Random signals

1 Elementary probability

Probability measure  We start with a few elementary (and simplified) definitions from the theory of probability. Let us fix a sample space $\Omega = [0, 1]$. A Borel set on $\Omega$ is a set that can be formed from open intervals of the form $(a,b), 0 \leq a < b \leq 1$, through the operations of countable union, countable intersection, and set difference. We will denote the collection of all Borel sets in $\Omega$ as $\Sigma$. It is pretty straightforward to show that $\Sigma$ contains the empty set, is closed under complement, and is closed under countable union. Such a set is known as a $\sigma$-algebra and its elements (subsets of $\mathbb{R}$) are referred to as events.

A probability measure $P$ on $\Sigma$ is a function $P : \Sigma \rightarrow [0, 1]$ satisfying $P(\emptyset) = 0$, $P(\mathbb{R}) = 1$ and additivity for every countable collection $\{E_n \in \Sigma\}$,

$$P\left(\bigcup_n E_n\right) = \sum_n P(E_n).$$

Random variables  A random variable $X$ is a measurable map $X : \Omega \rightarrow \mathbb{R}$, i.e., a function such that for every $a$, $\{X \leq a\} = \{\alpha : X(\alpha) \leq a\} \in \Sigma$. The map $X$ pushes forward the probability measure $P$; the pushforward measure $X_* P$ is given by

$$(X_* P)(A) = P(X^{-1}(A)),$$

where $X^{-1}(A) = \{\alpha : X(\alpha) \in A\}$ is the preimage of $A \subseteq \mathbb{R}$. (In short, we can write $X_* P = PX^{-1}$). This pushforward probability measure $X_* P$ is usually referred to as the probability distribution (or the law) of $X$.

When the range of $X$ is finite or countably infinite, the random variable is called discrete and its distribution can be described by the probability mass function (PMF):

$$f_X(x) = P(X = x),$$

which is a shorthand for $P(\{\alpha : X(\alpha) = x\})$. Otherwise, $X$ is called a continuous random variable. Any random variable can be described by the cumulative distribution function (CDF)

$$F_X(x) = P(X \leq x),$$

which is a shorthand for $F_X(x) = P(\{\alpha : X(\alpha) \leq x\})$. If $X$ is absolutely continuous, the CDF can be described by the integral

$$F_X(x) = \int_{-\infty}^x f_X(x')dx'.$$
where the integrand $f_X$ is known as the \textit{probability density function} (PDF)\(^1\).

\textbf{Expectation} \hspace{1em} The \textit{expected value} (a.k.a. the \textit{expectation} or \textit{mean}) of a random variable $X$ is given by

$$
\mathbb{E}X = \int_{\mathbb{R}} \text{id} \, d(X \ast P) = \int_{\Omega} X(\alpha) \, d\alpha,
$$

where the integral is the Lebesgue integral w.r.t. the measure $P$; whenever a probability density function exists, the latter can be written as

$$
\mathbb{E}X = \int_{\mathbb{R}} x f_X(x) \, dx.
$$

Note that due to the linearity of integration, the expectation operator $\mathbb{E}$ is linear. Using the Lebesgue integral notation, we can write for $E \in \Sigma$

$$
P(E) = \int_E dP = \int_{\mathbb{R}} 1_E \, dP = \mathbb{E}1_E,
$$

where

$$
1_E(\alpha) = \begin{cases} 
1 & : \alpha \in E \\
0 & : \text{otherwise}
\end{cases}
$$

is the \textit{indicator function} of $E$, which is by itself a random variable. This relates the expectation of the indicator of an event to its probability.

