CS 4980/6980: Data Science © Spring 2018

Lecture 18: Random variable generation using Covariance Matrix

INSTRUCTOR: DANIEL L. PIMENTEL-ALARCÓN Scribed by: Naga Jagadeesh Mutala

This is preliminary work and has not been reviewed by instructor. If you have comments about typos, errors, notation inconsistencies, etc., please email the scribers.

18.1 Introduction

A Covariance matrix is a square matrix that contains the variances and covariances associated with several variables. The diagonal elements of the matrix contain the variances of the variables and the off-diagonal elements contain the covariances between all possible pairs of variables.

18.2 Covariance and Covariance Matrix:

If x, y have means μ_x, μ_y respectively then Covariance is as follows:

$$cov(x, y) := \mathbb{E}[(x - \mu_x)(y - \mu_y)] := cov(y, x)$$

Covariance Matrix for random variables x and y represented by vector Z= $\begin{bmatrix} x \\ y \end{bmatrix}$ is given by:

$$cov(Z) = \begin{bmatrix} cov(x,x) & cov(x,y) \\ cov(y,x) & cov(y,y) \end{bmatrix}$$

$$= \begin{bmatrix} var(x) & cov(x,y) \\ cov(y,x) & var(y) \end{bmatrix}$$

18.3 Goal: Play GOD

The goal is to generate random variables x^* and y^* represented by vector Z= $\begin{bmatrix} x^* \\ y^* \end{bmatrix}$ for a given covariance cov(x, y) and covariance matrix cov(Z^*).

<u>Tricks to transform random variables</u>: -

- If Z has mean μ then Z + k has mean μ + k.
- If Z has covariance matrix $\mathbf{C}$, then $\mathbf{A}$Z has covariance matrix $\mathbf{ACA}^T$.

Solution:
Let $x, y \overset{iid}{\sim} N(0,1)$
NOTE: $\overset{iid}{\sim}$ indicates the variables x,y are drawn independently and indentically distributed.
Then build Z as follows: Z=$\begin{bmatrix} x \\ y \end{bmatrix}$
Since var(x) = var(y) = 1 and cov(x, y) = 0 as x is independent of y, Covariance matrix is given by:

$$cov(Z) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} := \mathbf{I}_2$$

Multiply Z by a matrix $\mathbf{A}$ to obtain,$Z^* = \mathbf{A}$Z such that

$$cov(Z^*) = \mathbf{AIA}^T$$
$$= \mathbf{AA}^T$$
$$\Rightarrow \mathbf{A} = \sqrt{\mathbf{C}^*}$$

where $\sqrt{\mathbf{C}}$ is the matrix such that $\sqrt{\mathbf{C}^*} * \sqrt{\mathbf{C}^*} = \mathbf{C}$. If $\mathbf{C}$ is symmetric then $\sqrt{\mathbf{C}}$ is also symmetric. i.e. $\mathbf{C} = \sqrt{\mathbf{C}^T}$.

$$\therefore cov(Z^*) = cov(\mathbf{A}Z) = \mathbf{AA}^T = \sqrt{\mathbf{C}^*} * (\sqrt{\mathbf{C}^*})^T = \sqrt{\mathbf{C}^*} * \sqrt{\mathbf{C}^*} = \mathbf{C}^*$$

18.4 Special Homework

Generate N random variables $\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} ... \begin{bmatrix} x_N \\ y_N \end{bmatrix}$ iid according to $N(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{bmatrix})$
Solution:

- Generate $Z_i' = \begin{bmatrix} x_i' \\ y_i' \end{bmatrix} \sim$ N(0,1)
- Multiply each of them with $\sqrt{\mathbf{C}}$ then:

$$Z_i = \sqrt{\mathbf{C}} * Z_i' \sim N(0, \mathbf{C}).$$

- Goal: Estimate μ and $\mathbf{C}$.

$$\widehat{\mu} = \frac{1}{N} \sum_{i=1}^{N} Z_i$$

$$\widehat{\mathbf{C}} = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{bmatrix}$$

As we know,

$$\sigma_x^2 = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{X}_i - \mu_x)^2$$

$$\sigma_y^2 = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{Y}_i - \mu_y)^2$$

$$\sigma_{xy} = \sigma_{yx} = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{X}_i - \mu_x)(\mathbf{Y}_i - \mu_y)$$

Therefore

$$\widehat{\mathbf{C}} = \begin{bmatrix} \frac{1}{N}\sum_{i=1}^{N}(\mathbf{X}_i-\mu_x)^2 & \frac{1}{N}\sum_{i=1}^{N}(\mathbf{X}_i-\mu_x)(\mathbf{Y}_i-\mu_y) \\ \frac{1}{N}\sum_{i=1}^{N}(\mathbf{X}_i-\mu_x)(\mathbf{Y}_i-\mu_y) & \frac{1}{N}\sum_{i=1}^{N}(\mathbf{Y}_i-\mu_y)^2 \end{bmatrix}$$

$$= \frac{1}{N}\sum_{i=1}^{N}\begin{bmatrix}\mathbf{X}_i-\mu_x \\ \mathbf{Y}_i-\mu_y\end{bmatrix}\begin{bmatrix}\mathbf{X}_i-\mu_x & \mathbf{Y}_i-\mu_y\end{bmatrix}$$

$$= \frac{1}{N}\sum_{i=1}^{N}(Z-\mu)(Z-\mu)^T$$

- Analyze accuracy as a function of N, $\mu - error = \|\widehat{\mu}-\mu\|_2^2$. C-error $= \|\widehat{C}-C\|_2^2$

18.5 Covariance in case of Multiple Random Variables

Suppose we have three random variables w,x,y such that Z= $\begin{bmatrix} w \\ x \\ y \end{bmatrix}$ then covariance of Z is as below:

$$cov(Z) = \begin{bmatrix} \sigma_w^2 & \sigma_{wx} & \sigma_{wy} \\ \sigma_{xw} & \sigma_x^2 & \sigma_{xy} \\ \sigma_{yw} & \sigma_{yx} & \sigma_y^2 \end{bmatrix} := \mathbf{C}$$

$$\widehat{\mathbf{C}} = \frac{1}{N}\sum_{i=1}^{N}\begin{bmatrix}\mathbf{W}_i \\ \mathbf{X}_i \\ \mathbf{Y}_i\end{bmatrix}\begin{bmatrix}\mathbf{W}_i\ \mathbf{X}_i\ \mathbf{Y}_i\end{bmatrix}$$

18.6 Conclusion:

We learned about covariance matrix and how to generate random variables for a given covariance and covariance matrix.