Lecture Notes in Stochastic Processes

Multivariate Gaussian Distribution

Rohith Krishna

Madras School of Economics
Chennai, India.

mailto: pgdm19rohith@mse.ac.in

Contents

1 Random Vectors **3**

1.1 Expected value of random vectors 3

1.2 Covariance Matrix for a random vector 4

1.3 Properties of the covariance matrix 5

2 Bivariate Gaussian Random Vectors **6**

2.1 Joint PDF of Bivariate Normal 8

3 Multivariate Gaussian Random Vector **10**

3.1 PDF of a Gaussian Vector 10

4 Memoryless Property **11**

4.1 Memoryless property of Exponential Distribution 11

4.2 Proving the memoryless property 12

4.3 Shifting of PDF leads to memoryless property 13

1 Random Vectors

The idea of ***random vectors*** turns out to be extremely handy when dealing with a number of random variables. In this section we define random vectors of order n, its expectation value and other entities such as the correlation and the covariance matrix for a random vector.

Definition 1.1 (Random Variable) Consider a given probability space $(\Omega, \mathscr{F}, \mathbb{P})$, where Ω is the sample space, $\mathscr{F}$ refers to σ-algebra, the event-space which is a collection of all possible events and $\mathbb{P}$ is the probability measure. A *random variable*, X is a function with domain Ω and counterdomain the real line. The random variable X is defined such that the set A_r defined by $A_r = \{\omega : X(\omega) \leq r\}$ belongs to $\mathscr{F}$ for every real number r.

The random variable function therefore maps every element in the sample space to the real number line. One can think of a random experiment where Ω is the totality of all outcomes of the random experiment. The random variable (RV) X takes every outcome of the experiment to a real number. The reason why we mention the event space, $\mathscr{F}$ in the definition is that we are usually in events defined by random variables. This is why we require a collection of ω's for which $X(\omega) \leq r$ to form an event; we would use the idea of random variable to describe events.

Definition 1.2 (Random Vector) We define a *random vector* $\vec{X}$ as a column vector with a collection of n random variables, say, $\{X_1, X_2, \ldots, X_n\}$ as its components. That is, $\vec{X}^T = [X_1 \;\; X_2 \;\; \ldots \;\; X_n]$.

Let x be a realization of the random vector $\vec{X}$. We define the cumulative distribution function (CDF) of $\vec{X}$ as:

$$\begin{aligned} F_{\vec{X}}(\vec{x}) &= F_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) \\ F_{\vec{X}}(\vec{x}) &= P\big[X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n\big] \end{aligned} \tag{1.1}$$

Now, if the X_i's are jointly continuous, then the probability density function (PDF) of $\vec{X}$, denoted by $f_{\vec{X}}(\vec{x})$ is the joint PDF of the X_i random variables.

$$f_{\vec{X}}(\vec{x}) = f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) \tag{1.2}$$

1.1 Expected value of random vectors

The expected value of a random vector is the vector with its entries as the expected values of component random variables. Thus, for a random vector $\vec{X}$ with n random variables, we have,

$$E[\vec{X}^T] = \Big[E[X_1] \;\; E[X_2] \;\; \ldots \;\; E[X_3] \;\Big] = \vec{\mu}_X^T \tag{1.3}$$

In extension, the expected value for a sum of random vectors, is the sum of individual expected values of each random vector.

$$E\left[\overrightarrow{X_1}^T + \overrightarrow{X_2}^T + \cdots + \overrightarrow{X_k}^T\right] = E\left[\overrightarrow{X_1}^T\right] + E\left[\overrightarrow{X_2}^T\right] + \cdots + E\left[\overrightarrow{X_k}^T\right] \tag{1.4}$$

$$E\left[\overrightarrow{X_1}^T + \overrightarrow{X_2}^T + \cdots + \overrightarrow{X_k}^T\right] = E\begin{bmatrix} X_{11} \\ X_{12} \\ \vdots \\ X_{1n} \end{bmatrix} + E\begin{bmatrix} X_{21} \\ X_{22} \\ \vdots \\ X_{2n} \end{bmatrix} + \cdots + E\begin{bmatrix} X_{k1} \\ X_{k2} \\ \vdots \\ X_{kn} \end{bmatrix}$$

$$E\left[\overrightarrow{X_1}^T + \overrightarrow{X_2}^T + \cdots + \overrightarrow{X_k}^T\right] = \begin{bmatrix} E[X_{11}] \\ E[X_{12}] \\ \vdots \\ E[X_{1n}] \end{bmatrix} + \begin{bmatrix} E[X_{21}] \\ E[X_{22}] \\ \vdots \\ E[X_{2n}] \end{bmatrix} + \cdots + \begin{bmatrix} E[X_{k1}] \\ E[X_{k2}] \\ \vdots \\ E[X_{kn}] \end{bmatrix}$$