\textbf{Moments} \hspace{1em} For any measurable function $g : \mathbb{R} \to \mathbb{R}$, $Z = g(X)$ is also a random variable with the expectation

$$
\mathbb{E}Z = \mathbb{E}g(X) = \int_{\mathbb{R}} g \, dP = \int_{\mathbb{R}} g(x) f_X(x) \, dx.
$$

Such an expectation is called a \textit{moment} of $X$. Particularly, the $k$-th order moment is obtained by setting $g(x) = x^k$,

$$
\mu_k(X) = \mathbb{E}X^k.
$$

The expected value itself is the first-order moment of $X$, which is often denoted simply as $\mu_X = \mu_1(X)$. The \textit{central} $k$-th order moment is obtained by setting $g(x) = (x - \mu_X)^k$,

$$
m_k(X) = \mathbb{E}(X - \mathbb{E}X)^k.
$$

A particularly important central second-order moment is the \textit{variance}

$$
\sigma_X^2 = \text{Var} X = m_2 = \mathbb{E}(X - \mathbb{E}X)^2 = \mu_2(X) - \mu_X^2.
$$

\(^1\)To be completely rigorous, the proper way to define the PDF is by first equipping the image of the map $X$ with the Lebesgue measure $\lambda$ that assigns to every interval $[a, b]$ its length $b - a$. Then, we invoke the Radon-Nikodym theorem saying that if $X$ is absolutely continuous w.r.t. $\lambda$, there exists a measurable function $f : \mathbb{R} \to [0, \infty)$ such that for every measurable $A \subset \mathbb{R}$, $(X \ast P)(A) = P(X^{-1}(A)) = \int_A f \, d\lambda$. $f$ is called the \textit{Radon-Nikodym derivative} and denoted by $f = \frac{d(X \ast P)}{d\lambda}$. It is exactly our PDF.
Joint distribution  A vector $\mathbf{X} = (X_1, \ldots, X_n)$ of random variables is called a random vector. Its probability distribution is defined as before as the pushforward measure $P = \mathbf{X}_* \lambda$. It is customary to treat $\mathbf{X}$ as a collection of $n$ random variables and define their joint CDF as

$$F_\mathbf{X}(\mathbf{x}) = P(\mathbf{X} \leq \mathbf{x}) = P(\{X_1 \leq x_1\} \times \cdots \times \{X_n \leq x_n\}).$$

As before, whenever the following holds

$$F_\mathbf{X}(\mathbf{x}) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_\mathbf{X}(x_1', \ldots, x_n') dx_1' \cdots dx_n',$$

the integrand $f_\mathbf{X}$ is called the joint PDF of $\mathbf{X}$. The more rigorous definition as the Radon-Nikodym derivative

$$f_\mathbf{X} = \frac{d(\mathbf{X}_* P)}{d\lambda}$$

stays unaltered, only that now $\lambda$ is the $n$-dimensional Lebesgue measure.

Marginal distribution  Note that the joint CDF of the sub-vector $(X_2, \ldots, X_n)$ is given by

$$F_{X_2,\ldots,X_n}(x_2, \ldots, x_n) = P(X_2 \leq x_2, \ldots, X_n \leq x_n) = P(\{X_1 \leq \infty\} \times \cdots \times \{X_n \leq x_n\}).$$

Such a distribution is called marginal w.r.t. $X_1$ and the process of obtaining it by substituting $x_1 = \infty$ into $F_\mathbf{X}$ is called marginalization. The corresponding action in terms of the PDF consists of integration over $x_1$,

$$f_{X_2,\ldots,X_n}(x_2, \ldots, x_n) = \int_\mathbb{R} f_{X_1,\ldots,X_n}(x_1, x_2, \ldots, x_n) dx_1.$$

Statistical independence  A set $X_1, \ldots, X_n$ of random variables is called statistically independent if their joint CDF is coordinate-separable, i.e., can be written as the following tensor product

$$F_{X_1,\ldots,X_n} = F_{X_1} \otimes \cdots \otimes F_{X_n}.$$

An alternative definition can be given in terms of the PDF (whenever it exists):

$$f_{X_1,\ldots,X_n} = f_{X_1} \otimes \cdots \otimes f_{X_n}.$$

We will see a few additional alternative definitions in the sequel.
Limit theorems

