Moments of More that One Random Variable

Charles B. Moss

July 14, 2010

I. Covariance and Correlation

A. **Definition 4.3.1** The covariance between two random variables X and Y can be defined as

$$\begin{aligned}\text{Cov}\left(X,Y\right) &= \text{E}\left[\left(\text{X}-\text{E}\left[\text{X}\right]\right)\left(\text{Y}-\text{E}\left[\text{Y}\right]\right)\right]\\ &= \text{E}\left[\text{XY}-\text{XE}\left[\text{Y}\right]-\text{E}\left[\text{x}\right]\text{Y}+\text{E}\left[\text{X}\right]\left[\text{Y}\right]\right]\\ &= \text{E}\left[\text{XY}\right]-\text{E}\left[\text{X}\right]\text{E}\left[\text{Y}\right]-\text{E}\left[\text{X}\right]\text{E}\left[\text{Y}\right]+\text{E}\left[\text{X}\right]\text{E}\left[\text{Y}\right]\\ &= \text{E}\left[\text{XY}\right]-\text{E}\left[\text{X}\right]\text{E}\left[\text{Y}\right]\end{aligned} \tag{1}$$

1. Note that this is simply a generalization of the standard variance formulation. Specifically, letting $Y \rightarrow X$ yields

$$\begin{aligned}\text{Cov}\left(XX\right) &= \text{E}\left[\text{XX}\right]-\text{E}\left[\text{X}\right]\text{E}\left[\text{X}\right]\\ &= \text{E}\left[\text{X}^2\right]-\left(\text{E}\left[\text{X}\right]\right)^2\end{aligned} \tag{2}$$

2. From a sample perspective, we have

$$\begin{aligned}V\left(X\right) &= \tfrac{1}{N}\textstyle\sum_{i=1}^{N} x_i^2-\bar{x}^2\\ \text{Cov}\left(X,Y\right) &= \tfrac{1}{N}\textstyle\sum_{i=1}^{N} x_i y_i-\bar{x}\bar{y}\end{aligned} \tag{3}$$

3. Together the variance and covariance matrices are typically written as a variance matrix

$$\Sigma = \begin{bmatrix} V(X) & \text{Cov}(X,Y) \\ \text{Cov}(Y,X) & V(Y) \end{bmatrix} = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{bmatrix} \tag{4}$$

Note that $\text{Cov}(X,Y) = \sigma_{xy} = \sigma_{yx} = \text{Cov}(Y,X)$.

4. Substituting the sample measures into the variance matrix yields

$$\begin{aligned} S = \begin{bmatrix} s_{xx} & s_{xy} \\ s_{yx} & s_{yy} \end{bmatrix} &= \begin{bmatrix} \frac{1}{N}\sum_{i=1}^{N} x_i x_i - \bar{x}\bar{x} & \frac{1}{N}\sum_{i=1}^{N} x_i y_i \bar{x}\bar{y} \\ \frac{1}{N}\sum_{i=1}^{N} y_i x_i - \bar{y}\bar{x} & \frac{1}{N}\sum_{i=1}^{N} y_i y_i - \bar{y}\bar{y} \end{bmatrix} \\ &= \frac{1}{N}\begin{bmatrix} \sum_{i=1} x_i x_i & \sum_{i=1}^{N} x_i y_i \\ \sum_{i=1}^{N} y_i x_i & \sum_{i=1}^{N} y_i y_i \end{bmatrix} - \begin{bmatrix} \bar{x}\bar{x} & \bar{x}\bar{y} \\ \bar{y}\bar{x} & \bar{y}\bar{y} \end{bmatrix} \end{aligned} \tag{5}$$

The sample covariance matrix can then be written as

$$S = \frac{1}{N}\begin{bmatrix} x_1 & \cdots & x_N \\ y_1 & \cdots & y_N \end{bmatrix} - \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}\begin{bmatrix} \bar{x} & \bar{y} \end{bmatrix}. \tag{6}$$

4. In terms of the theoretical distribution, the variance matrix can be written as

$$\Sigma = \begin{bmatrix} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_x)^2 f(x,y)\,dx\,dy & \int_{-\infty}^{\infty}\int_{\infty}^{\infty} (x-\mu_x)(y-\mu_y) f(x,y)\,dx\,dy \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_x)(y-\mu_y) f(x,y)\,dx,dy & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (y-\mu_y)^2\,dx\,dy \end{bmatrix} \tag{7}$$

5. Example 4.3.2. Compute the covariance for the data presented in Table 1. We start by computing the means from Table 1

Table 1: Discrete Sample

X Y	-1	0	1	Marginal Probability
1	0.167	0.083	0.167	0.417
0	0.083	0.000	0.083	0.167
-1	0.167	0.083	0.417	0.417
Marginal Probability	0.417	0.167	0.417	

$$\begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} = \begin{bmatrix} (0.167+0.083+0.167)(1) \\ +(0.083+0.000+0.083)(0) \\ +(0.417+0.083+0.167)(-1) \\ \\ (0.167+0.083+0.167)(1) \\ +(0.083+0.000+0.083)(0) \\ +(0.417+0.083+0.167)(-1) \end{bmatrix}$$
$$= \begin{bmatrix} (0.417)(1)+(0.167)(0)+(0.417)(-1) \\ (0.417)(1)+(0.167)(0)+(0.417)(-1) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{8}$$

