Covariances and correlations

18.650, Sept. 23, 2015

1 Covariances

Let X and Y be two real random variables on a probability space with $E(X^2) < \infty$ and $E(Y^2) < \infty$. Their variances are defined by $\mathrm{Var}(X) = E((X - EX)^2)$ and likewise for Y.

Lemma 1 *For any two such X and Y, $E(XY)$ is finite and satisfies $E(XY)^2 \leq E(X^2)E(Y^2)$.*

Proof. For any real t we have $0 \leq q(t) \equiv E((tX + Y)^2) = t^2 E(X^2) + 2tE(XY) + E(Y^2)$. The quadratic function $q(t)$ cannot have two distinct real roots, or it would become negative for some t. So its discriminant

$$b^2 - 4ac = 4(E(XY))^2 - 4E(X^2)E(Y^2) \leq 0.$$

Dividing by 4 gives he conclusion. Q.E.D.

The *covariance* of X and Y is defined by

$$\mathrm{Cov}(X, Y) = E((X - EX)(Y - EY)).$$

By Lemma 1 applied to $X - EX$ and $Y - EY$ we get

$$\mathrm{Cov}(X, Y)^2 \leq \mathrm{Var}(X)\,\mathrm{Var}(Y). \tag{1}$$

The *standard deviation* σ_X is defined as $(\mathrm{Var}(X))^{1/2}$. The *correlation* of X and Y is defined as

$$\rho_{X,Y} = \mathrm{Cov}(X, Y)/(\sigma_X \sigma_Y)$$

if the denominator is not 0. By (1) it satisfies $-1 \leq \rho_{X,Y} \leq 1$. For finite samples $x_1, ..., x_n$ and $y_1, ..., y_n$ with $n \geq 2$, x_j not all equal and y_j not all equal, the sample variances s_x^2 and s_y^2 are defined and positive. We have the sample covariance $scov(x, y)$ and the sample correlation $r_{x,y} = scov(x, y)/(s_x s_y)$. Just as for random variables we have

$$-1 \leq r_{x,y} \leq 1.$$