

Ma 322: Biostatistics

Homework Assignment 7

Prof. Wickerhauser

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

1. Plot the F densities with every pair of numerator, denominator degrees of freedom chosen from the list 3, 10, 50, over the interval $[0, 4]$. Arrange the graphs into a 3×3 grid. (Hint: modify the code on page 227 of our text.)
2. This problem will illustrate the Central Limit Theorem. Let X be a random variable taking real values $x \in [-1, 0] \cup [1, 2]$ with uniform pdf

$$f(x) = \begin{cases} 1/2, & \text{if } -1 \leq x \leq 0 \text{ or } 1 \leq x \leq 2; \\ 0, & \text{otherwise.} \end{cases}$$

- (a) Generate N samples from this pdf using `runif(N)+sample(c(-1,1),N,replace=TRUE)`. Do this with $N = 500$ and plot the histogram to see how little this pdf resembles the bell-shaped curve e^{-x^2} of the normal density.
- (b) What is the exact mean μ of X ? (Hint: do not use R or Calculus.)
- (c) What is the exact variance σ^2 of X ? (Hint: use Calculus.)
- (d) Fix $n = 3$ and $m = 200$. 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

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

where $S_i = \sum_{k=1}^n X_i(k)$, and μ and σ are from parts b and c. Plot the histogram of \bar{X}_i and the quantile-quantile plot `qqnorm()` against the normal pdf.

- (e) Repeat part d with $n = 50$ and $m = 200$.
3. 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	314	n_{AA}
Aa	531	n_{Aa}
aa	289	n_{aa}

4. Following are some samples from a population with unknown (but finite) mean μ and standard deviation σ :

6.92 11.9 8.94 3.18 10.3 9.90 9.22 5.61 6.73 6.66 9.86 5.50 8.53 5.46 4.95

- (a) Compute an estimate for σ .
 - (b) Compute an estimate for μ .
 - (c) Find the median of the samples.
 - (d) Find the quartile deviation of the samples.
5. This problem will illustrate nonparametric bootstrap estimation of sample variability. First, let MYSID be your student ID number and generate a 200 sample data set as follows:

```
set.seed(MYSID); 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.