1 Random Processes and Vectors

Till now we have dealt with sampling and representation of signals using different orthonormal bases. We have seen that we can recover a signal from its point-wise samples given that a signal is bound limited, i.e., is zero starting from a certain point in a given transform domain, e.g., the Fourier domain.

Notice that adding an assumption on the structure of the signal helps with its processing. Such assumptions are often denoted as signal priors or signal models. One common option for modeling signals is by having a prior on the distribution of the signals’ coefficients under a certain given basis. In this case the signals are looked at as random processes in the continuous case and random vectors in the discrete case. In this lecture we give a short introduction to random processes and vectors.

1.1 Sample Space and Probability Distribution

Let $\Omega$ be a sample space and $P$ a probability distribution. The sample space $\Omega$ defines the values that can be drawn in the random process and $P$ defines the distribution to draw each of the values. When the values in $\Omega$ are discrete the probability distribution $P$ defines a probability mass function (pmf), that set the probability to pick each of the values. In the case where the values in $\Omega$ are continuous, it is impossible to assign a probability for each value and thus weights are assigned to each by a probability density function (pdf). The integral over each fraction gives the probability of drawing a value from this fraction. Clearly, the integral over all values is 1.

We give several examples for $\Omega$ and $P$:

- **Bernoulli distribution** with parameter $p$: $\Omega = \{0, 1\}$ and the pmf is
  \[ f(k) = \begin{cases} p & k = 1 \\ 1 - p & k = 0 \end{cases}, \]
  i.e., the probability for getting 1 is $p$ and 0 is $1 - p$.

- **Binomial distribution** with parameters $p$ and $n$ ($B(n, p)$): $\Omega = \{0, 1, \ldots, n\}$ and the pmf is
  \[ f(k) = \binom{n}{k} p^k (1 - p)^{n-k}, \]
  where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

- **Poisson Distribution** with parameter $\lambda$ ($\text{Pois}(\lambda)$): $\Omega = \mathbb{N}$ and the pmf is
  \[ f(k) = \frac{\lambda^k e^{-\lambda}}{k!}. \]
• **Uniform distribution** with parameters $a$ and $b$ ($U[a,b]$): $\Omega = [a,b]$, where $-\infty < a < b < \infty$, and the pdf is

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{else} \end{cases} . \quad (4)$$

• **Exponential distribution** with parameter $\lambda$: $\Omega = \mathbb{R}^+$ and the pdf is

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases} . \quad (5)$$

• **Gaussian distributions** with parameters $\mu$ and $\sigma$ ($N(\mu,\sigma)$): $\Omega = \mathbb{R}$ and the pdf is

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (6)$$

• **Laplace distributions** with parameters $\mu$ and $b$: $\Omega = \mathbb{R}$ and the pdf is

$$f(x) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}} . \quad (7)$$

### 1.2 Random Variable

Given $\Omega$ and $P$, one can define a random variable. The value of this variable are selected from $\Omega$ with a probability defined by $\Omega$. Though a random variable is stochastic and does not have a single fixed value, we can still characterize its behavior. Two important properties of a random variable are its expected value, known also as the mean, and its variance.

The mean $\mu \triangleq E[x]$ is defined as

$$\mu = E[x] = \sum_{k \in \Omega} kf(k) \quad (8)$$

for discrete variables, and as

$$\mu = E[x] = \int_{\Omega} x f(x) dx \quad (9)$$

for continuous variables. Notice that the above sum and integral do not always defined and thus the mean does not always exists. When it does, it characterizes the value around which $x$ is expected to be. Another way to look at it is that if we will generate $x$ infinite times then $\mu$ will be the mean of the values we get.

The variance $\sigma^2 \triangleq \text{Var}(x) \triangleq E[(x-\mu)^2]$ is defined as

$$\text{Var}(x) \triangleq E[(x-\mu)^2] = \sum_{k \in \Omega} (k-\mu)^2 f(k) \quad (10)$$

for discrete variables, and as

$$\text{Var}(x) \triangleq E[(x-\mu)^2] = \int_{\Omega} (x-\mu)^2 f(x) dx \quad (11)$$

for continuous variables. The variance determines how much the values drawn in the random variable are concentrated around the mean. The smaller the variance the closer they are. Note that zero variance implies a deterministic $x$. Some may use the square root of the variance $\sigma$, known as the standard deviation, to measure the variation from the average.
It is easy to see that

\[
\text{Var}(x) = E[(x - E[x])^2] = E[x^2] - (E[x])^2 = E[x^2] - \mu^2.
\]  

(12)

The measure \(E[x^2]\) is denoted as the second moment of the random variable \(x\). In general, the \(n\)-th moment of a \(x\) is defined as

\[
\mu_n = E[x^n].
\]  

(13)

From this definition it can be seen that the first moment is actually the mean of the variable.

