Chapter II
Random Signals

This chapter provides a summary of definitions and fundamental concepts of probability spaces, random variables, and random processes. References [1]—[8] provide more extensive background and examples.

1 Kolmogorov's Axiomatic Definition of Probability

A probability space is a triple $(\Omega, \mathscr{F}, P)$ where Ω is the sample space, $\mathscr{F}$ is a σ-field defined on Ω, and P is a probability measure defined on $\mathscr{F}$.

- The *sample space* Ω represents the set of *outcomes* of a random experiment, e.g., $\Omega = \{\text{Head},\ \text{Tail}\}$ for a coin flip; or $\Omega = \mathbb{R}^+$ for the time to next child birth in Urbana, Illinois.
- The *sigma field* $\mathscr{F}$ is a collection of subsets of Ω (viewed as *events*) such that
  1. $\emptyset \in \mathscr{F}$ and $\Omega \in \mathscr{F}$
  2. If $\mathcal{E}_1, \mathcal{E}_2 \in \mathscr{F}$ then $\mathcal{E}_1 \cup \mathcal{E}_2 \in \mathscr{F}$
  3. If $\mathcal{E} \in \mathscr{F}$ then $\mathcal{E}^c \in \mathscr{F}$
  4. $\mathscr{F}$ is closed under countable set of unions and intersections, i.e.,
  $$\bigcup_{i=1}^{\infty} \mathcal{E}_i \in \mathscr{F} \quad \text{and} \quad \bigcap_{i=1}^{\infty} \mathcal{E}_i \in \mathscr{F}.$$
- The probability measure P assigns a number $P(\mathcal{E})$ to each $\mathcal{E} \in \mathscr{F}$, such that
  1. $P(\mathcal{E}) \geq 0$
  2. $P(\Omega) = 1$
  3. $P(\mathcal{E}_1 \cup \mathcal{E}_2) = P(\mathcal{E}_1) + P(\mathcal{E}_2)$ for disjoint sets: $\mathcal{E}_1 \cap \mathcal{E}_2 = \emptyset$
  4. $P\left(\bigcup_{i=1}^{\infty} \mathcal{E}_i\right) = \sum_{i=1}^{\infty} P(\mathcal{E}_i)$ if $\mathcal{E}_i \cap \mathcal{E}_j = \emptyset$ for all i, j

If Ω is a countable set, $\mathscr{F}$ may be the ensemble of *all* subsets of Ω. However this cannot be the case when Ω is uncountable, e.g., $\Omega = [a, b]$, because it would not be possible to define a probability measure (satisfying the above axioms) on $\mathscr{F}$. When $\Omega = [a, b]$, one may take $\mathscr{F}$ to be the Borel σ-field defined over $[a, b]$. The Borel σ-field is the smallest σ-field for $[a, b]$ and contains all the open subsets of $[a, b]$.

2 Random Variables

Given a probability space $(\Omega, \mathscr{F}, P)$, we may formally define a **real-valued random variable** X as a measurable function from Ω to $\mathbb{R}$, i.e.,

$$\{\omega \ : \ X(\omega) \leq x\} \in \mathscr{F}, \quad \forall x \in \mathbb{R}.$$

The cumulative distribution function (cdf) for X may then be defined as

$$F_X(x) \triangleq \Pr[X \leq x] = P(\{\omega \ : \ X(\omega) \leq x\}).$$

Observe that for every Borel set $\mathscr{B}$ in $\mathbb{R}$, the set $\{\omega \ : \ X(\omega) \in \mathscr{B}\}$ must correspond to an event $\mathcal{E} \in \mathscr{F}$, i.e., it must be in the domain of the probability function $P(\cdot)$. All sets of interest are unions or intersections of such sets.

A cdf is nondecreasing, right-continuous, and satisfies

$$\lim_{x\to-\infty} F_X(x) = 0 \quad \text{and} \quad \lim_{x\to\infty} F_X(x) = 1.$$

Two types of random variables are often encountered in practice:

- Discrete random variables, for which the probability of X is concentrated at a countable set of points $\{x_i,\, i \in \mathcal{I}\}$:
$$P(X \in \{x_i,\, i \in \mathcal{I}\}) = 1.$$
The cdf for discrete random variables is piecewise constant (staircase function). There is a jump at each point x_i; the amplitude of this jump is given by the probability mass function (pmf)
$$p_X(x_i) = P(X = x_i), \quad i \in \mathcal{I}.$$
The pmf is nonnegative and sums to 1.

