APPLIED MATHEMATICS AND STATISTICS LABORATORIES
STANFORD UNIVERSITY
CALIFORNIA

THE FUNCTION SPACE POINT OF VIEW IN TIME SERIES ANALYSIS

BY
EMANUEL PARZEN

TECHNICAL REPORT NO. 47
March 29, 1963

PREPARED UNDER CONTRACT Nonr-225(21)
(NR-042-993)
FOR
OFFICE OF NAVAL RESEARCH

THE FUNCTION SPACE POINT OF VIEW IN TIME SERIES ANALYSIS

by

Emanuel Parzen

TECHNICAL REPORT NO. 47

March 29, 1963

PREPARED UNDER CONTRACT Nonr-225(21)

(NR-042-993)

FOR

OFFICE OF NAVAL RESEARCH

APPLIED MATHEMATICS AND STATISTICS LABORATORIES
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

THE FUNCTION SPACE POINT OF VIEW IN TIME SERIES ANALYSIS*

by

Emanuel Parzen**

Stanford University

0. Summary

This is an expository paper which seeks to establish and show the value of the following assertion: the concept of a time series is equivalent to the idea of a probability measure on a function space.

1. What is a time series?

The general point of view adopted in analyzing a time series or a succession of observations $\{X(t), t \in T\}$, depending on the parameter t (often representing time), is the following.

A set T of values of t, called the index set of the time series, is preassigned; these are the times when observations are possible. The set T may be finite or infinite. It is possible to develop much of the theory of time series without placing any restriction on the nature of the index set T. However two important cases are when

$$T = \{0, \pm 1, \pm 2, \ldots\} \quad \text{or} \quad T = \{0, 1, 2, \ldots\},$$

---

*To be presented at the 1963 Joint Automatic Control Conference Workshop on "Stochastic Processes", to be held at the University of Minnesota during June of 1963. This paper was prepared with the partial support of the Office of Naval Research (Nonr-225-21). Reproduction in whole or in part is permitted for any purpose of the United States Government. This paper is based on material to appear in a forthcoming book [Parzen (1964)].

**It is with great pleasure that I dedicate this paper to Professor Charles Loewner on his 70th birthday.

in which case the time series is said to be a discrete parameter process, or when

$$T = \{t: -\infty < t < \infty\} \qquad \text{or} \qquad T = \{t: t \geq 0\} ,$$

in which case the time series is said to be a continuous parameter process.

At each point t in T, a number $X(t)$ may be observed. This number is a random variable in the sense that its value depends on chance and enjoys a probability distribution described by the one-dimensional distribution function

$$F_{X(t)}(x) = \text{Probability}[X(t) \leq x] , \qquad -\infty < x < \infty ,$$

or the one-dimensional characteristic function

$$\varphi_{X(t)}(u) = \int_{-\infty}^{\infty} \exp[iux] \, dF_{X(t)}(x) , \qquad -\infty < u < \infty .$$

More generally, for any integer n and n points $t_1, t_2, \ldots, t_n$ in T, the n observations $X(t_1), \ldots, X(t_n)$ which can be made at these times are jointly distributed random variables whose joint probability law is specified by either (i) the joint distribution function, given for all real numbers $x_1, \ldots, x_n$ by

$$F_{X(t_1),\ldots,X(t_n)}(x_1, \ldots, x_n)$$

(1.1)

$$= \text{Probability}[X(t_1) \leq x_1, X(t_2) \leq x_2, \ldots, X(t_n) \leq x_n]$$

or (ii) the joint characteristic function, given for all real numbers $u_1, \cdots, u_n$ by

$$\varphi_{X(t_1),\ldots,X(t_n)}(u_1, \cdots, u_n)$$

$$= E[\exp i\left(u_1X(t_1) + \ldots + u_nX(t_n)\right)]$$

(1.2)

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp i(u_1x_1 + \ldots + u_nx_n)$$

$$dF_{X(t_1),\ldots,X(t_n)}(x_1, \cdots, x_n) .$$

The distribution function in Equation (1.1) and the characteristic function in Equation (1.2) are said to be n-dimensional, since they represent the joint probability law of n random variables.

The point of view embodied in the foregoing discussion may be summarized as follows.

A time series is a jointly distributed family of random variables $\{X(t), t \in T\}$, indexed by a parameter t varying in an index set T. Time series analysis is concerned with the statements that can be made about a time series, knowing only all the finite dimensional distributions of the form of (1.1).

It should be noted that the phrase "a stochastic process" is often used to describe a numerical-valued random phenomenon that arises through a process which is developing in time in a manner controlled by probabilistic laws [see Parzen (1962)]. Mathematically, a stochastic process is represented by a collection of random variables $\{X(t), t \in T\}$. Thus, in a

sense, the notions of a time series and of a stochastic process are equivalent; every time series is a stochastic process and vice versa.

What distinguishes the theory of time series from the theory of stochastic processes is a certain difference in emphasis.

One approach to the problem of developing mathematical models for empirical phenomena evolving in accord with probabilistic laws is to characterize such phenomena in terms of the behavior of their first and second moments. This approach has found important applications in statistical communications and control theory and in time series analysis. When a stochastic process is being studied in terms of its moments it is often called a time series. Consequently one would make the following definition.

A time series $\{X(t), t \in T\}$ is a family of random variables (stochastic process) with finite second moments.

It is to be emphasized that it should not be assumed that a stochastic process which arises in practice necessarily has finite second moments and is therefore a time series. In particular, recent research [by Mandelbrot] has raised the question of whether certain series of economic observations involving price changes of stocks and commodities possess finite second moments.

Two important characteristics of a time series $\{X(t), t \in T\}$ are its mean value function $m(\cdot)$, defined for all t in T by

$$m(t) = E[X(t)] ,$$

and its covariance kernel $K(\cdot,\cdot)$, defined for all s and t in T

by*

$$K(s,t) = \mathrm{Cov}[X(s),X(t)] \ .$$

The importance of the mean value function and the covariance kernel derives from several facts:

(i) it is usually much easier to find the mean value function and the covariance kernel of a stochastic process than it is to find its complete probability law;

(ii) nevertheless, many important questions about a stochastic process can be answered on a basis of a knowledge only of its mean value function and covariance kernel;

(iii) for example, the continuity, differentiability and integrability properties of the covariance kernel lead to corresponding properties for the time series;

(iv) further, there exists an important class of time series, the normal stochastic processes, whose complete probability law is known once one knows its mean value function and covariance kernel.

