NASA CONTRACTOR REPORT

NASA CR-10

NASA CR-1016

GUIDANCE, FLIGHT MECHANICS AND TRAJECTORY OPTIMIZATION

Volume XVII - Guidance System Performance Analysis

by D. R. Grier

Prepared by
NORTH AMERICAN AVIATION, INC.
Downey, Calif.
for George C. Marshall Space Flight Center

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • APRIL 1968

NASA CR-1016

GUIDANCE, FLIGHT MECHANICS AND TRAJECTORY OPTIMIZATION

Volume XVII - Guidance System Performance Analysis

By D. R. Grier

Issued by Originator as Report No. SID 66-1678-9

Prepared under Contract No. NAS 8-11495 by
NORTH AMERICAN AVIATION, INC.
Downey, Calif.

for George C. Marshall Space Flight Center

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

For sale by the Clearinghouse for Federal Scientific and Technical Information
Springfield, Virginia 22151 – CFSTI price $3.00

FOREWORD

This report was prepared under contract NAS 8-11495 and is one of a series intended to illustrate analytical methods used in the fields of Guidance, Flight Mechanics, and Trajectory Optimization. Derivations, mechanizations and recommended procedures are given. Below is a complete list of the reports in the series.

Volume I	Coordinate Systems and Time Measure
Volume II	Observation Theory and Sensors
Volume III	The Two Body Problem
Volume IV	The Calculus of Variations and Modern Applications
Volume V	State Determination and/or Estimation
Volume VI	The N-Body Problem and Special Perturbation Techniques
Volume VII	The Pontryagin Maximum Principle
Volume VIII	Boost Guidance Equations
Volume IX	General Perturbations Theory
Volume X	Dynamic Programming
Volume XI	Guidance Equations for Orbital Operations
Volume XII	Relative Motion, Guidance Equations for Terminal Rendezvous
Volume XIII	Numerical Optimization Methods
Volume XIV	Entry Guidance Equations
Volume XV	Application of Optimization Techniques
Volume XVI	Mission Constraints and Trajectory Interfaces
Volume XVII	Guidance System Performance Analysis

The work was conducted under the direction of C. D. Baker, J. W. Winch, and D. P. Chandler, Aero-Astro Dynamics Laboratory, George C. Marshall Space Flight Center. The North American program was conducted under the direction of H. A. McCarty and G. E. Townsend.

TABLE OF CONTENTS

Page

1.0 STATEMENT OF THE PROBLEM

The final consideration of space guidance systems to be discussed in this series of monographs is that of system performance analysis. In general terms, system performance analysis implies both an assessment of system performance and an assessment of system requirements. In this context, system requirements refer to the specifications of system functions in order to achieve mission objectives and system performance refers to the results which can be expected. Of course, system performance and requirements are closely related and can generally be considered as equivalent in terms of analysis effort. That is, the analysis effort which establishes system performance also establishes system requirements, however, the relationship is not necessarily a one-to-one correspondence. In general, the ultimate objective of system performance analysis is a system configuration definition, or system design specification which is sufficient in terms of system functions that directly and significantly affect system performance. It is tacitly assumed that the required system performance can be stated in explicit terms as translated from mission objectives and/or requirements. Usually, system performance requirements must be deduced from mission objectives and then system functions are defined to achieve system performance within mission constraints. In addition to a sufficient system configuration definition an optimum system design is desirable wherein the least stringent set of sufficient requirements is specified. Therefore, sufficiency and optimality of system design are of primary consideration in system performance analysis. It is highly desirable to establish a "single-step" system design algorithm which would achieve the optimum sufficient system configuration definition in a direct and immediate manner. Unfortunately, due to the complexity of system function interrelationships, the number of possible alternatives, the lack of uniqueness in a set of requirements, and changes in mission requirements the system configuration definition evolves from an iterative design procedure which should be a uniformly convergent process that achieves the final sufficient and optimum system configuration in an intelligent and efficient manner. The primary function of system performance analysis can be considered as providing the convergence to the iterative design procedure.

The totality of efforts involved from mission conception and definition of objectives to finalized system configuration definition and/or design specifications for a space guidance system is a significant and formidable undertaking. The present effort does not consider the total system design effort. A significant portion of the total effort has been considered in the previous monographs in this series and the present effort is intended to supplement the previous efforts which define the basic relationships between system functions and performance. These relationships comprise the basic elements of the complete system model which is required to perform an overall system performance analysis. There exist two fundamental alternatives in the present effort. One alternative consists of a consideration of various specific cases which would supposedly be representative of guidance system performance analyses for general missions. The second

alternative consists of a consideration of methodology which has general application. The first of these alternatives has the serious deficiency of being only applicable to the cases considered with general usefulness severly limited, especially in an era of technological evolution and revolution. Both the types of techniques and their degree of utilization continually change with time and progress, and present conclusions must continuously be reviewed and revised as technology evolves. The second alternative has the primary merit of not being restricted to particular performance analyses; however, there exists the risk that generality will obscure the direct applicability. A compromise between these two alternatives with emphasis on the latter has been taken in this effort. The primary purpose of this effort is the consideration of the methodology of system performance analysis which forms the principles and techniques of direct applicability in assessing guidance system performance and requirements. The methodology is directly applicable to a guidance system which is dependent upon particular mission considerations.

In general, the performance of a system is affected by the behavior of the system's functions and the nature of the environment in which the system operates. It is generally possible to describe system performance in terms of a particular circumstance of system functions and environment. This aspect of the problem can be considered as the deterministic aspect of system performance analysis. A deterministic model of a system can be considered as the correspondence between given system functions and environment and resulting system performance. The basic elements of this model have been considered in the efforts of the previous monographs in this series. Unfortunately, both system functions and environment do not obey fixed deterministic rules of behavior and, therefore, system performance cannot be stated on an explicit basis. A deterministic model of a system is utilized to establish the nominal system requirements, but this model does not specify performance and final requirements for system operation in its natural environment. Both system functions and operating environment are characterized by elements of uncertainty which significantly affect system performance. Thus, system performance is characterized by uncertainty and final system performance and requirements must be assessed in accordance with the inherent uncertainty of the situation. Therefore, there exist two aspects of system performance analysis which can be defined as deterministic and statistical considerations. The deterministic considerations are often referred to as nominal considerations which follow directly from the deterministic model of the system. However, these considerations do not yield final system performance, configuration or requirements. Rather, a nominal system configuration is tentatively defined. The nominal design must be subjected to a comprehensive statistical analysis to assess expected system performance and to modify design to insure compatibility of final system configuration definition and system performance requirements. The efforts of the previous monographs in this series provide the basis for the nominal considerations of system performance. The present effort is concenred primarily with the statistical considerations of the problem.

It should be emphasized that system performance analyses are essentially statistical inferences. These inferences are always subjected to degrees of uncertainty which should be recognized and assessed. It is only in this manner that the final risk involved in commiting a particular system

configuration to development and deployment can be known and reduced to an acceptable level. It is apparent that technology and methodology for nominal system performance analysis is usually adequate and readily available. However, the statistical methodology is not as readily available or as completely understood in terms of applicability, utilization and limitations. Basic methodology is often utilized without regard to the effects of basic assumptions. On the other hand, useful methods are not utilized due to a lack of familiarity. It is the basic purpose of the present effort to present useful methods of analysis and to discuss applicability and limitations of methods. A brief description of the present effort is given below.

A basic premise of this effort is that the nominal performance of a system can be written in terms of a vector equation as follows:

$$\underline{y} = \underline{G}(\underline{x})$$

In this equation $\underline{y}$ is a vector of system performance parameters for which requirements are specified, $\underline{x}$ is a vector of system functions and environmental parameters which affect system performance, and $\underline{G}(\ )$ is a known function which is deduced from physical laws that the system obeys. In the general situation $\underline{x}$ is a set of random phenomena which reflects the uncertainty in the behavior of system functions and environment; hence, $\underline{y}$ becomes a set of random phenomena as reflected through the functional relationship of $\underline{G}(\ )$. Due to the random or uncertain nature of $\underline{x}$ statistical methods must be utilized in assessing system performance and also system requirements. It should be noted that there generally exists a degree of uncertainty even in the statistical nature of $\underline{x}$ and, therefore, $\underline{y}$. The present consideration of the problem of system performance analysis is concerned with the mathematical framework in which the assumptions implicit in such analyses can be appreciated and intelligent application of general methods can be made. This objective, in turn, can be realized through the theory of statistical inference since the problem being considered is embodied in a more general theory and structure in the extensive literature on the general subject. However, the complete theory is not required, thus, the present effort is concerned with presenting that portion of the general theory which is directly applicable to the problem of system performance analysis. A particular application is considered in terms of an error model and analysis of an Inertial Measurement Unit (IMU) which is of direct usefulness in guidance and navigation system performance analyses.

2.0 STATE OF THE ART

2.1 THE BASIC MODEL OF THE PROBLEM

The general problem of system performance analysis can be defined in the following terms. There generally exist two sets of parameters which can be considered as (1) performance parameters, denoted by $\underline{y}$, and (2) causal parameters, denoted by $\underline{x}$. In this context, performance parameters are generally associated with system state quantities that directly affect mission success, and causal parameters are associated with system functions and environmental factors that affect system performance. In general, there exists a known relationship between the two parameter sets $\underline{y}$ and $\underline{x}$, denoted by $\underline{y} = \underline{G}(\underline{x})$ where $\underline{G}(\ )$ is a known function which is deduced from physical laws that the system obeys. The explicit relationship between $\underline{y}$ and $\underline{x}$ is dependent upon a particular system configuration definition. In general, there exists a region in the space or domain of the set $\underline{y}$ which is conclusive to mission success, i.e., if $\underline{y} \in R_S$ then the mission is successful. Thus, R_S can be considered as a "region-of-success" or a "success" region for $\underline{y}$. Usually, the definition of $\underline{y}$ and R_S depends on mission type, objectives and constraints. Now, two basic purposes of system performance analyses can be defined which are (1) given $\underline{x}$ and $\underline{G}(\ )$, determine if $\underline{y} \in R_S$ and/or (2) determine the requirements for $\underline{x}$ and/or $\underline{G}(\ )$ such that $\underline{y} \in R_S$. The ultimate objective is to definitize the system configuration and specify tolerances or design requirements for system functions which are sufficient to achieve mission success. Within this objective the optimum system is sought which consists of the least stringent set of sufficient system requirements. That is, system configuration and function requirements for the achievement of mission objectives are generally not unique and there exist a number of alternatives. Although some alternatives are precluded by mission constraints there exist a number of possible alternatives of which some are sufficient and supposedly one is optimum.

If the system operating environment and system functions are known with certainty, then system performance and mission success could be stated with certainty. In such a situation the system could be "tailor-made" with absolute assurance of success and the optimum system could be readily defined. Unfortunately, this is not the usual situation. Both system environment and functions are not explicitly known entities, rather, they are generally random phenomena or random processes. That is, the causal parameter set $\underline{x}$ is a random vector and, hence, the performance parameter set $\underline{y}$ is a random vector. Fortunately, mission objectives usually allow some degree of uncertainty in system performance parameters, i.e., the success region R_S is not a single point. Due to the random nature of the situation, system performance analyses must consider the probability that $\underline{y}$ will lie in the region R_S. Alternatively, the task of system performance analysis is directly concerned with determining if the uncertainty in the system performance parameters is compatible with mission objectives; moreover, these tasks are concerned with determining an optimum system configuration definition which fulfills system performance requirements in accordance with a specified probability

of success, i.e., probability that $\underline{y} \in R_S$. To this end performance analysis is concerned with four general tasks which follow the definition of mission objectives and constraints.

First, the dependence of $\underline{y}$ upon $\underline{x}$ must be established. This can be considered as the deterministic aspect of the problem which is not of primary consideration in this effort.

Second, the uncertainty of the various system functions and environment must be specified. This can be considered as the fundamental statistical aspect of the problem and it is of primary consideration in this effort. This aspect of the problem is concerned with the statistical analysis of the random phenomena represented by the random vector $\underline{x}$. This task is a necessary, but not sufficient, effort in system performance analyses.

Third, with knowledge of the relationship $\underline{y} = \underline{G}(\underline{x})$ and with a knowledge of the uncertainty of the causal parameter set $\underline{x}$, the uncertainty of system performance is determined.

Fourth, assess the probability of mission success.

These tasks are usually accomplished within the system design iteration process to evolve the optimum system configuration definition which will fulfill mission objectives. The methodology is ultimately that of the general principles of statistical inference. The particular methods of the general principles which must be utilized are discussed in the following sections.

2.2 STATISTICAL METHODOLOGY

2.2.1 Introduction

System performance analysis is ultimately concerned with the analysis of random phenomena including system functions, operating environment and, finally, system performance. These random phenomena must be characterized or defined in statistical terms, i.e., adequate information of these phenomena must be obtained such that the nature of the randomness is sufficiently known. This is the subject of the general methods of statistical inference or statistical analysis. Those particular methods of statistical analysis which are directly applicable to system performance analysis are considered in this section. It is intended that the material presented herein is somewhat self-sufficient with adequate discussions presented so the applicability and limitations of methods are readily understood. On the other hand, an exhaustive treatment of the general subject is not presented nor intended since it is not required. An attempt has been made to provide sufficient and useful references where it is recognized that certain extensions of the basic methods will be required in certain cases. In addition, an extensive bibliography on the general subject of statistical inference is provided.

It is tacitly assumed that the reader is adequately familiar with the basic concepts of randomness and probability. This is a matter of convenience since a sizable treatise could be written on the conceptual aspects of these entities, however, this does not greatly serve the purpose of present application. Usually, for purposes of engineering application the basic concepts of randomness and probability suffice, although, there is often a lack of agreement with the more rigorous mathematical definitions of these concepts. Ultimately, the rigorous formulation is required, but this is not considered herein. Good discussions on this subject can be found in References 1, 2 and 3.

The discussions begin with basic definitions and properties which are frequently encountered. A Gaussian random variable is discussed in detail. The multivariate Gaussian probability density function is considered in detail in the definitions and Appendices B and C. Probabilities for Gaussian random vectors are specified for various regions of interest.

Functions of random variables are discussed with particular emphasis on transformations of probability density functions and statistical moments of functions of random variables. Several particular functions of Gaussian random variables are discussed with emphasis upon the probability density functions and statistical moments.

Several basic probability bounds are discussed which can generally be used to "bound" random variables when only lower order statistical moments are known. Similarly, several basic limiting theorems are discussed which generally concern the limiting behavior of sums of statistically independent random variables.

The determination of statistical properties is discussed with particular concern of estimating moments and examining the validity of assumptions concerning probability density functions. The particular case of estimating statistical moments for Gaussian random variables is considered in some detail. The basic methods of correlation and regression analyses are discussed. The use of confidence intervals is discussed and the method of hypothesis testing is considered.

2.2.2 Basic Definitions and Properties

2.2.2.1 Random Process

A random process can be defined as any phenomenon for which repeated observations, under a given set of conditions,do not yield identical results. In general, random processes are characterized by variations in outcomes for repeated equivalent trials. These variations in outcomes or observations are considered as the "randomness" of the process, which is equivalent to uncertainty in the outcome of the process. As a contrary example, consider a process whose behavior is completely described by a known system of differential equations. Theoretically, it is possible to completely determine the behavior of such a process if an adequate set of observations are made at some time. Such a process is said to possess deterministic regularity. However, until such time that all physical laws are explicitly established for the microscopic and infinitesimal domains, the concept of random physical processes must be admitted, accepted, and dealt with.

Alternatively, a random process could be defined as one which does not possess deterministic regularity and subsequent outcomes cannot be predicted with certainty from a set of observations of the process. However, a random process can possess definite properties of behavior which make possible a description on a statistical basis. Such random processes are said to possess statistical regularity. In such cases, even though particular outcomes of the process cannot be specified, it is possible to specify the relative frequency or probability of occurrence of outcomes for the process.

2.2.2.2 Random Variable

A random variable is defined as a real-valued function which is defined for each outcome of a random process. Of course, the outcomes for many random processes are actually random variables. Such random processes are quantitative or numerical processes, e.g., random voltages, pressure, errors, etc. On the other hand, random processes exist which are non-numerical, such as the tossing of a coin where the outcome is either a heads or tails. However, it is possible to define a random variable for this random process by assigning numbers to the outcomes or by defining the random variable to the number of heads in m tosses of a coin, etc.

The importance of the concept of a random variable lies in the fact that many of the arithmetic, algebraic and analytical operations which are defined for real-valued functions are meaningful for random variables, whereas they are not for the outcomes of all random processes. Thus,

additions, subtractions, multiplications, transformations, etc., are applicable to random variables.

2.2.2.3 Random Vector

In general, a random vector $\underline{x}$ of dimension n is an ordered set of n random variables, i.e.,

$$\underline{x} = (x_1, x_2, \ldots, x_i, \ldots, x_n)$$

where x_i is a random variable. The ordered set can be written as a row or column matrix or vector and convention seems to favor column vectors, i.e.,

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{bmatrix} \quad \text{and} \quad \underline{x}^T = \begin{bmatrix} x_1, & x_2, & \ldots, & x_i, & \ldots, & x_n \end{bmatrix}$$

where superscript T denotes transpose.

The basic property of random vectors, for engineering purposes, is that the domain of definition of each component is the set of real numbers, i.e.,

$$-\infty < x_i < \infty$$

for $i = 1, 2, \ldots, n$.

2.2.2.4 Probability Density and Distribution Functions

Let $P[\underline{x} \in R]$ denote the probability that the random vector $\underline{x}$ will lie in the region R, which is a subset of the domain of definition of $\underline{x}$. If a function of $\underline{x}$, $f(\underline{x})$, exists such that $P[\underline{x} \in R]$ is the multiple integral of $f(\underline{x})$ over the region or subset R then $f(\underline{x})$ is the probability density function of $\underline{x}$. That is, if $f(\underline{x})$ is the probability density function of $\underline{x}$ then

$$P[\underline{x} \in R] = \int_R f(\underline{x})\, d\underline{x}$$

where $\int_R (\ )\, d\underline{x}$ denotes a multiple integral over R. A basic

property of $f(\underline{x})$ is that if R is the domain of $\underline{x}$, i. e., the set of all possible values of $\underline{x}$ then

$$\int_{R=D(\underline{x})} f(\underline{x})\, d\underline{x} = 1.0$$

where $D(\underline{x})$ is the domain of $\underline{x}$. This follows from the fact that $D(\underline{x})$ is exhaustive for $\underline{x}$ and, hence, $P[\underline{x} \in D(\underline{x})] = 1$. Moreover, $f(\underline{x})$ is a positive semi-definite function of $\underline{x}$, or $f(\underline{x})$ is a non-negative function of $\underline{x}$; i. e.,

$$f(\underline{x}) \geq 0 \quad \text{for all } \underline{x} .$$

If the region R is defined by $-\infty < x_i \leq Z_i$ for $i = 1, 2, \ldots n$ then $f(\underline{x})$ integrated over R yields the "probability distribution function" for $\underline{x}$, denoted by $F_{\underline{x}}(\underline{z})$. That is,

$$F_{\underline{x}}(\underline{z}) = P\left[-\infty < \underline{x} \leq \underline{z}\right] = \int_{-\infty}^{\underline{z}} f(\underline{x})\, d\underline{x}$$

A basic property of $F_{\underline{x}}(\underline{z})$ is that $F_{\underline{x}}(\underline{z})$ is a monotonically non-decreasing function of $\underline{z}$, i. e., as $\underline{z}$ is "increased" over the domain of $\underline{x}$ the $P[-\infty < \underline{x} \leq \underline{Z}]$ cannot decrease. Moreover, it is apparent that $0 \leq F_{\underline{x}}(\underline{Z}) \leq 1$ for $\underline{z}$ over $D(\underline{x})$.

If $\underline{x}_1$ and $\underline{x}_2$ are two random vectors of dimensions n_1 and n_2, respectively, then the "joint" probability density function of $\underline{x}_1$ and $\underline{x}_2$ is simply the probability density function of $\underline{x}$ where $\underline{x}$ contains $\underline{x}_1$ and $\underline{x}_2$ as subvectors. Therefore, the probability density function $f(\underline{x})$ of any random vector $\underline{x}$ of dimension $n > 1$ is a joint probability density function and $f(\underline{x})$ can generally be writted as

$$f(\underline{x}) = f(\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_j, \ldots, \underline{x}_m)$$

where $\underline{x}_j$ are subvectors of $\underline{x}$ with dimensions n_j, respectively. Of course, the dimension of $\underline{x}$, n, is given by

$$n = \sum_{j=1}^{m} n_j$$

If $\underline{x}$ is composed of two subvectors $\underline{x}_1$ and $\underline{x}_2$ then the "marginal" probability density functions of $\underline{x}_1$ and $\underline{x}_2$, $f(\underline{x}_1)$ and $f(\underline{x}_2)$, respectively, are given by

$$f(\underline{x}_1) = \int_{D(\underline{x}_2)} f(\underline{x})\, d\underline{x}_2 = \int_{D(\underline{x}_2)} f(\underline{x}_1, \underline{x}_2)\, d\underline{x}_2$$

$$f(\underline{x}_2) = \int_{D(\underline{x}_1)} f(\underline{x})\, d\underline{x}_1 = \int_{D(\underline{x}_1)} f(\underline{x}_1, \underline{x}_2)\, d\underline{x}_1$$

The marginal probability density function $f(\underline{x}_1)$ determines the probability that $\underline{x}_1$ will lie in a region R_1 in the domain of $\underline{x}_1$ without regard to $\underline{x}_2$, i.e.,

$$P(\underline{x}_1 \in R_1) = \int_{R_1} f(\underline{x}_1)\, d\underline{x}_1$$

Similarly,

$$P(\underline{x}_2 \in R_2) = \int_{R_2} f(\underline{x}_2)\, d\underline{x}_2$$

It sould be noted that the marginal probability density function of a subvector $\underline{x}_1$ of $\underline{x}$ is independent of the subvector $\underline{x}_2$, where $\underline{x}$ is composed of $\underline{x}_1$ and $\underline{x}_2$.