### 1.3 Random Processes and Random Vectors

Given a sample space \(\Omega\) and a probability distribution, a random process \(X\), also denoted a stochastic process, is a collection of random variables

\[
\{X(t) : t \in T\}
\]  

(14)

that have a joint distribution, where \(T\) is the "time" of the process. Each time point \(t \in T\) corresponds to a random variable \(X(t)\).

Since a random process is a collection of random variables, we can define the same properties we have defined before for each. The mean of a random process is defined as

\[
\mu(t) = E[x(t)].
\]  

(15)

For higher moments notice that as we have more than one variable we are not only interested in the statistics of each alone but in the mutual correlations. Thus, we define the autocorrelation of a random process \(X\) as

\[
R_X(s, t) = E[X(s)X(t)].
\]  

(16)

Notice that the \(R_X(t, t)\) is the second moment of the variable at time point \(t\). In a similar way to how the variance is defined, we have the autocovariance

\[
C_X(s, t) = E[X(s)X(t)] - \mu(s)\mu(t).
\]  

(17)

A random process is uncorrelated if

\[
C_X(s, t) = 0 \quad \forall s \neq t \in T.
\]  

(18)

Every random process with independent random variables is also uncorrelated. However, note that the vice versa is not necessarily true.

We say that \(X\) is a stationary process if its mean is independent of the time and its autocorrelation is dependent only at the time difference, i.e., if

\[
\mu(t) = E[x(t)] = \mu,
\]  

(19)

and

\[
R_X(s, t) = R_X(s - t).
\]  

(20)

Notice that the above imply that also for the autocovariance

\[
C_X(s, t) = C_X(s - t).
\]  

(21)
A random vector is a finite random process, where the random variables are ordered in a vector form. For a random vector $\mathbf{x} \in \mathbb{R}^n$, the mean is a vector $\boldsymbol{\mu} \in \mathbb{R}^n$ and the autocorrelation and autocovariance turns to be the autocorrelation matrix

$$R_{\mathbf{x}} = E[\mathbf{x}\mathbf{x}^*]$$

and the covariance matrix

$$C_{\mathbf{x}} = E[ (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^* ] = E[\mathbf{x}\mathbf{x}^*] - E[\mathbf{x}]\boldsymbol{\mu}^* - \boldsymbol{\mu}E[\mathbf{x}] + \boldsymbol{\mu}\boldsymbol{\mu}^* = E[\mathbf{x}\mathbf{x}^*] - \boldsymbol{\mu}\boldsymbol{\mu}^*$$

If $C_{\mathbf{x}}$ is diagonal then we say that $\mathbf{x}$ in uncorrelated random vector. As we have mentioned before, if the variables in $\mathbf{x}$ are independent then $\mathbf{x}$ is also uncorrelated. However, uncorrelation does not imply necessarily independent.

### 1.4 Independent and Identically Distributed Random Processes and Vectors

A collection of random variables is mutually independent if their joint distribution equals to the product of each of them separately. For example, if $x_1, x_2, x_3, ..., x_n$ are mutually independent then

$$P(x_1 < a_1, x_2 < a_2, ..., x_n < a_n) = P(x_1 < a_1)P(x_2 < a_2) \cdots P(x_n < a_n).$$

The same holds also for the expectation

$$E[x_1x_2 \cdots x_n] = E[x_1]E[x_2] \cdots E[x_n].$$

A random process is independent and identically distribution (i.i.d) if any collection of random variables in it is independent and all have the same distribution. If $X(t)$ is i.i.d then for any collection of time points $t_1 < t_2 < \cdots < t_n$

$$P(X(t_1) < a_1, X(t_2) < a_2, ..., X(t_n) < a_n) = P(X(t_1) < a_1)P(X(t_2) < a_2) \cdots P(X(t_n) < a_n)$$

Note that this implies that $X(t)$ is stationary and

$$E[X(t_1)X(t_2) \cdots X(t_n)] = E[X(t_1)]E[X(t_2)] \cdots E[X(t_n)] = \mu^n,$$

and

$$C_X(\tau) = \begin{cases} \text{Var}(X(0)) & \tau = 0 \\ 0 & \tau \neq 0 \end{cases}.$$
1.5 Examples for Random Vectors

We hereby give several examples for random vectors with their mean and covariance matrix.