Normal processes: A stochastic process $\{X(t),\ t \in T\}$ is said to be a normal process if for any integer n and any subset $\{t_1,t_2,\ldots,t_n\}$ of T the n random variables $X(t_1), \ldots, X(t_n)$ are jointly normally

*In studying the general theory of time series, it is often convenient to admit complex valued random variables. The covariance kernel is then defined by

$$K(s,t) = E[X(s)\overline{X}(t)] - m(s)\overline{m}(t)$$

where $\overline{X}(t)$ denotes the complex conjugate of $X(t)$ and $\overline{m}(t)$ denotes the complex conjugate of $m(t)$.

distributed in the sense that their joint characteristic function is given by, for any real numbers $u_1, u_2, \ldots, u_n$,

$$\varphi_{X(t_1),X(t_2),\ldots,X(t_n)}(u_1,u_2,\ldots,u_n) = E[\exp i(u_1X(t_1)+\ldots+u_nX(t_n))]$$

$$= \exp\left\{i \sum_{j=1}^{n} u_j E[X(t_j)] - \frac{1}{2} \sum_{j,k=1}^{n} u_j u_k \operatorname{Cov}[X(t_j),X(t_k)]\right\}$$

$$= \exp\left\{i \sum_{j=1}^{n} u_j m(t_j) - \frac{1}{2} \sum_{j,k=1}^{n} u_j u_k K(t_j,t_k)\right\} .$$

Normal processes play a basic role in time series analysis for a number of reasons:

(i) Because of the central limit theorem, many random variables which arise in applications of probability theory may be considered to be approximately normally distributed; similarly, many stochastic processes can be approximated by normal processes;

(ii) Because of the mathematical tractibility of normal random variables, many questions can be more simply treated for normal processes than for other kinds of time series;

(iii) Normal processes have the useful closure property that any time series [such as $\int_0^t X(s)\,ds$, $X'(t)$, $X(t+1) - X(t)$] derived by means of linear operations on a normal process, is itself a normal process;

(iv) For a normal process, one obtains a knowledge of the complete probability law of the process from a knowledge of the mean value function $m(\cdot)$ and the covariance kernel $K(\cdot,\cdot)$. Conversely it may be shown that if $m(\cdot)$ and $K(\cdot,\cdot)$ are the mean value function and covariance kernel of some time series, then there is a (unique) normal process with this mean value function and covariance kernel.

2. A time series as a probability measure on function space.

A time series $\{X(t), t \in T\}$ is for many purposes best regarded as an observation on a random phenomenon each of whose possible outcomes is a real valued function with domain T. In other words, a time series $\{X(t), t \in T\}$ is a collection of real valued functions with domain T, one of which is observed whenever a sample is taken. The observed function is therefore called a sample function, or realization, of the time series.

Given an index set T, we let Ω_T denote the space of all real valued functions with domain T. A point ω belonging to Ω_T is a function on T whose value at a point t in T we denote by $\omega(t)$; consequently, we may write $\omega = \{\omega(t), t \in T\}$.

The problem of analyzing a time series $\{X(t), t \in T\}$ can be expressed as the problem of finding the probability function $P[\cdot]$, defined on suitable subsets of Ω_T called the measurable subsets, which describes the probability distribution of possible values of the time series (in the intuitive sense that for any measurable subset A of Ω_T, $P[A]$ is approximately the relative frequency of observations in a very long sequence of independent observations of the time series which are members of A.)

In this section we show how given the finite dimensional probability laws of a time series one can construct a probability measure on a suitable family of subsets of the function space Ω_T. This probability measure will enable us to define the notion of the probability density functional of a time series which plays a central role in modern time series analysis [see Parzen (1962, 1963)].

Let $\mathcal{A}$ be a collection of subsets of Ω_T which contains Ω_T as a member. A function P, with domain $\mathcal{A}$, is said to be a <u>probability</u> measure if it possesses the following properties:

Axiom 1. For every $A \in \mathcal{A}$, P[A] is well defined and is a non-negative real number; in symbols,

$$P[A] \geq 0 ;$$

Axiom 2. $P[\Omega_T] = 1$;

Axiom 3. For any sequence of disjoint sets $A_1, A_2, \ldots$ belonging to $\mathcal{A}$, whose union $\bigcup_{n=1}^{\infty} A_n$ belongs to $\mathcal{A}$,

$$P[\bigcup_{n=1}^{\infty} A_n] = \sum_{n=1}^{\infty} P[A_n] ;$$

the sets $\{A_n\}$ are said to be disjoint (or non-overlapping) if for any two distinct indices j and k the intersection of A_j and A_k is empty,

$$A_j A_k = \emptyset$$

where $\emptyset$ denotes the empty set.

Axiom 3 is referred to as the countable additivity or sigma-additivity property of the probability function P.

The sets belonging to $\mathcal{A}$ are called <u>measurable sets</u> or <u>events</u>. An event A is said to occur if the function representing the actual time series observed belongs to A.

In order to guarantee that the usual operations of analysis will lead to measurable sets, it is necessary to require that the family of measurable sets be a sigma-field.

A collection $\mathcal{A}$ of subsets of Ω_T is called a sigma-field if it has the following properties:

Axiom 1. Ω_T belongs to $\mathcal{A}$ (written symbolically: $\Omega_T \in \mathcal{A}$);

Axiom 2. If A belongs to $\mathcal{A}$ then the complement A^c belongs to $\mathcal{A}$ (written symbolically: $A \in \mathcal{A}$ implies $A^c \in \mathcal{A}$);

Axiom 3. For any sequence $A_1, A_2, \ldots$ belonging to $\mathcal{A}$, the union $\bigcup_{n=1}^{\infty} A_n$ belongs to $\mathcal{A}$ (written symbolically: $\{A_n\} \subset \mathcal{A}$ implies $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$).

In words, a sigma-field is a family of sets which contains the entire space Ω_T and is closed under the operations of forming complements and countable unions. It then follows that it is closed under the operation of forming countable intersections:

Property 4. For any sequence $A_1, A_2, \ldots$ belonging to $\mathcal{A}$, the intersection $\bigcap_{n=1}^{\infty} A_n \in \mathcal{A}$

An important example of a sigma-field is the family of all subsets of Ω_T.

The question naturally arises: cannot all subsets of Ω_T be made events so that it is never necessary to consider a sigma-field $\mathcal{A}$ smaller than the family of all subsets of Ω_T? Unfortunately the general answer to this question is in the negative; if we desire the probability

function P to be countably additive, it is usually the case that the family $\mathcal{A}$ of events cannot contain all subsets of Ω_T. For example, even in the simple case that T consists of a single point so that Ω_T is just the real line (the set of all real numbers ω satisfying $-\infty < \omega < \infty$), there is no probability function P defined on all subsets of Ω_T that agrees with the ordinary notion of length on the subintervals of Ω_T (see Halmos (1950), p. 70). Usually, when Ω_T is the real line, one adopts as the family of events the family $\mathcal{B}$ of Borel sets, where $\mathcal{B}$ is defined as the smallest sigma-field containing as members all intervals.

The idea that, for each t, $X(t)$ is a random variable can now be made precise by the following definition: for each t in T, $X(t)$ is a function on Ω_T whose value at a point ω in Ω_T, denoted by $X(t,\omega)$, is given by the value at t of the function ω:

$$X(t,\omega) = \omega(t) \ . \tag{2.1}$$

In order to regard $X(t)$ as a random variable, there must exist

(i) a family $\mathcal{A}$ of subsets of Ω_T, called the measurable sets, such that for all t in T and all real numbers x

$$\{\omega: X(t,\omega) < x\} \in \mathcal{A} \ , \tag{2.2}$$

and (ii) a probability measure P with domain $\mathcal{A}$.

A function X with domain Ω_T is said to be $\mathcal{A}$-<u>measurable</u> if for every real number x

$$\{\omega: X(\omega) < x\} \in \mathcal{A} .$$

The problem at hand is to find a sigma-field of subsets of Ω_T such that each random variable $X(t)$ is $\mathcal{A}$-measurable.

In the function space Ω_T, there is a smallest sigma-field of events which should belong to the sigma-field of measurable sets. Let $\mathcal{A}_T$ denote the smallest sigma-field of subsets of Ω_T which contains all sets of the form

$$\{\omega: \omega(t) < x\}$$

where t is a point in T and x is a real number. It is clear that, for every t in T, $X(t)$ is $\mathcal{A}_T$-measurable.