- Continuous random variables, for which the cdf is the integral of a function:
$$F_X(x) = \int_{-\infty}^{x} f_X(u) du$$
The function f_X is nonnegative and integrates to one; f_X is the probability density function (pdf) for X.

Other random variables are of a *mixed discrete/continuous* type, e.g., consider $X = 0$ with probability $\frac{1}{2}$ and X is uniformly distributed over the interval $[1, 2]$ with probability $\frac{1}{2}$.

The mathematical expectation of a random variable X is defined as

$$\mathbb{E}X \triangleq \int_{\mathbb{R}} x\, dF_X(x)$$

which may also be written as

$$\mathbb{E}X = \int_{\Omega} \omega\, dP(\omega).$$

These expressions simplify to the standard formulas $\mathbb{E}X = \sum_{i\in\mathcal{I}} x_i\, p_X(x_i)$ and $\mathbb{E}X = \int x\, f_X(x)\, dx$ for discrete and continuous random variables, respectively.

3 Convergence of Random Sequences

Let X_n, $1 \leq n < \infty$, be a sequence of random variables indexed by the integer n. We are interested in the convergence property $X_n \to X$. For a deterministic sequence, we would use the classical definition: $X_n \to X$ means that

$$\forall \epsilon > 0, \exists n(\epsilon) \ : \ |X_n - X| < \epsilon \quad \forall n > n(\epsilon).$$

For random sequences, the definition must involve the elements of Ω.

3.1 Types of Convergence

We shall encounter the following types of convergence:

1. $X_n \to X$ surely means
$$\lim_{n\to\infty} X_n(\omega) = X \quad \forall \omega \in \Omega.$$
2. $X_n \to X$ almost surely (with probability 1) means
$$P\left(\lim_{n\to\infty} X_n(\omega) = X\right) = 1.$$
3. $X_n \to X$ in probability means
$$\lim_{n\to\infty} P(|X_n(\omega) - X| < \epsilon) = 1.$$
4. $X_n \to X$ in the mean square (m.s.) means
$$\lim_{n\to\infty} \mathbb{E}|X_n(\omega) - X|^2 = 0.$$
5. $X_n \to X$ surely means
$$\lim_{n\to\infty} F_n(x) = F(x) \quad \text{(pointwise on } \mathbb{R})$$
where F_n and F are the distributions for X_n and X, respectively.

Convergence a.s. implies convergence surely.
Convergence in probability implies convergence almost surely.
Convergence in distribution implies convergence in probability and convergence m.s.

3.2 Law of Large Numbers

Let $X_i, 1 \le i < \infty$ be i.i.d. random variables with mean μ and finite variance. Then the sample average

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

converges in probability to μ (*weak law of large numbers*); as well, $\overline{X}_n$ converges a.s. to μ (*strong law of large numbers*).

3.3 Central Limit Theorem

Let $X_i, 1 \le i < \infty$ be i.i.d. random variables with mean μ and variance σ^2. Then the normalized sum

$$Z_n = \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n}(X_i - \mu)$$

converges to a normal distribution $\mathcal{N}(0,1)$ in distribution.

A generalization of this problem is the case of variable components. Let $X_i, 1 \le i < \infty$ be independent random variables with common mean μ but unequal variances $\sigma_i^2, 1 \le i < \infty$ [2, p. 213] [7, p. 518]. Define the sum of variances

$$S_n^2 = \sum_{i=1}^{n} \sigma_i^2.$$

Then the normalized sum

$$Z_n = \frac{1}{S_n}\sum_{i=1}^{n}(X_i - \mu)$$

converges to a normal distribution $\mathcal{N}(0,1)$ in distribution if Lindeberg's conditions are satisfied:

$$\forall \epsilon > 0, \exists n(\epsilon) \ : \ \frac{\sigma_i}{S_n} < \epsilon \quad \forall n > n(\epsilon),\, i = 1, \cdots, n.$$

We may interpret the ratio σ_i/S_n as the contribution of component i to the weighted sum Z_n. For the CLT to apply, Z_n must be sum of many asymptotically negligible components. One component is not allowed to dominate the sum.

4 Random Processes

Definition: A random process is a collection of random variables $\{X(t), t \in \mathcal{T}\}$, where $\mathcal{T}$ is the index set.

