

# Introduction to Uncertainty Quantification

## Module 2.3: Random Vectors & Random Processes

### 1 Random Vectors

Often, more than one random variables are required to describe uncertainty in a system. These random variables may be dependent on or independent of one another.

Examples

- *Dependent*: Strength, elastic modulus, and Poisson's ratio of a material are often dependent on one another.
- *Independent*: Strength of the material and load on the structure are generally independent of one another.

When multiple random variables are considered together in a system, we assemble them in a *random vector* and refer to them as jointly distributed random variables. Except under special conditions, jointly distributed random variables have probabilities that cannot be separated. That is, for two random variables  $X$  and  $Y$  the probability of  $X$  and  $Y$  depend on one another and therefore cannot be computed separately. These joint probabilities are defined through the *joint cumulative distribution function* and the *joint probability density function*, discussed next.

#### 1.1 Joint CDF and PDF

Consider two jointly distributed random variables  $X$  and  $Y$ .

The **Joint Cumulative Distribution Function** is defined as

$$F_{XY}(x, y) = P(X \leq x \cap Y \leq y) = P(X \leq x, Y \leq y) \quad (1)$$

That is, the joint CDF is the probability that both  $X \leq x$  and  $Y \leq y$ .

The **Joint Probability Density Function** is defined by the derivative of the joint CDF with respect to each variable as

$$f_{XY}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y} \quad (2)$$

Therefore, it follows that the joint CDF can be expressed in terms of the joint PDF as

$$F_{XY}(x, y) = \int_{-\infty}^y \int_{-\infty}^x f_{XY}(u, v) dudv \quad (3)$$

## Properties of Joint CDF and PDF

The joint CDF and PDF have the following properties:

- The joint CDF approaches zero as either variable (or both variables) approach negative infinity. That is:

$$F_{XY}(-\infty, -\infty) = F_{XY}(-\infty, y) = F_{XY}(x, -\infty) = 0$$

This can be interpreted to mean that as the probability of one variable approaches zero, the joint probability of both variables approaches zero.

- The joint CDF only approaches one when both variables approach infinity. That is:

$$F_{XY}(\infty, \infty) = 1$$

This can be interpreted to mean that the joint probability can only be high when both variables have large values. If one variable takes a small value, the joint probability cannot be high even if the other variable takes a large value.

- The joint PDF must integrate to one. That is:

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

As for random variables, this follows from the definition of the joint PDF and its relation to the joint CDF.

Illustrations of the joint PDF and CDF are shown in Figure 1 where we can see the above properties. We can see that the joint CDF approaches zero as either  $x$  or  $y$  take on small values and that it approaches one only when both  $x$  and  $y$  take large positive values.

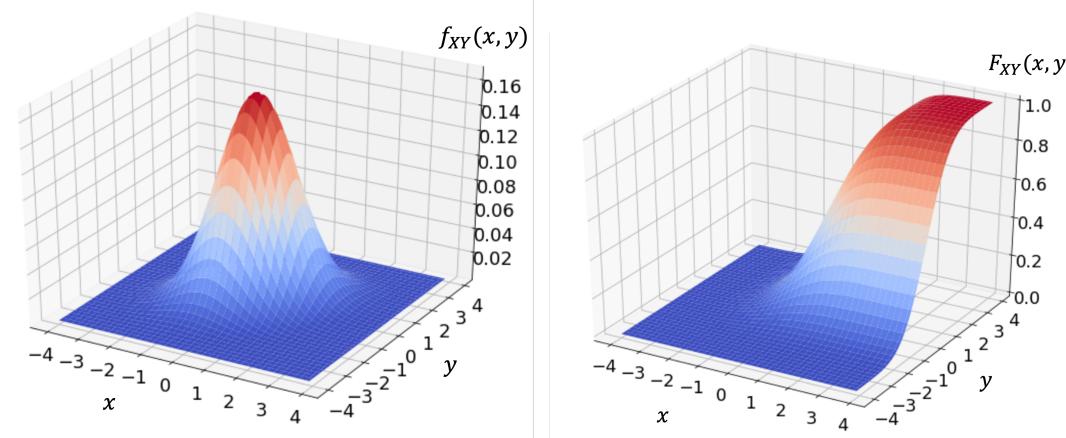


Figure 1: Illustration of the joint PDF (left) and joint CDF (right)

## High Dimensional Random Vectors

Very often we will deal with random vectors that contain a large number of random variables. In general, we will denote random vectors with bold and capital letters where, for an  $n$ -dimensional random vector

we have  $\mathbf{X} = [X_1, X_2, \dots, X_n]^T$ . The definitions of the joint PDF and CDF can be naturally extended to random vectors containing any arbitrary number of jointly distributed random variables,  $n$ , where the *joint CDF* is given by

$$F_{\mathbf{X}}(x_1, x_2, \dots, x_n) = P(X_1 \leq x_1 \cap X_2 \leq x_2 \cap \dots \cap X_n \leq x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) \quad (4)$$

and the *joint PDF* is given by

$$f_{\mathbf{X}}(x_1, x_2, \dots, x_n) = \frac{\partial^n F_{\mathbf{X}}(x_1, x_2, \dots, x_n)}{\partial x_1 \partial x_2 \dots \partial x_n} \quad (5)$$

From the joint PDF, any lower-dimensional density for a subset of components can be obtained by integrating over the excluded components as follows

$$f_{X_1, \dots, X_i, X_{i+k+1}, \dots, X_n}(x_1, \dots, x_i, x_{i+k+1}, \dots, x_n) = \underbrace{\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty}}_{k\text{-fold}} f_{\mathbf{X}}(x_1, x_2, \dots, x_n) dx_{i+1} \dots dx_{i+k} \quad (6)$$

where the components  $X_{i+1}$  to  $X_{i+k}$  have been integrated out.

## 1.2 Marginal Probability Distributions

The **Marginal Probability Density Function** for random variable  $X$  within a random vector  $\mathbf{X}$  is defined as the probability density of  $X$  considering any outcome of the other random variables. It is obtained by integrating the joint PDF over the remaining variables. For two random variables  $X$  and  $Y$ , this is given by the following expression:

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy \quad (7)$$

From the marginal probability density function, we can obtain the marginal cumulative distribution function through integration the standard way. That is:

$$F_X(x) = \int_{-\infty}^x f_X(u) du \quad (8)$$

The definition of the marginal probability density function can be extended to a single random variable  $X_i$  from a random vector of arbitrary dimension,  $n$ , by integrating over the other  $n - 1$  random variables as

$$f_{X_i}(x_i) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_{i-1} dx_{i+1} \dots dx_n \quad (9)$$

From this, the marginal CDF can be obtained from Eq. (8).

The marginal PDFs for two random variables,  $X_1$  and  $X_2$ , are illustrated in Figure 2.

## 1.3 Independent Random Variables

In general, the probabilities of jointly distributed random variables are dependent on one another and cannot be computed separately. Two random variables,  $X$  and  $Y$  are said to be **independent** if the joint probability of  $X$  and  $Y$  can be expressed as the product of the marginal probabilities. That is:

$$F_{XY}(x, y) = P(X \leq x \cap Y \leq y) = P(X \leq x)P(Y \leq y) = F_X(x)F_Y(y) \quad (10)$$

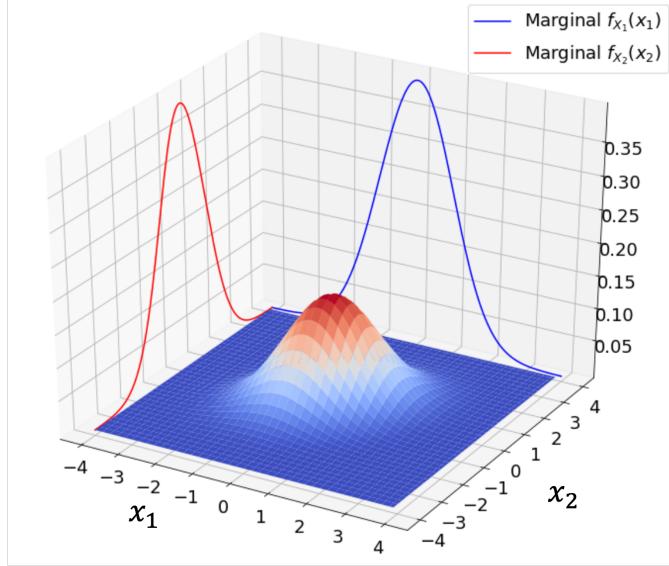


Figure 2: Marginal PDFs for two random variables  $X_1$  and  $X_2$  determined from integrating the joint PDF.

