INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Robust continuous-time detection in linear process noise

Srinivasa, Patibandla Rao, Ph.D.
Rice University, 1992
Robust Continuous-time Detection in Linear Process Noise

by

P. Srinivasa Rao

A Thesis Submitted
In Partial Fulfillment of the Requirements for the Degree
Doctor of Philosophy

Approved, Thesis Committee:

[Signatures]

Dr. Don H. Johnson, Chairman
Professor
Electrical and Computer Engineering

Dr. Behnaam Aazhang
Associate Professor
Electrical and Computer Engineering

Dr. Paul E. Pfeiffer
Professor
Mathematical Sciences

Dr. Katherine Ensor
Assistant Professor
Statistics

Houston, Texas
April, 1992
Robust Continuous-time Detection in Linear Process Noise

P. Srinivasa Rao

Abstract

Linear processes are suitable for modeling the random received waveforms in a scattering channel, which represents radar, sonar and multipath communication channels. We address the continuous-time detection problem where both the noise and signal-plus-noise waveforms are modeled as linear processes. Uncertainty in the nominal model is considered in the form of classes of probability distributions induced on the function space $L_2[0,T]$ by the processes under the two hypotheses. By embedding the linear processes in the larger class of infinitely divisible processes, and using an integral representation for the latter class, we identify the pair of distributions that are least favorable for the discrimination of the two linear processes; an optimal detector designed for these distributions is robust for the uncertainty classes considered. An investigation of a suboptimal performance criterion also leads us to the same robust detector.
I thank my thesis advisor Dr. Don Johnson for his generous support and guidance of this work. It is a privilege to have been initiated to research under his supervision. I also like to thank Dr. Behnaam Aazhang for his constant encouragement over the years, and Dr. Paul Pfeiffer and Dr. Katherine Ensor for serving on my thesis committee. Association and friendship with fellow graduate students has enriched my stay at Rice. Finally, I thank my wife Durga for her love and understanding.
Contents

Abstract ii
Acknowledgments iii
List of Illustrations vi

1 Introduction 1

2 Continuous-time Linear Processes 5
   2.1 Stochastic Process as a Family of Functions ................. 5
   2.2 Linear Processes ........................................... 8
   2.3 Infinitely Divisible Processes .............................. 14
   2.4 Linear Processes as Models in Detection Theory ............. 22

3 Optimal Detection in Linear Processes 25
   3.1 Continuous-time Detection ................................... 25
   3.2 Likelihood Ratio Detectors for Linear Processes ............. 27

4 Robust Optimal Detection in Linear Processes 32
   4.1 Minimax Robustness ........................................ 33
   4.2 Uncertainty Models for Linear Processes ..................... 35
   4.3 Robust Discrimination of Poisson Random Measures ........... 38
   4.4 Least Favorable Distributions for Linear Processes .......... 45
   4.5 Example ................................................... 46
5 Robust Suboptimal Detection in Linear Processes 49
  5.1 A Distance Criterion .............................................. 50
  5.2 Minimax Distance Linear Process Detection ..................... 53
  5.3 RKHS Methods for Robust Suboptimal Detection ................. 56

6 Conclusions 62

A Likelihood Ratio Formula for Linear Processes 64

B RKHS Approach to the Robust Detection of Linear Processes using the Distance Criterion 67

Bibliography 70
Illustrations

2.1 Classes of stochastic processes; IDP Infinitely divisible processes, LP Linear processes, IIP Independent increment processes, GP Gaussian processes, MP Markov processes. .......................... 15

2.2 A sample function of the non-Gaussian part of the independent increment process $x(t)$ and the corresponding realization of the Poisson random measure $\nu(dt, dv)$. ................................. 20

2.3 Transmitter T and receiver R with a common (shaded) region S of illuminated scatterers. ...................................................... 23

5.1 Analogy between the RKHS approaches to matched filtering and detection using distance criterion. ........................................ 60
Chapter 1

Introduction

Statistical models for optimal detection of signals in noise commonly treat the signal, or noise, or both as being Gaussian. However, in many important situations such as the radar and sonar signal processing, and satellite communication, the Gaussian assumption is not valid [33][34]. Unfortunately, it is often difficult to develop models for non-Gaussian processes that are accurate and yet analytically tractable. Classes of non-Gaussian processes that retain some of the analytical simplicity of Gaussian processes are important for this reason. For example, non-Gaussian processes with spherically invariant joint distributions are used to model under-ice acoustic noise [3]. Another important class is that of linear processes, which is suitable for modeling reverberation in a scattering channel. Because the scattering channel represents radar, sonar, seismic sounding, and multipath communication channels [17], we have focused on developing a detection theory for linear non-Gaussian processes.

The class of linear processes is closed under linear transformations and their finite-dimensional characteristic functions are known. Linear processes are essentially filtered versions of independent increment processes. Wiener process, Poisson process, compound Poisson processes (all of which have independent increments) and many Gaussian processes are linear. So too are the models of random received waveforms in a scattering channel. For example, in the case of active sonar and radar, the filter response corresponds to the signal reflection from a scatterer and the independent
increment process to the random parameters of the scatterers such as their position, size, etc.

The Gaussian and non-Gaussian components of an independent increment process and hence those of a linear process are independent. We consider here only linear processes without a Gaussian component. This restricted model is suitable when the scattering density is low and the receiver thermal noise can be neglected compared to the reverberation noise. As the density of the scatterers increases, the received process can be approximated as being Gaussian using the central limit theorem. A general study on some statistical distance measures between the finite-dimensional distributions of a discrete-time linear process and a Gaussian process was done by Mallows [39]. Briefly, as the bandwidth of the filter decreases, the linear process approaches a Gaussian process.

In this work, we consider detection problems where both the noise and the signal-plus-noise waveforms are modeled as linear processes without a Gaussian component. Wolff (cf. [8]) considered a related problem of detecting a known signal in discrete-time linear process noise. Eastwood and Lugannani [16] use the continuous-time linear process model and obtain an approximate likelihood ratio detector for a finite number of samples using a series approximation for the inversion of their joint characteristic functions. In contrast, we deal with continuous-time detection, using the entire observed waveform for decision making. The entire observation is treated as a random variable taking values in a function space \( L_2[0,T] \). It follows from Grenander [23] that a threshold test on the Radon-Nikodym derivative between the two probability measures on this function space is optimal for the Neyman-Pearson and minimum probability of error criteria, analogous to the likelihood ratio test in the finite-dimensional case.
The finite-dimensional distributions of a linear process can be shown to be infinitely divisible. Making use of an integral representation [40] for all processes with this property, the Radon-Nikodym derivative (which is also referred to as the likelihood ratio) between two probability measures corresponding to linear processes can be obtained [10][11]. However, this method only leads to what is referred to as a representation formula for the likelihood ratio that prescribes its distribution, but does not enable us to compute its value for a given sample function.

Optimal statistical procedures, such as the likelihood ratio test, are often quite sensitive to deviations from the assumed model (the probability measures here). Because in practice we rarely, if ever, have a knowledge of the actual model with sufficient accuracy, it is preferable to seek robust procedures that are not sensitive to small deviations from the assumed model. Robust statistical procedures have been studied extensively in the literature in recent years. See, for example, Huber [27], and for work related to signal processing, see the survey article by Kassam and Poor [32].

This thesis is concerned with robust detection schemes for linear processes. Instead of assuming uncertainty in different components of the problem (such as the marginal distribution, covariance function etc.) separately, as is often done, we take the approach of Kelly and Root [33] and consider uncertainty in the probability measures induced on the function space by the processes. This approach is well suited for the continuous-time detection problem. Most of the previous work on robust continuous-time detection however is concerned primarily with Gaussian or Poisson processes [33][18] and is not adequate for dealing with the general waveforms in a scattering channel. For structured uncertainty classes of probability measures, we seek detectors for linear processes that are robust with respect to the optimal Neyman-Pearson criterion. The integral representation and the likelihood ratio of all processes with infinitely divisible finite-dimensional distributions plays a key role in the solution
of this problem. The consideration of optimal performance criterion leads to detection structures that are often difficult and even impossible to implement in practice. Suboptimal criteria such as the signal-to-noise ratio are used in such cases to simplify the detector implementation. We also study detectors for linear processes that are robust with respect to a statistical distance criterion, which is a generalization of the often used signal-to-noise ratio. Detection problems can be posed in what are called reproducing kernel Hilbert spaces (RKHS), making possible a geometric approach to these problems. We extend some of the recent theory concerning RKHS approach for robust matched filtering, to the above robust distance detection problem for linear processes.
Chapter 2

Continuous-time Linear Processes

As mentioned in the introduction, we model the received waveform in the continuous-time detection problem as a linear process under both hypotheses. In this chapter, we develop the background material on linear processes. We begin with a general discussion on different ways of defining continuous-time stochastic processes, focusing on the method most suitable for our work. Then we define linear processes and discuss some of their properties. Next we introduce the class of infinitely divisible processes which encompasses linear processes, and outline an integral representation for infinitely divisible processes, which is instrumental in the derivation of both the optimal and robust detection results for linear processes. In the last section, we describe how linear processes are suitable as models in signal detection problems.

2.1 Stochastic Process as a Family of Functions

The classical theory of statistical inference for a finite number of samples was first extended by Grenander [23] to the case of continuous-time stochastic processes. Following Grenander, we describe three different approaches to the definition of stochastic processes. We are interested in continuous-time stochastic processes defined on the finite interval \( I \triangleq [0, T] \).

The first approach considers a stochastic process \( x(t) \) as a one parameter family of random variables \( \{x(t_0), t_0 \in I\} \). Assuming that \( E x^2(t_0) < \infty \) for all \( t_0 \in I \), each
random variable \( x(t_0) \) is considered as a point in the Hilbert space \( L_2(x) \) with inner product \( \langle y, z \rangle = Eyz \), consisting of all finite linear combinations of such random variables and their limits in the mean. The stochastic process itself becomes a curve in this abstract space (which is different from the underlying sample space \( \Omega \)). In studying the process, only the metric properties of the curve are usually used, disregarding the concrete meaning of the points on the curve. Note that this approach uses only the first two moments of the process and is suitable for application in linear problems. We will not use this approach here, although in Chapter 5 we deal with an abstract space of random variables obtained by nonlinear operations on the process.

To make statements of an inferential nature, the second approach by Doob of considering the process as a family of functions is preferable. As opposed to the previous approach where the sample space \( \Omega \) is an abstract space, here we take \( \Omega = \mathbb{R}^I \), the set of all real functions on \( I \). Define the stochastic process by \( x_\omega(t) = \text{value of } \omega \text{ at } t \). The map defined by \( x_\cdot(t) \) from \( \mathbb{R}^I \) to \( \mathbb{R} \) is called the coordinate function, because \( x_\omega(t) \) is the \( t \)th coordinate of \( \omega \) [54, p. 39]. Notice that in the current approach, \( \omega \) plays the role of the parameter. The smallest algebra with respect to which every \( x_\cdot(t) \) is measurable consists of sets of the form \( \{ \omega : (x_\omega(t_1), x_\omega(t_2), \ldots, x_\omega(t_n)) \in B \} \), where \( B \) is an \( n \)-dimensional Borel set. Such sets are referred to as \( n \)-dimensional cylinder sets and the smallest \( \sigma \) algebra \( \mathcal{B} \) containing all the cylinder sets is referred to as the Borel sets of \( \mathbb{R}^I \). Suppose we are given a family of finite-dimensional distributions \( \{ P_\theta, \theta \in \Theta \} \), the collection of all finite subsets of \( I \) that are compatible: If \( \theta \) and \( \theta' \) are two ordered finite subsets of \( I \) such that \( \theta \) contains \( \theta' \), then \( P_{\theta'} \) must be equal to \( P_\theta \) with the appropriate variables set to \( \infty \). We can then define a probability measure \( P' \) on the cylinder sets by setting

\[
P'(\{ \omega : (x_\omega(t_1), x_\omega(t_2), \ldots, x_\omega(t_n)) \in B \}) = \int_B dP_{t_1, t_2, \ldots, t_n}(x_1, x_2, \ldots, x_n)
\]
This measure can be uniquely extended to $\mathcal{B}$ [43, p. 82]. We thus have a probability space $(\mathbb{R}^I, \mathcal{B}, P')$. The usual definition of strict-sense stationarity given in terms of the finite-dimensional distributions can be expressed in the present function space approach as follows. The measure $P'$ must satisfy $P'(A) = P'(T_h A)$ for all sets $A \in \mathcal{B}$ and for all $h$, where $T_h$ is a translation operator defined by $T_h x(t) = x(t + h)$.

In practice, the sample space $\Omega$ can often be taken as a subset of $\mathbb{R}^I$, provided the distributions $\{P_{t_n}\}$ satisfy further conditions. For example, if a process $x(t)$ is separable and satisfies the Kolmogorov condition $E |x(t + h) - x(t)|^\alpha \leq C h^{1 + \beta}$ for some strictly positive constants $\alpha, \beta$, and $C$, its sample functions are almost surely continuous. We can then assume $\Omega = C(I)$, the space of continuous functions on $I$, and define a probability measure $P$ on this space [48][44, p. 216]. The constructive proof for the existence of $P$ begins with the probability space $(\mathbb{R}^I, \mathcal{B}, P')$ (constructed as above), and involves the definition of a measurable map of $\mathbb{R}^I$ into $C(I)$ which induces the probability measure $P$ on $C(I)$ such that

$$P(\{ x : x \in C(I), (x(t_1), x(t_2), \ldots, x(t_n)) \in B \}) = \int_B dP_{t_1,t_2,\ldots,t_n}(x_1, x_2, \ldots, x_n).$$

If $x(t)$ is the Brownian motion, this procedure leads to the well known Wiener measure on $C(I)$.

Suppose that $x(t)$ is continuous in probability (also referred to as stochastically continuous in the literature). Then it can be taken to be separable and measurable; if $x(t)$ does not have these properties, it can be modified to define a process $\hat{x}(t)$ that has the same finite-dimensional distributions as $x(t)$, and is separable and measurable.

If in addition $\int_I E x^2(t) \, dt < \infty$, i.e., the correlation function satisfies the condition $\int_I R(t, t) \, dt < \infty$, it follows from Fubini's theorem that the sample functions are square-integrable almost surely [54, p. 45]. Assuming that the process is defined on the abstract space $(\Omega, \mathcal{A}, \mu)$, we can define the map $\Gamma : (\Omega, \mathcal{A}) \to (L_2(I), \mathcal{B})$. 


by $\Gamma(\omega) = x(\cdot, \omega)$. It follows from the measurability of $x(t)$ and separability of $L_2(I)$ that $\Gamma$ is measurable and hence induces a probability measure $P$ on $(L_2(I), B)$ given by $P(B) = \mu \Gamma^{-1}(B) = \mu(\omega : x(\cdot, \omega) \in B)$ [49]. It should be remarked that the usual assumption of continuity in the (quadratic) mean of the process implies both continuity in probability and the continuity (and hence integrability) of $R(t, t)$ on the compact interval $I$, thus enabling us to define the probability measure on $L_2(I)$. Although we have described the construction of $P$ on $(L_2(I), B)$ for the sake of completeness, it is common in the literature to consider the specification of the measure $P$ itself (usually in terms of its characteristic functional, which is defined in Section 2.3) as a starting point for the study of the process. See, for example, Gikhman and Skorokhod [19].

The two approaches described above are evidently based on the fact that a stochastic process depends on two variables $t$, and $\omega$. The final approach is to consider the process as function of two variables $x(t, \omega), t \in I, \omega \in \Omega$, with a given probability measure on $\Omega$. This approach was shown to be equivalent to the second approach above and we shall not pursue it here.

2.2 Linear Processes

A process is said to have independent increments if its increments over disjoint intervals of time are independent random variables. (The names additive, decomposable and differential process are also used in the literature). Generally speaking, a linear process is an $L_2$ filtering of an independent increment process. Before giving a formal definition of linear processes, we introduce the important concept of infinite divisibility and discuss briefly some of the properties of independent increment processes that we need.
Definition 2.2.1 [22, p. 71] A random variable $x$ is *infinitely divisible* if, for each positive integer $n$, it can be represented as the sum of $n$ i.i.d. random variables $x_1, x_2, \ldots, x_n$ (whose distribution depends on $n$).

Clearly an equivalent definition of the infinite divisibility of $x$ is that its characteristic function $\Phi_x(u) = E e^{iuX}$, must satisfy $\Phi_x(u) = \left( \Phi_n(u) \right)^n$ for all $n$, $\Phi_n(u)$ being some valid characteristic function. A distribution function (without regard to any particular random variable) is said to be infinitely divisible if its characteristic function (Fourier transform) satisfies this condition. Infinite divisibility of random vectors and distributions on $\mathbb{R}^n$ is defined similarly. The class of infinitely divisible distributions coincides with that of limit distributions for sums of independent random variables, thus playing an important role in probability theory. Two well known examples of infinitely divisible random variables are the Gaussian with a characteristic function $\Phi_x(u) = \exp \{ -\sigma^2 u^2 / 2 \}$, and the centered Poisson with $\Phi_x(u) = \exp \{ \lambda (\exp(iu) - iu - 1) \}$.