Usually $\mathcal{T}$ is not a finite set. We shall distinguish between the case where $\mathcal{T}$ is a *countable* index set, e.g., $\{0,1\}$, $\mathbb{Z}$, and $\mathbb{Z}^d$, or an uncountable set, such as an interval on the real line.

One difficulty that arises when $|\mathcal{T}| = \infty$ is that the notion of pdf breaks down. Moreover, the definition of infinite-dimensional cdf's requires care.

Let us start with finite-dimensional cdf's. The definition of one-dimensional cdf's,

$$F_t(x) = P[X(t) \le x]$$

is naturally extended to two dimensions:

$$F_{t_1,t_2}(x_1,x_2) = P[X(t_1) \le x_1, X(t_2) \le x_2]$$

and to any finite number n of dimensions:

$$F_{t_1,\cdots,t_n}(x_1,\cdots,x_n) = P[X(t_1) \le x_1,\cdots, X(t_n) \le x_n].$$

This defines a hierarchy of multidimensional cdf's of all orders. They must satisfy the compatibility conditions

$$\lim_{x_2 \to \infty} F_{t_1,t_2}(x_1,x_2) = F_{t_1}(x_1), \quad \text{etc.}$$

In practice, fortunately, there exist simple mechanisms for generating probability distributions.

4.1 Kolmogorov's Extension Theorem

Having defined finite-dimensional distributions, we need extend this concept to the whole index set $\mathcal{T}$. Formally, we need to specify a probability space $(\Omega^{\mathcal{T}}, \mathscr{F}^{\mathcal{T}}, P^{\mathcal{T}})$ that is compatible with all finite-dimensional probability spaces $(\Omega^n, \mathscr{F}^n, P^n)$. If $\mathcal{T}$ is countable (say $\mathbb{N}$), $\mathscr{F}^{\mathcal{T}}$ is the smallest σ-field of subsets of $\mathbb{R}^{\mathcal{T}}$ containing all finite-dimensional rectangles [4, p. 21].

The existence of such a probability space $(\Omega^{\mathcal{T}}, \mathscr{F}^{\mathcal{T}}, P^{\mathcal{T}})$ is guaranteed by Kolmogorov's Extension Theorem [5, p. 16] [4, p. 24]. Furthermore, this extension is unique when $\mathcal{T}$ is countable.

Examples (with $\mathcal{T} = \mathbb{Z}^2$)

- $x(n_1,n_2)$ = i.i.d. random variables (spatially white noise)
- $x(n_1,n_2) \equiv A$ where A is a random variable
- $x(n_1,n_2) \in \{0,1\}$: binary random field

4.2 Uncountable Index Set

When the index $\mathcal{T}$ is countable, we may conceptually obtain the entire sample path of the random process by observing its samples for all values of t (we can "count" these samples). However, technical difficulties arise when $\mathcal{T}$ is uncountably infinite, e.g., $\mathcal{T} = [a, b]$. Intuitively, the collection of random variables $\{X(t), t \in \mathcal{T}\}$ may be "too large" and cannot be recovered by sampling. Such is the case, for instance, when $\mathcal{T} = \mathbb{R}$ and $X(t), t \in \mathcal{T}$, are i.i.d. samples.

A simple example is the process $X(t), t \in [0, 1]$, that is generated as follows: pick a random variable t_0 from the uniform distribution on $[0, 1]$ and let $X(t) = 0$ at all points except at $t = t_0$, where $X(t) = 1$. This process is *statistically indistinguishable* from the trivial process $\tilde{X}(t) \equiv 0$ in the following sense. The n-dimensional cdf's for the processes X and $\tilde{X}$ are identical for all values of n. Heuristically speaking, we may collect an arbitrary large number of samples of X, yet with probability 1 they will all have value 0, i.e., coincide with those for the process $\tilde{X}$.

In measure-theoretic terms, when $\mathcal{T}$ is uncountably infinite, Kolmogorov's extension is not necessarily unique. This difficulty does not arise when the random process is "continuous enough" that specifying all finite-order cdf's completely characterizes the distribution of the process. Such processes can be "sampled" and are said to be *separable*. Fortunately, all random processes of interest to us are in the category. The example given above describes a processes X that is not separable because not "continuous enough".