It follows that

$$f_{XY}(x, y) = f_X(x)f_Y(y) \quad (11)$$

In this case, the probabilities of  $X$  and  $Y$  are separable and do not affect one another. As before, this definition can be extended to define the joint CDF and joint PDF for a set of  $n$  independent random variables as:

$$F_{\mathbf{X}}(x_1, x_2, \dots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2) \cdots F_{X_n}(x_n) \quad (12)$$

and

$$f_{\mathbf{X}}(x_1, x_2, \dots, x_n) = f_{X_1}(x_1)f_{X_2}(x_2) \cdots f_{X_n}(x_n) \quad (13)$$

#### 1.4 Conditional Distributions

The **Conditional Probability Density Function** for the random variable  $X$ , conditioned on the outcome of random variable  $Y$  is defined by

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} \quad (14)$$

This follows from definition of conditional probability presented earlier. By integrating, we can obtain the **Conditional Cumulative Distribution Function**

$$F_{X|Y}(x|y) = \frac{\int_{-\infty}^x f_{XY}(z,y)dz}{f_Y(y)} \quad (15)$$

By integrating the conditional CDF weighted by  $f_Y(y)$ , we obtain again the *Law of Total Probability*

$$F_X(x) = \int_{-\infty}^{\infty} F_{X|Y}(x|y)f_Y(y)dy \quad (16)$$

The conditional probability density is illustrated in Figure 3 where we can see that the conditional PDF,  $f_{X|Y}(x|y)$ , corresponds to a slice through the joint probability density function at  $Y = y$  that has been

normalized by the integral of that slice (i.e. the marginal PDF evaluated at  $Y = y$ ,  $f_Y(y)$ ) so that it integrates to one.

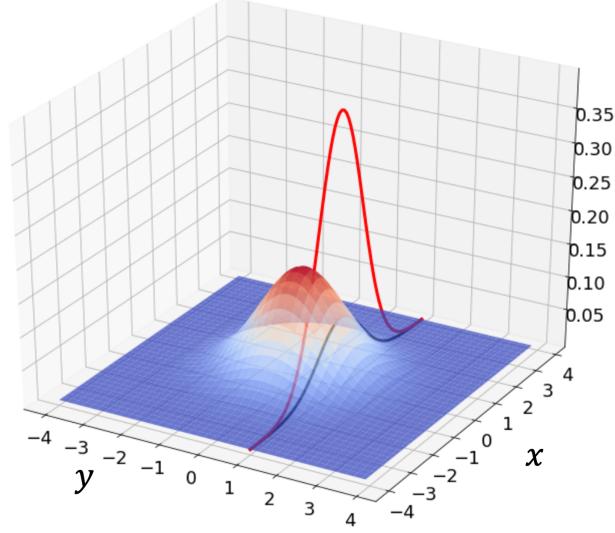


Figure 3: Illustration of the conditional PDF  $f_X(x|y = 1)$  (red curve) showing a slice through the joint PDF corresponding to  $f_{X,Y}(x, y = 1)$ .

Using conditional probabilities, we can say that  $X$  and  $Y$  are *independent* if

$$f_{X,Y}(x|y) = f_X(x) \quad (17)$$

## 2 Moments of Jointly Distributed Random Variables

In the same way we defined moments for random variables, we can likewise define moments for jointly distributed random variables. These moments describe both the individual random variables, similar to what we saw before, and the interaction between random variables. Again, we will define two types of moments.

### 2.1 Moments about the Origin

For two random variables,  $X$  and  $Y$ , we define the *joint moments about the origin* of order  $n + m$  as the expected value of the product  $X^n Y^m$ . This is given by:

$$\mu'_{nm} = \mathbb{E}[X^n Y^m] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^n y^m f_{X,Y}(x, y) dx dy \quad (18)$$

When either  $n$  or  $m$  is equal to zero, we recover the original moments for single random variables. Some important examples include the following

$$\begin{aligned} \mu'_{10} &= \mathbb{E}[X] = \mu_X \\ \mu'_{01} &= \mathbb{E}[Y] = \mu_Y \\ \mu'_{20} &= \mathbb{E}[X^2] \\ \mu'_{02} &= \mathbb{E}[Y^2] \end{aligned} \quad (19)$$

where, again,  $\mu_X, \mu_Y$  are the mean values of  $X$  and  $Y$ , respectively and  $\mathbb{E}[X^2], \mathbb{E}[Y^2]$  are the mean square of  $X$  and  $Y$ .

## 2.2 Central Moments

For two random variables,  $X$  and  $Y$ , we define the *joint central moments* (or moments about the mean) of order  $n + m$  as

$$\mu_{nm} = \mathbb{E}[(X - \mathbb{E}[X])^n(Y - \mathbb{E}[Y])^m] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^n(y - \mathbb{E}[Y])^m f_{XY}(x, y) dx dy \quad (20)$$

Again, when either  $n$  or  $m$  is equal to zero, we recover the original central moments for single random variables. Some important examples include the following

$$\begin{aligned} \mu_{20} &= \text{Var}(X) \\ \mu_{02} &= \text{Var}(Y) \\ \mu_{30} &= \sigma^3 \text{Skew}(X) \\ \mu_{03} &= \sigma^3 \text{Skew}(Y) \\ \mu_{40} &= \sigma^4 \text{Kurt}(X) \\ \mu_{04} &= \sigma^4 \text{Kurt}(Y) \end{aligned} \quad (21)$$

## 2.3 Covariance & Correlation

When neither  $n$  or  $m$  equals zero, the joint moments describe the interactions between random variables. We will only be concerned here with the joint moments up to second order, that is when  $n = m = 1$ . The **Covariance** of two random variables  $X$  and  $Y$  is defined as this second joint central moment given by:

$$\mu_{11} = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \text{Cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mathbb{E}[X])(y - \mathbb{E}[Y]) f_{XY}(x, y) dx dy \quad (22)$$

If  $X$  and  $Y$  are independent, with  $f_{XY}(x, y) = f_X(x)f_Y(y)$ , then

$$\begin{aligned} \text{Cov}(X, Y) &= \sigma_{XY} = \int_{-\infty}^{\infty} (x - \mathbb{E}[X]) f_X(x) dx \int_{-\infty}^{\infty} (y - \mathbb{E}[Y]) f_Y(y) dy \\ &= \left( \int_{-\infty}^{\infty} x f_X(x) dx - \mathbb{E}[X] \int_{-\infty}^{\infty} f_X(x) dx \right) \left( \int_{-\infty}^{\infty} y f_Y(y) dy - \mathbb{E}[Y] \int_{-\infty}^{\infty} f_Y(y) dy \right) \\ &= (\mathbb{E}[X] - \mathbb{E}[X])(\mathbb{E}[Y] - \mathbb{E}[Y]) \\ &= 0 \end{aligned} \quad (23)$$

The normalized covariance is referred to as the **Correlation Coefficient** and is given by

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}. \quad (24)$$

The correlation coefficient has bounds derived as follows. Consider an arbitrary constant,  $\alpha$ . The following conditions must hold:

$$\begin{aligned} \mathbb{E}[\{\alpha(X - \mathbb{E}[X]) + (Y - \mathbb{E}[Y])\}^2] &\geq 0 \\ \alpha^2 \mathbb{E}[(X - \mathbb{E}[X])^2] + 2\alpha \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] + \mathbb{E}[(Y - \mathbb{E}[Y])^2] &\geq 0 \\ \alpha^2 \sigma_X^2 + 2\alpha \text{Cov}(X, Y) + \sigma_Y^2 &\geq 0 \end{aligned} \quad (25)$$

Here, we have a quadratic equation in  $\alpha$  that is always positive. This means that the quadratic has imaginary roots. This implies that the discriminant  $\Delta = b^2 - 4ac \leq 0$  where  $b = 2\text{Cov}(X, Y)$ ,  $a = \sigma_X^2$ , and  $c = \sigma_Y^2$ . This yields the following relation:

$$\begin{aligned}\Delta &\leq 0 \\ \text{Cov}(X, Y)^2 - \sigma_X^2 \sigma_Y^2 &\leq 0\end{aligned}\tag{26}$$

This can only hold when

$$-1 \leq \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \leq 1\tag{27}$$

Hence,  $-1 \leq \rho_{XY} \leq 1$ .

The correlation coefficient is a measure of *linear dependence* between  $X$  and  $Y$ .

- $\rho_{XY} = 1$  implies  $Y = aX + b$ .
- $\rho_{XY} = -1$  implies  $Y = -aX + b$ .
- $\rho_{XY} = 0$  implies that  $X$  and  $Y$  are *linearly* uncorrelated. This does *not* imply that  $X$  and  $Y$  are independent. Although, as we showed,  $\rho_{XY} = 0$  if  $X, Y$  are independent.

### 3 Isoprobabilistic Transformations

In the previous modules, we emphasized the importance of Gaussian random variables. The same importance holds for Gaussian random vectors because they can be particularly convenient to work with (thanks in large part to the preservation of the Gaussian distribution under linear operations). As a result, it is common to employ certain transformations, referred to as **isoprobabilistic transformations** that preserve probabilities while mapping values from one distribution to another (usually from a non-Gaussian distribution to a Gaussian distribution).

Mathematically, this can be stated as follows. Consider the bijective transformation between random vectors  $\mathbf{X}$  and  $\mathbf{Y}$  having joint CDFs  $F_{\mathbf{X}}(\mathbf{x})$  and  $F_{\mathbf{Y}}(\mathbf{y})$  given by  $\mathbf{Y} = T(\mathbf{X})$  having unique inverse  $\mathbf{X} = T^{-1}(\mathbf{Y})$ . The transformation,  $T(\cdot)$  is an isoprobabilistic transformation if we can identify a pair of values  $(\mathbf{x}, \mathbf{y})$  satisfying:

$$\begin{aligned}F_{\mathbf{Y}}(\mathbf{y}) &= P(\mathbf{Y} \leq \mathbf{y}) \\ &= P(\mathbf{Y} \leq T(\mathbf{x})) \\ &= P(T^{-1}(\mathbf{Y}) \leq \mathbf{x}) \\ &= P(\mathbf{X} \leq \mathbf{x}) \\ &= F_{\mathbf{X}}(\mathbf{x})\end{aligned}\tag{28}$$

We are particularly interested in identifying isoprobabilistic transformations that allow us to map from a random vector  $\mathbf{X}$  having arbitrary probability distribution  $F_{\mathbf{X}}(\mathbf{x})$  to a standard normal random vector  $\mathbf{Y}$ .