If $\underline{x}$ is composed of two subvectors $\underline{x}_1$ and $\underline{x}_2$ then the "conditional" probability density function of $\underline{x}_1$, given $\underline{x}_2$, $f(\underline{x}_1/\underline{x}_2)$, is given by

$$f(\underline{x}_1/\underline{x}_2) = \frac{f(\underline{x}_1, \underline{x}_2)}{f(\underline{x}_2)}$$

Similarly,

$$f(\underline{x}_2/\underline{x}_1) = \frac{f(\underline{x}_1, \underline{x}_2)}{f(\underline{x}_1)}$$

From the foregoing it is seen that

$$f(\underline{x}) = f(\underline{x}_1, \underline{x}_2) = f(\underline{x}_1/\underline{x}_2)\, f(\underline{x}_2)$$
$$= f(\underline{x}_2/\underline{x}_1)\, f(\underline{x}_1)$$

Therefore,

$$f(\underline{x}_1/\underline{x}_2)\, f(\underline{x}_2) = f(\underline{x}_2/\underline{x}_1)\, f(\underline{x}_1)$$

Also,

$$\frac{f(\underline{x}_1/\underline{x}_2)}{f(\underline{x}_1)} = \frac{f(\underline{x}_2/\underline{x}_1)}{f(\underline{x}_2)}$$

For the sake of notation convenience, "pdf" will be used to denote "probability density function" and, similarly, "PDF" will be used to denote "probability distribution function" in the text. In this notation pdf of $\underline{x}$ = $f(\underline{x})$ and PDF of $\underline{x} = F_{\underline{x}}(\underline{z})$. It should be noted that if $\underline{x}_1$ and $\underline{x}_2$ are two different random vectors then, generally, the pdf of $\underline{x}_1 = f(\underline{x}_1)$ is not equal to the pdf of $\underline{x}_2 = f(\underline{x}_2)$, i. e., the notation $f(\underline{x}_1)$ and $f(\underline{x}_2)$ does not imply that $f(\underline{x}_1) = f(\underline{x}_2)$ or that $\underline{x}_1 = \underline{x}_2$.

2. 2. 2. 5 Statistical Independence

If the pdf of the random vector $\underline{x}$, which is composed of subvectors $\underline{x}_1$ and $\underline{x}_2$, can be written as

$$f(\underline{x}) = f(\underline{x}_1)\, f(\underline{x}_2)$$

then the subvectors $\underline{x}_1$ and $\underline{x}_2$ are defined to be "statistically" independent. In general, a random vector $\underline{x}$ is a statistically independent random vector if all of its components are statistically independent. In this case $f(\underline{x})$ can be written as the product of the n pdfs of the components

of $\underline{x}$, i. e. ,

$$f(\underline{x}) = f(x_1)\, f(x_2) \ldots f(x_i) \ldots f(x_n)$$

$$f(\underline{x}) = \prod_{i=1}^{n} f(x_i)$$

It should be noted that these definitions differentiate between statistically independent vectors $\underline{x}_1$ and $\underline{x}_2$ and a statistically independent vector $\underline{x}$. Also, in the latter case it is not necessary that the probability density functions for the vector components be the same.

2. 2. 2. 6 Mathematical Expectation

Let $y = g(\underline{x})$ be a scalar function of the random vector $\underline{x}$. If

$$\int_{D(\underline{x})} |g(\underline{x})|\, f(\underline{x})\, dx$$

exists, then the "mathematical expection" of y, denoted by E(y), is defined as follows.

$$E(y) = \int_{D(\underline{x})} g(\underline{x})\, f(\underline{x})\, d\underline{x}$$

In a similar manner, let $\underline{y}$ be a set of scalar functions of the random vector $\underline{x}$, i. e., $y_1 = g_1(\underline{x})$, $y_2 = g_2(x)$, ..., $y_m = g_m(\underline{x})$, or $\underline{y} = \underline{g}(\underline{x})$,

where $\underline{y}$ and $\underline{x}$ can be of dimensions m and n where $m \neq n$. If

$$\int_{D(\underline{x})} |g_i(\underline{x})|\, f(\underline{x})\, dx$$

exists for i = 1, 2, ... m, then

$$E(y_i) = \int_{D(\underline{x})} g_i(\underline{x})\ f(\underline{x})\ d\underline{x}$$

or

$$E(\underline{y}) = \int_{D(\underline{x})} \underline{g}(\underline{x})\ f(\underline{x})\ d\underline{x}$$

Hereafter "E( )" will denote "the expectation of" in accordance with the above definition. It is common practice to refer to E(y) as "the mean value" of y. For the special case of $\underline{y} = \underline{x}$ the mean value of $\underline{x}$ becomes

$$E(\underline{x}) = \int_{D(\underline{x})} \underline{x}\ f(\underline{x})\ d\underline{x}$$

Several basic properties of expectation exist which are noted below.

I. If C is a constant then E(C) = C.

II. If $y = g(\underline{x})$ and $z = c\,y$ then
$$E(z) = E(cy) = cE(y)$$

III. If $z = \sum_{i=1}^{m} c_i y_i$ then $E(z) = \sum_{i=1}^{m} c_i E(y_i)$

IV. If $y_1 = g_1(\underline{x}_1)$ and $y_2 = g_2(\underline{x}_2)$ and

if $\underline{x}_1$ and $\underline{x}_2$ are statistically independent vectors

then

$$E[(c_1 y_1)(c_2 y_2)] = c_1 c_2\ E(y_1)\,E(y_2)$$

Property IV can be established as follows. Let

$$y = c_1 y_1 c_2 y_2 = c_1 c_2 g_1(\underline{x}_1) g_2(\underline{x}_2) = c\, g(\underline{x})$$

Now E(y) can be written as follows:

$$E(y) = cE[g(\underline{x})] = c \int_{D(\underline{x})} g(\underline{x})\, f(\underline{x})\, d\underline{x}$$

$$= c \int_{D(\underline{x}_1)} \int_{D(\underline{x}_2)} g_1(\underline{x}_1)\, g_2(\underline{x}_2)\, f(\underline{x}_1)\, f(\underline{x}_2)\, d\underline{x}_1\, d\underline{x}_2$$

$$= c \int_{D(\underline{x}_1)} g_1(\underline{x}_1)\, f(\underline{x}_1)\, d\underline{x}_1 \int_{D(\underline{x}_2)} g_2(\underline{x}_2)\, f(\underline{x}_2)\, d\underline{x}_2$$

$$E(y) = c_1 c_2\, E(y_1) E(y_2)$$

In general, if $y = g(\underline{x})$ and if $\underline{x}$ is composed of two subvectors $\underline{x}_1$ and $\underline{x}_2$ it is possible to define the expectation of y with respect to $\underline{x}_1$ with $\underline{x}_2$ assumed constant. That is, if $y = g(\underline{x}) = g(\underline{x}_1, \underline{x}_2)$ then y will vary randomly even if $\underline{x}_2$ is a constant vector, hence, it is meaningful to consider the expectation of y on the condition that either $\underline{x}_1$ or $\underline{x}_2$ is a constant vector. Thus, the "conditional" expectation of y, given $\underline{x}_2 = \underline{c}$, denoted by $E(y/\underline{x}_2)$, is defined as follows.

$$E(y/\underline{x}_2) = \int_{D(\underline{x}_1)} g(\underline{x})\, f(\underline{x}_1/\underline{x}_2)\, d\underline{x}_1$$

where $f(\underline{x}_1/\underline{x}_2)$ is the conditional pdf as defined previously. Similarly,

$$E(y/\underline{x}_1) = \int_{D(\underline{x}_2)} g(\underline{x})\, f(\underline{x}_2/\underline{x}_1)\, d\underline{x}_2$$

2.2.2.7 Statistical Moments

Let a scalar function of $\underline{x}$, $g(\underline{x})$, be defined as follows:

$$g(\underline{x}) = x_1^{r_1} \cdot x_2^{r_2} \cdots x_i^{r_i} \cdots x_n^{r_n}$$

$$= \prod_{i=1}^{n} x_i^{r_i}$$

where $r_i = 0, 1, 2, 3, \ldots$ for all i. The general joint moment of $\underline{x}$, $m(\underline{r})$, of order $r_1 + r_2 + \ldots + r_i + \ldots + r_m$ is defined as

$$m(\underline{r}) = \int_{D(\underline{x})} g(\underline{x}) f(\underline{x}) \, d\underline{x}$$

where the vector r is the set $(r_1, r_2, \ldots, r_i, \ldots, r_m)$ and the integral is the multiple integral over the domain of $D(\underline{x})$. In a similar manner the general joint "central" moment, $\mu(\underline{r})$ is defined as

$$\mu(\underline{r}) = \int_{D(\underline{x})} g(\underline{x} - \underline{m}) \, f(\underline{x}) \, d\underline{x}$$

where $\underline{m}$ is the vector of joint moments of order 1, i.e.,

$$m_i = \int_{D(\underline{x})} x_i \, f(\underline{x}) \, d\underline{x} = \int_{D(x_i)} x_i \, f(x_i) \, dx_i$$

where $f(x_i)$ is the marginal pdf of x_i. Alternatively,

$$\underline{m} = \int_{D(\underline{x})} \underline{x} \, f(\underline{x}) \, d\underline{x}$$

In general, statistical moments can be considered as the **expectation** of the particular functions of $\underline{x}$ as defined above, i. e.,

$$m(\underline{r}) = E\left(\prod_{i=1}^{n} x_i^{r_i}\right)$$

and

$$\mu(\underline{r}) = E\left[\prod_{i=1}^{n} (x_i - m_i)^{r_i}\right]$$

where

$$m_i = E(x_i)$$

for $i = 1, 2, \ldots, n$.

Moments of particular interest are usually the first and second order joint moments and the second order central moments. (It is seen that the first order central moments are zero.) The first order joint moments are simply the mean value of the components of $\underline{x}$, i. e., $\underline{m} = E(\underline{x})$ as defined before. The second order moments consist of

$$E(x_i x_j) = \int_{D(\underline{x})} (x_i x_j) f(\underline{x}) d\underline{x}$$

for $i, j = 1, 2, \ldots, n$. Similarly, the second order central moments are given by

$$E\left[(x_i - m_i)(x_j - m_j)\right] = \int_{D(\underline{x})} (x_i - m_i)(x_j - m_j) f(\underline{x}) d\underline{x}$$

for $i, j = 1, 2, \ldots, n$. The second order central moments are usually referred as "variances" for i = j and "co-variances" for i ≠ j. The co-variances are usually denoted by μ_{ij} where

$$\mu_{ij} = E\left[(x_i - m_i)(x_j - m_j)\right]$$

The variances are denoted by $\sigma_i^2 = \mu_{ii}$, i. e. ,

$$\sigma_i^2 = E\left[(x_i - m_i)^2\right]$$

The variance for each x_i can be expressed in terms of its first and second moment, i. e. ,

$$\sigma_i^2 = E\left[x_i^2 - 2 x_i m_i - m_i^2\right] = E(x_i^2) - m_i^2 = m_{2i} - m_{1i}^2$$

$$= E(x_i^2) - E^2(x_i)$$

where m_{1i} and m_{2i} are the first and second moments of x_i , respectively.

A basic property of second central moments is the following inequality.

$$-\sigma_i \sigma_j \leq \mu_{ij} \leq +\sigma_i \sigma_j$$

or

$$\mu_{ij}^2 \leq \sigma_i^2 \sigma_j^2$$

This inequality can be established in the following manner. Let $g(x_i, x_j)$ be defined as follows:

$$g = (x_i, x_j) = \left[\left(\frac{x_i - m_i}{\sigma_i}\right) - \left(\frac{x_j - m_j}{\sigma_j}\right)\right]^2 \leq 0$$

Now $E\left[g(x_i, x_j)\right] \leq 0$ hence,

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x_i, x_j)\, f(x_i, x_j)\, dx_i\, dx_j \geq 0$$

where $f(x_i, x_j)$ is the joint pdf for x_i and x_j. Thus,

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \left[\frac{(x_i - m_i)^2}{\sigma_i^2} + \frac{(x_j - m_j)^2}{\sigma_j^2}\right] f(x_i, x_j)\, dx_i\, dx_j$$

$$\geq \frac{2}{\sigma_i\, \sigma_j}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x_i - m_i)(x_j - m_j)\, f(x_i, x_j)\, dx_i\, dx_j$$

It is apparent that the integral on the left side is simply 2 and that on the right side is μ_{ij}, hence,

$$+\sigma_i\, \sigma_j \geq \mu_{ij}$$

In a similar manner define

$$g(x_i, x_j) = -\left[\left(\frac{x_i - m_i}{\sigma_i}\right) + \left(\frac{x_j - m_j}{\sigma_j}\right)\right]^2 \leq 0$$

Now, $E\left[g(x_i, x_j)\right] \leq 0$ hence,

$$-\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left[\frac{(x_i - m_i)^2}{\sigma_i^2} + \frac{(x_j - m_j)^2}{\sigma_j^2}\right] f(x_i, x_j)\, dx_i\, dx_j$$

$$\leq \frac{2}{\sigma_i\,\sigma_j}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x_i - m_i)(x_j - m_j)\, f(x_i, x_j)\, dx_i\, dx_j$$

Thus,

$$-\sigma_i\,\sigma_j \leq \mu_{ij}$$

A basic property of first and second central moments of the scalar product $\underline{c}^T\underline{x}$ should be noted, where $\underline{c}$ is a constant vector and $\underline{x}$ is a random vector. Let $y = \underline{c}^T\underline{x}$ then

$$E(y) = E(\underline{c}^T\underline{x}) = \underline{c}^T E(\underline{x}) = \underline{c}^T\underline{m}$$

The second central moment or variance of y becomes

$$\sigma_y^2 = E\left[y - E(y)\right]^2 = E\left[\underline{c}^T\underline{x} - \underline{c}^T\underline{m}\right]^2 = E\left[\underline{c}^T(\underline{x} - \underline{m})\right]^2$$

$$= E\left[\sum_{j=1}^{n}\sum_{i=1}^{n} c_i\, c_j\, (x_i - m_i)(x_j - m_j)\right]$$

$$= \sum_{j=1}^{n}\sum_{i=1}^{n} c_i\, c_j\left[E(x_i - m_i)(x_j - m_j)\right]$$

$$\sigma_y^2 = \sum_{j=1}^{n}\sum_{i=1}^{n} c_i c_j \mu_{ij}$$

Also,

$$\sigma_y^2 = \sum_{i=1}^{n} c_i^2 \sigma_i^2 + 2\sum_{i \neq j}\sum c_i c_j \mu_{ij}$$

2.2.2.8 Co-variance Matrix

The second order central moments for a random vector $\underline{x}$ of dimension n comprise a set of n^2 elements. If these elements are arranged in a square matrix of order n x n with elements μ_{ij} then the resulting "matrix-of-covariances" is usually defined as the co-variance matrix. The co-variance matrix, Γ_x , for a random vector $\underline{x}$ can be written in vector notation as follows.

$$\Gamma_x = E[(\underline{x} - \underline{m})(\underline{x} - \underline{m})^T] = [\mu_{ij}]$$

where

$$\underline{m} = E(\underline{x})$$

The diagonal elements of Γ_x are the variances σ_i^2 of the components of the random vector $\underline{x}$. The trace of the co-variance matrix is the sum of its diagonal elements, therefore,

$$TRACE(\Gamma_x) = \sum_{i=1}^{n} \sigma_i^2 > 0$$

If $\underline{y} = A\underline{x}$ where A is a constant matrix of dimensions mxn then $\Gamma_y = A\,\Gamma_x\,A^T$. This can be shown as follows. If $y = A\underline{x}$ then

$$E(\underline{y}) = A\,E(\underline{x}) = A\underline{m}$$

$$\Gamma_y = E\left[(\underline{y} - A\underline{m})(\underline{y} - A\underline{m})^T\right] = E\left[(A\underline{x} - A\underline{m})(A\underline{x} - A\underline{m})^T\right]$$

$$= E\left[A(\underline{x} - \underline{m})(\underline{x} - \underline{m})^T\right]A^T = A\,E\left[(\underline{x} - \underline{m})(\underline{x} - \underline{m})^T\right]A^T$$

$$\Gamma_y = A\,\Gamma_x\,A^T$$

It should be noted that A is not necessarily square, however, Γ_y is a square symmetrical matrix of order m.

An important property of the co-variance matrix Γ_x for any random vector $\underline{x}$ is that Γ_x is a symmetrical positive definite matrix, i. e., the quadratic form $\underline{c}^T\,\Gamma_x\,\underline{c}$ is positive definite for $\underline{c} \neq \underline{0}$. This can be shown as follows. If $y = \underline{c}^T\underline{x}$ then y is a scalar with variance σ_y^2 which is always greater than zero if $\underline{c} \neq \underline{0}$, however, this is a special case of the matrix A above where $A = \underline{c}^T$. Thus,

$$\sigma_y^2 = \underline{c}^T\,\Gamma_x\,\underline{c} > 0$$

Another property of Γ_x is that the sum of all its elements is greater than zero. Simply let $\underline{c} = \underline{1}$, where $\underline{1}$ is a vector which has unity for each component, i. e.,

$$\underline{1}^T = (1, 1, \ldots, 1, \ldots, 1).$$

It is noted that if $\underline{c} = \underline{1}$, then $y = \underline{c}^T\underline{x} = \underline{1}^T\underline{x}$ is simply the sum of all the components of $\underline{x}$. In this case the variance of y, σ_y^2 , is the sum of the elements of the co-variance matrix for $\underline{x}$, i.e., if

$$y = \underline{1}^T\underline{x} = \sum_{i=1}^{n} x_i$$

then

$$\sigma_y^2 = \underline{1}^T \Gamma_x \underline{1} > 0$$

2.2.2.9 Correlation Coefficients

For any two components of a random vector a "correlation coefficient," ρ_{ij} , is defined as follows.

$$\rho_{ij} = \frac{\mu_{ij}}{\sqrt{\mu_{ii} \cdot \mu_{jj}}} = \frac{\mu_{ij}}{\sqrt{\sigma_i^2 \cdot \sigma_j^2}} = \frac{\mu_{ij}}{\sigma_i \cdot \sigma_j}$$

It is apparent that $\rho_{ij} = 1$ for $i = j$. A basic property of ρ_{ij} is the following inequality.

$$-1 \leq \rho_{ij} \leq +1$$

This follows directly from the inequality for second central moments given above, i.e.,

$$-\sigma_i \sigma_j \leq \mu_{ij} \leq +\sigma_i \sigma_j$$

2.2.2.10 Statistical Correlation and Orthogonality

In general, if the correlation coefficient, ρ_{ij}, for two components of a random vector is non-zero then the two components are referred to as being statistically correlated. Alternatively, two random variables are referred to as being uncorrelated if their second joint moment is equal to the product of their first moments, i.e., x_i and x_j are uncorrelated if

$$E(x_i x_j) = E(x_i) E(x_j) = m_i m_j$$

The co-variance for x_i and x_j, μ_{ij} is given by

$$\mu_{ij} = E\left[(x_i - m_i)(x_j - m_j)\right]$$

$$= E\left[x_i x_j - m_i x_i - m_j x_i + m_i m_j\right]$$

$$\mu_{ij} = E\left[x_i x_j\right] - m_i m_j$$

It is apparent that if $E(x_i x_j) = m_i m_j$, then $\mu_{ij} = 0$ and $\rho_{ij} = 0$.

Two components of a random vector are referred to as being statistically orthogonal if their second joint moment vanishes, i.e., if

$$E(x_i x_j) = 0$$

then x_i and x_j are statistically orthogonal.

It should be noted that statistical independence, correlation and orthogonality are related. That is, if x_i and x_j are statistically independent then they are statistically uncorrelated, however, the converse does not follow. Also, if x_i and x_j are uncorrelated and if at least one of their first moments vanishes then x_i and x_j are statistically orthogonal.

2. 2. 2. 11 Moment Generating Function

Let $\underline{s}^T\underline{x}$ denote the scalar product of the vectors $\underline{s}$ and $\underline{x}$, where $\underline{s}$ is a non-random vector. The moment generating function, $mgf_x(\underline{s})$, of the random vector $\underline{x}$ is defined as

$$mgf_x(\underline{s}) = E(e^{\underline{s}^T\underline{x}}) = \int_{D(\underline{x})} e^{\underline{s}^T\underline{x}} f(\underline{x})\, d\underline{x}$$

It is not difficult to show that the joint moments $m(r)$ of the random vector $\underline{x}$ can be determined from $mgf_x(\underline{s})$ by taking appropriate partial derivatives of $mgf_x(\underline{s})$ with respect to $\underline{s}$ and evaluating at $\underline{s} = \underline{o}$, i. e.,

$$m(r) = \frac{\partial^{r_1}}{\partial s_1^{r_1}} \cdot \frac{\partial^{r_2}}{\partial s_2^{r_2}} \cdot \ldots \cdot \frac{\partial^{r_i}}{\partial s_i^{r_i}} \cdot \ldots \frac{\partial^{r_n}}{\partial s_n^{r_n}} mgf_x(\underline{s}) \Big|_{\underline{s}=\underline{o}}$$

This follows from the fact that

$$\frac{\partial^{r_i}}{\partial s_i^{r_i}} mgf_x(\underline{s}) = \int_{D(\underline{x})} \frac{\partial^{r_i}}{\partial s_i^{r_i}} e^{\underline{s}^T\underline{x}} f(\underline{x})\, d\underline{x}$$

$$= \int_{D(\underline{x})} x_i^{r_i} e^{\underline{s}^T\underline{x}} f(\underline{x})\, d\underline{x}$$

Taking subsequent partials derivatives and setting $\underline{s} = \underline{o}$ yields the moments $m(\underline{r})$ since $e^0 = 1$.

It is easily seen that if the components of $\underline{x}$ are statistical independent then the moment generating function for $\underline{x}$ becomes the product of the moment generating functions of the components of $\underline{x}$, i. e., if

$$f(\underline{x}) = \prod_{i=1}^{n} f(x_i)$$

then

$$mg\, f_x(\underline{s}) = \prod_{i=1}^{n} mg\, f(s_i)$$

where $mfg(s_i)$ is the moment generating function for x_i . The converse is also true, i. e., if the moment generating function of a set of random variables factors into the product of functions of each random variable then the random variables are statistically independent. The same holds for subsets or subvectors of random vectors.

The most important property of the moment generating function for a random variable is that under rather general conditions the moment generating function and the probability density function is a unique integral transform pair. That is, probability density functions usually associated with "physical" random phenomena and their moment generating functions are uniquely related. This is readily illustrated by considering a positive definite random variable x such that $f(x) = 0$ for $x \leq 0$. In this case the moment generating function for x is equivalent to the Laplace transform of $f(x)$ where $s = -(\alpha + J\omega) = -s'$ and s' is the usual variable of transformation for the Laplace transform. Generally, for values of s for which the moment generating function converges, the moment generating function and the probability density function for a random variable are a unique integral trasnform pair. In general terms, a sufficient condition for uniqueness is that the probability density function is continuous. An alternate statement of the uniqueness of moment generating functions and probability density functions is as follows. Let x and y be two random variables with probability density functions $f(x)$ and $f(y)$, respectively. If the moment generating functions for x and y exist for $-\alpha^2 < s < +\alpha^2$ and are equal in this interval then x and y have equal probability density functions except possibly at points of discontinuities. The convergence, existence and uniqueness of moment generating functions is discussed in detail in References 1, 2, 3, and 4.