1.2 Covariance Matrix for a random vector

Consider a series of random variables $X_1, X_2, \dots X_n$. The realizations of all the random variables forms our sample data. Here, we consider each variable as a column and every realization adds a row to our sample dataset. Note that for every column X_i we can determine the variance - this is the $Var(X_i)$, where each value of random variable X_i is a realization. Likewise, between two columns, X_i and X_j one can determine the covariance, $Cov(X_i, X_j)$. Now we can construct the $n \times n$ covariance matrix C_X, such that the it contains variance terms along the diagonal, and covariance between all X_i's and X_j's.

The set of random variables $X_1, X_2, \dots X_n$ thought of as a random vector $\overrightarrow{X}$ can have a correlation and covariance matrix defined as follows. The correlation matrix for a random vector is defined as $R_{\overrightarrow{X}} = E\left[\overrightarrow{X}\overrightarrow{X}^T\right]$.

$$R_{\overrightarrow{X}} = E\left[\overrightarrow{X}\overrightarrow{X}^T\right] = \begin{bmatrix} E[X_1^2] & E[X_1X_2] & \dots & E[X_1X_n] \\ E[X_sX_1] & E[X_2^2] & \dots & E[X_2X_n] \\ \vdots & \vdots & \ddots & \vdots \\ E[X_nX_1] & E[X_nX_2] & \dots & E[X_n^2] \end{bmatrix} \tag{1.5}$$

We define variance of a random variable X_i as $Var(X_i) = E[(X_i - \mu_{X_i})^2]$ and the covariance between two random variables X_i and X_j as $E[(X_i - \mu_{X_i})(X_j - \mu_{X_j})]$. In a similar manner, we define the covariance matrix $(C_{\overrightarrow{X}})$ for a random vector $\overrightarrow{X}$ as:

$$C_{\overrightarrow{X}} = E\left[(\overrightarrow{X}_i - \vec{\mu}_{X_i})(\overrightarrow{X}_j - \vec{\mu}_{X_j})^T\right] = \begin{bmatrix} Var(X_1) & Cov(X_1, X_2) & \dots & Cov(X_1, X_n) \\ Cov(X_2, X_1) & Var(X_2) & \dots & Cov(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_n, X_1) & Cov(X_n, X_2) & \dots & Var(X_n) \end{bmatrix} \tag{1.6}$$

In a similar manner, we can also define for any two random vectors $\overrightarrow{X}$ and $\overrightarrow{Y}$, cross-correlation ($R_{\overrightarrow{X}\overrightarrow{Y}}$) and cross-covariance ($C_{\overrightarrow{X}\overrightarrow{Y}}$) matrices.

$$\begin{aligned} R_{\overrightarrow{X}\overrightarrow{Y}} &= E\Big[\overrightarrow{X}\overrightarrow{Y}^T\Big] \\ C_{\overrightarrow{X}\overrightarrow{Y}} &= E\Big[(\overrightarrow{X} - \vec{\mu}_X)(\overrightarrow{Y} - \vec{\mu}_Y)^T\Big] \end{aligned} \tag{1.7}$$

Note. Here, $\mu_{\overrightarrow{X}} = E[\overrightarrow{X}]$ is the expected value of random vector X and $\mu_{\overrightarrow{Y}} = E[\overrightarrow{Y}]$ is the expected value of random vector Y.

$$\vec{\mu}_X = E\big[\overrightarrow{X}\big] = E\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} \tag{1.8}$$

1.3 Properties of the covariance matrix

The covariance matrix generalizes the variance of a random variable to that of a random vector. In essence the covariance matrix for a random vector is equivalent to the covariance matrix generated by considering n random variables. This matrix exhibits certain interesting properties which we shall make use of in subsequent sections. These properties are first stated and subsequently proved here.

