Problem Set #3: Statistics

Economics 435: Quantitative Methods

Fall 2011

1 Averages

Let $x_1, x_2, \ldots, x_n$ be a random sample of size n on the random variable x. Let $\mu = E(x)$ and $\sigma^2 = var(x)$. Let $a_1, a_2, \ldots, a_n$ be a sequence of n numbers.

a) Let:

$$W_n = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$$

Find a condition on the a's that imply that W_n is an unbiased estimator of μ.

b) Find $var(W_n)$ as a function of the a's and of σ^2.

c) For any set of numbers $a_1, a_2, \ldots, a_n$ it is the case that:

$$\frac{(a_1 + a_2 + \cdots + a_n)^2}{n} \leq a_1^2 + a_2^2 + \cdots + a_n^2$$

Use this to show that if W_n is an unbiased estimator of μ, then $var(W_n) \geq var(\bar{X}_n)$, where $\bar{X}_n$ is the usual sample average.

2 Consistency and unbiasedness

Let x be a Bernouilli random variable with $0 < p < 1$, where $p \equiv \Pr(x = 1)$. Suppose we have a random sample of size n on x, and that $\bar{x}$ is the usual sample average.

a) Is $\bar{x}$ an unbiased estimator of p? Prove it.

b) Is $\bar{x}$ a consistent estimator of p? Prove it.

c) Suppose we are interested in estimating $\ln p$ rather than p. Our candidate estimator is $\ln \bar{x}$. Is $\ln \bar{x}$ an unbiased estimator of $\ln p$? Prove it.

d) Is $\ln \bar{x}$ a consistent estimator of $\ln p$? Prove it.

3 Estimating the covariance

Let (x, y) be a pair of random variables, and let $\{(x_i, y_i)\}_{i=1}^n$ be a random sample of size n on (x, y). The covariance of x and y is defined as:

$$cov(x, y) = E\left[(x - E(x))\,(y - E(y))\right]$$

a) Use the analog principle to define a statistic called $\hat{cov}(x, y)$ that is a consistent estimator for $cov(x, y)$.

Don't just look it up - I want to see you use the analog principle here. You may (or may not) find it useful to remember that:

$$cov(x, y) = E(xy) - E(x)E(y)$$

b) Prove that your estimator is consistent.

c) Find the expected value of your estimator. Is it biased or unbiased?

d) If it is biased, is there a way to correct the bias in a way that preserves consistency?

4 Basic statistics in R

Write an R script that does the following:

1. Generate a 1000-by-100 matrix of Bernoulli(0.75) random variables. You can use the `rbern` function you have already made. You do do not need to print this matrix out!
2. Generate the following four vectors of length 100.
   (a) For each column, the average of all 1,000 observations in that column.
   (b) For each column, the average of the first 25 observations in that column.
   (c) For each column, the average of the first 5 observations in that column.
   (d) For each column, the average of the first 1 observation in that column.

   You do not need to print these vectors.
3. Plot a histogram for each of these four vectors. Use the R function `hist`. You can get all four graphs on a single page by executing the R command `par(mfrow=c(2,2))` before generating the histograms.

Include a printout of the histograms in your assignment, and submit the R script by WebCT.

You should note that we have just used simulation to illustrate the law of large numbers and the central limit theorem. The law of large numbers says that the histogram should show the distribution narrowing to a small range around 0.75 as the number of observations goes up. The central limit theorem says that the distribution should start looking more and more like a normal distribution as the number of observations goes up.