Consequently, given a probability measure P on $\mathcal{A}_T$, one can define by (2.1) a time series $\{X(t), t \in T\}$ consisting of random variables with domain Ω_T and the finite dimensional probability distributions of the time series would be given by the formula

$$F_{X(t_1),\ldots,X(t_n)}(x_1,\ldots,x_n) = \text{Probability}[X(t_1) \leq x_1,\ldots,X(t_n) \leq x_n] \tag{2.3}$$
$$= P[\{\omega \in \Omega_T: \omega(t_j) \leq x_j \quad \text{for } j = 1, 2, \ldots, n\}] .$$

Basic to the theory of time series analysis is the fact that the converse holds (which follows from a celebrated theorem proved by Kolmogorov (1933).

Kolmogorov's celebrated existence theorem regarding the probability measure on function space Ω_T induced by a stochastic process $\{X(t), t \in T\}$.

Let $\{X(t),\ t \in T\}$ be a stochastic process with preassigned finite dimensional probability distributions. Then there exists a unique probability measure P_X on the sigma-field $\mathcal{A}_T$ satisfying (2.3).

To prove Kolmogorov's theorem, one considers a somewhat more general problem.

Let T be an index set. Given a family of finite dimensional characteristic functions,

$$\{\varphi_{t_1,\ldots,t_n}(u_1, \cdots, u_n), \quad n \text{ an integer and } t_1, \cdots, t_n \text{ points in } T\}, \tag{2.4}$$

what conditions need this family satisfy in order that there exist a stochastic process $\{X(t),\ t \in T\}$ whose true finite dimensional characteristic functions coincide with the given set (2.4):

$$\varphi_{X(t_1),\ldots,X(t_n)}(u_1, \cdots, u_n) = \varphi_{t_1,\ldots,t_n}(u_1, \cdots, u_n). \tag{2.5}$$

In view of (2.5), it is obvious that the given set (2.4) must be <u>mutually consistent</u> in the sense that (i) if $\alpha_1, \cdots, \alpha_n$ is a permutation of $1, \cdots, n$, then for any points $t_1, \cdots, t_n$ and real numbers $u_1, \cdots, u_n$,

$$\varphi_{t_{\alpha_1},\ldots,t_{\alpha_n}}(u_{\alpha_1}, \cdots, u_{\alpha_n}) = \varphi_{t_1,\ldots,t_n}(u_1, \cdots, u_n), \tag{2.6}$$

since the order in which the random variables are listed is irrelevant,

and (ii) if $m < n$,

$$(2.7) \quad \varphi_{t_1,\ldots,t_m}(u_1,\ldots,u_m) = \varphi_{t_1,\ldots,t_m,t_{m+1},\ldots,t_n}(u_1,\ldots,u_m,0,\ldots,0) .$$

The content of Kolmogorov's theorem is that these consistency conditions are the only conditions that need be imposed.

In order to give a precise statement of Kolmogorov's theorem we introduce the notion of a semi-infinite interval.

A subset C of Ω_T is called a <u>semi-infinite interval</u> if it is of the form

$$(2.8) \quad C = \{\omega \in \Omega_T: \omega(t_1) \leq x_1, \ \ldots \ , \omega(t_n) \leq x_n\}$$

for some integer n, points $t_1, \ \ldots \ , t_n$ belonging to T, and real numbers $x_1, \ \ldots \ , x_n$. Note that to specify a semi-infinite rectangle one must specify an integer n, n points $t_1, \ \ldots \ , t_n$ in the index set T, and n real numbers $x_1, \ \ldots \ , x_n$.

<u>Example 2A</u>.

<u>A semi-infinite interval</u>. Let $T = [0,1]$ so that Ω_T consists of all functions on the interval 0 to 1. Let $n = 11$, $t_j = j/10$ for $j = 0, 1, \ \ldots \ , 10$, and $x_j = j/2$ for $j = 0, 1, \ \ldots \ , 10$. Then

$$C = \{\omega \in \Omega(0,1): \omega(j/10) \leq j/2 \quad \text{for} \quad j = 0, 1, \ \ldots \ , 10\} .$$

Consider the following functions defined on the interval 0 to 1:

$$f_1(t) = t + 1$$

$$f_2(t) = t^2 ,$$

$$f_3(t) = \sin 5t .$$

It may be shown that $f_1(\cdot)$ does not belong to C, while $f_2(\cdot)$ and $f_3(\cdot)$ do belong to C.

Let P be a function defined on semi-infinite intervals in Ω_T as follows: for C given by (2.8)

$$P[C] = F_{t_1,\ldots,t_n}(x_1, \cdots, x_n) \tag{2.9}$$

where $F_{t_1,\ldots,t_n}(x_1, \cdots, x_n)$ is the distribution function corresponding to the characteristic function $\varphi_{t_1,\ldots,t_n}(x_1, \cdots, x_n)$.

Kolmogorov's theorem: There exists a unique probability measure P on $\mathcal{A}_T$ whose values on semi-infinite intervals satisfies (2.9). Further, define a family of functions $\{X(t), t \in T\}$ on Ω_T as follows: for each $\omega \in \Omega_T$, the value of $X(t)$ at ω, denoted $X(t,\omega)$, is given by

$$X(t,\omega) = \omega(t) , \qquad \text{the value at } t \text{ of the function } \omega .$$

Then $\{X(t), t \in T\}$ is a stochastic process, defined on the probability space $(\Omega_T, \mathcal{A}_T, P)$ whose finite dimensional characteristic functions satisfy (2.5).

The proof of Kolmogorov's theorem requires a background in measure theory which is beyond the scope of this paper [for a proof, see Kolmogorov (1933), p. 29, or Loève (1960), p. 93].

Example 2B.

An application of Kolmogorov's theorem. Consider a time series $\{X(t),\ t \in T\}$ defined by the formula

$$X(t) = \eta\,\varphi(t), \qquad t \in T, \tag{2.10}$$

where η is a random variable (with finite second moment) and $\varphi(\cdot)$ is a non-random function (for example, $\varphi(t) = t$).

The statement that η is a random variable is not completely explicit; we also want to know the space Ω on which η is defined as a function.

The random variables $X(t)$ defined by (2.10) are functions on the same space Ω as is η. Since this space may be unknown, it is sometimes useful to redefine $X(t)$ as follows.

Given the probability distribution of η, the family of random variables $\{X(t),\ t \in T\}$ induces a probability measure P on $\mathcal{A}_T$ by means of the formula

$$\begin{aligned} &P[\{\omega \in \Omega_T\colon \omega(t_1) \le x_1, \ \dots\ , \omega(t_n) \le x_n\}] \\ &\quad = \mathrm{Prob}[X(t_1) < x_1, \ \dots\ , X(t_n) < x_n] \qquad (2.11) \\ &\quad = \mathrm{Prob}[\eta\,\varphi(t_1) < x_1, \ \dots\ , \eta\,\varphi(t_n) < x_n]\ . \end{aligned}$$

Next, the random variables $\{X(t),\ t \in T\}$ can be redefined to be functions on Ω_T by the formula

$$X(t,\omega) = \omega(t) \quad . \tag{2.12}$$

The new family of random variables $\{X(t),\ t \in T\}$ obtained from the definition (2.12) can be identified with the old family obtained from definition (2.10) since they have the same finite dimensional distributions.