In this section, we will provide a brief review of Gaussian random vectors (extending what we already learned about Gaussian random variables) and then introduces two isoprobabilistic transformations, the Nataf transformation and the Rosenblatt transformation.

### 3.1 Gaussian Random Vectors

Two random variables  $X$  and  $Y$  are jointly Gaussian with mean values  $\mu_X, \mu_Y$ , standard deviations  $\sigma_X, \sigma_Y$  and correlation coefficient  $\rho_{XY}$  if their joint probability density function takes the form:

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{XY}^2}} \times \exp\left\{-\frac{1}{2(1-\rho_{XY}^2)} \left[ \left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)\rho_{XY} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 \right]\right\} \quad (29)$$

More generally, the random vector  $\mathbf{X} = [X_1, X_2, \dots, X_n]^T$  having mean vector  $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_n]^T$  and symmetric positive definite covariance matrix  $\boldsymbol{\Sigma}$  has  $n$ -dimensional joint probability density function:

$$f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} \det(\boldsymbol{\Sigma})^{-\frac{1}{2}} \exp\left\{\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right\} \quad (30)$$

where the terms of the covariance matrix  $\boldsymbol{\Sigma}$  are given by  $\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$ .

### Properties of Normal Random Vectors

Normal random vectors retain all of the convenient properties of their constituent Gaussian random variables, including the expression of their higher-order moments and their behavior under linear operations. Importantly, a Gaussian random vector is completely defined in terms of its mean vector  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$ . Consequently, a normal random vector can always be standardized such that there exists a standard normal random vector  $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{I})$  where  $\mathbf{0}$  is the zero vector and  $\mathbf{I}$  is the identity matrix such that:

$$\mathbf{X} = \mathbf{A}\mathbf{Z} + \boldsymbol{\mu} \quad (31)$$

where  $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T$  can be performed using either Cholesky decomposition or Eigen decomposition of the covariance matrix.

### 3.2 Nataf Transformation

Consider a random vector  $\mathbf{X} = [X_1, X_2, \dots, X_n]^T$  having arbitrary known *marginal* cumulative distribution functions  $F_{X_i}(x_i)$  (non-Gaussian) with prescribed mean vector  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$  having components  $\Sigma_{ij} = \xi_{ij}\sigma_{ij}$  where  $\xi_{ij}$  are the (non-Gaussian) correlation coefficients. The component-wise transformation given by

$$T(X_i) = \Phi^{-1}[F_{X_i}(X_i)], \quad i = 1, \dots, n \quad (32)$$

is known as the *Nataf Transformation*. Using this transformation for each component  $i$ , the standard normal random variable  $Z_i \sim N(0, 1)$  can be obtained from the following transformation:

$$Z_i = \Phi^{-1}[F_{X_i}(X_i)], \quad i = 1, \dots, n \quad (33)$$

We note that the Nataf Transformation operates component-wise. Given that we prescribe only the non-Gaussian covariance  $\boldsymbol{\Sigma}$  and do not prescribe the full joint probability density function, the Nataf Transformation makes the implicit assumption that, although the marginal distributions of  $X_i$  are not Gaussian, the components of  $X_i$  are jointly Gaussian (specifically having Gaussian copula [1]). That is, their dependence is modeled through a joint probability density function having the following form:

$$f_{X_i X_j}(x_i, x_j) = \phi_2(T(x_i), T(x_j), \rho_{ij}) \frac{f_{X_i}(x_i)f_{X_j}(x_j)}{\phi(T(x_i))\phi(T(x_j))} \quad (34)$$

where  $T(\cdot)$  is given by Eq. (32),  $\phi_2(\cdot)$  is the standard bivariate normal probability density function (Eq. (29) with zero mean and unit standard deviations), and  $\phi(\cdot)$  is the standard normal marginal probability density function.

The Nataf transformation operates by performing a marginal distribution transformation through the cumulative distribution functions. This is illustrated in Figure 4. Because it operates on the marginals

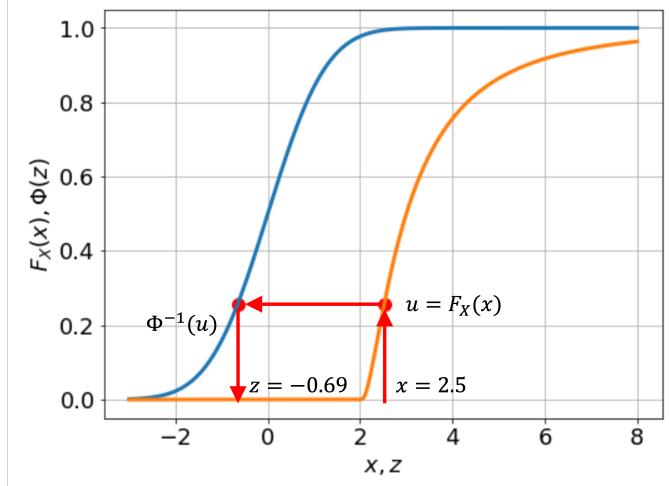


Figure 4: Illustration of the Nataf transformation showing the transformation from at a point  $x = 2.5$  for a point from a lognormal distribution (orange curve) to a corresponding value of  $z = -0.69$  for the standard normal.

independently under the assumption of Gaussian dependence, using the definition of the covariance we can express the non-Gaussian correlation coefficients as follows:

$$\begin{aligned}\xi_{ij} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left( \frac{x_i - \mu_i}{\sigma_i} \right) \left( \frac{x_j - \mu_j}{\sigma_j} \right) f_{X_i X_j}(x_i, x_j) dx_i dx_j \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left( \frac{x_i - \mu_i}{\sigma_i} \right) \left( \frac{x_j - \mu_j}{\sigma_j} \right) \phi_2(T(x_i), T(x_j), \rho_{ij}) \frac{f_{X_i}(x_i) f_{X_j}(x_j)}{\phi(T(x_i)) \phi(T(x_j))} dx_i dx_j\end{aligned}\quad (35)$$

By definition,  $f_{X_i}(x_i) = \frac{d}{dx_i}[F_{X_i}(x_i)]$ . Consider that  $x_i$  can be expressed through the *Inverse Nataf Transformation* by:

$$x_i = F_{X_i}^{-1}[\Phi(z_i)] \quad (36)$$

then

$$\begin{aligned}f_{X_i}(x_i) &= \frac{d}{dx_i} F_{X_i} \left[ F_{X_i}^{-1} [\Phi(z_i)] \right] \\ &= \frac{d}{dx_i} \Phi(z_i) \\ &= \frac{d\Phi(z_i)}{dz_i} \frac{dz_i}{dx_i} \\ &= \phi(z_i) \frac{dz_i}{dx_i}\end{aligned}\quad (37)$$

Repeating this for variable  $x_j$ , noting that  $x_i = T^{-1}(z_i)$  and plugging the result into Eq. (35) yields the following

$$\begin{aligned}\xi_{ij} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left( \frac{T^{-1}(z_i) - \mu_i}{\sigma_i} \right) \left( \frac{T^{-1}(z_j) - \mu_j}{\sigma_j} \right) \phi_2(z_i, z_j, \rho_{ij}) dz_i dz_j \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left( \frac{F_{X_i}^{-1}(\Phi(z_i)) - \mu_i}{\sigma_i} \right) \left( \frac{F_{X_j}^{-1}(\Phi(z_j)) - \mu_j}{\sigma_j} \right) \phi_2(z_i, z_j, \rho_{ij}) dz_i dz_j\end{aligned}\tag{38}$$

Eq. (38) expresses what is known as the *correlation distortion*. The correlation distortion expresses the change in correlation that occurs as we transform a pair of Gaussian random variables having correlation  $\rho_{ij}$  from the Gaussian to arbitrary non-Gaussian distributions  $F_{X_i}(x_i)$  and  $F_{X_j}(x_j)$  using the inverse Nataf Transformation with resulting correlation  $\xi_{ij} \neq \rho_{ij}$ . The correlation distortion has the following properties:

- $|\xi_{ij}| \leq |\rho_{ij}|$ . This means that applying the inverse Nataf transformation can never increase the correlation, and in most cases it causes a loss of correlation.
- When  $\rho_{ij} = 1$ ,  $\xi_{ij} = 1$ . This means that perfect positive correlation between Gaussian variables results in perfect positive correlation between non-Gaussian variables.
- When  $\rho_{ij} = 0$ ,  $\xi_{ij} = 0$ . This means that if the Gaussian variables are uncorrelated, the non-Gaussian variables will be uncorrelated.
- When  $\rho_{ij} = -1$ ,  $0 \leq \xi_{ij} \leq 1$ . This means that the perfectly negatively correlated random variables will not necessarily result in perfectly negatively correlated non-Gaussian random variables.
- In general, the correlation distortion in Eq. (38) is not invertible. That is, given a specified non-Gaussian correlation  $\xi_{ij}$  and arbitrary non-Gaussian marginal CDFs  $F_{X_i}(x_i)$  and  $F_{X_j}(x_j)$ , it is not always possible to identify a corresponding value of  $\rho_{ij}$  that, when mapped through Eq. (38) will yield the correct value of  $\xi_{ij}$ . Cases of  $\xi_{ij}$  and  $F_{X_i}(x_i)$  where inversion is not possible are referred to as incompatible with the Nataf transformation.

