Covariance and correlation

- For two random variables X and Y, the covariance between X and Y, denoted by $Cov(X, Y)$, is defined as

$$Cov(X, Y) = E\left(X - E(X)\right)\left(Y - E(Y)\right).$$

- The following properties follow from the definition of $Cov(X, Y)$ directly.
  (a) $Cov(X, X) = Var(X)$.
  (b) $Cov(X, Y) = Cov(Y, X)$.
  (c) $Cov(X_1 + X_2, Y) = Cov(X_1, Y) + Cov(X_2, Y)$.
  (d) $Cov(aX, Y) = aCov(X, Y)$ for a constant a.
  (e) $Cov(X + b, Y) = Cov(X, Y)$ for a constant b.

- The following property follows from Properties (a)–(e) listed above.

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y).$$

- Example 1. Suppose that $X \sim N(1, 4)$, $Y \sim N(-1, 1)$ and $Cov(X, Y) = 2$. Find $Var(0.5X - Y)$.

- When computing $Cov(X, Y)$ using the distribution of (X, Y), it is convenient to use

$$Cov(X, Y) = E(XY) - E(X)E(Y). \tag{1}$$

  Note that (1) implies that $Cov(X, Y) = 0$ if X and Y are independent and $E(X)$ and $E(Y)$ are finite.

- Example 2. Suppose that X and Y are independent, $X \sim N(0, 1)$ and $Y \sim N(0, 1)$. Find $Cov(X, (1 + X^2)Y)$.

- For two random variables X and Y such that $Var(X)$ and $Var(Y)$ are both positive, the correlation between X and Y, denoted by $Corr(X, Y)$, is defined as

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}.$$

- It can be showed that $|Corr(X, Y)| \leq 1$. When $|Corr(X, Y)| = 1$, we have $Y = a + bX$ (or $X = a + bY$) for some constant $b \neq 0$. $Corr(X, Y)$ is used to measure the strength of linear relation between X and Y.

  Example 3. Suppose that

$$P((X, Y) = (x, y)) = \begin{cases} 0.1 & \text{if } (x, y) = (1, -2); \\ 0.3 & \text{if } (x, y) = (2, -4); \\ 0.6 & \text{if } (x, y) = (3, -6); \\ 0 & \text{otherwise.} \end{cases}$$

  Find $Cov(X, Y)$ and $Corr(X, Y)$.

Sol. Note that X and Y are discrete random variables with joint PMF given in the problem. Computing $E(XY)$, $E(X)$, $E(Y)$, $E(X^2)$ and $E(Y^2)$ using

$$E(g(X,Y)) = \sum_{(x,y):P((X,Y)=(x,y))>0} g(x,y)P((X,Y)=(x,y))$$

gives

$$E(X) = 1 \times 0.1 + 2 \times 0.3 + 3 \times 0.6 = 2.5,$$
$$E(Y) = (-2) \times 0.1 + (-4) \times 0.3 + (-6) \times 0.6 = -5,$$
$$E(X^2) = 1^2 \times 0.1 + 2^2 \times 0.3 + 3^2 \times 0.6 = 6.7,$$
$$E(Y^2) = ((-2) \times 1)^2 \times 0.1 + ((-2) \times 2)^2 \times 0.3 + ((-2) \times 3)^2 \times 0.6 = 26.8,$$

and

$$E(XY) = (1)(-2) \times 0.1 + (2)(-4) \times 0.3 + (3)(-6) \times 0.6 = -13.4,$$

so

$$Cov(X,Y) = E(XY) - E(X)E(Y) = -13.4 - (2.5) \times (-5) = -0.9,$$
$$Var(X) = E(X^2) - (E(X))^2 = 6.7 - (2.5)^2 = 0.45,$$

and

$$Var(Y) = E(Y^2) - (E(Y))^2 = 26.8 - (-5)^2 = 1.8.$$

Thus

$$Corr(X,Y) = \frac{-0.9}{\sqrt{0.45 \times 1.8}} = -1.$$

- Best linear prediction. For two random variables X and Y such that $Var(X)$ and $Var(Y)$ are both finite and $Var(X) > 0$, the best linear predictor of Y based on X is $a_0 + b_0 X$, where