4.3 Moments of a Random Process

The moments of a real-valued random process include

- the mean $\mu_X(t), t \in \mathcal{T}$, where $\mu_X(t) \triangleq \mathbb{E}X(t)$;
- the correlation function $R_X(t_1, t_2), t_1, t_2 \in \mathcal{T}$, where $R_X(t_1, t_2) = \mathbb{E}[X(t_1)X(t_2)]$;
- the covariance function $C_X(t_1, t_2) = R_X(t_1, t_2) - \mu_X(t_1)\mu_X(t_2)$, which is the correlation function for the centered process $X(t)\mu_X(t)$;
- the n-th order moment $\mathbb{E}[X(t_1) \cdots X(t_n)]$.

The covariance function is symmetric and positive semidefinite, i.e., the $n \times n$ matrix $\{R_X(t_i, t_j), 1 \leq i, j \leq n\}$ is positive semidefinite for all choices of n-tuples $\{t_i\}$. A process is said to be a *second-order process* if $\mathbb{E}X^2(t) = R_X(t, t) < \infty$ for all t.

4.4 Gaussian Random Process

Define a mean function $\mu_X(t), t \in \mathcal{T}$ and a continuous covariance function $C_X(t_1, t_2)$. Let $F_{t_1,\cdots,t_n}(x_1, \cdots, x_n)$ be the classical n-dimensional distribution with mean vector $\{\mu_X(t_1), \cdots, \mu_X(t_n)\}$ and covariance matrix $\{C_X(t_i, t_j), 1 \leq i, j \leq n\}$. This defines a hierarchy of multidimensional distributions, and the consistency conditions are automatically satisfied. Note the simplicity of this construction.

5 Stationary Processes

A stationary process is a random process whose statistical properties do not change over time/space. Formally, the process $X(t), t \in \mathcal{T}$, is said to be stationary if

$$F_{t_1+\tau,\cdots,t_n+\tau}(x_1,\cdots,x_n) = F_{t_1,\cdots,t_n}(x_1,\cdots,x_n) \quad \forall\{t_i,x_i\} \in \mathcal{T}^n \times \mathbb{R}^n,\ \forall\tau \in \mathcal{T},\ \forall n.$$

The index set might be discrete or continuous time ($\mathcal{T} = \mathbb{R}$ or $\mathbb{Z}$, respectively) or d-dimensional space ($\mathcal{T} = \mathbb{R}^d$ or $\mathbb{Z}^d$).

A wide-sense stationary (WSS) process satisfies a similar invariance properties in terms of its first two moments:

$$\begin{aligned} \mu_X(t) &= \mu_X, \quad \forall t \in \mathcal{T} \\ R_X(t_1,t_2) &= R_X(t_1-t_2), \quad \forall t_1,t_2 \in \mathcal{T}. \end{aligned}$$

Stationarity implies wide-sense stationarity, but the converse is not true. A WSS process is also called weakly stationary.

When the index set for a WSS random process X is $\mathbb{R}^d$ or $\mathbb{Z}^d$ where $d \geq 1$, we define the *spectral density* of X as

$$S_X(f) = \mathcal{F}[C_X](f)$$

where f is a d-dimensional frequency vector, and $\mathcal{F}$ denotes the appropriate d-dimensional Fourier transform. When $\mathcal{T} = \mathbb{R}^d$, we have the Fourier transform pair

$$\begin{aligned} S_X(f) &= \int_{\mathbb{R}^d} C_X(t) e^{-j2\pi f\cdot t}\, dt, \quad f \in \mathbb{R}^d \\ C_X(t) &= \int_{\mathbb{R}^d} S_X(f) e^{j2\pi f\cdot t}\, df, \quad t \in \mathbb{R}^d \end{aligned}$$

where $f \cdot t$ denotes the dot product of the respective d-vectors. When $\mathcal{T} = \mathbb{Z}^d$, we have

$$\begin{aligned} S_X(f) &= \sum_{i\in\mathbb{Z}^d} C_X(i) e^{-j2\pi f\cdot i}, \quad f \in \left[-\frac{1}{2}, \frac{1}{2}\right]^d \\ C_X(t) &= \int_{\left[-\frac{1}{2},\frac{1}{2}\right]^d} S_X(f) e^{j2\pi f\cdot t}\, df, \quad t \in \mathbb{Z}^d. \end{aligned}$$

The (d-dimensional) spectral density represents the average distribution of energy across frequency. The spectral density is nonnegative (as proven by Herglotz (1911) for $\mathcal{T} = \mathbb{Z}$ and by Bochner (1932) for $\mathcal{T} = \mathbb{R}$). Existence of the spectral density is guaranteed if $\int |C_X| < \infty$ (in the case $\mathcal{T} = \mathbb{R}^d$) or $\sum_i |C_X(i)| < \infty$ (in the case $\mathcal{T} = \mathbb{Z}^d$).

