MASTER

STOCHASTIC POWER FLOW MODELING

JUNE 1980

Prepared by
General Electric Company
Schenectady, New York

DISCLAIMER

This book was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

for

U.S. Department of Energy
Division of Electric Energy Systems
Washington, D.C.

LEGAL NOTICE

Contract No. ET78-C-01-3246

There is no objection from the patent
point of view to the publication or
dissemination of the document(s)
referred to in this letter.

BROOKHAVEN PATENT GROUP

7/2 1980 By CHE

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

~~LEGAL NOTICE~~

~~This technical report is being transmitted in advance of DOE patent clearance and no further dissemination or publication shall be made of the report without prior approval of the DOE Patent Counsel.~~

STOCHASTIC POWER FLOW MODELING

ABSTRACT

The stochastic nature of customer demand and equipment failure on large interconnected electric power networks has produced a keen interest in the accurate modeling and analysis of the effects of probabilistic behavior on steady state power system operation. The principle avenue of approach has been to obtain a solution to the steady state network flow equations which adhere both to Kirchhoff's Laws and probabilistic laws, using either combinatorial or functional approximation techniques.

Clearly the need of the present is to develop sound techniques for producing meaningful data to serve as input. This research has addressed this end and serves to bridge the gap between electric demand modeling, equipment failure analysis, etc., and the area of algorithm development. Therefore, the scope of this work lies squarely on developing an efficient means of producing sensible input information in the form of probability distributions for the many types of solution algorithms that have been developed.

This report describes, in detail, two major areas of development: a decomposition of stochastic processes which gives hope of stationarity, ergodicity, and perhaps even normality; and a powerful surrogate probability approach using proportions of time which allows the calculation of joint events from one dimensional "probability" spaces.

Contract No. ET78-C-01-3246

JUNE 1980

NOTICE

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Department of Energy, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, mark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

TABLE OF CONTENTS

	<u>Page</u>
I INTRODUCTION	
Motivation and Background	1
Modeling Philosophy	1
Surrogate Probabilities	1
Additional Studies	2
II DECOMPOSITION OF STOCHASTIC PROCESSES	
Univariate Development	3
Univariate Summary	5
Multivariate Extension	6
III DIRECT TRANSFORMATION METHOD	
Development of Power Flow Equations	8
Application of Decomposition to Power Flow Equations	9
IV WINDOW TRANSFORMATION METHOD	
Development of Proportion of Time Distribution	11
Singularity of Proportion of Time Distribution	12
Development of the Singular Distribution	13
Application of the Singular Distribution	15
Proportion of Time Distribution of Events	16
V INTERVAL MAXIMUM TRANSFORMATION	18
VI NONLINEAR EFFECTS	
Basic Approach	19
Simulation Parameters	19
Performance Function Analysis	21
Output Distribution Deformations	21
Remarks	23
VII CONCLUSIONS AND RECOMMENDATIONS	
General Results	25
Recommendations for Further Investigation	25
Closing Remarks	26
APPENDIX A THEORETICAL DEVELOPMENT	
Density Shape	27
Moments	29
Stationarity and Ergodicity	31
APPENDIX B 14 BUS POWER SYSTEM DATA	38
REFERENCES	40

LIST OF FIGURES

<u>FIGURE</u>	<u>TITLE</u>	<u>PAGE</u>
4.1	Proportion of Time (Window) Transformation	11
4.2	Example Curve	13
4.3	Joint Proportion of Time Distribution for Example	14
4.4	Curve of Example Showing Time Dependence	14
5.1	Interval Maximum Transformation	18
6.1	Linear and Nonlinear Mappings	20
6.2	Simulation Flowchart	20
6.3	Performance Functions	21
6.4	Effects of Nonlinearity	21
6.5	Flow Density for Line 1-2	22
6.6	Flow Density for Line 9-14	22
6.7	Exact Density of x	23
6.8	Linearized Density of $x(y_0 = 1.5)$	23
6.9	Linearized Density of $x(y_0 = 2.0)$	23
6.10	Linearized Density of $x(y_0 = 1.0)$	24
6.11	Composite Linear Density of x	24

PREFACE

This is the final report for the DOE-funded contract entitled "Stochastic Power Flow Modeling." The contents addresses the problem of estimating the probability distributions necessary to serve as input for stochastic power flow algorithms. The report begins with a summarizing introduction followed by a chapter (2) on decomposition of nonstationary stochastic processes, and three chapters (3, 4, and 5) on various multivariate transformations of stochastic processes. Chapter 6 addresses the problem of nonlinearity in the network model, and Chapter 7 serves to conclude and suggest further work. A very detailed appendix follows which reviews the theoretical basis for the developments of Chapters 2 through 5. A second appendix contains the power system data used in Chapter 6.

The material on process decomposition and transformation was developed and written by Dr. William T. Tucker of General Electric's Corporate Research and Development, and the material of Chapter 6 on nonlinearity was prepared by A.M. Leite da Silva and Dr. Ron Allan of the University of Manchester Institute of Science and Technology (UMIST), Manchester, England.

We would like to acknowledge Dr. T.A. Trygar of DOE and Dr. L.H. Fink formerly of DOE for their contributions to the development of all phases of this research. We also thank Dr. H.H. Happ of General Electric for the original conception of this work.

Robert C. Burchett, Project Manager
General Electric Company

CHAPTER I INTRODUCTION

1.1 Motivation and Background

A stochastic power flow solution is a set of probabilistic information which describe the steady state uncertainty in electric power system voltages and power flows. This information is usually in the form of density functions, and is obtained by passing the input density function (system loads) through the nonlinear transformation of the power flow equations. The density functions describing voltage and power flow may then be used to calculate probabilities of unacceptable steady state operation, such as overloaded lines or under voltages.

The purpose of this research has been to develop techniques for constructing realistic models to be used for steady state stochastic analysis of interconnected electric power systems. A non-deterministic approach to steady state power system modeling is necessitated by the stochastic nature of customer demands and by the random occurrence of equipment failures. Virtually all of the work published in engineering literature to date has addressed the development of highly sophisticated algorithms to perform such an analysis, operating on the assumption that accurate, realistic models are available to serve as input. In fact, much of the efforts concerning steady state stochastic analysis of power systems in the literature have created stochastic models based on the numerical properties of their algorithms, rather than developing a methodology which accurately represents the true random behavior in the system. There is, therefore, high motivation for paying careful attention to the development of realistic approaches for representing the random steady state operating behavior of an electric power system.

1.2 Modeling Philosophy

Because of the time varying nature of power system demands, any study of power system uncertainty must begin by considering the stochastic processes which describe system loads and generations. The principle problem to be addressed is the construction of meaningful probability distribution functions (pdf) of random variables defined on these processes. In general, such distributions would be multivariate, time-varying functions, the construction-estimation of which requires several independent realizations of the process. To complicate the problem, these distributions are really needed in determining probabilities for future events.

In order to facilitate this difficult task, this report develops a decomposition of the stochastic processes representing system demand. The vector valued input processes are broken down into three components: a modeled mean, a perfectly predictable part, and a purely stochastic part. The modeled mean is that part of the process which can be accurately modeled over a specified frame of time. Such a model would produce a deterministic vector valued trajectory through time. This research assumes the availability of such a model and therefore this report does not develop or provide one.

The perfectly predictable part of the vector process is that portion of the remaining process (after mean extraction) which can be exactly determined in a linear manner given its initial value. For example, such functions may be represented as polynomials or transcendental functions with fixed random coefficients. This part of the process has zero mean, and may be combined with the modeled mean part to form the completely known or knowable part of the vector input process.

Upon removing all deterministic parts of the process, we are left with the perfectly stochastic part, which has zero mean, and contains no repeatable or predictable parts. As explained by this report, the purely stochastic part of the process shows good promise of stationarity and some properties of ergodicity which would permit the substitution of time average information for hard to get ensemble information. This would allow the construction of the desired probability density from a single observation of the process, rather than averaging several observations together. Such a property is extremely powerful since independent observations are rarely obtained. On the other hand, this property of ergodicity cannot be casually assumed without carefully removing all nonergodic parts of the process, which the proposed decomposition is designed to do. The existence of stationarity in the stochastic part coupled with the modeled and predictable parts would permit the calculation of probability distributions for short and medium term future events. We point out that this procedure differs from the least squares forecasting approach in that the variance about the projected mean would be constant over the range of consideration.

Using such a decomposition can be computationally challenging, as pointed out in the report, but the advantages and benefits are considerable. The decomposition offers the following features:

- o exploits stationarity and ergodicity properties that may be hidden by the non-stochastic or deterministic parts of the process,
- o permits the construction of instantaneous probability distributions based on time average information,
- o gives hope of normality and even independence in the vector valued purely stochastic part.

1.3 Surrogate Probabilities

The second major development of this report is the concept of "percentage of time" distributions, which describe the per unit time that the instantaneous value of the stochastic process is at or below a constant. As described in the text, such information is not a genuine instantaneous probability, in that

all reference to chronological time is lost. Only information describing the proportion of time in an interval the process is at or below a certain level can be produced, rather than the instantaneous probability that the process is at or below a certain level. For example, one could calculate the percentage of time in a one hour interval that a transmission component is overloaded. No information concerning the frequency and duration of each occurrence of overload can be calculated.

When this type of information is acceptable, it can be shown that the multivariate proportion of time distribution is completely singular. This is arrived at by considering the trajectory of the process in multidimensional space to be a curve which maps onto the time axis. As shown in the report, it is possible to represent the proportion of time distribution as a one dimensional histogram. This singularity therefore enables the very efficient calculation of high dimensional joint proportions of times. For example, it would be possible (indeed computationally practical) to calculate the proportion of time that all transmission components on a large system are within their ratings. Such a calculation has never been proposed as practical in the power system literature.

We also point out that the proportion of time distribution requires only an estimate of the expected future trajectory, and not higher order moments. This feature alone reduces the computational requirements tremendously. The total singularity of this distribution allows a simple transformation from input proportions to joint output proportions, even for very large systems. The calculation of surrogate probabilities (proportions of time) involves only the one dimensional integration of a function, rather than an

n-dimensional integration. In addition, the presence of stationarity and ergodicity are not as fundamentally important as they are with the direct transformation, which minimizes the need of complicated testing to verify the presence of these properties.

Proportion of time distributions for future events are obtained by one of two approaches. If several independent realizations of the stochastic process are available, it would be possible to study the changes in the proportion of time distribution with each realization. This suggests that there is a random distribution for proportion of time information. (This type is recognized in this report, but is not extensively studied in the technical development.) The expected distribution for future independent realizations could then be derived from this random distribution.

A second, possibly more promising technique, uses the before mentioned decomposition to obtain future realizations of the mean and perfectly predictable portions of the process. Indeed, if the model is valid, and the decomposition is successful, then the proportion of time distribution obtained will be in certain cases an approximation to the expected distribution of the future random distribution.

1.4 Additional Studies

Also included in this report is a simulation based examination of the effects of network model nonlinearity on the input-output transformation. Examples are provided. A computer program capable of performing a linearized A.C. stochastic power flow using traditional convolution techniques is provided in a separate document. This program and its documentation were developed at the University of Manchester.

CHAPTER II DECOMPOSITION OF STOCHASTIC PROCESSES

This chapter presents the extension to nonstationary processes of the Wold decomposition of second order stochastic processes. This three part decomposition splits the process into a deterministic mean function, a purely predictable stochastic part, and a purely stochastic part.

2.1 Univariate Development

We motivate the input decomposition by discussing a trivial example. Consider a simple DC network in which output voltage, V , can be expressed as a simple (linear) function of input power, P . We write this as

$$V_t = f_t(P_t), \quad t \text{ denotes time,}$$

and ignore the fact that P_t is unchanging with time. Write P_t as

$$P_t = \mu_t + x_t$$

where

μ_t is the deterministic part of P_t

and

x_t is the purely nondeterministic part of P_t .

We shall give a development of μ_t and x_t presently but for the time being one can think of μ_t as representing that part of P_t that can be predicted and x_t as that part that represents the purely stochastic nature of P_t .

Now

$$V_t \cong f_t(P_t(0)) + f'_t(P_t(0))\Delta P_t \quad (2.1)$$

where ' indicates the first derivative of f_t . In (2.1) we have ignored the remainder term of a Taylor's expansion of order one. Let (0) the point of evaluation of the Taylor's expansion be μ_t . Then (2.1) becomes

$$V_t \cong f_t(\mu_t) + f'_t(\mu_t) x_t \quad (2.2)$$

and we have a "nice" expression relating V_t to an overall mean level and a stochastic part. Even if higher order terms are required in (2.1) evaluating the Taylor's expansion about μ_t produces a separation into a "constant" term and stochastic terms.

In a general power flow analysis (2.1) becomes vector valued but the underlying reasoning remains the same. Also we note that f_t represents the inverse function (assuming such exists) to the usual equations that express power as a function of voltages and phase angles. We have purposefully indicated that f_t in (2.1) may vary in form with t since in the general case the actual network configuration can change as time changes. Thus, various power flow analyses may have to be carried out in the course of, say, the study of a day's time span.

We now consider the case of a univariate P_t in some detail. The argument for the general vector case follows the same lines but with some added complications. Let

$$E(P_t) = m_t < \infty$$

where

E is the expectation operator

and

t is integer.

Thus we restrict ourselves to the discrete time series case with a finite mean. Certain results can be derived for the continuous time situation but the method of analysis employed herein cannot be adapted to that case. We further assume that

$$E(P_t^2) < \infty$$

and so we are assuming that not only are power levels finite, but that so are their fluctuations. These assumptions are not unreasonable since we are only considering steady state operating situations in our analyses.

Now let

$$z_t = P_t - m_t.$$

Then

$$E(z_t) = 0$$

and

$$E(z_t^2) < \infty.$$

Note that we have not assumed that $E(z_t^2)$ is constant. We only assume that the stochastic processes we deal with are of second order. The decomposition we seek is given in terms of z_t and is just a second order version of the Wold decomposition for second order stationary processes. A second order version for continuous time was first given by Cramér (see Cramér and Leadbetter (1967) pages 107-108). We follow the intent of that theorem but adopt the methodology of the Wold decomposition as given by Hannan (1970) Theorem 2, page 137.

Since the development involves mean square (m.s.) convergence, we shall digress and briefly discuss this concept.

$$X_t \xrightarrow{m.s.} X \text{ as } t \rightarrow \infty \quad (t \text{ integer})$$

if (and only if)

$$E |X_t - X|^2 \rightarrow 0.$$

This is also called quadratic mean (q.m.) convergence.

Its nice property is that only moments (up to second order) of the process are involved. Furthermore, Loève (1963, pages 469-470) has shown that X_t converges to X in m.s. if and only if

$$E(X_s \bar{X}_t) \rightarrow \text{a finite limit}$$

as s and t tend independently to infinity. (The bar denotes the complex conjugate.) This result allows one to determine conditions and check out cases where mean square convergence holds.

Also it can be shown that the set of all random variables such that $E|X|^2 < \infty$ forms a complete vector (linear) space. Since

$$(X, Y) = E(X \bar{Y}),$$

with $E(X) = E(Y) = 0$, forms an inner product with norm

$$\|X\|^2 = E|X|^2$$

this space is also a Hilbert space. The beauty of this is that ordinary geometrical concepts can be employed. Indeed if $E(X\bar{Y}) = 0$ then X and Y are said to be orthogonal, $X \perp Y$. Also, if M denotes a subspace (a closed linear manifold), then $X \perp M$ if and only if X is orthogonal to every $Y \in M$. Moreover, the orthogonal projection, $\hat{X}$, of X on M is the unique (with probability one) element of M such that

$$(X - \hat{X}) \perp M \quad (2.3)$$

and

$$(X - \hat{X}, X - \hat{X}) \leq (X - Y, X - Y), Y \in M.$$

We are now in a position to consider the decomposition of z_t . Let M_t be the Hilbert (sub)space which is spanned by the z_s , $s \leq t$. Thus, M_t contains all finite linear combinations, $\sum a_j z_{t_j}$, $t_j < t$, as well as their limits in mean square. We put $M_\infty = \bigcup_{t=\infty} M_t$ and M_∞ equal to the Hilbert space of the process $\{z_t\}$, that is, M_∞ consists of all finite linear combinations of the z_t and their limits in m.s. Clearly,

$$M_\infty \subseteq M_s \subseteq M_t \subseteq M_\infty, s \leq t.$$

Also $\{z_t\}$ is (linearly) deterministic, nondeterministic, purely nondeterministic, according as

$$M_\infty = M_\infty, M_\infty \subset M_\infty, M_\infty = \{0\}$$

where

$\{0\}$ denotes the subspace consisting of the null random variable.

Let $\hat{z}_t$ be the projection of z_t onto M_{t-1} and put $\varepsilon_t = z_t - \hat{z}_t$ so that $\varepsilon_t \perp M_{t-1}$. Clearly from (2.3) $\hat{z}_t$ minimizes the mean square error of prediction (of z_t). Set $\sigma_t^2 = E(\varepsilon_t^2)$. In the usual second order stationary case σ_t^2 is not a function of t , that is, σ_t^2 is constant, say, σ^2 . Thus it can be that for some t , $\sigma_t^2 = 0$ and yet $\sigma_t^2 \neq 0$. Therefore, the relationship to the deterministic situation is not as straightforward as in the stationary case (if $\sigma^2 = 0$).

Now for $s < t$, $\varepsilon_t \perp M_s \subseteq M_{t-1}$ from (2.3) and since $\varepsilon_s \in M_s$ we have $\varepsilon_s \perp \varepsilon_t$, $s < t$. Thus the $\{\varepsilon_t\}$ process is uncorrelated. (The mean function is zero since $E(z_t) = 0$.) Consider the projection, x_t , of z_t on the subspace, $M_t(\varepsilon)$, of M_t spanned by the ε_s , $s \leq t$. By virtue of Loève's result and the fact that the ε_s are uncorrelated, this projection can be written as

$$x_t = \sum_{j=0}^{\infty} a_j(t) \varepsilon_{t-j} \quad (2.4)$$

where

$$\sum_{j=0}^{\infty} a_j^2(t) \sigma_{t-j}^2 < \infty$$

and the series defining x_t is taken in a m.s. sense. We have explicitly indicated that the a 's are in general a function of how much of the past is available, since z_t need not be stationary. We note that (2.4) expresses x_t as an infinite series (in m.s.) and gives the existence of the a 's but not their values.