2. 2. 2. 12 Characteristic Function

The characteristic function is essentially a special case of the moment generating function wherein the variable of transformation is taken as a vector of imaginary components, i. e., $\underline{s} = \sqrt{-1}\,\underline{\omega}$ where $\underline{\omega}$ is a vector of real components ω_i for $i = 1, 2, \ldots, n$. It is noted that the probability density function and the characteristic function are, essentially, Fourier transform pairs, except for a reversal of sign in the variable of transformation. In general terms, the moment generating function and characteristic function are equivalent in statistical analyses. The characteristic function for a random variable also yields the moments for the random variable by taking appropriate partial derivatives of the characteristic

function with respect to ω and evaluating at $\omega = 0$; however, it should be noted that the imaginary factor $\sqrt{-1}$ appears in the results and the partial derivative must be divided by a factor of $\sqrt{-1}$ raised to the order of the moment.

It should be noted that in the literature both moment generating functions and characteristic functions are used separately, i. e., either the moment generating function or the characteristic function will be used depending on the particular source.

2. 2. 2. 13 Gaussian and Normal Random Variables

Let x be a random variable whose pdf is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(x-m)^2}{2\sigma_x^2}}$$

where m and σ_x^2 are the mean and variance of x respectively. The random variable x is referred to as a Gaussian random variable and f(x) is defined as the Gaussian pdf. The moment generating function for a Gaussian random variable is given by

$$mg\, f_x(s) = \int_{-\infty}^{\infty} e^{sx} f(x)\, dx$$

$$= e^{sm} \int_{-\infty}^{\infty} e^{s(x-m)} f(x)\, dx$$

$$= \frac{e^{sm}}{\sqrt{2\pi\sigma_x^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma_x^2}[x-m-\sigma_x^2 s]}\, dx$$

$$mg\, f_x(s) = e^{+[ms+\frac{1}{2}\sigma_x^2 s^2]}$$

By evaluating the first and second derivatives of $mfg_x(s)$ at $s = o$ it is found that

$$m_1 = m$$

$$m_2 = m^2 + \sigma_x^2$$

where m_1 and m_2 are the first and second moments of x. It follows that

$$\sigma_x^2 = m_2 - m^2$$

These results verify that the terms m and σ_x^2 in the Gaussian pdf are actually the mean and variance of x, respectively.

The moment generating function can be differentiated repeatedly to determine the higher order moments of a Gaussian random variable. The results are given below

$$m(r) = E(x^r)$$

$$= r! \sum_{k=0}^{K} m^{r-2k} \frac{\sigma_x^{2k}}{2^k k!(r-k)!}$$

where $K=r/2$ if r is even and $K = \frac{1}{2}(r-1)$ if r is odd. If $m = 0$, i.e., $E(x)=0$, then $m(r) = 0$ for odd values of r. In this case the even moments become

$$m(r = 2k) = E(x^{2k}) = (2k)! \frac{\sigma_x^{2k}}{2^k(k!)}$$

for $k = 1, 2, 3, \ldots$. Also, it is noted that $m(r = 2k)$ are the central moments for a Gaussian random variable since the first moment is zero.

Let y be related to the Gaussian random variable x is the following manner.

$$y = \frac{x - m}{\sigma_x}$$

It is easily seen that $E(y) = 0$ and $\sigma_y^2 = 1$. The random variable y is a Gaussian random variable with zero mean value and unity variance. Such a random variable will be referred to as a "Normal" random variable. The higher order moments of y are given by

$$E(y^{2k}) = \frac{(2k)!}{2^k (k!)}$$

It should be noted that there exists a lack of consistency in the literature concerning the definition of Gaussian and Normal random variables. Often x above is referred to as a Normal random variable and y is referred to as a "standard" or "normalized" Normal random variable. This terminology appears somewhat redundant and inefficient, thus, the present definitions are used; i. e., x and y, as defined above, are Gaussian and Normal random variables, respectively. In this manner the Normal random variable is a "normalized" or special Gaussian random variable. The present definitions appear to be more efficient.

A Gaussian random vector can be defined in the following manner. If the marginal pdf for each component x_i of a random vector $\underline{x}$ is Gaussian then the vector $\underline{x}$ is a Gaussian random vector, i. e., if

$$f(x_i) = \frac{1}{\sqrt{2\pi\ \sigma_i^2}}\ e^{-\frac{(x_i - m_i)^2}{2\sigma_i^2}}$$

for $i = 1, 2, \ldots, n$, then the random vector $\underline{x}$ is a Gaussian random vector, where $m_i = E(x_i)$ and σ_i^2 is the variance of x_i. The definition of a Gaussian random vector refers only to the marginal pdf of each component. The joint pdf of a Gaussian random vector is given by

$$f(\underline{x}) = \frac{1}{\sqrt{(2\pi)^n |\Gamma_x|}}\ e^{-\frac{1}{2}(\underline{x} - \underline{m})^T \Gamma_x^{-1} (\underline{x} - \underline{m})}$$

where Γ_x is the co-variance matrix for $\underline{x}$, $|\Gamma_x|$ is the determinant of Γ_x, $\underline{m} = E(\underline{x})$ and n is the number of components or the dimension of $\underline{x}$. The pdf for a Gaussian random vector is usually referred to as a "multivariate" Gaussian pdf. It is seen that the joint pdf of a Gaussian random

vector is a function of only the first moments and the second central moments of all of the components of $\underline{x}$, i. e., only the mean values and all co-variances of the components of $\underline{x}$ are required to specify a multivariate Gaussian pdf. In general, the components of a Gaussian random vector are correlated. However, it should be noted that if the components of a Gaussian random vector are uncorrelated then the components are statistically independent, i. e., for a Gaussian random vector, statistical independence and "zero-correlation" are equivalent. This is not true in general. The multivariate Gaussian pdf is discussed in further detail in Appendix B. Therein it is shown that the marginal and conditional pdfs of any subset of the components of $\underline{x}$ are also Gaussian pdfs.

It should be noted that the basic properties of the joint Gaussian are dependent upon the quadratic form of the co-variance matrix Γ_X. It is apparent that the pdf is a function of the quadratic form of Γ_X^{-1}; however, the properties of this quadratic form are closely related to that of Γ_X and, hence, the behavior of the Gaussian pdf can be considered in terms of the quadratic form of Γ_X and its relationship to that of Γ_X^{-1}. It should be noted that Γ_X and Γ_X^{-1} are real symmetrical positive definite matrices which possess the same set of eigenvectors and reciprocal eigenvalues, i. e., if $\Gamma_X \underline{\phi} = \lambda \underline{\phi}$ then $\Gamma_X^{-1} \underline{\phi} = \lambda^{-1} \underline{\phi}$. In general, the set of eigenvectors for a real symmetrical matrix forms an orthogonal basis for an n dimensional space, where n is the order of the matrix. Moreover, the eigenvectors can be normalized to form an orthonormal basis for the space. Let M be the matrix of normalized eigenvectors of Γ_X, i. e., $M = [\underline{\phi}_1, \underline{\phi}_2, \ldots, \underline{\phi}_n]$ where

$$\underline{\phi}_i^T \underline{\phi}_j = 1 \text{ for } i = j \quad \text{and} \quad \underline{\phi}_i^T \underline{\phi}_j = 0 \text{ for } i \neq j$$

Thus, $M^T M = I$, $M^{-1} = M^T$ and $M^T \Gamma_X M = \Lambda$

where Λ is a diagonal matrix of the eigenvalues λ_i of Γ_X i. e., $\Gamma_X \underline{\phi}_i = \lambda_i \underline{\phi}_i$ for $i = 1, 2, \ldots\ldots, n$. It should be apparent that $\Gamma_X M = [\lambda \underline{\phi}_1, \lambda 2 \underline{\phi}_2, \ldots \ldots, \lambda_n \underline{\phi}_n]$ and, hence, $M^T \Gamma_X M = \Lambda$. The matrix M is usually referred to as a "modal" matrix. The modal matrix M for Γ_X is also an orthogonal matrix which represents a rotation of coordinates for which scalar products are invariant.

The modal matrix M and the matrix Λ of eigenvalues for Γ_X essentially characterize the behavior of the joint Gaussian pdf. In general, the set of points in n dimensional space for which a positive definite quadratic form is constant describes an n dimensional surface which is defined to be a "hyper-ellipsoid," or an ellipse and ellipsoid for $n = 2$ and 3. respectively. Thus, the joint Gaussian pdf for $\underline{x}$ is constant along some hyper-ellipsoidal surface in n dimensional space. The transformation $(\underline{x} - \underline{m}) = M \underline{z}$ essentially determines the hyper-ellipsoid

of constant probability density for $\underline{x}$. It is apparent that the hyper-ellipsoid is centered at $E(\underline{x}) = \underline{m}$ and has principal axes which coincide with the eigenvectors of Γ_x, since $\underline{Z}^T \Lambda^{-1} \underline{Z}$ is in diagonal form. It is generally possible to determine the probability that the random vector $\underline{x}$ will lie within a hyper-ellipsoid of constant probability density. This is discussed in further detail in Appendix C.

By definition, a Normal random vector is a random vector with statistically independent or uncorrelated Normal components, i.e., $\underline{y}$ is a Normal random vector if

$$f(y_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} y_i^2}$$

for i = 1, 2, ..., n and if

$$f(\underline{y}) = \prod_{i=1}^{n} f(y_i) = \frac{1}{\sqrt{2\pi}^n} e^{-\frac{1}{2} \underline{y}^T \underline{y}}$$

2.2.3 Several Particular Probability Density and Distribution Functions

There exist several probability density functions which are often used in statistical analyses. The Gaussian pdf defined above is perhaps the most often encountered pdf; however, the following ones are also encountered frequently.

2.2.3.1 Uniform Probability Density Function

The uniform pdf is constant over some interval of x and zero elsewhere, i.e.,

$$f(x) = \frac{1}{\beta - \alpha} \qquad \alpha < x < \beta$$

$$= 0 \quad \text{elsewhere}$$

The first moment, or expected value, of x is

$$E(X) = m_1 = \int_{-\infty}^{+\infty} X f(X)\, dX$$

$$= \frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} X\, dX$$

$$= \frac{\beta^2 - \alpha^2}{2(\beta - \alpha)}$$

$$E(X) = m_1 = 1/2\,(\beta + \alpha)$$

The second moment of x becomes

$$E(X^2) = \int_{-\infty}^{+\infty} X^2 f(X)\, dX$$

$$= \frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} X^2\, dX$$

$$= \frac{\beta^3 - \alpha^3}{3(\beta - \alpha)}$$

$$E(X^2) = 1/3\,(\beta^2 + \beta + \alpha^2)$$

Thus, the variance of x becomes

$$\sigma_x^2 = E(X^2) - E(X^2)$$

$$= \tfrac{1}{3}(\beta^2+\alpha\beta+\alpha^2) - \tfrac{1}{4}(\beta^2+2\alpha\beta+\alpha^2)$$

$$= \tfrac{1}{12}\left[4(\beta^2+\alpha\beta+\alpha^2) - 3(\beta^2+2\alpha\beta+\alpha^2)\right]$$

$$= \tfrac{1}{12}\left[\beta^2 - 2\alpha\beta + \alpha^2\right]$$

$$\sigma_x^2 = \frac{1}{12}(\beta-\alpha)^2$$

It is easy to determine the PDF for x since

$$P[x \leq z] = \int_{-\infty}^{z} f(x)\,dx = F(z)$$

$$= \int_{\alpha}^{z} f(x)\,dx$$

Thus,

$$F(z) = 0 \qquad \text{for} \quad z \leq \alpha$$

$$= \frac{z-\alpha}{\beta-\alpha} \qquad \text{for} \quad \alpha < z < \beta$$

$$= 1 \qquad \text{for} \quad z \geq \beta$$

It is noted that both f(x) and F(z) can be written in convenient form using the unit step function U(w) defined as follows.

$$U(W) = 0 \quad \text{for} \quad W < 0$$
$$= 1 \quad \text{for} \quad W \geq 0$$

Thus,

$$f(x) = \frac{1}{\beta - \alpha} U(x - \alpha)\, U(\beta - x)$$

$$F(x) = \frac{(x-\alpha)}{\beta - \alpha} U(x-\alpha)\, U(\beta - x) + U(x-\beta)$$

2.2.3.2 Gamma Probability Density Function

The Gamma pdf is defined in terms of two parameters, α and β, and is usually denoted by $f(x\ ;\ \alpha\ ,\ \beta\ )$. The definition is

$$F(x;\ \alpha, \beta) = \frac{x^{\alpha} e^{-x/\beta}}{\alpha!\ \beta^{\alpha+1}} U(x)$$

where $U(x)$ is the unit step function and $\beta > 0$ and $\alpha > -1$. The moment generating function for x is given by $m(s) = (1 - \beta s)^{-(\alpha + 1)}$ where $s < 1/\beta$. By differentiating $m(s)$ appropriately and setting $s = o$ the following moments are found.

$$E(X) = m_1 = \beta(\alpha+1)$$

$$E(X^2) = m_2 = \beta^2(\alpha+1)(\alpha+2)$$

The variance for x is determined from

$$E(X - m_1)^2 = \sigma^2 = m_2 - m_1^2$$

Thus,

$$\sigma^2 = \beta^2(\alpha+1)$$

The PDF for x is given as follows:

$$P(x \le z) = \int_{-\infty}^{z} f(x;\alpha,\beta)\,dx = F(z)$$

$$= U(z)\int_{0}^{z} \frac{x^{\alpha} e^{-x/\beta}}{\alpha!\,\beta^{\alpha+1}}\,dx$$

$$= U(z)\int_{0}^{z} \frac{1}{\alpha!}\left(\frac{x}{\beta}\right)^{\alpha} e^{-x/\beta}\,\frac{dx}{\beta}$$

$$F(z) = U(z)\int_{0}^{z/\beta} \frac{1}{\alpha!}\,\tau^{\alpha} e^{-\tau}\,d\tau$$

Now, if α is a positive integer, then the PDF for x can be obtained in closed form. This is done by successively integrating the integral by parts as follows:

$$I(\alpha, z/\beta) = \frac{1}{\alpha!}\int \tau^{\alpha} e^{-\tau}\,d\tau$$

$$= -\frac{\tau^{\alpha}}{\alpha!} e^{-\tau}\Big|_{0}^{z/\beta} + \int_{0}^{z/\beta} \frac{1}{(\alpha-1)!}\,\tau^{\alpha-1} e^{-\tau}\,d\tau$$

$$= -\frac{1}{\alpha!}\left(\frac{z}{\beta}\right)^{\alpha} e^{-z/\beta} + I(\alpha-1, z/\beta)$$

$$I(\alpha, \frac{z}{\beta}) = 1 - \left[1 + \left(\frac{z}{\beta}\right) + \frac{1}{2!}\left(\frac{z}{\beta}\right)^2 + \frac{1}{3!}\left(\frac{z}{\beta}\right)^3 + \cdots + \frac{1}{\alpha!}\left(\frac{z}{\beta}\right)^{\alpha}\right] e^{-z/\beta}$$

Thus,

$$F(z) = \left[1 - e^{-z/\beta} \sum_{k=0}^{k=\alpha} \frac{1}{k!}\left(\frac{z}{\beta}\right)^{k}\right] U(z)$$

In particular, for $\alpha = 0$, 1 and 2

$$P[x \le z] = [1 - e^{-z/\beta}]\, U(z) \qquad \alpha = 0$$

$$P[x \le z] = [1 - (1 + z/\beta)\, e^{-z/\beta}]\, U(z) \qquad \alpha = 1$$

$$P[x \le z] = \left\{1 - [1 + z/\beta + \tfrac{1}{2}(z/\beta)^2]\, e^{-z/\beta}\right\} U(z) \qquad \alpha = 2$$

It should be noted that for non-integer values of α the term $\alpha!$ must be defined such that $F(Z \longrightarrow \infty)$ is unity, therefore,

$$\alpha! = \int_0^\infty \tau^{\gamma-1} e^{-\tau}\, d\tau$$

However, the integral is the Gamma function for argument $\gamma = \alpha + 1$ which is defined as follows.

$$\Gamma(\gamma) = \int_0^\infty \tau^{\gamma-1} e^{-\tau}\, d\tau$$

Thus, in general,

$$\alpha! = \Gamma(\alpha + 1)$$

The following properties of $\Gamma(\lambda)$ are easily determined. (See Reference 1.)

$$\Gamma(\gamma) = (\gamma-1)\,\Gamma(\gamma-1) \qquad \gamma > 1$$

$$2\,\frac{\Gamma(2\gamma)}{\Gamma(\gamma)} = \frac{2^{2\gamma}}{\sqrt{\pi}}\,\Gamma(\gamma+\tfrac{1}{2}) \qquad \gamma > 0$$

$$\Gamma(\gamma+1) = \gamma! \quad \text{for } \gamma = + \text{ integer}$$

$$\Gamma(1) = 1$$

$$\Gamma(\tfrac{1}{2}) = \sqrt{\pi}$$

If α is not an integer then α can be written as $n+\delta$ where n is an integer and $0 \leq \delta < 1$. In this manner the integral $I(\alpha, Z/\beta)$ can be reduced as follows

$$I(\alpha, Z/\beta) = -e^{-Z/\beta} \sum_{k=0}^{n} \left[\frac{1}{\Gamma(k+\delta+1)} \left(\frac{Z}{\beta}\right)^{k+\delta} \right] + I(\delta-1, Z/\beta)$$

where

$$I(\delta-1, Z/\beta) = \int_0^{Z/\beta} \frac{1}{\Gamma(\delta)}\, \tau^{\delta-1} e^{-\tau} d\tau$$

A particular case of interest is that of $\delta = \frac{1}{2}$ for which $I(\delta - 1, Z/\beta) = I(-\frac{1}{2}, Z/\beta)$. In this case

$$I(-\tfrac{1}{2}, Z/\beta) = \int_0^{Z/\beta} \frac{1}{\Gamma(\frac{1}{2})}\, \tau^{-\frac{1}{2}} e^{-\tau} d\tau$$

By the change of variable $\tau = \frac{1}{2} u^2$ it is found that

$$I(-\tfrac{1}{2}, Z/\beta) = 2\int_0^{\ell} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\, du$$

where $\ell = +\sqrt{2\ Z/\beta}$. It is noted that since the integral is the pdf for a Normal random variable, $I\ (-\frac{1}{2},\ Z/\beta\ )$ is the probability that a Normal random variable will lie between $\pm\ \ell$, i. e.,

$$I(-\tfrac{1}{2}, Z/\beta) = P[|y| \leq \ell]$$

where y is a Normal random variable. The determination of $P\ [\ |y| \leq \ell\ ]$ is discussed in Appendix C.

2. 2. 3. 3 Beta Probability Density Function

The Beta pdf is defined in terms of two parameters, α and β , and is denoted by $f(x\ ;\ \alpha\ ,\ \beta\ )$. The definition is

$$f(x;\alpha,\beta) = \frac{(\alpha+\beta+1)!}{\alpha!\ \beta!} x^{\alpha}(1-x)^{\beta} U(x)\, U(1-x)$$

where $-1 < \alpha\ ,\ \beta$. It is noted that if $\alpha = \beta = 0$, the Beta pdf is the Uniform pdf over the interval $0 \leq x \leq 1$. It is possible to determine the rth moment, m_r , in terms of α and β , i. e.,

$$m_r = E(X^r) = \frac{(\alpha+\beta+1)!}{\alpha!\ \beta!} \int_0^1 x^{r+\alpha}\ (1-x)^{\beta}\, dx$$

$$= \frac{(\alpha+\beta+1)!\ (\alpha+r)!}{(\alpha+\beta+r+1)!\ \alpha!} \int_0^1 \frac{(\alpha+\beta+r+1)!}{(\alpha+r)!\ \beta!} x^{r+\alpha}\ (1-x)^{\beta}\, dx$$

$$m_r = \frac{(\alpha+\beta+1)!\,(\alpha+r)!}{(\alpha+\beta+r+1)!\,\alpha!}$$

Thus,

$$E(X) = m_1 = \frac{(\alpha+\beta+1)!\,(\alpha+1)!}{(\alpha+\beta+2)!\;\alpha!}$$

$$E(X) = \frac{(\alpha+1)}{(\alpha+\beta+2)}$$

$$E(X^2) = m_2 = \frac{(\alpha+1)(\alpha+2)}{(\alpha+\beta+2)(\alpha+\beta+3)}$$

The variance becomes

$$\sigma_x^2 = m_2 - m_1^2$$

$$= \frac{\alpha+1}{\alpha+\beta+2}\left[\frac{\alpha+2}{\alpha+\beta+3} - \frac{\alpha+1}{\alpha+\beta+2}\right]$$

$$\sigma_x^2 = \frac{(\alpha+1)\quad(\beta+1)}{(\alpha+\beta+2)^2(\alpha+\beta+3)}$$

2.2.4 Functions of Random Variables

In systems performance analysis, the general statistical problem can be adequately described by the following equation.

$$\underline{y} = \underline{F}(\underline{x})$$

where $\underline{F}(\ )$ is a "non-random" function , $\underline{x}$ is a random vector, and $\underline{y}$ is a random vector as a consequence of $\underline{x}$. Two general problems evolve in order to specify the statistical behavior of $\underline{y}$. First, the statistical behavior of $\underline{x}$ must be specified, and second, the behavior of $\underline{y}$ must be determined as a function of that of $\underline{x}$.

In general, either the probability density function of $\underline{y}$ or a sufficient set of moments of $\underline{y}$ is required. This requirement can be considered as a transformation of probability density functions or the expectation of functions of random variables.