P1. *The covariance matrix is symmetric.* $C_{\overrightarrow{X}} = C_{\overrightarrow{X}}^T$. Consequently, $C_{\overrightarrow{X}}$ can be diagonalized and *all the eigenvalues of* $C_{\overrightarrow{X}}$ *are real.*

P2. *If* $\overrightarrow{X}$ *is a real random vector* $\overrightarrow{X} \in \mathbb{R}^n$*, that is* ($X_i \in \mathbb{R}, \ \forall \ i$)*, then* $C_{\overrightarrow{X}}$ *is positive semidefinite (PSD).* This means that $\forall \ \vec{b} \neq \vec{0}, \ \vec{b}^T C_{\overrightarrow{X}} \vec{b} \geq 0$.

P3. *For a random vector* $\overrightarrow{X}$*, the covariance matrix* $C_{\overrightarrow{X}}$ *is positive definite (PD)* $\overset{iff.}{\Longleftrightarrow}$ *all eigenvalues of* $C_{\overrightarrow{X}} > 0$. Equivalently, $C_{\overrightarrow{X}}$ *is PD* $\overset{iff.}{\Longleftrightarrow}$ $det(C_{\overrightarrow{X}}) > 0$.

Theorem 1.1 *The covariance matrix is symmetric.* $C_{\overrightarrow{X}} = C_{\overrightarrow{X}}^T$. Consequently, $C_{\overrightarrow{X}}$ can be diagonalized and *all the eigenvalues of* $C_{\overrightarrow{X}}$ *are real.*

Proof. $C_{\overrightarrow{X}} = \big[C_{i,j}\big] = Cov(X_i, X_j) = Cov(X_j, X_i) = \big[C_{j,i}\big] = C_{\overrightarrow{X}}^T$. The covariance between two variables does not change when their positions are interchanges. As a consequence of symmetric nature of covariance matrix, we have the corollary that all the eigenvalues of $C_{\overrightarrow{X}}$ are real and its eigenvectors are orthogonal. ∎

Note. We say that a matrix A is positive semidefinite (PSD) if $\forall \ \vec{b} \neq \vec{0}, \ \vec{b}^T A \vec{b} \geq 0$. Likewise, A is positive definite (PD) if $\forall \ \vec{b} \neq \vec{0}, \ \vec{b}^T A \vec{b} > 0$.

Theorem 1.2 *If $\vec{X}$ is a real random vector $\vec{X} \in \mathbb{R}^n$, that is ($X_i \in \mathbb{R}, \ \forall \ i$), then $C_{\vec{X}}$ is positive semidefinite (PSD).*

Proof. Let $\vec{b} \in \mathbb{R}^N$ be a fixed vector with n components. We define a random variable $Y = \vec{b}^T[\vec{X} - E[\vec{X}]]$. Note that Y here is a random variable and not a random vector. Further, the random variable Y^2 is non-negative, ie., $E[Y^2] \geq 0$.

$$E[Y^2] \geq 0 \implies E[YY] \geq 0 \implies E[YY^T] \geq 0 \tag{1.9}$$

Note here that Y^T is not a vector but a number. Thus $Y^T = Y$. Substituting the value of Y in the above expression,

$$\begin{aligned}
E[YY^T] \geq 0 &\implies E\left[\left(\vec{b}^T[\vec{X} - E[\vec{X}]]\right)\left(\vec{b}^T[\vec{X} - E[\vec{X}]]\right)^T\right] \geq 0 \\
&\implies E\left[\vec{b}^T\left(\vec{X} - E[\vec{X}]\right)\left(\vec{X} - E[\vec{X}]\right)^T \vec{b}\right] \geq 0 \\
&\implies \vec{b}^T \underbrace{E\left[\left(\vec{X} - E[\vec{X}]\right)\left(\vec{X} - E[\vec{X}]\right)^T\right]}_{C_{\vec{X}},\ \text{covariance matrix of } \vec{X}} \vec{b} \geq 0 \\
&\implies \underbrace{\vec{b}^T C_{\vec{X}} \vec{b} \geq 0}_{\substack{\text{condition for positive} \\ \text{semidefiniteness (PSD)}}} \implies \boxed{C_{\vec{X}} \text{ is PSD.}}
\end{aligned} \tag{1.10}$$

■

2 Bivariate Gaussian Random Vectors

Definition 2.1 (Bivariate Gaussian) Two random variables X and Y are *bivariate Gaussian* or *jointly normal* if the variable $aX + bY$ has a normal distribution for all $a \in \mathbb{R}$ and $b \in \mathbb{R}$.