Three different canonical representations have been derived for the characteristic function of infinitely divisible random variables [22]. We will use Kolmogorov's *representation* for the zero mean, finite variance case

$$\ln \Phi_x(u) = \int_{-\infty}^{+\infty} \left( e^{iu} - iuv - 1 \right) \frac{1}{v^2} dM_x(v), \quad (2.1)$$

where $M_x(v)$ is non-decreasing and bounded in $v$, with $M_x(-\infty) = 0$. The integrand is defined by continuity at $v = 0$, where it takes the value $-u^2 / 2$. Note that for the Gaussian and centered Poisson examples, $M_x(v)$ has a single point of increase, at $v = 0$ and $v = 1$ respectively. As our primary concern will be distributions without a Gaussian component, the integration region is assumed to exclude the origin. The
function $M_x(v)$ is then uniquely determined by the relationship

$$-rac{d^2}{du^2} \ln \Phi_x(u) = \int_{-\infty}^{+\infty} e^{iuv} dM_x(v).$$

We will often be interested in situations when $M_x(v)$ can be written in the form $M_x(v) = \int_{-\infty}^{u} w^2 dG_x(w)$, where $G_x(v)$ is a scaled distribution function (similar to $M_x(v)$). We can then write 2.1 as

$$\ln \Phi_x(u) = \int_{-\infty}^{+\infty} (e^{iuv} - iuv - 1) dG_x(v). \quad (2.2)$$

Random variables whose characteristic functions have the above representation are referred to as elementary infinitely divisible random variables and all infinitely divisible random variables are limits of sequences of random variables of this type [44].

We will also write the above integral representation in terms of the finite measure on $R$ corresponding to $G_x(v)$, which we denote by the same symbol. The characteristic function of an elementary infinitely divisible random vector $\mathbf{x}$ has the analogous representation

$$\ln \Phi_{\mathbf{x}}(\mathbf{u}) = \int_{\mathbb{R}^n} (e^{i(\mathbf{u}, \mathbf{v})} - i(\mathbf{u}, \mathbf{v}) - 1) G_{\mathbf{x}}(d\mathbf{v}). \quad (2.3)$$

where $(\mathbf{u}, \mathbf{v}) = \sum_{k=1}^{n} u_k v_k$, and $G_{\mathbf{x}}(\cdot)$ is a finite measure on $\mathbb{R}^n$.

Let $\{x(t), t \in [0, T]\}$ be a zero-mean, second-order, independent increment process that is continuous in probability. Without loss of generality, we will assume that $x(0) = 0$ almost surely. Any increment $x(t) - x(t')$, $0 \leq t' < t \leq T$ of the process converges to zero in probability (becomes "asymptotically negligible") in the limit as $t'$ tends to $t$. Taking partitions of $[0, t]$, the increment $x(t) - x(0)$ can then be written as the limit of a sequence of sums of independent random variables that are uniformly asymptotically negligible, and hence by the central limit theorem must be infinitely divisible [37, p. 545].

It is interesting to note that Doob [14, p. 129] introduces the notion of infinite divisibility in the generalized sense where the random variables $x_1, \ldots, x_n$ in
Definition 2.2.1 may not be identically distributed, but then goes on to prove that an infinitely divisible distribution is an infinitely divisible distribution in the generalized sense and conversely. This fact enables us to see clearly why the increments of an independent increment process are infinitely divisible in general, and not just in the homogeneous (stationary increments) case which is discussed most often. See for example [21, p. 363] and [2, p. 195]. The close relation between independent increment processes and infinitely divisible random variables was described by Ito [29]; the increments of an independent increment process are infinitely divisible and given an infinitely divisible distribution, an independent increment process \( \{x(t); 0 \leq t \leq 1\} \) can be defined such that \( x(1) \) has this distribution. Historically, the canonical representations for the characteristic functions of infinitely divisible random variables have resulted from the study of independent increment processes.

As we shall see later, independent increment processes without a Gaussian component have sample functions that are constant except for jumps. If the number of jumps in each finite interval is finite, they are called compound Poisson or jump processes. For such processes, the characteristic function of \( x(t) \) has a representation of the form 2.2

\[
\ln \Phi_t(u) = \int_{-\infty}^{+\infty} (e^{iuv} - iuv - 1) dG_t(v),
\]

where \( G_t(v) \) (after normalization) is the distribution of jump heights at time \( t \).

Denoting the characteristic function of the increment \( x(t) - x(t') \), \( 0 \leq t' < t \leq T \) by \( \Phi_{t',t}(u) \), since the increments over disjoint intervals are independent, we must have

\[
\Phi_t(u) = \Phi_{t'}(u) \Phi_{t',t}(u).
\]  

It follows that the function \( G_{t',t}(v) \) in the representation of \( \Phi_{t',t}(u) \) satisfies \( G_{t',t}(v) = G_t(v) - G_{t'}(v) \) and hence that \( G_t(v) \) is nondecreasing in \( t \) also. We can then define a new function \( G(t,v) \) that is nondecreasing in both \( t \) and \( v \) such that \( G_t(v) = \)
\[ \int_0^t d_s G(s, v) \] and write
\[ \ln \Phi_t(u) = \int_0^t \int_{-\infty}^{+\infty} (e^{iuv} - iuv - 1) d_s d_v G(s, v). \]

The use of Kolmogorov's representation 2.1 for \( \Phi_t(u) \) instead of the special form 2.2 in the above argument would result in the monotonic function \( M(t, v) \). Next, suppose that \( V(t) \triangleq E x^2(t) \). It follows that \( V(t_2) - V(t_1) = E \{ x(t_2) - x(t_1) \}^2 \), for \( t_1 < t_2 \) and hence \( V(t) \) is a monotone, nondecreasing function. The functions \( V, M \) and \( G \) are related by
\[ V(t) = \frac{d^2}{du^2} \Phi_t(u) \bigg|_{u=0} = \int_0^t \int_{-\infty}^{+\infty} d_s d_v M(s, v) = \int_0^t \int_{-\infty}^{+\infty} d_s d_v G(s, v). \]

If the independent increment process is homogeneous, it follows from the equation 2.5 that \( M(t, v) = t M'(v) \) and \( V(t) = t M'(\infty) \) [14, p. 419]. Indeed, this condition is also sufficient for the process to be homogeneous. The nondecreasing functions \( M(t, v) \) and \( G(t, v) \) define finite measures on the product space \( I \times \mathbb{R} \), which we denote by the same symbols in future.

**Definition 2.2.2** A linear process \( y(t) \) is defined by the stochastic integral
\[ y(t) = \int_0^T f(t, s) \, dx(s), \quad t \in [0, T], \tag{2.6} \]
where \( x(s) \) is a zero-mean, second-order independent increment process that is continuous in probability.

The region of integration is assumed to be the interval \([0, T]\) simply for convenience in dealing with the parameters of the independent increment process \( x(t) \); in general any interval \([a, b], -\infty \leq a < b \leq \infty\), can be used. The stochastic integral is defined as a limit in the mean of a sequence of random variables whenever \( \int f^2(t, s) \, dV(s) < \infty \) [14, p. 426]. In fact, 2.6 defines an isometry between the space \( L_2(V) \) of functions
and the space $L_2(x)$ of random variables. It follows easily that the covariance function of the linear process is given by

$$K(t_1, t_2) = E y(t_1) y(t_2) = \int_0^T f(t_1, s) f(t_2, s) dV(s).$$  \hspace{1cm} (2.7)$$

Conversely, all stochastic processes with a covariance function of the above form have integral representations [13] similar to that in 2.6 but with $x(s)$ being an orthogonal increment process and hence are not necessarily linear according to our definition. For Gaussian processes however, a covariance function of the form 2.7 ensures their linearity; an orthogonal increment Gaussian process is also an independent increment process. A particular example is the class of mean-square continuous wide-sense stationary Gaussian processes with 2.7 corresponding to the Wiener-Khinchin theorem and 2.6 to the spectral representation of the process. If $f(t, s) = f'(t - s)$, $x(t)$ is homogeneous and $a = -\infty$ and $b = \infty$ then $y(t)$ is stationary. Brownian motion is an example of a nonstationary linear process.

The above definition of linear processes is commonly used and is analogous to the discrete-time version. A discrete-time linear process $\{Y_n\}$ is defined by $Y_n = \sum_{j=-\infty}^{+\infty} \alpha_j X_{n-j}$, where the random variables $X_n$ are i.i.d with zero mean and unit variance and $\sum_{j=-\infty}^{+\infty} \alpha_j^2 < \infty$. This class of processes encompasses [51, p. 39] the ARMA processes defined by the stochastic difference equation $\sum_{k=0}^{p} \beta_k Y_{n-k} = \sum_{j=0}^{q} \alpha_j X_{n-j}$. Analogously, a wide class of continuous-time processes defined in terms of stochastic differential equations are linear according to our definition [6, p. 148]. However, other definitions have been considered in the literature [24, 45]. Hida and Ikeda [24] define a linear process $y(t)$ to be one for which the linear manifold generated by $\{y(\tau), \tau \leq t\}$, and its orthogonal complement in the Hilbert space $L_2(y)$, are independent for all $t$. Clearly, all independent increment processes are linear according to both definitions. According to the definition of Hida and Ikeda,
all Gaussian processes are linear, without restriction on the form of their correlation functions. Their work deals with the class of Markov linear processes and the properties of their sample functions.

The class of linear processes is closed under linear transformations. Another important property that these processes share with Gaussian processes is that expressions for their finite-dimensional characteristic functions are easily obtained. Suppose \( f(t_0, s) \) is a piecewise-constant function of the form

\[
f(t_0, s) = f_1 I_{[0,s_1)}(s) + \ldots + f_m I_{[s_{m-1},s]_T}(s),
\]

where \( 0 = s_0 \leq s_1 \ldots \leq s_{m-1} \leq s_m = T \) is a partition of \([0,T]\) and \( I_A(\cdot) \) is the indicator function of set \( A \). The characteristic function of \( y(t_0) \) is easily seen to be

\[
\ln \Phi_{t_0}(u) = \sum_{j=1}^{m} \ln \Phi_{x_{j-1},x_j}(f_j u) = \sum_{j=1}^{m} \int_{s_{j-1}}^{s_j} \int_{-\infty}^{+\infty} (e^{if_j uv} - if_j uv - 1) G(ds, dv)
\]

\[
= \int_0^T \int_{-\infty}^{+\infty} (e^{if(t_0,s) uv} - if(t_0,s) uv - 1) G(ds, dv).
\]

Considering a sequence of functions of the form 2.8 converging to \( f(t_0, s) \) in \( L_2(V) \), it can be shown that the above form of the characteristic function is valid in general [14, p. 426][38]. Denote the characteristic function of the random vector \( y_\ell = (y(t_1), \ldots, y(t_n)) \) by \( \Phi_\ell(u) \) where \( \ell = (t_1, \ldots, t_n) \) and \( u = (u_1, \ldots, u_n) \). Noting that for fixed \( u \), this function is same as the characteristic function of \( y(t_0) \) with \( f(t_0,s)u \) replaced by \( w = \sum_{k=1}^{n} u_k f(t_k, s) \) and that this is true for all \( u \), we have

\[
\ln \Phi_\ell(u) = \int_0^T \int_{-\infty}^{+\infty} (e^{iwv} - iwv - 1) G(ds, dv).
\]

\[
(2.9)
\]

### 2.3 Infinitely Divisible Processes

Stochastic processes can often be approximated by a sum of a large number of independent components, each of which contributes an insignificant proportion of the
whole. An example is the point process of calls arriving at a telephone exchange, considered as the sum of a number of point processes of calls made by individual subscribers. Lee [35] first introduced and studied infinitely divisible processes as a model for such processes. A stochastic process is infinitely divisible if all of its finite-dimensional distributions are infinitely divisible.

Let \( \{\xi(t), t \in I\} \) be a zero-mean, second-order, infinitely divisible process that is continuous in probability and has no Gaussian component. Suppose that all finite-dimensional distributions of this process are elementary infinitely divisible. Denote by \( \Theta \) the collection of all finite subsets \( \theta = \{t_1, \ldots, t_n\} \) of \( I \). From 2.3, the characteristic function of the vector \( \xi_\theta = (\xi(t_1), \ldots, \xi(t_n)) \) has the representation

\[
\ln \Phi_\theta(u) = \int_{\mathbb{R}^\theta} \left( e^{i(u, \omega)} - i(u, \omega) - 1 \right) \Lambda_\theta(d\omega),
\]

with \( \Lambda_\theta(\cdot) \) being a finite measure on \( \mathbb{R}^\theta \).

For a linear process \( y(t) \) with finite-dimensional characteristic functions of the

\[\text{IDP} \supseteq \text{LP} \supseteq \text{IIP} \supseteq \text{GP} \supseteq \text{MP}\]

Figure 2.1: Classes of stochastic processes; IDP Infinitely divisible processes, LP Linear processes, IIP Independent increment processes, GP Gaussian processes, MP Markov processes.
form 2.9, define a measure on $\mathbb{R}^\theta$, the set of real functions on $\theta$, by

$$
\Lambda_\theta(A) = G((s,v): (v f(t_1,s), \ldots, v f(t_n,s)) \in A).
$$

(2.11)

Denoting $v_\theta = (v f(t_1,s), \ldots, v f(t_n,s))$, we can write the characteristic function of the vector $y_\theta = (y(t_1), \ldots, y(t_n))$ as

$$
\ln \Phi_\xi(y) = \int_0^T \int_{-\infty}^{+\infty} G(ds, dv) = \int_\mathbb{R}^\theta (e^{i(u,v_\theta)} - i(u,v_\theta) - 1) \Lambda_\theta(dv_\theta),
$$

which is of the form 2.10. Thus the linear processes are infinitely divisible [11]. The class of infinitely divisible processes clearly also contains all independent increment processes and Gaussian processes. Figure 2.1 depicts the relationship between various classes of processes.

Given the compatible family of measures $\{\Lambda_\theta(\cdot), \theta \in \Theta\}$ (which may not be finite in the general case) of an infinitely divisible process, Maruyama [40] constructs a measure $\Lambda(\cdot)$ on the function space $\mathbb{R}^f$ such that $\Lambda(g_{\theta f}^{-1} A) = \Lambda_\theta(A)$ for all measurable $A \subset \mathbb{R}^\theta$, where $g_{\theta f}(\cdot)$ is the projection mapping from $\mathbb{R}^f$ to $\mathbb{R}^\theta$ restricted to the complement of its null space. The measure $\Lambda(\cdot)$ is referred to as the projective limit of the collection $\{\Lambda_\theta(\cdot)\}$. The construction of $\Lambda(\cdot)$ involves an extension of Bochner’s [9] method for defining a probability measure on the function space from a compatible family of finite-dimensional distributions to the general case of infinite measures. For the case of interest to us, the projective limit $\Lambda(\cdot)$ of the finite measures $\{\Lambda_\theta(\cdot)\}$ is finite. Furthermore, we will assume that the process $\xi(t)$ is defined on $L_2(I)$ as described in section 2.1, and that $\Lambda(\cdot)$ is restricted to $L_2(I)$. Although we do not give a proof for this restriction of $\Lambda(\cdot)$ along the lines of Maruyama, suffice it to say that infinitely divisible processes are commonly defined as probability measures on
$\mathcal{X} \hat{=} L_2(I)$. These probability measures are specified in terms of their characteristic functionals, which are infinite-dimensional analogues of characteristic functions.

Suppose a process $x(t)$ induces a probability measure $P$ on $\mathcal{X}$. Then the characteristic functional of the process is defined by [19]

$$C(\psi) = E e^{i (x, \psi)} = \int_{\mathcal{X}} e^{i (x, \psi)} P(dx),$$

where $\psi \in \mathcal{X}$ and $(x, \psi) = \int_0^T x(t) \psi(t) \, dt$ is the inner product in $\mathcal{X}$. Depending on the nature of the process $x(t)$ (specifically the function space to which its sample functions belong), other definitions of characteristic functionals are also used. A characteristic functional completely defines the probability measure corresponding to a process. If $x(t)$ is Gaussian with mean $m(t)$ and covariance $K(t, s)$, then

$$\ln C(\psi) = i (m, \psi) - (K \psi, \psi)/2,$$

where $K$ denotes the integral operator on $\mathcal{X}$ with a kernel $K(t, s)$ [15]. Characteristic functionals, when they can be expressed in a closed form, provide a convenient method of characterizing non-Gaussian processes.

The nonnegative definiteness of characteristic functionals makes it possible to extend the reproducing kernel Hilbert space methods for detection to the non-Gaussian case [25][15], and in Section 5.3, we use this extension to characterize robust nonlinear statistics. Infinitely divisible processes without a Gaussian component are specified by characteristic functionals of the form

$$\ln C(\psi) = \int_{\mathcal{X}} \left[ e^{i (x, \psi)} - i (x, \psi) - 1 \right] \Lambda(dx),$$

where $\Lambda(\cdot)$ is a finite measure on $\mathcal{X}$ [19]. As in the case of the characteristic function of an elementary infinitely divisible distribution, we are considering here a special form of the general characteristic functional to suit our purpose. We use the above characteristic functional in Appendix A to provide a simple, alternative derivation of the likelihood ratio between two infinitely divisible processes. It follows from 2.11
that the projective limit measure of a linear process (which we assume is confined to $\mathcal{X}$) is given by [11]

$$\Lambda(A) = G((s,v) : vf(\cdot,s) \in A), \quad \forall A \in \mathcal{B},$$

where $\mathcal{B}$ is the class of Borel subsets of $\mathcal{X}$.

