Chris Piech
CS109

Lecture #18
Nov 3rd, 2017

Covariance and Sampling

Product of Expectations Lemma

Here is a lovely little lemma to get us started:

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)] \qquad \text{if and only if } X \text{ and } Y \text{ are independent}$$

Covariance

Covariance is a quantitative measure of the extent to which the deviation of one variable from its mean matches the deviation of the other from its mean. It is a mathematical relationship that is defined as:

$$\text{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])]$$

That is a little hard to wrap your mind around (but worth pushing on a bit). The outer expectation will be a weighted sum of the inner function evaluated at a particular (x,y) weighted by the probability of (x,y). If x and y are both above their respective means, or if x and y are both below their respective means, that term will be positive. If one is above its mean and the other is below, the term is negative. If the weighted sum of terms is positive, the two random variables will have a positive correlation. We can rewrite the above equation to get an equivalent equation:

$$\text{Cov}(X,Y) = E[XY] - E[Y]E[X]$$

Using this equation (and the product lemma) is it easy to see that if two random variables are independent their covariance is 0. The reverse is *not* true in general.

Properties of Covariance

Say that X and Y are arbitrary random variables:

$$\text{Cov}(X,Y) = \text{Cov}(Y,X)$$
$$\text{Cov}(X,X) = E[X^2] - E[X]E[X] = \text{Var}(X)$$
$$\text{Cov}(aX+b,Y) = a\text{Cov}(X,Y)$$

Let $X = X_1 + X_2 + \cdots + X_n$ and let $Y = Y_1 + Y_2 + \cdots + Y_m$. The covariance of X and Y is:

$$\text{Cov}(X,Y) = \sum_{i=1}^{n}\sum_{j=1}^{m}\text{Cov}(X_i,Y_j)$$
$$\text{Cov}(X,X) = \text{Var}(X) = \sum_{i=1}^{n}\sum_{j=1}^{n}\text{Cov}(X_i,X_j)$$

That last property gives us a third way to calculate variance. We can use it to, again, show how to get the variance of a Binomial.