By the procedure just described, a time series $\{X(t),\ t \in T\}$ defined by an explicit formula such as (2.10) can be regarded as being equivalent to a probability measure P on the function space Ω_T.

In particular, suppose that the random variable η is normal with mean μ and variance σ^2. Then $X(\cdot)$, defined by (2.10) is a normal process with mean value function

$$m(t) = E[X(t)] = \mu\, \varphi(t) \ , \tag{2.13}$$

and covariance kernel

$$K(s,t) = \sigma^2\, \varphi(s)\, \varphi(t) \ . \tag{2.14}$$

Any normal process $\{X(t),\ t \in T\}$ with the foregoing mean value function and covariance kernel induces the same probability measure P on function space Ω_T.

Using the representation theory of time series (Parzen (1961), p. 962), it may be shown that a time series $\{X(t),\ t \in T\}$ with mean value function and covariance kernel given by (2.13) and (2.14) respectively may be represented in the form of (2.10), where η is now a random variable

whose domain is the function space Ω_T. Consequently when considering time series of the form of (2.10) one may assume that all random variables are functions on Ω_T.

3. Probability density functionals and orthogonal measures.

The probability theory of time series is concerned with investigating the structure of a time series $\{X(t), t \in T\}$ whose corresponding probability measure P on the function space Ω_T is assumed known. The statistical theory of time series is concerned with a time series $\{X(t), t \in T\}$ whose probability measure P is not known exactly but is only known to belong to a class of probability measures $\{P_\theta, \theta \in \Phi\}$ on Ω_T; the class of possible probability measures P_θ can be assumed to be indexed by a parameter θ varying in a parameter set Φ. Consequently an important step in developing a theory of time series analysis is to examine the relations that can exist between two probability measures P_1 and P_2 with a common domain $\mathcal{A}$.

Absolutely continuous and orthogonal probability measures. Let Ω be a set (the sample description space) and let $\mathcal{A}$ be a sigma-field of subsets of Ω. Let P_1 and P_2 be probability measures with domain $\mathcal{A}$.

We say that P_1 is absolutely continuous with respect to P_2 if for every set A in $\mathcal{A}$

$$(3.1) \qquad P_2(A) = 0 \quad \text{implies} \quad P_1(A) = 0 .$$

We write $P_1 << P_2$ if P_1 is absolutely continuous with respect to P_2; the motivation for this notation is the idea that P_1 is small whenever P_2 is small.

We say that P_1 and P_2 are equivalent, denoted $P_1 \equiv P_2$, if each is absolutely continuous with respect to the other; in symbols, $P_1 \equiv P_2$ if and only if $P_1 << P_2$ and $P_2 << P_1$.

In order that P_1 not be absolutely continuous with respect to P_2 it is necessary and sufficient that there exist a set A in $\mathcal{A}$ such that

$$P_2(A) = 0 \quad \text{and} \quad P_1(A) > 0 . \tag{3.2}$$

We say that P_1 and P_2 are orthogonal (or perpendicular), denoted $P_1 \perp P_2$, if there exists a set A in $\mathcal{A}$ such that

$$P_2(A) = 0 \quad \text{and} \quad P_1(A) = 1 . \tag{3.3}$$

One can regard (3.2) as the extreme case of not being absolutely continuous.

It may be shown that if $P_1 << P_2$, then probability statements in terms of P_1 can be expressed in terms of P_2; more precisely, if $P_1 << P_2$, then there exists a function p, called the probability density function of P_1 with respect to P_2, such that for any A in $\mathcal{A}$,

$$P_1(A) = \int_A p \, dP_2 . \tag{3.4}$$

In words, (3.4) says that to evaluate $P_1(A)$ one integrates the function p with respect to the measure P_2 over the set A. More generally,

for any $\mathcal{A}$-measurable function g which is integrable with respect to P_1,

$$E_{P_1}[g] = \int_\Omega g \, dP_1 \tag{3.5}$$

is finite, then there holds the transformation formula

$$E_{P_1}[g] = \int_\Omega g \, p \, dP_2 \, . \tag{3.6}$$

Three problems preliminary to time series analysis. To develop a general theory of time series analysis, one must begin with an understanding of the relations that can exist between two probability measures P_1 and P_2 on a function space. Consequently, we may speak of three problems preliminary to time series analysis:

(i) determine whether two given probability measures P_1 and P_2 are orthogonal,

(ii) determine whether one is absolutely continuous with respect to the other,

(iii) if $P_2 << P_1$, determine the probability density functional or Radon-Nikodym derivative, denoted

$$p_{2,1} = \frac{dP_2}{dP_1} \, .$$

One aim of modern time series analysis is to develop ways of determining answers to these questions.

4. Signal detection and likelihood ratios.

In this section we examine the problem of detecting a signal in noise, and show how the proper formulation of this problem requires a consideration of the relations that exist between probability measures.

Many observed time series can be represented as the sum of a signal process and a noise process. More precisely, let T be a set of points, called the index set, such that at each t in T one has made (or one can make) an observation, denoted $X(t)$. The set of observations $\{X(t),\ t \in T\}$ is a function on T which is assumed to be the sum of two other functions $\{S(t),\ t \in T\}$ and $\{N(t),\ t \in T\}$:

$$X(t) = S(t) + N(t) , \qquad t \in T .$$

We call $\{S(t),\ t \in T\}$ the signal since it is supposed to represent the true value of the quantity being measured, while $\{N(t),\ t \in T\}$ is called the noise since it represents "errors", "fluctuations", or "residuals" by which the observed function $\{X(t),\ t \in T\}$ differs from the desired function $\{S(t),\ t \in T\}$.

The aim of time series analysis is to infer, from the observations, information about one or more features of the signal. The aspects of the signal in which one is interested depends on the assumptions one makes about the structure of the signal and noise processes. In this section we consider the important problem of detecting a signal in the presence of noise.

Let $\{S(t),\ t \in T\}$ and $\{N(t),\ t \in T\}$ be time series, called respectively the signal process and the noise process. Given an observed time series $\{X(t),\ t \in T\}$, one desires to test the hypothesis

$$H_0\colon X(\cdot) = N(\cdot) , \qquad \text{noise alone is present} ,$$

against the alternative hypothesis

$$H_1: X(\cdot) = S(\cdot) + N(\cdot) , \qquad \text{signal plus noise is present ,}$$

by choosing a subset R_1 of the sample space Ω_T, of possible realizations of the time series $\{X(t), t \in T\}$, which will be the rejection region for H_0; that is, one says signal plus noise is present if the observed time series $\{X(t), t \in T\}$ belongs to R_1, and one says that noise alone is present if $\{X(t), t \in T\}$ does not belong to R_1.

In this section, we suppose that the probability distribution of $\{X(t), t \in T\}$ under the hypotheses H_0 and H_1 are well defined; in this case we say that H_0 and H_1 are simple hypotheses. We may then introduce probability measures P_N and P_{S+N} defined on the measurable subsets B of the sample space Ω_T by

$$\begin{aligned} P_N[B] &= \text{Prob}\,[\{N(t), t \in T\} \in B] \\ &= \text{Prob}\,[\{X(t), t \in T\} \in B | H_0] , \end{aligned} \tag{4.1}$$

$$\begin{aligned} P_{S+N}[B] &= \text{Prob}\,[\{S(t) + N(t), t \in T\} \in B] \\ &= \text{Prob}\,[\{X(t), t \in T\} \in B | H_1] . \end{aligned} \tag{4.2}$$

Example 4A.