Using the measure $\Lambda$, Maruyama obtains an integral representation of infinitely divisible processes which generalizes the Lévy–Ito decomposition of independent increment processes. The sample functions of a continuous independent increment process $\{x(t), t \in I\}$ (almost surely) have jump discontinuities only. It is thus possible to decompose each of these sample functions uniquely as the sum of a continuous function and a piecewise-constant function. Analogously, it follows from the infinite divisible nature of the characteristic functions that the process itself can be decomposed as the sum of a Gaussian part with continuous sample paths and an independent non-Gaussian part whose paths are piecewise-constant functions. The non-Gaussian part can further be decomposed as a continuous sum (integral) of scaled independent Poisson processes corresponding to the different heights of the jump.

Let $\nu_t(A)$ denote the number of jumps of the independent increment process $x(t)$ occurring in $[0,t)$ with height in $A$, $A$ being any Borel subset of $\mathbb{R}$. It can be shown that $\nu_t(A)$ is a (nonhomogeneous) Poisson process and is independent of $\nu_t(A')$ whenever $A$ and $A'$ are disjoint [20, p. 261]. Further, $E\nu_t(A) = G_t(A)$. Since the number of jumps occurring during disjoint intervals of time are independent, $\nu([t_1,t_2] \times A) = \nu_{t_2}(A) - \nu_{t_1}(A)$ can be extended to a Poisson random measure with independent values on the Borel subsets of $I \times \mathbb{R}$, with $E \nu(dt, dv) = G(dt, dv)$; if $B_j$ are disjoint Borel subsets of $I \times \mathbb{R}$, $\nu(B_j)$ are independent Poisson random variables with means $G(B_j)$, and $\nu(\cdot, \cdot, \omega)$ is a measure on $I \times \mathbb{R}$ for almost all $\omega$. Random counting measures (such as the Poisson random measure $\nu$) are a convenient
tool for extending the concept of point processes on the real line to arbitrary spaces. For example, each realization of a Poisson process on the real line corresponds to a counting (integer-valued) measure which assigns to each Borel set in \( \mathbb{R} \) the number of points occurring in that set. Analogously, the process itself corresponds to a random counting measure. Making use of this correspondence, Poisson processes on arbitrary measurable spaces can be conveniently described in terms of Poisson random measures [31]. We will return to the topic of Poisson random measures in Section 4.3.

\( \nu(\cdot, \cdot) \) is quite similar in nature to the (real-valued) random measure induced on \( I \) by \( x(t) \), with respect to which the stochastic integral 2.6 is defined. Stochastic integration with respect to \( \nu(\cdot, \cdot) \) is defined similarly and the Lévy–Itô decomposition of \( x(t) \) is given by [19]

\[
x(t) = x^0(t) + \int_0^t \int_{\mathbb{R}} \nu(ds, dv) - G(ds, dv)
\]

(2.14)

where \( x^0(t) \) is a Gaussian independent increment process independent of \( \nu(\cdot, \cdot) \) and the equality holds with probability one. Because we will be dealing with processes without a Gaussian component, we ignore the Gaussian part henceforth. The Poisson random measure \( \nu \) corresponds to (or is identified with) a Poisson process on \( I \times \mathbb{R} \).

The realization of this process is given by the points \((t_k, v_k), k = 1, \ldots, M\) which are the times of occurrence and the heights of jumps in the sample function of the non-Gaussian part of \( x(t) \). This one-to-one correspondence between the realizations of \( x \) and \( \nu \) is illustrated in Figure 2.2. Note that the Lévy–Itô decomposition 2.14 expresses the value of the non-Gaussian component of the independent increment process at time \( t \) as the cumulative sum of all the jumps in the interval \([0, t)\):

\[
x(t) - x^0(t) = \sum_{k=1}^M x_k(t) - b(t)
\]
Figure 2.2: A sample function of the non-Gaussian part of the independent increment process $x(t)$ and the corresponding realization of the Poisson random measure $\nu(dt, dv)$.

where

$$x_k(t) = \begin{cases} 0 & t < t_k, \\ v_k & t \geq t_k, \end{cases} \quad t \in I,$$  \hspace{1cm} (2.15)$$

and $b(t) = \int_0^t \int_R v G(ds, dv)$.

Maruyama [40] extends the representation 2.14 to general infinitely divisible processes by considering a Poisson random measure $\Pi(\cdot)$ on the Borel sets of $\mathbb{R}^I$ with $E \Pi(A) = \Lambda(A)$. Defining stochastic integrals with respect to this random measure,
he shows that an infinitely divisible process $\xi(t)$ without a Gaussian component can be represented as

$$\xi(t) = \int h(t, z) [\Pi(dz) - \Lambda(dz)],$$

(2.16)

where $h(t, z)$ denotes the $t$th coordinate of $z(\cdot)$ (i.e., $h(t, z) = z(t)$), and the equality is now only in distribution [50]. Because we are interested in infinitely divisible processes inducing probability measures on the Hilbert space $\mathcal{X}$, we find the heuristic approach of Gikhman and Skorokhod [19] to obtain a representation of the form 2.16 for this case, more suitable. It can be shown that the characteristic function of a random variable $\int_{\mathcal{X}}(x, \psi)\Pi^*(dz)$, where $\Pi^* = \Pi - \Lambda$ is a centered Poisson random measure, is identical to the right hand side of 2.12 (see equation 4.14). We thus have

$$\xi(\cdot) = \int_{\mathcal{X}} z \Pi^*(dz),$$

(2.17)

where the equality is again in distribution. Note that the integrand is now written as an element of the function space, as opposed to being a functional as in 2.16. Equation 2.17 is purely symbolic in nature and denotes the equality in distribution of the random variables $(\xi, \psi)$ and $\int_{\mathcal{X}}(z, \psi)\Pi^*(dz)$ for all $\psi \in \mathcal{X}$. Similarly $\int_{\mathcal{X}} z \Lambda(dz)$ denotes the element of $\mathcal{X}$ satisfying $(\int_{\mathcal{X}} z \Lambda(dz), \psi) = \int_{\mathcal{X}}(z, \psi) \Lambda(dz)$ [44, p. 168]. Each realization of the Poisson random measure $\Pi$ is a finite set $\{z_k\}_{k=1}^M$ of functions in $\mathcal{X}$. Thus the integral representation 2.17 can also be expressed as

$$\xi(\cdot) = \sum_{k=1}^M z_k(\cdot) - c(\cdot),$$

where $c(\cdot) = \int z \Lambda(dz)$. Note that as opposed to the independent increment process case, it is not possible to directly relate a sample function of $\xi(t)$ to the realization of $\Pi$ in general. Although this fact limits the utility of the above integral representation, it is useful in the derivation of the likelihood ratio formula for infinitely divisible processes and consequently our robust detector for linear processes.
2.4 Linear Processes as Models in Detection Theory

Having discussed the important properties of linear processes, we next turn to their applicability in decision theoretic problems. In this section, we describe situations in which random phenomena of interest to us can be modeled as linear processes. We have noted that many Gaussian processes, Poisson, compound Poisson and filtered compound Poisson processes are linear. Middleton’s general model [42] for reverberation in a scattering channel is also a linear process. The scattering channel usually models the radar, sonar and multipath communication channels. We outline here the description of Middleton’s reverberation model as a linear process. See Eastwood’s thesis [17] for further details.

To deal with the scattering problem, instead of the physical approach of starting with a dynamical equation governing wave propagation in an inhomogeneous medium with random scattering introduced as a spatial perturbation, Middleton [42] uses what he refers to as the “quasi-phenomenological” approach; start with a dynamical equation again, but introduce inhomogeneities as an independent random distribution of point scatterers in a homogeneous and isotropic medium. This approach has the advantage of handling general signals (both narrowband and broadband) and providing higher-order statistics needed for system design and analysis. The principal disadvantage is “in the ultimately empirical nature of the impulse response function of the scatterers, which must be quantified by experiment” [42], which provides the motivation for our study of uncertainty models for linear processes.

Suppose that the transmitter illuminates a region of scatterers in space, a portion of which (denoted as $S$ in Figure 2.3) is also viewed by the receiver. Denote the random spatial coordinates of the $M$ scatterers in the region $S$ by $\mathbf{s}_k$, $k = 1 \ldots M$. It will be convenient to write $\mathbf{s}_k$ in polar coordinates as $\mathbf{s}_k = (t_k, \alpha_k, \beta_k) = (t_k, \theta_k)$,
where the radial coordinate is expressed in time units because of the assumption of constant propagation velocity. The resulting scatter process $y(t)$ is the superposition of the waveforms $U_k(t; \bar{x}_k, \bar{\ell}_k)$ due to the scatterer at $\bar{x}_k$, where $\bar{\ell}_k \in \Theta$ is a vector of random parameters such as scattering area, components of the velocity etc:

$$y(t) = \sum_{k=1}^{M} U_k(t; \bar{x}_k, \bar{\ell}_k).$$

Next suppose that the waveforms $U_k(t; \bar{x}_k, \bar{\ell}_k)$ are identical and are deterministic functions. The scatterers are assumed to have a Poisson distribution in space. That is, the probability that exactly $N$ of them lie in a small interval $\Delta \bar{x}_k$ around $\bar{x}_k$ is given by

$$P(N | \Delta \bar{x}_k, \bar{x}_k) = [\lambda(\bar{x}_k)|\Delta \bar{x}_k|]^{N} \cdot e^{-\lambda(\bar{x}_k)|\Delta \bar{x}_k|} / N!,$$

where $\lambda(\bar{x}_k)$ is the intensity of the Poisson process at $\bar{x}_k$, and $|\Delta \bar{x}_k|$ denotes the size (length, area etc.) of the interval $\Delta \bar{x}_k$. 

Figure 2.3: Transmitter T and receiver R with a common (shaded) region S of illuminated scatterers.
The Poisson process on $\mathcal{S}$ can be extended to $\mathcal{S} \times \Theta$ by defining a generalized Poisson process $x(t', s', \theta)$ with an intensity function $\lambda(s) p(\theta)$, where $p(\theta)$ is the probability density of the random parameter vector $\theta$. The reverberation process can then be expressed in the integral form

$$y(t) = \int_{\mathcal{S} \times \Theta} U(t; t', s', \theta) \, dx(t', s', \theta).$$

(2.18)

Note that since the Poisson process $x$ has independent increments, $y(t)$ is a generalized linear process. It can be analyzed along the same lines as the linear process defined in 2.6, simply by a suitable modification of the region of integration. For example, the $n$-dimensional characteristic function of $y(t)$ given by Middleton is similar to that in 2.9. Finally, for the case of scatterers on a line and when the only random parameter is the amplitude of basic waveform, 2.18 can be written as

$$y(t) = \int_t U(t; t') \, dx(t'),$$

where we have incorporated the random amplitude into the process $x(t')$, thus making it a compound Poisson process on an interval of the real line. This equation is precisely the form of the linear process discussed in Section 2.2, and we shall henceforth use it as a model for reverberation without loss of generality.
Chapter 3

Optimal Detection in Linear Processes

We describe here some of the previous work on detection schemes for linear processes, focusing primarily on optimal continuous-time procedures. We first discuss Grenander’s general framework [23], which extends the classical theory of statistical inference from finite-dimensional samples to the infinite-dimensional case.

3.1 Continuous-time Detection

Suppose $x_1, \ldots, x_n$ are the observed values of the components of a random vector $X$, with the probability distribution $P_0$ under the hypothesis $H_0$ and distribution $P_1$ under the alternative $H_1$. Also suppose that we reject $H_0$ when $(x_1, \ldots, x_n) \in W$, and that the probability of falsely rejecting $H_0$ (also referred to as the error of first kind) is $P_0(W) = \epsilon$. The region $W$ of the observation space is called a critical region of size $\epsilon$. The optimum Neyman-Pearson criterion is to choose from all tests (i.e., critical regions $W$) of size $\epsilon$ the one with a minimum error of the second kind $P_1(W^C)$. It is well known that the best critical region is given by the likelihood ratio test $W = \{l(x_1, \ldots, x_n) \geq c\}$, where $l(x_1, \ldots, x_n)$ is the likelihood ratio defined by

$$l(x_1, \ldots, x_n) = \frac{p_1(x_1, \ldots, x_n)}{p_0(x_1, \ldots, x_n)},$$

and $p_0$ and $p_1$ are the density functions corresponding to $P_0$ and $P_1$, respectively, and the constant $c$ being chosen to satisfy $P_0(W) = \epsilon$. 

25
Let us next consider statistical hypothesis testing for discrete and continuous-time stochastic processes. In both cases, the observation is assumed to be a sequence of real numbers \((x_1, x_2, \ldots) = \omega \in \Omega\). In the case of continuous-time processes, the sequence of coordinates can be derived from the waveform in different ways, depending on the nature of the process. For example, if the sample functions are a.s. continuous, the values of the process at time instants \(\{t_j\}\) that is a dense subset of \(I\) can be used as the coordinates. If the sample functions belong a.s. to \(\mathcal{X} = L_2(I)\), the coordinates can be taken as the Fourier coefficients with respect to some complete orthonormal set of functions. Suppose that the simple hypothesis \(H_0\) corresponds to a probability measure \(P_0\) on \(\Omega\) and the alternative \(H_1\) corresponds to \(P_1\). Assuming that \(P_1\) is absolutely continuous with respect to \(P_0\) (\(P_1(E) = 0\) whenever \(P_0(E) = 0\), denoted by \(P_1 \ll P_0\)), it follows from the Radon-Nikodym theorem [54, p. 215] that for every measurable set \(E \subset \Omega\),

\[
P_1(E) = \int_E l(\omega) \, dP_0(\omega),
\]

where the nonnegative integrable function \(l(\omega)\) is called the Radon-Nikodym derivative (or the density) of \(P_1\) with respect to \(P_0\) and is denoted by \((dP_1/dP_0)(\omega)\). Grenander [23] showed that \(l(\omega)\) plays the same role in the present case as the likelihood ratio does in the classical case; the best critical region in the Neyman-Pearson sense is of the form \(\{l(\omega) \geq c\}\). Furthermore, Grenander also provides the following constructive method of obtaining the Radon-Nikodym derivative. Assuming that the probability distributions of the finite set of coordinates \((x_1, \ldots, x_n)\) is for all \(n\) absolutely continuous with density functions \(q_0\) and \(q_1\) under \(H_0\) and \(H_1\) respectively, it is easy to see that the likelihood ratio \(l_n(x_1, \ldots, x_n)\) of \(q_1\) and \(q_0\) equals the conditional expectation \(E_0[l(\omega)|x_1, \ldots, x_n]\) a.s. \((P_0)\). The martingale sequence \(l_n\) converges to
\( l(\omega) \) a.s. with respect to both \( P_0 \) and \( P_1 \). That is
\[
l(\omega) = \lim_{n \to \infty} \frac{q_1(x_1, \ldots, x_n)}{q_0(x_1, \ldots, x_n)}.
\]

It is common to refer to the Radon-Nikodym derivative \( l(\omega) \) as the likelihood ratio of \( P_1 \) and \( P_0 \).

### 3.2 Likelihood Ratio Detectors for Linear Processes

We will be primarily concerned with the detection of a stochastic signal in the presence of additive random noise. Under both hypotheses (\( H_0 \): noise, \( H_1 \): signal-plus-noise), the observed waveform is modeled as a linear process, with different statistical properties. Our goal is then to decide between the hypotheses

\[
H_0 : \quad y(t) = \int f_0(t, s) \, dx_0(s),
\]
\[
H_1 : \quad y(t) = \int f_1(t, s) \, dx_1(s), \tag{3.1}
\]

according to an optimal (Neyman-Pearson, minimum probability of error) criterion. This problem set up is suitable for target detection in reverberation environments, multipath communication and channel sounding problems [17].

A similar problem of detecting a constant signal in discrete-time linear process noise was studied by Wolff [8]. He specifies the conditions under which the linear process is finitely invertible (\( X_n \) can be obtained by passing \( Y_n \) through an FIR filter) and studies the performance of nonparametric detection schemes under these conditions. Eastwood and Lugannani [16], on the other hand, use the continuous-time model above but assume that only a finite number of samples are observed instead of the entire waveform. Using a series approximation for the joint density of the samples obtained by the inversion of the characteristic functions 2.9, they derive an approximate likelihood ratio detector.
In contrast to the above approaches, we focus on the continuous-time detection of a random signal that along with the signal-plus-noise process is modeled as a linear process. Although most detectors are implemented in discrete-time, often the continuous-time model provides useful upper bounds on detection performance and leads to simplifying approximations to the actual implementation of the detector [4]. Further, to study robust detection, it is perhaps more meaningful to consider uncertainty models on the measures corresponding to the observation as a whole rather than the component-by-component uncertainty [33] because of the dependent nature of the noise samples and their dependence with those of the signal. This consideration also favors the continuous-time approach taken here.

In the hypothesis testing problem 3.1, suppose for the moment that the two filters \( f_j(t, s), j = 0, 1 \) are identical to \( f(t, s) \) which is causally invertible. If \( f(t, s) = h(t-s) \) and \( h(\cdot) \in L_2 \), the well known Paley-Wiener condition [54, p. 118]

\[
\int_{-\infty}^{+\infty} \frac{|\ln H^2(\omega)|}{1 + \omega^2} \, dw < \infty,
\]

where \( H(\cdot) \) is the Fourier transform of \( h(\cdot) \), is sufficient to guarantee the existence of a causal inverse filter. In the case of multipath communication (assuming BPSK, we can incorporate the phase factor in \( x_j \)) and channel sounding problems, \( f(t, s) \) is nearly the same as the transmitted signal and hence could be modified suitably to satisfy these conditions. By passing the observed waveform through the inverse filter, we recover the sample function of the independent increment process \( x_j(s) \).