Given independent identically distributed (i.i.d.) variables \(X_1, \ldots, X_n\) with mean \(\mu\) and variance \(\sigma^2\), we define their sample average as

\[
S_n = \frac{1}{n}(X_1 + \cdots + X_n).
\]

Note that \(S_n\) is also a random variable with \(\mu_{S_n} = \mu\) and \(\sigma^2_{S_n} = \frac{\sigma^2}{n}\). It is straightforward to see that the variance decays to zero in the limit \(n \to \infty\), meaning that \(S_n\) approaches a deterministic variable \(S = \mu\). However, a much stronger result exists: the (strong) law of large numbers states that in the limit \(n \to \infty\), the sample average converges in probability to the expected value, i.e.,

\[
P \left( \lim_{n \to \infty} S_n = \mu \right) = 1.
\]

This fact is often denoted as \(S_n \xrightarrow{p} \mu\). Furthermore, defining the normalized deviation from the limit \(D_n = \sqrt{n}(S_n - \mu)\), the central limit theorem states that \(D_n\) converges in distribution to \(N(0, \sigma^2)\), that is, its CDF converges pointwise to that of the normal distribution. This is often denoted as \(D_n \xrightarrow{D} N(0, \sigma^2)\).

A slightly more general result is known as the delta method in statistics: if \(g : \mathbb{R} \to \mathbb{R}\) is a \(C^1\) function with non-vanishing derivative, then by the Taylor theorem,

\[
g(S_n) = g(\mu) + g'(\nu)(S_n - \mu) + \mathcal{O}(|S_n - \mu|^2),
\]

where \(\nu\) lies between \(S_n\) and \(\mu\). Since by the law of large numbers \(S_n \xrightarrow{p} \mu\), we also have \(\nu \xrightarrow{p} \mu\); since \(g'\) is continuous, \(g'(\nu) \xrightarrow{p} g'(\mu)\). Rearranging the terms and multiplying by \(\sqrt{n}\) yields

\[
\sqrt{n}(g(S_n) - g(\mu)) = g'(\nu)\sqrt{n}(S_n - \mu) = g'(\nu)D_n,
\]

from where (formally, by invoking the Slutsky theorem):

\[
\sqrt{n}(g(S_n) - g(\mu)) \xrightarrow{D} N(0, g'(\mu)^2\sigma^2).
\]

Joint moments

Given a measurable function \(g : \mathbb{R}^n \to \mathbb{R}^m\), a (joint) moment of a random vector \(X = (X_1, \ldots, X_n)\) is

\[
\mathbb{E}g(X) = \int g(x) dP = \begin{pmatrix}
\int g_1(x) dP \\
\vdots \\
\int g_m(x) dP
\end{pmatrix} = \begin{pmatrix}
\int_{\mathbb{R}^n} g_1(x) f_X(x) dx \\
\vdots \\
\int_{\mathbb{R}^n} g_m(x) f_X(x) dx
\end{pmatrix};
\]

the last term might be undefined if the PDF does not exist. The mean of a random vector is simply \(\mu_X = \mathbb{E}X\). Of particular importance are the second-order joint moments of pairs of random variables,

\[
r_{XY} = \mathbb{E}XY.\]
and its central version

$$\sigma_{XY}^2 = \text{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}X)(Y - \mathbb{E}Y)) = r_{XY} - \mu_X \mu_Y.$$ 

The latter quantity is known as the covariance of $X$ and $Y$.

Two random variables $X$ and $Y$ with $r_{XY} = 0$ are called orthogonal\(^2\). The variables with $\sigma_{XY}^2 = 0$ are called uncorrelated. Note that for a statistically independent pair $(X, Y)$,

$$\sigma_{XY}^2 = \int_{\mathbb{R}^2} (x - \mu_X)(y - \mu_Y)d((X \times Y), P) = \int_{\mathbb{R}} (x - \mu_X)d(X, P) \int_{\mathbb{R}} (y - \mu_Y)d(Y, P)$$

$$= \mathbb{E}(X - \mathbb{E}X) \cdot \mathbb{E}(Y - \mathbb{E}Y) = 0.$$ 

However, the converse is not true, i.e., lack of correlation does not generally imply statistical independence (with the notable exception of normal variables).

