MOMENTS AND SAMPLE MOMENTS

A. Let $x_1,\ldots x_n$ be real numbers (the data).

The sample mean (or arithmetic mean) $\bar{x}$ is then defined as

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = (x_1 + x_2 + \ldots + x_n)/n$$

If $y_1,\ldots,y_n$ is another set of data then

$$s_{xy} = \frac{1}{n}\sum_{i=1}^{n} (x_i-\bar{x})(y_i-\bar{y})$$

is the sample covariance.

It is easy to check that

$$s_{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i(y_i-\bar{y}) = \frac{1}{n}\sum_{i=1}^{n} (x_i-\bar{x})y_i = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\,\bar{y}$$

Since s_{xy} depends on the units of measurement of x_i and y_i, i.e., $s_{\alpha x,\beta y} = \alpha\beta s_{xy}$ for $\alpha>0$, $\beta>0$, we often consider the sample correlation r_{xy} instead:

$$r_{xy} = \frac{s_{xy}}{s_x\, s_y} = \frac{s_{xy}}{\sqrt{s_x^2}\sqrt{s_y^2}}$$

where $s_x^2 = s_{xx} = \frac{1}{n}\sum_{i=1}^{n} (x_i-\bar{x})^2$ is the sample variance. (Note that clearly $s_x^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\bar{x})^2$ holds.)

Clearly, r_{xy} is not affected by a change in the units of measurement, i.e., $r_{\alpha x,\beta y}=r_{xy}$ holds ($\alpha>0,\beta>0$). Note also that $-1\le r_{xy}\le 1$ holds.

B. Let X be a random variable with discrete probability density $f(x_j)$, $j=1,\ldots,k$. That is, X takes the values $x_1,\ldots,x_k$ with $\Pr(X=x_j)=f(x_j)$.

The expected value of X is then defined as:

$$EX = \sum_{i=1}^{k} x_j f(x_j).$$

If X is a random variable with a continuous distribution having a probability density f(x), then the expected value is defined as

$$EX = \int_{-\infty}^{\infty} xf(x)dx.$$

Notation: Sometimes we write μ or μ_x for EX.

Note: If $X_1, X_2, \ldots, X_n$ is a sequence of random variable with the same discrete probability density $f(x_j)$, $j=1,\ldots,k$, we can form two different quantities:

1. $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ This is again a random variable, also sometimes called the sample mean.

2. $EX_i = \sum_{j=1}^{k} x_j f(x_j)$ This is a real number, the expected value of X_i.

(Of course, if $X_1, \ldots, X_n$ are independent and identically distributed, then $\bar{X}$ will serve as an estimator for EX_i.)

If X and Y are two random variables with discrete joint probability density $f(x_j, y_l)$ $j=1,\ldots,k$; $l=1,\ldots,m$, the covariance between X and Y is defined as

$$Cov(X,Y) = \sum_{l=1}^{m}\sum_{j=1}^{k}(x_j-EX)(y_l-EY)f(x_j,y_l) = \sum_{l=1}^{m}\sum_{j=1}^{k}(x_j-\mu_x)(y_l-\mu_y)f(x_j,y_l) =$$
$$E[(X-EX)(Y-EY)].$$

Notation: Sometimes we write σ_{xy} for Cov(X,Y).

If X and Y are two random variables with a continuous joint distribution having probability density f(x,y) then the covariance between X and Y is defined as

$$Cov(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_x)(y-\mu_y)f(x,y)dxdy = E[(X-EX)(Y-EY)].$$

Again Cov(X,Y) depends on the units of measurement, i.e.,

$$\text{Cov}(\alpha X,\beta Y) = \alpha\beta\text{Cov}(X,Y).$$

Herce, we define the correlation between X and Y as

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{\sigma_{xy}}{\sqrt{\sigma_x^2}\sqrt{\sigma_y^2}}$$

where $\sigma_x^2 = \text{Var}(X) = E[(X-EX)^2]$.

It is easy to check that the following holds:

$$\sigma_{xy} = E(X(Y-EY)) = E((X-EX)Y) = E(XY)-(EX)(EY) = E(XY)-\mu_x\mu_y,$$

$$\sigma_x^2 = E(X^2)-(EX)^2$$

Note: Again, do not confuse σ_{xy} with s_{xy}! σ_{xy} is the "population" analogue to s_{xy}.

Some properties of cov(X,Y):

If $\rho_{xy}=0$ we say that X and Y are uncorrelated. Clearly, $\rho_{xy}=0 \Leftrightarrow \text{cov}(X,Y)=0$.

1. Let $X_1,\ldots,X_n$ be uncorrelated with $EX_i=0$, then

$$E(\textstyle\sum_{i=1}^n X_i)^2 = \sum_{i=1}^n EX_i^2.$$

Proof:

$$E(\textstyle\sum_{i=1}^n X_i)^2 = E[\sum_{i=1}^n X_i^2 + \sum_{i\neq j} X_iX_j] = \sum_{i=1}^n EX_i^2 + \sum_{i\neq j} EX_iX_j =$$

$$\textstyle\sum_{i=1}^n EX_i^2 + \sum_{i\neq j} E\{(X_i-EX_i)(X_j-EX_j)\} = \qquad \text{(since } EX_i=0\text{)}$$

$$\textstyle\sum_{i=1}^n EX_i^2 + \sum_{i\neq j} \text{cov}(X_i,X_j) =$$

$$\textstyle\sum_{i=1}^n EX_i^2 \qquad \text{(since cov}(X_i,X_j)=0 \text{ by assumption)}$$

2. Let $Z_1,\ldots,Z_n$ be uncorrelated, then $\text{Var}(\sum_{i=1}^n Z_i) = \sum_{i=1}^n \text{Var}(Z_i)$.

More generally, if $Z_1,\ldots,Z_n$ are uncorrelated then

$$\text{Cov}(\textstyle\sum_{i=1}^n \alpha_i Z_i, \sum_{j=1}^n \beta_j Z_j) = \sum_{i=1}^n \alpha_i\beta_i \text{Var}(Z_i)$$

Proof:

$$Cov(\sum_{i=1}^{n}\alpha_i Z_i, \sum_{j=1}^{n}\beta_j Z_j) = E[(\sum_{i=1}^{n}\alpha_i Z_i - E(\sum_{i=1}^{n}\alpha_i Z_i))(\sum_{j=1}^{n}\beta_j Z_j - E(\sum_{j=1}^{n}\beta_j Z_j))]=$$

$$E[(\sum_{i=1}^{n}\alpha_i Z_i - \sum_{i=1}^{n}\alpha_i EZ_i)(\sum_{j=1}^{n}\beta_j Z_j - \sum_{j=1}^{n}\beta_j EZ_j)]=$$

$$E[\sum_{i=1}^{n}\alpha_i (Z_i - EZ_i)\sum_{j=1}^{n}\beta_j (Z_j - EZ_j)] = E[\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\beta_j (Z_i - EZ_i)(Z_j - EZ_j)] =$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\beta_j cov(Z_i, Z_j) = \sum_{i=1}^{n}\alpha_i\beta_i\ cov(Z_i, Z_i) = \sum_{i=1}^{n}\alpha_i\beta_i\ Var(Z_i),$$

since $cov(Z_i, Z_j)=0$ for $i \neq j$ by assumption.