From our earlier assumptions that \( x_j(s) \) has a finite measure \( G_j \) and no Gaussian component, we noted in Section 2.3 that it must be a compound Poisson process with sample functions which are constant except for a finite number of jumps. Denote by \( t = (t_1, \ldots, t_M) \) and \( v = (v_1, \ldots, v_M) \) the times of occurrence and the heights of these jumps. Clearly the vector \((L, v)\) is a sufficient statistic for the observation \( y \).
of the linear process and the problem 3.1 reduces to a likelihood ratio test for this vector. Assuming that the measures \( G_j \) corresponding to \( x_j(s) \) are absolutely continuous with densities \( g_j(t,v) \), it can be shown that the likelihood ratio of the joint densities of the vector \((t,v)\) under \( H_1 \) and \( H_0 \) is given by [17]

\[
l(t,v) = e^{-Q_1-Q_0} \prod_{k=1}^{M} \frac{g_1(t_k, v_k)}{g_0(t_k, v_k)},
\]

where \( Q_j = G_j(I \times \mathbb{R}) \). Noting that \( Q_j \) can also be expressed as \( Q_j = \int_0^T \lambda_j(t) \, dt \), with \( \lambda_j(t) = \int_{\mathbb{R}} g_j(t, v) \, dv \) being the intensity of points on the time axis, 3.2 can be seen to be analogous to the likelihood ratio for the discrimination of two Poisson processes with different intensity functions.

Often the filters \( f_j(t,s) \) are different or, \( f(t,s) \) is not known precisely, and the assumption of invertibility is not suitable in such situations. Our focus will be on the general problem 3.1, where given the realization of a linear process \( y(t) \), we must make an optimal decision between hypotheses \( H_0 \) and \( H_1 \). It is clear from the discussion in the previous section that it suffices to compute the Radon-Nikodym derivative \( (dP_1/dP_0)(\cdot) \), where \( P_0 \) and \( P_1 \) are the probability measures induced by \( y(t) \) on the function space under \( H_0 \) and \( H_1 \) respectively.

Since \( y(t) \) is also an infinitely divisible process, we have the projective limit measures \( \Lambda_0 \) and \( \Lambda_1 \) specified by

\[
\Lambda_j(A) = G_j((s,v): v f_j(\cdot,s) \in A), \quad \forall A \in B.
\]

corresponding to the two hypotheses [11]. The process \( y(t) \) has the integral representation

\[
y = \int_{\chi} x \left[ \Pi_j(dx) - \Lambda_j(dx) \right]
\]

under the hypothesis \( H_j \), where \( \Pi_j \) is a Poisson random measure with intensity \( \Lambda_j \). Under the condition that \( \Lambda_1 \ll \Lambda_0 \), (and \( \Lambda_0(.) \) and \( \Lambda_1(.) \) have identical means, i.e.,
\[ \int_{\mathcal{X}} x \Lambda_0(dx) = \int_{\mathcal{X}} x \Lambda_1(dx) \] Briggs [10] shows that \( P_1 \ll P_0 \) and

\[
\frac{dP_1}{dP_0}(y) = \exp \left\{ -(\Lambda_1(\mathcal{X}) - \Lambda_0(\mathcal{X})) + \int_{\mathcal{X}} \ln \frac{d\Lambda_1}{d\Lambda_0}(x) \Pi_0(dx) \right\} \tag{3.4}
\]

under hypothesis \( H_0 \). We provide an alternative derivation of this likelihood ratio formula in Appendix A, using the characteristic functional approach. Denoting the realization of the Poisson random measure \( \Pi_0 \) by the finite set of functions \( \{x_k\} \) as in Section 2.3, we can write the above likelihood ratio as

\[
\frac{dP_1}{dP_0}(y) = e^{-(\Lambda_1(x) - \Lambda_0(x))} \prod_{k=1}^{M} \frac{d\Lambda_1}{d\Lambda_0}(x_k) \tag{3.5}
\]

which is analogous to the likelihood ratio between probability measures induced by Poisson processes on the real line [15]. Because such processes correspond to Poisson random measures on \( \mathbb{R} \), the likelihood ratio formula in this case can thus be viewed as a special case of that for i.d.p.s given above.

It is important to note that 3.4 is only a representation formula for the likelihood ratio and is not always computable in terms of the observed waveform \( y(t) \). The only random component on the right hand side of the formula is the realization of the Poisson random measure \( \Pi_0(\cdot) \), and the integral in 3.3 (sum if written in terms of the realization \( \{x_k\} \)) that maps it into the function space is not one-to-one and hence may not be invertible. As noted in Section 2.3, this fact is in stark contrast to the situation for independent increment processes where the Poisson measure \( \nu \) on \( I \times \mathbb{R} \) follows uniquely from the realization of the process, thus making the implementation of the likelihood ratio detector possible.

It is interesting to compare the above representation formula with Kailath’s general likelihood ratio formula for the detection of random signals in white Gaussian noise [30]. The hypotheses considered are

\[ H_0 : y(t) = \int_0^t x(s)ds + w(t) \]
\[ H_1 : y(t) = w(t) \]

where \( t \in I \), \( x(t) \) is the random signal and \( w(t) \) is the Wiener process. The likelihood ratio for this problem is given by

\[ L(y) = \exp \left\{ \int_I \hat{x}(t)dy(t) - \frac{1}{2} \int_I \hat{x}^2(t)dt \right\} \]

where \( \hat{x}(t) = E[x(t) \mid y(s), 0 \leq s \leq t, H_1] \) is the conditional mean of \( x(t) \) given the observations up to the instant \( t \) assuming that \( H_1 \) is true. Although the estimator-correlator nature of this formula provides clues for engineering approximations to the optimum receiver, the causal least-squares estimate is quite difficult to determine for non-Gaussian signals. The above formula thus converts the detection problem into an estimation problem that is at least as difficult to solve [47, p. 446]. Analogously, the implementation of the formula 3.4 requires the knowledge of the realization of the Poisson random measure.
Chapter 4

Robust Optimal Detection in Linear Processes

Optimal signal detection procedures such as the one described in the previous chapter require complete statistical description of the data. The performance of such procedures often deteriorates rapidly when the actual model begins to differ from that assumed. Since uncertainty regarding the statistical model is frequent, decision procedures which are robust, that is, which perform well despite small variations from the assumed statistical model, are preferable to the optimum procedures. The choice of robust procedures clearly involves a tradeoff between the degree of robustness and performance degradation under the assumed model.

The study of robust procedures for the detection of linear processes is the main focus of this work. In this chapter, we consider the robustness of optimal detection schemes, i.e., detection schemes with optimal performance criteria such as the Neyman-Pearson, minimum probability of error etc., and in the next chapter we consider detection robustness under a suboptimal performance criterion. The words optimal and suboptimal are included in the titles of these chapters in order to make this distinction clear, although the usage robust optimal is not standard and perhaps confusing in light of the discussion above. Among the various approaches to quantitative robustness, minimax schemes are important and usually lead to practical design of systems. First, we state and discuss the minimax robust hypothesis testing problem formulated by Huber [26]. We then describe the uncertainty models that we use in the problem of robust detection in linear processes. Robust discrimination between
two Poisson processes defined on a general measurable space is discussed in the next section. Using these results and the integral representation 2.17, we obtain the minimax robust detector for linear processes in Section 4.4 and discuss an example in Section 4.5.

4.1 Minimax Robustness

In a robust decision problem, the actual statistics of the data are assumed to belong to a class, usually a small neighborhood of some nominal statistics. This class is often of the nonparametric function type. To compare the various robust schemes, we must first specify a measure of the overall performance of these schemes on the entire class of allowable statistics. One such measure is the worst case performance, and this choice leads us to schemes that optimize the worst case performance, i.e., minimax robust schemes [32]. While choosing the minimax schemes, we expect that the worst case performance of these schemes is acceptably good and that their performance at the nominal operating point is not very far below that of the nominally optimum scheme.

Suppose that we have an observation $x$ of the random variable $X$ with values in $\mathcal{X}$. On the basis of $x$ we wish to decide between the hypotheses

$$H_0 : X \sim P_0,$$

$$H_1 : X \sim P_1,$$

(4.1)

where $P_0$ and $P_1$ are probability distributions on the measurable space $(\mathcal{X}, \mathcal{B})$ satisfying $P_1 \ll P_0$.

Let $\phi : \mathcal{X} \to [0, 1]$ be any test between $H_0$ and $H_1$, accepting $H_1$ with a conditional probability $\phi(x)$ when $x$ is observed. Assuming that a cost $C_j > 0$ is incurred if
\(H_j (j = 0, 1)\) is falsely rejected, the expected costs or risks are given by

\[
R(P_0, \phi) = C_0 E_{P_0}[\phi] = C_0 \Pr[\phi \text{ accepts } H_1 | H_0 \text{ is true}]
\]

and

\[
R(P_1, \phi) = C_1 E_{P_1}[1 - \phi] = C_1 \Pr[\phi \text{ accepts } H_0 | H_1 \text{ is true}],
\]

where \(E_{P_j}\) denotes the expectation under the assumption that \(X \sim P_j\). The Neyman-Pearson criterion for an optimal test between \(H_0\) and \(H_1\) can then be stated as

\[
\min_{\phi \in \Psi} R(P_1, \phi) \text{ subject to } R(P_0, \phi) \leq \alpha,
\]

(4.2)

where \(\Psi\) denotes the class of all randomized tests, and \(0 \leq \alpha \leq C_0\). When the distributions \(P_0\) and \(P_1\) are known, the above problem is solved by the well-known likelihood ratio test given by

\[
\phi_l(x) = \begin{cases} 
1 & l(x) > \eta, \\
p & l(x) = \eta, \\
0 & l(x) < \eta,
\end{cases}
\]

(4.3)

where \(l(x) = (dP_1/dP_0)(x)\), and the randomization \(p\) and the threshold \(\eta\) are chosen to satisfy \(R(P_0, \phi) = \alpha\).

To formulate the robust version of the above hypothesis testing problem, we assume that the distributions \(P_0\) and \(P_1\) are known only to be members of some classes \(\mathcal{P}_0\) and \(\mathcal{P}_1\) respectively, of probability distributions on \((\mathcal{X}, \mathcal{B})\). For this case, Huber [26] poses the following minimax robust problem that is a modified version of 4.2:

\[
\min_{\phi \in \Psi} \sup_{P_1 \in \mathcal{P}_1} R(P_1, \phi) \text{ subject to } \sup_{P_0 \in \mathcal{P}_0} \sup_{\phi \in \Psi} R(P_0, \phi) \leq \alpha.
\]

(4.4)

Note that the desired test \(\phi\) minimizes the worst risk performance over the classes \(\mathcal{P}_0\) and \(\mathcal{P}_1\). A general approach to the above problem can be given in terms of least favorable distributions.
Definition 4.1.1 A pair of distributions $P_j' \in \mathcal{P}_j$ is said to be least favorable (for tests between $\mathcal{P}_0$ and $\mathcal{P}_1$) if for every likelihood ratio test $\phi'$ between $P_0'$ and $P_1'$ we have

$$R(P_j', \phi') \geq R(P_j, \phi') \quad \forall P_j \in \mathcal{P}_j, \ j = 0, 1.$$  \hspace{2cm} (4.5)

Clearly if a least favorable pair of distributions $P_j'$ exists, the minimax robust problem 4.4 is reduced to that of the optimum test 4.2 between $P_0'$ and $P_1'$. The classes $\mathcal{P}_0$ and $\mathcal{P}_1$ must be nonintersecting. If they do intersect, the least favorable pair is in the intersection of these two classes and will satisfy $P_0' = P_1'$ thus making the robust problem trivial, but uninteresting. We describe next the structure of the uncertainty classes $\mathcal{P}_0$ and $\mathcal{P}_1$ for detection in linear processes.

4.2 Uncertainty Models for Linear Processes

The common approach to dealing with a continuous-time observation in a detection problem is to derive a real-valued sequence from it, for example by sampling or by Karhunen-Loève decomposition. The components of the sequence are assumed to be independent, identically distributed, or at worst weakly dependent and stationary. Uncertainty is introduced into the problem by assuming that the distribution of a component lies in some class. Many robust detection solutions are asymptotic in nature and require that an infinite number of observations be available and that the signal have infinite energy [33]. This assumption is necessary in order to use the central limit theorem to obtain the asymptotic distribution of various statistics, but it implies that the detection problem is singular.

The samples of a linear process are not independent. Eastwood and Lugannani [16] use a series approximation of the joint probability density obtained by inverting the
characteristic function to derive approximate likelihood ratio detectors for linear processes. It is unreasonable to expect the validity and utility of such approximations under a change in the measure of the entire observation in the robust version of the problem. Due to this and the above reasons, we follow Kelly and Root [33] in dealing with the realization and assume uncertainty in the measure corresponding to the observation as a whole.

Instead of defining the uncertainty classes of distributions $\mathcal{P}_0$ and $\mathcal{P}_1$ on $(\mathcal{X}, \mathcal{B})$ directly as in [33], we consider classes of distributions induced by classes of projective limit measures $\Lambda_0$ and $\Lambda_1$ corresponding to the linear processes under the two hypotheses. The measures $\Lambda_j$ (recall that these are assumed to be finite) are more directly related to the statistics of the processes than the probability measures and also give a complete description of the processes. We find it more natural to specify the uncertainty classes of the measures $\Lambda_j$ first and then consider the robust discrimination of the classes of probability distributions induced by them. A similar approach is used in the robust discrimination of Poisson processes on general spaces, in which the hypotheses are characterized by uncertainty neighborhoods of intensity measures [18, 52]. We return to the subject of Poisson processes in the next section.

The uncertainty classes of projective limit measures we consider are small neighborhoods of nominal measures $\tilde{\Lambda}_0$ and $\tilde{\Lambda}_1$ that are assumed to be completely known. The $\epsilon$-contamination and total variation neighborhoods are commonly used classes of uncertainty for probability distributions. The $\epsilon$-contamination neighborhoods are defined by [26]

$$L_j^\epsilon = \left\{ \Lambda_j : \Lambda_j = (1 - \epsilon)\tilde{\Lambda}_j + \epsilon \hat{\Lambda}_j \right\}, \quad j = 0, 1, \quad (4.6)$$

where $\hat{\Lambda}_j$ is an arbitrary finite measure on $(\mathcal{X}, \mathcal{B})$ with $\hat{\Lambda}_j(\mathcal{X}) = \tilde{\Lambda}_j(\mathcal{X})$ so that every member of the class has the same total measure as $\tilde{\Lambda}_j$. The parameter $\epsilon$ with a value
between 0 and 1 is known and is a measure of the degree of uncertainty about the nominal measures. The total variation neighborhoods are defined by [26]

\[ \mathcal{L}_j^{tv} = \{ \Lambda_j : \| \Lambda_j - \tilde{\Lambda}_j \| \leq \epsilon \}, \quad j = 0, 1, \]  

(4.7)

where the total variation between \( \Lambda_j \) and \( \tilde{\Lambda}_j \) is given by

\[ \| \Lambda_j - \tilde{\Lambda}_j \| = \sup_{A \in \mathcal{B}} |\Lambda_j(A) - \tilde{\Lambda}_j(A)|. \]

Again we assume that \( \Lambda_j(\mathcal{X}) = \tilde{\Lambda}_j(\mathcal{X}) \). We will use \( \mathcal{L}_j \) to denote both the \( \epsilon \)-contamination and total variation neighborhoods. We also assume that \( \epsilon \) is small enough so that the classes \( \mathcal{L}_0 \) and \( \mathcal{L}_1 \) do not overlap.

The robust optimum detection problem for linear processes that we address is the minimax problem in 4.4 where the classes of probability distributions \( \mathcal{P}_0 \) and \( \mathcal{P}_1 \) are those induced on the function space by linear processes with projective limit measures in \( \mathcal{L}_0 \) and \( \mathcal{L}_1 \) respectively. Specifically, we seek measures \( \Lambda_j' \in \mathcal{L}_j \) such that the corresponding probability distributions \( P_j' \in \mathcal{P}_j \) are least favorable. Recalling from Chapter 2 that the projective limit of a linear process is given by

\[ \Lambda(A) = G((s, v) : v f(\cdot, s) \in A), \]

clearly the uncertainty in \( \Lambda \) is a result of uncertainty in both the filter \( f(t, s) \) and the measure \( G(s, v) \). It is difficult to identify the structure of uncertainty of these two components which would lead to the \( \epsilon \)-contamination or total variation neighborhoods of uncertainty in \( \Lambda \). However, when \( f(t, s) \) is assumed to be known precisely, \( \epsilon \)-contamination or total variation uncertainty in \( \Lambda \) result from uncertainty in \( G \) of the same structure. Since the linear processes have Maruyama’s representation in terms of Poisson random measures with the measures \( \Lambda_j \) as their intensities, we make use of Sadowsky’s results [52] concerning the least favorable distributions for the Poisson case.
4.3 Robust Discrimination of Poisson Random Measures

We require the generalization of a Poisson point process on the real line to the measurable space \((\mathcal{X}, B)\). This generalization is achieved by introducing the notion of a random measure. Consider a Poisson process on the interval \([0, T]\). For each realization of the process, the interval function, which assigns to each interval the number of events occurring in it, can clearly be extended to a counting measure on the Borel sets of \([0, T]\). Furthermore, given any such counting measure, there exists a countable set of points on which the measure is supported and this set of points constitutes a valid realization of a point process. Thus, each realization of a Poisson process corresponds to a counting measure and vice versa. The process itself can then be considered a random counting measure. For example, if \(\lambda(t)\) is the intensity function of the Poisson process, the measure of an interval \(I\) is a Poisson random variable with a parameter \(\int_I \lambda(t) \, dt\).