A specified signal in normal noise. One important case in which H_0 and H_1 are simple hypotheses is when the following assumptions hold.

Assumption on the index set: The index set T is a finite set $\{t_1, t_2, \ldots, t_n\}$.

Assumption on the noise process: The noise process $\{N(t), t \in T\}$ is assumed to possess finite second moments, to have zero means:

$$E[N(t)] = 0 , \tag{4.3}$$

and known covariance kernel K:

$$E[N(s)\, N(t)] = K(s,t) . \tag{4.4}$$

Further, $\{N(t), t \in T\}$ is a normal process; that is, the n random variables $N(t_1), \ldots, N(t_n)$ are jointly normally distributed so that their joint characteristic function is given by

$$\begin{aligned} \varphi_{N(t_1),\ldots,N(t_n)}(u_1, \ldots, u_n) &= E\left[\exp\left\{i \sum_{j=1}^{n} u_j\, X(t_j)\right\}\right] \\ &= \exp\left[-\frac{1}{2} \sum_{j,k=1}^{n} u_j\, K(t_j,t_k)\, u_k\right] . \end{aligned} \tag{4.5}$$

Assumption on the signal process: The signal $\{S(t), t \in T\}$ is a known non-random function. The signal plus noise process $\{X(t) = S(t) + N(t), t \in T\}$ is then a normal process with mean value function

$$E[X(t)] = S(t) \tag{4.6}$$

and covariance kernel

$$\text{Cov}[X(s),X(t)] = K(s,t) \ . \tag{4.7}$$

Under the assumptions of this example, P_N is the probability measure on function space induced by a normal process with mean value function identically zero and covariance kernel K, while P_{S+N} is the probability measure induced by a normal process with mean value function equal to the signal function $S(t)$ and covariance kernel K.

Perfect detectability and the singular detection problem: We say that the hypotheses H_0 and H_1 are perfectly detectable, or that the problem of detecting the signal process $S(\cdot)$ in the presence of the noise process $N(\cdot)$ is singular, if there exists a set A in the sample space Ω_T such that

$$P_N[A] = 0 \ , \qquad P_{S+N}[A] = 1 \ . \tag{4.8}$$

By choosing A as the rejection region R_1 for H_0 one has probability zero of incorrectly identifying noise as signal plus noise or signal plus noise as noise. Note that the probability measures P_N and P_{S+N} are orthogonal if (4.8) holds. By definition, then, the hypotheses H_0 and H_1 are perfectly detectable if and only if P_N and P_{S+N} are orthogonal.

The regular detection problem: The problem of detecting the signal process in the presence of the noise process $N(\cdot)$ is called regular if

P_N and P_{S+N} are not orthogonal. In this case, new principles must be introduced in order to optimally choose the rejection region R_1. First, one distinguishes two types of errors that can occur:

(i) a <u>false alarm</u> (or error of type I) occurs when one says that a signal plus noise is present when in fact noise alone is present;

(ii) a <u>detection failure</u> (or error of type II) occurs when one says that noise alone is present when in fact signal plus noise is present.

A rejection region R is then characterized by two numbers α and β, defined by

$$\alpha = \text{Prob [false alarm]}$$

$$= \text{Prob } [\{X(t),\ t \in T\} \in R | H_0] \ , \tag{4.9}$$

$$\beta = \text{Prob [detection failure]}$$

$$= 1 - \text{Prob } [\{X(t),\ t \in T\} \in R | H_1] \ . \tag{4.10}$$

In certain cases it may be possible to assign a numerical measure to the seriousness of a false alarm and of a detection failure; one denotes these costs by* L_1 and L_2 respectively. Further it may be possible to determine the fraction π_S of experimental situations in which signal plus noise is present; one calls π_S the <u>prior</u> probability that signal

*In the general theory of hypothesis testing, a false alarm is called a type I error, and L_1 is the cost of a type I error, while a detection failure is called a type II error and L_2 is the cost of a type II error.

is present. To each critical region one can assign a risk ρ, called the Bayes risk and defined as the expected cost of an incorrect decision:

$$\rho = \alpha(1 - \pi_S)\, L_1 + \beta\pi_S\, L_2 \,. \tag{4.11}$$

The Bayes rejection region R (or optimum rejection region according to the Bayes criterion) is defined as the region which minimizes ρ, the expected cost of an incorrect decision.**

It may be difficult to use a Bayes rejection region for one or both of the following reasons:

(i) because of difficulty in assigning the losses L_1 and L_2,

(ii) because of difficulty in assigning the prior probability π_S.

In these circumstances one may use the Neyman-Pearson rejection region (or optimum rejection region according to the Neyman-Pearson criterion) which is defined as the rejection region R minimizing β, the detection failure probability, subject to the restriction that α, the false alarm probability, is less than or equal to some desired level α_0.

We next show how one may determine the Bayes rejection region and the Neyman-Pearson rejection region by introducing probability density functionals.

Let us assume that there exists a measure Q on the measurable subsets of the sample space Ω_T, and functions p_N and p_{S+N} with domain Ω_T, with the property that, for every measurable subset B of Ω_T,

---

**Note that a rejection region which minimizes the average probability of error π_e, $\pi_e = \alpha(1 - \pi_S) + \beta\pi_S$, is the same as the Bayes rejection region with unit costs, $L_1 = L_2 = 1$.

$$P_N[B] = \int_B p_N \, dQ \tag{4.12}$$

$$P_{S+N}[B] = \int_B p_{S+N} \, dQ \, . \tag{4.13}$$

In order to emphasize that its argument is a function $\{X(t),\ t \in T\}$, we call p_N a functional, and sometimes denote it by $p_N(X(t),\ t \in T)$. It is called the probability density functional of P_N with respect to Q. The function p_N may be written symbolically as a derivative,

$$p_N = \frac{dP_N}{dQ} \, , \tag{4.14}$$

and is then called the Radon-Nikodym derivative of P_N with respect to Q. Similarly, p_{S+N} is called the probability density functional, or Radon-Nikodym derivative, of P_{S+N} with respect to Q.

In order for the probability density functionals p_N and p_{S+N} to exist, there must exist a measure Q with respect to which both P_N and P_{S+N} are absolutely continuous; recall that P_N is absolutely continuous with respect to Q if

$$Q(A) = 0 \quad \text{implies} \quad P_N(0) \quad \text{for all} \quad A \subset \Omega \, . \tag{4.15}$$

Such a measure is given by

$$Q = P_N + P_{S+N} \, . \tag{4.16}$$

Consequently there is no loss of generality in assuming that there exist functions p_N and p_{S+N} satisfying (4.12) and (4.13) respectively.

In terms of probability density functionals, the false alarm probability α and detection failure probability β of a rejection region are given by

$$\alpha = \int_R p_N \, dQ \tag{4.17}$$

$$\beta = 1 - \int_R p_{S+N} \, dQ \, . \tag{4.18}$$

Consequently, the Bayes risk ρ of a rejection region R is given by

$$\rho = \int_R \{(1 - \pi_S) L_1 p_S - \pi_S L_2 p_{S+N}\} \, dQ + \pi_S L_2 \, . \tag{4.19}$$

To minimize ρ, one should choose R as the set of observations $\{X(t), t \in T\}$ for which the integrand in (4.19) is negative, so that the Bayes rejection region R is given by

$$R = \left\{\Big(X(t), t \in T\Big): \frac{p_{S+N}}{p_N} > \frac{(1 - \pi_S) L_1}{\pi_S L_2}\right\} . \tag{4.20}$$

The ratio

$$\frac{p_{S+N}}{p_N} \tag{4.21}$$

of probability density functionals is called in classical statistical literature the <u>likelihood ratio</u> since $p_N(X(t), t \in T)$ is defined to be

the likelihood that noise alone is present given that the observed time series was $\{X(t),\ t \in T\}$, and $p_{S+N}(X(t),\ t \in T)$ is the likelihood that signal plus noise is present given that the observed time series was $\{X(t),\ t \in T\}$.

