STAT 269 – Winter 2009

Homework Assignment 1

Assignment 1 — due Friday, January 16 (in the class or in the mail-box for office 511 during the day)

Problem numbers refer to Freund's textbook.

1. Suppose a random sample $x_1, ..., x_n$ is drawn from the *exponential* distribution, see **Definition 6.3**, with an unknown parameter θ.

   a) What is the expected value μ corresponding to this distribution?

   b) What is the variance σ^2 of this distribution?

   c) Find an unbiased estimator $\hat{\theta}$ of the parameter θ.

   d) What is the variance of $\hat{\theta}$?

   e) What is the *mean square error* of $\hat{\theta}$?

   f) Is your estimator *consistent*? Explain.

2. Repeat Problem 1 when the distribution of the sample has density

$$f(x|\theta) = \begin{cases} \frac{x}{\theta^2} e^{-x/\theta} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

3. When $\hat{\theta}$ is not an *unbiased* estimator of a parameter θ, the extent of its bias is measured by $E(\hat{\theta}) - \theta$. Typically, this would depend on the parameter θ. It is called the *bias* of $\hat{\theta}$ and is denoted

$$b(\theta) = E(\hat{\theta}) - \theta.$$

   Show that the mean square error admits the following *variance-bias decomposition:*

$$E(\hat{\theta} - \theta)^2 = \text{var}(\hat{\theta}) + b^2(\theta).$$

4. Let $X_1, ...X_n$ be a random sample from a normal distribution with both parameters μ and σ^2 unknown. You have learned in the class that

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

   is an unbiased estimator of the variance σ^2. But can its *mean square error* be improved? In this problem you will answer this question. Consider a more general estimator of σ^2 of the form

$$\bar{\sigma}^2 = \varrho S^2$$

   where $\varrho > 0$ is some number.

a) What is the bias of $\bar{\sigma}^2$?

b) What is variance of this estimator?

c) What is its mean square error?

d) Find the value of ϱ for which the mean square error of $\bar{\sigma}^2$ is the smallest.

Hint: Problem 3 and Theorem 8.11(2).

5. **(Optional)** Let $X_1, X_2, ..., X_n$ be independent random variables with an unknown cumulative distribution function $F(x)$. Denote

$$\mathbf{I}(x) = \begin{cases} 0 & , \quad \text{if } x < 0, \\ 1 & , \quad \text{if } x \geq 0. \end{cases}$$

The *empirical cumulative distribution function (ecdf)* can be defined as

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{I}(x - X_i).$$

Prove that for any x and y,

a) $\hat{F}_n(x)$ is an unbiased estimator of $F(x)$;

b) $\text{var}(\hat{F}_n(x)) = \frac{1}{n} F(x)(1 - F(x))$;

c) $E(\hat{F}_n(x) - F(x))(\hat{F}_n(y) - F(y)) = \frac{1}{n}(F(\min(x, y)) - F(x)F(y))$;

d) $E(\hat{F}_n(x) - F(x))(\hat{F}_n(y) - F(y)) \geq 0$;

e) Conclude that $F_n(x)$ and $F_n(y)$ are positively correlated: if $F_n(x)$ overshoots $F(x)$, then $F_n(y)$ will tend to overshoot $F(y)$.

6. **Hands on R**

a) Generate a random sample of size $n = 50$ from the standard normal distribution.

b) Plot the *empirical cumulative distribution function* (ecdf) of the sample.

c) Superpose on the existing graph a plot of the actual cumulative distribution function (cdf) of the sample.

d) Add a title to your graph.

e) Comment on the visual accuracy of ecdf as an estimator of the true cdf.

f) Does your plot support the conclusion of the Problem 5e)?

Hint:

```
n=50
X=rnorm(n)
x=c(min(X)-1,sort(X),max(X)+1)
ECDF=c(0,(1:n)/n,1)
plot(x,ECDF,type="s")
lines(x,pnorm(x))
title("Empirical CDF vz Normal CDF")
```