Stuff we should know (in class 11/6/2000)

C. Rose

November 6, 2000

1 Expected Value of a Simple Variance Estimator

In class we derived a simple estimate for the mean of a random variable based on independent observations x_i as

$$\hat{\bar{x}} = \frac{1}{N}\sum_i x_i$$

also known as the sample mean. To estimate the variance we formed an innocent looking

$$\hat{\sigma_x^2} = \frac{1}{N}\sum_i (x_i - \hat{\bar{x}})^2$$

and then looked at the expected value of this estimate (and algebra reared its head). So let's do it again.

$$E[\hat{\sigma_x^2}] = \frac{1}{N}\sum_i (E[X_i^2] - 2E[X_i\hat{\bar{x}}] + E[\hat{\bar{x}}^2])$$

and further expanding we get after defining $E[X] = \eta$

$$E[\hat{\sigma_x^2}] = E[X^2] - \frac{2}{N^2}\left(N(N-1)\eta^2 + NE[X^2]\right) + \frac{1}{N^2}\left(N(N-1)\eta^2 + NE[X^2]\right)$$

which reduces to

$$E[\hat{\sigma_x^2}] = E[X^2] - \frac{1}{N^2}\left(N(N-1)\eta^2 + NE[X^2]\right) = \frac{N-1}{N}\left(E[X^2] - \eta^2\right) = \frac{N-1}{N}\sigma_x^2$$

Our problem in class stemmed from not recognizing that the final term in the sum was not a function of i (so we had an N^3 on the bottom).

PUNCHLINE: simple variance estimator is biased, but can correct simply by setting

$$\hat{\sigma_x^2} = \frac{1}{N-1}\sum_i (x_i - \hat{\bar{x}})^2$$

Another interpretation which I may or may not have mentioned is that by using a quantity estimated with the same data ($\hat{\bar{x}}$), we've effectively lost one of our observations (a degree of freedom). That is, instead of basing our estimate for the variance on all N terms, we've been reduced to $N-1$ terms.