If $X$ and $Y$ are uncorrelated and furthermore one of them is zero-mean, then they are also orthogonal (and the other way around).

## 2 Random signals

A random signal or a stochastic process is a collection of random variables $\{F(x) : x \in D\}$ indexed by some set $D$ (called the domain) and assuming values in some space $S$ (called the state space). For our purpose, we will assume $D$ to be either $\mathbb{R}^d$ (in this case we will refer to the process as to a continuous-domain stochastic process) or $\mathbb{Z}^d$ (discrete-domain process); the state space $D$ will be assumed either $\mathbb{R}$ (continuous state) or $\mathbb{Z}$ (discrete state). Informally, setting $D = \mathbb{R}^2$ and $S = \mathbb{R}$ we can think of $F$ as of a random image. When $d > 1$, it is customary to call the stochastic process a random field. We will henceforth use the term “random signal”. In what follows, we will almost always tacitly assume $D$ to be $\mathbb{R}^d$, the very same reasoning applies to discrete-domain signals mutatis mutandis.

Formally, a random signal $F$ is a function $F : D \times \Omega \to S$, where $\Omega$ denotes the sample space. The first argument $x \in D$ in $F(x, \alpha)$ sets the location in the domain, while the second argument $\alpha \in \Omega$ is responsible for the randomness. Fixing some $\alpha \in \Omega$, the resulting deterministic function $f(x) = F(x, \alpha)$ is called a realization or a sample function of $F$. Fixing some $x \in D$, we obtain a random variable $F(\alpha) = F(x, \alpha)$ describing the randomness of the signal sampled at a fixed location $x$.

Note that for a singletone $D = \{1\}$, the signal is just a random variable, while for a finite $D = \{1, 2, \ldots, n\}$ it is a random vector. Random signals can be therefore thought of as a generalization of random vectors to “random functions”. However, in such an infinitely-dimensional case it is not easy to define the standard notions such as density. Instead, we are\(^2\) in fact, $r_{XY}$ can be viewed as an inner product on the space of random variables. This creates a geometry isomorphic to the standard Euclidean metric in $\mathbb{R}^n$. Using this construction, the Cauchy-Schwarz inequality immediately follows: $|r_{XY}| \leq \sigma_X \sigma_Y$. 

---

\(^2\)In fact, $r_{XY}$ can be viewed as an inner product on the space of random variables. This creates a geometry isomorphic to the standard Euclidean metric in $\mathbb{R}^n$. Using this construction, the Cauchy-Schwarz inequality immediately follows: $|r_{XY}| \leq \sigma_X \sigma_Y$. 

5
going to sample the signal at a set of points \( \{ \mathbf{x}_1, \ldots, \mathbf{x}_n \} \subset D \) and describe the distribution of the random vector \( \mathbf{F} = (\mathcal{F}(\mathbf{x}_1), \ldots, \mathcal{F}(\mathbf{x}_n)) \). We will define the finite-dimensional CDF as
\[
F_{\mathbf{x}_1, \ldots, \mathbf{x}_n}(f_1, \ldots, f_n) = P(\mathcal{F}(\mathbf{x}_1) \leq f_1, \ldots, \mathcal{F}(\mathbf{x}_n) \leq f_n).
\]

**Stationarity** A random signal is said (strict sense) stationary (SSS), if its probability distribution is shift-invariant. In other words, if for any \( n \), any \( \{ \mathbf{x}_1, \ldots, \mathbf{x}_n \} \subset D \), and any \( \mathbf{p} \in D \), \((\mathcal{F}(\mathbf{x}_1 + \mathbf{p}), \ldots, \mathcal{F}(\mathbf{x}_n + \mathbf{p}))\) has the same distribution as \((\mathcal{F}(\mathbf{x}_1), \ldots, \mathcal{F}(\mathbf{x}_n))\), then \( \mathcal{F} \) is SSS.