2.2.4.1 Transformations of Probability Density Functions

Consider the case wherein $\underline{y} = \underline{F}(\underline{x})$ possesses a single real-valued inverse transformation $\underline{x} = \underline{F}^{-1}(\underline{y}) = \underline{G}(\underline{y})$. It is tacitly assumed that $\underline{y}$ is of the same dimension as $\underline{x}$. In this case, the pdf of $\underline{y}$ can be obtained in a manner similar to that of transforming variables in multiple integrals. The general result is simply

$$f(\underline{y}) = f\left[\underline{x} = \underline{G}(\underline{y})\right] \left|J(G)\right|$$

where $J(\underline{G})$ is the absolute value of the Jacobian of $\underline{G}(\underline{y})$. The Jacobian of $\underline{G}(\underline{y})$ is simply the determinant of the matrix of partial derivatives of $\underline{G}(\underline{y})$ with respect to the components of $\underline{y}$, i.e.,

$$J(G) = D_{ET}\left[\frac{\partial G_i}{\partial y_j}\right]$$

If the inverse transformation is multiple-valued then the pdf of $\underline{y}$ is given by

$$f(y) = \sum_{i=1}^{k} f\,[x = G_i(y)]\,|J_i(G)|$$

where $\underline{G}_i(\underline{y})$ is the ith solution for the inverse which has a total number of k solutions, and $J_i(G)$ is the Jacobian for the ith solution. For real random variables only real solutions of the inverse transformation are included. The pdf of y can also be written in terms of the Jacobian of $\underline{F}(\underline{x})$ as follows.

$$f(y) = \sum_{i=1}^{k} \frac{1}{|J_i(F)|}\, f\,[x = G_i(y)]$$

Thus, there exists a rather general method of essentially transforming the joint pdf of $\underline{x}$ into the joint pdf of $\underline{y}$. Usually the dimension of $\underline{y}$ is less than that of $\underline{x}$, however, the above method can still be used to determine the pdf of $\underline{y}$ by defining an augmented vector with $\underline{y}$ as a subvector such that the inverse transformation $\underline{x} = G\,(\underline{y}\,,\,\underline{y}_1)$ exists, and then determine the joint pdf of the augmented vector $(\underline{y}\,,\,\underline{y}_1)$. Now, the pdf of $\underline{y}$ is simply the marginal pdf which can be determined from $f(\underline{y}\,,\,\underline{y}_1)$. The procedure is as follows for a single-valued inverse.

$$\underline{y} = \underline{F}\,(\underline{x})$$

$$\underline{y}_1 = \underline{F}_1\,(\underline{x})$$

$$\underline{x} = \underline{G}(\underline{y}, \underline{y}_1)$$

$$f(\underline{y}, \underline{y}_1) = f\left[\underline{x} = \underline{G}(\underline{y}, \underline{y}_1)\right] \left| J(G) \right|$$

$$f(\underline{y}) = \int_{D(\underline{y}_1)} f(\underline{y}, \underline{y}_1)\, d\underline{y}_1$$

It is apparent that $\underline{F}_1(\underline{x})$ is not unique, but it should be selected on the basis of convenience in determining $\underline{x} = \underline{G}(\underline{y}, \underline{y}_1)$ and the marginal pdf of $\underline{y}$. Often it is most convenient to define the augmented components $\underline{y}_1$ as simply equal to x_i for $i = m+1, m+2, \ldots n$ where m and n are the dimensions of $\underline{y}$ and $\underline{x}$, respectively.

One of the fundamental properties of transformations of pdfs is that if $\underline{y}_1$ is a function of $\underline{x}_1$ and $\underline{y}_2$ is a function of $\underline{x}_2$ and if $\underline{x}_1$ and $\underline{x}_2$ are statistically independent random vectors then $\underline{Y}_1$ and $\underline{Y}_2$ are statistically independent random vectors. That is, if $\underline{Y}_1 = \underline{F}_1(\underline{x}_1)$ and $\underline{Y}_2 = F_2(\underline{x}_2)$ and $f(\underline{x}_1, \underline{x}_2) = f(\underline{x}_1) f(\underline{x}_2)$ then $f(\underline{Y}_1, \underline{Y}_2) = f(\underline{Y}_1) \cdot f(\underline{Y}_2)$. This can be shown in the following manner. Let $\underline{y}$ contain $\underline{y}_1$ and $\underline{y}_2$ as subvectors, i.e.,

$$\underline{y} = \begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \end{bmatrix} = \begin{bmatrix} \underline{F}_1(\underline{x}_1) \\ \underline{F}_2(\underline{x}_2) \end{bmatrix} = \underline{F}(\underline{x})$$

Now, the inverse relationship for $\underline{x}$ and $\underline{y}$ becomes

$$\underline{x} = \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \end{bmatrix} = \underline{F}^{-1}(\underline{y}) = \underline{G}(\underline{y}) = \begin{bmatrix} \underline{G}_1(\underline{y}_1) \\ \underline{G}_2(\underline{y}_2) \end{bmatrix}$$

The Jacobian for the relationship $\underline{y} = F(x)$ is simply the product of the Jacobians for the relationships $\underline{y}_1 = \underline{F}_1(\underline{x}_1)$ and $\underline{y}_2 = \underline{F}_2(\underline{x}_2)$, i.e.,

$$J(\underline{y}, \underline{x}) = J(\underline{y}_1, \underline{x}_1)\, J(\underline{y}_2, \underline{x}_2)$$

This follows from the fact that the matrix of partial derivatives between $\underline{y}$

and $\underline{x}$ can be partioned into two "non-null" matrices along the diagonal with all other terms zero. The determinant of such a matrix is simply the product of the determinants of the diagonal matrices. Thus,

$$f(y) = \frac{1}{|J(y, x)|} f[x = G(y)]$$

$$= \frac{f[x_1 = G_1(y_1)]}{|J(y_1, x_1)|} \cdot \frac{f[x_2 = G_2(y_2)]}{J(y_2, x_2)}$$

$$f(y) = f(y_1, y_2) = f(y_1) f(y_2)$$

The transformation of pdfs is discussed in further detail in References 5, 6, and 7.

2.2.4.2 Expectation of a Function of a Random Variable

Let $y = g(x)$ where x is a random variable and $g(\ )$ is a non-random function. Due to the dependence of y on x, y is also a random variable. The expectation of y, $E(y)$, is given by

$$E(y) = \int_{-\infty}^{\infty} y f(y) dy$$

where $f(y)$ is the pdf of y, which could be obtained from the pdf of x as indicated above. However, it is not necessary to obtain $f(y)$ if only $E(y)$ is required. The definition of expectation applies to any function of the random variable x, i.e., if $y = g(x)$, then

$$E[g(x)] = \int_{-\infty}^{-\infty} g(x) f x dx$$

Thus, if the expectation of a function $g(x)$ of a random variable x is required, it is generally not necessary to determine the pdf of $g(x)$ to obtain the expectation of $g(x)$. In the general case if $\underline{y} = \underline{g}(x)$ then

$$E(\underline{y}) = \int_{D(\underline{X})} g(\underline{X})\, f(\underline{X})\, d\underline{X}$$

It should be noted that there exists two definitions for the expectation of $\underline{y} = \underline{g}(\underline{x})$. However, the definitions are consistent since, in general, the transformation of probability density functions yields equivalent expectations, i. e. ,

$$f(\underline{y})\, d\underline{y} = f(\underline{X})\, d\underline{X}$$

This applies for any moment or expectation of y since if $y = g(x)$, then $y^r = g^r(x) = h(x)$, etc.

2. 2. 4. 3 Use of the Moment Generating Function

It is often convenient to use the moment generating function or characteristic function to determine the probability density function and/ or moments of a function of a random variable. That is, if y is a function of the random variable x, $y = F(x)$, then the moment generating function of y can be expressed in terms of the pdf of x as follows.

$$m_{gf\,y}(s) = E[e^{sy}] = \int_{-\infty}^{\infty} e^{sy} f(y)\, dy$$

$$= E[e^{sF(x)}]$$

$$mgf_y(s) = \int_{-\infty}^{\infty} e^{sF(x)} f(x)\,dx$$

The moments of y can be determined as discussed in Section 2.2.2.11. In order to determine the pdf of y it is essentially necessary to determine a probability density function which has a moment generating function corresponding to the one found for y. Usually this is accomplished by simply recognizing that the form of the moment generating function of y corresponds to one for which the probability density function is known. This is equivalent to employing moment generating functions and probability density functions as transform pairs. In case the corresponding probability density function cannot be recognized then by letting $s = +\sqrt{-1}\omega$ the characteristic function can be obtained which can be "inverted" by Fourier transform methods. Also, for positive definite random variables the Laplace transform can be used. The theory and methods of the Fourier and Laplace transforms are discussed in detail in References 8, 9, and 10.

2.2.4.4 Sums of Independent Random Variables

Consider the particular case for which y is a linear sum of a set of statistically independent random variables, i.e.,

$$y = \sum_{i=1}^{n} x_i = \underline{1}^T \underline{x}$$

where $\underline{x}$ is a statistically independent random vector as defined in Section 2.2.2.5. The moment generating function for y is given by

$$mgf_y(s) = E[e^{sy}] = E[e^{s\underline{1}^T\underline{x}}]$$

$$= \int_{D(\underline{x})} e^{s\underline{1}^T\underline{x}} f(\underline{x})\,d\underline{x}$$

$$= \prod_{i=1}^{n}\left[\int_{-\infty}^{\infty} e^{s x_i} f(x_i)\,dx_i\right]$$

$$= \prod_{i=1}^{n} E(e^{sx_i})$$

$$mgf_y(s) = \prod_{i=1}^{n} mgf_{x_i}(s)$$

Now, by setting $s = +\sqrt{-1}\,\omega$ it is seen that the characteristic function for y is the product of the characteristic functions for the random variables x_i . Therefore, by using the convolution theorem of Fourier transforms it is found that the pdf of y is the convolution of the pdfs of the x_i , i.e.,

$$f(y) = f(x_1) * f(x_2) * \cdots * f(x_n)$$

where * denotes the convolution operation.

2.2.4.5 Functions of Gaussian Random Variables

In system performance analyses Gaussian random variables are often encountered, and there exist several particular functions of a set of Gaussian random variables which arise frequently in statistical analyses of random processes, especially in the problem of estimating the statistical moments of a Gaussian probability density function from a set of samples. Generally, the probability density functions of these particular functions are required. In this section the probability density functions of several particular functions of Gaussian random are discussed which arise in system performance analyses.

2.2.4.5.1 Linear Functions

Linear functions of Gaussian random vectors are often encountered in statistical analyses and there exist several fundamental properties of these functions which are of direct usefulness. Let a random vector $\underline{y}$ be defined as follows.

$$\underline{y} = A\underline{x} + \underline{c}$$

where $\underline{x}$ is a Gaussian random vector, A is a constant matrix, and $\underline{c}$ is a constant vector. In this manner $\underline{y}$ is a linear function of $\underline{x}$ and, in general, the dimension of $\underline{y}$, m, and that of $\underline{x}$, n, can be different. A fundamental property of a linear function of a Gaussian random vector is that the resulting random vector is also a Gaussian random vector, i.e., the property of Gaussianness is invariant under a linear transformation. Moreover, the statistical moments of $\underline{y} = A(\underline{x} + \underline{c})$ are readily expressed in terms of those of $\underline{x}$, especially the covariance matrix and expectation of $\underline{y}$ which specify the pdf of $\underline{y}$. This is easily established as follows.

Using property III of Section 2.2.2.6, it is found that

$$E(\underline{y}) = E(A\underline{x}) + E(\underline{c})$$
$$= AE(\underline{x}) + \underline{c}$$

$$E(\underline{y}) = A\,\underline{m}_x + \underline{c}$$

where $\underline{m}_x = E(\underline{x})$. Also using the results of Section 2.2.2.8 it is found that

$$\Gamma_y = A\,\Gamma_x\,A^T$$

where Γ_x and Γ_y are the covariance matrices of $\underline{x}$ and $\underline{y}$, respectively. Thus, both the expectation and covariance matrix of $\underline{y}$ are determined directly in terms of those of $\underline{x}$ and the elements of the linear function A and $\underline{c}$.

The pdf of $\underline{y}$ is easily determined by use of the moment generating function. In Appendix B it is shown that the moment generating function for a Gaussian random vector $\underline{z}$ is as follows.

$$mgf_{\underline{z}}(\underline{s}) = \exp\left[\underline{s}^T \underline{m}_z + 1/2\, \underline{s}^T \Gamma_z \underline{s}\right]$$

The pdf of $\underline{z}$ is given by

$$f(\underline{z}) = \frac{1}{\sqrt{(2\pi)^n |\Gamma_z|}} \exp -1/2 (\underline{z} - \underline{m}_z)^T \Gamma_z^{-1} (\underline{z} - \underline{m}_z)$$

where $\underline{m}_z = E(\underline{z})$. Now, the moment generating function for $\underline{y}$ is determined by

$$mgf_y(\underline{s}) = E(e^{\underline{s}^T \underline{y}}) = E\left[e^{\underline{s}^T (A\underline{x} + \underline{c})}\right]$$

$$= e^{\underline{s}^T \underline{c}} E\left[e^{\underline{s}^T A \underline{x}}\right]$$

$$= e^{\underline{s}^T \underline{c}} \int_{D(\underline{x})} e^{\underline{s}^T A \underline{x}} f(\underline{x})\, d\underline{x}$$

$$= \frac{e^{\underline{s}^T \underline{c}}}{\sqrt{(2\pi)^n |\Gamma_x|}} \int_{D(\underline{x})} \exp\left[\underline{s}^T A \underline{x} - \frac{1}{2} (\underline{x} - \underline{m})^T \Gamma_x^{-1} (\underline{x} - \underline{m})\right] d\underline{x}$$

$$mgf_y(\underline{s}) = \frac{e^{\underline{s}^T (A\underline{m} + \underline{c})}}{\sqrt{(2\pi)^n |\Gamma_x|}} \int_{D(\underline{x})} \exp\left[\underline{s}^T A (\underline{x} - \underline{m}) - \frac{1}{2} (\underline{x} - \underline{m})^T \Gamma_x^{-1} (\underline{x} - \underline{m})\right] d\underline{x}$$

The integral can be evaluated using integral $I_1(\underline{s})$ of Appendix A with an appropriate definition of variables. The results are as follows.

$$mgf_y(\underline{s}) = \exp\left[\underline{s}^T(A\underline{m}+\underline{c}) + 1/2\,\underline{s}^T A \Gamma_x A^T \underline{s}\right]$$

$$mgf_y(\underline{s}) = \exp\left[\underline{s}^T \underline{m}_y + 1/2\,\underline{s}^T \Gamma_y \underline{s}\right]$$

Therefore, if $\underline{x}$ is a Gaussian random vector and $\underline{y} = A\underline{x} + \underline{c}$, then is a Gaussian random vector with

$$E(\underline{y}) = \underline{m}_y = A\underline{m}_x + \underline{c}$$

$$\Gamma_y = A\Gamma_x A^T$$

Of course, the pdf of $\underline{y}$ is as follows.

$$f(\underline{y}) = \frac{1}{\sqrt{(2\pi)^m |\Gamma_y|}} \exp -1/2\,(\underline{y}-\underline{m}_y)^T \Gamma_y^{-1} (\underline{y}-\underline{m}_y)$$

where m is the dimension of $\underline{y}$.

There exist two particular linear functions of interest. First, consider a translation and a rotation of coordinates such that $(\underline{x} - \underline{m}_x) = M\underline{z}$ where M is the modal matrix for Γ_x as discussed in Section 2.2.2.13. The vector $\underline{z}$ is the set of coordinates of $(\underline{x} - \underline{m}_x)$ in the orthonormal basis determined by the eigenvectors of Γ_x. Of course, $\underline{z} = M^{-1}(\underline{z} - \underline{m}_x) = M^T(\underline{x} -$ and, hence, $E(\underline{z}) = 0$ and $\Gamma_z = M^T \Gamma_x M = \Lambda$. Also, $\Gamma_z^{-1} = \Lambda^{-1} = M^T$ and $|\Lambda| = |\Gamma_z| = |M|\,|\Gamma_z|\,|M| = |\Gamma_z|$. Now, the joint pdf of $\underline{z}$ is simply

$$f(\underline{z}) = \frac{1}{\sqrt{(2\pi)^n |\Gamma_z|}} \exp -1/2\,(\underline{z}^T M^T \Gamma_x^{-1} M\underline{z})$$

$$= \frac{1}{\sqrt{(2\pi)^n |\Lambda|}} \exp -1/2\,(\underline{z}^T \Lambda^{-1} \underline{z}) = \prod_{i=1}^{n} f(x_i)$$

where

$$f(z_i) = \frac{1}{\sqrt{2\pi\lambda_i}}\, e^{-\frac{z_i^2}{2\lambda_i}}$$

Thus, the components of $\underline{z}$, z_i are statistically independent Gaussian random variables with variance λ_i and zero mean. In general, a rotation of coordinates by M will transform a Gaussian random vector $\underline{x}$ into a statistically independent Gaussian random vector. Now, consider a further transformation of the random vector $\underline{z}$, i.e., let $\underline{y} = D\underline{z}$, then $E(\underline{y}) = \underline{0}$ and $\Gamma_y = D\Lambda D^T$. Thus, if D is a diagonal matrix with elements equal to the reciprocal of the square root of the eigenvalues of Γ_x, then $D\Lambda D^T = I$ Therefore, if $\underline{x}$ is a Gaussian random vector, then $\underline{y} = DM^T(\underline{x} - \underline{m}_x) = A(\underline{x} - \underline{m}_x)$ is a Normal random vector, i.e., $E(\underline{y}) = \underline{0}$ and $\Gamma_y = I$. Thus, it is found that

a Normal random vector can be obtained from a Gaussian random vector by a translation and a linear transformation, i.e., if $\underline{Y} = A(\underline{x} - \underline{m}_x)$ where $A = DM^T$, then $\underline{Y}$ is a Normal random vector if $\underline{x}$ is a Gaussian random vector.

2.2.4.5.2 Chi Square, χ^2

Consider a statistically independent Gaussian random vector $\underline{x}$. Let $\underline{Y}$ be a Normal random vector defined by

$$y_i = \frac{x_i - m_i}{\sigma_i} \quad \text{for } i = 1, 2, \ldots, n$$

Clearly, each component of $\underline{Y}$ has unity variance and zero mean. Now, consider the "length" or modulus of the random vector $\underline{Y}$, χ, defined as follows:

$$\chi = \left| y^T y \right| = +\sqrt{y^T y}$$

The quantity χ^2 becomes

$$\chi^2 = y^T y = \sum_{i=1}^{n} y_i^2$$

It is apparent that χ^2 is the square of the length or modulus of the random vector $\underline{Y}$, which is referred to as "Chi-square."

The moment generating function for χ^2 is given by

$$mgf_{\chi^2}(s) = E\left[e^{\chi^2 s}\right]$$

$$= E\left[e^{s(y^T y)}\right]$$

$$mgf_{\chi^2}(s) = \int_{D(y)} e^{s(y^T y)} f(y)\, dy$$

However, since $\underline{Y}$ is a Normal random vector,

$$f(y) = \frac{1}{n\sqrt{2\pi}} e^{-1/2(y^T y)}$$

Thus,

$$mgf_{x^2}(s) = \eta \frac{1}{\sqrt{2\pi}} \int_{D(Y)} \exp\left[(s - 1/2)\, y^T y\right] dy$$

$$= \frac{1}{\sqrt{2\pi}} \int_{D(Y)} \exp - 1/2 \left[(1 - 2s)\, y^T y\right] dy$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^{\infty} \exp - 1/2 (1 - 2s)\, y^2 \right] dy$$

$$= \prod_{i=1}^{n} \left[\frac{\sqrt{(1-2s)}}{\sqrt{(1-2s)}\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2(1-2s)^{-1}}\, dy\right]$$

$$mgf_{x^2}(s) = \frac{1}{(1-2s)^{n/2}} \qquad s < 1/2$$

It is seen that $mgf_{\chi^2}(s)$ is the same as that for the Gamma pdf of Section 2.2.3.2 with $\alpha = n/2-1$ and $\beta = 2$; therefore, the pdf for χ^2 is the following pdf which has the parameter n.

$$f(x^2; n) = \frac{(x^2)^{n/2-1}}{\Gamma(n/2)\, n\sqrt{2}}\, e^{-1/2\, x^2}\, U(x^2) \quad .$$

The pdf $f(\chi^2; n)$ is referred to as the "Chi-square" pdf with "n "degrees of freedom." The first and second moments for χ^2 can be determined directly from those given in Section 2.2.3.2 for the Gamma random variable, i.e., let $\alpha = n/2-1$ and $\beta = 2$ for the moments given in Section 2.2.3.2. Thus, it is found that

$$E(x^2) = n$$

$$E[(x^2)^2] = n(n+2)$$

$$\sigma^2_{x^2} = 2n$$

As shown in Section 2.2.3.2, the pdf for the Gamma pdf can be obtained in closed-form solution for α equal to a positive integer. Therefore, a closed-form solution can be obtained for the PDF of χ^2 for an even number of degrees of freedom, n, i.e., since $\alpha = n/2-1$, $n = 2(\alpha + 1)$, hence, n is even for α any positive integer. Thus, for even n,

$$P[X^2 \le z] = 1 - \left[\sum_{i=1}^{n/2-1} \frac{1}{i!}\left(\frac{z}{2}\right)^i\right] e^{-z/2}$$

where $0 \le z$. Of course, $P[\chi^2 \le 0] = 0$. Some particular cases are given below.

$$P\left[X^2 \le z\right] = 1 - e^{-z/2} \; ; \; \alpha = 0, \; n = 2$$

$$P\left[X^2 \le z\right] = 1 - (1 + z/2)\, e^{-z/2} \; ; \; \alpha = 1, \; n = 4$$

$$P\left[X^2 \le z\right] = 1 - \left[1 + z/2 + 1/2\,(z/2)^2\right] e^{-z/2} \; ; \; \alpha = 2, \; n = 6$$

If n is an odd integer then α is not an integer since $\alpha = n/2-1$, however, α can be written as $k+\frac{1}{2}$ where $k = \frac{1}{2}(n-3)$. Hence, k is an integer for odd n and the results given in Section 2.2.3.2 can be used for odd n . Thus, it is found that

$$P[\chi^2 \le z] = \sqrt{2/\pi} \int_0^{\ell} e^{-1/2\,u}\, du - e^{-z/2} \sum_{K=0}^{1/2(n-3)} \frac{1}{\Gamma(K+3/2)} \left(\frac{z}{2}\right)^{K+1/2}$$

for $n \ge 3$ and $z \ge 0$ and where $\ell = +\sqrt{Z}$. Of course, $P[\chi^2 \le z] = 0$ for $z < 0$. For the special case of $n = 3$ it is found that

$$P[\chi^2 \le z] = \sqrt{2/\pi} \int_0^{\ell} e^{-1/2\,u}\, du - \frac{1}{\Gamma(3/2)} \left(\frac{z}{2}\right)^{1/2} e^{-z/2}$$

$$= \sqrt{2/\pi} \int_0^{\ell} e^{-1/2\,u}\, du - \sqrt{2/\pi}\,\sqrt{z}\, e^{-z/2}$$

$$P[\chi^2 \le z] = 2\left[\int_0^{\ell} \frac{1}{\sqrt{2\pi}}\, e^{-1/2\,u}\, du - \frac{\sqrt{z}}{\sqrt{2\pi}}\, e^{-z/2}\right]$$

for $z \le 0$ and zero otherwise. It is noted that this result is the same as that obtained in Appendix C for a three-dimensional Gaussian random vector. It is also noted that $P[\chi^2 \le z] = P[(\chi \le +\sqrt{z}]$, or $P[(\chi \le \ell] = P[\chi^2 \le \ell^2]$; therefore, the PDF for χ , rather than χ^2 , is easily determined as follows.

$$P[\chi \le \ell] = 1 - \left[\sum_{k=0}^{n/2-1} \frac{1}{k!}\left(\frac{\ell^2}{2}\right)^k\right] e^{-\frac{\ell^2}{2}}$$

for n even and for n odd

$$P[\chi \le \ell] = \sqrt{2/\pi}\int_0^{\ell} e^{-1/2u}\,du - e^{-\ell^2/2}\sum_{k=0}^{1/2(n-3)} \frac{1}{\Gamma(k+3/2)}\left(\frac{\ell^2}{2}\right)^{k+1/2}$$

for $\ell \ge 0$, and for $\ell < 0$, $P[\chi \le \ell] = 0$.