Note. If X and Y are Bivariate Gaussian and

- if $a = b = 0 \implies aX + bY = 0 \implies$ Constant zero is a normal random variable with mean 0 and variance 0.
- if $a = 1, \ b = 0 \implies aX + bY = X \implies X$ is a normal random variable.
- if $a = 0, \ b = 1 \implies aX + bY = Y \implies Y$ is a normal random variable.

Theorem 2.1 If $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ are jointly normal random variables, setting $a = b = 1 \implies aX + bY = X + Y$, then the resultant random variable $X+Y \sim \mathcal{N}(\mu_X+\mu_Y, \sigma_X^2+\sigma_Y^2+2\rho_{XY}\sigma_X\sigma_Y)$. Essentially, a linear combination fo two normal variables X and Y gives a joint normal variable $X + Y$.

Example 1 Let $Z_1 \sim \mathscr{N}(0,1)$ and $Z_2 \sim \mathscr{N}(0,1)$ be two independent random variables. Define $X = Z_1$ and $Y = \rho Z_1 + \sqrt{1-\rho^2} Z_2$, where $\rho \in (-1,+1)$. Show that X and Y are bivariate normal.

We know from the above theorem that since Z_1 and Z_2 are independent normal distributions, they are also jointly normal. We can therefore attempt to obtain the joint PDF of Z_1 and Z_2. Because, these are independent, we have,

$$\begin{aligned}
f_{Z_1,Z_2}(z_1,z_2) &= \underbrace{f_{Z_1}(z_1).f_{Z_2}(z_2)}_{\substack{\text{because } Z_1 \text{ and } Z_2 \\ \text{are independent RVs.}}} \\
f_{Z_1,Z_2}(z_1,z_2) &= \underbrace{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z_1^2}}_{f_{Z_1}(z_1)} \underbrace{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z_2^2}}_{f_{Z_2}(z_2)} \\
f_{Z_1,Z_2}(z_1,z_2) &= \frac{1}{2\pi} e^{-\frac{1}{2}(z_1^2+z_2^2)} \qquad (2.1)
\end{aligned}$$

In order to prove that X and Y are bivariate normal, we need to show first that $aX + bY$ is normal $\forall\ a, b \in \mathbb{R}$.

$$\begin{aligned}
aX + bY &= aZ_1 + b\left[\rho Z_1 + \sqrt{1-\rho^2} Z_2\right] \\
aX + bY &= \underbrace{(a + b\rho)}_{\text{constant coeff.}}\ Z_1 + \underbrace{\left(b\sqrt{1-\rho^2}\right)}_{\text{constant coefficient}}\ Z_2 \qquad (2.2)
\end{aligned}$$

From the above, we see that $aX + bY$ is expressed as a linear combination of standard normals Z_1 and Z_2. Thus it must be the case that $aX + bY$ is normal. ■

Example 2 Let $Z_1 \sim \mathscr{N}(0,1)$ and $Z_2 \sim \mathscr{N}(0,1)$ be two independent random variables. Define $X = Z_1$ and $Y = \rho Z_1 + \sqrt{1-\rho^2} Z_2$, where $\rho \in (-1,+1)$. Obtain the expected value and variance of X and Y.

$$\begin{aligned}
Var(X) &= Var(Z_1) = 1 \\
Var(Y) &= \underbrace{Var(\rho Z_1)}_{=\rho^2 Var(Z_1)} + \underbrace{Var(\sqrt{(1-\rho^2)})}_{=(1-\rho^2)Var(Z_2)} \\
E[X] &= E[Z_1] = 0 \\
E[Y] &= \rho \underbrace{E[Z_1]}_{=0,\text{ given}} + \sqrt{1-\rho^2}\ \underbrace{E[Z_2]}_{=0,\text{ given}} = 0
\end{aligned}$$

Recall the following properties about covariance of two variables:

- $Cov(X, Z + W) = Cov(X, Z) + Cov(X, W)$
- $Cov(Z, aW) = aCov(Z, W)$

$$
\begin{aligned}
\rho(X,Y) &= Cov(X,Y) = Cov(Z_1, \rho Z_1 + \sqrt{1-\rho^2}Z_2) \\
\rho(X,Y) &= Cov(Z_1, \rho Z_2) + Cov(Z_1, \sqrt{1-\rho^2}Z_2) \\
\rho(X,Y) &= \rho \underbrace{Cov(Z_1, Z_1)}_{=Var(Z_1)=1} + \sqrt{1-\rho^2} \underbrace{Cov(Z_1, Z_2)}_{\substack{=0\text{, since } Z_1 \text{ and } Z_2 \\ \text{are independent}}} \\
\rho(X,Y) &= \rho
\end{aligned}
$$

