EE 464

Spring 2003

Lecture Notes Part 11a

Christopher Wayne Walker, Ph.D.

11.0 Moments and Conditional Distributions

11.1 Joint Moments

Given random variables X and Y, let $Z = g(X, Y)$. The expected value of Z is given by

$$E(Z) = \int_{-\infty}^{\infty} z f_Z(z) dz$$

as usual.

Theorem:

$$E(Z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy.$$

Proof: Omitted.

Note that if Z is only a function of X, i.e., $Z = g(X)$, then

$$\begin{aligned} E(Z) = E(g(X)) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy \\ &= \int_{-\infty}^{\infty} g(x) \int_{-\infty}^{\infty} f(x, y) dy dx \\ &= \int_{-\infty}^{\infty} g(x) f_X(x) dx \end{aligned}$$

as expected.

If X and Y are discrete, then

$$E(Z) = E(g(X, Y)) = \sum_{i,k} g(x_i, y_k) p_{ik}.$$

Linearity: As in the 1-dim case, the 2-dim expectation operator is linear. In particular,

$$E(X + Y) = E(X) + E(Y).$$

We would like to measure in some meaningful way the degree of association between random variables X and Y. The covariance helps us do this.

Definition: The *covariance* of two random variables X and Y is

$$C_{XY} = E\left[(X - \mu_X)(Y - \mu_Y)\right],$$

where $\mu_x = E(X),\ \mu_Y = E(Y)$.

Note:

$$C_{XY} = E(XY) - \mu_X E(Y) - \mu_Y E(X) + \mu_X \mu_y$$

or

$$C_{XY} = E(XY) - E(X)E(Y).$$

One can write C for C_{XY} when clear to do so.

In order to compare the degree of association of different pairs of random variables it is convenient to normalize the covariance measure to produce the correlation coefficient.

Definition: The *correlation coefficient*, r_{XY}, of random variables X and Y is

$$r_{XY} = \frac{C_{XY}}{\sigma_X \sigma_Y},\ \sigma_X \sigma_y \neq 0.$$

Note: We define $C_{XY} = 0$ if $\sigma_X = 0$ or $\sigma_Y = 0$.

One can write r for r_{XY} when clear.

Note: Some authors use ρ_{XY} instead of r_{XY}.

Claim: $|r_{XY}| \leq 1$ (i.e., $|C_{XY}| \leq \sigma_X \sigma_Y$).

Proof: Consider

$$q(t) = E\left[(t(X - \mu_X) + (Y - \mu_Y))^2 \right].$$

Then $q(t) \geq 0$. Expanding we get

$$q(t) = E\left[t^2 (X - \mu_X)^2 + 2t(X - \mu_X)(Y - \mu_Y) + (Y - \mu_Y)^2 \right] \geq 0$$

or

$$q(t) = t^2 \sigma_X^2 + 2tC_{XY} + \sigma_Y^2 \geq 0.$$

This is a quadratic in t. Setting $q(t) = 0$ we find

$$t = \frac{-2C_{XY} \pm \sqrt{4C_{XY}^2 - 4\sigma_X^2 \sigma_Y^2}}{2\sigma_X^2}.$$

Since $q(t) \geq 0$, $q(t)$ cannot have two distinct real roots. Thus, the discriminant is less than or equal to zero. Hence,

$$4C_{XY}^2 - 4\sigma_X^2\sigma_Y^2 \leq 0 \Rightarrow |C_{XY}| \leq |\sigma_X\sigma_Y| \Rightarrow |C_{XY}| \leq \sigma_X\sigma_Y$$

$$\Rightarrow -\sigma_X\sigma_y \leq C_{XY} \leq \sigma_X\sigma_y \Rightarrow -1 \leq r_{XY} \leq 1 \Rightarrow |r_{XY}| \leq 1.$$

If $r_{XY}^2 = 1$, i.e., $r_{XY} = \pm 1$, we can establish a functional relationship between X and Y. We will now show this. We need the following theorem.