6 Isotropic Processes

Let $\mathcal{T} = \mathbb{R}^d$ or $\mathbb{Z}^d$. A d-dimensional isotropic process is a stationary process whose correlation function $R_X(t)$ is spherically symmetric:

$$R_X(t) = R_X(\|t\|)$$

where $\|\cdot\|$ denotes Euclidean norm. By the elementary properties of the Fourier transform, isotropy implies that the spectral density is also spherically symmetric:

$$S_X(f) = S_X(\|f\|).$$

Isotropic models are frequently used in image processing.

7 Wide-Sense Periodic Processes

A d-dimensional random process $X(t), t \in \mathcal{T}$ is wide-sense periodic if its mean function $\mu_X(t)$ and its correlation function $C_X(t_1, t_2)$ satisfy the following properties:

1. $\mu_X(t) = \mu_X(t + T_i)$ for $i = 1, \cdots, d$, where $T_1, \cdots, T_d$ are d *periodicity vectors*.
2. $C_X(t_1, t_2) = C_X(t_1 + T_i, t_2) = C_X(t_1, t_2 + T_i)$ for $i = 1, \cdots, d$.

For a wide-sense periodic process, it may be shown that

$$\mathbb{E}|X(t + T_i) - X(t)|^2 = 0 \quad \forall t \in \mathcal{T},\ i = 1, \cdots, d,$$

i.e., the realizations of this process are "mean-square periodic".

8 Continuity of Random Processes

Let $X(t), t \in \mathbb{R}^d$ be a second-order random process. The process is said to be mean-square (m.s.) continuous at t if

$$\lim_{s \to t} X(s) = X(t) \text{ m.s.}$$

i.e., $\mathbb{E}|X(t) - X(s)|^2 \overset{\text{m.s.}}{\to} 0$ as $s \to t$. One may similarly define processes that are continuous in probability, or in distribution, by replacing the m.s. limit by the appropriate limit.

None of this implies that the sample paths of the process X are continuous. For instance, define the following process: pick two independent random variables A and t_0 according to a normal distribution, and let

$$X(t) = \begin{cases} A & : \ t < t_0 \\ A + 1 & : \ t \geq t_0. \end{cases}$$

Any sample path of this process has one point of discontinuity (at the random location t_0). Yet it can be verified that X is m.s. continuous everywhere.

The notion of a.s. continuity of a random process is quite different from the three notions discussed above. A random process is a.s. continuous if the probability that a sample path is continuous is 1. The process X in our example above fails that stronger condition, because the probability that a sample path is continuous is 0.

The notion of m.s. continuity is useful in part because that property can be inferred by inspection of the correlation function. Indeed the following three properties are equivalent:

1. X is m.s. continuous
2. $R_X(t,s)$ is continuous over $\mathcal{T} \times \mathcal{T}$
3. $R_X(t,t)$ is continuous at all $t \in \mathcal{T}$.

For a WSS process, the correlation function is of the form $R_X(t)$, $t \in \mathcal{T}$, and the following three properties are equivalent:

1. X is m.s. continuous
2. $R_X(t)$ is continuous over $\mathcal{T}$
3. $R_X(t)$ is continuous at $t = 0$.

9 Stochastic Integrals

We often deal with random processes that are outputs of linear systems, e.g.,

$$Y(t) = \int_{\mathbb{R}^d} X(s)h(t-s)\,ds, \quad t \in \mathbb{R}^d.$$

Since X depends on the outcome ω of the random experiment, this integral may not always exist. What meaning can we attach to such an integral?