Given that the means for both variables are zero, we can compute the covariance as

$$\begin{aligned} \text{Cov}[X,Y] &= \frac{1}{9}\sum_{i\in\{1,0,1\}} x_i y_i P(x_i,y_i) \\ &= (1)(-1)\,0.167 + (0)(-1)\,0.083 + \cdots (-1)(1)\,0.167 = 0. \end{aligned} \tag{9}$$

B. **Theorem 4.3.2.** $V(X \pm Y) = V(X) + V(Y) \pm \text{Cov}(X,Y)$

$$\begin{aligned} V[X \pm Y] &= \mathrm{E}[(\mathrm{X} \pm \mathrm{Y})(\mathrm{X} \pm \mathrm{Y})] \\ &= \mathrm{E}[\mathrm{XX} \pm 2\mathrm{XY} + \mathrm{YY}] \\ &= \mathrm{E}[\mathrm{XX}] + \mathrm{E}[\mathrm{YY}] \pm \mathrm{E}[\mathrm{XY}] \\ &= V(X) + V(Y) \pm 2\text{Cov}(X,Y) \end{aligned} \tag{10}$$

1. Note that this result can be obtained from the variance matrix. Specifically, $X+Y$ can be written as a vector operation

$$\begin{bmatrix} X & Y \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = X+Y \tag{11}$$

Given this vectorization of the problem we can define the variance of the sum as

$$\begin{aligned} \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{xy} & \sigma_{yy} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} &= \begin{bmatrix} \sigma_{xx}+\sigma_{xy} & \sigma_{xy}+\sigma_{yy} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= \sigma_{xx} + 2\sigma_{xy} + \sigma_{yy} \end{aligned} \tag{12}$$

C. **Theorem 4.3.3.** Let X_i , $i = 1, 2, \cdots$ be pairwise independent. Then

$$V\left(\sum_{i=1}^{N} X_i\right) = \sum_{i=1}^{N} V(X_i) \tag{13}$$

The simplest proof to this theorem is to use the variance matrix. Note in the preceding example, if X and Y are independent, we have

$$\begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{xy} & \sigma_{yy} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \sigma_{xx} + \sigma_{yy} \tag{14}$$

if $\sigma_{xy} = 0$. Extending this result to three variables implies

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}' \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32}\sigma_{33} & \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \sigma_{11}+2\sigma_{12}+2\sigma_{13}+\sigma_{22}+2\sigma_{23}+\sigma_{33} \tag{15}$$

if the xs are independent, the covariance terms are zero and this expression simply becomes the sum of the variances.

E. **Definition 4.3.2.** The correlation coefficient for two variables is defined as

$$\text{Corr}(X,Y)=\frac{\text{Cov}(X,Y)}{\sqrt{\sigma_{xx}}\sqrt{\sigma_{yy}}} \tag{16}$$

1. Note that the covariance between any random variable and a constant is equal to zero. Letting Y equal to zero we have

$$\mathrm{E}\left[(\mathrm{X}-\mathrm{E}[\mathrm{X}])(\mathrm{Y}-\mathrm{E}[\mathrm{Y}])\right]=\mathrm{E}\left[(\mathrm{X}-\mathrm{E}[\mathrm{X}])(0)\right]=0 \tag{17}$$

2. It stands to reason the correlation coefficient between a random variable and a constant is also zero.

F. It now possible to derive the ordinary least squares estimator for a linear regression equation.

1. We define the ordinary least squares estimator as that set of parameters that minimizes the squared error of the estimate

$$\min_{\alpha,\beta}=\mathrm{E}\left[(\mathrm{Y}-\alpha-\beta\mathrm{X})^2\right]=\min_{\alpha,\beta}\mathrm{E}\left[\mathrm{Y}^2-2\alpha\mathrm{Y}-2\beta\mathrm{XY}+\alpha^2+2\alpha\beta\mathrm{X}+\beta^2\mathrm{X}^2\right] \tag{18}$$

The first order conditions for this minimization problem then becomes

$$\begin{aligned}\frac{\partial S}{\partial\alpha}&=-2\mathrm{E}[\mathrm{Y}]+2\alpha+2\beta\mathrm{E}[\mathrm{X}]=0\\ \frac{\partial S}{\partial\beta}&=-2\mathrm{E}[\mathrm{XY}]+2\alpha\mathrm{E}[\mathrm{X}]+2\beta\mathrm{E}\left[\mathrm{X}^2\right]=0\end{aligned} \tag{19}$$

