BIOL 701 Likelihood Methods in Biology: Homework #2

Due Wednesday, February 23, 2011.

You may work with others but the work you turn in must be your own. Make an earnest effort to solve all problems on your own.

1). For a sample of size n from a population with an exponential p.d.f., $\lambda e^{-\lambda t}, t > 0$, write down the likelihood and compute the maximum likelihood estimate of λ. Be sure to show that it is a maximum.

2). Two groups of probabilists develop probability models for random variables with binary outcomes. The first group derives the following p.f. $P(S_n = k|p) = \binom{N}{k} p^k (1-p)^{N-k}$ and the second group derives this p.f. $P(U = k|p) = \binom{N-1}{k-1} p^k (1-p)^{N-k}$. How do you interpret S_n and U? Both groups observe the same 20 trials and thus the same sequence of outcomes, (0,1,0,0,0,1,0,1,0,1,1,0,0,1,0,0,0,1,0,1). Both groups are interested in making inference about p. Will they reach the same conclusion about the "true" value of p? Justify your answers with the necessary calculations. (Hint: it's important that the last trial is a success, at least to one group).

3). For a random sample of size n from $\mathcal{N}(\mu, 1)$, write down the likelihood and compute the maximum likelihood estimate for μ, $\hat{\mu}$. Derive the likelihood ratio test statistic for testing $\mu \neq \hat{\mu}$. Show that in this case, $-2ln\Lambda$ can be expressed as the square of a Z-score. (This means that the distribution of $-2ln\Lambda$ is exactly $\chi^2(1)$ because if $X \sim \mathcal{N}(0,1), X^2 \sim \chi^2(1)$).

4). Consider a random sample from $\mathcal{U}(\theta - \frac{1}{2}, \theta + \frac{1}{2})$. Write the likelihood function and determine all the maximum likelihood states.

5). For any random sample we can order the observations from smallest $X_{(1)}$, to largest $X_{(n)}$, i.e. $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$. These are known as order statistics. The CDF of the largest of n observations in a random sample from a population with $F(x)$, the nth order statistic, is $F_{X_{(n)}}(v) = P(X_{(n)} \leq v) = P(X_{(1)} \leq v, X_{(2)} \leq v, \ldots, X_{(n)} \leq v) = [F(v)]^n$. We differentiate $F_{X_{(n)}}(v)$ to obtain the p.d.f. $f_{X_{(n)}}(v) = n[F(v)]^{n-1} f(v)$. Suppose you collect a sample of n bacteria primed for apoptosis, programmed cell death, with time to death distributed as $Exp(\lambda)$. Compute the pdf for the first death, considering it as the first order statistic or the minimum value in a random sample. (Hint: use the fact that $P(X > v) = 1 - P(X \leq v)$).

6). The empirical distribution function or empirical CDF is defined as $F_n(x) = \frac{\#\ of\ elements\ in\ sample \leq x}{n}$ Use a piece of software, preferably one that doesn't cost money, to generate 10, 20, and 100 Binomial (75,.4) random variates and plot the empirical CDF for each on the same figure. Also plot the CDF of the normal distribution with corresponding mean and variance on the same figure. How do they compare? If you use R, the following commands will be of use: `rbinom(10,75,.4)`, `plot(...,type='s')`, `pnorm()`, and `legend()`.