CSE 312: Foundations of Computing II
Instructor: Alex Tsun
Date: 7/24/20
Lecture Topics: 5.4 Covariance and Correlation

[**Tags:** Covariance]

1. The covariance matrix of a random vector $\boldsymbol{Z} = (Z_1, Z_2, \dots, Z_n)$ is defined to the $n \times n$ matrix Σ such that $\Sigma_{ij} = Cov(X_i, X_j)$. Don't be too intimidated - it's just a way to store all the information we need – we won't be doing any linear algebra with it! The examples will help ☺.

$$\Sigma = \begin{bmatrix} Cov(Z_1, Z_1) = Var(Z_1) & Cov(Z_1, Z_2) & \dots & Cov(Z_1, Z_n) \\ Cov(Z_2, Z_1) & Cov(Z_2, Z_2) = Var(Z_2) & \dots & Cov(Z_2, Z_n) \\ \vdots & \ddots & \ddots & \vdots \\ Cov(Z_n, Z_1) & \dots & \dots & Cov(Z_n, Z_n) = Var(Z_n) \end{bmatrix}$$

   a. Let X_1, X_2, X_3, X_4 be iid (independent and identically distributed) random variables with mean μ and variance σ^2. Let $\boldsymbol{X} = (X_1, X_2, X_3, X_4)$ be a random vector with the rvs X_i as its components. What is the (4×4) covariance matrix of $\boldsymbol{X}$?
   b. Define $\boldsymbol{Y} = (Y_1, Y_2, Y_3)$ as follows.
      - $Y_1 = X_1 + X_2$
      - $Y_2 = X_2 + X_3$
      - $Y_3 = X_3 + X_4$.

      What is the (3×3) covariance matrix of $\boldsymbol{Y}$?

Solution:

a. We have an 4×4 matrix where we need to handle the diagonal and off-diagonal separately. We have $Cov(X_i, X_i) = Var(X_i) = \sigma^2$ for each $i = 1, \dots, 4$, so the diagonal is all σ^2. However, for $i \neq j$, $Cov(X_i, X_j) = 0$ since they are independent rvs. Hence, our covariance matrix is

$$\begin{bmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{bmatrix} = \sigma^2 I_4$$

b. The diagonals are

$$Cov(Y_i, Y_i) = Var(Y_i) = Var(X_i + X_{i+1}) = Var(X_i) + Var(X_{i+1}) = 2\sigma^2$$

(variance adds for independent rvs X_i and X_{i+1}).

Then

$$Cov(Y_1, Y_2) = Cov(X_1 + X_2, X_2 + X_3) \quad \text{[def of } Y\text{]}$$
$$= Cov(X_1, X_2) + Cov(X_1, X_3) + Cov(X_2, X_2) + Cov(X_2, X_3) \quad \text{[like FOIL]}$$
$$= 0 + 0 + \sigma^2 + 0 = \sigma^2$$

We also have this equal to $Cov(Y_2, Y_1)$ because covariance ignores order. Also, note this is equal to $Cov(Y_2, Y_3)$ and $Cov(Y_3, Y_2)$ as well by symmetry .

Finally, $Cov(Y_1, Y_3) = Cov(X_1 + X_2, X_3 + X_4) = 0$ since $X_1 + X_2$ is independent of $X_3 + X_4$! Putting this all together gives:

$$\begin{bmatrix} Var(Y_1) & Cov(X_1, X_2) & Cov(X_1, X_3) \\ Cov(X_2, X_1) & Var(Y_2) & Cov(X_2, X_3) \\ Cov(X_3, X_1) & Cov(X_3, X_2) & Var(Y_3) \end{bmatrix} = \begin{bmatrix} 2\sigma^2 & \sigma^2 & 0 \\ \sigma^2 & 2\sigma^2 & \sigma^2 \\ 0 & \sigma^2 & 2\sigma^2 \end{bmatrix}$$

[**Tags:** Similar to PSet4 Q4, Covariance]

2. Suppose we throw 12 balls independently and uniformly into 7 bins. For $i = 1, \dots, 7$, let X_i be the indicator/Bernoulli rv of whether bin i is empty. Let $\boldsymbol{X} = (X_1, \dots, X_7)$ be the random vector of indicators.
    a. What is the covariance matrix of $\boldsymbol{X}$?
    b. Let $Y = \sum_{i=1}^{7} X_i$ be the number of empty bins. What is $Var(Y)$?

<u>**Solution**</u>:

a.

We have $X_i \sim Ber\left(p = \left(\frac{6}{7}\right)^{12}\right)$ as the probability a particular bin is empty. Then, $E[X_i] = p$ and $Var(X_i) = p(1-p) \approx 0.13253432$. So that gives us the diagonal entries.

To find $Cov(X_i, X_j)$ for $i \neq j$, we need to compute $Cov(X_i, X_j) = E[X_i X_j] - E[X_i]E[X_j]$. But, the range of the product of 1's and $0's$ $X_i X_j$ is just $\Omega_{X_i X_j} = \{0,1\}$, so $X_i X_j \sim Ber(q)$ for some q.

Hence $E[X_i X_j] = P(X_i X_j = 1) = P(X_i = 1, X_j = 1) = P(\text{bins } i \text{ and } j \text{ are both empty}) = \left(\frac{5}{7}\right)^{12}$ and

$$Cov(X_i, X_j) = E[X_i X_j] - E[X_i]E[X_j] = \left(\frac{5}{7}\right)^{12} - \left(\frac{6}{7}\right)^{12}\left(\frac{6}{7}\right)^{12} \approx -0.0071$$

Hence the covariance matrix has $\Sigma_{ii} \approx 0.13253432$ and $\Sigma_{ij} \approx -0.0071$.

b. We have that

$$\begin{aligned} Var(Y) &= Cov(Y, Y) \\ &= Cov\left(\sum_{i=1}^{7} X_i, \sum_{i=1}^{7} X_j\right) \\ &= \sum_{i=1}^{7}\sum_{j=1}^{7} Cov(X_i, X_j) \end{aligned}$$

$$= \sum_{i=1}^{7} Var(X_i) + 2\sum_{i<j} Cov(X_i, X_j)$$

$$\approx 7 \cdot 0.13253432 + 2\binom{7}{2}(-0.0071) \approx 0.62954$$