2.1 Joint PDF of Bivariate Normal

Let $Z_1 \sim \mathscr{N}(0,1)$ and $Z_2 \sim \mathscr{N}(0,1)$ be two independent random variables. Define random variables $X = Z_1$ and $Y = \rho Z_1 + \sqrt{1-\rho^2}Z_2$, where $\rho \in (-1,+1)$. In this case we have, $Var(X) = Var(Y) = 1$ and $E[X] = E[Y] = 0$. In such a case, X and Y have the bivariate normal distribution with correlation coefficient ρ if their joint PDF $f_{X,Y}(x,y)$ is:

$$
f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{\left(-\frac{1}{2(1-\rho^2)}\{x^2 - 2\rho xy + y^2\}\right)} \tag{2.3}
$$

In order to prove a very important result about the PDF of a bivariate normal distribution, we first construct two normal random variables $X \sim \mathscr{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathscr{N}(\mu_Y, \sigma_Y^2)$, such that $\rho(X,Y) = \rho$. Let $Z_1 \sim \mathscr{N}(0,1)$ and $Z_2 \sim \mathscr{N}(0,1)$ be independent random variables.

$$
\begin{aligned}
X &= \sigma_X Z_1 + \mu_X \\
Y &= \sigma_Y \left[\rho Z_1 + \sqrt{1-\rho^2} Z_2\right] + \mu_Y
\end{aligned} \tag{2.4}
$$

X and Y have a bivariate normal distribution if the joint PDF

$$
f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-u^2} \tag{2.5}
$$

$$
u^2 = \frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right] \tag{2.6}
$$

where, $\mu_x, \mu_Y \in \mathbb{R}$ and $\sigma_X > 0, \sigma_Y > 0$ and $\rho \in (-1,1)$ are all parameters of the PDF of the joint normal distribution $f_{X,Y}(x,y)$.

Theorem 2.2 Let X and Y be two bivariate normal random variable with joint PDF $f_{X,Y}(x,y)$ given by:

$$
f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]} \tag{2.7}
$$

Then, there exist independent standard normal variables Z_1 and Z_2 such that X and Y are given by:

$$\begin{aligned} X &= \sigma_X Z_1 + \mu_X \\ Y &= \sigma_Y \left[\rho Z_1 + \sqrt{1-\rho^2} Z_2\right] + \mu_Y \end{aligned} \tag{2.8}$$

Note. We can generate X and Y from the standard normal Z_1 and Z_2. Using this method, we can generate samples from the bivariate normal distribution.

On Independence of RVs

If X and Y are independent random variables, then $Cov(X, Y) = 0$, that is X and Y are uncorrelated. But the inverse is not true.

If X and Y are jointly normal random variables AND they are uncorrelated $Cov(X, Y) = 0$, then X and Y are independent random varibles.

Theorem 2.3 If X and Y are bivariate normal and are uncorrelated $Cov(X, Y) = 0$, then they are independent random variables.

Proof. If $\rho = \rho(X, Y) = Cov(X, Y) = 0$ AND $(X, Y) \sim$ *bivariate normal*, then from (2.6) and (2.5) we have,

$$\begin{aligned} u^2 &= \frac{1}{2}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right] \\ f_{X,Y}(x,y) &= \frac{1}{2\pi\sigma_X\sigma_Y} e^{-\frac{1}{2}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]} \\ f_{X,Y}(x,y) &= \underbrace{\frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{1}{2}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right]}}_{\substack{\text{marginal distribution of X} \\ f_X(x)}} \underbrace{\frac{1}{\sqrt{2\pi}\sigma_Y} e^{-\frac{1}{2}\left[\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]}}_{\substack{\text{marginal distribution of Y} \\ f_Y(y)}} \\ f_{X,Y}(x,y) &= f_X(x) f_Y(y) \quad \forall\ x, y \in \mathbb{R} \end{aligned} \tag{2.9}$$

Thus, X and Y are independent random variables. ■

In the subsequent sections, we extend the bivariate normal distribution to a case with n-jointly distributed random variables. We shall see precisely how the idea of *random vectors* aids us in this context. Here, we will also make use of functions of random vectors to showcase the same.