A basic property of the Chi-square random variable is that the sum of Chi-square random variables also has a Chi-square pdf. That is, let $\underline{y}$ be a Normal random vector which contains two Normal random subvectors $\underline{y}_1$, and $\underline{y}_2$ of dimensions n_1 and n_2, respectively. In this manner

$$\chi^2 = \underline{y}^T\underline{y} = \underline{y}_1^T\underline{y}_2 + \underline{y}_2^T\underline{y}_2$$

$$\chi^2 = \chi_1^2 + \chi_2^2$$

where $\chi_1^2 = \underline{y}_1^T\underline{y}$ and $\chi_2^2 = \underline{y}_2^T\underline{y}_2$. The random variables χ_1^2 and χ_2^2 are statistically independent, hence, the moment generating function for their sum is the product of their moment generating functions. Also, the pdfs of χ_1^2 and χ_2^2 are Chi-square with n_1 and n_2 degrees of freedom, respectively; hence,

$$mgf_{(\chi_1^2+\chi_2^2)}(s) = mgf_{\chi_1^2}(s) \cdot mgf_{\chi_2^2}(s)$$

$$= \frac{1}{(1-2s)^{n_1/2}} \cdot \frac{1}{(1-2s)^{n_2/2}}$$

$$= \frac{1}{(1-2s)^{(n_1+n_2)/2}} = \frac{1}{(1-2s)^{n/2}}$$

where $n = n_1 + n_2$. Of course, the resulting moment generating function is that for χ^2 with n degrees of freedom. In general, any sum of Chi-square random variables has a Chi-square pdf with degrees of freedom equal to the sum of the degrees of freedom of each term in the sum, i.e., if

$$\chi^2 = \sum_{i=1}^{k} x_i^2 = \sum_{i=1}^{k} y_i^T y_i \ .$$

where each χ_i^2 is a Chi-square random variable with n_i degrees of freedom then χ^2 is a Chi-square random variable with $n = \sum_{i=1}^{k} n_i$ degrees of freedom. However, it is important to note that each $\chi_i^2 = \underline{y}_i^T \underline{y}_i$ must be statistically independent.

The foregoing can be used to determine the pdf and PDF for the modulus of a statistically independent Gaussian random vector with components of equal variances and zero mean values. Let $\underline{x}$ be a Gaussian random vector with joint pdf as follows.

$$f(\underline{x}) = \frac{1}{n\sqrt{2\pi\sigma^2}} \, e^{-\frac{\underline{x}^T\underline{x}}{2\sigma^2}}$$

The modulus of the random vector $\underline{x}$ is $+\sqrt{\underline{x}^T\underline{x}}$. Also, it is seen that

$$\underline{x}^T\underline{x} = v^2 = \sigma^2\chi^2$$

The pdf of v^2 can be obtained from the pdf of χ^2 by a simple transformation of pdfs. That is, since

$$\chi^2 = (1/\sigma^2)v^2 \qquad \frac{\partial\chi^2}{\partial v^2} = 1/\sigma^2$$

it follows that

$$f(v^2) = \frac{\left(\frac{v^2}{\sigma^2}\right)^{\frac{n}{2}-1}}{\sigma^2\,\Gamma(n/2)\,n\sqrt{2}} \, e^{-1/2(v^2/\sigma^2)} \, U(v^2)$$

$$f(v^2) = \frac{(v^2)^{\frac{n-2}{2}}}{\sigma^n\,\Gamma(n/2)\,n\sqrt{2}} \, e^{-1/2(v^2/\sigma^2)} \, U(v^2)$$

Using the relationship $v = +\sqrt{\underline{x}^T\underline{x}} = +\sqrt{v^2}$ the pdf of v is easily determined to be as follows.

$$f(v) = 2\,\frac{v^{n-1}}{\sigma^n\,\Gamma(n/2)\,n\sqrt{2}} \, e^{-1/2(v^2/\sigma^2)} \, U(v)$$

The PDFs for v^2 and v can be determined from those for χ^2 and χ given above. That is,

$$P[v^2 \le z] = P[x^2 \le z/\sigma^2]$$

$$P[v \le \ell] = P[x \le \ell/\sigma]$$

It is noted that for $\sigma^2 = 1$, $v = \chi$; therefore, the pdf of χ becomes

$$f(x) = 2\frac{x^{n-1}}{\Gamma(n/2)\,n\sqrt{2}}\, e^{-1/2\,x^2}\, U(x)$$

Consider the sum of χ_1 and χ_2 where χ_1^2 and χ_2^2 are Chi-square random variables with n_1 and n_2 degrees of freedom, respectively, as before. That is, let $u = \chi_1 + \chi_2$ where $\chi_1 = +\sqrt{y_1^T y_1}$ and $\chi_2 = +\sqrt{y_2^T y_2}$. Note that $u \neq +\sqrt{\chi^2} = +\sqrt{\chi_1^2 + \chi_2^2}$. The pdf of u is not as easy to determine as that for $\chi^2 = \chi_1^2 + \chi_2^2$. However, define v as χ_2 to obtain the following relationships.

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

Since χ_1 and χ_2 are statistically independent, the joint pdf of χ_1 and χ_2 is given by

$$f(x_1, x_2) = f(x_1) f(x_2)$$

$$= \frac{4\, x_1^{n_1-1}\, x_2^{n_2-1}}{\Gamma(n_1/2)\,\Gamma(n_2/2)\,n\sqrt{2}}\, e^{-1/2\,x^2}\, U(x_1)\, U(x_2)$$

where $r = n_1 + n_2$ and $\chi^2 = \chi_1^2 + \chi_2^2$. The Jacobian of the transformation between u, v and χ_1, χ_2 is simply 1, hence, the joint pdf of u and v becomes $f(u, v) = f(\chi_1 = u - v)\, f(\chi_2 = v)$, thus

$$f(u, v) = \frac{4\,(u-v)^{n_1-1}\, v^{n_2-1}}{\Gamma(n_1/2)\,\Gamma(n_2/2)\,n\sqrt{2}}\, e^{-1/2\,[(u-v)^2 + v^2]}\, U(u-v)\, U(v)$$

Now, the pdf of $u = \chi_1 + \chi_2$ is the marginal pdf of u and v, i.e.,

$$f(u = \chi_1 + \chi_2) = \int_{-\infty}^{\infty} f(u, v)\, dv$$

$$= 4\int_{0}^{\infty} \frac{(u-v)^{n_1-1} v^{n_2-1}}{\Gamma(n_1/2)\Gamma(n_2/2)\sqrt{2}} \cdot e^{-1/2[(u-v)^2+v^2]} U(u-v)\, dv$$

$$f(u = \chi_1 + \chi_2) = 4\int_{u}^{\infty} \frac{(u-v)^{n_1-1} v^{n_2-1}}{\Gamma(n_1/2)\Gamma(n_2/2)\sqrt{2}} \cdot e^{-1/2(u^2-2uv+2v^2)} U(u)\, dv$$

The PDF for u is obtained by the following expression.

$$P[u \le z] = \int_{-\infty}^{z} f(u)\, du$$

$$= \int_{0}^{z}\int_{u}^{\infty} \frac{4(u-v)^{n_1-1} v^{n_2-1}}{\Gamma(n_1/2)\Gamma(n_2/2)\sqrt{2}} e^{-1/2(u^2-2uv+2v^2)}\, dv\, du$$

where $z \ge 0$ and $P[u < 0] = 0$.

2.2.4.5.3 Chi Square Ratio, Variance Ratio

Let $\underline{y}_1$ and $\underline{y}_2$ be two statistically independent Normal random vectors of dimensions m and n, respectively. Consider two random

variables u and v which are defined as follows:

$$u = \chi_1^2 = y_1^T y_1$$

$$v = \chi_2^2 = y_2^T y_2$$

It is apparent that u and v are Chi-square random variables with m and n degrees of freedom, respectively. Consider the ratio of u to v , i.e., let w be given by

$$w = \frac{u}{v} = \frac{\chi_1^2}{\chi_2^2}$$

The random variable w is the ratio of two Chi-square random variables, and the pdf of w can be obtained from the joint pdf of u and v . Since y_1 and y_2 are statistically independent the joint pdf of u and v is simply the product of the pdfs of χ_1^2 and χ_2^2 , hence,

$$f(u,v) = \frac{u^{1/2(m-2)} v^{1/2(n-2)}}{\Gamma(n/2)\Gamma(n/2)(\sqrt{2})^{m+n}} e^{-1/2(u+v)} U(u)U(v)$$

The joint pdf of w and v can be obtained by the transformation of variables w = u/v and v = v with an inverse transformation of u=wv and v=v with a Jacobian of v ; hence,

$$f(w,v) = \frac{v^{1/2(n-2)} (wv)^{1/2(m-2)}}{\Gamma(m/2)\Gamma(n/2)(\sqrt{2})^{m+n}} e^{-1/2(wv+v)} \cdot v\, U(wv) U(w)$$

$$f(w,v) = \frac{w^{1/2(m-2)} v^{1/2(m+n-2)}}{\Gamma(m/2)\Gamma(n/2)(\sqrt{2})^{m+n}} e^{-1/2(w+1)v} U(w)U(v)$$

Now f(w)is the marginal pdf of f (w, v) , i.e.,

$$f(w) = \left[\frac{w^{1/2(m-2)} U(w)}{\Gamma(m/2)\Gamma(n/2)(\sqrt{2})^{m+n}} \int_0^\infty v^{1/2(m+n-2)} e^{-1/2(w+1)v} dv\right]$$

$$f(w) = \frac{w^{1/2(m-2)}\, U(w)}{\Gamma(m/2)\Gamma(n/2)(\sqrt{2})^{m+n}} \int_0^\infty v^{\alpha-1} e^{-v/\beta}\, dv$$

where $\alpha = \frac{1}{2}(m+n)$ and $\beta = (2/w+1)$. The integral is seen to be of the form of the Gamma function discussed in Section 2.2.3.2; hence,

$$\int_0^\infty v^{\alpha-1} e^{-v/\beta}\, dv = (\alpha-1)!\,\beta^\alpha = \Gamma(\alpha)\beta^\alpha$$

Thus, the pdf of w becomes

$$f(w) = \frac{\Gamma\left(\frac{m+n}{2}\right) w^{1/2(m+2)}}{\Gamma(m/2)\Gamma(n/2)(\sqrt{2})^{m+n}} \left(\frac{2}{w+1}\right)^{1/2(m+n)} U(w)$$

$$= \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma(m/2)\Gamma(n/2)} \cdot \frac{w^{1/2(m-2)}}{(1+w)^{1/2(m+n)}}\, U(w)$$

The pdf of w can be generalized with respect to an arbitrary positive definite constant k, i.e., let r = kw where k > 0. The pdf of r is easily determined from the pdf of w since w=1/k r with a Jacobian of 1/k, hence,

$$f(r) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma(m/2)\Gamma(n/2)} \frac{k^{-m/2}\, r^{1/2(m-2)}}{(1+r/k)^{1/2(m+n)}}\, U(r)$$

where

$$r = k\,\frac{\chi_1^2}{\chi_2^2}$$

A special case of k is the ratio of n to m which is usually denoted as the random variable F, i.e.,

$$F = \frac{1/m\, \chi_1^2}{1/n\, \chi_2^2} = \frac{n\, \chi_1^2}{m\, \chi_2^2} = \frac{n}{m}\, w$$

The pdf of F is easily determined, i.e.,

$$f(F) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma(m/2)\,\Gamma(n/2)} \left(\frac{m}{n}\right)^{m/2} \frac{F^{1/2(m-2)}}{(1 + m/n\, F)^{1/2(m+n)}}\, U(F)$$

The random variable F is referred to as the "variance ratio" since u/m and v/n are of the form of sample variances which are often used to estimate variances. The PDF of F is used in statistical tests of equality of variances and it is tabulated extensively. Some useful tables are given in Reference 5, 11 and 12.

It should be noted that the PDF of F can be used to determine the PDF of r since they are related by a constant, i.e.,

$$r = k\, \frac{\chi_1^2}{\chi_2^2} = \frac{m}{n}\, k \left(\frac{n}{m}\right) \frac{\chi_1^2}{\chi_2^2}$$

$$r = k' F \qquad k' = m/n\, k$$

where $k' = m/n\, k$. Thus,

$$P[r \le z] = P[F \le 1/k'\, z]$$

Therefore, the PDF of F can be used to determine the PDF of the following ratio.

$$r = \frac{\underline{x}_1^T \underline{x}_1}{\underline{x}_2^T \underline{x}_2} = \frac{\sigma_1^2}{\sigma_2^2} \cdot \frac{\chi_1^2}{\chi_2^2} = \left(\frac{m}{n}\right)\left(\frac{\sigma_1^2}{\sigma_2^2}\right) F$$

where $\underline{x}_1$ and $\underline{x}_2$ are statistically independent Gaussion random variables with zero mean values and each component with equal variances of σ_1^2 and τ_2^2 for $\underline{x}_1$ and $\underline{x}_2$, respectively.

Consider a random variable q which is defined in terms of w as follows.

$$q = \frac{1}{1+w} = \frac{1}{1+\frac{\chi_1^2}{\chi_2^2}} = \frac{\chi_2^2}{\chi_1^2+\chi_2^2}$$

The pdf of q can be determined from the pdf of w with the inverse transformation $w = q^{-1} - 1$ with the Jacobian of $-q^{-2}$, hence,

$$f(q) = \frac{\Gamma(\frac{m+n}{2})}{\Gamma(m/2)\Gamma(n/2)} \frac{(q^{-1}-1)^{1/2(m-2)}}{q^{-1/2(m+n)} q^2} U(1/q - 1)$$

$$= \frac{\Gamma(\frac{m+n}{2})}{\Gamma(m/2)\Gamma(n/2)} (1-q)^{1/2(m-2)} q^{1/2(n-2)} U(q) U(1-q)$$

Thus, the random variable q has a Beta probability density function with parameters $\alpha = \frac{1}{2}$ n - 1 and $\beta = \frac{1}{2}$ m - 1 as defined in Section 2.2.3.3.

2.2.4.5.4 Student's,t

Let χ^2 be a Chi-square random variable and let y be a Normal random variable which is statistically independent of χ^2. The joint pdf of y and χ^2 becomes

$$f(y, \chi^2) = f(y) f(\chi^2)$$

Thus, defining $u = \chi^2$,

$$f(y, u) = \frac{1}{\sqrt{2\pi}} e^{-1/2 y^2} \frac{u^{(n/2-1)}}{\Gamma(n/2)\, n\sqrt{2}} e^{-1/2 u} U(u)$$

where n is the number of degrees of freedom for χ^2. Consider the random variable v which is defined as follows

$$v = \frac{y}{\chi} = \frac{y}{+\sqrt{u}}$$

The pdf of v can be determined from the joint pdf of y and u in the following manner. The joint pdf of v and u can be determined by a transformation of variables $v = y/\sqrt{u}$ and $u = u$ with an inverse transformation of $y = v\sqrt{u}$ and $u=u$ with Jacobian $\sqrt{u}$, hence,

$$f(v,u) = \frac{1}{\sqrt{2\pi}} e^{-1/2(v^2u)} \frac{u^{(n/2-1)}}{\Gamma(n/2)2^n\sqrt{2}} e^{-1/2u}\sqrt{u}\, U(u)$$

$$= \frac{(u/2)^{1/2(n-1)}}{\Gamma(n/2)2^n\sqrt{4\pi}} e^{-1/2(v^2+1)u}\, U(u)$$

Now, the pdf of v is the marginal pdf of v and u, hence,

$$f(v) = \int_{-\infty}^{\infty} f(v,u)\,du$$

$$= \frac{2}{\Gamma(n/2)\sqrt{4\pi}} \int_0^{\infty} (u/2)^{1/2(n-1)} e^{-1/2(v^2+1)u} \frac{du}{2}$$

$$= \frac{1}{\Gamma(n/2)\sqrt{\pi}} \int_0^{\infty} \tau^{1/2(n-1)} e^{-(v^2+1)\tau} d\tau$$

$$= \frac{1}{\Gamma(n/2)\sqrt{\pi}} \int_0^{\infty} \tau^{\alpha-1} e^{-\tau/\beta}\, d\tau$$

$$f(v) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma(n/2)\sqrt{\pi}} \left[1+v^2\right]^{-1/2(n+1)}$$

The pdf of $w = kv$ is easily determined, i.e.,

$$f(w) = \frac{1}{k\sqrt{\pi}} \cdot \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma(n/2)} \left[1 + \frac{w^2}{k^2}\right]^{-1/2(n+1)}$$

For the special case of $k = \sqrt{n}$ the random variable w is referred to as "Student's" t random variable which is defined as follows.

$$t = \frac{y}{\sqrt{\frac{1}{n}\chi^2}}$$

The pdf of t becomes

$$f(t) = \frac{1}{n\pi} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma(n/2)} \left[1 + \frac{t^2}{n}\right]^{-1/2(n+1)}$$

The PDF of t is used in statistical tests of estimating and it is tabulated extensively. Some useful tables are given in references 5, 6 and 12.

It should be noted that the PDF of t can be used to determine the PDF of w since they are related by a constant, i.e.,

$$w = kv = k\frac{y}{\sqrt{\chi^2}} = \frac{k}{\sqrt{n}}\, t = k't$$

where $k' = k/\sqrt{n}$. Thus,

$$P[w \le z] = P[t \le 1/k'\, z]$$

The PDF of t can be used to determine the PDF for the following ratio.

$$w = \frac{y}{\sqrt{\underline{x}^T\underline{x}}} = \frac{y}{\sigma\chi} = \frac{1}{\sigma} v$$

where $\underline{x}$ is a statistically independent Gaussian random vector with zero mean value and each component of equal variance σ^2. In this manner, $k = \sigma^{-1}$ and $k' = 1/\sqrt{n\sigma^2}$, hence,

$$P[w \le z] = P[t \le \sqrt{n\sigma^2}\, z]$$

2.2.4.5.5 Quadratic Form of a Gaussian Random Vector

Let y be a quadratic form of the random vector $\underline{x}$, i.e.,

$$y = \underline{x}^T Q \underline{x}$$

where Q is a positive definite symmetrical matrix. Of course, y is a scalar random variable. The moment generating function for y is given by

$$mgf_y(s) = E[e^{sy}]$$

$$= E[e^{s(\underline{x}^T Q \underline{x})}]$$

$$mgf_y(s) = \int_{D(\underline{x})} e^{s(\underline{x}^T Q \underline{x})} f(\underline{x}) d\underline{x}$$

Now, if $\underline{x}$ is a Gaussian random vector the pdf of $\underline{x}$ is as follows.

$$f(\underline{x}) = \frac{1}{\sqrt{(2\pi)^n |\Gamma_x|}} \exp -1/2 (\underline{x}-\underline{m})^T \Gamma_x^{-1} (\underline{x}-\underline{m})$$

where Γ_x is the covariance matrix for $\underline{x}$ and $\underline{m} = E(\underline{x})$. For this case mgfy (s) becomes

$$mgf_y(s) = \frac{1}{\sqrt{(2\pi)^n |\Gamma_x|}} \int_{D(\underline{x})} \exp [G(\underline{x}, s)] d\underline{x}$$

where

$$G(\underline{x}, s) = s(\underline{x}^T Q \underline{x}) - 1/2 (\underline{x}-\underline{m})^T \Gamma_x^{-1} (\underline{x}-\underline{m})$$

Now, since

$$\underline{x}^T Q \underline{x} = (\underline{x}-\underline{m})^T Q (\underline{x}-\underline{m}) + 2\underline{m}^T Q \underline{x} - \underline{m}^T Q \underline{m}$$

the function $G(\underline{x}, s)$ becomes

$$G(\underline{x}, s) = [2s\underline{m}^T Q(\underline{x}-\underline{m}) + s\underline{m}^T Q \underline{m} - (\underline{x}-\underline{m})^T A (\underline{x}-\underline{m})]$$

where $A = -s Q + \frac{1}{2} \Gamma_x^{-1}$. Thus,

$$mgfy(s) = \frac{e^{s(m^T Q m)}}{\sqrt{(2\pi)^n |\Gamma_x|}} \int_{D(x)} \exp\left[2s m^T Q (x-m) - (x-m)^T A (x-m)\right] dx$$

The integral can be evaluated using $I_1(\underline{s})$ of Appendix A with an appropriate definition of terms. The results are as follows.

$$mgfy(s) = \frac{1}{\sqrt{2^n |A| |\Gamma_x|}} \exp\left[s m^T Q m + s^2 m^T Q A^{-1} Q m\right]$$

The matrix A can be written in the following form.

$$A = 1/2\, \Gamma_x^{-1} (I - 2s \Gamma_x Q)$$

Hence, the determinant and inverse of A become

$$|A| = 2^{-n} |\Gamma_x|^{-1} \cdot |I - 2s \Gamma_x Q|$$

$$A^{-1} = 2 (I - 2s \Gamma_x Q)^{-1} \Gamma_x = 2B\Gamma_x$$

where $B = (I - 2s\ \Gamma_x Q)^{-1}$. Therefore,

$$mgfy(s) = \frac{1}{\sqrt{|I - 2s\Gamma_x Q|}} \exp\left[s m^T Q (I + 2s B \Gamma_x Q) m\right]$$

In general, the pdf of y is difficult to obtain and usually only approximate and limiting forms can be obtained for the pdf of y . However, it is possible to obtain the statistical moments of y from mgfy(s) . This is accomplished by taking the appropriate partial derivatives with respect to s and evaluating at $s = 0$ as discussed previously. It is necessary to obtain expressions for the terms of mgfy(s) for which the derivatives can be determined. It is convenient to use the logarithm of mgfy(s) , i.e.,

$$\ln\left[mgfy(s)\right] = -1/2 \ln\left[|I - 2s\Gamma_x Q|\right] + s\left[m^T Q (I + 2s B \Gamma_x Q) m\right]$$

It is noted that

$$\frac{d}{ds} \ln [mgfy(s)] = \frac{1}{mgfy(s)} \cdot \frac{d}{ds} [mgfy(s)]$$

Therefore,

$$\frac{d}{ds} [mgfy(s)]\Big|_{s=0} = mgfy(s)\Big|_{s=0} \cdot \frac{d}{ds} \ln [mgfy(s)]\Big|_{s=0}$$

However

$$mgfy(s=0) = 1$$

hence

$$\frac{d}{ds} [mgfy(s)]\Big|_{s=0} = \frac{d}{ds} \ln [mgfy(s)]\Big|_{s=0}$$

Thus,

$$E(y) = \frac{d}{ds} [mgfy(s)]\Big|_{s=0}$$

$$= \frac{d}{ds} \ln [mgfy(s)]\Big|_{s=0}$$

$$E(y) = m^T Q m - \tfrac{1}{2} \frac{d}{ds} \ln [|I - 2s\Gamma_x^T Q|]\Big|_{s=0} + 2 \frac{d}{ds} [s^2 m^T Q B \Gamma_x^T Q m]\Big|_{s=0}$$

It should be noted that in taking the derivative of the last term the matrix B is a function of s .