An illustrative example of the correlation distortion is shown in Figure 5, which shows the correlations distortion for the transformation from Normal random variables to a certain set of Lognormal random variables by the blue curve. Shown for reference (orange) is the curve  $\rho = \xi$  corresponding to no correlation distortion. From this graph, we can see several of the properties of the correlation distortion. Specifically, we can see that  $|\xi| \leq |\rho|$  across the entire domain and that when  $\rho = 0$ , we have  $\xi = 1$  and when  $\rho = 1$ , we have  $\xi = 1$ . We further see the Nataf model incompatibility arise in the regime of negative correlation. Note that  $\min(\xi) \approx -0.35$ . Hence, for this distribution the Nataf model cannot produce correlated random variables with  $\xi < -0.35$ .

### 3.3 Rosenblatt Transformation

Consider now a random vector  $\mathbf{X} = [X_1, X_2, \dots, X_n]^T$  having known *complete* joint cumulative distribution function  $F_{\mathbf{X}}(\mathbf{x})$ . Using the tools we developed in Section 1, we can compute the marginal distributions, lower-dimensional joint distributions, and the conditional distributions. This represents the complete probability information for the random vector, and it is typically only known in special cases. Nonetheless, if this information is given we can define the standard normal random vector  $\mathbf{V}$  component-wise through the following sequence of recursive transformations

$$V_i = \Phi^{-1} [F_{X_i|X_{i-1} \dots X_1}(X_i|X_{i-1} \dots X_1)]\tag{39}$$

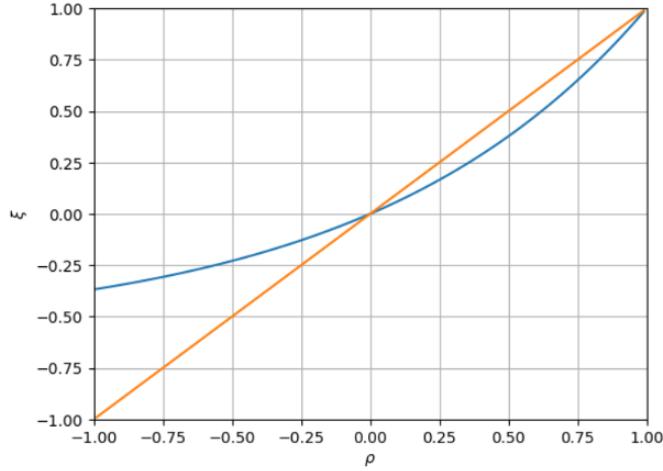


Figure 5: Illustration correlation distortion. Shown here is the correlation distortion from a Lognormal distribution having shape parameter equal to 1 and scale parameter equal to 1 to the standard normal distribution (blue curve). For reference, the case of no correlation distortion  $\rho = \xi$  is shown in orange.

where  $F_{X_i|X_{i-1}\dots X_1}(X_i|X_{i-1}\dots X_1)$  is the conditional cumulative distribution function for component  $X_i$  given all preceding components  $X_1, X_2, \dots, X_{i-1}$ . In the case of component  $X_1$ , the conditional distribution reduces the marginal distribution, i.e.  $V_1 = \Phi^{-1}[F_{X_1}(X_1)]$  and is thus given by the Nataf transformation for that component.

## 4 Random Processes & Random Fields

Random processes and random fields are the next object in the progression of random quantities in increasing dimension. We will present two alternative, but equivalent definitions of a random process (random field) that appeal to different interpretations.

**Definition 1:** A random process  $\{X(t), t \in T\}$  is a parameterized family of random variables with the parameter belonging to the indexed set  $T$ .

According to the first definition, we interpret the random process as being composed of a *set of random variables* that are indexed on a set  $T$  by a variable  $t$ . When  $T$  is a countable set, the process is called a *discrete random process*. When  $T$  is an interval of the real line, the process is referred to as a *continuous random process*. As we'll see, these random variables have a joint probability structure that depends on  $t$ .

**Definition 2:** A random process  $X(t)$  is an ensemble of functions that can be characterized probabilistically.

According to the second definition, the elements of the random process are considered to be functions residing in a function space.

The main difference between the two definitions lies in the interpretation of the elements of the random process. In the first definition, the random process is thought to be an ensemble of random variables. That is,  $X(t_1), X(t_2), \dots, X(t_n)$  are random variables and the random process is the collection of these random

variables. In the second definition the random process is thought to be composed of a functions that vary probabilistically. That is, let  $X_i(t)$  be a realization, or *sample functions*, of  $X(t)$ . The collection of all possible  $X_i(t)$  together constitutes the random process.

Next, we provide a few notes on terminology. Mathematically, *random processes* and *random fields* are identical. The difference is that they are indexed on different sets. Random processes are typically indexed on *time* (denoted  $T$ ), while random fields are indexed on Cartesian *spatial coordinates* (different notations may be used). We will not make the distinction in defining them and will only make the distinction when a particular application calls for it. We will then use the term random process in most cases, with the understanding that this is synonymous with random field. Furthermore, the terms *random process*, *stochastic process*, and *random function* can be used interchangeably. To the extent possible, we will remain consistent in terminology and use the term random process.

## 4.1 PDF & CDF

The probability structure of a random process  $X(t)$  is defined through its complete joint probability density function. In order of completeness, we can define the following series of probability densities:

### Marginal Density

The marginal probability density at a given time instant  $t$  is given by:

$$f_X(x, t) \quad (40)$$

### Two-Point Joint Density

The joint probability density between any two points in time  $X(t_i)$  and  $X(t_j)$  is given by:

$$f_X(x_i, t_i; x_j, t_j) \quad (41)$$

### Three-Point Joint Density

The joint probability density between any three points in time  $X(t_i)$ ,  $X(t_j)$ , and  $X(t_k)$  is given by:

$$f_X(x_i, t_i; x_j, t_j; x_k, t_k) \quad (42)$$

### General $n$ -Point Joint Density

The joint probability density between any  $n$  points in time  $X(t_i)$ ,  $X(t_j)$ , ...,  $X(t_n)$ , is given by:

$$f_X(x_i, t_i; x_j, t_j; \dots; x_n, t_n) \quad (43)$$

### Cumulative Distribution Functions

As is standard, the cumulative distribution functions can be obtained through integration.

## 5 Moments of Random Processes

In the same way we defined moments for random variables and random vectors, we can likewise define moments for random processes. These moments describe the individual random variables in the random process as well as their interactions through time. Again, we will define two types of moments.

## 5.1 Moments about the Origin

Here, rather than defining moments in the general sense, we will define specific moments of interest.

### Mean Value

The mean value, or mean function, of a random process is given by:

$$\mu_X(t) = \mathbb{E}[X(t)] = \int_{-\infty}^{\infty} xf_X(x, t)dx \quad (44)$$

Notice that the mean value, in general, depends on time  $t$  and is not constant for the process. That is  $\mu_X(t_1) \neq \mu_X(t_2)$  if  $t_1 \neq t_2$ .

### Mean Square

The mean square of a random process is given by:

$$\mathbb{E}[X^2(t)] = \int_{-\infty}^{\infty} x^2 f_X(x, t)dx \quad (45)$$

Notice that the mean square also depends on time  $t$  and is not constant.

### Autocorrelation Function

The autocorrelation function (sometimes simply called the correlation function) is given by:

$$R_{XX}(t_i, t_j) = \mathbb{E}[X(t_i)X(t_j)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_i x_j f_X(x_i, t_i; x_j, t_j) dx_i dx_j \quad (46)$$

Notice that the autocorrelation function, in general, is a function of *both* time instants  $t_i$  and  $t_j$ . Furthermore, we can see that  $R_{XX}(t, t) = \mathbb{E}[X^2(t)]$ .

### 3-Point Correlation Function

The 3-point correlation function is given by:

$$R_{XX}(t_i, t_j, t_k) = \mathbb{E}[X(t_i)X(t_j)X(t_k)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_i x_j x_k f_X(x_i, t_i; x_j, t_j; x_k, t_k) dx_i dx_j dx_k \quad (47)$$

Notice that the 3-point correlation function is a function of *three* time instants  $t_i$ ,  $t_j$ , and  $t_k$ . Using the general  $n$ -dimensional density for the random process, this concept can be extended to define any general  $n$ -point correlation function.

### Properties of Autocorrelation Functions

In general, autocorrelation functions have the following properties:

- *Symmetry*: The autocorrelation function is a symmetric function having:

$$R_{XX}(t_i, t_j) = R_{XX}(t_j, t_i) \quad (48)$$

- *Non-negative Definite*: The autocorrelation function is a non-negative definite function (also called positive semi-definite) such that for any arbitrary function  $h(t)$ :

$$\sum_{j=1}^n \sum_{k=1}^n R_{XX}(t_j, t_k) h(t_j) h(t_k) \geq 0. \quad (49)$$

Similar properties can be derived for higher-order correlation functions, although these will not be discussed here.

## 5.2 Central Moments

Likewise, we will similarly define specific central moments of interest here.