The likelihood ratio (4.21) has a probabilistic meaning which shows that the measure Q used in defining the likelihood ratio plays no role. Let

$$(4.22) \qquad A = \left\{\Big(X(t),\ t \in T\Big):\ p_{S+N} > 0 \qquad \text{and} \qquad p_N = 0\right\}.$$

If $P_{S+N}[A] = 0$, then P_{S+N} is absolutely continuous with respect to P_N, and the Radon-Nikodym derivative of P_{S+N} with respect to P_N is given by the likelihood ratio:

$$(4.23) \qquad \frac{dP_{S+N}}{dP_N} = \frac{p_{S+N}}{p_N}.$$

The Radon-Nikodym derivative of P_{S+N} with respect to P_N is called the probability density functional of signal plus noise with respect to noise; where no ambiguity can arise it is denoted by p. To summarize, the probability density functional p of signal plus noise with respect to noise is a function on the sample space Ω_T satisfying

$$(4.24) \qquad P_{S+N}[B] = \int_B p\, dP_N, \qquad B \subset \Omega,$$

and exists if and only if P_{S+N} is absolutely continuous with respect to P_N, which holds if and only if $P_{S+N}[A] = 0$, where A is defined

by (4.22); it then follows that p is given by the likelihood ratio

$$p = \frac{P_{S+N}}{P_N} . \tag{4.25}$$

If $P_{S+N}[A] > 0$, instead of (4.24), one has for every measurable subset B of Ω,

$$P_{S+N}[B] = \int_B p \, dP_N + P_{S+N}[AB] , \tag{4.26}$$

where p is still given by (4.25); (4.26) is an example of the Lebesgue decomposition theorem which states that to the probability measures P_N and P_{S+N} there exists a function p on Ω and set A such that $P_N[A] = 0$ and (4.26) holds. We have made this assertion concrete by showing how one may find p and A.

Optimum detectors: A random variable U (that is, a function on the sample space Ω) which has the property that the rejection region which is optimum according to a certain criterion may be expressed in terms of the values of U is called an optimum detector according to that criterion.

From (4.20) it follows that the likelihood ratio is an optimum Bayes detector. Indeed the optimum rejection region according to the Bayes criterion consists of all observations $\{X(t), t \in T\}$ for which the likelihood ratio is above a certain threshold value Λ_0 given by

$$\Lambda_0 = \frac{(1 - \pi_s) L_1}{\pi_s L_2} . \tag{4.27}$$

Similarly, it may be shown that the likelihood ratio is an optimum Neyman-Pearson detector. Indeed the optimum Neyman-Pearson rejection region according to the Neyman-Pearson criterion consists of all observations $\{X(t),\ t \in T\}$ for which the likelihood ratio is above a certain threshold Λ_0, determined by the condition that the false alarm probability be equal to α_0:

$$P_N\left[\frac{p_{S+N}}{p_N} > \Lambda_0\right] = \alpha_0 \ . \tag{4.28}$$

To prove this assertion, one uses the fact that the Neyman-Pearson rejection region R is that region which subject to the condition

$$\int_R p_N \, dQ \leq \alpha \tag{4.29}$$

maximizes

$$\int_R p_{S+N} \, dQ \ . \tag{4.30}$$

Intuitively, one sees that the optimum rejection region R should contain those sample points which have the highest value of the likelihood ratio (4.21) since for a given contribution to the integral in (4.29) these points make a maximum contribution to the integral in (4.30). A formal proof of this assertion is easily given, using the fundamental lemma of Neyman and Pearson.

Example 4B.

Detection of a specified signal in normal noise. Consider the signal and noise processes described in example 4A. Assume that the covariance matrix

$$\underset{\sim}{K} = \begin{bmatrix} K(t_1,t_1) & \cdots & K(t_1,t_n) \\ \vdots & & \\ K(t_n,t_1) & \cdots & K(t_n,t_n) \end{bmatrix}$$

is non-singular with inverse denoted

$$\underset{\sim}{K}^{-1} = \begin{bmatrix} K^{-1}(t_1,t_1) & \cdots & K^{-1}(t_1,t_n) \\ \vdots & & \\ K^{-1}(t_n,t_1) & \cdots & K^{-1}(t_n,t_n) \end{bmatrix} .$$

In words, $K^{-1}(t_i,t_j)$ does not denote the reciprocal of $K(t_i,t_j)$, but rather denotes the (i,j)-th element of the inverse matrix K^{-1}. The signal and noise process and the noise process then both possess probability density functionals p_{S+N} and p_N with respect to Lebesgue measure Q:

$$p_N\Big(X(t),\ t \in T\Big) = \left\{(2\pi)^n\ |\underset{\sim}{K}|\right\}^{-\frac{1}{2}} \exp\left[-\frac{1}{2} \sum_{i,j=1}^{n} X(t_i)\ K^{-1}(t_i,t_j)\ X(t_j)\right] ,$$

$$p_{S+N}\Big(X(t),\ t \in T\Big) = \left\{(2\pi)^n\ |\underset{\sim}{K}|\right\}$$

$$\exp\left[-\frac{1}{2} \sum_{i,j=1}^{n} \{X(t_i)-S(t_i)\}\ K^{-1}(t_i,t_j)\{X(t_j)-S(t_j)\}\right]$$

where $|\underset{\sim}{K}|$ is the determinant of the matrix $\underset{\sim}{K}$.

The likelihood ratio is given by

$$\frac{p_{S+N}}{p_N} = \exp\left[\sum_{i,j=1}^{n} X(t_i)\, K^{-1}(t_i,t_j)\, S(t_j) - \frac{1}{2}\sum_{i,j=1}^{n} S(t_i)\, K^{-1}(t_i,t_j)\, S(t_j)\right] .$$

In order to write this expression more compactly, let us introduce the notation (which will play an important role in the sequel)

$$(f,g)_K = \sum_{i,j=1}^{n} f(t_i)\, K^{-1}(t_i,t_j)\, g(t_j) ,$$

defined for any two functions f and g on T. Then

$$\frac{p_{S+N}}{p_N} = \exp\left[(X,S)_K - \frac{1}{2}(S,S)_K\right] .$$

Since the likelihood ratio is a monotone increasing function of $(X,S)_K$, it follows that $(X,S)_K$ is an optimum (Bayes or Neyman-Pearson) detector. Indeed, the rejection region for testing H_0 against H_1 can be expressed as the set of observations $\{X(t),\ t \in T\}$ for which $(X,S)_K$ is above a certain threshold Λ_1, say. A detector of the form

$$(X,S)_K = \sum_{i=1}^{n} X(t_i) \left\{\sum_{j=1}^{n} K^{-1}(t_i,t_j)\, S(t_j)\right\}$$

is said to be a "correlation detector" or a "matched filter" since it is obtained by "correlating" or "matching" the specified signal shape $S(t)$ with the observed time series $X(t)$.