A measure \(\pi\) on \(B\) is a counting measure if \(\pi\) is integer valued and is finite on bounded sets of \(B\). Such a measure is clearly \(\sigma\)-finite and there exists a countable set of points \(\mathcal{X}_\pi \subset \mathcal{X}\) such that for all \(B \in B\), \(\pi(B) = \sum_{x \in \mathcal{X}_\pi} \pi(\{x\}) \delta_x(B)\), where \(\delta_x(\cdot)\) denotes the unit mass distribution on the point \(x\) so that \(\delta_x(B) = I_B(x)\), the indicator function of \(B\). The set \(\mathcal{X}_\pi\) can be viewed as the set of points where the "events" occur and \(\pi(\{x\}), x \in \mathcal{X}_\pi\), as the number of events occurring at \(x\). Let \(N\) be the class of all counting measures on \(B\) with an associated \(\sigma\)-algebra \(\mathcal{N}\) that is the smallest \(\sigma\)-algebra that makes the maps \(\pi \rightarrow \pi(B)\) from \(N\) to the nonnegative integers measurable for each \(B \in B\).

**Definition 4.3.1** [41, p. 11] A random counting measure (point process)

\(\Pi\) is a random element taking values in the measurable space \((N, \mathcal{N})\).
Analogous to $\mathcal{X}_\pi$ above, we now have a random set of points $\mathcal{X}_\Pi$ where the events occur and $\Pi(\{x\})$, the random number of events occurring at $x$.

Let $\Lambda$ be a finite measure on $\mathcal{B}$. Then there exists a probability measure $P_\Lambda$ on $\mathcal{N}$ and a random counting measure $\Pi$ with distribution $P_\Lambda$, such that for any disjoint collection of bounded sets $\{B_k, k = 1, 2, \ldots\} \subset \mathcal{B}$, $\{\Pi(B_k)\}$ is a collection of independent Poisson random variables, $\Pi(B_k)$ being Poisson with parameter $\Lambda(B_k)$ [41, p. 57]. $\Pi$ is referred to as the Poisson random measure with the Poisson distribution $P_\Lambda$ with intensity measure $\Lambda$.

Consider the optimum discrimination between Poisson random measures. Suppose the hypotheses are simple and correspond to the Poisson distributions $P_{\Lambda_0}$ and $P_{\Lambda_1}$ with finite intensity measures $\Lambda_0$ and $\Lambda_1$ respectively. Given the observation of the event $\{\Pi = \pi\}$, the Neyman-Pearson optimum test $\phi : N \to [0, 1]$ which chooses $H_1$ with probability $\phi(\pi)$ is of the form 4.3. Note that the observation space is now $N$ instead of $\mathcal{X}$. If $\Lambda_1 \ll \Lambda_0$, it is known that $P_{\Lambda_1} \ll P_{\Lambda_0}$ and [41, p. 60][12]

$$
\frac{dP_{\Lambda_1}}{dP_{\Lambda_0}}(\pi) = \exp \left\{ -[\Lambda_1(\mathcal{A}) - \Lambda_0(\mathcal{A})] + \int_{\mathcal{A}} \ln \frac{d\Lambda_1}{d\Lambda_0}(x) \pi(dx) \right\} \quad (4.8)
$$

Consider next the robust discrimination of Poisson random measures when their intensities are not completely known. We assume that the intensity measures belong to the $\epsilon$-contamination neighborhoods $\mathcal{L}_j^\epsilon$ or the total variation neighborhoods $\mathcal{L}_j^{tv}$ (defined in 4.6 and 4.7 respectively) of the nominal measures $\bar{\Lambda}_j$. Let the nominal measures $\bar{\Lambda}_0$ and $\bar{\Lambda}_1$ have densities $\bar{\lambda}_0$ and $\bar{\lambda}_1$ with respect to some measure $\mu$, e.g., $\mu = \bar{\Lambda}_0 + \bar{\Lambda}_1$. For the case of $\epsilon$-contamination neighborhoods, define measures $\Lambda_j^\epsilon$ by their densities with respect to $\mu$ as follows:

$$
\lambda_j^\epsilon(x) = \begin{cases} 
(1 - \epsilon)\bar{\lambda}_0(x), & \bar{\lambda}_1(x)/\bar{\lambda}_0(x) < \epsilon'' \\
(1/\epsilon'')(1 - \epsilon)\bar{\lambda}_1(x), & \text{otherwise,}
\end{cases}
$$
\[ \lambda_i'(x) = \begin{cases} 
(1 - \epsilon)\bar{\lambda}_1(x), & \frac{\bar{\lambda}_1(x)}{\bar{\lambda}_0(x)} > c', \\
c'(1 - \epsilon)\bar{\lambda}_0(x), & \text{otherwise}, 
\end{cases} \tag{4.9} \]

where the constants \(0 \leq c' < c'' \leq \infty\) are determined so that \(\Lambda_j'(\mathcal{X}) = \bar{\Lambda}_j(\mathcal{X})\). It is easily verified that \(\Lambda_j' \in \mathcal{L}_j\). Similarly for the total variation neighborhood case, define \(\Lambda_j'\) by

\[ \lambda_0'(x) = \begin{cases} 
\frac{[1/(1 + c')]}{\bar{\lambda}_0(x)}(\bar{\lambda}_0(x) + \bar{\lambda}_1(x)), & \frac{\bar{\lambda}_1(x)}{\bar{\lambda}_0(x)} < c', \\
\bar{\lambda}_0(x), & c' < \frac{\bar{\lambda}_1(x)}{\bar{\lambda}_0(x)} < c'', \\
\frac{[1/(1 + c'')]}{\bar{\lambda}_0(x)}(\bar{\lambda}_0(x) + \bar{\lambda}_1(x)), & \frac{\bar{\lambda}_1(x)}{\bar{\lambda}_0(x)} > c'', 
\end{cases} \]

\[ \lambda_1'(x) = \begin{cases} 
\frac{[c'/(1 + c')]}{\bar{\lambda}_1(x)}(\bar{\lambda}_0(x) + \bar{\lambda}_1(x)), & \frac{\bar{\lambda}_1(x)}{\bar{\lambda}_0(x)} < c', \\
\bar{\lambda}_1(x), & c' < \frac{\bar{\lambda}_1(x)}{\bar{\lambda}_0(x)} < c'', \\
\frac{[c''/(1 + c'')]}{\bar{\lambda}_1(x)}(\bar{\lambda}_0(x) + \bar{\lambda}_1(x)), & \frac{\bar{\lambda}_1(x)}{\bar{\lambda}_0(x)} > c'', \tag{4.10} 
\end{cases} \]

where again the constants \(c'\) and \(c''\) are determined so that \(\Lambda_j'(\mathcal{X}) = \bar{\Lambda}_j(\mathcal{X})\) and \(\| \Lambda_j' - \bar{\Lambda}_j \| = \epsilon\).

**Lemma 4.1** [26] Let \(\mathcal{L}_j, j = 0, 1\) be the \(\epsilon\)-contamination or total variation neighborhoods of intensity measures \(\bar{\Lambda}_j\), with a sufficiently small \(\epsilon\).

Then the measures \((\Lambda_0', \Lambda_1') \in \mathcal{L}_0 \times \mathcal{L}_1\) defined above satisfy

\[ \Lambda_0'(\{l'(x) > t\}) \geq \Lambda_0(\{l'(x) > t\}) \quad \text{and} \]

\[ \Lambda_1'(\{l'(x) > t\}) \leq \Lambda_1(\{l'(x) > t\}) \tag{4.11} \]

for all \(\Lambda_j \in \mathcal{L}_j\) and \(t \in \mathbb{R}\), where \(l'(x) = \lambda_1'(x)/\lambda_0'(x)\).

The above definition of the measures \(\Lambda_j'\) and Lemma 4.1 are a simple generalization to finite measures of Huber's results [26] characterizing the least favorable probability distributions for \(\epsilon\)-contamination and total variation uncertainty classes. The assumption that the parameter \(\epsilon\) is small ensures the existence of constants \(c'\) and \(c''\)
satisfying \( c' < c'' \) and that the classes \( \mathcal{L}_j \) do not overlap (after normalizing all their members to be probability measures). Note that the likelihood ratio between \( \Lambda'_1 \) and \( \Lambda'_0 \) is a censored version of that between the nominal measures \( \bar{\Lambda}_1 \) and \( \bar{\Lambda}_0 \):

\[
\frac{d\Lambda'_1}{d\Lambda'_0}(x) = \frac{\lambda'_1(x)}{\lambda'_0(x)} = \begin{cases} 
    c', & \bar{\lambda}_1(x)/\bar{\lambda}_0(x) < c', \\
    \bar{\lambda}_1(x)/\bar{\lambda}_0(x), & c' < \bar{\lambda}_1(x)/\bar{\lambda}_0(x) < c'', \\
    c'', & \bar{\lambda}_1(x)/\bar{\lambda}_0(x) > c''.
\end{cases}
\tag{4.12}
\]

which can also be expressed in the concise form

\[
\frac{d\Lambda'_1}{d\Lambda'_0}(x) = \min \left\{ c'', \max \left\{ c', \frac{\bar{\lambda}_1(x)}{\bar{\lambda}_0(x)} \right\} \right\}.
\]

Consider random variables of the form

\[
U_g(\Pi) \overset{\Delta}{=} \int_\mathcal{X} g(x)\Pi(dx) = \sum_{x \in \Lambda_n} g(x)\Pi(\{x\}),
\tag{4.13}
\]

where \( g : \mathcal{X} \to \mathbb{R} \) is Borel measurable and \( \Pi \) is a Poisson random measure with a finite mean \( \Lambda \). Note that the likelihood ratio in 4.8, and the integral representation for infinitely divisible processes in 2.17 involve such random variables with \( g = \ln(d\lambda_1/d\lambda_0) \) and \( g = (\cdot, \psi), \psi \in \mathcal{X} \) respectively. The characteristic function of \( U_g \) is given by [52]

\[
E_{\mathbb{P}_A} \left[ \exp(iuU_g) \right] = \exp \left\{ \int_{\mathcal{X}} \left[ \exp(iug(x)) - 1 \right] \Lambda(dx) \right\}.
\tag{4.14}
\]

when \( g = I_A \), the indicator function of \( A \in \mathcal{B} \), \( U_g \) is clearly a Poisson random variable with mean \( \Lambda(A) \). In this case, 4.13 is easily seen to reduce to the familiar characteristic function of a Poisson random variable. The cases of a scaled indicator function \( (g = aI_A) \) and a simple function \( (g = \sum_{k=1}^N a_k I_{A_k}, A_k \text{ disjoint}) \) are similar. The proof of 4.13 in the general case follows by writing the measurable function \( g \) as the limit of a sequence of simple functions. Using the integral representation
of an infinitely divisible process, its characteristic functional can easily be obtained from 4.13.

Assume that $X_1$ and $X_2$ are two random variables with their distribution functions satisfying $F_1(x) \geq F_2(x)$ for all $x$. Then $\text{Pr}[X_1 > x] \leq \text{Pr}[X_2 > x]$, so that $X_2$ tends to have larger values than $X_1$. $X_2$ is said to be stochastically larger than $X_1$ [36, p. 84]. The following lemma deals with the stochastic ordering of the random variables $U_g$ under different distributions of $\Pi$.

**Lemma 4.2** [52] Suppose $\Lambda$ and $\tilde{\Lambda}$ are finite measures on $\mathcal{B}$ such that

$$\Lambda(\mathcal{X}) = \tilde{\Lambda}(\mathcal{X})$$

and $g : \mathcal{X} \to \mathbb{R}$ is a Borel measurable function such that for all $t \in \mathbb{R}$, $\tilde{\Lambda} \{g(x) > t\} \geq \Lambda \{g(x) > t\}$. Then we have

$$P_{\tilde{\Lambda}} \{U_g(\Pi) > v\} \geq P_{\Lambda} \{U_g(\Pi) > v\} \quad \forall v \in \mathbb{R}. \quad (4.15)$$

In other words, since $\tilde{\Lambda}$ distributes more mass than $\Lambda$ in regions where $g$ is larger, $P_{\tilde{\Lambda}}$ makes $U_g(\Pi)$ stochastically larger than it is under $P_{\Lambda}$. For the case $g = I_A$, 4.15 follows from the easily verifiable fact that if $\Upsilon$ is a Poisson random variable with parameter $\lambda$, $\text{Pr}[\Upsilon > n]$ is a strictly increasing function of $\lambda$. When $g$ is a simple function, Lemma 4.2 can be proved using the fact that if $\{U_m\}$ is a set of random variables, independent under two distributions $P$ and $\tilde{P}$ such that each $U_m$ is stochastically larger under $\tilde{P}$ than it is under $P$, then $\sum U_m$ is also stochastically larger under $\tilde{P}$ than under $P$. Making use of a limiting argument again, Lemma 4.2 is proved for the general case.

As in the case of 2-alternating capacity neighborhoods (described briefly at the end of this section) considered by Sadowsky [52], Lemmas 4.1 and 4.2 point to a stochastic ordering of the Poisson distributions corresponding to the $\epsilon$-contamination and total variation neighborhoods $L_j$, induced by the random variable $U_{\ln(d\Lambda_1/d\Lambda_0)}$.
with $\Lambda_j'$ as in 4.9 and 4.10. This stochastic ordering, coupled with the fact that the likelihood ratio of two Poisson distributions involves such a random variable, leads us to the identification of least favorable Poisson distributions.

**Theorem 4.1** Let $L_j, j = 0, 1$ be the $\epsilon$-contamination or total variation neighborhoods of nominal intensity measures $\tilde{\Lambda}_j$, with a sufficiently small $\epsilon$. Denote by $P_{\Lambda_j}$ the classes of Poisson distributions generated by the intensity measures in $L_j$. The pair of distributions $(P_{\Lambda_0'}, P_{\Lambda_1'}) \in P_{\Lambda_0} \times P_{\Lambda_1}$ corresponding to $(\Lambda_0', \Lambda_1')$ defined in 4.9 and 4.10 is least favorable for discrimination between $P_{\Lambda_0}$ and $P_{\Lambda_1}$.

**Proof** We can conclude from Lemmas 4.1 and 4.2 that $P_{\Lambda_0'}$ makes $U_{\ln \nu}$ stochastically larger than does any other distribution $P_{\Lambda_0} \in P_{\Lambda_0}$, and $P_{\Lambda_1'}$ makes $U_{\ln \nu}$ stochastically smaller than does any other distribution $P_{\Lambda_1} \in P_{\Lambda_1}$. Then there exists a random variable $Y$ and nondecreasing functions $f_0 \leq f_0' \leq f_1' \leq f_1$ such that the probability distributions of $f_j(Y)$ and $f_j'(Y)$ coincide with the distributions of $U_{\ln \nu}$ under $P_{\Lambda_j}$ and $P_{\Lambda_j'}$ respectively [26][36, p. 84]. Also it follows from 4.8 that any likelihood ratio test between $P_{\Lambda_0'}$ and $P_{\Lambda_1'}$ is of the form

$$\phi'(\pi) = \begin{cases} 
1 & U_{\ln \nu}(\pi) > \gamma, \\
q & U_{\ln \nu}(\pi) = \gamma, \\
0 & U_{\ln \nu}(\pi) < \gamma,
\end{cases} \quad (4.16)$$

where $l'(x) = \lambda_1'(x)/\lambda_0'(x)$. We then have

$$R(P_{\Lambda_0'}, \phi') = C_0 \{ P_{\Lambda_0'}(U_{\ln \nu} > \gamma) + qP_{\Lambda_0}^{-1}(U_{\ln \nu} = \gamma) \}$$

$$= C_0 \{ (1-q)P_{\Lambda_0'}(U_{\ln \nu} > \gamma) + qP_{\Lambda_0'}(U_{\ln \nu} \geq \gamma) \}$$

$$= C_0 \{ (1-q)\Pr(f_0'(Y) > \gamma) + q\Pr(f_0'(Y) \geq \gamma) \}$$

$$\geq C_0 \{ (1-q)\Pr(f_0(Y) > \gamma) + q\Pr(f_0(Y) \geq \gamma) \}$$
\[ R(P_{\lambda_0}, \phi') \]
for any \( P_{\lambda_0} \in \mathcal{P}_{\lambda_0} \). Similarly we can show that \( R(P_{\lambda_1}, \phi') \geq R(P_{\lambda_1}, \phi') \).

It follows from the theorem and equations 4.8 and 4.12 that the problem of minimax discrimination between the Poisson distribution classes \( \mathcal{P}_{\lambda_0} \) and \( \mathcal{P}_{\lambda_1} \) is solved by the test statistic

\[
\frac{dP_{\lambda_1}(\pi)}{dP_{\lambda_0}(\pi)} = \exp \left\{ -K + \int_{\mathcal{X}} \ln \min \left\{ c'' \max \left\{ c', \frac{d\lambda_1}{d\lambda_0}(x) \right\} \right\} \pi(dx) \right\},
\]

where \( K = \lambda_1(\mathcal{X}) - \lambda_0(\mathcal{X}) \).