### Covariance Function

The covariance function is given by:

$$C_{XX}(t_i, t_j) = \mathbb{E}[(X(t_i) - \mu_X(t_i))(X(t_j) - \mu_X(t_j))] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_i - \mu_X(t_i))(x_j - \mu_X(t_j)) f_X(x_i, t_i; x_j, t_j) dx_i dx_j \quad (50)$$

Notice that the covariance function, in general, is a function of *both* time instants  $t_i$  and  $t_j$ . It follows from Eqs. (50) and (46) that the covariance can be expressed:

$$C_{XX}(t_i, t_j) = R_{XX}(t_i, t_j) - \mu_X(t_i)\mu_X(t_j) \quad (51)$$

From this we can see that the covariance and correlation functions are identical when the means are zero. Furthermore, we can see that the **variance** of the process is given by:

$$C_{XX}(t, t) = \mathbb{E}[(X(t) - \mu_X(t))^2] = \int_{-\infty}^{\infty} (x - \mu_X(t))^2 f_X(x, t) dx = \sigma_X^2(t) \quad (52)$$

and can be expressed as:

$$\sigma_X^2(t) = \mathbb{E}[X^2(t)] - \mu_X^2(t) \quad (53)$$

From the above expressions, we define the *correlation coefficient* (or normalized correlation function) as:

$$\rho_{XX}(t_i, t_j) = \frac{C_{XX}(t_i, t_j)}{\sigma_X(t_i)\sigma_X(t_j)} \quad (54)$$

## 6 Stationary Random Processes

A random process is said to be *strongly stationary* if its complete probability structure is independent of a shift in the parametric origin. That is, the complete joint probability density function is unchanged by a shift in time as follows:

$$f_X(x_1, t_1; x_2, t_2; \dots; x_n, t_n) = f_X(x_1, t_1 + \alpha; x_2, t_2 + \alpha; \dots; x_n, t_n + \alpha) \quad (55)$$

for any arbitrary shift,  $\alpha$ .

A random process is said to be *weakly stationary* if the first marginal probability density and the two-point joint probability density are independent of a shift in the parametric origin. That is,

$$f_X(x, t) = f_X(x, t + \alpha) \quad (56)$$

and

$$f_X(x_i, t_i; x_j, t_j) = f_X(x_i, t_i + \alpha; x_j, t_j + \alpha) \quad (57)$$

Again, these relations hold for arbitrary  $\alpha$ . Therefore, we can show two important properties of stationary random processes.

- Let  $\alpha = -t$  in Eq. (56) then

$$f_X(x, t) = f_X(x, 0) \rightarrow \mathbb{E}[X(t)] = \text{constant in time} \quad (58)$$

That is, for stationary random processes the mean value is constant in time.

- Let  $\alpha = -t_1$  in Eq. (57) then

$$f_X(x_i, t_i; x_j, t_j) = f_X(x_i, 0; x_j, t_j - t_i) \rightarrow R_{XX}(t_i, t_j) = R_{XX}(t_j - t_i) = R_{XX}(\tau) \quad (59)$$

where  $\tau = t_j - t_i$  is the *time lag*. This means that the second-order density and the autocorrelation function do not depend on both  $t_i$  and  $t_j$ , but depend only on their difference  $t_j - t_i$ . Therefore, the correlation depends only on how far apart two points are in time, and not on where those two points arise in the process.

## 6.1 Stationary Correlation Functions

For stationary random processes, we have demonstrated that the correlation function depends only on a single variable, the time lag  $\tau$  separating the points. Consequently, stationary autocorrelation functions have the following properties:

- *Symmetry*: The stationary autocorrelation function is a symmetric function having:

$$R_{XX}(\tau) = R_{XX}(-\tau) \quad (60)$$

- *Non-negative Definite*: The stationary autocorrelation function is a non-negative definite function (also called positive semi-definite) such that for any arbitrary function  $h(t)$ :

$$\sum_{j=1}^n \sum_{k=1}^n R_{XX}(t_j - t_k) h(t_j) h(t_k) \geq 0. \quad (61)$$

- *Inequality*: The following inequality holds for all stationary autocorrelation functions:

$$|R_{XX}(\tau)| \leq R_{XX}(0) \quad (62)$$

This means that the autocorrelation function for a stationary random process is maximized at zero lag,  $\tau = 0$ .

- *Limiting Behavior as  $\tau \rightarrow 0$* : As the time lag grows very small, the stationary autocorrelation asymptotically approaches the mean square. That is:

$$\lim_{\tau \rightarrow 0} R_{XX}(\tau) = \mathbb{E}[X^2(t)] \quad (63)$$

- *Limiting Behavior as  $\tau \rightarrow \infty$* : As the time lag grows very large, the stationary autocorrelation asymptotically approaches zero. That is:

$$\lim_{\tau \rightarrow \infty} R_{XX}(\tau) = 0 \quad (64)$$

Note that this asymptotic property does not hold for *periodic* random processes.

### Examples of Stationary Autocorrelation Functions:

Figures 6 and 7 at the end of this document shows several examples of stationary autocorrelation functions. These correlation functions are given by the following expressions.

#### Constant Correlation

$$R_{XX}(\tau) = c^2 \quad (65)$$

where  $c$  is a constant value.

#### White Noise

$$R_{XX}(\tau) = a\delta(\tau) \quad (66)$$

where  $a$  is a constant and  $\delta(\tau)$  is the Dirac delta function.

#### Low-Pass White Noise

$$R_{XX}(\tau) = aB \left( \frac{\sin 2\pi B\tau}{2\pi B\tau} \right) \quad (67)$$

where  $a$  and  $B$  are constants.

#### Band Pass White Noise

$$R_{XX}(\tau) = aB \left( \frac{\sin \pi B\tau}{\pi B\tau} \right) \cos 2\pi f_0 \tau \quad (68)$$

where  $a, B$  are constants and  $f_0$  is the central frequency of the noise.

#### Exponential

$$R_{XX}(\tau) = e^{-a|\tau|} \quad (69)$$

where  $a$  is a constant.

#### Exponential Cosine

$$R_{XX}(\tau) = e^{-a|\tau|} \cos(2\pi f_0 \tau) \quad (70)$$

where  $a$  is a constant and  $f_0$  is a central frequency.

## 6.2 Ergodic Random Processes

For random process theory to be practically useful, we must be able to estimate the properties of a random process from data. For most random processes, we cannot know the full joint probability density function. Instead, we typically aim to estimate its first and second moments, mean and autocorrelation function, to characterize the process. The exception, as we'll see, is the Gaussian random process which is completely defined by the mean and autocorrelation function. These functions often must be estimated from realizations of the random process. We can do this in two ways:

### Ensemble Averaging

Consider that we have  $n$  sample functions of the random process  $X(t)$ . The *ensemble mean value* is estimated by:

$$\mathbb{E}[X(t_i)] = \frac{1}{n} \sum_{j=1}^n X_j(t_i), \quad \forall t_i \quad (71)$$

Likewise, the *ensemble autocorrelation function* is estimated by:

$$\mathbb{E}[X(t_i)X(t_j)] = \frac{1}{n} \sum_{k=1}^n X_k(t_i)X_k(t_j), \quad \text{for all combinations of } t_i, t_j \quad (72)$$

### Temporal Averaging

Ensemble averaging requires a large set of  $n$  realizations. For many datasets, only a small number of sample functions are available (or perhaps even only a single sample function is available). In these cases, we may resort to estimating the mean and autocorrelation functions through temporal averaging. Consider that we have a single sample function of the random process  $X(t)$  having duration  $T$ , the *temporal average mean value* is estimated as:

$$\langle X(t) \rangle = \frac{1}{T} \int_0^T X(t) dt \quad (73)$$

Likewise, the *temporal average autocorrelation function* is estimated as:

$$\langle X(t)X(t+\tau) \rangle = \frac{1}{T-\tau} \int_0^{T-\tau} X(t)X(t+\tau) dt \quad (74)$$

### Ergodic Random Processes

In general, an *ergodic random process* is a *stationary* random process whose ensemble and temporal properties are equivalent.

A random process,  $X(t)$  is said to be **ergodic in the mean** when:

$$\mathbb{E}[X(t)] = \langle X(t) \rangle, \quad \text{as } T \rightarrow \infty \quad (75)$$

That is, the ensemble mean value is equal to the temporal mean value. The random process can be determined to be ergodic in the mean if the following necessary and sufficient conditions hold:

- $\mathbb{E}[X(t)]$  is constant
- $\mathbb{E}[X(t)X(t+\tau)]$  depends only on  $\tau$
- $\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T R_{XX}(\tau) d\tau = 0$

The first two conditions are the conditions for the process to be stationary. The third condition is unique to ergodic processes.

A random process  $X(t)$  is said to be **ergodic in correlation** when:

$$\mathbb{E}[X(t)X(t+\tau)] = \langle X(t)X(t+\tau) \rangle, \quad \text{as } T \rightarrow \infty \quad (76)$$

That is, the ensemble autocorrelation is equal to the temporal autocorrelation. The random process can be determined to be ergodic in the mean if the following necessary and sufficient conditions hold:

- $\mathbb{E}[X(t)X(t+\tau)]$  depends only on  $\tau$
- The quantity  $Q_{XX}(\tau, u) = \mathbb{E}[(X(t)X(t+\tau) - R_{XX}(\tau))(X(t)X(t+\tau+u) - R_{XX}(\tau+u))]$  is independent of  $t$ .
- $\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T Q_{XX}(\tau, u) du = 0$

### 6.3 Power Spectral Density Function

Random processes are often characterized in terms of their mean and autocorrelation functions. An alternative means of characterization uses the mean value and the *Power Spectral Density Function*, which is an alternate means of representing the second moment function of the random process. The Power Spectral Density Function of a stationary random process  $X(t)$  is defined as the Fourier transform of the autocorrelation function given by:

$$S_{XX}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R_{XX}(\tau) e^{-i\omega\tau} d\tau \quad (77)$$

Inverting this expression, we can determine the autocorrelation function from the power spectral density function by:

$$R_{XX}(\tau) = \int_{-\infty}^{\infty} S_{XX}(\omega) e^{i\omega\tau} d\omega \quad (78)$$

The relations in Eqs. (77) and (78) are known as the *Wiener-Khintchine Relations*.