  $$(a_0, b_0) = \arg\min_{(a,b)} E(Y - (a + bX))^2.$$

  It can be shown that

  $$E(Y - (a + bX))^2 = Var(Y - bX) + [E(Y) - (a + bE(X))]^2$$

  and

  $$Var(Y - bX) \geq Var(Y - bX)|_{b=Cov(X,Y)/Var(X)},$$

  so $b_0 = Cov(X,Y)/Var(X)$, $a_0 = E(Y) - b_0 E(X)$ and

  $$\begin{aligned} E(Y - (a_0 + b_0 X))^2 &= Var(Y - b_0 X) \\ &= \frac{Var(X)Var(Y) - (Cov(X,Y))^2}{Var(X)} \qquad (2) \\ &= Var(Y)\left(1 - (Corr(X,Y))^2\right) \qquad (3) \end{aligned}$$

- When $E(Y|X)$ is a linear function of X, $E(Y|X)$ is the best linear predictor of Y and $E(Y - E(Y|X))^2 = E(Var(Y|X))$ is the right-hand side of the equality in (2).
- (3) implies that $|Corr(X,Y)| \leq 1$.

- Example 4. Suppose that (X,Y) has joint PDF $f_{X,Y}$, where

$$f_{X,Y}(x,y) = \left(\frac{\sqrt{3}}{2\pi}\right) e^{-(2x^2+2y^2+2xy)/2}.$$

(a) Find $Corr(X,Y)$.

(b) Find the best linear predictor of X based on Y.

(c) Find $E(X|Y)$.

Sol. Note that

$$\begin{aligned} f_{X,Y}(x,y) &= \left(\frac{\sqrt{3}}{2\pi}\right) e^{-(2x^2+2y^2+2xy)/2} \\ &= g(x,y)h(y), \end{aligned}$$

where

$$g(x,y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x+0.5y)^2/(2\sigma^2)} \tag{4}$$

with $\sigma^2 = 0.5$ and

$$h(y) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-y^2/(2\sigma_1^2)}$$

with $\sigma_1^2 = 2/3$. Since $g(\cdot, y)$ is a PDF of $N(-0.5y, \sigma^2)$, we have

$$\int_{-\infty}^{\infty} g(x,y)dx = 1$$

so $f_Y(y) = \int f_{X,Y}(x,y)dx = h(y) > 0$ for all $y \in R$. For $y \in R$, take $f_{X|Y=y}(x) = f_{X,Y}(x,y)/f_Y(y) = g(x,y)$ for $x \in R$. Then $\{f_{X|Y=y} : y \in R\}$ is a version of conditional PDF of X given Y.

(a) We first compute $E(Y)$, $E(Y^2)$, $E(X)$, $E(X^2)$ and $E(XY)$. Since $f_Y = h$ is a PDF of $N(0, \sigma_1^2)$, we have $E(Y) = 0$ and $E(Y^2) = Var(Y) = \sigma_1^2 = 2/3$. Moreover,

$$f_X(x) = \int f_{X,Y}(x,y)dy = \int f_{X,Y}(y,x)dy = h(x)$$

for all x, so $f_X = h = f_Y$, which implies that $E(X) = E(Y) = 0$ and $E(X^2) = Var(X) = Var(Y) = 2/3$. Finally,

$$\begin{aligned} E(XY) &= \int_{R^2} xy f_{X,Y} d(x,y) d(x,y) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyg(x,y)h(y)dxdy \\ &= \int_{-\infty}^{\infty} yh(y)\left(\int_{-\infty}^{\infty} xg(x,y)dx\right)dy \\ &= \int_{-\infty}^{\infty} yh(y)\left(-\frac{y}{2}\right)dy \\ &= -\frac{E(Y^2)}{2} = -\frac{1}{2}\left(\frac{2}{3}\right) = -\frac{1}{3}. \end{aligned}$$

Therefore,

$$Cov(X,Y) = E(XY) - E(X)E(Y) = -\frac{1}{3},$$

and

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{-1/3}{\sqrt{(2/3)(2/3)}} = -\frac{1}{2}.$$

(b) $Cov(X,Y) = E(XY) - E(X)E(Y) = -1/3$, $Var(Y) = 2/3$, $E(X) = 0 = E(Y)$. Take $b_0 = Cov(X,Y)/Var(Y) = -1/2$ and $a_0 = E(X) - b_0E(Y) = 0$, then the best linear predictor of X based on Y is $a_0 + b_0Y = -0.5Y$.