**Auto-correlation and cross-correlation** Moments of random signals can be defined by considering random vectors obtained from sampling the signal at some set of locations. Of particular importance will be the following first- and second-order moments: the mean function
\[
\mu_{\mathcal{F}}(\mathbf{x}) = \mathbb{E}\mathcal{F}(\mathbf{x})
\]
and the auto-correlation function
\[
R_{\mathcal{F}}(\mathbf{x}_1, \mathbf{x}_2) = \mathbb{E}\mathcal{F}(\mathbf{x}_1)\mathcal{F}(\mathbf{x}_2).
\]

Given two random signals \( \mathcal{F} \) and \( \mathcal{G} \), we can similarly define their cross-correlation as
\[
R_{\mathcal{F}\mathcal{G}}(\mathbf{x}_1, \mathbf{x}_2) = \mathbb{E}\mathcal{F}(\mathbf{x}_1)\mathcal{G}(\mathbf{x}_2).
\]

It follows from definition that \( R_{\mathcal{F}}(\mathbf{x}_1, \mathbf{x}_2) = R_{\mathcal{F}}(\mathbf{x}_2, \mathbf{x}_1) \) and \( R_{\mathcal{G}\mathcal{F}}(\mathbf{x}_1, \mathbf{x}_2) = R_{\mathcal{G}\mathcal{F}}(\mathbf{x}_2, \mathbf{x}_1) \). Two random signals are said to be orthogonal if their covariance function vanishes at every point.

**Wide-sense stationarity** A random signal \( \mathcal{F} \) is called wide-sense stationary (WSS) if its mean and auto-correlation functions are shift-invariant, i.e., for every \( \mathbf{p} \in D \) \( \mu_{\mathcal{F}}(\mathbf{x} + \mathbf{p}) = \mu_{\mathcal{F}}(\mathbf{x}) \) for every \( \mathbf{x} \in D \), and \( R_{\mathcal{F}}(\mathbf{x}_1 + \mathbf{p}, \mathbf{x}_2 + \mathbf{p}) = R_{\mathcal{F}}(\mathbf{x}_1, \mathbf{x}_2) \) for every \( \mathbf{x}_1, \mathbf{x}_2 \in D \). These conditions immediately translate to demanding \( \mu_{\mathcal{F}} = \text{const} \) and \( R_{\mathcal{F}}(\mathbf{x}_1, \mathbf{x}_2) = R_{\mathcal{F}}(\mathbf{x}_1 - \mathbf{x}_2) \). Two WSS random signals \( \mathcal{F} \) and \( \mathcal{G} \) are called jointly WSS if their cross-correlation is shift-invariant, i.e., \( R_{\mathcal{F}\mathcal{G}}(\mathbf{x}_1, \mathbf{x}_2) = R_{\mathcal{G}\mathcal{F}}(\mathbf{x}_1 - \mathbf{x}_2) \).

**Power spectrum density** We start our discussion with the more familiar domain of deterministic signals. In the signal processing jargon, the energy of a deterministic signal \( f \) is defined as
\[
E = \| f \|_{L^2(\mathbb{R}^d)}^2 = \int_{\mathbb{R}^d} |f(\mathbf{x})|^2 d\mathbf{x}.
\]

Due to Parseval’s identity,
\[
E = \| \mathcal{F} f \|_{L^2(\mathbb{R}^d)}^2 = \int_{\mathbb{R}^d} |\mathcal{F}(\boldsymbol{\xi})|^2 d\boldsymbol{\xi}.
\]
(the same is true mutatis mutandis for discrete-domain signals). We can therefore think of $|F(\xi)|^2$ as of the energy density of $f$ per unit of frequency.