#### Properties of the Power Spectral Density Function:

The Power Spectral Density Function is defined only for stationary random processes. Furthermore, it has the following properties.

- $S_{XX}(\omega)$  is a real and even function of frequency  $\omega$ . That is:

$$S_{XX}(\omega) = S_{XX}(-\omega) \quad (79)$$

As a result, when we exploit the Euler equation  $e^{ix} = \cos(x) + i\sin(x)$ , the Wiener-Khintchine relations can be expressed as:

$$S_{XX}(\omega) = \frac{1}{\pi} \int_0^{\infty} R_{XX}(\tau) \cos(\omega\tau) d\tau \quad (80)$$

and

$$R_{XX}(\omega) = 2 \int_0^{\infty} S_{XX}(\omega) \cos(\omega\tau) d\omega \quad (81)$$

Furthermore, the evenness of the Power Spectral Density allows us to define a one-sided Power Spectral Density Function as:

$$G_{XX}(\omega) = 2S_{XX}(\omega) \quad \text{for } 0 < \omega < \infty \quad (82)$$

- $S_{XX}(\omega)$  is strictly non-negative, i.e.

$$S_{XX}(\omega) \geq 0 \quad \forall \omega \quad (83)$$

This property is equivalent to the property that the autocorrelation function is non-negative definite. This result is known as Bochner's Theorem.

- The integral of the Power Spectral Density Function is equal to the mean square of the process. That is,

$$R_{XX}(0) = \mathbb{E}[X^2(t)] = \int_{-\infty}^{\infty} S_{XX}(\omega) \quad (84)$$

This can be interpreted to mean that  $S_{XX}(\omega)$  describes the distribution of the mean square (variance for a zero mean process) over the frequency domain.

- If the process has finite mean square, i.e.  $\mathbb{E}[X^2(t)] < \infty$ , then  $S_{XX}(\omega)$  goes to zero as  $\omega \rightarrow \pm\infty$  faster than  $\frac{1}{\omega}$ .

### Examples of Power Spectral Density Functions:

Figures 6 and 7 at the end of this document shows several examples of one-sided power spectral density functions  $G(f)$  defined for  $f \geq 0$  where frequency  $f = \frac{\omega}{2\pi}$ . These power spectral density functions are given by the following expressions.

#### Constant Correlation

$$G_{XX}(f) = c^2\delta(f) \quad (85)$$

where  $c$  is a constant and  $\delta(f)$  is the Dirac delta function.

#### White Noise

$$G_{XX}(f) = 2a \quad (86)$$

where  $a$  is a constant.

#### Low-Pass White Noise

$$G_{XX}(f) = a; \quad 0 \leq f \leq B; \quad \text{otherwise zero} \quad (87)$$

where  $a$  and  $B$  are constants.

#### Band Pass White Noise

$$G_{XX}(f) = a; \quad f_0 - \frac{B}{2} \leq f \leq f_0 + \frac{B}{2}; \quad \text{otherwise zero} \quad (88)$$

where  $a, B$  are constants and  $f_0$  is the central frequency of the noise.

#### Exponential

$$G_{XX}(f) = \frac{4a}{a^2 + 4\pi^2 f^2} \quad (89)$$

where  $a$  is a constant.

#### Exponential Cosine

$$G_{XX}(f) = 2a \left[ \frac{1}{a^2 + 4\pi^2(f + f_0)^2} + \frac{1}{a^2 + 4\pi^2(f - f_0)^2} \right] \quad (90)$$

where  $a$  is a constant and  $f_0$  is a central frequency.

## 6.4 Gaussian Random Processes

A random process,  $X(t)$  is a Gaussian random process if the parameterized family of random variables that comprises the process follow the joint Gaussian distribution. A Gaussian random process is completely defined by its mean value  $\mu_X(t)$  and autocorrelation function  $R_{XX}(t_i, t_j)$ . For the special case of stationary Gaussian random processes, the process can be completely defined by it's constant mean value  $\mu_X$  and autocorrelation function  $R_{XX}(\tau)$ . As a result of these two properties, if a Gaussian process is weakly stationary, then it is also strongly stationary. In fact, all higher-order moments of a Gaussian random process  $X(t)$  can be obtained from the mean and correlation.

**Linearity Properties** As we saw for Gaussian random variables and Gaussian random vectors, Gaussian random processes remain Gaussian under linear operations. For example, if  $X(t)$  is a Gaussian random process then,

- $Z(t) = aX(t) + b$  is also Gaussian (for  $a$  and  $b$  being deterministic constants).
- $\frac{dX(t)}{dt}, \frac{d^2X(t)}{dt^2}, \dots, \frac{d^nX(t)}{dt^n}$  are also Gaussian random processes (if they exist).
- $Z(t^*) = \int_a^b X(t)h(t, t^*)dt$  is also Gaussian for any bounded deterministic function  $h(t, t^*)$ .

## 7 Markov Chains / Markov Processes

Consider a stochastic process  $\{X_t, t = 0, 1, 2, \dots\}$  that takes on a countable number of possible values defined as the state space,  $S$ . If  $X_t = i$ , the process is said to be in state  $i$  at time  $t$ . Given the complete history of the process up to time  $n$ , we can define the conditional probability of the process moving to state  $j$  at time  $t+1$  as:

$$P(X_{t+1} = j | X_t = i, X_{t-1} = i_{t-1}, \dots, X_0 = i_0) \quad (91)$$

for all states  $i_0, i_1, \dots, i_{t-1}, i$  and for all  $t \geq 0$ . The stochastic process  $\{X_t\}$  has the *Markov property* or is said to be *Markovian* if we can define a fixed probability  $P_{ij}$ , referred to as the *transition probability*, such that

$$P_{ij} = P(X_{t+1} = j | X_t = i, X_{t-1} = i_{t-1}, \dots, X_0 = i_0) = P(X_{t+1} | X_t = i) \quad (92)$$

for all states  $i_0, i_1, \dots, i_{t-1}, i$  and for all  $t \geq 0$ . That is, whenever the process is in state  $i$ , the conditional probability of transition to state  $j$  in the next step depends only on the current state and is independent of all previous states. In other words, the process is said to be *memoryless*. Any process satisfying the Markov property in Eq. (92) is called a *Markov chain* or a *Markov process*. The transition probability,  $P_{ij}$  has the following properties

$$\begin{aligned} P_{ij} &\geq 0 \quad \forall i, j \geq 0 \\ \sum_{j=0}^{\infty} P_{ij} &= 1, \quad \forall i \end{aligned} \quad (93)$$

and is often expressed by the matrix

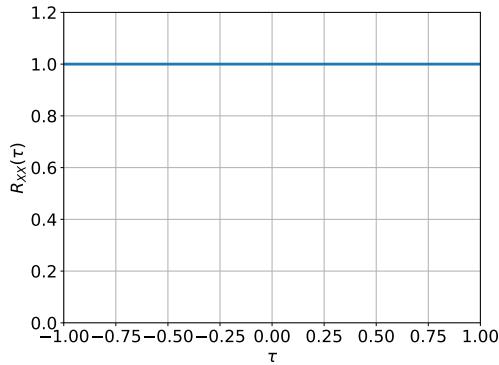
$$\mathbf{P} = \begin{bmatrix} P_{00} & P_{01} & P_{02} & \cdots \\ P_{10} & P_{11} & P_{12} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (94)$$

where the second property in Eq. (93) implies that the sum of the terms in each row must be equal to one.

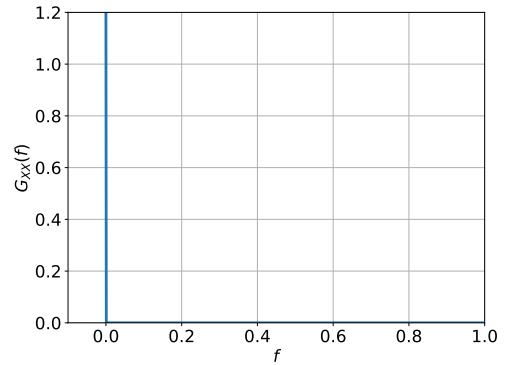
We note that the notation and conventions used here denote a *discrete-time* Markov chain operating on a countable (i.e. discrete) state space,  $S$ . We will maintain these conventions with the recognition that Markov process theory can be generalized to both continuous-time processes and continuous measureable state spaces. In the case of continuous-time Markov chains we adopt a continuous time indexing on the transition probabilities as

$$P_{ij} = P(X(t+s) = j | X(s) = i) \quad (95)$$

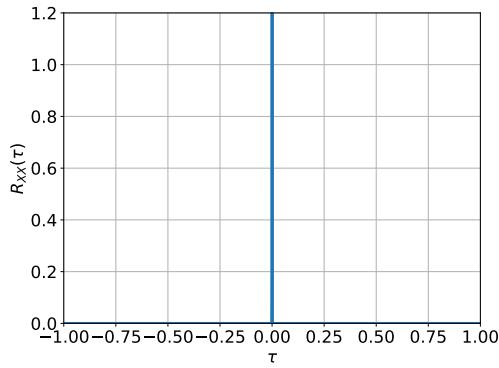
where we retain a countable state space. The more interesting case, for our purposes, is the case where the state space is continuous, in which case the transition probability is no longer defined as a probability mass for state  $i$  given we are currently in state  $j$ . Instead, the discrete transition probability generalizes as a *Markov kernel*, which in simplistic terms can be viewed as a probability density on the continuous state space conditioned on the current state.