Now put $y_t = z_t - x_t$ and let $M_t(x)$ be the subspace of M_t spanned by the x_s , $s \leq t$ and $M_t(y)$ the subspace of M_t spanned by the y_s , $s \leq t$. It is evident from (2.3) that $y_t \perp \varepsilon_s$, $s \leq t$. But $y_t \in M_t$ and $\varepsilon_s \perp M_t$ for $s > t$ so that

$$E(\varepsilon_s y_t) = 0. \quad (2.5)$$

Let $u \in M_t(x)$. Then u is a linear combination of certain x_s , $s \leq t$ or their limits in m.s. Thus there exists a sequence of finite linear combinations $\sum_j b_j x_{s_j}$ that has m.s. limit u . But each x_{s_j} is a m.s. series in the ε_s , $s \leq t$. Then from the orthogonality of the ε and Loève's result it follows that $u \in M_t(\varepsilon)$. Therefore from (2.5)

$$M_t(x) \perp M_t(y) \quad (2.6)$$

and since $z_t = x_t + y_t$ we have

$$M_t \subseteq M_t(x) \oplus M_t(y).$$

But $M_t(x)$ and $M_t(y)$ are subspaces of M_t by their construction. Therefore

$$M_t = M_t(x) \oplus M_t(y). \quad (2.7)$$

We also note that x_t and y_t are unique (with probability one) since x_t was defined as the projection of z_t on $M_t(\varepsilon) \subseteq M_t$.

Equations (2.5) and (2.7) allow us to obtain an explicit expression for the a 's in the series given by (2.4). Let $s \leq t$ and consider

$$\begin{aligned} E(z_t \varepsilon_s) &= E\{(x_t + y_t) \varepsilon_s\} \\ &= E(x_t \varepsilon_s) \\ &= E\{(\sum a_j(t) \varepsilon_{t-j}) \varepsilon_s\} \\ &= E(a_{t-s}(t) \varepsilon_s^2) \\ &= a_{t-s}(t) E(\varepsilon_s^2). \end{aligned}$$

Thus

$$E(z_t \varepsilon_{t-j}) = a_j(t) E(\varepsilon_{t-j}^2), j \geq 0 \quad (2.8)$$

and

$$a_j(t) = E(z_t \varepsilon_{t-j}) / \sigma_{t-j}^2, j \geq 0$$

if $\sigma_{t-j}^2 > 0$. If $\sigma_{t-j}^2 = 0$, then $a_j(t)$ is arbitrary. In this case we set $a_j(t) = 0$ which in the vector case would be equivalent to determining the a 's by taking the generalized inverse of σ_{t-j}^2 . We select the a 's to solve (2.8) that are of minimum norm. Clearly this cannot alter the convergence of (2.4) and is thus an acceptable way to determine the a 's. We note in passing that in general the a 's are not unique since σ_{t-j}^2 can be zero for some j .

Finally we would like to show that x_t is purely nondeterministic and y_t is deterministic. Now

$$M_t(x) \subseteq M_t \quad (2.9)$$

and

$$M_t(y) \subseteq M_t \quad (2.10)$$

as we have previously noted. Then $y_t \in M_t$ from (2.10) and since

$$M_t = M_{t-1} \oplus [\varepsilon_t]$$

where $[\varepsilon_t]$ denotes the one dimensional subspace spanned by ε_t we have that $y_t \in M_{t-1}$ (recall (2.5), (2.6) and (2.7)). But $M_{t-1} = M_{t-2} \oplus [\varepsilon_{t-1}]$ so that $y_t \in M_{t-2}$. Continuing by induction we have that

$$y_t \in M_{t-j}, j \geq 0$$

and, therefore,

$$y_t \in M_{-\infty}$$

as $M_s \subseteq M_t$ for $s \leq t$. Denote the projection on $M_{-\infty}$ by $P_{-\infty}$. Then since $y_t \in M_{-\infty}$ and x_t and y_t are orthogonal we must have

$$y_t = P_{-\infty} z_t.$$

Therefore, for every t , $x_t \perp M_{-\infty}$ and thus

$$M_t(x) \perp M_{-\infty}. \quad (2.11)$$

We now employ (2.9), (2.10) and (2.11) to complete the argument. Allow t to approach $-\infty$ in (2.9) and (2.11) to obtain the result that $M_{-\infty}(x) = \{0\}$ for every t so that x_t is purely nondeterministic. Moreover, from (2.7) we have that

$$M_{-\infty} \subseteq M_t(x) \oplus M_t(y)$$

for all t so that with (2.11) we conclude that

$$M_{-\infty} \subseteq M_t(y).$$

Thus the fact that $y_s \in M_{-\infty}$ for every $s \leq t$ yields the result that

$$M_t(y) = M_{-\infty}.$$

Finally, from (2.10) we have

$$M_{-\infty}(y) = M_{-\infty} = M_t(y)$$

for all t and thus y_t is deterministic.

2.2 Univariate Summary

In summary we express P_t as

$$P_t = m_t + x_t + y_t \quad (2.12)$$

where

$$m_t = E(P_t)$$

x_t is the purely nondeterministic part of $\{z_t\}$

y_t is the deterministic part of $\{z_t\}$

and

$$z_t = P_t - m_t.$$

Now let

$$\mu_t = m_t + y_t \quad (2.13)$$

so that

$$P_t = \mu_t + x_t. \quad (2.14)$$

In (2.13) we have combined the (stochastic) deterministic part of $\{P_t\}$ with the mean function of $\{P_t\}$.

The two parts of (2.13) represent two kinds of determinism. The mean function m_t is the nonstochastic part of $\{P_t\}$. It may not be deterministic in the sense that m_t is perfectly predictable from past values of m_t . However, it is repeatable in that the long run average of independent realizations of $\{P_t\}$ results in m_t . Rationally there must be a cause for m_t . Thus we associate m_t with a model of $\{P_t\}$ and assume that m_t is correctly modeled. The process $\{y_t\}$ is deterministic in the sense that once past values of y_t are known then y_t is known, that is, y_t can be predicted perfectly (in a linear manner). Thus μ_t represents the known or knowable part of $\{P_t\}$.

The purely stochastic part of $\{P_t\}$ is given (in (2.14)) by the process $\{x_t\}$ which contains no repeatable or perfectly predictable (in a linear manner) parts. The process $\{x_t\}$ is of further interest to us. By virtue of (2.4), x_t is expressed as an infinite series (in m.s.) in the ε_t where the ε_t have a zero mean function and are uncorrelated. Thus, for example, if $\{\varepsilon_t\}$ is normal, then so must $\{x_t\}$ due to the linearity of (2.4). Even if $\{\varepsilon_t\}$ is not normal, it is possible to obtain various moments of $\{x_t\}$ from those of the $\{\varepsilon_t\}$ and, hence, to ultimately use certain series expansions (e.g., Gram-Charlier) to approximate the output distributions (see Deutsch (1962) or Sauer (1977)). Thus the nature of the ε_t is also of importance to the decomposition approach that we propose.

In order to analyze either x_t or ε_t we must first obtain z_t . We assume at a minimum that the form of the model of m_t is known. In addition, we assume that the generation dispatch algorithm is part of the "load model" so that any singularity associated with a relationship between load and generation can be removed. Thus in the vector case we are assuming that P_t is a vector, P_t , with entries the injections at various buses in the system which is nonsingular in a probabilistic sense, that is, the covariance matrix is positive definite for (almost) all t .

In the case where certain parameters of m_t must be estimated, we must employ general time series regression methodologies (see Fuller (1976) Chapter 9 or Hannan (1970) Chapter VII). Unfortunately in the development of these methodologies it is assumed that the residual, z_t in our study, is (second order) stationary. Thus some generality is lost. In any event we obtain an m_t and then z_t by setting

$$z_t = P_t - m_t.$$

Now from Doob (1963) theorem 7.4 page 167 we have that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N \hat{z}_t(N) = z_t$$

with probability one where $\hat{z}_t(N)$ is the projection of z_t on N past values of z_t or on the subspace generated by $z_{t-N}, \dots, z_{t-1}$. We propose that $\hat{z}_t$ be approximated by $\hat{z}_t(N)$ with N decided in practice based on a trade-off between adequacy of fit of the regression determining $\hat{z}_t(N)$ and the total amount of data available. An approximation to ε_t can then be obtained by forming

$$\hat{\varepsilon}_t = z_t - \hat{z}_t(N).$$

If N is sufficiently large, then $\hat{\varepsilon}_t$ will be a good approximation to ε_t and will, hence, be uncorrelated. Thus in order to test for stationarity of $\{\hat{\varepsilon}_t\}$ (and $\{\varepsilon_t\}$) we need only check for constancy of the variance function. (Determining the ε_t as above tacitly assumes that the process is asymptotically stationary (Hannan (1970) Chapter II, Section 6).) One way to do this (see Bendat and Piersol (1971) pages 234 - 237) is to break the data (the $\hat{\varepsilon}_t$) into subsets and check for constancy of variance among these subsets. In practice one would want to consider more than one subdivision of the data to, hopefully, avoid a failure to detect possible periodic trends in the variance function. Bendat and Piersol propose the use of the nonparametric runs tests so that distributional assumptions may be avoided. Finally, subsets that under test show no difference in their variances can be pooled and the pooled subset(s) then tested for normality by well known procedures. Normality of the $\hat{\varepsilon}_t$ will give asymptotic normality of the ε_t and, hence, of the purely stochastic part of the decomposition given in (2.14) and is thus of great practical (and theoretical) interest. We note that even if $\{\varepsilon_t\}$ is stationary that $\{x_t\}$ may not be since the a 's may vary with t . Therefore unless $\{\varepsilon_t\}$ is normal any final power flow analyses may be only in terms of moments.

In the event that parameters of m_t need be estimated, one might want to take a different approach since in this situation it is assumed that z_t is stationary. Again, the deterministic part of z_t must be accounted for and so procedures like those of Fuller or Hannan will need to be applied to z_t . Also, one could follow the methodology developed by Box and Jenkins (1976) and employ differencing operators on z_t to remove the deterministic part. Finally, one is left with an $\{x_t\}$ process that can be modeled in terms of an $\{\varepsilon_t\}$ process and the resulting ε_t can be tested for normality, etc.

In summary, we propose that the analysis of P_t follow the procedure:

- (1) Remove m_t from P_t by estimating unknown parameters if necessary.

- (2) Correct $z_t = P_t - m_t$ for any deterministic part.
- (3) Determine residuals, $\hat{\varepsilon}_t$, etc., that make up the purely nondeterministic part of z_t .
- (4) Test the residuals for stationarity, normality and any other characteristics that could be pertinent to a power flow analysis.
- (5) As a result of (4) (and the preceding steps) determine how the analysis should proceed. For example if normality is a reasonable assumption (based on the analysis through step (4)) then any further analysis is fairly straightforward. However, if the assumption of both normality and stationarity is unreasonable then one might want to, say, consider a new m_t function and start the process over again. In this viewpoint we follow Box and Jenkins who view the model building process as an iterative scheme with diagnostic checking of the adequacy of any proposed model as an integral and important part of the overall procedure.

2.3 Multivariate Extension

We now finally need to investigate (directly) the vector situation. The resulting procedure for the vector case is the same as for the univariate case (steps (1) - (5) just enumerated) but some additional complications need be considered. As we have already mentioned we assume that enough is known about the dispatch algorithm so that any singularity due to a relationship between load and generation can be properly accounted for. Even so there can be times where certain linear combinations of entries in P_t can be predicted perfectly (in a linear manner). (Recall the situation concerning the prediction error variance in the univariate case.)

Let

$$P_t = \begin{bmatrix} P_{t1} \\ P_{t2} \\ \vdots \\ P_{tq} \end{bmatrix}$$

where

P_{tj} , $j=1, 2, \dots, q$, are the injections at time t at the q buses that are of interest in the system.

Now we let M_∞ be the Hilbert space spanned by the P_{tj} , $j=1, 2, \dots, q$, all t and M_t be the subspace spanned by the P_{sj} , $j=1, 2, \dots, q$, $s \leq t$. Again $M_\infty = \bigcap_{t=-\infty}^{\infty} M_t$. Now $\hat{z}_t$, the projection of z_t onto M_{t-1} involves each entry of z_t where

$$z_t = \begin{bmatrix} P_{t1} - m_{t1} \\ P_{t2} - m_{t2} \\ \vdots \\ P_{tg} - m_{tg} \end{bmatrix}$$

and $m_{tj} = E(P_{tj})$, $j=1, 2, \dots, q$. Put $\varepsilon_t = z_t - \hat{z}_t$.

Then $\underline{\varepsilon}_t \perp M_{t-1}$ where by orthogonality we mean that each entry in $\underline{\varepsilon}_t$ is orthogonal to M_{t-1} . Set $\underline{G}_t = E(\underline{\varepsilon}_t \underline{\varepsilon}'_t)$. Then as in the univariate case one can show that, $\underline{x}_t$, the projection of $\underline{z}_t$ onto $M_t(\varepsilon)$ can be expressed as

$$\underline{x}_t = \sum_{j=0}^{\infty} \underline{A}_j(t) \underline{\varepsilon}_{t-j} \quad (2.15)$$

where

$$\sum_{j=0}^{\infty} \underline{A}_j(t) \underline{G}_{t-j} \underline{A}'_j(t) < \infty$$

and the series defining $\underline{x}_t$ is taken in a m.s. sense. (The ' denotes matrix transpose.)

Arguing as in the univariate case one may then show that

$$E(\underline{z}_t \underline{\varepsilon}'_{t-j}) = \underline{A}_j(t) \underline{G}_{t-j}, \quad j \geq 0 \quad (2.16)$$

but now $\underline{G}_j(t)$ can just be singular for some t (and not necessarily null as in the univariate case). We set

$$\underline{A}_j(t) = E(\underline{z}_t \underline{\varepsilon}'_{t-j}) \underline{G}_{t-j}^+, \quad j \geq 0 \quad (2.17)$$

where $+$ denotes the generalized inverse. Again this

cannot affect the convergence in (2.15) since the generalized inverse of $\underline{G}_{t-j}$ minimizes the norm of $\underline{A}_j(t)$. Note that the A 's need not be unique since $\underline{G}_t$ can be singular for some t . However, the situation is more complex than that in the univariate case since all diagonal entries in $\underline{G}_t$ can be nonzero and $\underline{G}_t$ still be singular.

If we put $\underline{y}_t = \underline{z}_t - \underline{x}_t$ then, finally, it can be shown that

$$\underline{P}_t = \underline{m}_t + \underline{x}_t + \underline{y}_t \quad (2.18)$$

where

$$\underline{m}_t = E(\underline{P}_t)$$

$\underline{x}_t$ is the purely nondeterministic part of $\{\underline{z}_t\}$

$\underline{y}_t$ is the deterministic part of $\{\underline{z}_t\}$

and

$$\underline{z}_t = \underline{P}_t - \underline{m}_t.$$

Also $\underline{x}_t$ is determined by (2.15) and $\underline{y}_t = \underline{z}_t - \underline{x}_t$.

We now have a decomposition in the vector situation that is analogous to that in the univariate situation. Furthermore, we can proceed as in that case since all tests, etc., involved can either be applied in a univariate manner or have vector extensions.

CHAPTER III DIRECT TRANSFORMATION METHOD

This chapter presents a development of the well known power flow equations. However, the Taylor's expansion is made about μ_t instead of an arbitrary operating point. The usefulness of this approach is that if x_t is either stationary or normal (or both), then sample information can be used to determine ensemble averages (population or process moments). These moments can then be used to approximate the instantaneous probability distributions associated with the process.

3.1 Development of Power Flow Equations

The direct transformation is no more than a consideration of a stochastic power flow analysis about an operating point, μ_t . The goal is to be able to make instantaneous probability statements about any voltages, etc. or set of these quantities that is of interest. Consider the usual power flow equations:

$$P_j = |V_j| \sum_{k=1}^n |Y_{jk}| |V_k| \cos(\delta_j - \delta_k - \theta_{jk})$$

$$= g_j(v_1, \dots, v_n, \delta_1, \dots, \delta_n) \quad (3.1)$$

$$Q_j = |V_j| \sum_{k=1}^n |Y_{jk}| |V_k| \sin(\delta_j - \delta_k - \theta_{jk})$$

$$= b_j(v_1, \dots, v_n, \delta_1, \dots, \delta_n) \quad (3.2)$$

where

V_j , the complex voltage at bus j , $= |V_j| \angle \delta_j$

$v_j = |V_j|$

$$Y_{jk} = |Y_{jk}| \angle \theta_{jk} = \begin{cases} - \text{admittance between bus } j \text{ and } k, & j \neq k \\ + \sum \text{admittances connected to bus } j, & j = k \end{cases}$$

and

n = the number of buses.

In any analysis the total number of equations, N , employed is a function of the number of buses and generators. For most buses (load buses) the unknowns are v_j and δ_j , but this is not always the case. In any event we assume that the number of equations is suitably reduced from $2n$ (the maximum number possible) to N so that locally the system is nonsingular (we shall amplify this shortly). Also let any "known" (independent) variable be denoted by u_j , $j=1, 2, \dots, N$, any "unknown" (dependent) variable be denoted by w_j , $j=1, 2, \dots, N$, and the functional forms g_j and b_j be denoted by just g_j , $j=1, 2, \dots, N$. We then have a simultaneous system of equations given by

$$w_j = g_j(u_1, \dots, u_N) \quad j = 1, 2, \dots, N. \quad (3.3)$$

From (3.1) and (3.2) it follows that partial derivatives of all orders of (3.3) exist and are continuous. Our singularity assumption is that the Jacobian of the transformation given by (3.3) is nonsingular for all values of u_j , $j=1, 2, \dots, N$ in a domain of interest.

We have put known and unknown in quotes since in practice the w_j are initially known and the system (3.3) must be inverted to obtain the u 's as functions of the w 's. It may result that certain u values that are solutions to (3.3) are not feasible as steady state operating points due to overloads, etc. From our viewpoint this is acceptable since we are ultimately concerned with the probability of these kinds of events (among other things). Therefore, we shall assume that any feasible solution of (3.3) is possible as a steady state and for example that no change in the dispatch is made during the solution as would be done in practice.

Now set

$$F_j = g_j(u_1, \dots, u_N) - w_j, \quad j = 1, 2, \dots, N.$$

Then we have $F_j \equiv 0$ over the domain of interest or

$$F_j(u_1, \dots, u_N, w_1, \dots, w_N) \equiv 0.$$

Thus from the Implicit Function Theorem we have that the inverse system of functions to (3.3) exists and is uniquely defined in the domain of possible w values. Furthermore, the system of first partial derivatives with respect to the w_j exists and is continuous in the N w_j . Also the Jacobian matrix in this system is nonsingular. We may then write

$$u_j = f_j(w_1, \dots, w_N) \quad j=1, 2, \dots, N \quad (3.4)$$

where the f_j denote the inverse functions and $\partial f_j / \partial w_k$ exists and is continuous for $j, k=1, 2, \dots, N$ in the domain defined by the possible values of the w_j for the study of interest. (Recall our earlier remarks concerning infeasible u and w .) In the sequel, we make the further assumption that the system (3.4) has continuous partial derivatives of order m ($m=2, \dots$, etc.) at each point in the domain of possible w values. The continuity assumption is valid if a continuous model is used to represent discrete changes in transformer tap settings and controllable reactive power sources. Network components are also considered to be in continuous service.