3 Multivariate Gaussian Random Vector

Definition 3.1 (Multivariate Gaussian) Random variables $\{X_1, X_2, \ldots X_n\}$ are said to be *multivariate Gaussian* (or jointly normal) if $\forall\ a_1, a_2, \ldots a_n in \mathbb{R}$, the random variable W (which is a linear combination of all X_i's) is a normal random variable. Here,

$$W = \underbrace{a_1 X_1 + a_2 X_2 + \cdots + a_n X_n}_{\substack{W \text{ is a linear combination} \\ \text{of all } n \text{ random variables.}}} \tag{3.1}$$

Expressed in terms of vector notation, we have,

$$W = \vec{a}^T \vec{X} \tag{3.2}$$

where, $\vec{a}^T = [a_1\ a_2\ \ldots\ a_n]$ is a constant coefficient vector and $\vec{X}^T = [X_1\ X_2\ \ldots\ X_n]$ is the random vector. Note that in the trivial case if $\vec{a} = \vec{0} \implies W \sim \mathcal{N}(0, 0)$.

Definition 3.2 (Gaussian Vector) A random vector $\vec{X}$ is a Gaussian (or normal) random vector if the n-random variables $\{X_1, X_2, \ldots X_n\}$ (that make up its components) are jointly normal.

3.1 PDF of a Gaussian Vector

Consider the standard normal vector $\vec{Z}$ comprised of n identical but independent standard normal random variables Z_i. That is $Z_i \sim \mathcal{N}(0, 1)$ and Z_i's are independent. This implies,

$$\begin{aligned}
f_{\vec{Z}}(\vec{z}) &= f_{Z_1, Z_2, \ldots, Z_n}(z_1, z_2, \ldots, z_n) \\
f_{\vec{Z}}(\vec{z}) &= \underbrace{f_{Z_1}(z_1).f_{Z_2}(z_2)\ldots f_{Z_n}(z_n)}_{\substack{\text{Since each of } Z_i \text{ are independent} \\ \text{of the others } Z_j,\text{ where } i \neq j}} \\
f_{\vec{Z}}(\vec{z}) &= \prod_{i=1}^{n} f_{Z_i}(z_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z_i^2} \\
f_{\vec{Z}}(\vec{z}) &= \underbrace{\frac{1}{(2\pi)^{(n/2)}} e^{-\frac{1}{2}\sum_{i=1}^{n} z_i^2}}_{\substack{\text{product becomes sum} \\ \text{within the exponent.}}} = \underbrace{\frac{1}{(2\pi)^{(n/2)}} e^{-\frac{1}{2}\vec{z}^T.\vec{z}}}_{\substack{\text{expressed in} \\ \text{vector form.}}} \\
f_{\vec{Z}}(\vec{z}) &= \frac{1}{(2\pi)^{(n/2)}} e^{-\frac{1}{2}\vec{z}^T.\vec{z}}
\end{aligned} \tag{3.3}$$

The above equation (3.3) is the PDF of a ***Standard Gaussian*** vector. We may now, extend this PDF for a general normal random vector $\vec{X}$ with mean vector $\vec{m}$ and covariance matrix $C_{\vec{X}}$

Conversion from standard normal to general normal

Let Z be a standard normal random variable such that $Z \sim \mathcal{N}(0,1)$, and let X be a general normal random variable with finite mean μ and variance σ. We can obtain X from Z using the transformation:

$$X = \sigma Z + \mu$$

Likewise, one can extend this to the case of random vector $\vec{X}$ with mean vector $\vec{m}$, where $\vec{m} = E[\vec{X}]$ and covariance matrix $C_{\vec{X}}$ such that

$$C_{\vec{X}} = E\big[(\vec{X} - \vec{m})(\vec{X} - \vec{m})^T\big]$$

Our objective now is to find such a transformation. It turns out that if $C_{\vec{X}}$ can be decomposed into $C_{\vec{X}} = AA^T$, then

$$\vec{X} = A\vec{Z} + \vec{m}$$

The matrix A can be thought of as a kind of standard deviation matrix, containing the roots of variances and covariances.

4 Memoryless Property

The memoryless property is a unique property exhibited by the exponential distribution and its discrete analog, the geometric distribution. We shall observe in this section how the distribution is forgetful of its past and at any given moment, the processes modelled by this distribution behaves as though it is time zero.