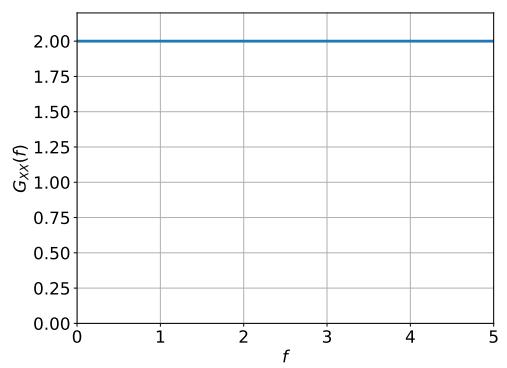
(a) Constant Correlation  
 $c = 1$



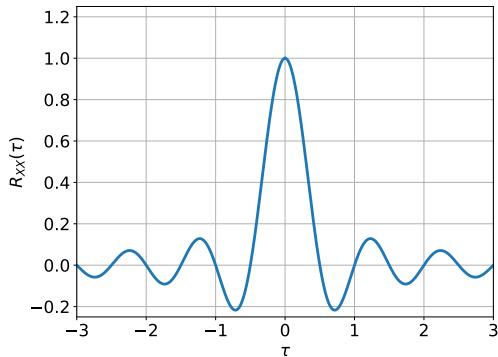
(b) Constant Correlation Power Spectrum  
 $c = 1$



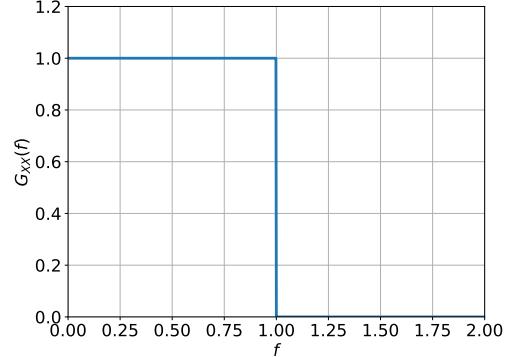
(c) White Noise Correlation  
 $a = 1$



(d) White Noise Power Spectrum  
 $a = 1$

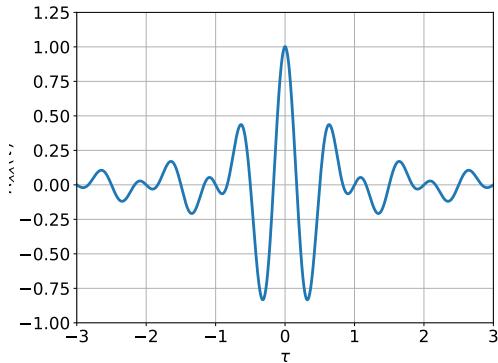


(e) Low-Pass White Noise Correlation  
 $a = 1, B = 1$

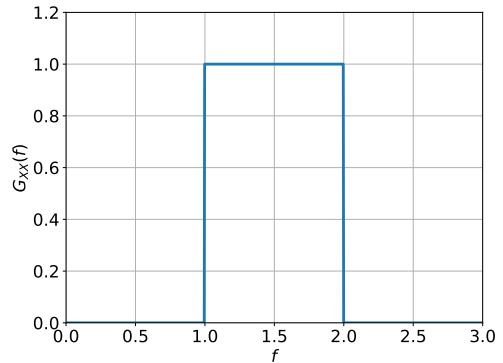


(f) Low-Pass White Noise Power Spectrum  
 $a = 1, B = 1$

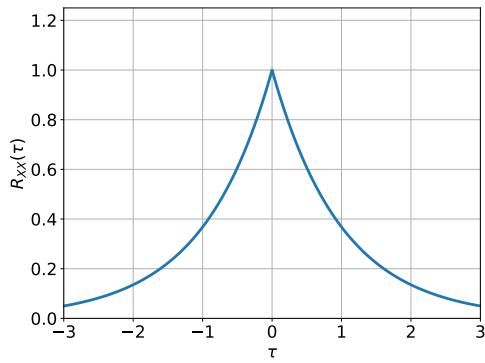
Figure 6: Left (a, c, e): Some common stationary correlation functions. Right (b, d, f): Corresponding One-Sided Power Spectral Density Functions.



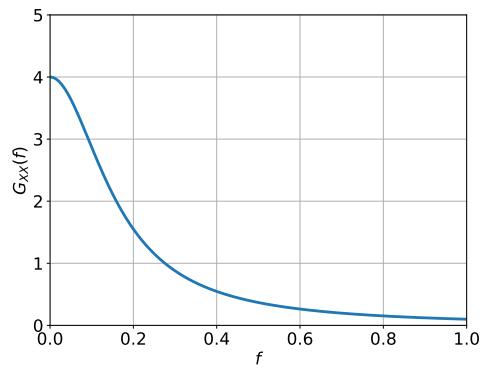
(a) Bandpass White Noise Correlation  
 $a = 1, B = 1, f_0 = 1.5$



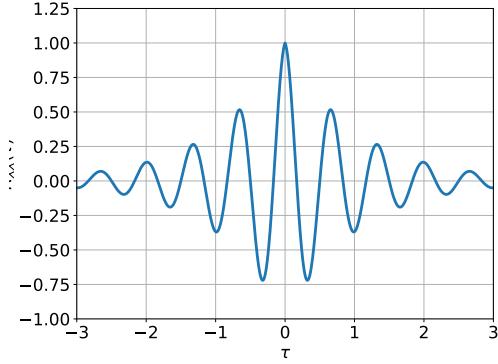
(b) Bandpass White Noise Power Spectrum  
 $a = 1, B = 1, f_0 = 1.5$



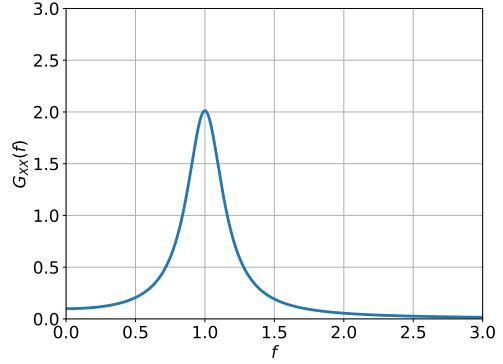
(c) Exponential Correlation  
 $a = 1$



(d) Exponential Power Spectrum  
 $a = 1$



(e) Exponential Cosine Correlation  
 $a = 1, f_0 = 1$



(f) Exponential Cosine Power Spectrum  
 $a = 1, f_0 = 1$

Figure 7: Left (a, c, e): Some common stationary correlation functions. Right (b, d, f): Corresponding One-Sided Power Spectral Density Functions.

## 7.1 Chapman-Kolmogorov Equations and Limiting Probabilities

Let us consider now that we take  $n$  steps of the Markov chain from an initial state  $X_k = i$ . We denote the probability of transitioning to state  $j$  after  $n$  steps as

$$P_{ij}^n = P(X_{k+n} = j | X_k = i), \quad (96)$$

which we refer to as the *n-step transition probability*.

Next, let us consider that we take a total of  $n + m$  steps starting from a state  $X_0 = i$ , without loss of generality. We can then express the  $n + m$ -step transition probability as:

$$\begin{aligned}
P_{ij}^{n+m} &= P(X_{n+m} = j | X_0 = i) \\
&= \sum_{k=0}^{\infty} P(X_{n+m} = j, X_n = k | X_0 = i) \\
&= \sum_{k=0}^{\infty} P(X_{n+m} = j | X_n = k, X_0 = i) P(X_n = k | X_0 = i) \\
&= \sum_{k=0}^{\infty} P(X_{n+m} = j | X_n = k) P(X_n = k | X_0 = i) \\
&= \sum_{k=0}^{\infty} P_{kj}^m P_{ik}^n
\end{aligned} \tag{97}$$

These relations are known as the *Chapman-Kolmogorov equations* and can be interpreted as the summation of the probabilities of all paths that start in state  $i$  and end in state  $j$  after  $n + m$  steps. This is illustrated in Figure ?? for simple two and three state Markov chains. [Create this figure](#).