References

P. Halmos. Measure Theory. Van Nostrand: Princeton. 1950.

A. Kolmogorov. Foundations of the Theory of Probability. Chelsea, New York, 1950. Translation of Grundbegriffe der Wahrscheinlichkeitsrechnung, 1933.

M. Loève. Probability Theory. Second Edition. Van Nostrand: Princeton. 1960.

E. Parzen. "An approach to time series analysis," Annals of Math. Stat. Vol. 32 (1961), pp. 951-989.

E. Parzen. Stochastic Processes. Holden-Day. San Francisco. 1962.

E. Parzen. "Extraction and detection problems and reproducing kernel Hilbert spaces," SIAM Journal of Control. Vol. 1 (1962), pp. 35-62.

E. Parzen. "Probability density functionals and reproducing kernel Hilbert spaces," Proceedings of a Symposium on Time Series Analysis, edited by Murray Rosenblatt. John Wiley. New York. 1963.

E. Parzen. "A new approach to the synthesis of optimal smoothing and prediction systems," Proceedings of a Symposium on Optimization Techniques, edited by E. F. Beckenbach and R. Bellman. University of California Press. 1963.

E. Parzen. Foundations of Time Series Analysis. Holden-Day. San Francisco. To be published in 1965.

Exercises

1. In the notation defined in Example 2A, show that $f_1(\cdot)$ does not belong to C, while $f_2(\cdot)$ and $f_3(\cdot)$ do belong to C.

2. Let H_0 and H_1 be simple hypotheses about a time series $\{X(t),\ 0 \le t \le 1\}$. Under both hypotheses, $\{X(t),\ 0 \le t \le 1\}$ is a normal process with covariance kernel

$$K(s,t) = \varphi(s)\,\varphi(t)$$

and mean value function respectively given by

$$H_0:\ E[X(t)] \equiv 0\ ,$$

$$H_1:\ E[X(t)] = S(t)\ .$$

Assume that $S(\cdot)$ and $\varphi(\cdot)$ are orthogonal,

$$\int_0^1 S(t)\,\varphi(t)\,dt = 0\ .$$

Show that H_0 and H_1 are perfectly detectable.

Hint: $U = \int_0^1 S(t)\,X(t)\,dt$ is a perfect detector, since

$P_1[U = 0] = 1$, $P_2[U = \int_0^1 S^2(t)\,dt \neq 0] = 1$.

3. Let H_0 and H_1 be simple hypotheses about a time series $\{X(t),\ 0 \le t \le 1\}$. Under both hypotheses, $\{X(t),\ 0 \le t \le T\}$ is of the form

$$X(t) = \eta\, \varphi(t)$$

where $\varphi(t)$ is a non-random function and η is a normal random variable,

$$\text{Under } H_0\text{: } E[\eta] = 0\ , \qquad \mathrm{Var}[\eta] = 1$$

$$\text{Under } H_1\text{: } E[\eta] = m\ , \qquad \mathrm{Var}[\eta] = 1\ .$$

Show that an optimum detector for testing H_0 against H_1 is given by

$$\exp\left[-\frac{1}{2}\left\{\left(\eta - m\right)^2 - \eta^2\right\}\right] = \exp\left\{m\eta - \frac{1}{2}m^2\right\}$$

and therefore is given by η. A possible formula for η is

$$\eta = \frac{X(t_1)}{\varphi(t_1)}$$

for any point t_1 in $0 \le t_1 \le 1$ such that $\varphi(t_1) \neq 0$.

STANFORD UNIVERSITY
TECHNICAL REPORTS DISTRIBUTION LIST
CONTRACT Nonr-225(21)
(NR-042-993)

Armed Services Technical
Information Agency
Arlington Hall Station
Arlington 12, Virginia — 10

Bureau of Supplies and Accounts
Code OW
Department of the Navy
Washington 25, D. C. — 1

Head, Logistics and Mathematical
Statistics Branch
Code 436
Office of Naval Research
Washington 25, D. C. — 3

Commanding Officer,
Office of Naval Research Branch Office
346 Broadway
New York 13, New York
Attn: J. Laderman — 1

Commanding Officer,
Office of Naval Research Branch Office
1030 East Green Street
Pasadena 1, California (Dr. A.R. Laufer) — 1

Commanding Officer,
U. S. Army Signal Electronic Research
P. O. Box 205 - Major Adornetto
Mountain View, California — 1

Commanding Officer
U. S. Navy Powder Factory
Attn: F. Freshman, R & D
Indianhead, Maryland — 1

Director, Naval Research Lab.
Attn: Tech. Infor. Officer
Washington 25, D. C. — 6

Director,
National Security Agency
Attn: REMP-1
Fort George G. Meade, Maryland — 2

Dept. of Mathematical Statistics
University of North Carolina
Chapel Hill, North Carolina — 1

Department of Statistics
University of California
Berkeley 4, California — 1

Department of Statistics
Michigan State University
East Lansing, Michigan — 1

Library,
Institute for Defense Analyses
Communications Research Division
Von Neumann Hall
Princeton, New Jersey — 1

Library
Department of Industrial Engr.
& Operations Research
College of Engr.
New York University
University Heights
New York 53, New York — 1

Office of the Ass't Naval
Attache for Research
American Embassy
Navy No. 100, Fleet P. O.
New York, New York — 2

Scientific Section
Office of Naval Research
1000 Geary Street
San Francisco 9, California — 2

Statistical Engineering Lab.
National Bureau of Standards
Washington 25, D. C. — 1

Statistical Laboratory
University of Washington
Seattle 5, Washington — 1

Superintendent,
U. S. Navy Postgraduate School
Attn: Library
Monterey, California — 1

U. S. Naval Avionics Facility
Attn: Library
Indianapolis 18, Indiana — 1

Dr. Stephen G. Allen
Stanford Research Institute
333 Ravenswood Avenue
Menlo Park, California — 1

Professor T. W. Anderson
Dept. of Mathematical Statistics
Columbia University
New York 27, New York — 1

Professor Fred C. Andrews
Department of Mathematics
University of Oregon
Eugene, Oregon — 1

Professor Robert Bechhofer
Dept. of Industrial and Eng. Admin.
Sibley School of Mechanical Eng.
Cornell University
Ithaca, New York — 1

Mr. J. Berkowitz
Quality Control Supervisor
NORDEN Division of United Aircraft Corp.
Ketay Department - Helen Street
Norwalk, Connecticut — 1

Dr. Julian H. Bigelow
School of Mathematics
Institute for Advanced Study
Princeton, New Jersey — 1

Professor Z. W. Birnbaum
Laboratory of Statistical Research
Department of Mathematics
University of Washington
Seattle 5, Washington — 1

Professor David Blackwell
Department of Mathematical Sciences
University of California
Berkeley 4, California — 1

Dr. Julius R. Blum
Sandia Corporation
Albuquerque, New Mexico — 2

Dr. Paul Blunk
Box 532
Fair Oaks, California — 1

Dr. Charles Boll
Aerospace Corp. F-2813
P. O. Box 95085
Los Angeles 45, California — 1

Professor G. E. P. Box
Department of Statistics
University of Wisconsin
Madison, Wisconsin — 1

Professor Ralph A. Bradley
Department of Statistics
Florida State University
Tallahassee, Florida — 1