Solving the first equation for α yields

$$\alpha=\mathrm{E}[\mathrm{Y}]-\beta[\mathrm{X}] \tag{20}$$

Substituting this expression into the second first order condition yields

$$\begin{gathered}-\mathrm{E}[\mathrm{XY}]+(\mathrm{E}[\mathrm{Y}]-\beta\mathrm{E}[\mathrm{X}])\,\mathrm{E}[\mathrm{X}]+\beta\mathrm{E}\left[\mathrm{X}^2\right]=0\\ -\mathrm{E}[\mathrm{XY}]+\mathrm{E}[\mathrm{Y}]\,\mathrm{E}[\mathrm{X}]+\beta\left(\mathrm{E}\left[\mathrm{X}^2\right]-(\mathrm{E}[\mathrm{X}])^2\right)=0\\ -\text{Cov}(X,Y)+\beta V(X)=0\\ \Rightarrow\beta=\frac{\text{Cov}(X,Y)}{V(X)}\end{gathered} \tag{21}$$

G. **Theorem 4.3.6.** The best linear predictor (or more exactly, the minimum mean-squared-error linear predictor) of Y based on X is given by $\alpha^*+\beta^* X$, where α^* and β^* are the least square estimates.

II. Conditional Mean and Variance

A. **Definition 4.4.1** Let (X,Y) be a bivariate discrete random variable taking values (x_i,y_j) $i,j = 1,2,\cdots$. Let $P(y_j|X)$ be the conditional probability of $Y=y_j$ given X. Let $\phi(x_i,y_j)$ be an arbitrary function. Then the conditional mean of $\phi(X,Y)$ given X, denoted $\mathrm{E}[\phi(\mathrm{X},\mathrm{Y})|\mathrm{X}]$ or by $\mathrm{E}_{\mathrm{Y|X}}[\phi(\mathrm{X},\mathrm{Y})]$, is defined by

$$\mathrm{E}_{\mathrm{Y|X}}[\phi(\mathrm{X},\mathrm{Y})] = \sum_{\mathrm{i}=1}^{\infty} \phi(\mathrm{X},\mathrm{y_i})\,\mathrm{P}(\mathrm{y_i}|\mathrm{X}) \tag{22}$$

B. **Definition 4.4.2.** Let (X,Y) be a bivarite continuous random variable with conditional density $f(y|x)$. Let $\phi(x,y)$ be an arbitrary function. Then the conditional mean of $\phi(X,Y)$ given X is defined by

$$\mathrm{E}_{\mathrm{Y|X}}[\phi(\mathrm{X},\mathrm{Y})] = \int_{-\infty}^{\infty} \phi(\mathrm{X},\mathrm{y})\,\mathrm{f}(\mathrm{y}|\mathrm{X})\,\mathrm{dy} \tag{23}$$

C. **Theorem 4.4.1** (Law of Iterated Means) $\mathrm{E}[\phi(\mathrm{X},\mathrm{Y})] = \mathrm{E_x}\mathrm{E}_{\mathrm{Y|X}}[\phi(\mathrm{X},\mathrm{Y})]$. (Where the symbol $\mathrm{E_x}$ denotes the expectation with respect to X).

D. **Theorem 4.4.2**

$$V(\phi(X,Y)) = \mathrm{E_X}\left[\mathrm{V}_{\mathrm{Y|X}}[\phi(\mathrm{X},\mathrm{Y})]\right] + \mathrm{V_x}\left[\mathrm{E}_{\mathrm{Y|X}}[\phi(\mathrm{X},\mathrm{Y})]\right] \tag{24}$$

Proof:

$$V_{Y|X}[\phi] = \mathrm{E}_{\mathrm{Y|X}}\left[\phi^2\right] - \left(\mathrm{E}_{\mathrm{Y|X}}[\phi]\right)^2 \tag{25}$$

Implies

$$\mathrm{E_X}\left[\mathrm{V}_{\mathrm{Y|X}}(\phi)\right] = \mathrm{E}\left[\phi^2\right] - \mathrm{E_X}\left[\mathrm{E}_{\mathrm{Y|X}}[\phi]\right] \tag{26}$$

By definition of conditional variance

$$V_X\left(\mathrm{E}_{\mathrm{Y|X}}\left[\phi\right]\right)=\mathrm{E}_{\mathrm{X}}\left[\mathrm{E}_{\mathrm{Y|X}}\left[\phi\right]\right]^2-\left(\mathrm{E}\left[\phi\right]\right)^2 \tag{27}$$

Adding these expressions yields

$$\mathrm{E}_{\mathrm{X}}\left[\mathrm{V}_{\mathrm{Y|X}}\left(\phi\right)\right]+\mathrm{V}_{\mathrm{X}}\left(\mathrm{E}_{\mathrm{Y|X}}\left[\phi\right]\right)=\mathrm{E}\left[\phi^2\right]-\left(\mathrm{E}\left[\phi\right]\right)=\mathrm{V}\left(\phi\right) \tag{28}$$

E. **Theorem 4.4.3** The best predictor (or the minimum mean-squared-error predictor) of Y based on X is given by $\mathrm{E}\left[\mathrm{Y|X}\right]$.