Theorem: Suppose $\sigma_X^2 = 0$. Then $P(X = \mu_X) = 1$ (we say $X = \mu_X$ with probability 1).

Proof: Recall Tchebycheff's inequality

$$P\left(|X - \mu_X| \geq \epsilon\right) \leq \frac{\sigma_X^2}{\epsilon^2}, \quad \forall\, \epsilon > 0.$$

So,

$$P\left(|X - \mu_X| \geq \epsilon\right) = 0 \Rightarrow P\left(|X - \mu_X| < \epsilon\right) = 1$$
$$\Rightarrow P(-\epsilon < X - \mu_X < \epsilon) = 1.$$

Let $\epsilon \to 0$ to get $P(X - \mu_X = 0) = 1$ or $P(X = \mu_X) = 1$.

Theorem: $r_{XY}^2 = 1$ if and only if $Y = aX + b$ for some constants a, b.

Proof:
"$\Longrightarrow$" (Necessity or "only if" part).

Assume $r_{XY}^2 = 1$. Recall,

$$q(t) = E\left[(t(X - \mu_X) + (Y - \mu_Y))^2\right]$$

or

$$q(t) = t^2\sigma_X^2 + 2tC_{XY} + \sigma_Y^2 \geq 0.$$

If $q(t) > 0$, then the discriminant

$$4C_{XY}^2 - 4\sigma_X^2\sigma_Y^2 < 0 \Rightarrow \frac{C_{XY}^2}{\sigma_X^2\sigma_Y^2} < 1 \Rightarrow r_{XY}^2 < 1$$

contrary to the assumption $r_{XY}^2 = 1$. So if $r_{XY}^2 = 1$ there exists some $t_0 \in \mathbf{R}$ such that $q(t_0) = 0$. Thus,

$$q(t_0) = E\left[(t_0(X - \mu_X) + (Y - \mu_Y))^2\right] = 0.$$

Note that

$$E\left[t_0(X - \mu_X) + (Y - \mu_Y)\right] = 0.$$

So

$$E\left[(t_0(X - \mu_X) + (Y - \mu_Y))^2\right] = Var\left(t_0(X - \mu_X) + (Y - \mu_Y)\right) = 0.$$

The last theorem implies

$$P\left(t_0(X - \mu_X) + (Y - \mu_Y) = 0\right) = 1$$

which imples

$$t_0(X - \mu_X) + (Y - \mu_Y) = 0 \text{ with probability 1.}$$

We suppress the expression "with probability 1" in equations. Thus,

$$Y = -t_0 X + t_0 \mu_X + \mu_Y \Rightarrow Y = ax + b$$

with $a = -t_0$ and $b = t_0 \mu_X + \mu_Y$.

"$\Longleftarrow$" (Sufficiency or "if" part).

Assume $Y = aX + b$. Then,

$$E(Y) = aE(X) + b$$

and

$$Var(Y) = a^2 Var(X).$$

Also,

$$E(XY) = E\left[X(aX + b)\right] = aE(X^2) + bE(X).$$

Thus,

$$r_{XY}^2 = \frac{\left[E(XY) - E(X)E(Y)\right]^2}{Var(X)Var(Y)}$$

$$= \frac{\left[aE(X^2) + bE(X) - E(X)(aE(X) + b)\right]^2}{Var(X)a^2Var(X)}$$

$$= \frac{\left[aE(X^2) + bE(X) - a(E(X))^2 - bE(X))\right]^2}{a^2(Var(X))^2}$$

$$= \frac{a^2\left[E(X^2) - (E(X))^2)\right]^2}{a^2(Var(X))^2} = 1.$$

Note that if $a > 0$, $r_{XY} = 1$ and if $a < 0$, $r_{XY} = -1$.

Observe that the correlation coefficient is a measure of the degree of linearity between X and Y: $|r_{XY}| \approx 1$ indicates a high degree of linearity while $|r_{XY}| \approx 0$ indicates a lack of linearity. Positive values of r_{XY} means Y tends to increase with increasing X (and vice versa) while negative values of r_{XY} means that Y tends to decrease with increasing X (and vice versa).