When the signal has infinite energy, we can still define its average power by windowing the signal and normalizing its energy within the window by the volume of the latter,

$$W = \lim_{T \to \infty} \frac{1}{T^d} \int_{[-\frac{T}{2}, \frac{T}{2}]^d} |f(x)|^2 dx = \lim_{T \to \infty} \left\| \frac{1}{T^\frac{d}{2}} \text{rect}_T \cdot f \right\|_{L^2(\mathbb{R}^d)}^2$$

where $\text{rect}_T(x) = \text{rect}(x/T)$. Defining the windowed Fourier transform

$$F_T(\xi) = \frac{1}{T^\frac{d}{2}} \mathcal{F}(\text{rect}_T \cdot f)$$

and invoking Parseval’s identity, we have

$$W = \lim_{T \to \infty} \|F_T(\xi)\|_{L^2(\mathbb{R}^d)}^2 = \lim_{T \to \infty} \int_{\mathbb{R}^d} |F_T(\xi)|^2 d\xi.$$ 

$\lim_{T \to \infty} |F_T(\xi)|^2$ can be interpreted as the density of power per unit of frequency and is often referred to as the power spectral density (PSD).

The same reasoning can be repeated for WSS random processes. While a random process $\mathcal{F}$ has infinite energy, it generally has finite average power and one can define the PSD as

$$S_{\mathcal{F}}(\xi) = \lim_{T \to \infty} \frac{1}{T^d} \mathbb{E} \left| \mathcal{F}(\text{rect}_T \cdot \mathcal{F}) \right|^2$$

$$= \lim_{T \to \infty} \frac{1}{T^d} \mathbb{E} \left( \int_{[-\frac{T}{2}, \frac{T}{2}]^d} f(x)e^{2\pi i \xi^T x} dx \int_{[-\frac{T}{2}, \frac{T}{2}]^d} f(x')e^{-2\pi i \xi^T x'} dx' \right)$$

$$= \lim_{T \to \infty} \frac{1}{T^d} \int_{[-\frac{T}{2}, \frac{T}{2}]^{2d}} \mathbb{E}f(x)f(x') e^{-2\pi i \xi^T (x-x')} dx dx'.$$

Changing the integration variable $x'$ to $y = x - x'$, one obtains the following result known as the Wiener-Khinchin theorem:

$$S_{\mathcal{F}} = \mathcal{F} R_{\mathcal{F}}$$

This is a profound conclusion relating the PSD of a random signal to the Fourier transform of its auto-correlation.

**Cross-spectral density** The result can be generalized to a pair of jointly WSS random signals $\mathcal{F}$ and $\mathcal{G}$. We define the cross-spectral density as

$$S_{\mathcal{F},\mathcal{G}}(\xi) = \lim_{T \to \infty} \frac{1}{T^d} \mathbb{E} \left( \mathcal{F}(\text{rect}_T \cdot \mathcal{F}) \right)^* \mathcal{F}(\text{rect}_T \cdot \mathcal{G})$$.

The Wiener-Kinchin theorem states

$$S_{\mathcal{F},\mathcal{G}} = \mathcal{F} R_{\mathcal{F},\mathcal{G}}$$

where the cross-correlation is defined as $R_{\mathcal{F},\mathcal{G}}(x) = \mathbb{E}\mathcal{F}(0)\mathcal{G}(x)$. 

7
LSI systems  Let $\mathcal{F}$ be a WSS signal passing through a linear shift-invariant system $\mathcal{H}$ with the impulse response $h$. We define the output signal as $\mathcal{G} = h \ast \mathcal{F}$. Straightforwardly, the mean function of $\mathcal{G}$ is given by

$$
\mu_\mathcal{G}(x) = \mathbb{E}\mathcal{G}(x) = \mathbb{E} \int_{\mathbb{R}^d} h(x') \mathcal{F}(x-x') dx' = \int_{\mathbb{R}^d} h(x') \mu_\mathcal{F}(x-x') dx' = \int_{\mathbb{R}^d} h(x') \mu_\mathcal{G}(x-x') dx'
$$

where $H(0)$ is usually called the DC response of $\mathcal{H}$. Note that the mean function is constant.