Before concluding this section, we should remark that Sadowsky [52] considers the classes \( \mathcal{L}^{\text{cap}}_j \) of intensity measures bounded by 2-alternating capacities following Geraniotis and Poor [18]. 2-alternating capacities are generalized notions of measure: A set function \( v : \mathcal{B} \rightarrow [0, \infty] \) is a \( 2\)-alternating capacity when \( v(\phi) = 0 \), \( A \subset B \) implies \( v(A) \leq v(B) \), \( v \) is sequentially continuous for increasing sequences of sets and for decreasing sequences of closed sets and \( v(A \cup B) + v(A \cap B) \leq v(A) + v(B) \) [28].

The uncertainty classes are then given by

\[ \mathcal{L}^{\text{cap}}_j = \{ \Lambda_j | \Lambda_j(A) \leq v_j(A), \forall A \subset \mathcal{B}; \Lambda_j(\mathcal{X}) = v_j(\mathcal{X}) \} \]

with \( v_j, j = 0, 1 \) being 2-alternating capacities. If the space \( \mathcal{X} \) is compact, these capacity classes are very general and contain many common models of uncertainty including the \( \epsilon \)-contamination and total variation models. In this case, the results of Huber and Strassen [28] can be used to establish the existence of least favorable Poisson distributions as was done by Sadowsky [52]. However, since our interest lies in Poisson processes defined on the noncompact function space \( L_2[0, T] \), we cannot make use of this result directly. Thus we have reformulated it in terms of \( \epsilon \)-contamination and total variation uncertainty classes. Huber's theory [26] for these classes is appli-
cable to arbitrary spaces, and provides us not only the existence but also the above explicit characterization of the least favorable distributions.

4.4 Least Favorable Distributions for Linear Processes

Using the integral representation 2.17 for linear processes without a Gaussian component and the results of the previous section concerning Poisson random measures, we now obtain the minimax robust detector for linear processes.

**Theorem 4.2** Let $L_{j}, j = 0, 1$ be the $\epsilon$-contamination or total variation neighborhoods of nominal measures $\Lambda_{j}$, with a sufficiently small $\epsilon$. Denote by $P_{j}$ the classes of probability measures induced on the function space $X = L_{2}[0, T]$ by linear processes with projective limit measures in $L_{j}$.

Then the pair of distributions $(P_{0}, P_{1}) \in P_{0} \times P_{1}$ corresponding to $(\Lambda_{0}, \Lambda_{1})$ in 4.9 and 4.10 is least favorable for discrimination between $P_{0}$ and $P_{1}$.

**Proof** Suppose $P_{0}$ and $P_{1}$ are probability measures induced on $X$ by linear processes with projective limit measures $\Lambda_{0}$ and $\Lambda_{1}$. It is clear from 3.4 and 4.8 that under the hypothesis $H_{0}$ (and similarly under $H_{1}$), the likelihood ratio $(dP_{1}/dP_{0})(y)$ is identical in distribution to $(dP_{\Lambda_{1}}/dP_{\Lambda_{0}})(\pi)$ for some Poisson random measure $\pi$ satisfying $y = \int x[\pi(dx) - \Lambda_{0}(dx)]$. This equality in distribution is true in spite of the fact that the map from $N$ to $X$ defined by $\int x[\pi(dx) - \Lambda_{0}(dx)] = y$ is not one-to-one. Thus

$$E_{P_{0}}(\phi') = E_{P_{\Lambda_{0}}}(\phi') \geq E_{P_{0}}(\phi') = E_{P_{1}}(\phi'),$$

where $\phi'$ denotes a likelihood ratio test with the statistic $l'(y) = (dP_{1}'/dP_{0}') (y)$ and the inequality follows from Theorem 4.1. Similarly we have $E_{P_{1}}(1 - \phi') \geq E_{P_{1}}(1 - \phi')$. Thus $(P_{0}, P_{1})$ is least favorable. \qed
Thus the robust detector for linear processes corresponding to the classes $\mathcal{P}_0$ and $\mathcal{P}_1$ is the optimal detector for the linear processes with distributions $P'_0$ and $P'_1$ and has the representation formula
\[
\frac{dP'_1}{dP'_0}(y) = \exp \left\{ -K + \int_{\mathcal{X}} \ln \min \left\{ c'', \max \left\{ c', \frac{d\bar{\Lambda}_1}{d\Lambda_0}(x) \right\} \right\} \pi(dx) \right\}, \tag{4.18}
\]
where again $K = \bar{\Lambda}_1(\mathcal{X}) - \bar{\Lambda}_0(\mathcal{X})$.

4.5 Example

As an application of the above theorem, we consider the special case of robust discrimination between independent increment processes. These are infinitely divisible processes for which the projective limit measure $\Lambda$ is concentrated on the scaled and shifted unit step functions
\[
x_{t_0,v_0}(t) = \begin{cases} 
0 & t < t_0, \\
v_0 & t \geq t_0,
\end{cases} \quad t \in I. \tag{4.19}
\]

Denote by $A$ the set of all functions of the form 4.19 with $t_0 \in [t_1,t_2]$ and $v_0 \in B$, where $B$ is any Borel set in $\mathbb{R}$. Then
\[
\Lambda(A) = G([t_1,t_2] \times B) \tag{4.20}
\]
defines the projective limit in this case [19].

Suppose $G_0$ and $G_1$ are the measures corresponding to two independent increment processes with $G_1 \ll G_0$, and $\Lambda_0$ and $\Lambda_1$ are their projective limits. It can be shown that
\[
\frac{d\Lambda_1}{d\Lambda_0}(x_{t_0,v_0}) = \frac{dG_1}{dG_0}(t_0,v_0). \tag{4.21}
\]
As described in Section 2.3, any realization $x$ of an independent increment process without a Gaussian component can be expressed uniquely as the sum of a finite set of
functions $\{x_{t_k,v_k}\}_{k=1}^M$ of the form 4.19. Specifically, $(t_k,v_k)$ are the times of occurrence and heights of jumps in $x$. The set $\{x_{t_k,v_k}\}$ constitutes $\pi$, the realization of the Poisson random measure. Denoting by $P_0$ and $P_1$ the probability distributions induced on the function space by the processes with measures $G_0$ and $G_1$ respectively, from 3.5 and 4.21 we have

$$\frac{dP_1}{dP_0}(x) = e^{-(\Lambda_1(x) - \Lambda_0(x))} \prod_{k=1}^M \frac{dG_1}{dG_0}(t_k,v_k).$$

(4.22)

Suppose $G_j$ are absolutely continuous (with respect to the Lebesgue measure on $I \times \mathbb{R}$) with densities $g_j$. Because $\Lambda_j(x) = G_j(I \times \mathbb{R})$, the above equation can be seen to be identical to 3.2. This fact also follows directly from the one-to-one correspondence between a realization $x$ and the vector $(t,v)$.

It is clear from 4.20 that $\epsilon$-contamination and total variation neighborhoods of $\Lambda$ result from neighborhoods of $G$ of the same structure. Also the measures $\Lambda_j'$ in Lemma 4.1 correspond to the measures $G_j'$ identified directly via Huber's theory. Thus the likelihood ratio between $G_0'$ and $G_1'$ is a censored version of that between the nominal measures $\bar{G}_0$ and $\bar{G}_1$ and from Theorem 4.2, the robust detector for independent increment processes is of the form

$$\frac{dP_1'}{dP_0'}(x) = e^{-K} \prod_{k=1}^M \min \left\{ c'', \max \left\{ c', \frac{d\bar{G}_1}{d\bar{G}_0}(t_k,v_k) \right\} \right\}.$$

(4.23)

This form of the robust detector can also be inferred by using the Lévy-Ito representation of $x$ and the resulting one-to-one correspondence between $x$ and the Poisson random measure $\nu$ on $I \times \mathbb{R}$ and then applying Theorem 4.1. Note that this detector, as opposed to that in 4.18 for general linear processes, can actually be implemented. Again, this is because the times of occurrence of the jumps $t_k$ and their heights $v_k$ are directly observable from $x$ as illustrated in Figure 2.2. The above detector easily specializes to the Poisson case, with $\nu = 1$ and $\bar{G}_j(t,1) = \int_0^t \bar{\lambda}_j(s) ds$. 
Suppose for example that the two independent increment processes are homogeneous, with the nominal measures given by

\[
\bar{g}_0(t,v) = \frac{1}{\sqrt{2\pi}} e^{-v^2/2},
\]
\[
\bar{g}_1(t,v) = \frac{1}{\sqrt{2\pi}} e^{-v^2/8}.
\]

The jump height distributions in the noise alone case and signal-plus-noise case are both Gaussian with a variance of 1 and 4 respectively, while the rate of occurrence of the jumps are \(\bar{\lambda}_0(t) = \int \bar{g}_0(t,v) dv = 1\) and \(\bar{\lambda}_1(t) = 2\) respectively. Note that although the jump height distributions are taken to be Gaussian, the amplitude distributions of the independent increment processes are not; it follows from the characteristic function formula 2.4 for \(x(t)\) that its amplitude distribution is an infinite mixture of Gaussian distributions with different variances [16]. Assuming \(\epsilon\)-contamination uncertainty neighborhoods for the measures \(G_j\), it follows from 4.9 that the least favorable densities are of the form

\[
g_0'(t,v) = \begin{cases} 
\frac{(1-\epsilon)}{\sqrt{2\pi}} e^{-v^2/2} & v^2 < 8 \ln \epsilon''/3, \\
\frac{(1-\epsilon)}{\epsilon''\sqrt{2\pi}} e^{-v^2/8} & \text{otherwise},
\end{cases}
\]

\[
g_1'(t,v) = \begin{cases} 
\frac{(1-\epsilon)}{\sqrt{2\pi}} e^{-v^2/8} & v^2 > 8 \ln \epsilon'/3, \\
\frac{c(1-\epsilon)}{\sqrt{2\pi}} e^{-v^2/2} & \text{otherwise},
\end{cases}
\]

The constants \(\epsilon'\) and \(\epsilon''\) are obtained in terms of \(\epsilon\) from \(\int \int g_0'(t,v) dt dv = 1\) and \(\int \int g_1'(t,v) dt dv = 2\) respectively. For example, \(\epsilon = 0.1\) leads to \(\epsilon' = 1.45\) and \(\epsilon'' = 4.06\). The robust detector for this case is then given by

\[
\frac{dP_1}{dP_0}(x) = e^{-1} \prod_{k=1}^{M} \min \left\{ 4.06, \max \left\{ 1.45, e^{3v^2/8} \right\} \right\}.
\]

The least favorable densities and the robust detector for the total variation uncertainty case can be obtained in a similar fashion.
Chapter 5

Robust Suboptimal Detection in Linear Processes

Suboptimal performance criteria are often used for the evaluation of detection procedures, because of their greater tractability compared to the optimal criteria. We consider robust linear process detection with a suboptimal distance criterion in this chapter. This criterion is a simple generalization of the commonly used signal-to-noise ratio and was studied in the minimax robustness framework by Poor [46]. As in the case of robustness with respect to risk discussed in the previous chapter, the approach here is again to identify least favorable distributions.

The problem of detecting a known signal in additive Gaussian noise can be given a geometric interpretation in a Hilbert space of functions that is uniquely determined by the covariance function of the noise process. This Hilbert space, which is called a reproducing kernel Hilbert space (RKHS), is isomorphic to the linear Hilbert space of random variables generated by the process. Thus most statistical problems concerning Gaussian processes, and linear problems involving non-Gaussian processes, can be treated in this RKHS framework. By introducing the RKHS associated with the characteristic functional of a process, Duttweiler and Kailath [15] develop a general theory for non-Gaussian detection. In Section 5.3, we extend this approach to the robust distance detection problem by using Barton and Poor's RKHS theory for robust matched filtering [7].
5.1 A Distance Criterion

As in section 4.1, we consider the problem of testing the hypotheses

\[ H_0 : X \sim P_0, \]
\[ H_1 : X \sim P_1, \]  \hspace{1cm} (5.1)

where \( X \) is a random variable with values in \( \mathcal{X} \) and \( P_0 \) and \( P_1 \) are probability distributions on \( (\mathcal{X}, \mathcal{B}) \) satisfying \( P_1 \ll P_0 \). We restrict the class of decision rules to those randomized tests of the form

\[ \phi_g(x) = \begin{cases} 
1 & g(x) > \eta, \\
p & g(x) = \eta, \\
0 & g(x) < \eta,
\end{cases} \] \hspace{1cm} (5.2)

where \( \phi_g(x) \) is the probability of accepting \( H_1 \) when \( x \) is observed, the statistic \( g : \mathcal{X} \to \mathbb{R} \) is any measurable map and the constants \( \eta \) and \( p \) are chosen to give the desired risk or error probability performance.

As a performance measure for the above hypothesis testing problem, we use the statistical distance between \( H_0 \) and \( H_1 \) at the output of the statistic \( g \) for the test \( \phi_g(x) \) defined by [46]

\[ S_g(P_0, P_1) = \frac{(E_{P_1}[g] - E_{P_0}[g])^2}{\text{Var}_{P_0}(g)}, \] \hspace{1cm} (5.3)

when \( \text{Var}_{P_0}(g) > 0 \) and 0 otherwise. We consider only those statistics \( g \) for which the statistical averages in the above definition are finite. The functional \( S \) is a generalization of the common definition of signal-to-noise ratio. Since a high signal-to-noise ratio usually implies a low probability of error, a suboptimum performance criterion is to maximize \( S \) over the class of all valid statistics \( g \). Rewriting \( S \) as

\[ S_g(P_0, P_1) = \frac{(E_{P_0}[g] - E_{P_0}[g]E_{P_0}[l])^2}{\text{Var}_{P_0}(g)} = \frac{\text{Cov}_{P_0}^2(g, l)}{\text{Var}_{P_0}(g)}, \] \hspace{1cm} (5.4)
where \( l(x) = (dP_1/dP_0)(x) \), and applying the Cauchy-Schwarz inequality yields

\[
S_g(P_0, P_1) \leq S_\nu(P_0, P_1) = \text{Var}_{P_0}(l) \tag{5.5}
\]

provided \( l \) is a valid statistic. Thus, if \( P_0 \) and \( P_1 \) are known, the likelihood ratio test \( \phi_l(x) \) is the decision rule that maximizes the distance \( S \) and hence the suboptimum distance criterion leads to the same test as the optimum Neyman-Pearson and minimum error probability criteria.

Suppose next that the distributions \( P_0 \) and \( P_1 \) are known only to be members of some classes \( \mathcal{P}_0 \) and \( \mathcal{P}_1 \) respectively. Analogous to the robust hypothesis testing problem formulated in terms of risk in Section 4.1, we now consider the minimax problem

\[
\sup_{g \in \mathcal{G}} \inf_{(P_0, P_1) \in \mathcal{P}} S_g(P_0, P_1), \tag{5.6}
\]

where \( \mathcal{P} = \mathcal{P}_0 \times \mathcal{P}_1 \) and \( \mathcal{G} = \{ g : \text{Var}_{P_0}(g) < \infty, E_{P_0} g < \infty, \forall (P_0, P_1) \in \mathcal{P} \} \). The notion of least favorability in terms of the distance criterion plays an important role in the solution of the above problem.

**Definition 5.1.1** [46] A pair of distributions \( P'_j \in \mathcal{P}_j \) is said to be least favorable in terms of \( S \) (for tests between \( \mathcal{P}_0 \) and \( \mathcal{P}_1 \)) if \( l' \in \mathcal{G} \) and

\[
S_{l'}(P_0, P'_1) \leq S_{l'}(P_0, P_1), \quad \forall (P_0, P_1) \in \mathcal{P} \tag{5.7}
\]

where \( l''(x) = (dP'_1/dP_0)(x) \).

This definition is analogous to the Definition 4.1.1 of least favorability in terms of risk. If \( (P'_0, P'_1) \in \mathcal{P} \) is least favorable in terms of \( S \), then we have

\[
S_g(P'_0, P'_1) \leq S_{l'}(P'_0, P'_1) \leq S_{l'}(P_0, P_1), \quad \forall \{g, (P_0, P_1)\} \in \mathcal{G} \times \mathcal{P}. \tag{5.8}
\]
Then \( \{l', (P'_0, P'_1)\} \) is called a saddle point solution \([5]\) to the minimax problem 5.6. The existence of a saddle point implies that

\[
\sup_{g \in \mathcal{G}} \left( \inf_{(P_0, P_1) \in \mathcal{P}} S_g(P_0, P_1) \right) = \inf_{(P_0, P_1) \in \mathcal{P}} \sup_{g \in \mathcal{G}} S_g(P_0, P_1) = \inf_{(P_0, P_1) \in \mathcal{P}} \text{Var}_{P_0}(l). \quad (5.9)
\]

Thus if a least favorable pair exists, it can be obtained by searching for a pair \((P'_0, P'_1)\) whose likelihood ratio variance under \(P'_0\) is a minimum over the class \(\mathcal{P}\). The following Lemma gives the sufficient conditions under which the result of such a search is in fact the least favorable pair.