In order to minimize the notational burden, we have dropped any reference to t in equations (3.3) and (3.4). We shall continue this, although a relationship to t exists. Since any admittance associated with two buses that are not connected is zero the form of (3.3) can change as the system is switched into different configurations. Also the w change as t changes. In view of our assumptions, it is possible to express each f_j in a Taylor's expansion of order $m-1$ with a remainder that is a function of the m th differential. In practice one needs to study the remainder term in order to ascertain the possible

convergence of the Taylor's series. Thus we make the (valid) assumption concerning $\underline{m}$. In our initial simple example there was only one w which was the injected real power, P_t . We now have N w 's which in general comprise both real and reactive powers.

Now let

$$\underline{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix} = \begin{bmatrix} w_{1t} \\ w_{2t} \\ \vdots \\ w_{Nt} \end{bmatrix}$$

Employing the Wold decomposition (equation (2.18)) we can write $\underline{w}$ as

$$\begin{aligned} \underline{w} &= \underline{m} + \underline{x} + \underline{y} \\ &= \underline{\mu} + \underline{x} \end{aligned}$$

where

$$\underline{m} = \underline{m}_t = E(\underline{w})$$

$$\underline{x} = \underline{x}_t = \text{the purely nondeterministic part of } \underline{z}$$

$$\underline{y} = \underline{y}_t = \text{the deterministic part of } \underline{z}$$

$$\underline{z} = \underline{z}_t = \underline{w} - \underline{m}$$

and

$$\underline{\mu} = \underline{\mu}_t = \underline{m} + \underline{y}.$$

Then

$$f_j(\underline{w}) \approx f_j(\underline{\mu}) + \sum_{k=1}^{m-1} \frac{1}{k!} d^k f_j(\underline{\mu}; \underline{x}), \quad j=1, 2, \dots, N \quad (3.5)$$

from the multivariate version of the Taylor's expansion. In (3.5) $d^k f_j(\underline{\mu}; \underline{x})$ is the k th differential of f_j and is given by

$$d^1 f_j(\underline{\mu}; \underline{x}) = \sum_{\ell=1}^N D_{\ell} f_j(\underline{\mu}) x_{\ell}$$

$$d^2 f_j(\underline{\mu}; \underline{x}) = \sum_{\ell=1}^N \sum_{\ell'=1}^N D_{\ell, \ell'} f_j(\underline{\mu}) x_{\ell} x_{\ell'},$$

.

and

$$D_{\ell} f_j(\underline{\mu}) = \left. \frac{\partial f_j}{\partial w_{\ell}} \right|_{\underline{\mu}}$$

$$D_{\ell, \ell'} f_j(\underline{\mu}) = \left. \frac{\partial^2 f_j}{\partial w_{\ell} \partial w_{\ell'}} \right|_{\underline{\mu}}$$

.

Also

$$f_j(\underline{\mu}) = f_j(\mu_1, \mu_2, \dots, \mu_N),$$

$$f_j(\underline{w}) = f_j(w_1, w_2, \dots, w_N),$$

etc.

In (3.5) we have ignored the remainder term. Again as in the univariate case we have a nice expression relating any $u_j = f_j(\underline{w})$ to an overall mean level and a stochastic part.

3.2 Application of Decomposition to Power Flow Equations

Recall our complete decomposition procedure. We first remove $\underline{m}$ and then the deterministic part of $\underline{z}$. This can be done as in the Wold decomposition which as we have already indicated tacitly assumes that the process is asymptotically stationary or we can employ statistical procedures that assume $\underline{x}$ is stationary from the outset. In either case the method(s) of analysis is such that practically the resulting $\underline{x}$ is stationary. Speaking loosely the analysis places as much as possible of any nonstationary behavior in the deterministic part. This is enhanced by iterating the whole process and changing the original $\underline{m}$. Now the iteration process can either halt or continue indefinitely. Practically in the later case one would not continue on but stop after some "large" number of iterations. In this situation the methodology of this section would not be applicable in determining instantaneous probability statements since we cannot produce a suitable stationary and/or ergodic decomposition. However, one should note that in the former case the resulting analysis may be off unless some form of stationarity holds for $\underline{x}$, since one could obtain a spurious decomposition (even with a large sample size).

Now suppose that a decomposition can be found. It may turn out in many cases that the entries in $\underline{x}$ are uncorrelated since we are removing the deterministic part in a vector manner. If in addition $\underline{x}$ is normal and the first order differential in the Taylor's expansion gives an adequate approximation then the distribution of $f_j(\underline{w})$ is also (approximately) normal. But since the entries in $\underline{x}$ are uncorrelated it is straightforward to obtain the distribution of $f_j(\underline{w})$. Even if the entries in $\underline{x}$ are correlated if linearity and normality hold (approximately) then the problem can be considered in a vector sense and the resulting distribution of $\underline{u} = (u_1, \dots, u_N)' = (f_1(\underline{w}), \dots, f_N(\underline{w}))'$ is normal. However, now the computation of moments is more difficult.

If linearity is not reasonable then we can add higher order terms in (3.5). Then normality of $\underline{x}$ is no longer as important as in the linear case. In this situation if $\underline{x}$ is stationary then it is also ergodic (Hannan (1970) Chapter IV) so that moments can be determined by time averages. Equation (3.5) can be employed to transform input moments into output moments and then series expansions (Deutsch (1962) and Sauer (1977)) can be used to approximate the output distribution.

It may be that a suitable decomposition can be found but that the resulting $\underline{x}$ is neither stationary or normal. At first thought one might conclude that this is a situation of nonconvergence of the iterative procedure. However, it may be that the $A_j(t)$ do vary with t and the resulting $\underline{x}_t$ is not normal. The most straightforward way to determine this would be to have a sample with multiple independent realizations but it none the less could be the case. If indeed this is the situation, then normal theory methods should not be employed unless their efficacy is demonstrated since normality does not hold. Without repeated realizations (either independent or with their dependency known) it appears that there is no way to determine an instantaneous input distribution. Hence there

is no way to directly employ (3.5) in determining output distributions. This situation is a practical sign of nonconvergence of the iterative procedure. Indeed we shall assume that this is the only terminal state of the iterative procedure and that a suitable decomposition can always be found. It just may be that the resulting $\underline{x}$ is neither stationary nor normal. Then (if repeated realizations or not available) either the window or interval maximum transformations may be employed. In any event we have a qualitative decomposition of the $\{\underline{u} = \underline{u}_t\}$ process.

The purpose of the methodology presented in this section is to allow one to make instantaneous probability statements about any $\underline{u}$, etc. that is of interest. We have employed the Wold decomposition and a Taylor's expansion since we believe that there is a sufficient number of practical situations in which this methodology can be successfully employed. Indeed if work in other fields is any indication then the expectation of normality (and even independence) of $\underline{x}$ is not unreasonable. Thus analysis of typical power flow data sets is of real importance. Without such analysis there is no way to assess the usefulness of the proposed methodology.

CHAPTER IV WINDOW TRANSFORMATION METHOD

In this chapter we discuss an analysis methodology based on what we have termed the window transformation. This method can be employed in those situations where (due to nonstationarity and nonnormality) the direct transformation method is not feasible in determining instantaneous probability statements. Also it may be employed when one does not desire instantaneous probability information. In certain cases the output from this analysis can serve as a surrogate for probability information. Since the power flow equations, assumptions, etc. remain the same for this chapter as in the previous we shall not repeat them here and shall only discuss the actual methodology associated with the window transformation.

4.1 Development of Proportion of Time Distribution

For the time being consider a particular output $u_j = f_j(\underline{w})$ (voltage, power flow, etc.). We assume that a particular time frame is given during which there is interest in the nature of u_j . We call this time frame of interest the window. Now consider Figure 4.1 in which a possible trajectory of u_j is shown. The window goes from a to b . If arbitrary levels are selected and then inverted into the t axis as shown one obtains the proportion of time during the window that the trajectory was at or below the selected level. We assume as the Figure indicates that u_j is continuous. (Since we are only treating the discrete time situation in this study we automatically have continuity of any observed trajectory. What is important is the sampling rate which we shall discuss shortly.) If one considers all possible levels, then

Figure 4.1 Proportion of Time (Window) Transformation

a distribution is obtained that gives the proportion of time in the window that the trajectory was at or below any given level. Thus the output of this analysis is not a probability distribution that reflects the stochastic nature of the $\{u_j\}$ process but is a probability distribution that measures proportions of times a particular trajectory is at or below various levels.

The trajectory can be a realization of the $\{u_j\}$ process and hence the resulting distribution is random since it changes from realization to realization. However, the trajectory can be given by, say, μ_j and is in a sense nonrandom. The trajectories given by μ are similar to conditional expectations. In any event, if we consider only a particular past history, then μ is fixed and various kinds of proportion of times distributions can be obtained from the analysis.

In order to handle the vector case, one has to generalize the method to obtain joint distributions. However, this is not as straightforward as one might like. Conceptually there are at least two ways the procedure could be carried out, namely,

- o Directly transform the input vector to the output vector, that is, obtain u and then make a transformation to the joint distribution of proportion of times for voltages, etc.
- o Transform the input vector, w , into a joint distribution of proportion of times and then transform this distribution into an output distribution of proportion of times for voltages, etc.

Also a definition of what is meant by a multivariate or joint proportion of time distribution must be given.

We now investigate these matters. First we only allow windows that are closed intervals, say, $[a, b]$. Consider any trajectory, say, $u_j = f_j(\underline{w}) = u_j(t)$ and let Lebesgue measure be denoted by m . Set

$$A_j = A_j(\tilde{u}_j) = \{t \mid u_j(t) \leq \tilde{u}_j\}. \quad (4.1)$$

Now A_j is measurable since $u_j(t)$ is measurable. This follows since we have assumed that the $w_k = w_k(t)$ are continuous. (A continuous function of a measurable function is measurable.) We let the marginal proportion associated with u_j be defined by $(1/(b-a))m(A_j)$. If we had considered an input trajectory we would have an (input) marginal proportion defined by $(1/(b-a))m(B_k)$ where

$$B_k = B_k(\tilde{w}_k) = \{t \mid w_k(t) \leq \tilde{w}_k\} \quad (4.2)$$

We now define a joint (input proportion distribution on the w 's called Q and generated by $(1/(b-a))m\{B_k\}$ where

$$\bigcap_{k=1}^N B_k = \{t \mid t \in B_1 \text{ and } \dots \text{ and } B_N\} \quad (4.3)$$

and each B_k is defined as in (4.2). The three measures associated with (4.1), (4.2) and (4.3) are generated in the usual way from the measurable sets generated by (4.1), (4.2) and (4.3). If all realizations are as in Figure 1 then we see that (4.3) gives the proportion of time in the interval $[a, b]$ that all conditions, $w_k(t) \leq \tilde{w}_k$, are met as a measure of the length of the intersected intervals of these conditions in $[a, b]$. The generated measure is a generalization of this and is hence intuitively appealing. However, if this joint measure is to be useful it must be consistent with the output measure generated from (4.1).

We consider this. Set

$$C_j = C_j(\tilde{u}_j) = \{u_j \mid u_j \leq \tilde{u}_j\}.$$

Recall that

$$f^{-1}(C_j) = \{w \mid f_j(w_1, \dots, w_N) \in C_j\}.$$

Let P be the "probability" proportion of C_j generated from the input joint distribution. Then

$$P\{C_j\} = Q\{f^{-1}(C_j)\}$$

from probability theory. Thus

$$P\{C_j\} = (1/(b-a)) \int_a^b \{t \mid f_j(w_1(t), \dots, w_N(t)) \in C_j\}$$

from the definition of Q .

But

$$A_j = \{t \mid u_j(t) \leq \tilde{u}_j\}$$

so that

$$A_j = \{t \mid f_j(w_1(t), \dots, w_N(t)) \in C_j\}$$

Therefore we have the result that

$$P\{C_j\} = (1/(b-a)) \int_a^b A_j$$

and P is indeed consistent with the measure generated from (4.1). In a similar manner one can show that the joint output and input proportions are consistent if defined in the manner that we have indicated. Thus either of the two methods that we have proposed can be (theoretically) employed with the given definitions.

4.2 Singularity of Proportion of Time Distribution

We have just shown that both the input and output joint proportion of time distributions can be consistently generated from Lebesgue measure on the interval $[a, b]$. Either of these can be transformed into a measure on or associated with the curve (we consider this curve in more detail shortly) generated by the trajectories. The measure associated with the curve is just the measure generated by the inverse transformation (generally not a function) from the curve to the interval $[a, b]$. The measure of any measurable set in the space of the curve is the measure of the set in $[a, b]$ that would map (under the curve transformation) into the given measurable set. This follows in a straightforward way from probability theory.

Now while the theory is straightforward, the derivation of these joint proportion of time distributions is not as straightforward. In view of our assumptions, both input and output trajectories are

continuous. Speaking loosely this gives the result that a joint proportion of time distribution (as defined herein) is singular. Consider the input situation. The set of $w_k(t)$, $t \in [a, b]$, $k = 1, 2, \dots, N$ define parametrically a continuous curve in N -dimensional space. It is straightforward to show that this curve comprises a compact, separable, connected set in R^N . Furthermore its topological dimension is 1. From topology it then follows that the curve is topologically contained in (homeomorphic to some subset of) the cube I^3 . Thus if $N > 3$, then the joint distribution must be singular and this may be the case for $N < 3$. Unfortunately we must make further assumptions concerning the trajectories if we are to resolve the situation $N \leq 3$ since space (plane) filling continuous curves exist. We also need to avoid the space filling situation if we hope to obtain computational algorithms.

We shall additionally assume that the trajectories are of bounded variation. While this is not quite equivalent to assuming the trajectories are indefinite integrals, it comes close. In fact, if a function is of bounded variation on an interval, then its derivative exists a.e. (Lebesgue measure on the interval). And if a function is continuous and of bounded variation and has no singular part (in a measure theory sense), then it is an indefinite integral. Since the input trajectories are loads or controlled voltages, the assumption of bounded variation is not unreasonable.

Now consider an output trajectory, $u_j = f_j(w)$. We have for a partition, $a = t_0 < t_1 < \dots < t_n = b$ of $[a, b]$

$$\begin{aligned} & \sum_{\ell=1}^n [f_j(w_1(t_\ell), \dots, w_N(t_\ell)) - f_j(w_1(t_{\ell-1}), \\ & \quad \dots, w_N(t_{\ell-1}))] \\ &= \sum_{\ell=1}^n \sum_{m=1}^N \left. \frac{\partial f_j}{\partial w_m} \right|_{t'_m} [w_m(t_\ell) - w_m(t_{\ell-1})] \end{aligned}$$

where $t'_m \in (t_{\ell-1}, t_\ell)$ from the mean value theorem. But all partials are continuous on $[a, b]$ and hence bounded. Thus there exists a bound, A , such that

$$\left| \frac{\partial f_j(t)}{\partial w_m} \right| \leq A, \quad t \in (a, b) \text{ and } m = 1, 2, \dots, N.$$

Then

$$\begin{aligned} & \sum_{\ell=1}^n |f_j(w(t_\ell)) - f_j(w(t_{\ell-1}))| \\ & \leq A \sum_{m=1}^N \sum_{\ell=1}^n |w_m(t_\ell) - w_m(t_{\ell-1})| \end{aligned}$$

But each w is of bounded variation by assumption. Thus the output trajectories are of bounded variation since the input trajectories are.

With these assumptions the curve determined by the trajectories is nice in that it is not space (plane) filling. Also, the curve is rectifiable. All of the proportion of time measure must be associated with this curve since it determines the coordinate values that occur together at any time. Thus the distribution is singular for $N > 2$. The computation of this distribution is complicated by the fact that

values of the curve can be repeated (and this possibility is only enhanced by sampling). If a way can be found to take into account the repeats, then it should be possible to compute the distribution in a straightforward manner. Since the distribution is singular for $N > 2$, one is in effect only dealing with a one dimensional distribution.

4.3 Development of the Singular Distribution

Since the proportion of time distribution is in effect one dimensional, one should be able to compute proportion of times as is done for univariate random variables. We investigate these possibilities in a simple example. Consider the input situation and let $N=2$. Suppose that

$$w_1(t) = \frac{1}{2}t$$

and

$$w_2(t) = -\frac{1}{2}t + \frac{1}{2}$$

with

$$t \in [0, 1].$$

We see that

$$w_1(t) + w_2(t) = \frac{1}{2}$$

for all t and, hence, for $t \in [0, 1]$. In general, this simple a relationship among the trajectories would not exist. If we consider the 2-dimensional space of values of the curve, say, $\tilde{w}_1 \times \tilde{w}_2$, the curve is a straight line that connects $(0, \frac{1}{2})$ to $(\frac{1}{2}, 0)$ as shown in Figure 4.2. We note that this example has no repeat values of $(\tilde{w}_1, \tilde{w}_2)$.

Figure 4.2 Example Curve

We now need to find the joint proportion of time distribution for various values of $\tilde{w}_1$ and $\tilde{w}_2$ and then see how it is related to the curve. Let

$$0 \leq \tilde{w}_1 \leq \frac{1}{2}$$

and

$$0 \leq \tilde{w}_2 \leq \frac{1}{2}$$

and consider $\text{Prop}(\tilde{w}_1 \leq \tilde{w}_1 \text{ and } \tilde{w}_2 \leq \tilde{w}_2)$

where

Prop = proportion of time

and

$\tilde{w}_j$ = a random variable whose values are w_j
 $j = 1, 2$.

Now $w_1(t) \leq \tilde{w}_1$ if and only if $0 \leq t \leq 2\tilde{w}_1$. Also, $w_2(t) \leq \tilde{w}_2$ if and only if $(1 - 2\tilde{w}_2) \leq t \leq 1/2$. (Recall that $t \in [0, 1]$ in the example.) Thus

$$\begin{aligned} \text{Prop}(\tilde{w}_1 \leq \tilde{w}_1 \text{ and } \tilde{w}_2 \leq \tilde{w}_2) &= \frac{m\{[t | 0 \leq t \leq 2\tilde{w}_1] \cap [t | (1 - 2\tilde{w}_2) \leq t \leq 1]\}}{1} \\ &= \begin{cases} 2(\tilde{w}_1 + \tilde{w}_2) - 1, & \tilde{w}_1 + \tilde{w}_2 \geq \frac{1}{2} \\ 0, & \tilde{w}_1 + \tilde{w}_2 \leq \frac{1}{2} \end{cases} \end{aligned}$$

If either (or both) $\tilde{w}_1$ or $\tilde{w}_2$ are less than zero, then the joint distribution is zero. Also, if both $\tilde{w}_1$ and $\tilde{w}_2$ are greater than $\frac{1}{2}$, then the joint distribution is one. The case $\tilde{w}_2 > \frac{1}{2}$ gives the marginal for $\tilde{w}_1$ and the case $\tilde{w}_1 > \frac{1}{2}$ gives the marginal for $\tilde{w}_2$. These marginals are

$$\text{Prop}(\tilde{w}_1 \leq \tilde{w}_1) = \begin{cases} 1, & \tilde{w}_1 > \frac{1}{2} \\ 2\tilde{w}_1, & 0 \leq \tilde{w}_1 \leq \frac{1}{2} \\ 0, & \tilde{w}_1 < 0 \end{cases}$$

and

$$\text{Prop}(\tilde{w}_2 \leq \tilde{w}_2) = \begin{cases} 1, & \tilde{w}_2 > \frac{1}{2} \\ 2\tilde{w}_2, & 0 \leq \tilde{w}_2 \leq \frac{1}{2} \\ 0, & \tilde{w}_2 < 0 \end{cases}$$

We indicate the results in Figure 4.3.