The auto-correlation function of $\mathcal{G}$ is given by

$$
R_\mathcal{G}(x, x-y) = \mathbb{E}\mathcal{G}(x)\mathcal{G}(x+y) = \mathbb{E} \left( \int_{\mathbb{R}^d} h(x') \mathcal{F}(x-x') dx' \int_{\mathbb{R}^d} h(x'') \mathcal{F}(x-y-x'') dx'' \right) = \int_{\mathbb{R}^d} h(x'') \left( \int_{\mathbb{R}^d} h(x') \mathbb{E}\mathcal{F}(x-x') dx' \right) dx'' = \int_{\mathbb{R}^d} h(x'') (h * R_\mathcal{F})(y+x'') dx'' = \int_{\mathbb{R}^d} \overline{h}(y'') (h * R_\mathcal{F})(y-y'') dy'' = (h * R_\mathcal{F} * \overline{h})(y),
$$

where $\overline{h}(x) = h(-x)$. Note that the auto-correlation function is shift-invariant (i.e., it does not depend on $x$). These two results imply that $\mathcal{G}$ is WSS (not surprising: a signal with shift-invariant moments passing through a shift-invariant system leads to a signal with shift-invariant moments).

The input-output cross-correlation is given by

$$
R_{\mathcal{F}\mathcal{G}}(x, x-y) = \mathbb{E}\mathcal{F}(x)\mathcal{G}(x-y) = \mathbb{E} \left( \mathcal{F}(x) \int_{\mathbb{R}^d} h(x') \mathcal{F}(x-y-x') dx' \right) = \int_{\mathbb{R}^d} h(x') \mathbb{E}\mathcal{F}(x) \mathcal{F}(x-y-x') dx' = \int_{\mathbb{R}^d} h(x') R_{\mathcal{F}}(y+x') dx' = (\overline{h} * R_\mathcal{F})(y).
$$

Again note that the function is shift-invariant, implying that $\mathcal{F}$ and $\mathcal{G}$ are jointly WSS. Clearly, since $R_{\mathcal{G}\mathcal{F}} = R_{\mathcal{F}\mathcal{G}} = h * R_\mathcal{F} = h * R_\mathcal{F}$.

Translating the latter expressions to the frequency domain, we obtain the following rela-
\[ R_g = h \ast R_f \ast \overline{h} \quad \overset{\mathcal{F}}{\longleftrightarrow} \quad S_g = |H|^2 \cdot S_f \]
\[ R_{fg} = \overline{h} \ast R_f \quad \overset{\mathcal{F}}{\longleftrightarrow} \quad S_{fg} = H^* \cdot S_f \]
\[ R_{gH} = h \ast R_f \quad \overset{\mathcal{F}}{\longleftrightarrow} \quad S_{gH} = H \cdot S_f \]

**White and colored noise** A random signal \( N \) with constant PSD, \( S_N(\xi) = \sigma_N^2 \) is usually called *white noise*, by analogy with white light that has approximately flat spectrum\(^3\). Its auto-correlation is given by \( R_N(x) = \sigma_N^2 \delta \). When white noise passes through an LSI system \( \mathcal{H} \), the spectrum at the output, \( S_{\mathcal{H}N}(\xi) = |H(\xi)|^2 \sigma_N^2 \), is shaped by the power response \( |H(\xi)|^2 \) of the system. This phenomenon is called as *coloring*. The auto-correlation function of colored noise is given by \( R_{\mathcal{H}N} = \sigma_N^2 h \ast \delta \ast \overline{h} = \sigma_N^2 h \ast \overline{h} \).

\(^3\)This appears to be false when color perception is examined more closely!