**Lemma 5.1** \([46]\) Let \(\mathcal{P}_0\) and \(\mathcal{P}_1\) be two convex, nonoverlapping classes of probability distributions on \((\mathcal{X}, \mathcal{B})\) such that for each \((P_0, P_1) \in \mathcal{P}\), we have \(P_1 \ll P_0\) and \(\text{Var}_{P_0}(l) < \infty\). Suppose further that a pair \((P'_0, P'_1) \in \mathcal{P}\) exists such that \(l' \in \mathcal{G}\). Then \((P'_0, P'_1)\) is least favorable in terms of \(S\) if and only if

\[
\inf_{(P_0, P_1) \in \mathcal{P}} \text{Var}_{P_0}(l) = \text{Var}_{P'_0}(l'). \quad (5.10)
\]

Note that the necessity of the condition 5.10 follows directly from 5.9. The fact that this condition is also sufficient when \(\mathcal{P}\) is convex, depends on the convexity properties of \(S\) and the quantity \(\text{Var}_{P_0}(l)\). Another sufficient condition, in more familiar terms, for the existence of a least favorable pair is given in the next Lemma.

**Lemma 5.2** \([46]\) Suppose the classes \(\mathcal{P}_0\) and \(\mathcal{P}_1\) satisfy the conditions of Lemma 5.1. If \((P'_0, P'_1)\) is least favorable in terms of risk, then it is also least favorable in terms of \(S\).

The proof of Lemma 5.2 involves establishing \((P'_0, P'_1)\) as the minimizer of \(\text{Var}_{P_0}(l)\) over \(\mathcal{P}\), and then using Lemma 5.1.
5.2 Minimax Distance Linear Process Detection

We now consider robust discrimination (with respect to the distance criterion $S$) of probability measures induced on the function space by linear processes. Although the robust suboptimal detector for linear processes can be obtained directly by using the results of Section 4.4, our approach here, considering Poisson random measures first and using the integral representation of linear processes to arrive at the same detector, parallels that in the previous chapter. This method enables us to generalize similar results for Poisson processes on the real line.

Let $\mathcal{L}_j, j = 0, 1$ be the $\epsilon$-contamination or total variation neighborhoods of nominal intensity measures $\tilde{\Lambda}_j$ with a sufficiently small $\epsilon$. Denote by $\mathcal{P}_{\Lambda_j}$ the classes of Poisson distributions generated by the intensity measures in $\mathcal{L}_j$.

For the case of Poisson distributions, explicit expressions for the distance measure $S$ can be obtained, making plausible the analytical derivation of least favorable distributions directly from Definition 5.7. Denoting $l' = dP_{\Lambda_1}/dP_{\Lambda_0}$ and $l'_\Lambda = d\Lambda'_1/d\Lambda'_0$, it follows from 4.8 and 4.14 that

$$E_{P_{\Lambda_j}}[(l')^t] = E_{P_{\Lambda_j}}[\exp \{-tK + tU_{l\Lambda_j}\}] = \exp \left\{-tK + \int_X \left[(l'_\Lambda(x))^t - 1\right] \Lambda_j(dx) \right\}, \quad (5.11)$$

where $t$ is any real number and $K = \Lambda'_1(\mathcal{X}) - \Lambda'_0(\mathcal{X})$. Using the above equation for the cases $t = 1$ and $t = 2$, we have

$$S_{P}(P_{\Lambda_0}, P_{\Lambda_1}) = \frac{\exp \left(\int_X \left[l'_\Lambda(x) - 1\right] \Lambda_1(dx)\right) - \exp \left(\int_X \left[l'_\Lambda(x) - 1\right] \Lambda_0(dx)\right)}{\exp \left(\int_X \left[(l'_\Lambda(x))^2 - 1\right] \Lambda_0(dx)\right) - \exp \left(2\int_X \left[l'_\Lambda(x) - 1\right] \Lambda_0(dx)\right)}. \quad (5.12)$$

Geraniotis and Poor [18] point out that the above functional is convex in $\Lambda_1$ but not in $(\Lambda_0, \Lambda_1)$ and hence that only uncertainty in $\Lambda_1$ can be treated analytically. For Poisson processes on a compact interval of the real line, they obtain a least favorable (in terms of $S$) intensity measure $\Lambda'_1$ in a capacity class.
An indirect approach to deal with uncertainty in both $\Lambda_0$ and $\Lambda_1$ can be considered by applying Lemma 5.1. The minimization problem 5.10 in the lemma is tractable for Poisson distributions. It can be shown using 5.11 that $\text{Var}_{P_{\Lambda_0}}(I) = a \exp\{b \text{Var}_{\hat{\Lambda}_0}(\hat{I}_\Lambda)\} - 1$, where $\hat{\Lambda}_j = \Lambda_j / \Lambda_j(\mathcal{X})$ are the probability measures obtained by normalizing $\Lambda_j$ and the nonnegative constants $a$ and $b$ are given by $a = \exp\{K^2/\Lambda_0(\mathcal{X})\}$ and $b = \Lambda_1^2(\mathcal{X})/\Lambda_0(\mathcal{X})$. Hence the solution of the problem

$$\inf_{(P_{\Lambda_0}, P_{\Lambda_1}) \in \mathcal{P}_{\Lambda_0} \times \mathcal{P}_{\Lambda_1}} \text{Var}_{P_{\Lambda_0}}(I)$$

corresponds to that of

$$\inf_{(\Lambda_0, \Lambda_1) \in \mathcal{L}_0 \times \mathcal{L}_1} \text{Var}_{\Lambda_0}(\hat{I}_\Lambda).$$

Since the classes of probability measures obtained by normalizing the members of $\mathcal{L}_j$ are convex, it can be inferred from Lemmas 5.1 and 5.2 that the latter problem is solved by $(\Lambda_0', \Lambda_1')$ defined in 4.9 and 4.10. Another application of Lemma 5.1 to conclude the least favorability of the corresponding distributions $(P_{\Lambda_0'}, P_{\Lambda_1'})$ is not possible; though the uncertainty classes $\mathcal{L}_j^v$ and $\mathcal{L}_j^y$ of the intensity measures $\Lambda_j$ are convex, the Poisson distribution classes $\mathcal{P}_{\Lambda_j}$ are not. The convex combination of two Poisson distributions may not even be Poisson.

Another indirect approach to deduce least favorability in terms of $S$ for uncertainty in $\Lambda_0$ and $\Lambda_1$ is via Lemma 5.2 which we explore next. For any set $A$ of a linear vector space, denote by $CO(A)$ the convex cover of $A$: the intersection of all convex sets containing $A$. $CO(A)$ is thus the smallest convex set which contains $A$.

**Lemma 5.3** [5] $CO(A)$ coincides with the set of all convex combinations of elements in $A$:

$$CO(A) = \left\{ \sum_{j=1}^{n} a_j x_j; n = 1, 2, \ldots, a_j \geq 0, \sum_{j=1}^{n} a_j = 1, x_j \in A \right\}.$$
Let \((P_{\Lambda_0'}, P_{\Lambda_1'})\) be the pair of distributions in Theorem 4.1 that are least favorable with respect to risk.

**Theorem 5.1** \((P_{\Lambda_0'}, P_{\Lambda_1'})\) are also least favorable with respect to \(S\) for discrimination between \(\mathcal{P}_{\Lambda_0}\) and \(\mathcal{P}_{\Lambda_1}\).

**Proof** Let \(\mathcal{M}\) be the space of all probability measures on \((\mathcal{X}, \mathcal{B})\) and \(\mathcal{M}'\) be the set of finite signed measures on \((\mathcal{X}, \mathcal{B})\), that is the linear space generated by \(\mathcal{M}\). The convex covers \(CO(\mathcal{P}_{\Lambda_0})\) and \(CO(\mathcal{P}_{\Lambda_1})\) of the classes \(\mathcal{P}_{\Lambda_0}\) and \(\mathcal{P}_{\Lambda_1}\) clearly belong to the subspace \(\mathcal{M}\).

It is easy to see that \((P_{\Lambda_0'}, P_{\Lambda_1'})\) are also least favorable with respect to risk for discrimination between \(CO(\mathcal{P}_{\Lambda_0})\) and \(CO(\mathcal{P}_{\Lambda_1})\). Suppose \(P_0\) is an arbitrary distribution in \(CO(\mathcal{P}_{\Lambda_0})\). From Lemma 5.3, we can write \(P_0 = \sum_{j=1}^{n} a_j P_{\Lambda_0'}\), where \(P_{\Lambda_0'} \in \mathcal{P}_{\Lambda_0}\).

\[
R(P_0, \phi') = C_0 \sum_{j=1}^{n} a_j E_{P_{\Lambda_0'}} [\phi'] 
\leq C_0 \sum_{j=1}^{n} a_j E_{P_{\Lambda_1}} [\phi'] = R(P_{\Lambda_0'}, \phi'),
\]

where \(\phi'\) is a likelihood ratio test between \(P_{\Lambda_0'}\) and \(P_{\Lambda_1'}\). Similarly we have \(R(P_1, \phi') \leq R(P_{\Lambda_1'}, \phi')\) for any \(P_1 \in CO(\mathcal{P}_{\Lambda_1})\).

Finally it follows from Lemma 5.2 that \((P_{\Lambda_0'}, P_{\Lambda_1'})\) are also least favorable with respect to \(S\) for discrimination between \(CO(\mathcal{P}_{\Lambda_0})\) and \(CO(\mathcal{P}_{\Lambda_1})\) and since \((P_{\Lambda_0'}, P_{\Lambda_1'}) \in \mathcal{P}_{\Lambda_0} \times \mathcal{P}_{\Lambda_1}\), the proof is complete. \(\square\)

We have assumed that the classes \(\mathcal{P}_{\Lambda_0}\) and \(\mathcal{P}_{\Lambda_1}\) are sufficiently further apart so that \(CO(\mathcal{P}_{\Lambda_0})\) and \(CO(\mathcal{P}_{\Lambda_1})\) are nonoverlapping.

**Theorem 5.2** The pair of distributions \((P_0', P_1') \in \mathcal{P}_0 \times \mathcal{P}_1\) induced on \(\mathcal{X}\) by linear processes with projective limits \((\Lambda_0', \Lambda_1')\) in 4.9 and 4.10 is least favorable with respect to \(S\) for discrimination between \(\mathcal{P}_0\) and \(\mathcal{P}_1\).
Proof Suppose $P_j \in \mathcal{P}_j$, $j = 0, 1$ is any probability measure induced on the function space $\mathcal{X}$ by a linear process with projective limit $\Lambda_j$. As in the proof of Theorem 4.2, the distribution of the likelihood ratio $l' = dP_1'/dP_0'$ under $P_j$ is identical to that of $dP_{\Lambda_1}/dP_{\Lambda_0}$ under $P_{\Lambda_j}$. Equating the first and second order moments of these two distributions, and using Theorem 5.1 we obtain the result.

5.3 RKHS Methods for Robust Suboptimal Detection

Suppose $\{x(t), t \in I\}$ is a zero-mean, second-order stochastic process. In Chapter 2, we considered the Hilbert space $L_2(x)$ spanned by $x(t)$ consisting of all finite linear combinations $\sum_{k=1}^{n} c_k x(t_k)$ and their limits in the norm $\|u\|^2 = E|u|^2$. $L_2(x)$ has the inner product $(u_1, u_2)_{L_2} = E[u_1^* u_2]$ where $^*$ denotes a complex conjugate, and contains all finite variance random variables obtained by linear operations on $x(t)$. In order to distinguish this space from the space of random variables obtained by nonlinear operations on the process, we will denote it by $LL_2(x)$ here (for linear $L_2$).

Define a Hilbert space $H(K)$ of functions by the mapping

$$J(u) = E[ux(s)] \triangleq f(s)$$

and norm $\|f\|_{H(K)}^2 = E|u|^2$, where $u \in LL_2(x)$ and $K$ is the covariance function of $x(t)$. It can be shown that $J$ is one-to-one and defines an isometry between $LL_2(x)$ and $H(K)$; If the functions $f_1$ and $f_2$ are the images of $u_1$ and $u_2$ under $J$, $(f_1, f_2)_{H(K)} = (u_1, u_2)_{LL_2}$. Comparing the equation $E[x(s)x(t)] = K(s, t)$ with 5.13 we note that $K(\cdot, t) \in H(K)$ and indeed the functions $\{K(\cdot, t), t \in I\}$ span the space $H(K)$. Furthermore, since $(f(\cdot), K(\cdot, t))_{H(K)} = E[ux(t)] = f(t)$, $H(K)$ is a reproducing kernel Hilbert space (RKHS) with a reproducing kernel $K(s, t)$.

Next consider the Hilbert space $NL_2(x)$ (for nonlinear $L_2$) generated by $x(t)$ consisting of all finite products $\prod_{k=1}^{n} x(t_k)$, all linear combinations of these products,
and their limits in the norm $\| v \|^2 = E|v|^2$. Again, the resulting inner product is $\langle v_1, v_2 \rangle_{NL_2} = E[v_1 \cdot v_2]$. Define a Hilbert space $H(C)$ of functionals by the mapping

$$\tau(v) = E[v \exp i(x, \eta)] \triangleq g(\eta)$$

and norm $\| g \|^2_{H(C)} = E|v|^2$, where $v \in NL_2(x)$, $(x, \eta) = \int_I x(t)\eta(t)dt$ and $C$ is the characteristic functional of $x(t)$ defined by $C(\psi) = E \exp i(x, \psi)$. The map $\tau$ is also one-to-one and isometric [15]; If the functionals $g_1$ and $g_2$ are the images of $v_1$ and $v_2$ under $\tau$, $\langle g_1, g_2 \rangle_{H(C)} = \langle v_1, v_2 \rangle_{NL_2}$. Since $E[\exp\{i(x, \eta)\}\exp\{-i(x, \psi)\}] = C(\eta - \psi)$, we have

$$\langle g(\cdot), C(\cdot - \psi) \rangle_{H(C)} = \langle v, \exp\{-i(x, \psi)\} \rangle_{NL_2} = E[v \exp i(x, \psi)] = g(\psi).$$

Hence $H(C)$ is an RKHS with a reproducing kernel $C'(\eta, \psi) = C(\eta - \psi)$. Also note that since $\tau(1) = C(\psi)$, we have $C(\psi) \in H(C)$.

Before discussing the RKHS approach to robust matched filtering, we introduce some of the terminology needed. For further details, see Barton and Poor [7]. A filter $h$ is defined as a pair $\{(h_k(n), t_k(n))\}_{k=1}^n, n = 1, 2, \ldots$ consisting of a double sequence of coefficients and a double sequence of time instants. Then, for a covariance function $K$, the function $hK \in H(K)$ is defined by

$$hK(\cdot) = \lim_{n \to \infty} \sum_{k=1}^n h_k K(\cdot, t_k).$$

where we dropped the argument $n$ of $h_k$ and $t_k$ for convenience. Denote by $F(K)$ the class of all filters for which the above limit exists in $H(K)$. For any $s \in H(K)$

$$\langle s, hK \rangle_{H(K)} = \lim_{n \to \infty} \sum_{k=1}^n h_k s(t_k).$$

If $x(t)$ is a process with a covariance function $K$, $h(x) \in LL_2(x)$ is defined by

$$h(x) = \lim_{n \to \infty} \sum_{k=1}^n h_k x(t_k).$$

(5.17)
It is easy to see that \( J(h(x)) = hK(\cdot) \).

Consider the hypothesis testing problem

\[
H_0 : \quad x(t) \text{ has zero-mean and covariance } K, \\
H_1 : \quad x(t) \text{ has mean } s(t) \text{ and covariance } K,
\]

where \( s(t) \) is a known signal in \( H(K) \). In the linear detection problem, the test statistic \( h(x) \) is of the form 5.17. The signal-to-noise ratio (SNR) for this problem is given by

\[
\Delta_h(s, K) \triangleq \frac{(E_1[h(x)])^2}{\text{Var}_0 h(x)} = \frac{(\lim_{n \to \infty} \sum_{k=1}^{n} h_k s(t_k))^2}{\lim_{n \to \infty} \sum_{k,k'=1}^{n} h_k h_{k'} K(t_{k'}, t_k)} = \frac{\langle s, hK \rangle^2_{H(K)}}{\langle hK, hK \rangle_{H(K)}},
\]

(5.19)

Any filter \( \hat{h} \) that satisfies

\[
\Delta_{\hat{h}}(s, K) = \sup_{h \in F(K)} \Delta_h(s, K)
\]
is referred to as a \textit{matched filter} and it follows from the Cauchy-Schwarz inequality that \( \hat{h}K = cs \) for some constant \( c \) and the maximum SNR is given by \( \Delta_{\hat{h}}(s, K) = \langle s, s \rangle_{H(K)} \).