We now note that for this example if we wish to compute the joint proportion of time for any $\tilde{w}_1 \times \tilde{w}_2$ set, we take the intersection of said set with the curve and use the time points at the boundaries of the intersection to compute the proportion of time. For example, suppose we want $\text{Prop}(\tilde{w}_1 \leq \frac{3}{8} \text{ and } \tilde{w}_2 \leq \frac{1}{4})$. Then by examining Figure 4.3 we see that this is given by

$$\begin{aligned} \text{Prop}(\tilde{w}_1 \leq \frac{3}{8} \text{ and } \tilde{w}_2 \leq \frac{1}{4}) &= 2(\frac{3}{8} + \frac{1}{4}) - 1 \\ &= \frac{1}{4} \end{aligned}$$

Figure 4.3 Joint Proportion of Time Distribution for Example

Now the intersection of the set

$$\{(\tilde{w}_1, \tilde{w}_2) | 0 \leq \tilde{w}_1 \leq \frac{3}{8} \text{ and } 0 \leq \tilde{w}_2 \leq \frac{1}{4}\}$$

with the curve gives the line segment from $(\frac{1}{4}, \frac{1}{4})$ to $(\frac{3}{8}, \frac{1}{8})$. The time point associated with $(\frac{1}{4}, \frac{1}{4})$ is $t = \frac{1}{2}$ and the time point associated with $(\frac{3}{8}, \frac{1}{8})$ is $t = \frac{3}{4}$. Now all points on the curve occur between the intersection endpoints and since the curve is continuous, all time points between $t = \frac{1}{2}$ and $t = \frac{3}{4}$ must meet the criterion $0 \leq \tilde{w}_1 \leq \frac{3}{8}$ and $0 \leq \tilde{w}_2 \leq \frac{1}{4}$. Thus the proportion of time measure is given by $(\frac{3}{4} - \frac{1}{2})/1 = \frac{1}{4}$.

This procedure (as we have indicated) works in general for this example. However, this is so since there is a nice progression of time points along the curve and the curve is without repeats. Also, the fact that the curve is rectifiable means that it cannot oscillate too wildly. Indeed, for this example there are no oscillations. These properties produce another interesting result. The length of the intersected curve can be used to determine the proportion of time measure as follows. Consider the length of the line segment from $(\frac{1}{4}, \frac{1}{4})$ to $(\frac{3}{8}, \frac{1}{8})$. It is easily computed to be $\sqrt{2}/8$. Moreover, the length of the curve is $\sqrt{2}/2$. We then have

$$\begin{aligned} (\sqrt{2}/8)/(\sqrt{2}/2) &= \frac{1}{4} \\ &= \text{Prop}(\tilde{w}_1 \leq \frac{3}{8} \text{ and } \tilde{w}_2 \leq \frac{1}{4}). \end{aligned}$$

Since there is a constant proportionality between time points in $[0,1]$ and points along the curve, this procedure works for this example. Unfortunately, this constant proportionality does not hold in general. We need to associate a measure with the curve that reflects the amount of time taken in going from point to point.

Before we investigate a time measure associated with the curve, we formulate a more general definition of the curve that accounts for repeats of $\tilde{w}_k$ values.

Since time values are unique, we consider the curve in the $(N+1)$ -dimensional space of the values of the curve and the time axis. For the example this gives the curve shown in Figure 4.4. This curve automatically takes care of any repeats of $\tilde{w}_k$ values. Since the curve in the N -dimensional space is rectifiable, it follows that the curve in the $(N+1)$ -dimensional space which considers time values is also rectifiable. Furthermore, the mapping from $[a,b]$ to $S = \{w_1(t), \dots, w_N(t), \text{ to } t \in [a,b]\}$ (for the input situation) is one-to-one, onto, and continuous. Also the inverse mapping to this mapping is continuous. Therefore, the two sets are homeomorphic. If we let the measurable sets in $[a,b]$ consist of the sigma algebra of sets (the Borel sets) generated by the open sets, then the measurable sets in S also consist of the Borel sets in S . For these developments we let the metric in $[a,b]$ be the usual metric and in S employ arc length of the curve in the $(N+1)$ -dimensional space as the metric.

Figure 4.4 Curve of Example Showing Time Dependence

We need to now investigate the possibility of assigning a measure to the curve in $N+1$ dimensions (hereafter called the curve) that reflects proportion of time. We shall approach the problem from essentially a sampling viewpoint, but before proceeding we note that:

- (i) The curve maintains an ordered time progression.

Again we employ the example. The curve now goes from $(0, \frac{1}{2}, 0)$ to $(\frac{1}{2}, 0, 1)$ and is easily shown to have length $\sqrt{6}/2$. Suppose the time interval is broken into ten equally spaced increments, each of length $\frac{1}{10}$ second. (For this there would be 11 samples starting at $t=0$ and ending at $t=1$.) Each corresponding increment on the curve has length $\sqrt{6}/20$ since the curve is a straight line. In general, this would not be the case and the curve incremental lengths would vary.

Now form the ratio $(\frac{1}{10})/(\sqrt{6}/20)$. This ratio gives the per incremental proportion of time consumed in going from start to finish of any increment on the curve. The area (amount of time consumed) is given by

$$[(\frac{1}{10})/(\sqrt{6}/20)] (\sqrt{6}/20) = \frac{1}{10}.$$

In general, we let

$h_j \ell_j$ = the time consumed in going from start to finish of increment j on the curve
 = a constant c with uniform sampling.

Then we have

h_j = c/ℓ_j
 = the height of a histogram approximation to the proportion of time distribution on the curve.

In this derivation ℓ_j is the incremental length along a polygonal approximation to the curve. We see that in order to compute the cumulative proportion of time consumed in going to any sampling instant we form

$$\sum_{j=1}^J h_j \ell_j$$

where J denotes the final sampling instant. We have that

$$\begin{aligned} \sum_{j=1}^J h_j \ell_j &= \sum_{j=1}^J \left(\frac{c}{\ell_j}\right) \ell_j \\ &= Jc \end{aligned}$$

= the exact amount of time consumed in getting to sampling instance J proceeding along the curve.

We record this property as

- (ii) The histogram approximation to the proportion of time distribution is exact at sampling instants along the curve. This is not true in general for points in between sampling instants.

We finally note that

- (iii) Any point on the curve is approximated by a sequence of sampling points if we allow the sampling rate to approach infinity. (Recall we are assuming uniform sampling.)

Thus we conclude from (i), (ii), and (iii) that a sequence of histogram approximations with a sampling rate that approaches infinity converges in distribution to a proportion of time measure on the curve that reflects the amount of time taken in going from point to point on the curve.

We now use this measure (or the approximation to the measure) as we did for the example with the curve in N dimensions. However, now we must, in effect, consider the projection of any point on the curve into the plane defined by the N -dimensional space of $\tilde{w}_k$ values. This will take care of repeated $\tilde{w}_k$ values.

Thus, in the example, we must consider all points on the curve that are such that $\tilde{w}_1 \leq 3/8$ and $\tilde{w}_2 \leq 1/4$.

This encompasses all points between $(1/4, 1/4, 1/2)$ and $(3/8, 1/8, 3/4)$. Now in the example the proportion of time measure is directly proportional to length so that we only need to compute the length between these two points and divide by $\sqrt{6}/2$. The length from $(1/4, 1/4, 1/2)$ to $(3/8, 1/8, 3/4)$ is

$$\sqrt{\left(-\frac{1}{8}\right)^2 + \left(\frac{1}{8}\right)^2 + \left(-\frac{1}{4}\right)^2} = \sqrt{6}/8.$$

Then

$$\begin{aligned} \text{Prop}(\tilde{w}_1 \leq \frac{3}{8} \text{ and } \tilde{w}_2 \leq \frac{1}{4}) &= (\sqrt{6}/8)/(\sqrt{6}/2) \\ &= \frac{1}{4} \end{aligned}$$

which is the correct answer. In general, the curve is not straight and has curvature. In this situation we employ a histogram approximation to the true measure and proceed as outlined. We know that the approximation can be made more and more exact by increasing the sampling rate. Also as we increase the sampling rate, the polygonal approximation to the curve and its length become increasingly better.

4.4 Application of the Singular Distribution

Investigation of the two proposed methods is now possible. If method one is used then either realizations of output voltages, etc. must be observed or a number of power flow analyses made. Normally observations are not available for many of the voltages, etc. that are of interest and this is one reason why power flow analyses are made. Since the equivalent of a set of complete trajectories must be known in order to apply this method, in theory a power flow solution must be made at each sampling instant. However, if no input value changes from one sampling instant to the next then no new solution is required since the form of the equations is constant over $[a, b]$.

If method two is used, then the input proportion of time distribution must be transformed into the output distribution. Since we are approximating the input distribution by a histogram (which is a density), we can employ the change of variable theorem. The input space must be broken into regions wherein the determinant of the Jacobian matrix is of constant sign. But the Jacobian matrix (for a fixed set of equations) is nonsingular for all u and w values in view of our assumptions. Thus the determinant cannot change sign and all that is required is that the input histogram be multiplied by the magnitude of the Jacobian determinant to obtain the output joint distribution. Then by proper summing in a one dimensional space the distribution of any u_j or set of u_j 's can be obtained. Thus we see that in practice the critical tradeoff between method one and two is between the number of power flow solutions required and the amount of summation required. We note that technically speaking for this method a power flow solution is required at each sampling instant since the Jacobian determinant must be evaluated. However, in practice it may be possible to approximate the determinant in a way that does not require power flow solutions at each instant of time. We believe that method two is probably the better of the two.

4.5 Proportion of Time Distribution of Events

We now consider the nature of the proportion of time distribution as defined herein. Since an actual (or potentially actual) realization is employed in defining the proportion of time distribution, it turns out that there is a random distribution for proportion of time information. At least two approaches look promising in dealing with this distribution of a distribution. One has a sampling basis and the other is more-or-less theoretical in nature and employs the decomposition developed in this report. We call these two methodologies approaches. Each could be carried out or conceptually carried out by two possible procedures: (i) obtain a proportion of time distribution for each sample (or conceptual sample) and then average these distributions, or (ii) average the samples (or conceptual samples) first to obtain an average curve (this estimates a mean curve) and then determine a proportion of time distribution from the average curve. As we eventually show, (i) gives an unbiased estimate of the "true" but unknown proportion of time distribution. Procedure (ii) is thus biased, but when employed with approach two, is in certain cases approximately unbiased.

If several (the number could be determined from sampling theory) independent realizations of the stochastic process are available, one could form estimates of the "true" but unknown proportion of time distribution. This is analogous to estimating, say, the unknown mean of a population. However, in this situation we are now estimating an unknown function and not just an unknown constant.

One possible way to carry out this approach is to assume a certain set of days, say, all the Mondays in a particular quarter, for example, are random samples from a possible population of Mondays. Since the proportion of time distribution represents a nonlinear transformation, one should obtain the proportion of time distribution for each sample Monday and then average these distributions to determine the estimated proportion of time distribution. The procedure: average the data first to obtain an average curve and then determine the proportion of time distribution, could possibly be used, but is may not be equivalent to the given procedure. We shall investigate this further shortly. In any event, one can obtain an average proportion of time distribution and from sampling theory determine confidence intervals, etc.

The second approach can be thought of as having a sampling background. However, the population in this approach differs from that of the first approach. In order to explain this population let us consider a simple example of the decomposition in a nonstationary setting (eqs. 2.12 - 2.14). Suppose

$$z_t = z_{t-1} + e_t$$

with

$$z_0 = \text{a given initial value that is independent of } e_t$$

e_t = a white noise process.

In this example $t=0$ represents the infinite or remote past. Clearly z_t is second order and nonstationary. Indeed z_t is just one of the many models of the random walk in one dimension.

It is straightforward to show that

$$\hat{z}_t = z_{t-1}$$

$$e_t = z_t - \hat{z}_t = e_t$$

and

$$x_t = e_t + e_{t-1} + \dots + e_1.$$

Thus

$$y_t = z_t - x_t = z_0.$$

Now for any particular realization the value of z_0 is fixed for all future time. (This is the inherent nature of the perfectly predictable part of the process.) If one is at $t=\text{now}$, then it must happen that the value of z_0 is perpetuated. Also any nonzero mean function must continue in the future according to its functional representation. Thus we take these two quantities as fixed and the only future variation we consider is that due to x_t .

Then the population we consider is the set of values generated by x_t for all t and producing the variation about $\mu_t = m_t + y_t$. We note that the mean of this population is μ_t and is not m_t the mean of the process, or (if normality holds) the conditional mean of the process given past values up to t . The conditional mean is the mean square (or least square) error forecast and when normality holds given by the linear decomposition. The conditional mean takes into account any correlation among the x_t at differing time points. We in effect let the x_t part of the process be started over again from sample to sample and keep μ_t fixed from sample to sample. This population is indicative of what could happen given the past up to t but only using part of that past.

We note that if this approach is to make practical sense then the bulk of the variation in the process must be associated with μ_t . In the example this means that we are considering a random walk that with respect to the complete data set has not drifted far from z_0 . In some situations one might want to take into account past e_t values. Here μ_t is kept fixed from sample to sample and x_t is generated conditional on past e_t values. Again in terms of the example this means we consider all future values to be a drift about z_{t-1} . If normality holds this is just the MSE forecast of future values for the random walk and is the conditional mean of the process. In general the linear decomposition does not give the conditional mean. None-the-less with a suitable interpretation μ_t or μ_t with conditioning on past e_t values can be thought of as the mean of a particular population.

Thus the decomposition either gives or is the basis for a future mean function. This future mean function can be employed to generate the curve in $(N+1)$ - dimensional space that is the basis of a proportion of time distribution. We note that this proportion of time distribution may not be the one obtained if one could actually sample the defined populations, determine a proportion of time distribution for each sample, and then average these distributions. (Recall our earlier comments.) We employ the approach just explained to obtain a proportion of time distribution associated with a future mean function.

We now investigate the two possible procedures of estimating a proportion of time distribution (as opposed to the two general approaches). We consider

the input situation and let

$$\eta_{ij}(t) = \begin{cases} 1, & w_{ij}(t) \leq \tilde{w}_i \quad i = 1, 2, \dots, N, \\ 0, & \text{otherwise} \quad j = 1, 2, \dots, n \end{cases}$$

where n is the sample size, j indicates sample number, and the other quantities have their earlier meanings. Set

$$\eta_j(t, \underline{w}) = \prod_{i=1}^N \eta_{ij}(t).$$

Then $\eta_j(t, \underline{w}) = 1$ if and only if $w_{ij}(t) \leq \tilde{w}_i$ for every $i, i=1, 2, \dots, N$ and

$$Z_j(T, \underline{w}) = \frac{1}{T} \int_0^T \eta_j(t, \underline{w}) dt$$

gives the proportion of time in $[0, T]$ that $w_{ij}(t) \leq \tilde{w}_i$ for every i . Also

$$\begin{aligned} \bar{Z}(T, \underline{w}) &= \frac{1}{n} \sum_{j=1}^n Z_j(T, \underline{w}) \\ &= \frac{1}{n} \sum_{j=1}^n \frac{1}{T} \int_0^T \eta_j(t, \underline{w}) dt \\ &= \frac{1}{T} \int_0^T \frac{1}{n} \sum_{j=1}^n \eta_j(t, \underline{w}) dt. \end{aligned}$$

Now consider

$$\bar{\eta}_i(t) = \begin{cases} 1, & \frac{1}{n} \sum_{j=1}^n w_{ij}(t) \leq \tilde{w}_i \\ 0, & \text{otherwise} \end{cases} \quad i = 1, 2, \dots, N$$

which is associated with the average curve. Let

$$\bar{\eta}_{..}(t, \underline{w}) = \prod_{i=1}^N \bar{\eta}_i(t)$$

which is one if and only if each $\bar{\eta}_i(t)$ is one or if and only if the average curve is such that $\frac{1}{n} \sum_{j=1}^n w_{ij}(t) \leq \tilde{w}_i$ for every $i, i=1, 2, \dots, N$. Thus the

proportion of time distribution for the average curve is given by

$$\bar{Z}_{..}(T, \underline{w}) = \frac{1}{T} \int_0^T \bar{\eta}_{..}(t, \underline{w}) dt.$$

Now if we require that $\bar{Z}_{..}(T, \underline{w}) = \bar{Z}(T, \underline{w})$ for all possible T values then it must be that

$$\bar{\eta}_{..}(t, \underline{w}) = \frac{1}{n} \sum_{j=1}^n \eta_j(t, \underline{w})$$

for every t . Unfortunately, this is not true in general.

We have that

$$E \bar{Z}(T, \underline{w}) = E \left\{ \frac{1}{T} \int_0^T \frac{1}{n} \sum_{j=1}^n \eta_j(t, \underline{w}) dt \right\}$$

where $\xi_i(t)$ is a stochastic process for $i=1, 2, \dots, N$ given by one of the sampling schemes and populations that we are studying. Thus

$$E \bar{Z}(T, \underline{w}) = \frac{1}{T} \int_0^T P\{\xi_1(t) \leq \tilde{w}_1, \dots, \xi_N(t) \leq \tilde{w}_N\} dt \quad (4.4)$$

so that $\bar{Z}(T, \underline{w})$ is an unbiased estimator. If we standardize the integrand we obtain

$$\begin{aligned} &P\{\xi_1(t) \leq \tilde{w}_1, \dots, \xi_N(t) \leq \tilde{w}_N\} \\ &= P\left\{ \frac{\xi_1(t) - \mu_{1t}}{\sigma_{1t}} \leq \frac{\tilde{w}_1 - \mu_{1t}}{\sigma_{1t}}, \dots, \frac{\xi_N(t) - \mu_{Nt}}{\sigma_{Nt}} \leq \frac{\tilde{w}_N - \mu_{Nt}}{\sigma_{Nt}} \right\}. \end{aligned}$$

Let

$$\sigma_M = \max \{\sigma_{1t}, \dots, \sigma_{Nt} \mid t \in [0, T]\}.$$

Then if $\sigma_M \rightarrow 0$, we see that

$$\begin{aligned} &P\{\xi_1(t) \leq \tilde{w}_1, \dots, \xi_N(t) \leq \tilde{w}_N\} \\ &\rightarrow \begin{cases} 1, & \tilde{w}_i - \mu_{it} > 0 \text{ for every } i \\ 0, & \tilde{w}_i - \mu_{it} < 0 \text{ for at least one } i \end{cases} \\ &\approx \bar{\eta}_{..}(t, \underline{w}) \text{ for large } n. \end{aligned}$$

Finally we see that for large n and the bulk of the variation associated with μ_t that the decomposition approach is approximately unbiased. Note, however, for this approach that no sample is actually taken and this is part of the usefulness of the decomposition approach.

