

Moments of More than One Random Variable: Lecture XI

Charles B. Moss

September 16, 2026



Reading Question

- Given a uniform distribution function $f(x_1, x_2) = 1$, what is the expected value of

$$g(x_1, x_2) = x_1^2 - 3x_1x_2 + x_2^2 \quad (1)$$

- 1 Covariance and Correlation
- 2 Conditional Mean and Variance

Covariance and Correlation

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

$$\begin{aligned}\text{Cov}(X, Y) &= E[(X - E[X])(Y - E[Y])] \\ &= E[XY - XE[Y] - E[X]Y + E[X]E[Y]] \\ &= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y] \\ &= E[XY] - E[X]E[Y]\end{aligned}\tag{2}$$

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

$$\begin{aligned}\text{Cov}(XX) &= E[XX] - E[X]E[X] \\ &= E[X^2] - (E[X])^2\end{aligned}\tag{3}$$

Continued

- From a sample perspective, we have

$$V(X) = \frac{1}{N} \sum_{i=1}^N x_i^2 - \bar{x}^2 \quad (4)$$

$$\text{Cov}(X, Y) = \frac{1}{N} \sum_{i=1}^N x_i y_i - \bar{x} \bar{y}$$

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

$$\Sigma = \begin{bmatrix} V(X) & \text{Cov}(X, Y) \\ \text{Cov}(Y, X) & V(Y) \end{bmatrix} = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{bmatrix} \quad (5)$$

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

Continued

- Substituting the sample measures into the variance matrix yields

$$\begin{aligned}
 S &= \begin{bmatrix} s_{xx} & s_{xy} \\ s_{yx} & s_{yy} \end{bmatrix} = \begin{bmatrix} \frac{1}{N} \sum_{i=1}^N x_i x_i - \bar{x}\bar{x} & \frac{1}{N} \sum_{i=1}^N x_i y_i - \bar{x}\bar{y} \\ \frac{1}{N} \sum_{i=1}^N y_i x_i - \bar{y}\bar{x} & \frac{1}{N} \sum_{i=1}^N y_i y_i - \bar{y}\bar{y} \end{bmatrix} \\
 &= \frac{1}{N} \begin{bmatrix} \sum_{i=1}^N x_i x_i & \sum_{i=1}^N x_i y_i \\ \sum_{i=1}^N y_i x_i & \sum_{i=1}^N y_i y_i \end{bmatrix} - \begin{bmatrix} \bar{x}\bar{x} & \bar{x}\bar{y} \\ \bar{y}\bar{x} & \bar{y}\bar{y} \end{bmatrix} \quad (6)
 \end{aligned}$$

- The sample covariance matrix can then be written as

$$S = \frac{1}{N} \begin{bmatrix} x_1 & \cdots & x_N \\ y_1 & \cdots & y_N \end{bmatrix} \begin{bmatrix} x_1 & y_1 \\ \vdots & \vdots \\ x_N & y_N \end{bmatrix} - \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} \begin{bmatrix} \bar{x} & \bar{y} \end{bmatrix}. \quad (7)$$

Continued

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

$$\Sigma = \begin{bmatrix} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)^2 f(x, y) dx dy & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y) f(x, y) dx dy \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y) f(x, y) dx dy & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - \mu_y)^2 f(x, y) dx dy \end{bmatrix} \quad (8)$$

Discrete Sample

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

Continued

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

$$\begin{aligned} \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} &= \begin{bmatrix} (0.167 + 0.083 + 0.167 = 0.417) \times (1) \\ + (0.083 + 0.000 + 0.083 = 0.167) \times (0) \\ + (0.167 + 0.083 + 0.167 = 0.417) \times (-1) \\ (0.167 + 0.083 + 0.167 = 0.417) \times (1) \\ + (0.083 + 0.000 + 0.083 = 0.167) \times (0) \\ + (0.167 + 0.083 + 0.167 = 0.417) \times (-1) \end{bmatrix} \quad (9) \\ &= \begin{bmatrix} (0.417)(1) + (0.167)(0) + (0.417)(-1) \\ (0.417)(1) + (0.167)(0) + (0.417)(-1) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \end{aligned}$$

Continued

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

$$\begin{aligned}\text{Cov}[X, Y] &= \frac{1}{9} \sum_{i \in \{-1, 0, 1\}} \sum_{j \in \{-1, 0, 1\}} x_i y_j P(x_i, y_j) \\ &= (1)(-1)0.167 + (0)(-1)0.083 + \cdots + (-1)(1)0.167 = 0.\end{aligned}\tag{10}$$

- Theorem 4.3.2.** $V(X \pm Y) = V(X) + V(Y) \pm 2\text{Cov}(X, Y)$

$$\begin{aligned}V[X \pm Y] &= E[(X \pm Y)(X \pm Y)] \\ &= E[XX \pm 2XY + YY] - (E[X \pm Y])^2 \\ &= E[XX] + E[YY] \pm 2E[XY] - E[X]^2 \mp 2E[X]E[Y] - E[Y]^2 \\ &= V(X) + V(Y) \pm 2\text{Cov}(X, Y)\end{aligned}\tag{11}$$