Suppose now that the pair \((s, K)\) is known only to belong to an uncertainty class \( \mathcal{U} \) which is convex and and has the property that for all pairs \((s, K) \in \mathcal{U}, s \in H(K)\). Denote by \( \mathcal{K} (\mathcal{U}) \) the class of all covariance functions in \( \mathcal{U}; \)

\[
\mathcal{K} (\mathcal{U}) = \{ K : (s, K) \in \mathcal{U} \text{ for some } s \}.
\]

The goal of robust matched filtering is to find a linear detector \( \hat{h}' \) that satisfies

\[
\inf_{(s,K) \in \mathcal{U}} \Delta_{\hat{h}'}(s, K) = \sup_{h \in F(\mathcal{K}(\mathcal{U}))} \inf_{(s,K) \in \mathcal{U}} \Delta_h(s, K),
\]

(5.20)
and it is often achieved by searching for a saddle point solution \( \{ \hat{h}', (s', K') \} \in F(\mathcal{K}(U)) \times U \) satisfying

\[
\Delta_{h}(s', K') \leq \Delta_{h'}(s, K') \leq \Delta_{h}(s, K) \quad \forall \{ h, (s, K) \} \in F(\mathcal{K}(U)) \times U. \quad (5.21)
\]

Let us next formulate the problem of robustness with respect to the distance criterion in RKHS terms. Denote by \( C_0 \) and \( C_1 \) the characteristic functionals of the distributions \( P_0 \) and \( P_1 \) in the hypothesis testing problem 5.1, and by \( NL_2(P_0) \) the nonlinear \( L_2 \) space of random variables under \( P_0 \). Note that if \( l = dP_1/dP_0 \) satisfies \( \text{Var}_{P_0} l < \infty \), then \( l \in NL_2(P_0) \) and

\[
\tau_0(l) = E_{P_0}[l \exp i(x, \psi)] = E_{P_1}[\exp i(x, \psi)] = C_1(\psi). \quad (5.22)
\]

From 5.4,

\[
S_g(C_0, C_1) \triangleq S_g(P_0, P_1) = \frac{\langle g - E_{P_0} g, l - 1 \rangle_{NL_2(P_0)}}{\langle g - E_{P_0} g, g - E_{P_0} g \rangle_{NL_2(P_0)}} = \frac{\langle \tau_0(g - E_{P_0} g), C_1 - C_0 \rangle_{H(C_0)}}{\langle \tau_0(g - E_{P_0} g), \tau_0(g - E_{P_0} g) \rangle_{H(C_0)}} \quad (5.23)
\]

using the isometry between \( NL_2(P_0) \) and \( H(C_0) \) induced by the map \( \tau_0 \) given by

\[
\tau_0(\nu) = E_{P_0}[\nu \exp i(x, \psi)] = \langle \nu, \exp -i(x, \psi) \rangle_{NL_2(P_0)}. \quad (5.24)
\]

Note that since \( \tau_0(l - 1) = C_1 - C_0 \), any nonlinear statistic of the form \( c(l - 1) \), \( c \) being a constant, maximizes \( S_g(C_0, C_1) \). We set \( c = 1 \) without loss of generality (cf. also Theorem B.1).

The robust suboptimal detection problem 5.6 can be stated in terms of the characteristic functionals as

\[
\sup_{g \in g} \inf_{(C_0, C_1) \in C} S_g(C_0, C_1) \quad (5.25)
\]

where \( C = C_0 \times C_1 \) and the classes of characteristic functionals \( C_0 \) and \( C_1 \) correspond to the classes \( \mathcal{P}_0 \) and \( \mathcal{P}_1 \) of distributions respectively. \( C_j, j = 0, 1 \) is isomorphic to \( \mathcal{P}_j \) due to the linearity of the expectation in the definition of a characteristic functional.
Comparing 5.23 with 5.19, we can draw the following analogy between the distance detection and matched filtering problems. The functional $C_1 - C_0$ corresponds to the signal $s$, the nonlinear statistic $g(x) = E_{P_0} g(x)$ to the linear statistic $h(x)$ and $C_0$ to the covariance $K$. Further, note that the image of $g(x) - E_{P_0} g(x)$ under the map $\tau_0$ is determined by $C_0(\psi)$, just as the image of $h(x)$ under $J$, (which is $hK$) is determined by $K$. Since $l$ is assumed to satisfy $\text{Var}_{P_1} l < \infty$ (for all pairs $(P_0, P_1)$), from 5.22, $C_1$ clearly belongs to $H(C_0)$ and so does $C_1 - C_0$. The analogy described here is illustrated in Figure 5.1.

![Diagram](image)

Figure 5.1: Analogy between the RKHS approaches to matched filtering and detection using distance criterion.

When $C_0$ and $C_1$ are convex, necessary and sufficient conditions analogous to Theorem 4.4 of Barton and Poor [7] for the existence of a saddle point solution to the problem 5.25 are stated in Appendix B. When $C_0$ and $C_1$ are not convex (as in the
case of linear processes corresponding to $\epsilon$-contamination and total variation neighborhoods of projective limits), the idea of taking convex covers as in Section 5.2 proves to be useful. In Appendix B, we prove the RKHS version of Theorem 5.2 concerning the least favorability with respect to the distance criterion of linear processes.
Chapter 6

Conclusions

We obtained the least favorable pair of distributions for robust discrimination between linear processes. The robust detector for Neyman-Pearson criterion involves a likelihood ratio test between these two distributions. Although we used the representation formula for the likelihood ratio as a tool to derive this result, its scope is not limited by that of the representation formula itself; if any other likelihood ratio formula (e.g. one that is explicitly computable in terms of the observation) is known, the robust detector still involves applying this formula to our least favorable distributions.

The robust detector for the distance criterion is identical to that for the optimal criterion. Thus in this case, consideration of the suboptimal criterion does not lead to a system with reduced complexity. However, the difficulty of implementing the robust detector in continuous-time points to the study of approximate discrete-time detectors, a task that might be simpler for the suboptimal case. In light of the fact that continuous-time robust detectors for the two criteria are the same, it can be hoped that the performance of approximate discrete-time implementations for the optimal and suboptimal criteria is not very different.

Although the notion that the characteristic functional plays the same role for non-Gaussian processes as the covariance function does for Gaussian processes is elegant, its utility is often limited by the difficulty in obtaining expressions for the characteristic functionals. The characteristic functional approach is quite useful for
linear processes, as demonstrated by our derivation of the likelihood ratio formula and the RKHS approach to the robust distance detection of linear processes.

Because we used only those properties of linear processes that they share with infinitely divisible processes in obtaining the results, most of the work here applies equally well to the general class of infinitely divisible processes. One of the limitations of our uncertainty model, as with the others for Poisson processes and random measures referred here, is that it is too structured; because we assume $c$-contamination and total variation neighborhoods of uncertainty in the projective limit, the corresponding neighborhoods in the probability measures only include those of linear processes.

This thesis addressed some of the more theoretical aspects of robust detection for linear processes. Further study of this problem can be carried out in different directions. On the theoretical side, the derivation of an implementable likelihood ratio formula is of prime importance. Another important problem from the standpoint of implementation is that of finding the optimal linear detector for the discrimination of linear processes, and studying its robustness properties. Consideration of other uncertainty models, e.g., the more general Prokhorov neighborhoods of uncertainty for the projective limit, and neighborhoods of uncertainty defined directly in the space of probability measures, is also relevant. The least favorable distributions identified here may also be robust for the sequential detection problem (cf. [26][52]). On the practical side, derivation of discrete-time approximations of the likelihood ratio test would be interesting. The distribution of the likelihood ratio is known [11] and thus the performance of the approximate versions could easily be compared with that of the optimal.
Appendix A

Likelihood Ratio Formula for Linear Processes

In this appendix, we provide an alternative derivation of the likelihood ratio formula for the discrimination of the linear processes $y_0(t)$ and $y_1(t)$. Using the integral representation for infinitely divisible processes, we can write

$$y_j = \int_{\mathcal{X}} x \left[ \Pi_j(dx) - \Lambda_j(dx) \right] \quad j = 0, 1 \quad (A.1)$$

where $\Pi_j$ is a Poisson random measure on $\mathcal{X} \triangleq L_2[0,T]$ with intensity $\Lambda_j$. Recall from Section 2.3 that the above equation is only symbolic; $\int_{\mathcal{X}} x \Pi_j(dx)$ denotes any random process such that the random variables $(\int_{\mathcal{X}} x \Pi_j(dx), \psi)$ and $\int_{\mathcal{X}} (x, \psi) \Pi_j(dx)$ are equal in distribution for all $\psi \in \mathcal{X}$, and $\int_{\mathcal{X}} x \Lambda_j(dx)$ is an element of $\mathcal{X}$ satisfying $(\int_{\mathcal{X}} x \Lambda(dx), \psi) = \int_{\mathcal{X}} (x, \psi) \Lambda(dx)$. Here $(x, \psi) = \int_0^T x(t) \psi(t) \, dt$ is the inner product in $\mathcal{X}$.

The characteristic functional of the process $y_j$ is given by

$$C_j(\psi) = E e^{i \left( \int_{\mathcal{X}} x \Pi_j(dx) - \Lambda_j(dx) \right),\psi} = e^{-i \int_{\mathcal{X}} (x, \psi) \Lambda_j(dx)} E e^{i \int_{\mathcal{X}} (x, \psi) \Pi_j(dx)}$$

Note that the expectation involves random variables of the form 4.13 and hence from 4.14, we have

$$C_j(\psi) = \exp \left\{ \int_{\mathcal{X}} \left[ e^{i(x,\psi)} - 1 - i(x,\psi) \right] \Lambda_j(dx) \right\} \quad (A.2)$$

Suppose $P_j$ is the probability measure induced by $y_j$ on $\mathcal{X}$. Note that if $l(y)$ is
the likelihood ratio between \( y_1(t) \) and \( y_0(t) \), i.e., \( l(y) = (dP_1 / dP_0)(y) \), then we have

\[
E_{P_0} \ l(y) \ e^{i(y, \psi)} = \int \chi e^{i(y, \psi)} \frac{dP_1}{dP_0}(y) P_0(dy) = C_1(\psi).
\]

On the other hand, any random variable \( l(y) \) satisfying the equation

\[
E_{P_0} \ l(y) \ e^{i(y, \psi)} = C_1(\psi)
\]

for all \( \psi \) in \( \chi \) must be the likelihood ratio between \( y_1(t) \) and \( y_0(t) \) [19, 15]. This follows from the fact that all \( B \) measurable functions can be approximated by finite combinations of functions \( e^{i(y, \psi)} \) with different \( \psi \).

Denote by \( l(y) \) the random variable

\[
l(y) = \exp \left\{ -K + \int \chi \ln \frac{d\Lambda_1}{d\Lambda_0}(x) \Pi_0(dx) \right\}
\]

(A.4)

where \( K = \Lambda_1(\chi) - \Lambda_0(\chi) \). We show that \( l(y) \) is the likelihood ratio between \( y_1(t) \) and \( y_0(t) \) under the hypothesis \( H_0 \), by verifying that equation A.3 is true. A similar proof for the case of independent increment processes is given by Gikhman and Skorokhod [19]. Substituting for \( l(y) \) from the above equation,

\[
E_{P_0} \ l(y) \ e^{i(y, \psi)} = E \left\{ \exp \left\{ -K + \int \chi \ln \frac{d\Lambda_1}{d\Lambda_0}(x) \Pi_0(dx) + \right. \right.
\[
\left. \left. i \left( \int \chi [\Pi_0(dx) - \Lambda_0(dx)], \psi \right) \right\} \right\}
\]

\[
= \exp \left\{ -K - \int \chi i(x, \psi) \Lambda_0(dx) \right\} .
\]

\[
E \left\{ \exp \left\{ \int \chi \left[ \ln \frac{d\Lambda_1}{d\Lambda_0}(x) + i(x, \psi) \right] \Pi_0(dx) \right\} \right\}
\]

By assumption (see p. 30), the intensity measures \( \Lambda_0(.) \) and \( \Lambda_1(.) \) have identical means, i.e., \( \int \chi (x, \psi) \Lambda_0(dx) = \int \chi (x, \psi) \Lambda_1(dx) \) for all \( \psi \) in \( \chi \). Also evaluating the expectation using an argument similar to that used above for obtaining the characteristic functional formula, we get
$$E_{P_0} I(y) e^{i(y, \psi)} = \exp \left\{ -K - \int_{\mathcal{X}} i(x, \psi) \Lambda_1(dx) \right\} \exp \left\{ \int_{\mathcal{X}} \left[ e^{\ln(d\Lambda_1/d\Lambda_0)(x)} + i(x, \psi) \right] - 1 \right\} \Lambda_0(dx)$$

Since $e^{\ln(d\Lambda_1/d\Lambda_0)(x)} \cdot \Lambda_0(dx) = \Lambda_1(dx)$, it follows that

$$E_{P_0} I(y) e^{i(y, \psi)} = \exp \left\{ \int_{\mathcal{X}} \left[ e^{i(x, \psi)} - 1 - i(x, \psi) \right] \Lambda_1(dx) \right\} = C_1(\psi)$$

and hence equation A.3 is verified.
Appendix B

RKHS Approach to the Robust Detection of Linear Processes using the Distance Criterion

Making use of the analogy established in Section 5.3 between the robust matched filtering problem and the robust distance detection problem, we prove here the RKHS version of Theorem 5.2. Analogous to Theorem 4.4 of Barton and Poor [7], we have

**Theorem B.1** \( \{l', (C'_0, C'_1)\} \in \mathcal{G} \times \mathcal{C} \), where \( \mathcal{C} \) is convex and \( C_1 \in H(C_0) \)

\( \forall (C_0, C_1) \in \mathcal{C} \), is a saddle point solution for the problem

\[
\sup_{g \in \mathcal{G}} \inf_{(C_0, C_1) \in \mathcal{C}} S_g(C_0, C_1)
\]

if and only if

\[
2(C_1 - C_0, C_1' - C_0') - (C_1' - C_0', C_1' - C_0') - (\tau_0(l' - 1), C_1' - C_0') \geq 0 \quad (B.1)
\]

for all \( (C_0, C_1) \in \mathcal{C} \).

All inner products in B.1 are taken in the Hilbert space \( H(C'_0) \). The proof of the theorem is identical to that of Theorem 4.4 in [7] and is omitted.

We now apply the above general theorem to the robust detection of linear processes using the distance criterion. Denote by \( \mathcal{C}_0 \) and \( \mathcal{C}_1 \) the classes of characteristic functionals corresponding to the neighborhoods \( \mathcal{P}_0 \) and \( \mathcal{P}_1 \) of probability measures induced by linear processes on the function space \( \mathcal{X} \). Let \( (C'_0, C'_1) \) be the characteristic functionals of the risk least favorable distributions \( (P'_0, P'_1) \) in Theorem 4.2.

**Theorem B.2** \( (C'_0, C'_1) \) is least favorable with respect to \( S \) for discrimination between \( \mathcal{C}_0 \) and \( \mathcal{C}_1 \).
Proof. The assumption that $\text{Var}_{P_0}(l) < \infty$ for all $(P_0, P_1) \in \mathcal{P}$ (cf. Lemma 5.1) implies that $C_1 \in H(C_0)$ for all $(C_0, C_1) \in C_0 \times C_1$. Then all members of $CO(C_0) \times CO(C_1)$ also have this property. We show this in two steps. First, suppose $D_0 \in CO(C_0)$. Then from Lemma 5.3, we can write $D_0 = \sum_{k=1}^{N} a_k C_{0k}$ with $C_{0k} \in C_0$. Since any $C_1 \in C_1$ belongs to $H(C_{0k})$ for all $k$, it follows from a theorem (Part I, Section 6) of Aranzazj [1] that $C_1 \in H(D_0)$. Next, writing any element of $D_1 \in CO(C_1)$ as $D_1 = \sum_{k=1}^{M} b_k C_{1k}$ with $C_{1k} \in C_1$, we note that since each $C_{1k}$ is in $H(D_0)$, we have $D_1 \in H(D_0)$, $\forall (D_0, D_1) \in CO(C_0) \times CO(C_1)$.

To show that $(C_0', C_1')$ is least favorable, from Theorem B.1 it is sufficient to verify equation B.1. $(P_0', P_1')$ is least favorable with respect to risk for discrimination between $CO(P_0) \times CO(P_1)$. From the comment after Lemma 5.2,

$$\inf_{(P_0, P_1) \in CO(P_0) \times CO(P_1)} \text{Var}_{P_0}(l) = \text{Var}_{P_0'}(l').$$

From the proof of Theorem 1 of Poor [46], this implies that

$$\inf_{(P_0, P_1) \in CO(P_0) \times CO(P_1)} [2 E_{P_1}(l') - E_{P_0}(l')] = 2 E_{P_1}(l') - E_{P_0}(l').$$

Rewriting the above equation, we obtain

$$2 E_{P_0}[(l-1)(l'-1)] - E_{P_0}[(l'-1)^2] \geq E_{P_0}[(l'-1)^2]. \quad (B.2)$$

for all $(P_0, P_1) \in CO(P_0) \times CO(P_1)$. Denoting $\xi = \frac{dP_0}{dP_0'}$, we have

$$E_{P_0}[(l-1)(l'-1)] = E_{P_0'}[\xi(l-1)(l'-1)]$$

$$= \langle \xi(l-1), l'-1 \rangle_{NL_2(P_0')},$$

$$= \langle \tau_{0}'(\xi(l-1)), C_1' - C_0' \rangle_{H(C_0')}$$

$$= \langle \tau_{0}(l-1), C_1' - C_0' \rangle_{H(C_0')}$$

$$= \langle C_1 - C_0, C_1' - C_0' \rangle_{H(C_0')}.$$
using the isometry between the spaces $NL_2(P_0')$ and $H(C'_0)$, and the fact that $(l' - 1)$ is a real random variable. Similarly we have $E_{P_0'}[(l' - 1)^2] = \langle \pi_0(l' - 1), C'_1 - C'_0 \rangle_{H(C'_0)}$ and $E_{P_0'}[(l' - 1)^2] = \langle C'_1 - C'_0, C'_1 - C'_0 \rangle_{H(C'_0)}$. The condition B.1 now follows easily from B.2. \qed
Bibliography


[22] B. V. Gnedenko and A. N. Kolmogorov. Limit Distributions for Sums of Indep-

[23] U. Grenander. Stochastic processes and statistical inference. Arkiv For Matem-

86, 1961.


[28] P. J. Huber and V. Strassen. Minimax tests and the Neyman-Pearson lemma for