- i.i.d Rademacher random vector: Each variable in the vector equals ±1 w.p. (with probability) 0.5. All variables are mutually independent. The mean is zero and the covariance matrix is
  \[ \mathbf{C_x} = \mathbf{I}. \] (31)

- i.i.d uniformly distributed random vector:
  \[ \forall i, x_i \sim U[a, b] \] (32)
  All variables are mutually independent. The mean is \( \frac{1}{2} (a + b) \) and the covariance matrix is
  \[ \mathbf{C_x} = \frac{1}{12} (b - a)^2 \mathbf{I}. \] (33)

- i.i.d Gaussian random vector:
  \[ \forall i, x_i \sim N(\mu, \sigma) \] (34)
  All variables are mutually independent. The pdf is:
  \[ f(x) = f(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{\sum_{i=1}^{n} (x_i-\mu)^2}{2\sigma^2}}. \] (35)
  The mean is \( \mu \) and the covariance matrix is
  \[ \mathbf{C_x} = \sigma^2 \mathbf{I}. \] (36)

- Gaussian random vector with mean \( \mu \) and covariance matrix \( \Sigma \) \( (N(\mu, \Sigma)) \). The pdf is:
  \[ f(x) = f(x_1, x_2, \ldots, x_n) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} e^{-\frac{1}{2}(x-\mu)^\top \Sigma^{-1} (x-\mu)}, \] (37)
  where \( |\Sigma| \) is the determinant of \( \Sigma \). Notice that if \( \mathbf{x} \) is i.i.d, i.e., \( \mu = \mu \mathbf{1} \) and \( \Sigma = \sigma^2 \mathbf{I} \), we get back (35).

1.6 Random Vectors and Signals

The theory of random process and vectors can be used to model signals. Instead of having a deterministic coefficients for a given a signal, these can be modeled using a random vector.

Given a set of orthonormal functions \( \{\psi_i(t)\}_{i=1}^{N} \) in the continuous case, we can write a signal as
\[ \varphi(t) = \sum_{i=1}^{N} \varphi_i \psi_i(t). \] (38)
where \( \varphi = [\varphi_1, \varphi_2, \ldots, \varphi_N]^\top \) is a random vector. Notice that now, the signal \( \varphi(t) \) is a random process.

The same can be done for discrete signals. For a given orthonormal basis \( \{\psi_i\}_{i=1}^{N} \), a signal \( \varphi \) can be modeled as
\[ \varphi = \sum_{i=1}^{N} \varphi_i \psi_i, \] (39)
with \( \varphi = [\varphi_1, \varphi_2, \ldots, \varphi_N]^\top \) a random vector. Notice that in this case the signal \( \varphi \) is a random vector with its own statistics.
1.6.1 i.i.d. White Noise

In many cases we have a random noise in the system. Unlike natural signals which are expected to have some structure, the noise components are usually independent of each other and have more or less the same distribution.

Thus, we model noise as an i.i.d white process, which means that the noise is zero mean, independent and has the same variance everywhere. The reason we can assume that a noise has a zero expected value is that if this is not the case it will have a constant value due to the i.i.d property. This mean is easy to estimate and we can always subtract it from the system ending up with a zero mean noise.

1.7 Random Vectors and Orthonormal Bases

Let $\phi$ be a given random vector representing a signal under the standard basis:

$$\phi = \sum_{i=1}^{N} \phi_i e_i. \quad (40)$$

Given the mean

$$\mu_\phi = E[\phi], \quad (41)$$

the autocorrelation matrix

$$R_\phi = E[\phi\phi^*] \quad (42)$$

and the covariance matrix

$$C_\phi = E[\phi\phi^*] - \mu_\phi\mu_\phi^*, \quad (43)$$

we ask how does they look if we move to another orthonormal basis.

Let $\Psi$ be such a basis (the $i$-th column of $\Psi$ is the $i$-th basis vector $\psi_i$). We can write $\phi$ under this basis as

$$\phi = I_\phi = \Psi (\Psi^* \phi) = \sum_{i=1}^{N} \langle \psi_i, \phi \rangle \psi_i = \sum_{i=1}^{N} \phi_i^\Psi \psi_i. \quad (44)$$

Therefore, the expected value of a given representation coefficient under this basis is

$$E[\phi_i^\Psi] = E[\langle \psi_i, \phi \rangle] = \langle \psi_i, E[\phi] \rangle = \langle \psi_i, \mu_\phi \rangle, \quad (45)$$

where we have used the linearity property of the expectation. Looking at the expected value of the whole representation vector we have

$$\mu_\phi^\Psi = E[\phi^\Psi] = \Psi^* \mu_\phi. \quad (46)$$

Notice that the above implies that a centered process (with zero mean) stays centered under any orthonormal basis.

