Math 319: Applied Probability and Stochastic Processes for Biology
Fall 2008. Third Assignment.

Instructor: Peter Thomas, Yost 209 / pjt9@case.edu

Due Friday 10/3/2008 to my mailbox (Yost Hall) or office (Yost 210) by noon.

1. **The exponential distribution has "no memory"** in the sense that if X is exponentially distributed with parameter λ then

$$\Pr\{X > (s+t)|X > t\} = \Pr\{X > s\}.$$

   In other words, if the length of time that customer service keeps people on hold is exponentially distributed, then the fact that you've already waited t hours to speak to someone doesn't improve your chances of speaking to anyone within the next s hours. Grim thought.

   Show that the geometric distribution has the same property. Recall that if $X \sim \text{geom}(p)$ then

$$\Pr\{X = k\} = (1-p)^{k-1}p.$$

   Prove that

$$\Pr\{X > (k+l)|X > l\} = \Pr\{X > k\}.$$

   Thus if success or failure is a series of Bernoulli trials, then the fact that something has worked the last l times does not increase the chance of it working the next k times.

2. Here is **another maximum likelihood problem**, more closely related to a real biological situation. Assume that with probability p, a randomly selected individual in a given population is infected with a parasite (say malaria). If an individual is infected, then the individual will have an amount X of the parasite in the blood, where X is exponentially distributed with unknown parameter λ. Suppose that N individuals are examined from the population and that n of them have positive levels of the parasite, measured to be $\{x_1 > 0, \cdots, x_n > 0\}$. The remaining $N-n$ individuals have measurements $\{x_{n+1} = 0, \cdots, x_N = 0\}$. Find an expression for maximum likelihood estimates of the two unknown parameters p and λ, in terms of N, n and the values $\{x_1, \cdots x_n\}$.

   Extra credit problem, not for the faint of heart. Suppose that the x_i cannot be measured directly, but are themselves subject to measurement noise that adds a random quantity to each, giving observed variables $y_i = x_i + \eta_i$, where the η_i are exponentially distributed with known mean μ. Find expressions for the unknowns p, λ and n (the number actually infected) in terms of the y_i.

3. **Correlation does not imply independence.**

(a) Let X and Y be two continuous random variables with joint density $\rho(x, y)$. Assume that both random variables have well defined densities $\rho_x(x) = \int_y \rho(x, y)\, dy$, $\rho_y(y) = \int_x \rho(x, y)\, dx$, and that all first and second moments exist (*i.e.* the means $\mathbb{E}[x], \mathbb{E}[y]$ and the (co)variances $\mathbb{E}[x^2], \mathbb{E}[xy], \mathbb{E}[y^2]$ all exist and are finite). Show that if X and Y are independent (*i.e.* $\rho(x, y) = \rho_x(x)\rho_y(y)$) then the two variables are uncorrelated.

(b) Now consider two random variables X, Y with joint distribution equal to

$$\rho(x, y) = \begin{cases} 1/\pi, & x^2 + y^2 \leq 1 \\ 0, & x^2 + y^2 > 1 \end{cases}$$

i.e. they are uniformly distributed over the unit circle.

i. Show that $\mathbb{E}[X] = \mathbb{E}[Y] = 0$.

ii. Show that $\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = 0$.

iii. Show that X and Y are not independent.

4. **Floating point numbers in matlab** (Numerical Computing in Matlab problem 1.33). Modify the m-file `floatgui.m` (available in the NCM code set) by changing its last line from a comment to an executable statement and changing the question mark to a simple expression that counts the number of floating-point numbers in the model system.

5. **More on floating point numbers.** NCM problem 1.35. What does each of these programs do? How many lines of output does each program produce? What are the last two values of `x` printed in terms of `eps`, `realmin` and `realmax`?

(a) `x=1; while 1+x>1,x=x/2,end`

(b) `x=1; while x+x>x,x=2*x,end`

(c) `x=1; while x+x>x,x=x/2,end`

Late homework policy: In general, homework that is n weeks late will be accepted, but its total score will be multiplied by p^n, where p is a random variable drawn from the uniform distribution on the interval $[0, 1]$. For example, if an assignment is one week late the average penalty is 50%. But it could be as high as 100 % or as low as nothing.