CHAPTER V INTERVAL MAXIMUM TRANSFORMATION

In this method of analysis the only data recorded is the maximum value of a trajectory in some interval. If input loads are considered then as shown in Figure 5.1 one generally loses any temporal relationship since the maxima do not occur at the same time. Thus there is no way to transform input into output and therefore use is restricted to output data sets. This implies that possibly many power flow analyses must be made. As the analysis of the previous section indicated this may not rule out the use of this transformation but care must be employed in its use.

Another caution is of interest as well. The problem of obtaining the distribution of the maximum

over an interval of a stochastic process from the distribution of the process itself is an extremely difficult problem. Possibly the most complete account is given by Cramér and Leadbetter (1967), especially Chapter 12. Generally only asymptotic results are available and these for only the univariate case. Thus more than likely one will only consider sample results obtained from observing the maxima over a number of windows. This has the possibility of further increasing the computational burden if information is required on many voltages, etc. in the system. Thus this transformation will most often be used in association with, say, daily peak loads, etc. and will not be employed in studying daily peak voltages, etc. Therefore, we shall not consider it further herein.

Figure 5.1 Interval Maximum Transformation

CHAPTER VI NONLINEAR EFFECTS

The objective of this chapter is to analyze the errors introduced by the linearization of the power flow equations used in the stochastic power flow problem. The linear model is compared with the exact one through a Monte Carlo Simulation. Several results are presented and discussed.

6.1 Basic Approach

The stochastic power flow (spf) model which is well known (Barkowska, 1974; Allan and Al-Shakarchi, 1976; Dapazo, et. al., 1975; Sauer, 1977) may be interpreted as an infinite number of linear load flows combined appropriately (probabilistically). This combination is done in a very practical way by convolution technique. The question addressed here is how well does this linear model perform?

Since the load flow equations are linearized around the expected value of the input quantities, any movement away from this region produces a linearization error. Consequently the error is a function of the region of uncertainty of the input quantities which can be measured by the variance. The effects of the bounds of the region is determined by the non-linear behavior of the system power flows and voltages, as depicted in Figure 6.1.

One practical way of comparing the linear model and the exact one is through a Monte Carlo simulation which consists of running and probabilistically processing several cases of individual load flows where the data are generated by pseudo-random numbers. A linearized Monte Carlo solution and an exact model Monte Carlo solution are computed for comparison. The computational procedure is charted in Figure 6.2. Using the strong law of large numbers, it can be shown that the exact probabilistic solution R_E will converge in probability to the Monte Carlo solution using the exact model R_e within the accuracy of the random number generator. Also the linear solution using the spf approach R_L will converge in probability to the Monte Carlo solution using the linear model R_ℓ . This can be expressed as follows:

$$\lim_{N \rightarrow \infty} P \{ |R_E - R_e| < \epsilon \} = 1$$

$$\lim_{N \rightarrow \infty} P \{ |R_L - R_\ell| < \epsilon \} = 1$$

where N is the number of trials used to evaluate R_e and R_ℓ , and ϵ is any positive real number. It should be noted that R is some measurable quantity relating to the uncertainty of the random variable, such as mean, variance, higher moments, etc.

In practical terms, it is impossible to evaluate the difference between R_L and R_E because of the computation time and the finite length of the pseudo-random series. It is therefore necessary to establish the largest possible number N that is realistic and practical to use. This requires:

- i - A comparison of a large number of possible cases.
- ii - Avoidance of introducing another error by the fact that the point of linearization will not in general coincide with the mean value of the pseudo-random Monte Carlo input data.

6.2 Simulation Parameters

This section defines various measurement parameters used to evaluate the performance of the linearized simulation and the nonlinear simulation. The power system under consideration is a 14 bus network described by the data in Appendix B.

The controlled variable in both simulations is a scaling factor ρ which adjusts the variance of each input random variable. As this scaling factor is increased, the range of input uncertainty stretches further away from the point of linearization. The behavior of the performance functions may then be analyzed as the scaling is increased or diminished.

The performance functions are defined to measure the cumulative behavior differences between the linearized and exact simulations. The input variables in this simulation are 21 active and reactive powers injected at the buses. The output variables under consideration are the 12 active power flows in the transmission system. The performance functions are defined below.

Input Variance Level:

$$\sigma_i^2 = (\rho \sigma_i^0)^2$$

Output Cumulative Differences:

$$\epsilon_\sigma = \frac{1}{n} \sum_{i=1}^n (| \text{exact } \sigma_i - \text{linear } \sigma_i | / \text{exact } \sigma_i) \times 100\%$$

$$\epsilon_\mu = \frac{1}{n} \sum_{i=1}^n (| \text{exact } \mu_i - \text{linear } \mu_i | / \text{exact } \mu_i) \times 100\%$$

where

m = number in input random variables (nodal injections)

n = number of output random variables (active line flows)

σ_i^2 = variance of input or output variable i

μ_i = expected value of input or output variable i

ρ = scaling factor for input variance

σ_i^0 = base case input standard deviation

The following two sections present the behavior of the performance functions as ρ is varied, and the distortions of output density functions due to non-linearity.

Figure 6.1 Linear and Non Linear Mappings

Figure 6.2 Simulation Flowchart

6.3 Performance Function Analysis

In this simulation, the value of the scaling factor is increased from 0 (deterministic simulation) to the value of 15. The two performance functions, which measure the cumulative differences between output mean and variance calculated from linear and exact simulations, are plotted in Figure 6.3. The number of Monte Carlo trials for each ρ is 5000.

Figure 6.3 Performance Functions

The average error produced by assuming linearity in the network model was less than 4% for the expected value and less than 2% for the standard deviation. Certainly these errors are system dependent and data dependent, however, they do provide certain insights into the behavior of the linearity error as a function of the variance of the input distributions. It is clearly seen here that the error increases dramatically as input variance is scaled upward (higher values of ρ). This is a natural result of linearization about an operating point, as depicted in Figure 6.4. This figure portrays a linear approximation of a single variable non-linear function. At the point of linearization, the curves have the same value and the same first derivative. If the value of the input variable x is restricted to small deviations about the point of linearization x_0 , then the error of the linearized representation of y is kept small. Larger deviations (higher values of ρ) permit larger deviations between the actual characteristic and its linearized approximation.

It can also be shown from Figure 6.4 that the expected value should be influenced by the linearization more so than the higher moments. As seen from the dashed lines, the linear approximation produces a definite shift in the range of y for the same domain of the variable x . Although this shift is nonlinear, it tends to affect the position of the distribution of y more than the shape of the distribution.

Again this result is data dependent. The shape as well as range of the distribution of x must be considered, since the largest deformation occurs at the extreme regions of x . If the density of x in these regions has small area (probability), a smaller linearization error could be expected compared to the case where significant probability mass exists in the tail regions.

Figure 6.4 Effects of Non Linearity

6.4 Output Distribution Deformations

In addition to performance functions, the effects of non-linearity can be shown in the probability distributions of the output variates as well. This section presents a comparison of distributions of power flow in the line connecting bus 1 with bus 2, and the line connecting bus 9 with bus 14.

The comparisons are presented graphically in Figure 6.5 for line 1-2, and Figure 6.6 for line 9-14. Each probability distribution shown was generated from 5000 Monte Carlo trials based on the same random sequence. The scaling factor for the input variance for these results is unity.

The curves in Figure 6.5 show that the linear model closely copies the performance of the actual nonlinear network model over the entire domain of the distribution. In fact, when these curves are integrated to calculate probabilities, there is little visible difference between them. It can be shown that as the scaling factor is increased, the deformation becomes more noticeable in the tail regions. This effect is minimal, however, because most of the probability mass is centered around the point of linearization (expected value).

Figure 6.5 Flow Density for Line 1-2 ($\rho=1$)

Figure 6.6 Flow Density for Line 9-14 ($\rho=1$)

The curves of Figure 6.6 demonstrate the effects of having a high concentration of probability mass in the extreme regions of the domain. It is clear that most of the distortion appears in these tail regions, while the region near the point of linearization shows minimal difference. Also note that the effect of linearization is primarily a shift, and not as much a change in shape. It is also interesting to note that while this distribution is the result of a linear combination of twenty-one random variables (all but one are normal), the form of the density is grossly non-normal.

6.5 Remarks

The results of this section serve as a warning, rather than an analysis of the nonlinearity of the input-output transformation. Faced with this, two alternatives come to mind: develop an efficient way to account for nonlinearity, or ignore nonlinear effects while using proper precautions.

Several articles have addressed the first alternative with some success. The approach that most have taken is to include the quadratic terms of the multi-variable Taylor's expansion of equation (3.5). Although this does not include the total effect of the nonlinearity of equations (3.1) and (3.2), it does represent all second order nonlinearity. It should be noted, however, that such a formulation is an awesome computational feat. The size of the second order differential array for the inverse function is $(n^2 + 3n^2 + 2n)/6$ considering symmetry. Any formulation of this type would experience great difficulty with even moderately sized power systems of 50 equations or more.

Other techniques for including nonlinearity in the transformation are suggested here, but have not appeared in engineering literature to date, and have not been seriously studied in this research. One suggestion is to consider multiple points of linearization. This idea can be demonstrated by a simple example. Consider the relationship

$$y = \sqrt{x} = g(x) \quad (6.1)$$

where y is a uniformly distributed random variable between $y=1$ and $y=2$. We can determine the exact distribution of x for this simple case using the change of variable theorem. The density of x can be shown to be:

$$f_x = \begin{cases} \frac{0.5}{\sqrt{x}} & 1 \leq x \leq 4 \\ 0 & \text{Otherwise} \end{cases}$$

This density is shown below.

Figure 6.7 Exact Density of x

To approximate this density with a linearized model, the nonlinear relationship of (6.1) is linearized by the following:

$$y_o = E(y) = 1.5$$

$$x_o = g^{-1}(y_o) = 2.25$$

$$y = y_o + \left. \frac{\partial g}{\partial x} \right|_{x_o} (x - x_o) \quad (6.2)$$

where the subscript "o" designates the point of linearization. Solving the linear equation (6.2) for x ,

$$x = x_o + \left[\left. \frac{\partial g}{\partial x} \right|_{x_o} \right]^{-1} (y - y_o) \quad (6.3)$$

$$x = 2.25 + 3(y - 1.5), \quad .75 \leq x \leq 3.75$$

The density of the linearized x is shown in the following figure.

Figure 6.8 Linearized Density of x ($y_o = 1.5$)

Now linearize equation (6.1) at the maximum value of y :

$$y_o = 2$$

$$x_o = 4$$

$$\left. \frac{\partial g}{\partial x} \right|_{x_o}^{-1} = 4$$

$$x = 4 + 4(y - 2), \quad 0 \leq x \leq 4$$

The density of the new linearized x is shown below.

Figure 6.9 Linearized Density of x ($y_o = 2.0$)

Now linearize equation (6.1) at the minimum value of y :

$$y_0 = 1$$

$$x_0 = 1$$

$$\left. \frac{\partial g}{\partial x} \right|_{x_0}^{-1} = 2$$

$$x = 1 + 2(y-1) \quad 1 \leq x \leq 3$$

The density of this linearized version of x is shown below.

Figure 6.10 Linearized Density of x ($y_0 = 1.0$)

Now if the three density functions of each linearization are combined appropriately, we get the following density curve:

The combination was made by equally sharing the three regions of linearization and excluding any values outside the domain of $1 \leq x \leq 4$. The composite curve was scaled by 12/13 to restore unity area.

Such a combination for multivariate problems is not in general such a simple task. Much development work is needed to determine key points of linearization and how to consistently combine the density components. This methodology would certainly require very skillful programming to be efficient.

Considering the other alternative of ignoring nonlinearity, it would be highly desirable to be able to determine an upper bound on the error of the density functions (or probabilities) to be computed. This research did not investigate how to compute such a bound, although the problem may be addressed by either limited simulation studies in the tail regions of the distributions, or by an analytical approach using the network power equations.

Figure 6.11 Composite Linear Density of x

CHAPTER VII

CONCLUSIONS AND RECOMMENDATIONS

This final chapter reviews the resulting contributions of this research and draws conclusions concerning their applicability to the stochastic power flow problem. As with most research endeavors, this effort has exposed many unexplored areas for future study, many of which are offered here.

7.1 General Results

This research has made contributions and developments to achieve the following:

- i. A nonstationary multivariate discrete time extension is made for the stationary Wold decomposition. This decomposition is fundamental to our approach of obtaining instantaneous probability information concerning events of interest in system operation. We point out that some form of decomposition such as this is necessary to remove nonstationarity and nonergodicity from the random processes to facilitate the estimation of ensemble averages with time averages.
- ii. A rigorous treatment of the stochastic power flow problem (direct transformation) is given in view of this decomposition. From the Implicit Function Theorem, the existence of the transformation is shown. As stated, the usefulness of this transformation depends wholly on the success of the decomposition in obtaining a stationary, ergodic stochastic part. Although these prospects (and more) are not unreasonable, it is impossible to make conclusive statements regarding the performance of such without investigating actual data.
- iii. A detailed development of the concept of "proportion of time" as a surrogate for instantaneous probability is given. The mechanics of estimating the distributions from time trajectories is shown. The basic contribution of this technique has been to show that the joint proportion of time distribution is totally singular, thus making it computationally practical to calculate joint proportions of time for high dimensional problems. The information sacrificed by using this transformation is the concept of chronological time. Probability statements can only be made concerning finite intervals of time, not specific instants.
- iv. Approaches and procedures are developed for estimating the expectation of the random proportion of time distribution. Proportion of time information is random since a distribution results from every independent realization of the process. Because the transformation is nonlinear, care must be taken to provide an unbiased estimator of the true distribution.
- v. The interval maximum transformation is introduced for discussion only. This method, which selects the maximum value of each process over a frame of time, is not considered meaningful because in general the occurrence of these maxima are not simultaneous.
- vi. A brief examination of the effects of nonlinearity in the network model are presented. These effects are examined by a comparison of Monte Carlo simulations employing both linear and exact (nonlinear) network models. Since the work is nonanalytical, no general statements can be made concerning nonlinearity, however, some insight into the general effects are drawn out.
- vii. An appendix containing the theoretical framework for most of the developments of this project is presented. This material contains theoretical background on properties of moments, density forms, stationarity, ergodicity, and correlation of stochastic processes.

7.2 Recommendations for Further Investigation

In the course of this work, several areas of interest were uncovered and not investigated in thorough detail. These topics are discussed below.

- i. What other decompositions could be developed as an alternative to the Wold decomposition? The goal here would be to create a representation which would be less dependent on a sophisticated demand model to track the mean function.
- ii. What is the mean square error forecast of the decomposition developed in Chapter 2? For 2nd order processes (not limited to 2nd order stationary processes) the MSE forecast has been developed for the single variate continuous time case, but not for the vector valued discrete time version. It can be shown that if the purely stochastic part of a process (not, in general, the whole of the process) is stationary, then the variance of this part is constant about $\mu(t)$ which is the mean plus predictable parts. From Box and Jenkins, it is known that the variance about the MSE forecast grows as one proceeds into the future. What is the relationship between these two variances?
- iii. Concerning the proportion of time distributions, how do the two procedures for obtaining a distribution (biased and unbiased) compare on a MSE basis?
- iv. Are there better ways of computing the expected value of the proportion of time distribution given by equation (4.4) without independent realizations or decomposition?
- v. What are the properties of the sampling distribution of the proportion of time distribution? Little is known about the convergence of this distribution concerning how it is affected by sample size, sample correlation, etc.
- vi. How do standard probability operations (expectation, variance, conditional events, etc.) apply to the proportion of time distribution and how are they interpreted?
- vii. How could the singular multivariate proportion of time distribution be programmed on a large scale

general purpose digital computer. Although the concept of this distribution is straightforward, the programming may not be. The development of a pilot program would be an important contribution.

7.3 Closing Remarks

This investigation has shown that there is a very fertile field of exploration that separates the load modeller from the stochastic power flow developer. It has been shown that careful attention must be paid in estimating probability models for demands and generations, and that the task is far removed from the trivial compilation of raw system data. This research has produced a few key concepts upon which a sound methodology can be founded. It appears conclusive that the necessary ensemble averages required for instantaneous probability statements are in no way accessible from raw data unless the troublesome properties of nonstationarity and non-ergodicity are carefully and systematically removed. Although the validation of the decomposition offered here can be computationally burdensome if stationarity and ergodicity are the goal, the technique appears to be feasible for

practical problems. However, even if one is successful in achieving a stochastic part that is ergodic and stationary, the problem of integrating in a multidimensional probability space of output variables is still a tremendous obstacle, even with the presence of normality.

This research has shown that such a barrier can be effectively eliminated with the concept of the proportion of time distribution as a substitute for instantaneous probabilities. The total singularity of this proportionality ("probability") distribution permits the integration to be in one dimension only, even for joint event calculations. Much work lies ahead in the investigation of this distribution concerning its convergence properties and the interpretation of its higher moments.

Indeed the modeling of stochastic behavior of nodal power injections (even in the quasi steady state) for stochastic power flow analysis is a complex problem. The time has arrived for the careful consideration of how to obtain and how to interpret an accurate collection of models suitable for stochastic power flow analysis.

APPENDIX A THEORETICAL DEVELOPMENT

This appendix compiles the required mathematical and statistical theory for examining the following process characteristics:

- a. stationarity,
- b. ergodicity,
- c. correlation (auto- and cross-),
- d. density shape (marginal and joint for a particular time frame).

Since the discussion of these items is ancillary to the development of stochastic power flow analyses, we have placed this material in an appendix. As the family of joint distributions (densities), to be defined in section A.1, figures intimately in the definition and characteristics of stochastic process article d will be covered first (section A.1). Various moments of a stochastic process including correlations will be defined in section A.2. Certain transformations of a stochastic process will also be considered. Since stationarity and ergodicity have certain relationships and sometimes misunderstandings, they will be studied jointly in section A.3. Also certain simplifications and characterizations that result for various moments when stationarity holds will be given in section A.3.

In the main, the discussion will involve a univariate real valued stochastic process which is by definition a collection of random variables $X(t)$, $t \in T$ where T is a specified set. In those cases where difficulties arise or confusion can result in considering vector and/or complex valued stochastic processes, more detail will be given.