In matrix notation, let  $\mathbf{P}^{(n)}$  denote the  $n$ -step transition probability matrix. The Chapman-Kolmogorov equations state:

$$\mathbf{P}^{(n+m)} = \mathbf{P}^{(n)} \mathbf{P}^{(m)} \tag{98}$$

For example, we can express the two-step transition probability matrix by:

$$\mathbf{P}^{(2)} = \mathbf{P}^{(1+1)} = \mathbf{P}^{(1)} \mathbf{P}^{(1)} = \mathbf{P}^2 \tag{99}$$

Likewise, we can express the  $n$ -step transition probability matrix by:

$$\mathbf{P}^{(n)} = \mathbf{P}^{(n-1+1)} = \mathbf{P}^{(n-1)} \mathbf{P}^{(1)} = \mathbf{P}^n \tag{100}$$

In other words, the  $n$ -step transition matrix can be obtained by multiplying  $\mathbf{P}$  by itself  $n$  times.

Next, let us consider two states  $i, j$  such that  $P_{ij}^n > 0$  for  $n \geq 0$ . Here, we say that state  $j$  is *accessible* from state  $i$  because it is possible to transition from state  $i$  to state  $j$  in an arbitrary number of steps,  $n$ . Two states that are accessible to each other are said to *communicate* and states that communicate satisfy the following properties:

1. Any state  $i$  always communicates with itself since  $P_{ii}^0 = P(X_0 = i | X_0 = i) = 1$ .
2. If state  $i$  communicates with state  $j$ , then state  $j$  communicates with state  $i$ .
3. If state  $i$  communicates with state  $j$  and state  $j$  communicates with state  $k$ , then state  $i$  communicates with state  $k$ .

A Markov chain is said to be *irreducible* if all states communicate, i.e.  $P_{ij}^n > 0, \forall i, j$ . In other words, all states are accessible to all other states.

**Theorem 1** For an irreducible ergodic Markov chain,  $\pi_j = \lim_{n \rightarrow \infty} P_{ij}^n$  exists, is independent of the initial state  $i$ , and is the unique solution of

$$\begin{aligned}
\pi_j &= \sum_{i=0}^{\infty} \pi_i P_{ij}, \quad j \geq 0 \\
\text{with } \sum_{j=0}^{\infty} \pi_j &= 1
\end{aligned} \tag{101}$$

The probability  $\pi_j$  is referred to as the *limiting probability* that the process will be in state  $j$ . This is the long-run proportion of time that the process is in state  $j$  and is also commonly referred to as the *stationary probability* because it represents the probability mass (density for continuous state space) for the state  $j$ .

## 7.2 Time Reversible Markov Chains

Consider a stationary ergodic Markov chain with transition probabilities  $P_{ij}$  and stationary probabilities  $\pi_i$ . Let us consider that we now start marching backward in time. That is, from state  $X_n$  we move to states  $X_{n-1}, X_{n-2}, \dots$ . The reverse process is also a Markov chain (for proof, see [2]) with transition probabilities

$$Q_{ij} = \frac{\pi_j P_{ji}}{\pi_i} \quad (102)$$

If  $Q_{ij} = P_{ij}, \forall i, j$ , the Markov chain is said to be *time reversible* and we therefore have

$$\pi_i P_{ij} = \pi_j P_{ji} \quad (103)$$

We can interpret the product  $\pi_i P_{ij}$  as the rate at which the process goes from state  $i$  to state  $j$  since it is the probability of residing in state  $i$  and from there moving to state  $j$ . The condition in Eq. (103) states that, for a reversible Markov chain, the rate at which the process moves from  $i$  to  $j$  is equal to the rate at which the process moves from  $j$  to  $i$ . This will be an important property for the development of Markov chain Monte Carlo methods later.

## 7.3 Example

Consider a simple two-state weather model in which the chance of rain tomorrow depends only on whether it is raining today. If it rains today, there is a probability  $\alpha$  that it will rain tomorrow. If it does not rain today, there is a probability  $\beta$  that it will rain tomorrow. This can be described through the transition probability matrix:

$$\mathbf{P} = \begin{bmatrix} \alpha & 1 - \alpha \\ \beta & 1 - \beta \end{bmatrix} \quad (104)$$

Let us consider  $\alpha = 0.7$  and  $\beta = 0.4$  and determine the probability that it will rain in 2 days, 4 days, and 10 days. Using the Chapman-Kolmogorov equations, we obtain

2 Days

$$\mathbf{P}^{(2)} = \mathbf{P}^2 = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} = \begin{bmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{bmatrix} \quad (105)$$

4 Days

$$\mathbf{P}^{(4)} = \mathbf{P}^4 = \begin{bmatrix} 0.5749 & 0.4251 \\ 0.5668 & 0.4332 \end{bmatrix} \quad (106)$$

10 Days

$$\mathbf{P}^{(4)} = \mathbf{P}^4 = \begin{bmatrix} 0.5714 & 0.4286 \\ 0.5714 & 0.4286 \end{bmatrix} \quad (107)$$

Next, we can solve for the limiting probabilities by solving the following system of equations:

$$\begin{aligned} \pi_0 &= \alpha\pi_0 + \beta\pi_1 \\ \pi_1 &= (1 - \alpha)\pi_0 + (1 - \beta)\pi_1 \\ 1 &= \pi_0 + \pi_1 \end{aligned} \quad (108)$$

which yields the solution

$$\pi_0 = \frac{\beta}{1 + \beta - \alpha} = 0.5714; \quad \pi_1 = \frac{1 - \alpha}{1 + \beta - \alpha} = 0.4286 \quad (109)$$

Finally, we can show that this Markov chain is reversible by solving for each of the transition probabilities of the reverse process  $Q_{ij} = \frac{\pi_j P_{ji}}{\pi_i}$  as

$$\begin{aligned} Q_{00} &= \frac{\pi_0}{\pi_0} P_{00} = P_{00} \\ Q_{01} &= \frac{\pi_1}{\pi_0} P_{10} = \frac{\frac{1 - \alpha}{1 + \beta - \alpha}}{\frac{\beta}{1 + \beta - \alpha}} \beta = \frac{1 - \alpha}{\beta} \beta = 1 - \alpha = P_{01} \\ Q_{10} &= \frac{\pi_0}{\pi_1} P_{01} = \frac{\frac{\beta}{1 - \alpha}}{\frac{1 + \beta - \alpha}{1 + \beta - \alpha}} (1 - \alpha) = \frac{\beta}{1 - \alpha} (1 - \alpha) = \beta = P_{10} \\ Q_{11} &= \frac{\pi_1}{\pi_1} P_{11} = P_{11} \end{aligned} \quad (110)$$

## References

- [1] R. Lebrun, A. Dutfoy, An innovating analysis of the nataf transformation from the copula viewpoint, Probabilistic Engineering Mechanics 24 (3) (2009) 312–320.
- [2] S. M. Ross, Introduction to probability models, Academic press, 2014.

## Nomenclature

### Functions

- $P(\cdot)$  Probability measure
- $f_X(\cdot)$  Probability density function (PDF) of a random variable  $X$
- $f_{XY}(\cdot)$  Joint cumulative density function of the random variables  $X$  and  $Y$
- $f_{\mathbf{X}}(\cdot)$  Joint probability density function of the random vector  $\mathbf{X}$
- $f_{X|Y}(\cdot|\cdot)$  Conditional probability density function of the random variable  $X$  conditioned on the random variable  $Y$
- $F_X(\cdot)$  Cumulative distribution function (CDF) of a random variable  $X$
- $F_{XY}(\cdot)$  Joint cumulative distribution function of the random variables  $X$  and  $Y$
- $F_{\mathbf{X}}(\cdot)$  Joint cumulative distribution function of the random vector  $\mathbf{X}$
- $F_{X|Y}(\cdot|\cdot)$  Conditional cumulative distribution function of the random variable  $X$  conditioned on the random variable  $Y$

- $\phi(\cdot)$  Probability Density Function of a standard normal random variable  
 $\Phi(\cdot)$  Cumulative distribution function of a standard normal random variable  
 $\delta(\cdot)$  Dirac delta function  
 $\mathbb{E}[\cdot]$  Expected value of a random variable. Also denoted  $\mu_X \triangleq \mathbb{E}[X]$   
 $\mathbb{E}[X^n]$  The  $n^{th}$  moment about the origin of a random variable  $X$   
 $\langle X(\cdot) \rangle$  Temporal average mean value of a random process  $X(\cdot)$   
 $\langle X(\cdot)X(\cdot) \rangle$  Temporal average autocorrelation function of a random process  $X(\cdot)$   
 $\text{Cov}(\cdot)$  The covariance of the random variable input  
 $\text{Var}(\cdot)$  Variance of the random variable. Also denoted  $\sigma_X^2 \triangleq \text{Var}(X)$   
 $C_{XX}(\cdot)$  Covariance function of the random process  $X(\cdot)$   
 $G_{XX}(\cdot)$  One-sided power spectral density function of the random process  $X(\cdot)$   
 $R_{XX}(\cdot)$  Autocorrelation function of the random process  $X(\cdot)$   
 $S_{XX}(\cdot)$  Power Spectral Density Function of the random process  $X(\cdot)$

## Operators

- $\cap$  Intersection  
 $\cup$  Union

## Variables

- $X$  A random variable  
 $\mathbf{X}$  A random vector in  $\mathbb{R}^n$   
 $X(\cdot)$  A random process, or a random field  
 $\rho_{XY}$  Correlation Coefficient of the random variables  $X$  and  $Y$   
 $\mu'_{nm}$  Joint moment about the origin of order  $n + m$   
 $\mu_{nm}$  Joint central moments, also known as moments about the mean, of order  $n + m$   
 $\Sigma$  The  $n$  by  $n$  covariance matrix of a random vector  $\mathbf{X} \in \mathbb{R}^n$