4.1 Memoryless property of Exponential Distribution

Consider a random variable X such that it follows an exponential ditribution with λ as the parameter, $X \sim \mathscr{E}xp(\lambda)$. We know that the PDF and CDF of the exponential distribution are given by:

$$f_X(x) = \lambda e^{-\lambda x} u(x) \tag{4.1}$$
$$F_X(x) = (1 - e^{-\lambda x}) u(x) \tag{4.2}$$

where $u(x)$ is the unit step function which is given by:

$$u(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{4.3}$$

We also note that the expected value and variance of the exponential distribution is,

$$E[X] = \frac{1}{\lambda}; \qquad E[X^2] = \frac{2}{\lambda^2}; \qquad Var(X) = \frac{1}{\lambda^2} \tag{4.4}$$

Consider the x to represent a time period. Starting from any time zero on the timeline, there exists a positive number a on the timeline, and after a duration x has passed the process reaches point $x + a$ on the timeline. This is shown in Figure 1.

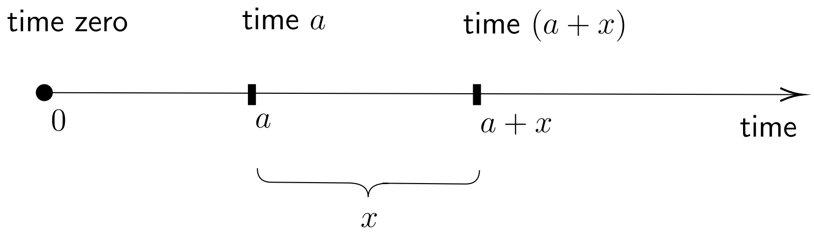

Figure 1. Timeline

We can consider the event that the random variable X (modelling time in this case) has a value $X > x + a$, given that the event $X > a$ has occurred. The memoryless property is stated as,

Memoryless Property
$$P[X > (x + a) \mid (X > a)] = P[X > a] \tag{4.5}$$

Given that the event $X > a$ has already occurred, then the conditional probability of $X > (x + a)$ is same as $P(x)$, that is, the process assumes as if event $X > a$ has not even occurred, like when it started at time zero.

Note that the right-hand side of the expression $P[X > x]$ is same as the event $P[X > x|X > 0]$. And the event that $X > 0$ refers to the entire given the sample space; naturally for $P(E) = P(E|\Omega)$ implies $E(X) = E(X|\Omega)$.

4.2 Proving the memoryless property

Consider the following scenario where one has to model whether a customer has arrived or not, or whether a car engine has failed. Let X be the random variable that denotes the time to failure of a car. The event $E = (X > t)$ is the event that the car engine has not failed by time t. Given that this event has already occurred we ask if there is any change in the probability that the car engine would fail after a further period s. This is shown in the Figure 2.

The required probability is given event $E = (X > t)$, what is the conditional probability of event $X \leq (t + s)$. This is given by, the probability of intersection of the two events over the probability of E.

$$P[(X \leq t + s) \leq (X > t)] = \frac{P[(X \leq t + s) \cap (X > t)]}{P[X > t]} \tag{4.6}$$

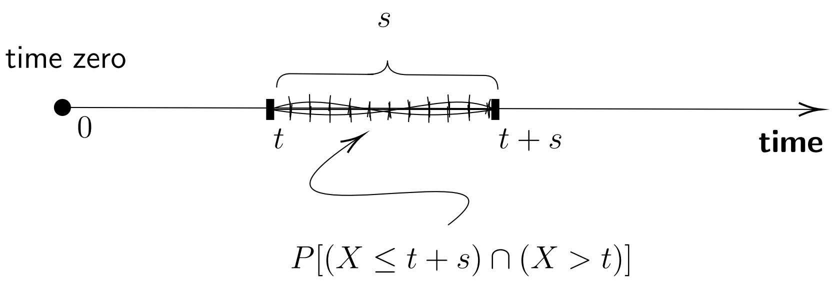

Figure 2. Time till car engine fails modelled by random variable X

Since we know the CDF, we can calculate the tail probability as,

$$F_X(x) = P[X \le x] = (1 - e^{-\lambda x})u(x)$$
$$P[X > x] = 1 - F_X(x) = 1 - (1 - e^{-\lambda x})u(x) = e^{-\lambda x}u(x) \tag{4.7}$$