A.1 DENSITY SHAPE

Consider the stochastic process $\{X(t), t \in T\}$. Then, for example if $t' \in T$, $X(t')$ is a random variable, that is, a measurable function on some space say, Ω . As a result, the notation $X(t', \omega)$ is also used. But $X(t', \omega)$ must be a random variable for every $t' \in T$ and if T is uncountable complications in considering various mathematical operations (limits, etc.) over or in T can arise. Also it is not obvious that a measurable space exists for which a probability measure can be given in which all of the random variables are defined (measurable). Recall that a measurable space is a pair $(X, \mathcal{B})$ where X is a space and $\mathcal{B}$ is a particular class of subsets of X . $\mathcal{B}$ is a σ -field (Borel field). This just means that $\mathcal{B}$ is closed under certain set operations. The elements of $\mathcal{B}$ are called events and probabilities are only given for events. If a suitable $(X, \mathcal{B})$ exists then, hopefully, a probability space $(X, \mathcal{B}, \Pi)$ can be formed and the probabilistic and statistical features of the process investigated (as opposed to that of the individual random variables).

Now $X(t)$ being a random variable for every $t \in T$ implies that for given $t_1, t_2, \dots, t_n$ the joint probability distribution of $X(t_j)$, $j=1, 2, \dots, n$ is prescribed. Thus $F(x_1, x_2, \dots, x_n) = \Pr(X(t_1) \leq x_1,$

$X(t_2) \leq x_2, \dots, X(t_n) \leq x_n)$ is known. However, this distribution might be obtained in other ways. Clearly, the event $X(t_1) \leq x_1, X(t_2) \leq x_2, \dots, X(t_n) \leq x_n$ is equivalent to the event $X(t_{j_1}) \leq x_{j_1}, X(t_{j_2}) \leq x_{j_2}, \dots, X(t_{j_n}) \leq x_{j_n}$ where $j_1, j_2, \dots, j_n$ is a permutation of $1, 2, \dots, n$. Thus we should have $F(x_1, x_2, \dots, x_n) = F(x_{j_1}, x_{j_2}, \dots, x_{j_n})$ and if not, we have an inconsistency. Also we might start from the time points $t_1, t_2, \dots, t_n, t_{n+1}$, obtain the joint distribution associated with these $n+1$ time points, and then obtain the marginal on the original n time points. If this gives a different distribution from that originally obtained, then the two prescriptions are inconsistent. If we rule out these two forms of inconsistency, then a celebrated theorem of Kolmogorov shows that a probability space $(X, \mathcal{B}, \Pi)$ can be constructed on which all the random variables $X(t)$, $t \in T$ may be defined. All probability statements concerning the $X(t)$ and events associated with them are given by Π which is defined on $\mathcal{B}$. The set of all distributions for every n and $t_1, t_2, \dots, t_n$ is called the family of (finite-) joint distributions of the process and when the two consistency requirements are met uniquely defines Π . For further discussion, see Billingsley (1968) or Cramér and Leadbetter (1967).

The space X can be thought of as the space of all "realizations" of $X(t)$, that is, a point in X is a function $x(t)$ that consists of each observed value of $X(t)$, $t \in T$. This motivates the notation $X(t, \omega)$. When ω is fixed and t varies, we obtain a particular realization. As ω varies, different realizations are obtained. In this sense, a unique realization is associated with each ω and in any application if we knew ω we would know the complete history for the given observation. This mathematical setup does not belie experience since in practice we only observe part of a realization and have no way to determine the exact ω point that gives the observation. X is called the sample space. Thus a stochastic process or random function is a measurable function from the underlying probability space $(\Omega, \mathcal{A}, P)$ to the measurable space $(X, \mathcal{B})$ and the Kolmogorov theorem shows how to construct $\mathcal{B}$ and gives a probability measure Π that is defined on $\mathcal{B}$ and is consistent with P .

The construction of $\mathcal{B}$ warrants some discussion. An open interval in X is the set of all points in X , that is, all finite, real-valued functions $x(t)$, which satisfy a finite set of inequalities of the form

$$a_j < x(t_j) < b_j, \quad j = 1, 2, \dots, n$$

where n is arbitrary, the t_j are any points in T and a_j, b_j are real constants (possibly infinite). Closed intervals are defined by replacing the $<$ by $\leq$. Half closed intervals, etc. are defined accordingly. These intervals are used to form the base of a topology in X in the usual way. In this topology, a sequence of points x_n in X converges to a limit x if for every

fixed $t \in T$, the real number sequence $x_n(t)$ tends as an ordinary limit to $x(t)$.

All finite sums (unions) of intervals from a field B_0 of sets in X . The smallest σ -field including B_0 , denoted by B , may be regarded as a generalization of the Borel sets in R^n (R denotes the real line). Thus the sets of B are sometimes called the Borel sets of X . A unique probability measure, Π , is defined on B_0 by the family of joint distributions of the process and the difficult part of the Kolmogorov theorem is in showing that Π can be extended (uniquely) to B . Indeed the theorem shows more in that the two consistency requirements are necessary and sufficient for the existence of a stochastic process with a given family of joint distributions (and, hence, a measure Π that is consistent with the joint family).

The Kolmogorov theorem allows the meaningful investigation of various analytical-probabilistic properties of the process in terms of distributions associated with the process. In practical applications, all that one can ever hope to have in terms of probabilistic information about a process is that contained in the distributions of $X(t)$ for various $t \in T$. We just do not observe Π . Indeed it is probably only infrequently that repeated observations (different ω) of a process can be made. Unfortunately, assumptions will generally have to be made in order to investigate the analytical properties of a process.

We shall pursue these matters by investigating various forms of limits associated with a stochastic process. Ultimately, we would like to make probabilistic statements about various limits associated with a process and this is where the difficulty lies. We shall say that two stochastic processes $X(t)$ and $Y(t)$ are equivalent if $X(t) = Y(t)$ with probability 1 (for either P measure or Π measure) for every $t \in T$. Notice that the null set, N_t , on which $X(t) \neq Y(t)$ can vary with t . Moreover, if $X(t) = Y(t)$ except on N^c for all $t \in T$ then we say that $X(t)$ and $Y(t)$ are a.s. (almost sure or almost surely) equal processes or random functions. Notice here that N does not vary with t (N^c denotes the complement of N). Now the N_t are in A for every t but $N = \bigcup_{t \in T} N_t$ and thus N may not be in A . By assumption in the definition of a.s. equality N is in A and thus a.s. equality is in general stronger than equivalence. Note that two equivalent processes are equivalent in a probabilistic sense since they have the same family of joint distributions. Evidently, two processes that are a.s. equal are also equivalent in a probabilistic sense.

Consider now an example from Loeve (1963). Let $T = [0, 1]$, take for the probability space the Lebesgue interval $[0, 1]$, and consider stochastic processes defined as follows: $X(t, \omega) = 0$ except at $\omega = \omega(t) = \omega_t$ where $X(t, \omega_t) = 1$. The ω_t for each $t \in T$ can be varied in any way from one process to another without altering their equivalence since the Lebesgue measure of a set with a single point is zero. Consider $Z = \sup_{t \in T} X(t, \omega)$, which takes at most two values 0 and 1, and set $A = \{\omega | Z=1\} = \{\omega_t | t \in T\}$. Now the ω_t can be selected so that A can be any set in Ω . For instance, $A = \emptyset$ and $Z \equiv 0$ for the process with every $\omega_t \in \emptyset$, $A = \Omega$ and $Z = 1$ for the process with $\omega_t = t$, and Z is not measurable for a process with ω_t ranging over a nonmeasurable set.

Furthermore, for the first example process, all

realizations are continuous and, hence, the process is a.s. sample continuous, that is, almost all realizations are continuous. For the second example process each realization has a jump discontinuity at $\omega_t = t$ and is otherwise continuous. Thus this process, is a.s. continuous, that is, $X(t') \xrightarrow{a.s.} X(t) (\equiv 0)$ as $t' \rightarrow t$ for every $t \in T$. Almost sure continuity and a.s. sample continuity can be described in other terms. When $X(t')$ does not converge a.s. to $X(t)$ as $t' \rightarrow t$, we say that t is a fixed discontinuity point of the process. A discontinuity point $t = t(\omega)$ which is not a fixed discontinuity point of the process is called a moving discontinuity point of the process. (Recall the second example). Thus a.s. continuity means no fixed discontinuity points and a.s. sample continuity means neither fixed nor (outside a null event) moving discontinuity points.

Unfortunately, as the example shows two equivalent processes can have very different analytical properties from a probabilistic viewpoint. Indeed various limits in T might not even be measurable. Thus in studying the analytic properties of stochastic processes defined on R or intervals of R , etc., we cannot consider the equivalence classes produced by equating equivalent processes. Only subclasses of the equivalence classes must, therefore, be considered and in such a way that, hopefully, limits in T will always be measurable. If a suitable equivalent process exists and has limits in T that are measurable, then there is hope of expressing various analytical properties in terms of the probabilities associated with the process. One choice is given by the "separable" processes, a concept due to Doob. For further discussion of these matters, see Doob (1953) or Loeve (1963). Also various analytical properties are investigated in Cramer and Leadbetter (1967).

The intent of these considerations is to show that not unreasonable restrictions on a family of joint distributions is all that is required to express uniquely the probability properties of equivalent stochastic processes. However, when considering the analytical properties of a process defined on, say, the real line knowledge of the family of joint distributions (even when consistent) does not suffice. In order to guarantee the measurability of the various analytical limiting operations in T further restrictions on the process must be made. One possibility is to assume that the process is separable. Separability involves the consideration of limits over a countable but infinite subset of T . Thus a process defined on the integers is automatically separable. However, this is not true in general and thus the assumption of separability is a restriction. Since it involves a countably infinite set, it cannot be directly tested in practice. The full implications of this are probably never known in practice since for a given stochastic process $X(t)$, $t \in T$, there exists a separable for closed sets stochastic process $\tilde{X}(t)$, $t \in T$, equivalent to $X(t)$ and defined on $X(t)$. (For a more complete discussion of this, see Doob or Loeve). There appears to be no way of easily determining whether an observed process is a separable version of a given process or not.

We mention another possible approach. Instead of considering separability we can consider for a given limiting operation in T , say, sample continuity, deriving sufficient conditions on a process that guarantee that there exists an equivalent version for which the realizations are continuous with probability one (or other probabilistic modes of convergence). This approach is followed in Cramer and Leadbetter (1967) Chapter 4. (See also their discussion in

Section 3.6). The theorems and approach are somewhat simpler than those in considering separability but different theorems are required for each limiting operation considered. Of practical importance is the fact that sufficient conditions can be expressed in terms of certain moments of the process and this can be more readily checked than the conditions of separability.

A.2 MOMENTS

As in the study of (individual) random variables, the study of moments associated with a stochastic process is important. Not only are various moments, especially the lower order moments, of interest in their own right they can be useful as indicated in Section 2 in determining characteristics of a particular process. This section will be primarily devoted to definitions of various moments of a stochastic process. All moments are defined in terms of the family of joint distributions and, hence, two equivalent processes have the same set of moments. (One should not assume that if two processes have the same set of moments for all orders and any $t \in T$ that the processes have the same family of joint distributions. It is well known that the sequence of moments for a univariate distribution do not in general uniquely determine the distribution.)

The moments of a stochastic process are a direct carryover from the random variable situation. For example, the mean function of a stochastic process is by definition

$$m(t) = E \{X(t)\}, \quad t \in T$$

where E denotes the expectation operator. Thus $m(t)$ exists as a real-valued function of t if and only if

$$E |X(t)| < \infty, \quad t \in T.$$

Note that $E X(t)$ always exists but may be infinite. Also recall that

$$E \{X(t)\} = \int_L X(t, \omega) dP(\omega) \quad (A.1)$$

and

$$E |X(t)| = \int_L |X(t, \omega)| dP(\omega), \quad (A.2)$$

so that expectations are defined in a measure theory context as integrals over Ω . In the case of real-valued random variables, it is well known that (3.1) and (3.2) can be expressed as (Lebesgue-) Stieltjes integrals. Indeed

$$E \{g(X(t))\} = \int_{-\infty}^{\infty} g(x) dF(x) \quad (A.3)$$

where $g(\cdot)$ is a finite, (Borel) measurable function on R and

$F(x)$ is the distribution function associated with P by

$$F(x) = P\{(-\infty, x]\}$$

where

$$(-\infty, x] = \{\omega | X(t, \omega) \leq x\}, \quad t \text{ fixed,}$$

in the sense that if either side of the expression exists, so does the other, and then they are equal. Expressions similar to (A.3) hold for sets of t -values so that the family of joint distributions is again important. In what follows, we shall interpret all integrals in the sense of (A.3).

The covariance function $R(t, s)$ of a process is by definition

$$R(t, s) = E \{[X(t) - m(t)][X(s) - m(s)]\}. \quad (A.4)$$

Note that in some works no assumption about the mean function is made and $R(t, s)$ is defined without a mean correction. For example, see Doob (1953) page 95. Of course, in the definition of a covariance function it is implicitly assumed that

$$E |X(t)|^2 < \infty. \quad (A.5)$$

In the complex case, the definition becomes

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

where the bar indicates the complex conjugation. When (A.5) holds the process is said to be of second order.

It is clear that $R(t, t) = \sigma^2(t)$ is the variance function of the process. In the distribution notation of (A.3), the covariance, (A.4), is

$$R(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 - m_1)(x_2 - m_2) dF(x_1, x_2), \quad (t_1 \neq t_2)$$

where

$$m_1 = \int_{-\infty}^{\infty} x_1 dF(x_1)$$

and

$$m_2 = \int_{-\infty}^{\infty} x_2 dF(x_2).$$

Note that

$$F(x_1) = F(x_1, \infty)$$

and

$$F(x_2) = F(\infty, x_2)$$

by the consistency conditions.

The correlation function associated with a covariance function is, say

$$r(t_1, t_2) = R(t_1, t_2) / \sqrt{\sigma^2(t_1)\sigma^2(t_2)} \quad (A.6)$$

Unfortunately, many authors do not make a distinction between (A.4) and (A.6) and call both correlation functions. Also notation different from the pair $R(t_1, t_2)$, $r(t_1, t_2)$ is used (sometimes without clear distinction between them).

When more than one process is considered certain cross-moments are frequently of interest. Consider two second order processes $X(t)$ and $Y(t)$, $t \in T$. Then,

$$R_{XY}(t, s) = E \{[X(t) - m_X(t)][Y(s) - m_Y(s)]\} \quad (A.7)$$

is the (cross-) covariance between $X(t)$ and $Y(t)$. If the processes are complex, then the second factor on the right-hand-side of (A.7) is replaced by its complex conjugate. In (A.7)

$$m_X(t) = \int_{-\infty}^{\infty} x dF_X(x)$$

and

$$m_Y(t) = \int_{-\infty}^{\infty} x dF_Y(x)$$

where both distributions are obtained at time point t . In the case of real-vector processes, say, consisting of $X(t)$ and $Y(t)$, we have the various moments

$$R_{XX}(t,s) = E\{[X(t) - m_X(t)][X(s) - m_X(s)]\}$$

$$R_{XY}(t,s) = E\{[X(t) - m_X(t)][Y(s) - m_Y(s)]\}$$

$$R_{YX}(t,s) = E\{[Y(t) - m_Y(t)][X(s) - m_X(s)]\}$$

and

$$R_{YY}(t,s) = E\{[Y(t) - m_Y(t)][Y(s) - m_Y(s)]\}$$

Note that in the complex case

$$R_{XY}(t,s) = \bar{R}_{YX}(s,t).$$

In vector notation we have

$$\underline{X}(t) = \begin{bmatrix} X(t) \\ Y(t) \end{bmatrix},$$

$$E\{\underline{X}(t)\} = \underline{m}(t) = \begin{bmatrix} m_X(t) \\ m_Y(t) \end{bmatrix}$$

and

$$\text{Cov}[\underline{X}(t), \underline{X}(s)] = E\{[\underline{X}(t) - \underline{m}(t)][\underline{X}(s) - \underline{m}(s)]'\}$$

$$= \Gamma(t,s) = \begin{bmatrix} R_{XX}(t,s) & R_{XY}(t,s) \\ R_{YX}(t,s) & R_{YY}(t,s) \end{bmatrix}$$

The "'" denotes the vector transpose and $\Gamma(t,s)$ is called a covariance matrix. (Again, some authors call $\Gamma(t,s)$ a covariance matrix without the various mean removals; see Hannan (1970) page 4.) A correlation matrix has the $R_{xx}(t,s)$, etc. replaced by $r_{xx}(t,s)$, etc. where

$$r_{XX}(t,s) = R_{XX}(t,s) / \sqrt{\sigma_X^2(t)\sigma_X^2(s)}$$

$$r_{XY}(t,s) = R_{XY}(t,s) / \sqrt{\sigma_X^2(t)\sigma_Y^2(s)}$$

$$r_{YX}(t,s) = R_{YX}(t,s) / \sqrt{\sigma_Y^2(t)\sigma_X^2(s)}$$

$$r_{YY}(t,s) = R_{YY}(t,s) / \sqrt{\sigma_Y^2(t)\sigma_Y^2(s)}$$

In the cases of a p -variate vector process where $p > 2$ an index notation is frequently used where X is replaced by, say, i , and Y is replaced by j , etc.

While moments higher than the second order are sometimes considered a great deal of study involves only the second order properties of a process. This is possibly so since the class of random variables, X , such that $E|X|^2 < \infty$ from a Hilbert space with inner product defined by

$$(X,Y) = E\{X\bar{Y}\}$$

and the more common operations that transform one process into another or other random variables are linear.

We shall leave the discussion of higher order moments to the literature and consider various relationships among moments brought about by transformations, specifically linear transformations. We shall consider only one example, that of integration.

The concept of integration of a stochastic process has been studied from two basically different viewpoints. Both give valid characterizations of integration and their use in a specific problem depends on the application.

On the one hand a stochastic integral has been defined as a random variable that is the limit in quadratic mean (q.m.) of a sequence of approximating sums in much the same sense as various forms of Riemann integration are defined. Thus the q.m. integral of a process is just an element in the Hilbert space associated with the process. (Recall that a q.m. limit is just a limit in mean square (m.s.).) Study of this type of integral is covered under the second order calculus of Loeve. See Loeve (1963) Chapter X or Cramer and Leadbetter (1967) Chapter 5 or Hannan (1970) Chapter I. Typical of what can be shown for this type of integration is the following: Consider

$$\int_a^b g(t)X(t)dt = \text{l.i.m.} \sum_{j=1}^m g(t'_j)X(t'_j)(t_{j+1}-t_j)$$

where

$$t'_j \in (t_j, t_{j+1})$$

and

$$a = t_1 < t_2 < \dots < t_{m+1} = b$$

is a partition of the interval $I = [a,b]$. The limit in mean square (l.i.m.) indicated is taken over all partitions such that

$$m \rightarrow \infty$$

and

$$\max(t_{j+1} - t_j) \rightarrow 0$$

and the integral is defined only when this l.i.m. exists.