In a similar manner, it is found that the variance of y can be determined directly from the second derivative of $\ell_n[mgfy(s)]$, i.e.,

$$\frac{d^2}{ds^2} \ln [mgfy(s)] = \frac{d}{ds} \left\{ mgfy^{-1}(s) \cdot \frac{d}{ds} [mgfy(s)] \right\}$$

$$= -mgfy^{-2}(s) \left[\frac{d}{ds} mgfy(s) \right]^2 + \cdots$$

$$\cdots + mgfy^{-1}(s) \frac{d^2}{ds^2} [mgfy(s)]$$

Evaluating at $s = 0$, it is found that

$$\frac{d^2}{ds^2} \ln [mgfy(s)] \Big|_{s=0} = \frac{d^2}{ds^2} mgfy(s) \Big|_{s=0} - \left[\frac{d}{ds} mgfy(s) \Big|_{s=0} \right]^2$$

$$= E(y^2) - E^2(y)$$

Therefore,

$$\sigma_y^2 = E(y^2) - E^2(y)$$

$$= \frac{d^2}{ds^2} \ln [mgfy(s)] \Big|_{s=0}$$

$$\sigma_y^2 = -1/2 \frac{d^2}{ds^2} \ln [|I - 2s\Gamma_x Q|] \Big|_{s=0} + \cdots$$

$$\cdots + 2 \frac{d^2}{ds^2} [s^2 m^T Q B \Gamma_x Q m] \Big|_{s=0}$$

The derivatives of the two terms involving s can be determined in the following manner. The determinant of $(I - 2s \Gamma_x Q)$ can be expressed as a polynomial in s by using the orthogonal transformation defined by the modal matrix for $\Gamma_x Q$, i.e., let M be the matrix of normalized eigenvectors of $\Gamma_x Q$. In this manner,

$$M^T \Gamma_x Q M = \Lambda$$

$$M^T M = I$$

$$|M| = 1$$

where Λ is a diagonal matrix of the eigenvalues of $\Gamma_x Q$. Now, the determinant of the product of a set of matrices is equal to the product of the determinants of the matrices in the set, hence,

$$|M^T (I - 2s \Gamma_x Q) M| = |M^T| \cdot |I - 2s \Gamma_x Q| \cdot |M|$$

However, since $|M| = |M^T| = 1$, it follows that

$$|I - 2s \Gamma_x Q| = |M^T (I - 2s \Gamma_x Q) M|$$

$$= |M^T I M - 2s M^T \Gamma_x Q M|$$

$$|I - 2s \Gamma_x Q| = |I - 2s \Lambda|$$

Now, the matrix $(I - 2s \Lambda)$ is a diagonal matrix, therefore,

$$|I - 2s \Gamma_x Q| = \prod_{i=1}^{n} (1 - 2s \lambda_i)$$

where λ_i are the eigenvalues of $\Gamma_x Q$ for $i = 1, 2, \ldots, n$. Thus,

$$\ln [|I - 2s \Gamma_x Q|] = \ln \left[\prod_{i=1}^{n} (1 - 2s \lambda_i) \right]$$

$$= \sum_{i=1}^{n} \ln (1 - 2s \lambda_i)$$

It becomes apparent that

$$\frac{d}{ds} \ln\left[\left|I - 2s\Gamma_x^{-1}Q\right|\right]\Big|_{s=0} = -2\sum_{i=1}^{n} \lambda_i$$

$$\frac{d^2}{ds^2} \ln\left[\left|I - 2s\Gamma_x^{-1}Q\right|\right]\Big|_{s=0} = -4\sum_{i=1}^{n} \lambda_i^2$$

Now, the modal matrix M can be used to determine the derivative of the matrix B as needed. That is,

$$B = (I - 2s\Gamma_x^{-1}Q)^{-1}$$

$$B^{-1} = (I - 2s\Gamma_x^{-1}Q)$$

$$M^T B^{-1} M = M^T (I - 2s\Gamma_x^{-1}Q) M$$
$$= (I - 2s\Lambda)$$

Thus,

$$B^{-1} = M(I - 2s\Lambda)M^T$$

$$B = M(I - 2s\Lambda)^{-1}M^T$$

The matrix $(I - 2s\Lambda)$ is diagonal with elements $1 - 2s\lambda_i$, hence, $(I - 2s\Lambda)^{-1}$ is diagonal with elements $(1 - 2s\lambda_i)^{-1}$. Therefore,

$$\frac{d}{ds} B\Big|_{s=0} = 2M\Lambda M^T$$

$$\frac{d^2}{ds^2} B\Big|_{s=0} = 8M\Lambda^2 M^T$$

In this manner it is found that

$$\frac{d}{ds}\left[s^2 m^T Q B \Gamma_x Q m\right]\Big|_{s=0} = 0$$

$$\frac{d^2}{ds^2}\left[s^2 m^T Q B \Gamma_x Q m\right]\Big|_{s=0} = 2 m^T Q \Gamma_x Q m$$

Using the foregoing results the following is obtained for the first and second moments of y .

$$E(y) = m^T Q m + \sum_{i=1}^{n} \lambda_i$$

$$\sigma_y^2 = 2\sum_{i=1}^{n} \lambda_i^2 + 4 m^T Q \Gamma_x Q m$$

$$E(y^2) = 4 m^T Q \Gamma_x Q m + 2\sum_{i=1}^{n} \lambda_i^2 + \left[m^T Q m + \sum_{i=1}^{n} \lambda_i\right]^2$$

For the special case of $m = 0$, these results become

$$E(y) = \sum_{i=1}^{n} \lambda_i$$

$$\sigma_y^2 = 2\sum_{i=1}^{n} \lambda_i^2$$

$$E(y^2) = \left(\sum_{i=1}^{n} \lambda_i\right)^2 + 2\sum_{i=1}^{n} \lambda_i^2$$

It is noted that the sum of the eigenvalues of $\Gamma_x Q$ is equal to the "trace" of $\Gamma_x Q$ which is the sum of the diagonal elements of $\Gamma_x Q$. Similarly, the sum of the squares of the eigenvalues of $\Gamma_x Q$ is the trace of the square of $\Gamma_x Q$. Thus, the eigenvalues of $\Gamma_x Q$ are not needed to define the first and second moments of y , i.e., these moments can be expressed if the traces of $\Gamma_x Q$ and $(\Gamma_x Q)^2$ as follows.

$$E(y) = \underline{m}^T Q \underline{m} + TR[\Gamma_x Q]$$

$$\sigma_y^2 = 2\,TR[(\Gamma_x Q)^2] + 4\,\underline{m}^T Q \Gamma_x Q \underline{m}$$

where TR[ ] is the trace of the matrix within the brackets, which is the sum of the diagonal elements of the matrix.

2.2.5 Probability Bounds

The probability density function $f(\underline{y})$ of a random vector $\underline{y}$ contains the complete information required to specify the probability that $\underline{y}$ will lie in some region of space which defines the domain of $\underline{y}$. However, quite often the explicit form of $f(\underline{y})$ is not known or it is not easily determined. On the other hand, the lower order moments of $\underline{y}$ are known they are easily determined. In such cases it is convenient to use the available moments of $\underline{y}$ to specify domains of $\underline{y}$ and associated probabilities. Of course, explicit statements are not to be expected since the lower order moments do not contain all of the information concerning the probability of occurrence of $\underline{y}$. On the other hand, useful bounds for the probability of occurrence can be obtained in terms of the moments of $\underline{y}$. Several such bounds are given below.

2.2.5.1 Tchebycheff Inequality

Let x be a random variable with probability density function $f(x)$. The second central moment, or variance, σ_x^2, of x is given by

$$\sigma_x^2 = \int_{-\infty}^{\infty} (x - m)^2 f(x)\, dx$$

where m = first moment of x, or $m = E(x)$.

The integral can be divided into three ranges as follows.

$$\sigma_x^2 = \int_{-\infty}^{m-\alpha} (x-m)^2 f(x)\, dx + \int_{m-\alpha}^{m+\alpha} (x-m)^2 f(x)\, dx + \int_{m+\alpha}^{+\infty} (x-m)^2 f(x)\, dx$$

where $\alpha > 0$.

By neglecting the second integral, the following inequality is obtained.

$$\sigma_x^2 \geq \int_{-\infty}^{m-\alpha} (x-m)^2 f(x)\,dx + \int_{m+\alpha}^{+\infty} (x-m)^2 f(x)\,dx$$

It is easily seen that $(x - m)^2 \geq \alpha^2$ for $-\infty < x \leq m - \alpha$ and $m + \alpha \leq x < +\infty$; therefore,

$$\sigma_x^2 > \alpha^2 \left[\int_{-\infty}^{m-\alpha} f(x)\,dx + \int_{m+\alpha}^{+\infty} f(x)\,dx \right]$$

The two integrals within the brackets yield the probability that x does not lie in the interval from $m - \alpha$ to $m + \alpha$ or that $(x - m)^2 \geq \alpha^2$, i.e.,

$$P\left[(x-m)^2 \geq \alpha^2\right] = \int_{-\infty}^{m-\alpha} f(x)\,dx + \int_{m+\alpha}^{+\infty} f(x)\,dx$$

Thus,

$$\frac{\sigma_x^2}{\alpha^2} > P\left[(x-m)^2 \geq \alpha^2\right]$$

where $\alpha > 0$. This inequality essentially bounds the probability in terms of the second central moment. Obviously, the inequality can be written in several equivalent forms; i.e., let $\alpha = k\ \sigma_x$, then

$$\frac{1}{k^2} > P\left[(x-m)^2 \geq k^2 \sigma_x^2\right]$$

Also,

$$\frac{1}{k^2} > P\left[|x-m| \geq k\,\sigma_x\right]$$

It is apparent that

$$P\left[(x-m)^2 < k^2\sigma_x^2\right] + P\left[k^2\sigma_x^2 \leq (x-m)^2\right] = 1$$

therefore,

$$P\left[(x-m)^2 < k^2\sigma_x^2\right] > 1 - \frac{1}{k^2}$$

Also,

$$P\left[|x-m| < k\sigma_x\right] > 1 - \frac{1}{k^2}$$

2.2.5.2 An Inequality for a Positive Random Variable

Let x be a random variable with probability density function $f(x)$ such that

$$f(x) = 0 \quad \text{for } x < 0$$

The random variable x is non-negative or positive semi-definite. Now, the first moment of x, $E(x)$, is given by

$$E(X) = \int_0^{+\infty} X f(X)\, dX$$

$$= \int_0^{\alpha} X f(X)\, dX + \int_{\alpha}^{+\infty} X f(X)\, dX$$

where $\alpha \geq 0$. By neglecting the first integral, the following inequality is obtained.

$$E(X) \geq \int_{\alpha}^{+\infty} X f(X)\, dX$$

Now, $x > 0$ for the range from α to $+\infty$; thus,

$$E(X) > \alpha \int_{\alpha}^{+\infty} f(X)\,dX$$

The integral is simply the probability that x will lie in the interval from α to $+\infty$; thus,

$$\frac{E(X)}{\alpha} > P\left[X \geq \alpha\right]$$

The inequality essentially bounds the probability that x will exceed α in terms of the first moment of x , where $x \geq 0$. Of course, α can be taken as $k\,E(x)$ in which case the following inequality is obtained.

$$\frac{1}{k} > P\left[X \geq k\,E(X)\right]$$

where $f(x) = 0$ for $x < 0$

2.2.5.3 Frechet Inequality

The Tchebycheff inequality determines a probability bound for an interval which is symmetrical placed about the first moment of a random variable. It is possible to determine a similar bound for an interval which is not symmetrical about the first moment. Consider the interval I from $m - k\sigma_x$ to $m + k_2\sigma_x$ where $k_1, k_2 \geq 0$ and m and σ_x^2 are the mean and variance for the random variable x . The length ℓ and center c of the interval I are given by

$$\ell = m + k_2\,\sigma_x - m + k_1\,\sigma_x$$

$$\ell = (k_1 + k_2)\,\sigma_x$$

$$c = m - k_1\,\sigma_x + 1/2\,\ell$$

$$c = m + 1/2\,(k_2 - k_1)\,\sigma_x$$

Now, if x lies outside of I, then $|x - c| > \frac{1}{2}\ell$ or, equivalently,

$$(x-c)^2 > \left(\frac{k_1 + k_2}{2}\right)\sigma_x^2$$

Let $y = (x - c)^2 \geq 0$. Since y is a positive random variable, the previous inequality shows that

$$P\left[y > \left(\frac{k_1 + k_2}{2}\right)^2 \sigma_x^2\right] < \frac{E(y)}{\left(\frac{k_1 + k_2}{2}\right)\sigma_x^2}$$

Equivalently,

$$P\left[(x-c)^2 > \left(\frac{k_1 + k_2}{2}\right)^2 \sigma_x^2\right] < \frac{E\left[(x-c)^2\right]}{\left(\frac{k_1 + k_2}{2}\right)^2 \sigma_x^2}$$

However, the mean of y is given by

$$\begin{aligned} E(y) &= E\left[(x-c)^2\right] \\ &= E\left[(x-m) - 1/2\,(k_2 - k_1)\,\sigma_x\right]^2 \\ &= E\left[(x-m)^2 - (x-m)(k_2 - k_1)\,\sigma_x + \left(\frac{k_2 - k_1}{2}\right)^2 \sigma_x^2\right] \\ &= \sigma_x^2 + \left(\frac{k_2 - k_1}{2}\right)^2 \sigma_x^2 \\ E(y) &= \left[1 + \left(\frac{k_2 - k_1}{2}\right)^2\right]\sigma_x^2 \end{aligned}$$

Therefore,

$$P\left[(x-c)^2 > \left(\frac{k_1 + k_2}{2}\right)^2 \sigma_x^2\right] < \frac{1 + \left(\frac{k_2 - k_1}{2}\right)^2}{\left(\frac{k_1 + k_2}{2}\right)^2}$$

2.2.5.4 Bienayme Inequality

Let $y = |x - a|^n$ where x is a random variable and a and n are constants. Clearly, $y \geq 0$; thus,

$$P(y \geq \alpha) < \frac{E(y)}{\alpha}$$

It follows that

$$P\left[|x-a|^n \geq \alpha\right] < \frac{E(|x-a|^n)}{\alpha}$$

2.2.5.5 The Law of Large Numbers

Let $\underline{x}$ be a statistically independent random vector such that

$$f(\underline{x}) = \prod_{i=1}^{n} f(x_i)$$

and

$$E(x_i) = m_i$$

$$E\left[(x_i - m_i)^2\right] = \sigma_i^2$$

Consider the arithmetic mean, s, of the sum of the components of $\underline{x}$, i.e.,

$$s = \frac{1}{n}\,\underline{1}^T\underline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

It follows that

$$E(s) = \frac{1}{n}\sum_{i=1}^{n} m_i$$

$$\sigma_s^2 = E\left[s - E(s)\right]^2 = \frac{1}{n^2}\sum_{i=1}^{n} \sigma_i^2$$

If the variance of s, σ_s^2, tends to zero as n becomes indefinitely large, then it can be shown that s will approach $E(s)$. More explicitly, if

$$\lim_{n\to\infty}\left[\frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2\right] = 0$$

then there exists an n such that

$$P\left[|S-E(S)| < \epsilon\right] > 1-\eta_\epsilon$$

where ϵ and η_ϵ are arbitrary positive numbers which determine a suitable n. This result is referred to as the "Law of Large Numbers."

The foregoing can be established in the following manner. Using the Tchebycheff inequality it follows that

$$P\left[|S-E(S)| < k\,\sigma_S\right] > 1-\frac{1}{k^2}$$

By letting $k^2 = \eta_\epsilon^{-1}$ it is found that

$$P\left[|S-E(S)| < \eta_\epsilon^{-1/2}\,\sigma_S\right] > 1-\eta_\epsilon$$

Thus, for all $n \geq N(\epsilon, \eta_\epsilon)$ such that $\epsilon \geq \eta_\epsilon^{-\frac{1}{2}}\sigma_S$ or $\epsilon\sqrt{\eta_\epsilon} \geq \sigma_S$ it follows that

$$P\left[|S-E(S)| < \epsilon\right] > 1-\eta_\epsilon$$

The bound for n, $N(\epsilon, \eta_\epsilon)$ is determined by the smallest n such that

$$\epsilon^2\eta_\epsilon \geq \sigma_S^2 = \frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2$$

The condition that

$$\lim_{n\to\infty}\left[\frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2\right] = 0$$

assures that for each ϵ and η_ϵ there exists a number $N(\epsilon, \eta_\epsilon)$ such that for $n \leq N(\epsilon, \eta_\epsilon)$ then

$$P\left[|S-E(S)| < \epsilon\right] > 1-\eta_\epsilon$$

The bound for n can be written as

$$N(\epsilon, \eta_0) = \frac{1}{\epsilon \eta_0^{1/2}} \left[\sum_{i=1}^{n} \sigma_i^2 \right]^{1/2}$$

if σ_s^2 uniformly converges to zero for increasing n.

A particular case of interest is that for which each x_i has the same mean and variance; i.e.,

$$E(x_i) = m$$

$$E\left[(x_i - m)^2\right] = \sigma^2$$

In this case $E(s) = m$ and $\sigma_s^2 = \frac{1}{2} n \sigma^2$ and

$$P\left[|s - m| < \epsilon\right] > 1 - \eta_0$$

where $\epsilon^2 \eta_0 \geq 1/2\ \sigma^2$ or

$$n \geq \frac{\sigma^2}{\epsilon^2 \eta_0}$$

This particular case is referred to as the "Weak" Law of Large Numbers, which implies that the conditions stated are sufficient but not necessary for convergence.

2.2.6 Limiting Theorems

Some of the most useful results of mathematical considerations of probability consist of "limiting" theorems which, in general terms, describe the behavior of a random variable that is the sum of a large number of statistically independent random variables. Alternatively, limiting theorems can be considered to be a study of the properties of the results of the repeated convolution of probability density functions. It is somewhat remarkable that under rather general conditions the repeated convolution of arbitrary probability density functions approaches the Gaussian probability density function in the limit. The result is often applied in various statistical analyses; however, there exist certain requirements of conditions of validity for these basic results. The most basic and useful limiting theorems are discussed in this section for the primary purpose of understanding the conditions of validity and the useful applications of the results.

2.2.6.1 Central Limit Theorem

One of the most basic results of mathematical considerations of probability is the "central limit theorem" which states, in general terms, that

the sum of n statistically independent random variables, with identical probability density functions, approaches a Gaussian random variable as n becomes large. An alternate statement is that the repeated convolution of an arbitrary probability density function approaches a Gaussian probability density function in the limit. An explicit formulation of the theorem is given below.

Let $y = \underline{1}^T \underline{X}$ where $\underline{X}$ is a statistically independent random vector, i.e.,

$$y = \sum_{i=1}^{n} x_i$$

and

$$f(\underline{x}) = \prod_{i=1}^{n} f(x_i)$$

Further, let the probability density function for each X_i be the same, but arbitrary; i.e., the random vector can be considered as a set of independent samples taken from a random process with an arbitrary probability density function. It follows that the mean and variance of each X_i are equal; i.e., $E(X_i)=m$ and $E(X_i - m_i)^2 = \sigma^2$, for all i. Thus, the mean and variance of y become

$$E(y) = nE(x) = nm$$

$$\sigma_y^2 = n\sigma^2$$

Now, consider a random variable Z defined by

$$z = \frac{1}{\sqrt{n}}\left[y - E(y)\right]$$

$$= \frac{1}{\sqrt{n}}\left[\sum_{i=1}^{n} x_i - nm\right]$$

$$z = \frac{1}{\sqrt{n}}\left[\sum_{i=1}^{n} (x_i - m)\right]$$

It is apparent that $E(Z) = 0$ and $\sigma_Z^2 = \sigma^2$; i.e., Z has zero mean and variance equal to that of X.

