Ma 322: Biostatistics
Homework Assignment 7

Prof. Wickerhauser

Due Friday, March 23rd, 2012

Read Chapter 13, "Foundations of Statistical Inference," pages 217–239 of our text.

1. Plot the χ^2 densities with 2, 3, and 7 degrees of freedom, over the interval $[0, 10]$.

2. Plot the F densities with every pair of numerator, denominator degrees of freedom chosen from the list 3, 10, 50, over the interval $[0, 4]$. (Hint: modify the code on page 227 of our text.)

3. This problem will illustrate the Central Limit Theorem. Let X be a random variable taking real values $0 \le x < \infty$ with exponential pdf $f(x) = \frac{1}{10}e^{-x/10}$. This has random sample function `rexp(rate=1/10)` in `R`. Note how little this pdf resembles the bell-shaped curve e^{-x^2} of the normal density.

   (a) What is the mean μ of X?

   (b) What is the variance σ^2 of X?

   (c) Fix $n = 5$ and $m = 50$. Generate m vectors $\{X_i : i = 1, \dots, m\}$ of n random samples $X_i(1), \dots, X_i(n)$ of X and form m normalized averages

   $$\bar{X}_i \stackrel{\text{def}}{=} \frac{S_i - n\mu}{\sigma\sqrt{n}}, \qquad i = 1, \dots, m.$$

   where $S_i = \sum_{k=1}^{n} X_i(k)$, and μ and σ are from parts and b. Plot the histogram of $\bar{X}_i$.

   (d) Repeat part c with $n = 100$ and $m = 1000$.

4. This problem will illustrate the Strong Law of Large Numbers. Let X be a random variable taking real values $-\infty < x < \infty$ with pdf $f(x) = \frac{1}{\pi(1+x^2)}$. Recall from HW6,Ex1 that X has the same pdf as $-1/\tan(\pi U)$ where U is a uniform random variable on $[0, 1]$.

   (a) What is the median m of X?

   (b) Show that the mean μ is undefined and the variance σ^2 of X is infinite.

   (c) Fix $n = 5$ and $m = 50$. Generate m vectors $\{X_i : i = 1, \dots, m\}$ of n random samples $X_i(1), \dots, X_i(n)$ of X and form m averages

   $$\bar{X}_i \stackrel{\text{def}}{=} \frac{1}{n}\sum_{k=1}^{n} X_i(k), \qquad i = 1, \dots, m.$$

   Plot the histogram of $\bar{X}_i$.

   (d) Repeat part c with $n = 1000$ and $m = 1000$.

5. Alleles A and a are present in a population in unknown proportions p and $1 - p$. Assuming a Hardy-Weinberg equilibrium distribution of the resulting diploid genotypes, find the maximum likelihood estimator for p given the following experimental results:

Genotype Count Data for One Allele

Genotype	Count Data	Variable
AA	193	n_{AA}
Aa	216	n_{Aa}
aa	889	n_{aa}

6. Following are 14 samples from a normal population with mean μ and unknown standard deviation σ:

```
2.59 2.67 2.16 1.95 2.61 1.11 2.62 2.06 2.06 1.66 2.16 3.35 2.46 2.55
```

   (a) Compute an estimate for σ.

   (b) Compute an estimate for μ.

   (c) Find the 95% confidence interval for μ derived from the 14 samples.

   (d) Find the 50% confidence interval for μ derived from the 14 samples.

7. This problem will illustrate nonparametric bootstrap estimation of sample variability. First generate a 200 sample data set as follows:

```
set.seed(12345); data<- c(rnorm(90,mean=3,sd=2), rexp(110,rate=1));
```

   (a) Plot the histogram of `data`.

   (b) Find the mean and standard deviation of `data`.

   (b') Estimate the "standard error" of a 200-sample mean by $s/\sqrt{200}$ using the standard deviation from part b.

   (c) Find the median and the 1st and 3rd quartile values of `data`.

   Now apply the bootstrap method: generate 100 replications of 200 samples of `data`, with replacement, and calculate their means and medians.

   (d) Calculate the mean and standard deviation of the 100 bootstrap means.

   (d') Which is bigger, the bootstrap standard deviation of the means, or the "standard error" from part b'?

   (e) Calculate the median and the 1st and 3rd quartile values of the 100 bootstrap medians.

   (e') Compute the ratio of the differences between the 3rd and 1st quartiles for the bootstrap medians and the original data.