Now $g(t)$ such that the Riemann integral

$$\int_a^b \int_a^b g(t)g(s)R(t,s)dtds$$

exists is necessary and sufficient that

$$\int_a^b g(t)X(t)dt$$

exists and when this integral exists then

$$E \int_a^b g(t)X(t)dt = \int_a^b g(t)E\{X(t)\}dt$$

and

$$E \left\{ \int_a^b g(t)X(t)dt \int_a^b g(s)X(s)ds \right\} = \int_a^b \int_a^b g(t)g(s)R(t,s)dtds.$$

On the other hand, one could consider a stochastic integral from the sample realization viewpoint. From this point of view, the previous definition suffers from two defects. First the limit is in the m.s. sense and this does not imply a.s. convergence. However, this might not be too serious since the distributional properties of a stochastic process are only given by the family of joint distributions. A more serious objection is that the integral

is defined only in a random variable sense. Often in practice one contemplates the integration of individual realizations and it is not clear how the previous definition relates to this. To be specific, suppose $g(t) = 1$, then the integral under consideration is

$$\int_a^b X(t)dt = \int_a^b X(t, \omega)dt$$

where by the notation, we mean that the integral is considered for each realization and is understood to be a Lebesgue integral.

In order to pursue this further, we need to consider another definition of a stochastic process. So far we have studied two definitions, namely,

- o The random variable definition: A stochastic process is a function of T to the space R of random variables on the probability space (Ω, A, P) .
- o The sample (realization) definition: A stochastic process is a measurable function on the probability space (Ω, A, P) to a sample space (X, B) .

These two definitions are equivalent as is readily shown. One emphasizes the random variable nature of a stochastic process while the other emphasizes the realization or sample nature of a process. We follow Doob (1953) Chapter II in the following definition. (Loève (1963) Chapter XI gives a somewhat more general definition and then specializes to the Doob case.)

- o Measurable stochastic process definition: A stochastic process is measurable if the parameter set T is Lebesgue measurable and if $X(t, \omega)$ defines a function to (X, B) measurable in the pair of variables (t, ω) .

This definition clearly includes the previous two but places a restriction on the stochastic process since neither of the original definitions guarantees joint measurability in the pair (t, ω) . The purpose of the definition is to place the study of stochastic processes squarely into the field of measure theory. This in conjunction with separability has interesting implications.

We now state three theorems due to Doob. Loève also gives these theorems, but with slight variations from the Doob versions.

Theorem. Let $X(t)$, $t \in T$ be a separable process with a Lebesgue measurable parameter set T . Suppose that there is a t set T_1 of Lebesgue measure 0 such that,

$$P\{\lim_{s \rightarrow t} X(s, \omega) = X(t, \omega)\} = 1, \quad t \in T - T_1.$$

Then the process $X(t)$ is measurable.

The probability requirement in the theorem is somewhat weaker than the requirement of a.s. continuity since it does not have to hold for every $t \in T$. Also note that the continuity requirement is only in a random variable sense as opposed to, say, a.s. sample continuity.

Theorem. Let $X(t)$, $t \in T$ be a separable process with a Lebesgue measurable parameter set T . Suppose that there is a t set T_1 of Lebesgue measure 0 such that

$$p \lim_{s \rightarrow t} X(s) = X(t), \quad t \in T - T_1.$$

Then there is a process $\tilde{X}(t)$, $t \in T$, defined on the same ω space, which is separable relative to the closed sets, measurable, and equivalent to $X(t)$.

Comments similar to those following the first theorem also hold here with the exception that the continuity requirement is in terms of continuity in probability as opposed to a.s. continuity. The importance of these theorems is that they give conditions under which a process is either measurable or has an equivalent measurable version. The importance of measurability is given by the third theorem:

Theorem. Let $X(t)$, $t \in T$ be a measurable stochastic process. Then almost all sample functions of the process are Lebesgue measurable functions of t . If $E X(t, \omega)$ exists for $t \in T$, it defines a Lebesgue measurable function of t . If A is a Lebesgue measurable parameter set and if

$$\int_A E|X(t, \omega)| dt < \infty,$$

then almost all sample functions are Lebesgue integrable over A . It follows directly from the theorem that

$$E \int_A X(t, \omega) dt = \int_A E X(t, \omega) dt.$$

If $A = [a, b]$, then these theorems give sufficient conditions for the existence of the integral

$$\int_a^b X(t) dt$$

as a Lebesgue integral. Now if

$$\int_a^b \int_a^b R(t, s) dt ds$$

exists as a Riemann integral then

$$\int_a^b X(t) dt$$

exists as a q.m. integral. (We tacitly assume that the process is of second order.) Thus we have two definitions of an integral of the process. It is of interest to know under what conditions they are equal. If it is further assumed that $R(t, s)$ is continuous then Loève (1963) Chapter XI has shown that (among other things) the two integrals coincide.

Irrespective of which form of integration (q.m. or a.s. sample) is considered it turns out that the expectation of the integral equals the integral of the expectation and this gives a transformation rule for the expectation operator and integration. Analogous results hold for certain other linear transformations such as differentiation and we refer the reader to Cramér and Leadbetter (1967), Doob (1953), Hannan (1970), or Loève (1963) for the details.

A.3 STATIONARITY AND ERGODICITY

In this section we shall investigate the concepts of stationarity and ergodicity as they are employed in the study of stochastic processes. Also certain simplifications and characterizations that result for various moments when stationarity holds will be pre-

sented. The origins of these concepts are from physics, especially statistical mechanics and from the study of Markov processes (which itself has benefited from physical considerations).

We first consider stationarity and ergodicity in the context of a finite Markov chain with discrete parameter (time). A Markov chain is characterized by a set of states and a probability structure that governs the transition among the states. If the states are designated by $1, 2, \dots, N$ then one way to do this is to specify a matrix

$$P_t = [p_{ij}(t)] \quad , t=1, 2, \dots$$

of transition probabilities with row sums equal to one that give the (conditional) probability of a (one-step) transition from state i to state j at time point t and an initial distribution vector

$$P_0 = (p_i)$$

that gives the probability of initially being in state i , $i = 1, 2, \dots, N$ with $\sum p_i = 1$. For this setup it is straightforward to generate a family of joint distributions and, hence, from Kolmogorov's theorem there exists a process with the given family of joint distributions. Thus, formally the initial vector and transition matrix define a process. Unfortunately, as we have seen Kolmogorov's theorem gives no direct information about the realizations of a process. Provided we deal with events that are defined on a finite number of trials then the generality of Kolmogorov's theorem can be avoided. This approach is taken in Feller Vol. 1 (1968) and even works for a countable state space. However, when we go to general state spaces (Markov processes) and/or continuous time, then we must employ Kolmogorov's theorem and the study becomes very difficult.

When

$$P_t = P$$

the process is said to be temporarily homogeneous or stationary or to have stationary transition probabilities. The latter two are more commonly employed. If we think of the process as describing a physical particle, then stationarity implies that the particle has no memory with respect to time translations; it only remembers its most recent position and this for only the immediate transition. Most work has been carried out for stationary Markov chains. Doob (1953) page 175 has shown that for a finite stationary Markov chain there exists a so called stationary (absolute) distribution vector Π where

$$\Pi = (\pi_i) \quad , \quad \sum \pi_i = 1$$

and is such that

$$\Pi' P = \Pi$$

If this Π is the initial vector, then we see that the probability of being in any state is invariant with respect to time translations. Indeed the joint family produced by this initial vector Π and transition matrix P also has this property. This property can hold for other processes as well. In fact, if a process has the property that the distribution of $x_{t_1+h}, x_{t_2+h}, \dots, x_{t_n+h}$ is independent of h for every t_i , $i=1, 2, \dots, n$, and n such that $t_i+h \in T$, $i=1, 2, \dots, n$ then the process is said to be strictly stationary. When no confusion can arise the modifier strictly is frequently omitted.

Strictly stationary processes have interesting properties. Consider again a finite stationary Markov chain and let $f_j^{(n)}$ be the probability that the first return to state j occurs at time n . Since for each j and n this "event" can be defined by a finite number of trials the event is well defined as a random variable. Set

$$f_j = \sum_{n=1}^{\infty} f_j^{(n)}$$

which is defined as a limit. We emphasize that the f_j may have no meaning as probabilities of an event since we are not considering infinite sample spaces or realizations of the process. Irrespective of this a realization must either return to j or not thus

$$f_j \leq 1.$$

If $f_j = 1$ the state is persistent and if $f_j < 1$, the state is transient (Feller Vol. 1 (1968)). The respective terminology in Doob (1953) is non-transient and transient. Unfortunately, other authors use even different terminology; so beware.

While these definitions do not directly involve the consideration of infinite sample spaces, there is an implicit consideration. Even this can be essentially avoided for chains with discrete time as shown by Doob (1953). Call j a consequent of i if there exists an n such that $p_{ij}^{(n)}$ (the probability of going from i to j in n finite transitions) is greater than zero. This terminology follows Doob. Karlin (1966) and Parzen (1967) call j accessible from i while Feller Vol. 1 (1968) says j can be reached from i . The states of the chain can be classified as transient if the state has a (at least one) consequent of which it is not itself a consequent or as non-transient if the state is a consequent of every one of its consequents. These are the only possibilities. The non-transient states can be further divided into ergodic classes by putting two states in the same class if they are consequents of each other. These are called closed communicating classes by Parzen. A communicating class (a set of states such that for each pair of states i and j , j is a consequent of i and i is a consequent of j) can have transient states. Also a non-empty class is closed if no state outside the class is a consequent of any state in the class.

With these considerations the transition matrix of a chain can be put into the following form:

$$\begin{array}{c} E_1 \quad E_2 \quad \dots \quad E_k \quad F \\ \begin{array}{c} E_1 \\ E_2 \\ \vdots \\ E_k \\ F \end{array} \left[\begin{array}{ccccc} P_{11} & & & & \\ & P_{22} & & & \\ & & \ddots & & \\ & & & P_{kk} & \\ P_{k+1,1} & P_{k+1,2} & \dots & P_{k+1,k+1} & \end{array} \right] \end{array}$$

The E_i , $i = 1, 2, \dots, k$ represent the ergodic classes and F represents the class of transient states. Only elements in a submatrix P_{kk} are positive. The ele-

ments of class F have at least one member of an ergodic class which is a consequent of the element in F. It can be further shown (check the given references Doob, Feller, Karlin, or Parzen) that an ergodic class can be divided into cyclical subclasses in which a transition is always made from one subclass into the next, etc. The period of the phenomenon is called the period of the class and the period of the individual states (which is the same as the class period). If the period is one, the class or state is called aperiodic. With this taken into consideration we see that once in an ergodic class the chain remains there (with probability one) and can make transitions to any member of the class subject to a possible periodicity. Also a finite chain cannot stay in the transient states indefinitely (with nonzero probability). This phenomenon is consistent with $f_j = 1$ or $f_j < 1$ and in fact can be shown to be equivalent. Thus the Doob and Feller approaches are basically not different.

Also an ergodic class has a sort of permanence in that ultimately transitions to every state in the class must occur. This notion is in keeping with the general concept of "the ergodic hypothesis". If the states (in an ergodic class) have physical meaning attached to the designations 1, 2, ..., N then one might be interested in the "average value" of the states. This is the mean value or expectation and is constant since the process is strictly stationary. The question arises: Can this mean be estimated in some sense from a single realization? In the context of general stochastic processes (as opposed to Markov chains) an ergodic hypothesis is the statement that a certain time (realization) average equals a certain expectation (ensemble average). This is related to an ergodic class (in a Markov chain) as we will eventually show. In terms of the Markov chain example the statement would be:

$$\sum_{i=1}^N i \pi_i = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n X_j$$

where $\lim$ remains to be defined and X_j denotes the j th random variable in the discrete time Markov chain. The intent of the statement is intuitively clear: Since the chain can leave an ergodic class only with probability zero it is not unreasonable to suppose that the average sojourn in any state is proportional to the probability of being in that state.

Before we attack the general situation, we shall consider Feller's definition of an ergodic state (and, hence, class). To do this we need one more definition. Since in the case of a persistent state the $f^{(n)}$ define a probability distribution one can consider the mean of this distribution, that is, the mean recurrence time of the persistent state. If this mean is finite (exists), then the state is said to be non-null; otherwise null. Feller Vol. I (1968) calls a persistent, aperiodic, non-null state ergodic. This is clearly more restrictive than the Doob definition. The usefulness of the Feller definition stems from the fact that in an irreducible (a chain with only one closed set of states) aperiodic Markov chain (denumerable state space and not just finite) if one state is ergodic then all are, and a unique stationary distribution exists which is obtained from

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \pi_j \text{ (ordinary limit).}$$

Thus in this case the results about the stationary distribution hold for any discrete parameter chain and they are stronger than the Doob results for a finite chain. However, the Doob definition of ergodicity for

Markov chains is closer to the mathematical-probabilistic definition. But in Feller's considerations there is a unique limiting or steady-state distribution which is approached asymptotically and is independent of the initial distribution. There are relationships between the two concepts of ergodicity which we shall not pursue.

However, along these lines the following example is interesting. Let

$$P = \begin{bmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \end{bmatrix}$$

The chain is clearly irreducible and periodic with transitions from {1,2} to {3,4} and {3,4} to {1,2} and, hence has period 2. It is straightforward to show that

$$P = [p_{ij}^{(3)}] = [p_{ij}^{(5)}] = [p_{ij}^{(7)}] = \dots$$

and

$$P^2 = [p_{ij}^{(2)}] = [p_{ij}^{(4)}] = [p_{ij}^{(6)}] = \dots$$

and that there is a unique stationary distribution given by $\pi_1 = \pi_2 = \pi_3 = \pi_4 = 1/4$ but this distribution is not a limiting distribution since

$$p_{ij}^{(n)} \nrightarrow \text{a limit}$$

as

$$P^2 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

We have seen that the definition of stationarity in a Markov chain is equivalent to the definition of strict stationarity for a general stochastic process. We shall now investigate the relationship of an ergodic class (Doob's definition) to the definition of ergodicity in a more general setting. Ergodicity and ergodic theorems have a background in statistical mechanics, as first proposed by Gibbs, which deals with the study of mechanical systems with many degrees of freedom. The system is described by a set of generalized coordinates, q , and momenta, p , and set of equations of motion. The states of the system are represented by points $w = (q, p)$ in a $2N$ -dimensional space - the phase space of the system where N is the degrees of freedom (of the system). As time goes on the state of the system changes in accordance with the differential equations of motion thereby tracing a curve or trajectory in the phase space. Instead of attempting to study an individual trajectory statistical mechanics attempts to investigate, say, the probability of being in a specified subset of the phase space at some point in time. Thus the interest is on ensembles of states and probabilistic behavior.

Now the motion in the phase space can be represented as a group called a flow as follows. If the influence of the equations of motion is denoted by T and initially the system is in state X_0 , then at time

t the system is given by $X_t = T_t X_0$. T_t gives a one parameter group of transformations since $T_{s+t} = T_s T_t$. Liouville's theorem states that if the coordinates used in the description of the phase space are appropriately chosen, then the flow in the phase space leaves all volumes invariant. As a result, the total energy of a conservative system does not change as the flow progresses and, hence, the term ergodic.

It follows that a general problem of statistical mechanics is the study of families of measure-preserving transformations and specifically asymptotic properties of these families. Asymptotic properties are of interest since it is assumed that the time to take measurements is long compared to the microscopic changes in the system and the measure-preserving restriction results from Liouville's theorem. From a theoretical viewpoint the problem is still not solved since the time-averages must be evaluated. To circumvent this difficulty the hypothesis is made that time-averages equal phase-averages (ensemble averages); this is the ergodic hypothesis. From a mathematical-probabilistic point of view the hypothesis is false as a general statement. Thus restrictions must be imposed and the point of attack changed somewhat. Ergodic theorems in "probability" theory assert conditions under which sequences of time-averages converge in some sense and then conditions under which the limits are constant (recall the Feller approach.) In the "mathematical" approach similar kinds of theorems are proven but the interest is on measure-preserving transformations. One possible connection between the two approaches is given in Doob (1953) Chapter X.

Paralleling the concepts, the ergodic hypothesis and ergodic theorems in the context of ergodic theory, is a third concept-ergodicity. In the strict technical sense in either the probabilistic or mathematical approach to ergodic theory ergodicity has a very narrow and precise meaning. In the probabilistic theory the fundamental transformation is that of translation. This is usually coupled with the assumption of strict stationarity which implies that translations are measure-preserving transformations, that is, $m(S) = m(T^{-1} S)$ where S is a measurable set, T denotes the transformation, T^{-1} denotes an inverse operation, and m is the measure of interest. In the mathematical approach measure-preserving transformations are considered from the outset. Given a measure-preserving transformation T a set S is said to be invariant under T if and only if $T^{-1} S = S$ a.e. (with probability one in the stochastic process context). Note that T^{-1} is considered in an operator sense and this does not imply that T is invertible; T may not be one-to-one. A transformation is said to be ergodic or metrically transitive if and only if the only invariant sets are the trivial ones, that is, if and only if $m(S)=0$ or $m(X-S)=0$ whenever S is a measurable set invariant under T and X denotes the space of interest. In the probabilistic context with T denoting translation (T has an inverse here) this is stated as S has probability zero or one (again S is invariant). Also in this situation the process is said to be ergodic or metrically transitive.

The implications of these considerations for a finite Markov chain are interesting. Consider a finite chain and let the initial distribution be a stationary distribution such that $\pi_i > 0$ if i is in an ergodic class. This can always be done. This defines a strictly stationary process in which all ergodic classes will be visited with non-zero probability.

Let E be an ergodic class and set

$$S = \{ X_0(\omega) \in E \}$$

Since $X_n(\omega) \in E$ implies that $X_{n+1}(\omega) \in E$ and conversely, neglecting ω sets of probability zero, it follows that S is an invariant set. Now

$$\Pr \{S\} = \sum_{i \in E} \pi_i > 0.$$

If there is another ergodic class, this procedure defines a second invariant set with positive probability and which has no points in common with S . Thus, $0 < \Pr\{S \cap F\}$ and, therefore, if there is more than one ergodic class the process is not metrically transitive or ergodic. Conversely Doob (1953), page 460, has shown that if there is only one ergodic class, then the process is metrically transitive or ergodic. Thus a finite Markov chain with discrete time is ergodic if and only if it has a single ergodic class. In this context the Feller approach would rule out periodic chains and would yield a stationary distribution that is unique and is also a limiting distribution.