(c) Since $\{f_{X|Y=y} = g(\cdot,y) : y \in R\}$ is a version of conditional PDF of X given Y, we have

$$E(X|Y=y) = \int x f_{X|Y=y}(x)dx = \int xg(x,y)dx = -0.5y \qquad (5)$$

for $y \in R$. Here $\int xg(x,y)dx = -0.5y$ since $g(\cdot,y)$ is a PDF of $N(-0.5y,\sigma^2)$. From (5), $E(X|Y) = -0.5Y$.

- Definition of the expectation of a matrix of random variables. Suppose that W is an $n \times m$ matrix whose (i,j)-th element is a random variable $W_{i,j}$ for $1 \le i \le n$, $1 \le j \le m$, then $E(W)$ is the $n \times m$ matrix whose (i,j)-th element is $E(W_{i,j})$ for $1 \le i \le n$, $1 \le j \le m$.

- Definition of a covariance matrix. Suppose that $(X_1,\ldots,X_m)$ is a random vector. Let Σ be the $m \times m$ matrix whose (i,j)-th element is $Cov(X_i,X_j)$ for $1 \le i,j \le m$. Then Σ is called the covariance matrix (or the variance-covariance matrix, 共變異數矩陣) of $(X_1,\ldots,X_m)$. Σ can also be defined as $E(\boldsymbol{Y}\boldsymbol{Y}^T)$ with $\boldsymbol{Y} = \boldsymbol{X} - E(\boldsymbol{X})$ and $\boldsymbol{X} = (X_1,\ldots,X_m)^T$.

- Example 5. Suppose that $Z_1, \ldots, Z_m$ are IID[1] random variables and $Z_1 \sim N(0,1)$. Find the covariance matrix of $(Z_1, \ldots, Z_m)$.

  Ans. The covariance matrix of $(Z_1, \ldots, Z_m)$ is I_m: the $m \times m$ identity matrix.

- Fact 1 Suppose that W, W_1, W_2 are $n \times m$ matrices of random variables, and A and B are matrices of constants of sizes $\ell \times n$ and $m \times k$ respectively. Then
$$E(W_1 + W_2) = E(W_1) + E(W_2),$$
  $E(AW) = AE(W)$ and $E(WB) = E(W)B$. Moreover, if W is a matrix of constants, then $E(W) = W$.

- Fact 2 Suppose that $\boldsymbol{X} = (X_1, \ldots, X_m)^T$ is a random vector with covariance matrix Σ. Suppose that A is a $n \times m$ matrix of constant and $\boldsymbol{b}$ is a $n \times 1$ vector of constants. Then the covariance matrix of $A\boldsymbol{X} + \boldsymbol{b}$ is $A\Sigma A^T$.

- Example 6. Suppose that (X, Y) is a random vector with covariance matrix Σ, where
$$\Sigma = \begin{pmatrix} 1 & 3 \\ 3 & 25 \end{pmatrix}.$$
  Find $Var(X + Y)$.

  Sol. Let $\boldsymbol{b} = (1,1)^T$ and $W = (X,Y)^T$, then the covariance matrix of $\boldsymbol{b}^T W = (X + Y)$ is $\boldsymbol{b}^T \Sigma \boldsymbol{b} = (32)$, so $Var(X + Y) = 32$.

- Note.
  - When we define a matrix to represent a vector, we often take the matrix to be a matrix with one column. For instance, in Example 6, we define $W = (X,Y)^T$, which is a 2×1 random matrix.
  - For a 1×1 matrix $A = (a_{1,1})$, where $a_{1,1}$ is a real number, we sometimes treat A as the number $a_{1,1}$. For instance, in the solution to Example 6, we sometimes write $\boldsymbol{b}^T \Sigma \boldsymbol{b} = 32$.

[1] Recall that IID= Independent and Identically distributed.