The moment-generating function for Z becomes

$$mg f_z(\Delta) = E\left[e^{\Delta z}\right]$$

$$= E\left\{ \exp \Delta\left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (x_i - m)\right]\right\}$$

$$= E\left\{ \prod_{i=1}^{n} \exp \Delta\left[\frac{1}{\sqrt{n}} (x_i - m)\right]\right\}$$

Since the X_i are statistically independent, the moment generating function for Z can be expressed as the nth power of the moment generating function of $1/\sqrt{n}$ $(X_i - m)$; i.e.,

$$mg f_z(\Delta) = \int_{D(x)} \left\{ \prod_{i=1}^{n} \exp \Delta\left[\frac{1}{\sqrt{n}} (x_i - m)\right]\right\} f(\underline{x})\, d\underline{x}$$

$$= \int_{D(x)} \prod_{i=1}^{n} \left\{ \exp \Delta\left[\frac{1}{\sqrt{n}} (x_i - m)\right] f(x_i)\right\} d\underline{x}$$

$$= \prod_{i=1}^{n} \int_{-\infty}^{+\infty} \exp \Delta\left[\frac{1}{\sqrt{n}} (x_i - m)\right] f(x_i)\, dx_i$$

$$mg f_z(\Delta) = \prod_{i=1}^{n} \left[mg f_i(\Delta)\right]$$

where

$$mg f_i(\Delta) = \int_{-\infty}^{+\infty} \exp \Delta\left[\frac{1}{\sqrt{n}} (x_i - m)\right] f(x_i)\, dx_i$$

It is seen that $mgf_i(\Delta)$ is the moment generating function for $1/\sqrt{n}$ $(X_i - m)$ which is the same for each i; thus,

$$mg f_z(\Delta) = \left[mg f_i(\Delta)\right]^n$$

Since the probability density function $f(X_i)$ is arbitrary, $mgf_i(\lambda)$ is not explicitly known; however, $mgf_i(\lambda)$ can be expanded in a power series in terms of the central moments of X_i, i.e.,

$$m_{ir} = E\left[\frac{1}{\sqrt{n}}(x_i - m)\right]^r$$

$$= \frac{1}{n^{r/2}} E\left[(x_i - m)^r\right]$$

$$m_{ir} = \frac{1}{n^{r/2}} \mu_r$$

where μ_r is the rth central moment of X_i, and m_{ir} is the rth moment of $1/\sqrt{n}\ (X_i - m)$. Of course, $\mu_1 = 0$ and $\mu_2 = \sigma^2$. Expanding $mfg_i(\lambda)$ in a power series for $\exp\lambda[1/\sqrt{n}\,(X_i - m)]$, it is found that

$$mgf_i(\lambda) = E\left[1 + \frac{\lambda}{\sqrt{n}}(x_i - m) + \frac{\lambda^2}{2n}(x_i - m)^2 + \ldots + \frac{\lambda^r}{r!\,n^{r/2}}(x_i - m)^r + \ldots\right]$$

$$= \left[1 + \lambda m_{i1} + \frac{\lambda^2}{2} m_{i2} + \ldots + \frac{\lambda^r}{r!} m_{ir} + \ldots\right]$$

$$= \left[1 + 0 + \frac{\lambda^2\sigma^2}{2n} + \ldots + \frac{\lambda^r \mu_r}{(r!)\,n^{r/2}} + \ldots\right]$$

$$mgf_i(\ ) = \left[1 + \frac{1}{n}\left(\frac{\lambda^2\sigma^2}{2} + \ldots + \frac{\lambda^r \mu_r}{(r!)(\sqrt{n})^{r-2}} + \ldots\right)\right]$$

Thus,

$$mgf_z(\lambda) = \left[1 + \frac{1}{n}\left(\frac{\lambda^2\sigma^2}{2} + \ldots + \frac{\lambda^r \mu_r}{(r!)(\sqrt{n})^{r-2}} + \ldots\right)\right]^n$$

$$mgf_z(\lambda) = \left[1 + \frac{1}{n} u(n)\right]^n$$

where

$$u(n) = \left[\frac{\lambda^2\sigma^2}{2} + \ldots + \frac{\lambda^r \mu_r}{(r!)(\sqrt{n})^{r-2}} + \ldots\right]$$

It is seen that as n becomes large, u(n) approaches a finite limit and for sufficiently large n, 1/n u(n) becomes arbitrarily small. Thus, as n becomes large, $mgf_Z(\omega)$ can be expanded in a convergent power series as follows.

$$mgf_z(\omega) = \left[1 + u(n) + \frac{(n-1)}{2!\,n}u^2(n) + \ldots + \frac{n!\,u^k(n)}{n^k(n-k)!\,k!} + \ldots\right]$$

Since

$$\lim_{n\to\infty}\left[\frac{n!}{n^k(n-k)!}\right] = \lim_{n\to\infty}\left[\frac{(n-1)(n-2)\ldots(n-k)}{n^k}\right]$$

$$= \lim_{n\to\infty}\prod_{i=1}^{k}\frac{(n-i)}{n}$$

$$= \prod_{i=1}^{k}\lim_{n\to\infty}\left(1-\frac{i}{n}\right)$$

$$\lim_{n\to\infty}\left[\frac{n!}{n^k(n-k)!}\right] = 1$$

It follows that

$$\lim_{n\to\infty}\left[mgf_z(\omega)\right] = e^{+\frac{\Delta\sigma^2}{2}}$$

Therefore, as n increases without bound, the moment generating function of Z approaches that for a Gaussian random variable with zero mean value and variance equal to σ^2. Alternatively, for sufficiently large n the probability density function for Z approaches the Gaussian density function with zero mean and variance σ^2.

From the foregoing it can be concluded that the arithmetic mean of a sufficiently large set of statistically independent samples of a random variable will be distributed as a Gaussian random variable in the limit. That is, let

$$S = \frac{1}{n}y = \frac{1}{n}\underline{1}^T\underline{x} = \frac{1}{n}\sum_{i=1}^{n}x_i = \frac{1}{n}Y$$

It follows that

$$E(s) = \frac{1}{n}E(y) = m$$

and

$$\sigma_s^2 = \frac{1}{n^2}\sigma_y^2 = \frac{1}{n}\sigma^2$$

Furthermore, if n is sufficiently large, then S will be approximately Gaussian; i.e., the probability that S will deviate from the mean value of X can be determined by considering the behavior of a Gaussian random variable for sufficiently large n. This represents an application of the central limit theorem.

2.2.6.2 Local Limit Theorem

One of the most useful results of mathematical probability is often referred to as the "local limit" theorem. The results of this theorem establish a convenient limiting expression for the probability density function of an independent trials process. This result has application in statistical methods of hypothesis testing. The conditions of validity of the theorem should be understood; thus, the basis of the theorem is considered below.

Consider a random process which has m distinct possibilities, or possible outcomes, each with probability pi for $i = 1, 2, \cdots, m$. Moreover, let the process be independent in trials such that in n trials, the probability of a particular sequence of outsomes is given by

$$P(n) = \prod_{\ell=1}^{n} p(s_{i\ell})$$

where $p(s_{i\ell})$ denotes the probability of the particular outcome in the ℓ th element of the sequence. In a total of n trials the possible outcomes can be repeated and, in general, each outcome can occur k_i times in n trials; hence, P(n) can be written as follows.

$$P(n) = \prod_{i=1}^{m} p_i^{k_i}$$

where

$$\sum_{i=1}^{m} k_i = n \quad \text{and} \quad \sum_{i=1}^{m} p_i = 1$$

The set of m possible outcomes is referred to as being mutually exclusive and exhaustive. Now, for a given set of k_i there exists a total of $N(\underline{k})$ sequences of outcomes where $\underline{k} = (k_1, k_2, \cdots k_m)$. Thus, the probability of a particular $\underline{k}$ in n trials can be written as

$$p(n, \underline{k}) = N(\underline{k}) \prod_{i=1}^{m} p_i^{k_i}$$

The number $N(\underline{k})$ is the number of ways in which n elements can be arranged into m ordered sets with k_i elements in the ith set for $i = 1, 2, \cdots m$. From combinational analysis it is known that

$$N(\underline{k}) = \frac{n!}{\prod_{i=1}^{m} (k_i)!}$$

Thus,

$$p(n, \underline{k}) = \frac{n!}{\prod_{i=1}^{m} (k_i)!} \prod_{i=1}^{m} p_i^{k_i}$$

It is apparent that $p(n, \underline{k})$ is the probability density function for the random vector $\underline{k}$, i.e., the probability of occurrence of a particular $\underline{k}$ in n trials is $p(n, \underline{k})$. It should be noted that since p_i is the probability of occurrence of the ith outcome, the expected number of occurrences of the ith outcome is np_i, i.e.,

$$E(k_i) = n p_i$$

In general, $p(n, \underline{k})$ is difficult to evaluate; however, if each k_i is sufficiently large, then the factorials in $p(n, \underline{k})$ can be accurately approximated by use of Striling's formula for factorials, i.e., (see Reference 1)

$$\alpha! = \Gamma(\alpha+1) \simeq \sqrt{2\pi\alpha}\,(\alpha^{\alpha})\, e^{-\alpha}$$

In this manner, it is found that

$$p(n, \underline{k}) \simeq \frac{\sqrt{2\pi}\, n^{n+\frac{1}{2}}\, e^{-n}}{\sqrt[m]{2\pi}\left[\prod_{i=1}^{m} k_i^{k_i+\frac{1}{2}}\right] e^{-\sum_{i=1}^{m} k_i}} \left(\prod_{i=1}^{m} p_i^{k_i}\right)$$

$$\simeq \frac{\prod_{i=1}^{m} \left(\frac{np_i}{k_i}\right)^{(k_i+\frac{1}{2})}}{(2\pi n)^{\frac{m-1}{2}} \left[\prod_{i=1}^{m} p_i\right]^{\frac{1}{2}}}$$

$$p(n,\underline{k}) \simeq \frac{\prod_{i=1}^{m}\left(1+\frac{k_i - np_i}{np_i}\right)^{-(k_i+\frac{1}{2})}}{\sqrt{n}\,(2\pi)^{\frac{m-1}{2}}\left(\prod_{i=1}^{m}\sqrt{np_i}\right)}$$

Taking the natural logarithm of the numerator, it is found that

$$Ln\left[\prod_{i=1}^{m}\left(1+\frac{k_i - np_i}{np_i}\right)^{-(k_i+\frac{1}{2})}\right] = -\sum_{i=1}^{m}(k_i+\tfrac{1}{2})Ln\left(1+\frac{k_i-np_i}{np_i}\right)$$

Now, if $\left|(k_i - np_i)/np_i\right| < 1$, then

$$Ln\left[\prod_{i=1}^{m}\left(1+\frac{k_i - np_i}{np_i}\right)^{-(k_i+\frac{1}{2})}\right] = -\sum_{i=1}^{m}(k_i+\tfrac{1}{2})\sum_{j=1}^{\infty}\frac{(-1)^{j}}{j}\left(\frac{k_i-np_i}{np_i}\right)^{j}$$

$$= -\sum_{j=1}^{\infty}\frac{(-1)^{j+1}}{j}\sum_{i=1}^{m}(k_i - np_i + np_i + \tfrac{1}{2})\left(\frac{k_i-np_i}{np_i}\right)^{j}$$

$$= -\sum_{j=1}^{\infty}\frac{(-1)^{j+1}}{j}\sum_{i=1}^{m}\left[\frac{(k_i-np_i)^{j+1}}{(np_i)^{j}} + (np_i+\tfrac{1}{2})\left(\frac{k_i-np_i}{np_i}\right)^{j}\right]$$

$$= -\sum_{i=1}^{m}\tfrac{1}{2}\left[\frac{(k_i-np_i)^{2}}{np_i} + np_i\left(\frac{k_i-np_i}{np_i}\right)\right] + O(1/n)$$

$$= -\frac{1}{2}\sum_{i=1}^{m}\frac{(k_i-np_i)^{2}}{np_i} + O(1/n)$$

$$\ln\left[\prod_{i=1}^{m}\left(1+\frac{k_i-np_i}{np_i}\right)^{-(k_i+\frac{1}{2})}\right]\simeq -\frac{1}{2}\sum_{i=1}^{m}\frac{(k_i-np_i)^2}{np_i}$$

Therefore,

$$\prod_{i=1}^{m}\left(1+\frac{k_i-np_i}{np_i}\right)^{-(k_i+\frac{1}{2})}\simeq e^{-\frac{1}{2}\sum_{i=1}^{m}\frac{(k_i-np_i)^2}{np_i}}$$

Also,

$$p(n,\underline{k})\simeq\frac{\sqrt{n}\;e^{-\frac{1}{2}\sum_{i=1}^{m}\frac{(k_i-np_i)^2}{np_i}}}{(2\pi)^{\frac{m-1}{2}}\prod_{i=1}^{m}\sqrt{np_i}}$$

Alternatively,

$$p(n,\underline{k})\simeq Ke^{-\frac{1}{2}T}$$

where

$$K=\frac{\sqrt{n}}{(2\pi)^{\frac{m-1}{2}}\prod_{i=1}^{m}\sqrt{np_i}}$$

$$T=\sum_{i=1}^{m}\frac{(k_i-np_i)^2}{np_i}$$

This result is often referred to as the "local limit" theorem. It is important to note that $p(n, \underline{k})$ generally decreases as $(k_i - np_i)^2$ increases and the maximum probability occurs for $k_i = np_i = E(k_i)$ for all i. This implies that in a large number of trials the number of occurrences of each possibility should equal the expected number np_i with maximum probability; i.e., as k_i deviates from the expected number of occurrences np_i then $p(n, \underline{k})$ decreases.

It is important to note the conditions of validity of the above expression for $p(n, \underline{k})$ for finite n. Two fundamental approximations are made which require that each k_i be "sufficiently" large and that $|k_i - np_i| < np_i = E(k_i)$. Usually, if $k_i \geq 20$, then Stirling's formula is quite accurate. The value of n should be adequately large such that $|k_i - np_i| < np_i$ for $k_i \geq 20$ for all i; hence, the value of n will depend upon p_i. Generally, the expression for $p(n, \underline{k})$ is adequately accurate in the neighborhood of $E(k_i)$ if n is such that $E(k_i)$ is large for all i.

2.2.6.3 DeMoivre-Laplace Theorem

A special case of an independent trials process is that of two possible outcomes which is usually referred to as a Bernoulli trials process. For this case the local limit theorem shows that the probability density function approaches that for a Gaussian random variable. This result was first established by DeMoivre for the special case of equal probabilities and was later generalized by Laplace; hence, the name of the theorem. This theorem can be considered as a special case of the more general local limit theorem as shown below.

For the special case of two possible outcomes the results of Section 2.2.6.2 become

$$k_1 + k_2 = n$$

$$p_1 + p_2 = 1$$

$$K = \frac{\sqrt{n}}{\sqrt{2\pi}\sqrt{n^2 p_1 p_2}} = \frac{1}{\sqrt{2\pi\, n\, p_1 p_2}}$$

$$T = \frac{(k_1 - np_1)^2}{np_1} + \frac{(k_2 - np_2)^2}{np_2}$$

$$T = \frac{1}{np_1 p_2}(k_1 - np_1)^2$$

$$p(n, k_1) \simeq \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{x^2}{2\sigma^2}}$$

where $\sigma^2 = np_1p_2$ and $X = (k_1 - np_1)$. Thus, in the limit as n becomes large, k_1 has a Gaussian pdf with mean value np_1 and variance np_1p_2.

Now, it can be shown that these two moments are those for k_1 for any n, *i.e.*,

$$p(n, \underline{k}) = \frac{n!}{k_1!\, k_2!}\, p_1^{k_1} p_2^{k_2} = \frac{n!}{k_1!\,(n-k_1)!}\, p_1^{k_1} p_2^{n-k_1}$$

$$p(n, \underline{k}) = C_k^n\, p_1^k\, p_2^{n-k} = p_n(k)$$

where $k = k_1$ and C_k^n is the binomial coefficient, i.e.,

$$(a+b)^n = \sum_{k=0}^{n} C_k^n\, a^k b^{n-k}$$

The moment generating function for k becomes

$$mgf_k(s) = E[e^{sk}]$$

$$= \sum_{k=0}^{n} e^{sk}\, p_n(k)$$

$$= \sum_{k=0}^{n} C_k^n\, e^{sk}\, p_1^k\, p_2^{n-k}$$

$$= \sum_{k=0}^{n} C_k^n\, (e^s p_1)^k\, p_2^{n-k}$$

$$mgf_k(s) = (e^s p_1 + p_2)^n$$

Thus,

$$E(k) = n(e^s p_1 + p_2)^{n-1}\, p_1 e^s \Big|_{s=0} = n p_1$$

Similarly,

$$E(k^2) = n(n-1)p_1^2 + np_1$$

$$\sigma_k^2 = E(k^2) - E^2(k) = np_1p_2$$

These results have a special and important significance. Let y be defined as follows:

$$y = \frac{k - np_1}{\sqrt{np_1p_2}}$$

Since $E(k) = np_1$ and $\sigma_k^2 = np_1p_2$, the random variable y has zero mean value and unity variance. Moreover, using the foregoing results, in the limit as n increases the pdf of y approaches that for a normal random variable; i.e.,

$$f(y) \simeq \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$$

for sufficiently large n. Now define a random variable Z_j for the jth trial of a Bernoulli trials process such that if the outcome with probability p_1 occurs, then $Z_j = 1$; otherwise, $Z_j = 0$. Further, define u as the sum of Z_j for n trials; i.e.,

$$u = \sum_{i=1}^{n} z_i$$

It follows that

$$E(u) = E\left(\sum_{i=1}^{n} z_i\right) = \sum_{i=1}^{n} E(z_i)$$

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

where σ_i^2 = variance of Z_j. However, u is simply k; hence,

$$E(u) = np_1$$

$$\sigma_u^2 = np_1p_2$$

and

$$y = \frac{u - E(u)}{\sigma_u}$$

$$= \frac{\left[\sum_{i=1}^{n} z_i - E\left(\sum_{i=1}^{n} z_i\right)\right]}{\left\{E\left[\sum_{i=1}^{n} z_i - E\left(\sum_{i=1}^{n} z_i\right)\right]^2\right\}^{1/2}}$$

$$y = \frac{\sum_{i=1}^{n} \left[z_i - E(z_i)\right]}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}}$$

From the last expression y can be considered as a normalized sum of the statistically independent random variables $[Z_j - E(Z_j)]$. According to the foregoing, this sum approaches a normal random variable in terms of its probability density function for large n. This observation leads to the conjecture that, in general, a sum of statistical independent random variables will approach a Gaussian random variable in probability density function if each of the contributions to the sum of each random variable is uniformly small. This limiting property is considered further in the following section.

2.2.6.4 Lindeberg Condition, Liapounov's Theorem

The previous limiting theorems suggest that the sum of statistically independent random variables approach a Gaussian random variable in the limit. The central limit theorem and the DeMoivre-Laplace theorem are special cases of this property. It is of general interest and practical importance to consider the general conditions for which the sum of statistically independent random variables approach a Gaussian random variable. This problem was first investigated by Laplace and the first regorous proof of the sufficiency of certain conditions was given by Liapounov. However, a more general set of sufficient conditions was established by Lindeberg, which includes the conditions considered by Liapunov; thus, the results of Liapounov can be obtained from the results of Lindeberg. In the interest of generality, the results of Lindeberg are considered first. These results are generally referred to as the Lindeberg condition which is shown to be sufficient for a sum of statistically independent random variables to approach a Gaussian random variable in the limit. This condition is discussed below.

The Lindeberg condition can be stated in the following forms. Let $\underline{X}$ be a statistically independent random vector whose components have arbitrary probability density functions, and let m_i and σ_i^2 be the mean and variance of each component of $\underline{X}$, i.e.,

$$m_i = E(x_i)$$

$$\sigma_i^2 = E(x_i - m_i)^2$$

Define a random vector $\underline{Z}$ with each component $Z_i = X_i - m_i$, i.e., $\underline{Z} = \underline{X} - \underline{m}$. Consider a random variable u which is the sum of the components of $\underline{Z}$, i.e.,

$$u = \underline{1}^T \underline{z} = \sum_{i=1}^{n} z_i$$

where n is the dimension of $\underline{X}$ and $\underline{Z}$. Clearly

$$E(u) = 0$$

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

Define a random variable v as follows:

$$v = \frac{1}{\sigma_u} \underline{1}^T \underline{z}$$

It is apparent that $E(v) = 0$ and $\sigma_v^2 = 1$.

The Lindeberg condition is as follows:

$$\lim_{n \to \infty} \left(\frac{1}{\sigma_n^2}\right) \sum_{i=1}^{n} \int_{|x_i - m_i| > \tau \sigma_n} (x_i - m_i)^2 f(x_i)\, dx_i = 0$$

where $\sigma_n^2 = \sigma_u^2$ and $\tau > 0$. The notation σ_n^2 is used for σ_u^2 to denote the dependence on n. If this limit is satisfied for a random vector $\underline{X}$, then $\underline{X}$ is said to satisfy the Lindeberg condition. The following interpretation of the Lindeberg condition should be considered. It is noted that

$$\int_{|x_i - m_i| > \tau \sigma_n} (x_i - m_i)^2 f(x_i)\, dx_i \geq (\tau \sigma_n)^2 \int_{|x_i - m_i| > \tau \sigma_n} f(x_i)\, dx_i$$

The integral on the right is simply the probability that $|X_i - m_i|$ will exceed $\tau \sigma_n$, thus,

$$P\left[|x_i - m_i| > \tau \sigma_n\right] \leq \frac{1}{\tau^2 \sigma_n^2} \int_{|x_i - m_i| > \tau \sigma_n} (x_i - m_i)^2 f(x_i)\, dx_i$$

Now, the probability that the maximum of $|X_i - m_i|$ for all i exceeds $\tau \sigma_n$ is bounded by the sum of $P\left[|X_i - m_i| \geq \tau \sigma_n\right]$, i.e.,

$$P\left[MAX_i\, |x_i - m_i| \geq \tau \sigma_n\right] \leq \sum_{i=1}^{n} P\left[|x_i - m_i| > \tau \sigma_n\right]$$

where $MAX_i |X_i - m_i|$ is taken over all i. Therefore,

$$P\left[MAX_i\, |x_i - m_i| \geq \tau \sigma_n\right] \leq \frac{1}{\tau^2 \sigma_n^2} \sum_{i=1}^{n} \int_{|x_i - m_i| > \tau \sigma_n} (x_i - m_i)^2 f(x_i)\, dx_i$$

The Lindeberg condition requires that the right-hand member approach zero as n increases without bound, hence, the Lindeberg condition is equivalent to the following

$$\lim_{n \to \infty} P\left[MAX_i |x_i - m_i| \geq r \sigma_n \right] = 0$$

Thus, the Lindeberg condition requires that each random variable of the random vector $\underline{X}$ be uniformly small. In somewhat equivalent terms, the Lindeberg condition requires that none of the components of $\underline{X}$ "dominate" in a sum of the components. Alternatively, if the limit of σ_n exists for $n \to \infty$, then the Lindeberg condition is not satisfied. Thus, a requirement for the Lindeberg condition is as follows:

$$\lim_{n \to \infty} \sigma_n = \infty \quad \text{(Does not exist)}$$

It is possible to show that if the random vector $\underline{X}$ satisfies the Lindeberg condition, then the probability density function of the sum v approaches that for a Normal random variable. The essential steps in this proof are discussed below. The discussion given below is rather heuristic; a detailed rigorous proof is given in Reference 2.

