Assignment #6 — Due Wednesday, March 4, 2009, by 5:00 P.M.

Turn in homework in lecture, discussion, or your TA's mailbox. Indicate the discussion section in which you expect to attend to pick up this assignment on the assignment.

311: Monday 1:20–2:10 **312:** Monday 12:05–12:55

1. The file `wi-tornado.txt` contains the total number of tornadoes in Wisconsin for each year from 1950 to 1995. Use the bootstrap to find a 95% confidence interval for the upper quartile (the 0.75 quantile) of the distribution of the number of tornadoes in Wisconsin per year during a time period with comparable climate. You will need to download the R package `boot` and follow examples from lecture and the file `stat310.R` to complete this problem.

2. The following set of problems have you explore different methods for computing confidence intervals from exponential samples. Suppose that $X_1, \ldots, X_n \sim$ i.i.d Exponential(θ). Recall that the MLE is $\hat{\theta} = 1/\bar{x}$ and that the distribution of $\sum_{i=1}^n X_i$ is Gamma(n, θ). For concreteness, consider a sample where $n = 20$ and $\bar{x} = 4.2$ and find a numerical confidence interval for each method below.

   (a) **Exact Confidence Interval:** Show that $Y = \theta\bar{X} \sim$ Gamma(n, n), a distribution which does not depend on θ. (Recall from last semester if $Y = h(X)$, then $f_Y(y) = f_X(h^{-1}(y))/|h'(h^{-1}(y))|$ when h is a differentiable one-to-one function, as in section 2.6.) If a and b are the 0.025 and 0.975 quantiles of the Gamma(n, n) distribution, then $\mathsf{P}\left(a < \theta\bar{X} < b\right) = 0.95$. Solve this relationship to find a method for computing a confidence interval with exact 0.95 coverage probability.

   (b) **Confidence Interval using MSE:** The MSE of $\hat{\theta}$ is $\theta^2(n+2)/((n-1)(n-2))$. Find a confidence interval of the form $\hat{\theta} \pm z\sqrt{\text{MSE}(\hat{\theta})}$.

   (c) **Confidence Interval Using Fisher Information:** Find an expression for the Fisher Information for the sample. Estimate it with both the observed Fisher Information and the plug-in estimate. Construct a confidence interval of the form $\hat{\theta} \pm z/\sqrt{nI(\hat{\theta})}$.

   (d) **Confidence Interval using Unbiased Estimator:** Verify that $\mathsf{E}(\hat{\theta}) = (n/(n-1))\theta$, which implies that $\tilde{\theta} = ((n-1)/n)\hat{\theta} = (n-1)/(n\bar{x})$ is an unbiased estimate of θ. Construct a confidence interval of the form $\tilde{\theta} \pm z\sqrt{\text{MSE}(\tilde{\theta})}$.

   (e) For the numerical example, which of these four confidence intervals is the shortest (has the smallest range — the right endpoint minus the left endpoint).

3. Use R and `exponential.sim()` in the file `exponential.R` to simulate 10,000 samples of size 5 from an Exponential(1) distribution. For each sample, compute the coverage probability for each of the methods in the previous problem, and compute the average range of the interval for each method. Are the actual coverage probabilities close to 0.95 for each method? Briefly explain.

4. Repeat the previous problem using a sample size of 100. (This may take a little while.) How do the results from this larger sample compare to the results from the previous problem with $n = 5$?

Work to do, but not turn in.

- Read sections 7.1–7.3.