Chapter 1

Introduction

Generally speaking, signal *detection* and *estimation* is the area of study that deals with information processing: conversion, transmission, observation and information extraction. The main area of applications of detection and estimation theory are radar, sonar, analog or digital communications, but detection and estimation theory becomes also the main tool in other area such as radioastronomy, geophysics, medical imaging, biological data processing, etc.

In general, detection and estimation applications involve making *inferences* from observations that are distorted or corrupted in some unknown way or too complicated to be modelled in a deterministic way. Moreover, sometimes even the information that one wishes to extract from such observations is not well determined. Thus, it is very useful to cast detection and estimation problems in a *probabilistic framework* and *statistical inference*. But using the probability theory and the statistical inference tools does not forcibly means that the corresponding physical phenomena are necessarily random.

In statistical inference, the goal is not to make an immediate decision, but is instead to provide a summary of the statistical evidence which the future users can easily incorporate into their decision process. The task of decision making is then given to the *decision theory*.

Signal detection is inherently a decision making task. In signal estimation also we need often to make decisions. So, for detection and estimation we need not only the probability theory and statistical inference tools but also the decision and hypothesis testing tools. The main common tool with which we have to start is then the probabilistic and stochastic description of the observations and the unknown quantities.

Once again a probablistic or stochastic description models the effect of causes whose origin and nature are either unknown or too complex to be described deterministically.

The simplest tool of a probabilistic model for a quantity is a scalar *random variable* X which is fully described by its probability distribution $F(x) = \Pr\{X \leq x\}$. The next simplest model is a *random vector* $\boldsymbol{X} = [X_1, \cdots, X_n]^t$, where $\{X_j\}$ are random variables. A random vector is fully described by its probability distribution $F(\boldsymbol{x}) = \Pr\{\boldsymbol{X} \leq \boldsymbol{x}\}$. The next and the most general stochastic model for a quantity is a *random function* $X(\boldsymbol{r})$, where $\boldsymbol{r}$ is a finite dimensional independent variable and where for every fixed values $\boldsymbol{r} = \boldsymbol{r}_j$, the scalar quantity $X_j = X(\boldsymbol{r}_j)$ is a scalar random variable. For example, when $\boldsymbol{r} = (x, y)$ represents the spatial coordinates in a plane, then $X(x, y)$ is called a *random field* and when $\boldsymbol{r} = t$ represents the time variable, then $X(t)$ is called a *stochastic process*. In the rest of these notes, we consider only this last model.

A stochastic process $X(t)$ is completely described by the probability distribution

$$F(x_1, \cdots, x_n; t_1, \cdots, t_n) = F(\boldsymbol{x}; \boldsymbol{t}) = \Pr\{X(t_j) \leq x_j; j = 1, \cdots, n\}$$

for every n and every time instants $\{t_j\}$. The stochastic process is *discrete-time* if it is described only by its realizations on a countable set $\{t_j\}$ of time instants. Then, time is counted by the indices j, and the stochastic process is fully described by the random vectors $\boldsymbol{X}_j = [X_j, X_{j+1}, \cdots, X_{j+n}]^t$.

A stochastic process $X(t)$ is said *well known*, if the distribution $F(x_1, \cdots, x_n; t_1, \cdots, t_n)$ is precisely known for all n, every set $\{t_j\}$ and every vector value $\boldsymbol{x}$. The process is instead said *parametrically known*, if there exists a finite dimensional parameter vector $\boldsymbol{\theta} = [\theta_1, \cdots, \theta_m]^t$ such that the conditional distribution $F(x_1, \cdots, x_n; t_1, \cdots, t_n|\boldsymbol{\theta})$ is precisely known for all n, every set $\{t_j\}$, every vector value $\boldsymbol{x}$ and a fixed given value of $\boldsymbol{\theta}$. A stochastic process $X(t)$ is *non parametrically* described, if there is no vector parameter $\boldsymbol{\theta}$ of finite dimensionality such that the distribution $F(\boldsymbol{x}, \boldsymbol{t}|\boldsymbol{\theta})$ is completely described for all values of the vector $\boldsymbol{\theta}$ and for all $n, \boldsymbol{t}$ and $\boldsymbol{x}$. As an example, a stationary, discrete time process $\{X_i\}$ where the random variables X_i have finite variances is a nonparametrically described process. In fact, this description represents a whole class of stochastic processes. If we assume now that this process is also Gaussian, then it becomes *parametrically known*, since only its mean and spectral density functions are needed for its full description. When these two quantities are also provided, the process becomes *well known*.

From now, we have the main necessary ingredients to give a general scope of the detection and estimation theory. Let consider a case where the observed quantity is modelled by a stochastic process $X(t)$ and the observed signal $x(t)$ is considered as a realization of the process, *i.e.*, an observed waveform generated by $X(t)$.