Let us consider the following generic problem. Denote by I the integral $\int_\Lambda X(t)\,dt$, where $\Lambda \subset \mathcal{T}$. Define a set of intervals $\{\Delta_i,\, 1 \le i \le n\}$ that form a partition of Λ, i.e.,

$$\bigcup_{i=1}^{\infty} \Delta_i = \Lambda, \quad \Delta_i \cap \Delta_j = \emptyset \quad \forall i \ne j.$$

Assume that $\max_i |\Delta_i| \to 0$ as $n \to \infty$. Choose some t_i in each interval Δ_i and define the sum

$$I_n = \sum_{i=1}^{n} X(t_i)|\Delta_i|.$$

For a classical (Riemann) integral, we have $I \triangleq \lim_{n\to\infty} I_n$.

For a stochastic integral, the *mean-square integral* I exists when

$$\lim_{n\to\infty} \mathbb{E}(I - I_n)^2 = 0, \quad \text{i.e., } I_n \overset{\text{m.s.}}{\to} I.$$

Properties of the mean-square integral.

- Existence: guaranteed if $\int_\Lambda \int_\Lambda R_X(t,s)\,dt\,ds < \infty$;
- Linearity: $\int [aX(t) + bY(t)]\,dt = a\int X(t)\,dt + b\int Y(t)\,dt$ provided that $\int X$ and $\int Y$ exist;
- Moments: $\mathbb{E}[\int X(t)\,dt] = \int \mathbb{E}[X(t)]\,dt$ and $\mathbb{E}(\int X(t)\,dt)^2 = \int\int R_X(t,s)\,dt\,ds$.

LSI Filtering. Consider a Linear Shift Invariant (LSI) system with impulse response $h(t)$, and denote by $R_h(u) \triangleq \int h(t)h(t+u)\,dt$ the *deterministic autocorrelation function* of h. If the process $X(t)$, $t \in \mathbb{R}^d$ is input to this sytem, the output is

$$\begin{aligned}
Y(t) &= \int_{\mathbb{R}^d} X(s)h(t-s)\,ds, \quad t \in \mathbb{R}^d \\
\mu_Y(t) &= \mathbb{E}Y(t) = \int \mathbb{E}X(s)h(t-s)\,ds = \int \mu_X(s)h(t-s)\,ds \\
R_Y(t,s) &= \mathbb{E}[Y(t)Y(s)] \\
&= \mathbb{E}\left[\int X(t')h(t-t')\,dt' \int X(s')h(s-s')\,ds'\right] \\
&= \int\int \mathbb{E}[X(t')X(s')]h(t-t')h(s-s')\,dt'\,ds' \\
&= \int\int R_X(t',s')h(t-t')h(s-s')\,dt'\,ds'.
\end{aligned}$$

If the process X is WSS, these expressions simplify to

$$\begin{aligned}
\mu_Y(t) &= \mu_X \int h(t)\,dt \\
R_Y(t-s) &= (R_X \star R_h)(t-s).
\end{aligned}$$

Special Case: Filtering *white noise* $X(t)$ through a LSI system.
Taking $R_X(t) = \delta(t)$, we obtain

$$R_Y(t) = (R_X \star R_h)(t) = R_h(t).$$

The stochastic m.s. integral $Y(t) = \int X(s)h(t-s)\,ds$ exists if $R_h(0) = \int |h(t)|^2\,dt < \infty$. Compare with the traditional *Bounded Input Bounded Output* (BIBO) condition for stability of LSI systems: $\int |h(t)|\,dt < \infty$. Which condition is stronger?

Note that “white noise” on $\mathbb{R}$ is not a separable process, but the filtered noise is a separable process. We could have been more careful in the derivation above and selected a separable process X_ϵ as the input, where X_ϵ becomes “white noise” as $\epsilon \to 0$. Specifically, choose some correlation function $R(t)$, continuous at $t = 0$ and integrating to 1, and define $R(t;\epsilon) \triangleq \epsilon^{-1}R(\epsilon^{-1}t)$ which forms a resolution of the identity:

$$\int f(t)\delta(t)\,dt \triangleq \lim_{\epsilon\to 0}\int f(t)R(t;\epsilon)\,dt = f(0)$$

for any function f that is continuous at $t = 0$. Then if a WSS process X_ϵ with correlation function $R(t;\epsilon)$ is input to the LSI system, the correlation function of the output is given by $R_h(t)$ in the limit as $\epsilon \to 0$.