In a more general setting the story is not quite as crisp but is broader. Actually a process need not be strictly stationary to satisfy an ergodic theorem. While we have not yet discussed it there are two types of ergodic theorem. These theorems can be thought of as analogues of theorems regarding laws of large numbers in the non-stochastic process setting. Thus it is possible to consider more than one type of convergence. In ergodic theory two forms of convergence, namely, a.e. or a.s. and in m.s., have been considered. The first (a.s.) results in the "individual ergodic theorem" due to Birkhoff and the second (m.s.) results in the "statistical ergodic theorem" due to von Neumann. Even if one considers the strong form of convergence (a.s.) then a process need not be strictly stationary to satisfy an ergodic theorem. (We shall discuss the m.s. case in more detail presently.)

We now consider the individual ergodic theorem and state three theorems for discussion purposes.

Theorem (Halmos (1956)). If T is a measure-preserving (but not necessarily invertible) transformation on a space X (with possibly infinite measure) and if $f \in L_1$, then $\frac{1}{n} \sum_{j=0}^{n-1} f(T^j x)$ converges almost everywhere. The limit function f^* is integrable and invariant (that is, $f^*(Tx) = f^*(x)$ a.e.). If $m(X) < \infty$, then $\int f^*(x) dx = \int f(x) dx$.

Stationarity is not involved in the statement of this theorem and is not (directly) required. However, T must be measure-preserving. As mentioned, Doob (1953) Chapter X, makes a connection between the mathematical ergodic theorem and applications in stochastic processes. This is accomplished by considering (strictly) stationary processes. In any event some connection to applications in stochastic processes must be made so that the statement about stationarity and ergodic theorems is not enhanced by the theorem. It is presented to indicate the full generality of an ergodic theorem.

Theorem (Loeve (1963)). If

$$(i) \sum_{k=1}^n \int \frac{X_k}{Y^n} \rightarrow 0 \text{ and}$$

$$\sum_{k=1}^n \int \frac{Y_k}{A_k Y^n} \rightarrow 0 \text{ as } n \rightarrow \infty \text{ and then } A_k \neq \emptyset,$$

$$(ii) \int \frac{X_{n+k}}{Y^n} \rightarrow 0 \text{ and } \int \frac{Y_{n+k}}{Y^n} \rightarrow 0, \text{ for every fixed } k \text{ as } n \rightarrow \infty,$$

then the sequence

$$\frac{X^n}{Y^n}$$

converges a.s.. Assumption (i) implies that from some n on the sum therein exists.

In the theorem the A_k are defined as events on the X_k 's and Y_k 's and

$$X^n = \sum_{k=1}^n X_k, Y^n = \sum_{k=1}^n Y_k.$$

Also it is assumed that

$$Y_k > 0, Y_k \neq \infty$$

with probability one and that the X_k and Y_k are such that the indicated integrals exist. Thus the Loève approach only considers sequences of random variables as processes and takes translation as the transformation of interest at the outset. To see this let $Y_k = 1$ for every k . Moreover, it is easy to construct examples of processes that meet the conditions of the theorem and are not strictly stationary. However, it is not clear what additional, general assumption(s) (outside of independence of the Y_k) is required to guarantee that the a.s. limit of X^n/Y^n is a constant. In the case of stationarity, translation is measure preserving and the condition is that the process be ergodic (metrically transitive). From an applications viewpoint the theorem is too general. However, it does make the point that an ergodic theorem need not be tied to an assumption of stationarity.

Theorem (Doob (1953)). Let $X_n, n \geq 0$ be a strictly stationary process, with $E\{X_0\} < \infty$, and let F be the σ -field of invariant sets. Then

$$\lim_{n \rightarrow \infty} \frac{X_0 + \dots + X_n}{n+1} = E\{X_0|F\}$$

with probability one. In particular, if the process is metrically transitive, the right-hand of the limit expression can be replaced by $E\{X_0\}$.

This is the more typical statement of the individual ergodic theorem in probability theory or the study of stochastic processes. The limit expectation is a conditional expectation and, hence, may not be constant. It is straightforward to show that the class of invariant events is a σ -field and thus that this expectation is well defined. With the additional assumption that the process is metrically transitive (ergodic) the expectation becomes a constant. This is the real importance of the ergodic condition. Unfortunately, outside of special cases such as finite

Markov chains, it appears that determining if a stationary process is also metrically transitive is very difficult. Thus in many applications the assumption is made that the process is metrically transitive or ergodic.

In the case of strict stationarity it is possible to consider ergodicity from a somewhat more general viewpoint as is shown by Rozanov (1967) Chapter IV. This viewpoint indicates the restrictive nature of the ergodic concept. We restrict ourselves to the discrete time case and present results analogous to the Doob theorem. (For the continuous time case see Rozanov.) Let η be any random variable measurable with respect to B (the σ -field of the process) and having finite expectation $E|\eta| < \infty$. Set $\eta_n = T_n \eta$ with $\eta_0 = \eta$, then

$$\lim_{n \rightarrow \infty} \frac{\eta_0 + \eta_1 + \dots + \eta_n}{n+1} = E\{\eta|F\} \quad (A.8)$$

with probability one. (This includes the Doob result.) If the process is ergodic (metrically transitive), then $E\{\eta|F\} = E\{\eta\}$ a constant. Conversely Rozanov shows that if the limit in (A.8) equals $E\{\eta\}$ for any random variable (measurable with respect to B) with (finite) expectation then the process is ergodic. Clearly this includes the convergence of all sample moments of a process to the population moments (assuming that they exist) and is indeed a much stronger condition. Even in the presence of strict stationarity (which is itself a strong requirement), the condition of ergodicity is quite strong and is clearly a severe restriction on the process.

Another possible approach to this difficulty is to consider conditions which are, hopefully, more readily verified and at the same time imply ergodicity. Two such conditions are the mixing condition and the uniform mixing condition. If

$$\lim_{n \rightarrow \infty} P\{A \cap T^{-n}B\} = P(A)P(B) \text{ for } A, B \in \sigma\text{-field of the underlying probability space}$$

the process is said to be mixing. That mixing implies ergodicity is seen by taking $A = B =$ an invariant set so that we obtain $P(A)P(B) = P^2(A) = P(A)$ and $P(A) = 0$ or 1 . Speaking loosely the condition says that events determined by sets of X_n far apart in time are near to independent. Unfortunately, this condition is difficult to check directly when one considers all possible A and B . Now consider uniform mixing. Suppose there exists a positive function g defined on the integers such that $g(n) \rightarrow 0$ as $n \rightarrow \infty$ and that

$$|P(A \cap B) - P(A)P(B)| < g(q-p), A \in B_p, B \in F_q, q > p$$

where B_p is the σ -field of events determined by $X_n, n \leq p$ and F_q is the σ -field of events determined by $X_n, n > q$. Then X_n is said to satisfy a uniform mixing condition. We imply that events separated in time by $(q-p)$ approach independence uniformly in $A \in B_p$ and $B \in F_q$ as $(q-p) \rightarrow \infty$. This condition obviously implies the mixing condition and hence the ergodicity of the process. The uniform mixing condition is rather strong but has hope of being checked if it is possible to obtain more than a single realization and the A and B are suitably chosen. In any event, it may be more pleasing to assume some form of asymptotic independence holds than to assume that ergodicity holds.

We now consider the m.s. concept of ergodicity. A second order process in which the covariance is a function of only time differences is said to be second order stationary. In our earlier notation this means that

$$R(t,s) = R(t-s).$$

We note that this form of the definition is not always taken and sometimes we have

$$R(t,s) = R(s-t).$$

This latter approach is followed by Hannan (1970) page 18 while the former is followed by Doob (1953) page 95 and Loeve (1963) page 482. In the case of a real process the two definitions coincide but for a complex process the complex conjugate makes a difference.

Also of further interest is the fact that various authors treat the process mean differently in the definition of second order stationarity. The most common definition is to assume that $m(t)$ is constant. Thus a second order process is said to be second order stationary if

- o Its mean function is constant with time:

$$E\{X(t)\} = m(t) = c$$

- o Its covariance (autocovariance) function is a function of only time differences:

$$R(t,s) = R(t-s).$$

However, Doob (1953) page 95 explicitly makes no assumption about the mean function and calls such processes wide sense stationary. Furthermore, Loeve (1963) page 465 shows that the assumption of a constant mean is irrelevant and thus without loss of generality assumes that the mean function is zero. We adopt the stated definitions in this discussion. This is consistent with definitions of higher order stationarity as given by Bendat and Piersol (1971) in terms of moments and Papoulis (1965) in terms of the finite joint distributions. We point out that higher order stationarity for all moments does not imply strictly stationarity as is sometimes stated. Note that if a process is strictly stationary and is also of second order then it is automatically second order stationary (and that the converse is not true). In the case of real normal process (each member of the family of joint distributions has a multivariate normal distribution) second order stationarity (constant mean version) and strict stationarity coincide.

With the assumption of second order stationarity, it is well known that the process has a decomposition in terms of sines and cosines (harmonic decomposition). This decomposition is best shown in the decomposition of the autocovariance function as

$$R(t-s) = \int d^{i2\pi} (t-s)\lambda dF(\lambda) \quad (A.9)$$

where the integral is a Stieltjes integral and the limits of integration are $(-\infty, \infty)$ if t is real, $(-1/2, 1/2)$ if t takes on only integral values and λ is frequency in cycles per time unit (the frequency domain). F is called the spectral distribution function and the set of frequencies the spectrum. Thus $R(t-s)$ is what is known as a statistical characteristic function, and therefore all of the properties of characteristic functions apply; namely, there is a one-to-one relationship between $F(\cdot)$ and $R(\cdot)$ and certain operations in the time domain have simple counterparts in the frequency domain. With further assumptions, the relationship expressed in (4.2) can be given as a Fourier transform, and in this case the

methods and theory of Fourier analysis can be employed. Fourier methods have long been applied in time series problems by engineers (electrical and communications, in particular) and applied scientists (especially in the earth sciences).

As a result of (A.9) it can also be shown that there exists a stochastic process $Z(\lambda)$ such that

$$X(t) = \int e^{i2\pi t\lambda} dZ(\lambda)$$

where the integral is defined in a m.s. sense. In case time is continuous, it must be further assumed that $X(t)$ is mean square continuous. For these results see Hannan (1970) Chapter II or Loeve (1963) pages 482-485.

We now restrict ourselves to the discrete time case and consider a second order ergodic theorem following Hannan (1970) Chapter IV, Section 3.

Theorem. Let X_n be second order stationary then

$$\lim_{n \rightarrow \infty} \frac{X_1 + X_2 + \dots + X_n}{n} = Z(0) - Z(0^-).$$

Thus we see that if X_n is second order stationary with mean m , then the sample mean converges in m.s. to m if and only if there is no jump in the spectrum at the origin. The major difference between this, the statistical ergodic theorem, and the individual ergodic theorem is the mode of convergence. However, we also note that the process need not be strictly stationary, only second order stationary for the theorem to hold. The counterpart of ergodicity in this setup is the assumption of no jump in the spectrum at the origin.

There are interesting implications when the process is normal with zero mean since then second order stationarity is equivalent to strict stationarity. In this situation Rozanov (1967) pages 163-166 has shown that a necessary and sufficient condition for the process to be ergodic is that the spectral distribution function be continuous. Now consider a process given by

$$X(t) = A \sin 2\pi ft + B \cos 2\pi ft$$

where A and B are independent normal random variables with zero means and equal variances,

$$0 < f \leq 1/2, \text{ and}$$

$$t \in (-\infty, \infty).$$

It is straightforward to show that $X(t)$ is a normal, second order stationary process with a zero mean and a spectrum concentrated at f . Thus the process is not ergodic. Yet it is clear that every sample mean converges to zero and, hence, the convergence is with probability one. The process satisfies a strong ergodic theorem but is not ergodic.

As a result of the difficulties associated with ergodicity and practical considerations some authors have chosen to consider other definitions of ergodicity. In some of these works a clear distinction is not made between the concepts: (the) ergodic hypothesis, ergodic theorem, and ergodicity. Indeed a common definition of ergodicity is:

$X(t)$ is ergodic if time averages equal ensemble averages.

See, for example, Bendat and Piersol (1971) Chapter 1,

Papoulis (1965) Chapter 9, and Parzen (1967) Chapter 3. In Bendat and Piersol, while it is not stated, the form of convergence (of the time averages) is a.s. and in Papoulis and Parzen it is in m.s. Papoulis and Parzen treat conditions under which the variance of sample averages approach zero. Under suitable restrictions which are met in their cases this is equivalent to convergence in m.s. Parzen's results point out the fact that a process need not even be first order stationary to satisfy a first order ergodic theorem. The mean of a process can change with time and yet the sample mean can approach a constant (in m.s.). Indeed one can construct examples based on independent random sequences in which sample averages

approach a constant with probability one by employing Kolmogorov's proposition (see Loève (1963) page 238). Thus even for first order ergodicity there is no "nice" relationship with first order stationarity. The theorems of Papoulis and Parzen concerning ergodicity can be thought of as analogues of the kind of theorems given by Cramér and Leadbetter (1967) concerning properties of sample functions. In both Papoulis and Parzen the theorems involve population moments and there is some hope in practice of verifying properties of moments. In this context Hannan (1970) Chapter IV, Section 3 also gives various theorems concerning convergence of second order sample moments in terms of population moments, spectra, etc.

APPENDIX B
FOURTEEN BUS POWER SYSTEM DATA

The power system bus and line data for the simulation studies of Chapter VI are provided in the following tables;

Table B.1 Network Transmission Data

<u>Sending Bus</u>	<u>Receiving Bus</u>	<u>% Resistance</u>	<u>% Reactance</u>	<u>% Line Charging</u>	<u>% Tap Ratio</u>
1	2	1.9380	5.9170	2.6400	
1	5	5.4030	22.3040	2.4600	
2	3	4.6990	19.7970	2.1900	
2	4	5.8110	17.6320	1.8700	
2	5	5.6950	17.3880	1.7000	
3	4	6.7010	17.1030	1.7300	
4	5	1.3350	4.2110	.6400	
4	7	0.0000	20.9120	0.0000	-2.2
4	9	0.0000	55.6180	0.0000	-3.1
5	6	0.0000	25.2020	0.0000	-6.8
6	11	9.4980	19.8900	0.0000	
6	12	12.2910	25.5810	0.0000	
6	13	6.6150	13.0270	0.0000	
7	8	0.0000	17.6150	0.0000	
7	9	0.0000	11.0010	0.0000	
9	10	3.1810	8.4500	0.0000	
9	14	12.7110	27.0380	0.0000	
10	11	8.2050	19.2070	0.0000	
12	13	22.0920	19.9880	0.0000	
13	14	17.0930	34.8020	0.0000	

Table B.2 Bus Data

<u>BUS</u>	<u>TYPE</u>	<u>VOLTAGE</u>	<u>% SHUNT CAP.</u>
1	slack	1.060	
2	generator	1.045	
3	generator	1.010	
4	load	1.000	
5	load	1.000	
6	generator	1.070	
7	load	1.000	
8	generator	1.090	
9	load	1.000	19.00
10	load	1.000	
11	load	1.000	
12	load	1.000	
13	load	1.000	
14	load	1.000	

System MVA Base = 100 MVA

Table B.3 Probabilistic Data

Variable	Bus	Type	Active Power		Probability
			Power (P.U.)	% Stand. Dev.	
1	1	Normal	2.300	9.33	
2	2	Normal	.400	9.00	
3	2	Normal	-.217	9.00	
4	3	Normal	-.942	10.00	
5	4	Normal	-.478	11.00	
6	5	Normal	-.076	5.00	
7	6	Normal	-.122	1.00	
	7	Fixed	.000		
	8	Fixed	.000		
8	9	Normal	-.295	1.00	
9	10	Normal	-.090	10.00	
10	11	Normal	-.035	9.50	
11	12	Normal	-.061	1.00	
12	13	Normal	-.135	1.00	
13	14	Discrete	-.626		.20
	14	Discrete	-.150		.45
	14	Discrete	-.125		.35
Reactive Power					
14	4	Normal	-.039	9.70	
15	5	Normal	-.016	5.00	
	7	Fixed	.000		
16	9	Normal	-.166	5.00	
17	10	Normal	-.058	10.00	
18	11	Normal	-.018	9.50	
19	12	Normal	-.016	8.60	
20	13	Normal	-.058	9.50	
21	14	Normal	-.050	8.60	

All but one of the 21 random variables that serve as input are normally distributed. The active power load at bus 14 is a three point discrete distribution. The percent standard deviations shown are a percent of the mean value, and are scaled with $\rho=1$ (see Chapter 6).

REFERENCES

- Allan, R.N. and Al-Shakarchi, M.R.G. (1976)
 "Probabilistic A.C. Load Flow", Proceedings IEE, Vol. 124, No. 2, February, p. 154.
- Bendat, J.S. and A.G. Piersol (1971)
 Random Data: Analysis and Measurement Procedures, John Wiley and Sons, New York.
- Billingsley, P. (1968)
Convergence of Probability Measures, John Wiley and Sons, New York.
- Borkowska, B. (1974)
 "Probabilistic Load Flow", IEEE Transactions on Power Apparatus and Systems, PAS-93, No. 3, pp. 752-759.
- Box, G.E.P. and Jenkins, G.M. (1976)
Time Series Analysis Forecasting and Control, Holden-Day, Inc., San Francisco.
- Cramér, H. and M.R. Leadbetter (1967)
Stationary and Related Stochastic Processes, John Wiley and Sons, New York.
- Deutsch, R. (1962)
Nonlinear Transformations of Random Processes, Prentice-Hall, Englewood Cliffs, New Jersey.
- Dobb, J.L. (1963)
Stochastic Processes, John Wiley and Sons, New York.
- Dopazo, J. , Klittin, O., and Sasson, A.M. (1975)
 "Stochastic Load Flows", IEEE Transactions on Power Apparatus and Systems, PAS-94, No. 2, pp. 299-309.
- Feller, W. (1968)
An Introduction to Probability Theory and Its Applications, Volume I, 2nd Ed., John Wiley and Sons, New York.
- Fuller, W.A. (1976), Introduction to Statistical Time Series, John Wiley and Sons, New York.
- Halmos, P.R. (1956), Lectures on Ergodic Theory, Chelsea Publishing Company, New York.
- Hannan, E.J. (1970)
Multiple Time Series, John Wiley and Sons, New York.
- Karlin, S. (1966)
A First Course in Stochastic Processes, Academic Press, New York.
- Loève, M. (1963)
Probability Theory, 3rd Ed., Van Nostrand Reinhold, New York.
- Papoulis, A. (1965)
Probability, Random Variables, and Stochastic Processes, McGraw-Hill Book Company, New York.
- Parzen, E. (1967)
Stochastic Processes, 3rd Printing, Holden-Day Inc., San Francisco.
- Rozanov, Y.A. (1967)
Stationary Random Processes, Holden-Day, Inc., San Francisco.
- Sauer, P.W. (1977)
A Generalized Stochastic Power Flow Algorithm, Purdue University, TR-EE 77-32, PCTR 47-77.