1.1 Basic definitions

- Probability spaces:
  The probability theory starts by defining an *observation set* Γ and a class of subsets $\mathcal{G}$ of it, called *observation events*, to which we wish to assign probabilities. The pair $(\Gamma, \mathcal{G})$ is termed the *observation space.*

  For analytical reasons we will always assume that the collection $\mathcal{G}$ is a σ-algebra; that is, $\mathcal{G}$ contains all complements relative to Γ and denumerable unions of its members, *i.e.*;

  $$\begin{array}{lll} \text{if } A \in \mathcal{G} & \longrightarrow & A^c \in \mathcal{G} \\ \text{and} & & \\ \text{if } A_1, A_2, ... \in \mathcal{G} & \longrightarrow & \cup_i A_i \in \mathcal{G} \end{array} \tag{1.1}$$

  Two special cases are of interest:

  - Discrete case: $\Gamma = \{\gamma_1, \gamma_2, \cdots\}$
    In this case $\mathcal{G}$ is the set of all subsets of Γ which is usually denoted by 2^Γ and is called the *power set* of Γ.
    For this case, probabilities can be assigned to subsets of Γ in terms of a *probability mass function*, $p : \Gamma \longrightarrow [0, 1]$, by

    $$P(A) = \sum_{\gamma_i \in A} p(\gamma_i), \quad A \in 2^\Gamma. \tag{1.2}$$

    Any function mapping Γ to $[0, 1]$ can be a probability mass function provided that it satisfies the condition of normality

    $$\sum_i p(\gamma_i) = 1. \tag{1.3}$$

  - Continuous case: $\Gamma = \mathbf{R}^n$, the set of n-dimensional vectors with real components.
    In this case we want to assign the probabilities to the sets

    $$\{\boldsymbol{x} = (x_1, \cdots, x_n) \in \mathbf{R}^n | a_1 < x_1 < b_1, \cdots, a_n < x_n < b_n\} \tag{1.4}$$

    where the a_i's and b_i's are arbitrary real numbers. So, in this case, $\mathcal{G}$ is the smallest σ-algebra containing all of these sets with the a_i's and b_i's ranging throughout the reals. This σ-algebra is usually denoted $\mathcal{B}^n$ and is called the class of *Borel sets* in $\mathbf{R}^n$.
    In this case the probabilities can be assigned in terms of a *probability density function*, $p : \mathbf{R}^n \longrightarrow \mathbf{R}^+$, by

    $$P(A) = \int_A p(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}, \quad A \in \mathcal{B}^n. \tag{1.5}$$

    Any integrable function mapping $\mathbb{R}^n$ to $\mathbf{R}^+$ can be a probability density function provided that it satisfies the condition

    $$\int_{\mathbb{R}^n} p(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} = 1. \tag{1.6}$$

- For compactness, we may use the term *density* for both probability density function and probability mass function and use the following notation when necessary

$$P(A) = \int_A p(\boldsymbol{x}) \mu(\,\mathrm{d}\boldsymbol{x}) \tag{1.7}$$

for both the summation equation (1.2) and the integration equation (1.5).

- Random variable:
$X = X(\omega)$ is a function $\omega \mapsto \mathbb{R}$, where ω represents elements on the probability space.

- Probability distribution:
$F(x)$ is a function $\mathbf{R} \mapsto [0, 1]$ such that

$$F(x) = \Pr\{X \leq x\} \tag{1.8}$$

- Probability density function:
$f(x)$ is a function $\mathbb{R} \mapsto \mathbf{R}^+$ such that

$$\begin{aligned} f(x) &= \frac{\partial F(x)}{\partial x}, \\ F(x) &= \Pr\{X \leq x\} = \int_{-infty}^{x} f(t)\,\mathrm{d}t \end{aligned} \tag{1.9}$$

- For a real function g of the random variable X, the *expected value* of $g(X)$, denoted $\mathrm{E}[g(X)]$, is defined by any of the followings:

$$\mathrm{E}[g(X)] = \sum_i g(\gamma_i)\, p(\gamma_i) \tag{1.10}$$

$$\mathrm{E}[g(X)] = \int_{\mathbf{R}} g(x)\, p(x)\,\mathrm{d}x \tag{1.11}$$

$$\mathrm{E}[g(X)] = \int_{\Gamma} g(x)\, p(x)\, \mu(\,\mathrm{d}x) \tag{1.12}$$

- Random vector or a vector of random variables:
$X = [X_1, \cdots, X_n]$ where X_i are scalar random varaibles.

- Joint probability distribution:

$$\begin{aligned} F(x_1, \cdots, x_n) &= \Pr\{X_1 \leq x_1, \cdots, X_n \leq x_n\} \\ F(\boldsymbol{x}) &= \Pr\{\boldsymbol{X} \leq \boldsymbol{x}\} \end{aligned} \tag{1.13}$$

- Stochastic process

- Stochastic process: $X(t) = X(t, \omega)$, where $X(t, \omega)$ is a scalar random variable for all t.