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

$$\begin{bmatrix} X & Y \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = X + Y \quad (12)$$

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

$$\begin{aligned} \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{xy} & \sigma_{yy} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} &= \begin{bmatrix} \sigma_{xx} + \sigma_{xy} & \sigma_{xy} + \sigma_{yy} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= \sigma_{xx} + 2\sigma_{xy} + \sigma_{yy} \end{aligned} \quad (13)$$

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

$$V\left(\sum_{i=1}^N X_i\right) = \sum_{i=1}^N V(X_i) \quad (14)$$

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

$$\begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{xy} & \sigma_{yy} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \sigma_{xx} + \sigma_{yy} \quad (15)$$

if $\sigma_{xy} = 0$.

- Extending this result to three variables implies

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}' \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{matrix} \sigma_{11} + 2\sigma_{12} + 2\sigma_{13} + \\ \sigma_{22} + 2\sigma_{23} + \sigma_{33} \end{matrix} \quad (16)$$

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

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

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\sigma_{xx}}\sqrt{\sigma_{yy}}} \quad (17)$$

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

$$E[(X - E[X])(Y - E[Y])] = E[(X - E[X])(0)] = 0 \quad (18)$$

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

- It is now possible to derive the ordinary least squares estimator for a linear regression equation.
 - We define the ordinary least squares estimator as that set of parameters that minimizes the squared error of the estimate

$$\begin{aligned} \min_{\alpha, \beta} &= E \left[(Y - \alpha - \beta X)^2 \right] \\ \Rightarrow \min_{\alpha, \beta} E & \left[Y^2 - 2\alpha Y - 2\beta XY + \alpha^2 + 2\alpha\beta X + \beta^2 X^2 \right] \end{aligned} \quad (19)$$

The first order conditions for this minimization problem then becomes

$$\begin{aligned} \frac{\partial S}{\partial \alpha} &= -2E[Y] + 2\alpha + 2\beta E[X] = 0 \\ \frac{\partial S}{\partial \beta} &= -2E[XY] + 2\alpha E[X] + 2\beta E[X^2] = 0 \end{aligned} \quad (20)$$

Solving the first equation for α yields

$$\alpha = E[Y] - \beta E[X] \quad (21)$$

- Substituting this expression into the second first order condition yields

$$\begin{aligned}
 -E[XY] + (E[Y] - \beta E[X]) E[X] + \beta E[X^2] &= 0 \\
 -E[XY] + E[Y] E[X] + \beta (E[X^2] - (E[X])^2) &= 0 \\
 -\text{Cov}(X, Y) + \beta V(X) &= 0 \\
 \Rightarrow \beta &= \frac{\text{Cov}(X, Y)}{V(X)}
 \end{aligned}
 \tag{22}$$

- 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.

Conditional Mean and Variance

- **Definition 4.4.1** Let (X, Y) be a bivariate discrete random variable taking values (x_i, y_j) $i, j = 1, 2, \dots$. 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 $E[\phi(X, Y) | X]$ or by $E_{Y|X}[\phi(X, Y)]$, is defined by

$$E_{Y|X}[\phi(X, Y)] = \sum_{i=1}^{\infty} \phi(X, y_i) P(y_i|X) \quad (23)$$

- **Definition 4.4.2.** Let (X, Y) be a bivariate 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

$$E_{Y|X} [\phi(X, Y)] = \int_{-\infty}^{\infty} \phi(X, y) f(y|X) dy \quad (24)$$

- **Theorem 4.4.1** (Law of Iterated Means)

$E[\phi(X, Y)] = E_X E_{Y|X}[\phi(X, Y)]$. (Where the symbol E_x denotes the expectation with respect to X).

- **Theorem 4.4.2**

$$V(\phi(X, Y)) = E_X [V_{Y|X}[\phi(X, Y)]] + V_X [E_{Y|X}[\phi(X, Y)]] \quad (25)$$

- *Proof:*

$$V_{Y|X}[\phi] = E_{Y|X}[\phi^2] - (E_{Y|X}[\phi])^2 \quad (26)$$

Implies

$$E_X [V_{Y|X}(\phi)] = E[\phi^2] - E_X [(E_{Y|X}[\phi])^2] \quad (27)$$

- By definition of conditional variance

$$V_X (E_{Y|X} [\phi]) = E_X \left[(E_{Y|X} [\phi])^2 \right] - (E [\phi])^2 \quad (28)$$

- Adding these expressions yields

$$E_X [V_{Y|X} (\phi)] + V_X (E_{Y|X} [\phi]) = E [\phi^2] - (E [\phi])^2 = V (\phi) \quad (29)$$

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

Readings

- Moss 4.5
- Popper – Chapter 4: Falsifiability