Caution: $r_{XY} = 0$ does not mean there is no relationship between X and Y, it only means there is no linear relationship. There could still be a nonlinear relationship.

Example: Suppose $X \sim N(0, 1)$. Let $Y = X^2$. Here Y is very much related to X, in fact, Y is a (nonlinear) function of X. Now

$$r_{XY} = \frac{C_{XY}}{\sigma_X \sigma_Y}$$

but

$$C_{XY} = E(XY) - E(X)E(Y) = E(X^3) - E(X)E(X^2) = 0$$

since $E(X^3) = E(X) = 0$. Thus $r_{XY} = 0$.

Definition: Two random variables X and Y are called *uncorrelated* if their covariance is zero, i.e., $C_{XY} = 0$.

Definition: Two random variables X and Y are called *orthogonal* if $E(XY) = 0$.

Notation: $X \perp Y$ means X and Y are orthogonal.

Note: If X and Y are uncorrelated then $(X - \mu_X) \perp (Y - \mu_Y)$.

Theorem: If X and Y are independent then they are uncorrelated.

Proof: We will show $E(XY) = E(X)E(Y)$ which proves the theorem.

$$E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x,y)dxdy$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf_X(x)f_Y(y)dxdy$$

$$= \int_{-\infty}^{\infty} xf_X(x)dx \int_{-\infty}^{\infty} yf_Y(y)dy = E(X)E(Y).$$

Note: If X and Y are independent then $g(X)$ and $h(Y)$ are also independent and $E\left[g(X)h(Y)\right] = E\left[g(X)\right]E\left[h(Y)\right]$ (so $g(X)$ and $h(Y)$ are also uncorrelated) but if X and Y are just uncorrelated it does not follow that $g(X)$ and $h(Y)$ are necessarily uncorrelated. Also, random variables can be uncorrelated without being independent, but in the normal case we that independence implies uncorrelated and uncorrelated implies independence.

Variance of $X + Y$

Consider $Z = X + Y$. Then $E[Z] = \mu_Z = \mu_X + \mu_Y$.

$$Var(Z) = \sigma_Z^2 = E\left[(Z - \mu_Z)^2\right] = E\left[((X - \mu_X) + (Y - \mu_Y))^2\right]$$

$$= E\left[(X - \mu_X)^2\right] + 2E\left[(X - \mu_X)(Y - \mu_Y)\right] + E\left[(Y - \mu_Y)^2\right]$$

or

$$\sigma_Z^2 = \sigma_X^2 + 2r_{XY}\sigma_X\sigma_Y + \sigma_Y^2.$$

If X and Y are uncorrelated then $r_{XY} = 0$ and

$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2.$$

It then follows that if X and Y are independent

$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2.$$

Moments

Definition: A *joint moment* of the random variables X and Y of order $k + r = n$ is

$$m_{kr} = E\left(X^k Y^r\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^k y^r f(x, y) dx dy.$$

Note: The first order moments are

$$m_{10} = E(X)$$

$$m_{01} = E(Y)$$

and the second order moments are

$$m_{20} = E(X^2)$$

$$m_{11} = E(XY)$$

$$m_{02} = E(Y^2).$$

Definition: The *joint central moments* of the random variables X and Y are the moments of $(X - \mu_X)$ and $(Y - \mu_Y)$, i.e.,

$$\mu_{kr} = E\left[(X - \mu_X)^k (Y - \mu_Y)^r\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \mu_X)^k (y - \mu_Y)^r f(x, y) dx dy.$$

Here

$$\mu_{10} = 0,\ \mu_{01} = 0,\ \mu_{11} = C_{XY},\ \mu_{20} = \sigma_X^2,\ \mu_{02} = \sigma_Y^2.$$

In general, the joint density of X and Y is required in order to determine the joint statistics. However, often for many applications it is sufficient to use only the first- and second-order moments which can be easily estimated numerically given sampled data.