in the interval from 0 to t, and $S_2 > t - S_1$. Remember that S_2 is independent of S_1, so we can compute the probability that $S_2 > t - S_1$ in terms of t, S_1, and λ just thinking of S_1 as a parameter. Find an integral over s_1 involving the probability density for S_1 and the probability that $S_2 > t - S_1$ that represents $f_1 = \Pr(N(t) = 1)$. Work this integral to get a simple formula for f_1.

c. Use the same reasoning to get a formula for $f_k(t) = \Pr(N(t) = k)$ as an integral involving the density for S_k and $f_{k-1}(t - s_k)$. If you work this integral, you will find a formula for $f_k(t)$.

3. Suppose that X_1, …, X_L are independent random variables each uniformly distributed in the interval $[0, 1]$. Since there are L numbers uniformly (but randomly) sprinkled inside the unit interval, it makes sense that we have to wait about $1/L$ to get to the first (smallest) one. For that reason, define

$$T = \frac{1}{L} \min_j X_j \ . \qquad (1)$$

Use the approximation

$$\left(1 - \frac{a}{L}\right)^L \approx e^{-a}$$

to show that T is approximately an exponential random variable when L is large.

Extra credit challenge: Let T_1 be the T defined in (1) Define T_n to be the n^{th} smallest of the X_j, again normalized with the factor $1/L$. Show that when L is large and n is not too large, the T_n are approximately a Poisson process. You can do this by calculating the probabilities $f_n(t) = \Pr(N(t) = n)$.

4. This exercise asks you to verify that the standard normal random numbers produced by Matlab have the right probability density. For this, we choose a "bin size", Δx, and define the "bin centers" $x_j = j\Delta x$. The j^{th} bin, B_j, is the interval of length Δx centered at x_j. A number, X, is in bin j if it is closer to x_j than to any of the other bin centers. The probability that a particular X lands in bin B_j is approximately $f_j = f(x_j)\Delta x$. Therefore, if I generate n independent X_i, the number of them landing in B_j is approximately $n \cdot f_j = n \cdot f(x_j)\Delta x$. Use Matlab to generate n independent standard normal random variables and put them into bins. It is extremely unlikely that a standard normal could be as large as 6, so make the bin centers run from about -6 to 6. Choose Δx something like .1 or .2. Choose a range of n values with the largest being the largest you computer can handle. Make a plot of the actual bin counts and the expected bin counts. Comment on the results. There are some hints on how to do this posted with this assignment. You should read them even if you are not using Matlab. If your system does not have a source of standard normal random variables, it certainly has a source of uniformly distributed random variables. Call them Y_i, and generate exponential random variables using $X_i = -\log(Y_i)$.