10 Ergodic Processes

Let $\mathcal{T} = \mathbb{R}^d$ or $\mathbb{Z}^d$. In practice a statistical description of images (or any other random process) is not available, so these quantities have to be estimated from the data. Consider for instance the problem of estimating the mean μ of a stationary random process given data $X(t), t \in \Lambda$. The mean can be estimated using the stochastic integral

$$\hat{\mu} = \frac{1}{|\Lambda|}\int_\Lambda X(t)\,dt.$$

In many cases, $\hat{\mu} \to \mu$ as $|\Lambda| \to \infty$. (Example: spatially white noise) In this case, the process $X(t)$ is said to be ergodic in the mean. Formally:

Definition: A WSS process $X(t), t \in \mathcal{T}$, is (m.s.) ergodic in the mean iff

$$\hat{\mu} \overset{\text{m.s.}}{\to} \mu \quad \text{as} \quad |\Lambda| \to \infty.$$

What does it take to have ergodicity?

Intuitively, $\hat{\mu}$ should be an average of many *independent* observations (so that a law of large numbers type result applies). This means that $X(t)$ should decorrelate rapidly, i.e., the “correlation area” should be small enough.

Example #1 (Worst-case scenario): $X(t) = A$ (random variable) for all $t \in \mathcal{T}$. Then $\hat{\mu} = A$ no matter how large the observation window Λ is, and $\hat{\mu}$ does not converge to μ.

Example #2 $X(t) = A\cos(2\pi t)$ for all $t \in \mathcal{T}$. For a symmetric window $\Lambda = [-n-\tau, n+\tau]$ where $n \in \mathbb{N}$ and $\tau \in [0, 1)$, we obtain $\hat{\mu} = \frac{2A\sin(2\pi\tau)}{|\Lambda|}$ which vanishes as $|\Lambda| \to \infty$.

The expected value of $\hat{\mu}$ is given by

$$\mathbb{E}\hat{\mu} = \frac{1}{|\Lambda|}\int_\Lambda \mathbb{E}X(t)\,dt = \mu,$$

i.e., $\hat{\mu}$ is an unbiased estimator of μ. The variance of $\hat{\mu}$ is given by

$$\mathbb{E}(\hat{\mu} - \mu)^2 = \frac{1}{|\Lambda|^2} \int_\Lambda \int_\Lambda \mathbb{E} C_X(t - t')\, dt\, dt'.$$

If this variance vanishes as $|\Lambda| \to \infty$, we say that the estimator $\hat{\mu}$ is m.s. consistent.

Slutskys Ergodic Theorem [Yaglom, p. 218]: X is ergodic in the mean iff

$$\frac{1}{|\Lambda|} \int_\Lambda C_X(t)\, dt \to 0 \quad \text{as } |\Lambda| \to \infty$$

The condition above is certainly satisfied if $C_X(t) \to 0$ as $|t| \to \infty$ or even if $C_X(t)$ contains sinusoidal components (as in the case of periodic processes such as X in Example #2 above). But $C_X(t)$ should not contain a d.c. component, which implies that $S_X(f)$ may not contain an impulse at $f = 0$! (as in Example #1 above).

Other types of ergodicity can be defined similarly. For instance:

Definition: A WSS process X is ergodic in correlation at the shift s if

$$\hat{R}_X(s) \to R_X(s) \quad \text{m.s. as} \quad |\Lambda| \to \infty$$

where

$$\hat{R}_X(s) = \frac{1}{|\Lambda|} \int_\Lambda X(t + s) X(t)\, dt.$$

Definition: A WSS process X is ergodic in correlation if the condition above holds for all shifts.

Analysis of ergodicity conditions is similar to analysis for ergodicity in the mean, where the variance of $\hat{R}_X(s)$ should tend to zero as $|\Lambda| \to 0$.

For Gaussian processes we have the following result [5, p. 136].

Theorem (Maruyama 1949). Let X be a real, stationary, Gaussian process with continuous covariance function. Then X is ergodic $\Leftrightarrow$ $S_X(f)$ is continuous.