- A stochastic process is completely defined by

$$\begin{aligned} F(x_1, \cdots, x_n; t_1, \cdots, t_n) &= \Pr\{X_i(t_i) \leq x_i, i = 1, \cdots, n\} \\ F(\boldsymbol{x}; \boldsymbol{t}) &= \Pr\{\boldsymbol{X}(\boldsymbol{t}) \leq \boldsymbol{x}\} \end{aligned} \tag{1.14}$$

- A stochastic process is stationary if

$$\begin{aligned} F(x_1, \cdots, x_n; t_1, \cdots, t_n) &= F\left(x_1, \cdots, x_n; (t_1 + \tau), \cdots, (t_n + \tau)\right) \\ F(\boldsymbol{x}; \boldsymbol{t}) &= F(\boldsymbol{x}; \boldsymbol{t} + \tau \mathbf{1}) \end{aligned} \tag{1.15}$$

- A stochastic process is memoryless or white if

$$F(x_1, \cdots, x_n; t_1, \cdots, t_n) = \prod_i F\left(x_i; t_i\right) \tag{1.16}$$

- Discrete time stochastic process:

$$F(\boldsymbol{x}) = \Pr\{\boldsymbol{X} \leq \boldsymbol{x}\} \tag{1.17}$$

- Memoryless discrete time stochastic process:

$$F(\boldsymbol{x}) = \Pr\{\boldsymbol{X} \leq \boldsymbol{x}\} = \prod_{i=1}^{n} F_i(x_i) \tag{1.18}$$

- Memoryless and stationary discrete time stochastic process:

$$F(\boldsymbol{x}) = \prod_{i=1}^{n} F_i(x_i) \quad \text{and} \quad F_i(x) = F_j(x) = F(x), \forall i, j \tag{1.19}$$

- A memoryless and stationary discrete time stochastic process generates in time *independent and identically distributed* (i.i.d.) random variables.

- Well known stochastic process:
  A stochastic process is well known if the distribution $F(\boldsymbol{x}, \boldsymbol{t})$ is known for all $n, \boldsymbol{t}$ and $\boldsymbol{x}$.

- Parametrically well known stochastic process:
  A stochastic process is parametrically well known if there exists a finite dimentional vector parameter $\boldsymbol{\theta} = [\theta_1, \cdots, \theta_m]$ such that the conditional distribution $F(\boldsymbol{x}, \boldsymbol{t}|\boldsymbol{\theta})$ is known for all $n, \boldsymbol{t}$ and $\boldsymbol{x}$.

- Non parametric description of a stochastic process:
  A stochastic process $X(t)$ is non parametrically described, if there is no vector parameter $\boldsymbol{\theta}$ of finite dimensionality such that the distribution $F(\boldsymbol{x}, \boldsymbol{t}|\boldsymbol{\theta})$ is completely described for every given vector $\boldsymbol{\theta}$ and for all $n, \boldsymbol{t}$ and $\boldsymbol{x}$.

- Observed data:
  samples of $x(t)$ a realization of $X(t)$ in some time interval $[0, T]$.

1.2 Summary of notations

X	A random variable
x	A realization of a random variable
$\boldsymbol{x} = \{x_1, \cdots, x_n\}$	n samples (realizations) of a random variable
$X(t)$	A random function or a stochastic process
$x(t)$	A realization of a random function
$\boldsymbol{X}$	A random vector or a discrete-time stochastic process
$\boldsymbol{x}$	A realization of a random vector or a discrete-time stochastic process
$\boldsymbol{x}_n = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$	n samples (realizations) of a random vector or a discrete-time stochastic process
$F(x)$	A probability distribution of a scalar random variable
$f(x)$	A probability density function of a scalar random variable
$F_{\boldsymbol{\theta}}(x)$	A parametrical probability distribution
$f_{\boldsymbol{\theta}}(x\|\boldsymbol{\theta})$	A parametrical probability density function
$F(x\|\boldsymbol{\theta})$	A conditional probability distribution
$f(x\|\boldsymbol{\theta})$	A conditional probability density function
$F(\boldsymbol{\theta}\|\boldsymbol{x})$	A posterior probability distribution of $\boldsymbol{\theta}$ conditioned on the observations $\boldsymbol{x}$
$f(\boldsymbol{\theta}\|\boldsymbol{x})$	A posterior probability density function of $\boldsymbol{\theta}$ conditioned on the observations $\boldsymbol{x}$
Θ	A random scalar parameters
θ	A sample of Θ
$\boldsymbol{\Theta}$	A random vector of parameters
$\boldsymbol{\theta}$	A sample of $\boldsymbol{\Theta}$
Γ	The space of possible values of x
$\boldsymbol{\Gamma}$	The space of possible values of $\boldsymbol{x}$
$\mathcal{T}$	The space of possible values of θ
$\boldsymbol{\mathcal{T}}$	The space of possible values of $\boldsymbol{\theta}$