From figure 2 the probability of the event $P[(X \le t + s) \cap (X > t)]$ is written as,

$$P[(X \le t+s) \le (X > t)] = \frac{P[(X \le t+s) \cap (X > t)]}{P[X > t]}$$
$$P[(X \le t+s) \le (X > t)] = \frac{P[t < X \le (t+s)]}{P[X > t]}$$
$$P[(X \le t+s) \le (X > t)] = \frac{F_x(t+s) - F_X(t)}{1 - -F_X(t)} = \frac{(1 - e^{-\lambda(t+s)}) - (1 - e^{-\lambda t}}{1 - (1 - e^{-\lambda t}))}$$
$$P[(X \le t+s) \le (X > t)] = \frac{e^{-\lambda t} - e^{-\lambda(t+s)}}{e^{-\lambda t}} = 1 - e^{-\lambda s} = F_X(s)$$
$$P[(X \le t+s) \le (X > t)] = P[X \le s] \tag{4.8}$$

Thus, this indicates that the random process only remembers the present, but not the past, hence its name - memoryless property.

4.3 Shifting of PDF leads to memoryless property

Consider the conditional CDF and PDF of the event that $X \le x$ given $X > t$,

$$F_{X|X>t}(x|x > t) = P[(X \le x) \mid (X > t)]$$
$$= 1 - e^{-\lambda s} = 1 - e^{-\lambda(x-t)} \tag{4.9}$$
$$f_{X|X>t}(x|x > t) = \frac{d}{dx}\left[1 - e^{-\lambda(x-t)}\right] = \lambda e^{-\lambda(x-t)} \tag{4.10}$$

Recall the PDF of the original exponential distribution if of the same form. That is,

$$f_X(x) = \lambda e^{-\lambda x}$$
$$f_{X|X>t}(x|x > t) = \lambda e^{-\lambda(x-t)} \tag{4.11}$$

Thus, as depicted in Figure 3, the shifting property of the exponential distribution's PDF causes the memoryless property. We see from the figure, that the PDF is merely shifted by t in the exponent, whereas the distribution suffers no change.

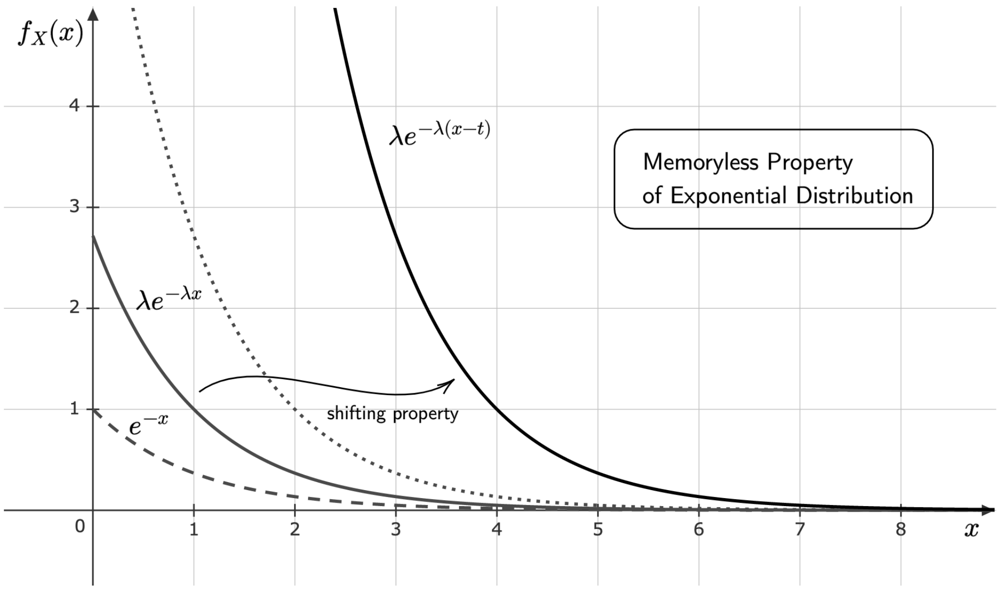

Figure 3. Memoryless Property of the exponential distribution arises from the shifting nature of its PDF.

References

[1] Stochastic Processes Course notes of Prof. Rakesh Nigam.

[2] Ross, Sheldon. A first course in probability. Pearson, 2014.

[3] Karlin. Samuel, and Howard E. Taylor. *A second course in stochastic processes.* Elsevier, 1981.

[4] Karlin, S., and H. M. Taylor. *A first course in stochastic processes, Acad.* Press, New York. 1966.