Using the fact that $\Psi^\ast \Psi = I$, we have

$$R_\phi = E[\phi\phi^*] = E[\Psi\Psi^* \phi\phi^* \Psi\Psi^*] = \Psi E[(\Psi^* \phi)(\phi^* \Psi)] \Psi^* = \Psi E[(\Psi^* \phi)(\Psi^* \phi)^*] \Psi^* = \Psi R_\phi^\Psi \Psi^*. \quad (47)$$
Thus, we have
\[ R_\varphi = \Psi R_\varphi \Psi^*, \quad \text{(48)} \]
\[ R_\varphi^\psi = \Psi^* R_\varphi \Psi. \quad \text{(49)} \]

Note that we have \( R_\varphi \) and \( R_\varphi^\psi \) are similar.

Plugging (46) and (48) in (43), we have that the covariance matrices are also similar:
\[ C_\varphi = \Psi C_\varphi \Psi^*, \quad \text{(50)} \]
\[ C_\varphi^\psi = \Psi^* C_\varphi \Psi. \quad \text{(51)} \]

1.7.1 i.i.d. White Noise and Orthonormal Bases

Let \( \eta \) be an i.i.d white noise:
\[ \mu_\eta = E[\eta] = 0 \quad \text{(52)} \]

and
\[ C_\eta = R_\eta = \sigma^2 I. \quad \text{(53)} \]

Let \( \Psi \) be an orthonormal bases and \( \eta^\Psi = \Psi^* \eta \) the noise representation under the noise basis. Using (46) and (48) it is easy to see that the noise has the same statistics also in the new basis:
\[ \mu_{\eta^\Psi} = E[\eta^\Psi] = 0 \quad \text{(54)} \]

and
\[ C_{\eta^\Psi} = R_{\eta^\Psi} = \sigma^2 I. \quad \text{(55)} \]

1.8 Properties of the Autocorrelation and the Covariance matrix

We provide some properties of the autocorrelation matrix. The same holds also for the covariance matrix.

1. The autocorrelation matrix matrix \( R_\varphi \) is positive semi-definite, i.e., for any vector \( x \in \mathbb{R}^N \):
\[ x^* R_\varphi x \geq 0. \]

   \textbf{Proof.} Let \( x \in \mathbb{R}^N \). Then
\[ x^* R_\varphi x = E[x^* \varphi \varphi^* x] = E[(x^* \varphi)(\varphi^* x)] = E[(x^* \varphi)^2] \geq 0. \quad (56) \]

2. The autocorrelation matrix \( R_\varphi \) is symmetric
\[ R_\varphi = R_\varphi^*. \quad \text{(57)} \]

3. The preceding two properties implies that \( R_\varphi \) has a diagonalizable matrix with real non-negative eigenvalues and a corresponding set of eigenvectors that form an orthonormal basis: There exists a set of pairs \( \{\lambda_i, u_i\}_{i=1}^N \) such that
\[ R_\varphi u_i = \lambda_i u_i, \quad \text{(58)} \]
and
\[ u_i u_k = \delta_{i,k}. \] (59)

Summing everything together we have
\[ R_{\mathbf{\phi}} [u_1 u_2 \ldots u_N] = [u_1 u_2 \ldots u_N] \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_N). \] (60)

Denoting \( U = [u_1 u_2 \ldots u_N] \) and \( \Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_N) \) we have
\[ R_{\mathbf{\phi}} U = U \Lambda. \] (61)

4. It is possible to select an orthonormal basis \( U \) such that \( U^* \mathbf{\phi} \) has a diagonal autocorrelation matrix \( \Lambda \). The basis is the eigenvectors of \( R_{\mathbf{\phi}} \) and the diagonal autocorrelation matrix contains the corresponding eigenvalues. This can be easily seen from (61). By multiplying it by \( U^* \) from the lhs we have
\[ R_{\mathbf{\phi}} = U \Lambda U^*. \] (62)

We get that \( \Lambda \) is the autocorrelation matrix of \( U^* \mathbf{\phi} \).