Examples of ergodic processes

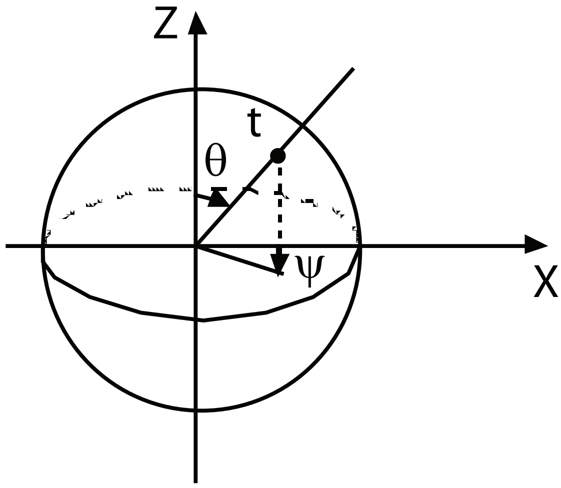
1. $\{X(t),\, t \in \mathbb{Z}^2\}$ = i.i.d. random variables (spatially white noise)
2. $Y(t) = (X \star h)(t)$ where X is as above, and h is FIR.
3. $Y(t) = (X \star h)(t)$ where X is as above, and $\|h\| < \infty$ (possibly IIR filter)

11 Random Processes on the Sphere

[3, Ch. 22.5] [5, p. 129] [Grenander, 1963]

Let $\mathcal{T}$ be the unit sphere in $\mathbb{R}^3$. A random process $X(t), t \in \mathcal{T}$, may be thought of as an image defined on the sphere. This type of imagery is encountered in 3-D imaging (e.g., electric surface fields and imaging of planets. Moreover, 3-D scenes are often represented using spherical projections.

The figure below depicts the unit sphere in $\mathbb{R}^3$. A point t on the sphere may be represented using two angles: the latitude $\theta \in [0, \pi]$ and the longitude $\psi \in [0, 2\pi]$.



For 2-D images, the notion of stationarity implies that images are defined over infinite planar domains, but that assumption is clearly artificial. Modeling images as wide-sense periodic processes, as discussed earlier, is a possible solution. The representation of images as random fields defined on the sphere is a useful alternative because one may view such images as projections of a physical 3-D image.

Now, how can we define stationary and isotropic processes on the sphere?

On the sphere, stationarity and isotropicity mean invariance to rotations:

- $\mathbb{E}X(t) = \mu_X$ is independent of t;
- $\mathbb{E}[X(t)X(s)] = R_X(t, s) = R_X(\phi)$ where ϕ is the *spherical angle* (a distance) between t and s.

The general form of a nonnegative definite kernel $R_X(t, s)$ that depends on t and s only via ϕ

is as follows [Grenander p. 130]:

$$\begin{aligned}R_X(\phi) &= \sum_{m=0}^{\infty} S_X(m) P_m(\cos\phi) \\ S_X(m) &= \int_{\text{sphere}} R_X(\phi) P_m(\cos\phi)\, dt\end{aligned}$$

where $S_X(m) \geq 0$, and $P_m(\cdot)$ are Legendre polynomials. These expressions are reminiscent of the Fourier transform formulas, except that the basis functions are not complex exponentials but Legendre polynomials.

12 Random Processes on Arbitrary Algebraic Structures

We have studied random processes whose correlation function $R_X(t,s)$ is invariant to shifts (in $\mathbb{Z}^d$ and $\mathbb{R}^d$) and rotations (on the sphere). Can this be further generalized?

The answer is positive if we consider an index set $\mathcal{T}$ with group structure [6]. One can then introduce the concept of invariance to the group operation (e.g., translation or rotation) and define a spectral density which is a Fourier transform defined on the group. The "basis functions" for such Fourier transforms are *group characters* $\gamma(t)$ which satisfy the property $\gamma(t_1 + t_2) = \gamma(t_1) + \gamma(t_2)$.

References

[1] B. Hajek, *Notes for ECE534: An Exploration of Random Processes for Engineers*, available from `http://www.ifp.uiuc.edu/~hajek/Papers/randomprocJuly06.pdf`, 2006.

[2] H. Stark and J. W. Woods, *Probability, Random Processes, and Estimation Theory for Engineers*, Prentice-Hall, Upper Saddle River, NJ, 1994.

[3] A. M. Yaglom, *Correlation Theory of Stationary and Related Random Functions I: Basic Results*, Springer-Verlag, 1987.

[4] L. Breiman, *Probability*, SIAM, 1992.

[5] U. Grenander, *Abstract Inference*, Wiley, New York, 1981.

[6] U. Grenander, *Probabilities on Algebraic Structures*, Wiley, New York, 1963.

[7] W. Feller, *An Introduction to Probability Theory and Its Applications*, Wiley, New York, 1971.

[8] J. Doob, *Stochastic Processes*, Wiley, 1953.