MATHEMATICS E-156, SPRING 2014
Mathematical Foundations of Statistical Software

Date: February 6, 2014
Name: Susan Coombs
Proof of the Week #2:

Theorem: Let $X_1, X_2, \cdots X_n$ be independent random variables from a distribution with $\mathrm{Var}[X_i] = \sigma^2 < \infty$.

We do not know the expectation μ, although we know from the previous result that the expectation of $\overline{X}$ is equal to μ.

We also do not know the variance. We try to estimate it by using the usual formula but, not knowing μ, we can do no better than to use $\overline{X}$ in its place.

$$\text{Prove that } E[\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2] = \frac{n-1}{n}\sigma^2.$$

It follows that $S^2 = E[\dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i - \overline{X})^2] = \sigma^2$. This is what var() computes.

Proof:

1. By algebra:

$$\begin{aligned}
\sum_{i=1}^{n}(X_i - \overline{X})^2 &= \sum_{i=1}^{n}(X_i^2 - 2\overline{X}X_i + \overline{X}^2) \\
&= \sum_{i=1}^{n} X_i^2 - 2\overline{X}\sum_{i=1}^{n} X_i + \overline{X}^2\sum_{i=1}^{n} 1 \\
&= \sum_{i=1}^{n} X_i^2 - 2\overline{X}(n\overline{X}) + n\overline{X}^2 \\
&= \sum_{i=1}^{n} X_i^2 - n\overline{X}^2
\end{aligned}$$

2. Take expectation; by linearity (also by $E[a_1X_1 + a_2X_2] = a_1E[X_1] + a_2E[X_2]$ in all cases, from class 1 Proof of the Week on p.3):

$$E[\sum_{i=1}^{n}(X_i - \overline{X})^2] = E[\sum_{i=1}^{n} X_i^2] - nE[\overline{X}^2]$$

3. Now, by class 1 Proof of the Week:

$$\mathrm{V}ar[X_i] = E[X_i^2] - (E[X_i])^2$$

so by definition:

$$\sigma^2 = E[X_i^2] - \mu^2$$

and:

$$\mathrm{V}ar[\overline{X}] = E[\overline{X}^2] - (E[\overline{X}])^2$$

Using the mathematical notes for class 2, $Var[\overline{X}] = \frac{1}{n}\sigma^2$:

$$\frac{1}{n}\sigma^2 = E[\overline{X}^2] - (E[\overline{X}])^2$$

$$\frac{\sigma^2}{n} = E[\overline{X}^2] - \mu^2$$

Thus

$$E[X_i^2] = \sigma^2 + \mu^2$$

while

$$E[\overline{X}^2] = \frac{\sigma^2}{n} + \mu^2$$

and

$$E[\sum_{i=1}^{n} X_i^2] - nE[\overline{X}^2] = n(\sigma^2 + \mu^2) - n(\frac{\sigma^2}{n} + \mu^2) = (n-1)\sigma^2$$

Divide by n:

$$E[\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2] = \frac{n-1}{n}\sigma^2$$

Alternative: divide by n-1:

$$S^2 = E[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2] = \sigma^2$$