Professor W. G. Cochran
Department of Statistics
Harvard University
2 Divinity Avenue, Room 311
Cambridge 38, Massachusetts — 1

Professor Lee Cronbach
Bureau of Education Research
1007 S. Wright
Champaign, Illinois — 1

Professor Cyrus Derman
Dept. of Industrial Engineering
Columbia University
New York 27, New York — 1

Dr. Joseph Daly
Bureau of the Census
Washington 25, D. C. — 1

Dr. Francis Dresch
Stanford Research Institute
333 Ravenswood Avenue
Menlo Park, California — 1

Mr. John A. Dutton
Department of Meteorology
University of Wisconsin
Madison, Wisconsin — 1

Professor Meyer Dwass
Department of Mathematics
Northwestern University
Evanston, Illinois — 1

Professor W. T. Federer
Cornell University
Department of Plant Breeding
Biometrics Unit
Ithaca, New York — 1

George Fishman
Logistics Department
Rand Corporation
1700 Main Street
Santa Monica, California — 1

Dr. Merrill M. Flood
Mental Health Research Institute
205 North Forest Avenue
Ann Arbor, Michigan — 1

Dr. Donald P. Gaver
Westinghouse Research Labs.
Beulah Rd. - Churchill Boro.
Pittsburgh 35, Pennsylvania — 1

Mr. Murray A. Geisler
Logistics Section
The RAND Corporation
1700 Main Street
Santa Monica, California — 1

Professor H. P. Goode
Dept. of Industrial and Engineering
Administration
Cornell University
Ithaca, New York — 1

Professor E. J. Gumbel
Industrial Engr. Department
409 Engineering Building
Columbia University
New York 27, New York — 1

Professor Donald Guthrie, Jr.
Stanford Research Institute
333 Ravenswood Avenue
Menlo Park, California — 1

Professor Bernard Harris
Department of Mathematics
The University of Nebraska
Lincoln, Nebraska — 1

Dr. Theodore E. Harris
The RAND Corporation
1700 Main Street
Santa Monica, California — 1

Leon H. Herbach
Dept. of Industrial Engineering
and Operations Research
College of Engineering
New York University
New York 53, New York — 1

Professor W. Hirsch
Institute of Mathematical Sciences
New York University
New York 3, New York — 1

Professor Harold Hotelling
Associate Director,
Institute of Statistics
University of North Carolina
Chapel Hill, North Carolina — 1

Professor Stanley Isaacson
4706 Lakeview Drive
Des Moines 12, Iowa — 1

Professor Oscar Kempthorne
Statistics Laboratory
Iowa State College
Ames, Iowa — 1

Professor Solomon Kullback
Department of Statistics
George Washington University
Washington 7, D. C. — 1

Dr. Eugene Lukacs
Department of Mathematics
Catholic University
Washington 17, D. C. — 1

Professor G. W. McElrath
Department of Mechanical Engineering
University of Minnesota
Minneapolis 14, Minnesota — 1

Professor Paul Meier
Department of Statistics
University of Chicago
Chicago, Illinois — 1

Professor James M. Moore
Dept. of Industrial Engineering
Northeastern University
Boston 12, Massachusetts — 1

D. E. Newnham
Chief, Industrial Eng. Division
Comptroller
Hdqrs., San Bernardino Air Materiel Area
Norton Air Force Base, California — 1

Professor J. Neyman
Department of Statistics
University of California
Berkeley 4, California — 1

Mr. Edward Paulson
72-10 41st Avenue
Woodside 77
New York, New York — 1

Mr. W. J. Peterson
Dept. 65-24
Lockheed Missile Systems Division
Sunnyvale, California — 1

Dr. Richard Post
Department of Mathematics
San Jose State College
San Jose, California — 1

Professor Ronald Pyke
Department of Mathematics
University of Washington
Seattle 5, Washington — 1

Professor George J. Resnikoff
Department of Industrial Engineering
Illinois Institute of Technology
Chicago 16, Illinois — 1

Professor Herbert Robbins
Mathematical Statistics Dept.
Columbia University
New York 27, New York — 1

Professor Murray Rosenblatt
Department of Mathematics
Brown University
Providence 12, Rhode Island — 1

Professor Herman Rubin
Department of Statistics
Michigan State University
East Lansing, Michigan — 1

Professor I. R. Savage
Department of Statistics
University of Statistics
Minneapolis, Minnesota — 1

Professor George J. Schick
Department of Mathematics & Statistics
Sacramento State College
Sacramento, California — 1

Professor Seymour Sherman
Department of Mathematics
Wayne State University
Detroit 2, Michigan — 1

Professor W. L. Smith
Department of Statistics
University of North Carolina
Chapel Hill, North Carolina — 1

Dr. Milton Sobel
Department of Statistics
University of Minnesota
Minneapolis, Minnesota — 1

Professor Frank Spitzer
Department of Mathematics
Cornell University
Ithaca, New York — 1

Professor Gerald L. Thompson
Carnegie Institute of Technology
Schenley Park
Pittsburgh 13, Pennsylvania — 1

Professor Donald Truax
Department of Mathematics
University of Oregon
Eugene, Oregon — 1

Mr. Harry Weingarten
U.S. Arms Control and Disarmament Agency
State Dept. Bldg.
21st St. and Virginia Ave., N. W.
Washington 25, D. C. — 1

Capt. Burton L. Weller
AFPR Office
Martin Aircraft Corporation
Denver, Colorado — 1

Professor M. B. Wilk
Statistics Center
Rutgers-The State University
New Brunswick, New Jersey — 1

Dr. John D. Wilkes
Office of Naval Research
Code 200
Washington 25, D. C. — 1

Professor S. S. Wilks
Department of Mathematics
Princeton University
Princeton, New Jersey — 1

Professor J. Wolfowitz
Department of Mathematics
Cornell University
Ithaca, New York — 1

Professor M. A. Woodbury
Department of Mathematics
New York University
New York 53, New York — 1

Distribution via ONR London

Commading Officer
Branch Office
Navy No. 100
Fleet Post Office
New York, New York — 1

Dr. Dennis V. Lindley
Statistical Laboratory
University of Cambridge
Cambridge, ENGLAND — 1

Other Foreign Addresses

Professor J. E. Moyal
Institute of Advanced Studies
Australian National University
Canberra A. C. T.
AUSTRALIA — 1

Dr. A. R. Roy
Statistical Wing
Indian Council of Agricultural Research
Linlithgow Avenue
New Delhi, INDIA — 1

Professor Tosio Kitagawa
Mathematical Institute
Faculty of Science
Kyusyu University
Fukuoka, JAPAN

Mr. Cesareo Villegas
Institute de Mathematica y Estadistica
Av. J. Herrerra y Reissig 565
Montevideo, URUGUAY — 1

Professor G. S. Watson
Department of Mathematics
University of Toronto
Toronto 5, Ontario,
CANADA — 1

Professor D. A. S. Fraser
Department of Mathematics
University of Toronto
Toronto 5, Ontario
CANADA — 1

Professor H. Witting
Institut fur Mathematische Statistik
University of Munster
Schlossplatz 2
44 Munster/Westf.,
GERMANY — 1

Mr. Geoffrey Gregory
10 Morris Avenue
Box Hill North
E 12 Victoria
AUSTRALIA — 1

Additional Copies for Project Leader and Assistants, and Reserve for Future Requirements — 50

Contract Nonr-225(21)
April 1963