The sum v can be written as follows

$$V = \frac{1}{\sigma_n} \underline{1}^T \underline{Z} = \underline{1}^T \underline{V} = \sum_{i=1}^{n} v_i$$

where

$$v_i = \frac{1}{\sigma_n} z_i$$

for i = 1, 2, ..., n. It is important to note that v is a random variable with zero mean value and unity variance, i.e.,

$$E(v) = 0$$

$$E(v^2) = \sigma_v^2 = 1$$

It should also be noted that if the random vector $\underline{X}$ satisfies the Lindeberg condition, then the random vector $\underline{V}$ satisfies the following condition.

$$\lim_{n \to \infty} \sum_{i=1}^{n} \int_{|v_i| > \tau} v_i^2 f(v_i)\, dv_i = 0$$

Now, the random vector $\underline{V}$ is statistically independent, hence, the moment generating function of v is given by

$$mgfv(s) = E\left[e^{sv}\right] = \prod_{i=1}^{n} mgf_i(s)$$

where $mgf_i(s)$ is the moment generating function of v_i for $i = 1, 2, \ldots, n$. It is noted that $mgf_i(s)$ is dependent upon n since $v_i = \sigma_n^{-1} Z_i$. Also, it is noted that

$$E(v_i) = 0$$

$$\lim_{n \to \infty} E(v_i^2) = \lim_{n \to \infty} \sigma_{v_i}^2 = \lim_{n \to \infty} \frac{\sigma_i^2}{\sigma_n^2} = 0$$

Thus, as n increases without bound, the variance of each v_i approaches zero, and the probability density function of v_i, $f(v_i)$, approaches a positive "pulse" of unit area and infinitesimal width, i.e., as $n \to \infty$, $f(v_i)$ approaches the unit impulse function which is often used in engineering analysis. Now, the Fourier transform of an infinitesimally narrow unit area pulse at the origin approaches unity, hence, it is possible to find a sufficiently large n such that

$$\left| mgf_i(s) - 1 \right| < 1/2$$

By taking the logarithm of $mgf_v(s)$ it is found that

$$Ln\left[mgf_v(s)\right] = \sum_{i=1}^{n} Ln\left[mgf_i(s)\right]$$

$$= \sum_{i=1}^{n} Ln\left[1 + mgf_i(s) - 1\right]$$

$$= \sum_{i=1}^{n} \sum_{k=1}^{\infty} \frac{(-1)^k}{k}\left[mgf_i(s) - 1\right]^k$$

$$Ln\left[mgf_v(s)\right] = \sum_{i=1}^{n}\left[mgf_i(s) - 1\right] + R_n$$

where

$$R_n = \sum_{i=1}^{n}\sum_{k=2}^{\infty} \frac{(-1)^k}{k}\left[mgf_i(s) - 1\right]^k$$

It is apparent that R_n is bounded as follows

$$\left|R_n\right| \le \sum_{i=1}^{n}\sum_{k=2}^{\infty} 1/2 \left|mgf_i(s) - 1\right|^k$$

$$\le 1/2 \sum_{k=1}^{n} \frac{\left|mgf_i(s) - 1\right|^2}{1 - \left|mgf_i(s) - 1\right|}$$

$$\left|R_n\right| \le \sum_{k=1}^{n} \left|mgf_i(s) - 1\right|^2$$

It is noted that since $E(v_i) = 0$,

$$mgf_i(s) - 1 = E\left[e^{sv_i} - 1 - sv_i\right]$$

By letting $s = \sqrt{-1}\,\omega = J\omega$ it is seen that

$$|mgf_i(s) - 1| \leq \frac{w^2}{2} E(v_i^2)$$

thus,

$$\sum_{i=1}^{n} |mgf_i(s) - 1| \leq \frac{w^2}{2} \sum_{i=1}^{n} E(v_i^2)$$

$$\leq \frac{w^2}{2} E\left[\sum_{i=1}^{n} v_i^2\right]$$

$$\leq \frac{w^2}{2} E(v^2)$$

$$\sum_{i=1}^{n} |mgf_i(s) - 1| \leq \frac{w^2}{2}$$

By multiplying both sides by the maximum of $|mgf_i(s) - 1|$, it is found that

$$|R_n| \leq \frac{w^2}{2} MAX_i |mgf_i(s) - 1|$$

Since $mgf_i(s)$ approaches unity as n increases without bound, it follows that

$$\lim_{n \to \infty} |R_n| = 0$$

Therefore,

$$\lim_{n \to 0} \left\{ Ln\left[mgf_{sr}(Jw) \right] \right\} = \lim_{n \to \infty} \sum_{i=1}^{n} \left[mgf_i(Jw) - 1 \right]$$

Now, the summation can be written as follows.

$$\sum_{i=1}^{n}\left[mgf_i(Jw) - 1 \right] = \sum_{i=1}^{n} E\left[e^{Jwv_i} - 1 - Jwv_i \right]$$

$$= \sum_{i=1}^{n} \int_{-\infty}^{\infty} \left(e^{Jwv_i} - 1 - Jwv_i \right) f(v_i)\, dv_i$$

$$= \sum_{i=1}^{n} \int_{-\infty}^{\infty} \sum_{k=2}^{\infty} \frac{(Jwv_i)^k}{k!} f(v_i)\, dv_i$$

$$= -\sum_{i=1}^{n} \int_{-\infty}^{\infty} \frac{w^2 v_i^2}{2} f(v_i)\, dv_i + \rho_n$$

$$= -\frac{w^2}{2} E(v^2) + \rho_n$$

$$\sum_{i=1}^{n}\left[mgf_i(Jw) - 1 \right] = -\frac{w^2}{2} + \rho_n$$

where

$$\rho_n = \sum_{i=1}^{n} \int_{-\infty}^{\infty} \sum_{k=3}^{\infty} \frac{(Jwv_i)^k}{k!} f(v_i)\, dv_i$$

Therefore,

$$\lim_{n\to\infty} \left\{ Ln\left[mgf_v(Jw) \right] \right\} = -\frac{w^2}{2} + \lim_{n\to\infty} \rho_n$$

However,

$$|\rho_n| \le \sum_{i=1}^{n} \left[\int_{|v_i| \le \epsilon} \frac{|wv_i|^3}{6} f(v_i)\, dv_i + \int_{|v_i| > \epsilon} (wv_i)^2 f(v_i)\, dv_i \right]$$

$$\leq \frac{|w|^3}{6} \sum_{i=1}^{n} \int_{|v_i| \leq \epsilon} |v_i|^3 f(v_i)\, dv_i + w^2 \sum_{i=1}^{n} \int_{|v_i| > \epsilon} (v_i)^2 f(v_i)\, dv_i$$

$$\leq \frac{|w|^3}{6} \epsilon \sum_{i=1}^{n} \int_{|v_i| \leq \epsilon} v_i^2 f(v_i)\, dv_i + w^2 \sum_{i=1}^{n} \int_{|v_i| > \epsilon} v_i^2 f(v_i)\, dv_i$$

$$|\rho_n| \leq \frac{|w|^3}{6} \epsilon + w^2 \left(1 - \frac{|w|}{6} \epsilon\right) \sum_{i=1}^{n} \int_{|v_i| > \epsilon} v_i^2 f(v_i)\, dv_i$$

where ϵ is an arbitrary positive number. Now, by virtue of the Lindeberg condition, the second term approaches zero for any arbitrarily small ϵ as n increases without limit, therefore,

$$\lim_{n \to \infty} \rho_n = 0$$

Thus,

$$\lim_{n \to \infty} \left\{ \operatorname{Ln} \left[m_g f_v(Jw) \right] \right\} = -\frac{w^2}{2}$$

Alternatively,

$$\lim_{n \to \infty} \left[m_g f_v(Jw) \right] = e^{-\frac{w^2}{2}}$$

The terms of the right-hand sides are simply the moment generating function of a Normal random variable where $\alpha = J\omega$, hence, the sum v approaches a Normal random variable as n increases without limit.

Liapounov established the foregoing results under a different condition which was that

$$\lim_{n\to\infty} \frac{1}{\sigma_n^{2+\delta}} \sum_{i=1}^{n} E\left|X_i - m_i\right|^{2+\delta} = 0$$

where $\delta > 0$. To prove that this condition is also sufficient for the above results, it is only necessary to show that if this condition holds, then the Lindeberg condition holds. This is easily done by the following inequalities

$$\frac{1}{\sigma_n^2} \sum_{i=1}^{n} \int_{|X_i - m_i| > \tau\sigma_n} (X_i - m_i)^2 f(X_i)\, dX_i \leq$$

$$\leq \frac{1}{\sigma_n^2 (\tau\sigma_n)^{\delta}} \sum_{i=1}^{n} \int_{|X_i - m_i| > \tau\sigma_n} |X_i - m_i|^{2+\delta} f(X_i)\, dX_i \leq$$

$$\leq \left(\frac{1}{\tau^{\delta}}\right) \frac{1}{\sigma_n^{2+\delta}} \sum_{i=1}^{n} \int_{-\infty}^{\infty} |X_i - m_i|^{2+\delta} f(X_i)\, dX_i$$

$$= \frac{1}{\tau^{\delta}} \cdot \frac{1}{\sigma_n^{2+\delta}} \sum_{i=1}^{n} E\,|X_i - m_i|^{2+\delta}$$

Thus, if the condition of Liapounov is satisfied, then the Lindeberg condition is satisfied. A direct proof of the Liapounov theorem is given in Reference 4.

2.2.7 Determination of Statistical Properties

In the design and performance analysis of Navigation and Guidance systems it is necessary to have available certain knowledge of the statistical properties of random variables, which are usually error sources that adversely affect system performance. The statistical properties of such error sources are known, then it is generally possible to define "optimum" estimation and control procedures which mitigate the adverse effects of these error sources. Also, in order to assess final system performance, it is necessary to know the statistical properties of all factors which influence system behavior. Usually, optimum navigation and guidance procedures are defined with the tacit assumption that all statistical properties which affect the procedures are known. A similar situation often exists in system performance analyses. That is, optimum procedures are usually defined on the basis of certain information being available concerning statistical properties of error sources, also, system performance statements are usually made assuming statistical properties of error sources. Of course, the usefulness and validity of such efforts and results is dependent upon the possibility of ultimately obtaining the required information of statistical properties. However, it is generally necessary to either verify or determine the required statistical properties from a set of observations of error sources or, generally, from sets of samples of random processes. In general, the required statistical properties cannot be determined explicitly, rather, they must be "estimated" from a set of samples of random variables. Therefore, it becomes necessary to consider the methods of estimation of statistical properties in both the design and performance analysis of Navigation and Guidance systems.

Generally, there exists two major aspects of the problem of "determining" or, actually, estimating statistical properties of a random variable which are: (1) estimating the required set of statistical moments which specify the probability density function; and (2) the determination of the particular type of probability density function for the random variable. Usually, the type of probability density function is assumed and the statistical moments which specify the probability density function are estimated from a set of samples of the random variable. It becomes apparent that two areas of concern exist which are: (1) the accuracy of estimating statistical moments from sample sets; and (2) the validity of assumptions concerning the types of probability density functions. These two aspects of determining statistical properties are considered below.

The problem of estimating statistical properties can be considered as equivalent to the problem of parameter estimation which has been considered in detail in a previous monograph concerning state estimation (see Reference 15). However, there exists a fundamental difference between the two problems which essentially changes the approach. In the problem of parameter estimation it is assumed that the randomness of the observation process is specified statistically which represents a'priori information that is available for estimation of the parameters of interest. In the present situation, it is this a'priori information which is being sought and there is usually no a'priori information available; that is, the randomness which is usually assumed known must now be determined. It should be pointed out that in parameter estimation as considered previously, the parameters were usually physically identifiable "state" quantities such as position and velocity deviations of a spacecraft from a reference

trajectory. Optimum estimation procedures for these state parameters require the use of the statistical properties of the observation uncertainty or random errors and, also, those of the parameters being estimated. The required statistical properties are usually the statistical moments of various error sources. For example, in the case of Gaussian error sources, the mean vector and covariance matrix are used in the optimum state parameter estimation procedure. These statistical parameters, i.e., means and covariances of error sources, must be determined ultimately from sample data sets of various error sources. That is, the statistical properties of the randomness of the state observation process must be determined to perform an optimum estimation of the state parameters, also, those of the state parameters must be determined.

It should be apparent that there exists a salient distinction between state parameter estimation and the problem of estimating statistical properties of a random process or variable. In the latter, a sufficient set of moments is usually sought, or estimates of, which specifies the random process. In the present discussions, parameters usually refer to statistical moments. Fortunately, most random processes are specified by only first and second statistical moments, e.g., Gaussian random variables, and the problem is often reduced to estimating these two moments.

2.2.7.1 Estimation of Statistical Moments

The most often encountered problem in the estimation of statistical moments is that of estimating the first moments and second central moments of a joint Gaussian probability density function; i.e., the elements of the mean vector and the covariance matrix must be determined in order to specify the probability density function. In general, there exists n random variables and n sets of samples are available to estimate the required parameters or moments. The general problem can be considered in terms of a fundamemtal problem which involves only two random variables. It should be noted that in cases of non-Gaussian random variables, the first and second statistical moments are usually sufficient to specify probability density functions. That is, in the case of non-Gaussian random variables, the corresponding probability density functions are explicit functions of parameters which are not necessarily the first and second moments of the random variables; however, the first and second moments are unique explicit functions of the parameters which specify the probability density function and, hence, these moments implicitly specify the probability density functions. It can be stated that, generally, the first and second statistical moments are adequate to specify known probability density functions of practical interest. Usually, the first moments and second central moments are adequate.

Let x and y be two random variables with the following moments which are assumed to be a sufficient set of parameters to specify the joint pdf of x and y ; e.g., in the case of Gaussian random variables.

$$E(x) = m_x$$

$$E(y) = m_y$$

$$E(x - m_x)^2 = \sigma_x^2$$
$$E(y - m_y)^2 = \sigma_y^2$$
$$E\left[(x - m_x)(y - m_y)\right] = \mu_{xy}$$

The correlation coefficient, ρ_{xy}, for the random variables is defined as

$$\rho_{xy} = \frac{\mu_{xy}}{\sigma_x \sigma_y}$$

In general terms, x and y denote two random variables that are generated from two random processes which have marginal probability density functions f(x) and f(y), respectively. The means and variances of x and y specify f(x) and f(y), respectively. If x and y are statistically independent random variables, then $\mu_{xy} = 0$ and the joint pdf of x and y is simply f(x) f(y). In this case, the moments m_x, m_y, σ_x^2 and σ_y^2 specify f(x, y). However, in the more general case the two first moments and the three second-central moments are required to specify f(x, y). Generally, the problem of determining these statistical moments, or estimating these parameters, is considered in terms of two problems being: (1) the analysis of mean and variance; and (2) the analysis of correlation. It is generally assumed that sets of sample data are available from which the required moments can be estimated. The sets of sample data for the random variables will be denoted by the vectors $\underline{x}$ and $\underline{y}$, respectively, of dimensions n, where n is the number of samples. There exist two basic problems in estimating statistical moments which concern: (1) the functions of $\underline{x}$ and $\underline{y}$ to be used as estimates for the required moments; and (2) an assessment of the accuracy of the estimates.

In general, the estimates of the required moments are denoted by $\hat{m}_x(\underline{x})$, $\hat{\sigma}_x^2(\underline{x})$, $\hat{m}_y(\underline{y})$, $\hat{\sigma}_y^2(\underline{y})$ and $\hat{\mu}_{xy}(\underline{x}, \underline{y})$, which denote that the estimates are functions of the sample sets $\underline{x}$ and $\underline{y}$. Often, the functional dependence of the estimates and the sample sets is understood and it is not explicitly denoted. It is important to note that the estimates for the required moments are functions of random variables and, hence, the estimates are random variables. There exist two basic criteria for the estimates which are: (1) the expected value of an estimate for a moment should be equal to the moment, e.g., $E[\hat{m}_x(\underline{x})] = m_x$, etc.; and (2) the statistical variation of the estimates from the moments should decrease as the sample set size increases. These criteria are usually referred to as: (1) unbiasness and (2) consistency. The basic concern in assessing the accuracy of the estimates is to determine or assure, if possible, that the error in an estimate will be limited to a prescribed value with a certain probability. This usually requires consistency in terms of the estimate variance decreasing uniformly as sample size increases. This is considered in further detail in the following sections.

2.2.7.1.1 Statistical Analysis of Mean and Variance

Let the random vector $\underline{x}$ denote a set of n samples from a random process with probability density function f(x) which has mean and variance as follows:

$$E(x) = \int_{-\infty}^{+\infty} f(x)\,dx = m$$

$$E[(x-m)^2] = \int_{-\infty}^{+\infty} (x-m)^2 f(x)\,dx = \sigma^2$$

Also, let s and Λ^2 denote the sample mean and variance as defined below:

$$s = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\,(\underline{1}^T\underline{x})$$

$$\Delta^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - s)^2 = \frac{1}{n}\,(\underline{v}^T\underline{v})$$

where $\underline{v} = \underline{x} - s\,\underline{1}$. It is seen that there exists two means and variances which refer to the random process x and the random sample $\underline{x}$. It is necessary to differentiate between these two means and variances. By convention m and σ^2 are usually referred to as the "population" mean and variance, respectively, whereas s and Λ^2 are referred to as the sample mean and variance, respectively. The essential difference is that s and Λ^2 are random variables, whereas m and σ^2 are not.

Consider the expected value of the sample mean s, i.e.,

$$E(s) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = m$$

Thus, the expected value of the sample mean is equal to the expected value of x, or the population mean, m. On this basis, the sample mean is used as an estimate for the mean value of x or the population mean, i.e., $\hat{m} = s$ where $\hat{m}$ denotes an estimate of m, the expected value of x. Now consider the variance of s, σ_s^2 ; i.e.,

$$\sigma_s^2 = E\left[(s-m)^2\right]$$

$$= E\left[\frac{1}{n^2}\,(\underline{1}^T\underline{x} - n\,m)^2\right]$$

$$= \frac{1}{n^2}\,E\left[\underline{1}^T(\underline{x} - m\,\underline{1})\right]^2$$

$$\sigma_s^2 = \frac{1}{n^2}\,\underline{1}^T R_x\,\underline{1}$$

where Γ_x is the covariance matrix for the sample set $\underline{x}$. The variance of s, σ_s^2, represents a measure of the accuracy of the sample mean s as an estimate for the population mean m. It is important to note that σ_s^2 is the variance of the sample mean s and not the population variance σ^2. If the sample set is statistical independent, or uncorrelated, then $\underline{1}^T \Gamma_x \underline{1} = n\sigma^2$, where σ^2 is the population variance, and

$$\sigma_s^2 = \frac{1}{n} \sigma^2$$

In this case, it is apparent that

$$\lim_{n \to \infty} \sigma_s^2 = 0$$

Thus, for an uncorrelated sample set $\underline{x}$ the sample mean is a consistent estimate for the population mean, m. It should be noted that an uncorrelated sample set is a sufficient condition for the consistency of s as an estimate for m. The necessary condition is that, uniformly,

$$\lim_{n \to \infty} (\sigma_s^2) = \lim_{n \to \infty} \left(\frac{1}{n^2} \underline{1}^T \Gamma_x \underline{1} \right) = 0$$

From the foregoing, it becomes apparent that if s is a consistent estimate for m, then it is possible to determine a sufficiently large n, number of samples, such that the sample mean is as close to the population mean as desired with a specified probability. This follows directly from the "law of large numbers" as discussed previously. Consider the uncorrelated sample set such that $\sigma_s^2 = (1/n)\sigma^2$. From the weak law of large numbers, it follows that

$$P_{ROB} \left[|s - m| < \epsilon \right] > 1 - \eta$$

where $\epsilon^2 \eta \geq (1/n)\sigma^2$. Thus, for a given ϵ and η if the number of samples n is given by

$$n \geq \frac{\sigma^2}{\epsilon^2 \eta}$$

then $|s - m| < \epsilon$ with probability $1 - \eta$. It is apparent that if the population variance, σ^2, is known, then the required sample size n could be directly determined without knowledge of the particular probability density function f(x). However, in the general case, the population variance σ^2 is not known and it must also be estimated from the sample set $\underline{x}$, which is considered below. Nonetheless, without explicity knowledge of the probability density function of x, f(x), or the population variance σ^2, it is known that the sample mean s is an unbiased estimate of the population mean m, and if $1/n^2 \, \underline{1}^T \Gamma_x \underline{1}$ converges

uniformly to zero for increasing n then s is a consistent estimate for m, which is true for an uncorrelated sample set.

In a similar manner, consider the expected value of Δ^2, i.e.,

$$E(\Delta^2) = \frac{1}{n} E(\underline{V}^T \underline{V})$$

$$= \frac{1}{n} E\left[(\underline{X} - s\underline{1})^T (\underline{X} - s\underline{1})\right]$$

$$= \frac{1}{n} E\left[(\underline{X} - m\underline{1} + m\underline{1} - s\underline{1})^T (\underline{X} - m\underline{1} + m\underline{1} - s\underline{1})\right]$$

$$= \frac{1}{n} E\left\{\left[(\underline{X} - m\underline{1}) - (s-m)\underline{1}\right]^T \left[(\underline{X} - m\underline{1}) - (s-m)\underline{1}\right]\right\}$$

$$= \frac{1}{n} E\left[(\underline{X} - m\underline{1})^T - s - m\underline{1}) - 2(s-m)(\underline{X} - m\underline{1})^T \underline{1} + (s-m)^2 \underline{1}^T \underline{1}\right]$$

$$= \frac{1}{n} E\left[\sum_{i=1}^{n} (x_i - m)^2 - 2(s-m)(ns - nm) + n(s-m)^2\right]$$

$$= \frac{1}{n} E\left[\sum_{i=1}^{n} (x_i - m)^2 - n(s-m)^2\right]$$

$$E(\Delta^2) = \sigma^2 - \frac{1}{n^2} \underline{1}^T \Gamma_X \underline{1}$$

Now, if the sample set is uncorrelated, then $\underline{1}^T \Gamma_X \underline{1} = n\sigma^2$ and

$$E(\Delta^2) = \sigma^2 - \frac{1}{n} \sigma^2$$

$$= \left(1 - \frac{1}{n}\right) \sigma^2$$

$$E(\Delta^2) = \left(\frac{n-1}{n}\right) \sigma^2$$

It is seen that $E(\Delta^2)$ is not equal to σ^2, i.e., Δ^2 is a biased estimate of σ^2; however,

$$E\left[\left(\frac{n}{n-1}\right)\Delta^2\right] = \sigma^2$$

Thus, $\hat{\sigma}^2$ is an unbiased estimate for σ^2, where

$$\hat{\sigma}^2 = \frac{n}{n-1}\Delta^2$$

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - s)^2$$

The consistency of $\hat{\sigma}^2$ can be considered in terms of the variance of Δ^2, $\sigma^2(\Delta^2)$, which is given by

$$\sigma^2(\Delta^2) = E\left[\Delta^2 - E(\Delta^2)\right]^2$$

$$= E\left[\frac{1}{n}(\underline{V}^T\underline{V}) - E(\Delta^2)\right]$$

$$\sigma^2(\Delta^2) = \frac{1}{n^2}E(\underline{V}^T\underline{V})^2 - E^2(\Delta^2)$$

For a statistically independent sample set, it can be shown after considerable algebraic manipulation, that (see References 1 and 6)

$$\sigma^2(\Delta^2) = \left[\left(\frac{\mu_4 - \sigma^4}{n}\right) - \frac{2}{n^2}(\mu_4 - 2\sigma^4) + \frac{\mu_4 - 3\sigma^2}{n^3}\right]$$

$$= \frac{1}{n^3}\left[(n^2 - 2n + 1)\mu_4 - (n^2 - 4n + 3)\sigma^4\right]$$

$$\sigma^2(\Delta^2) = \left(\frac{n-1}{n^3}\right)\left[(n-1)\mu_4 - (n-3)\sigma^4\right]$$

where μ_4 is the fourth central moment of x, i.e.,

$$\mu_4 = E\left[(x - m)^4\right]$$

Thus,

$$\sigma^2(\hat{\sigma}^2) = \frac{n^2}{(n-1)^2}\sigma^2(\Delta^2)$$

$$= \frac{1}{n(n-1)}\left[(n-1)\mu_4 - (n-3)\sigma^4\right]$$

$$\sigma^2(\hat{\sigma}^2) = \frac{1}{n}\left[\mu_4 - \left(\frac{n-3}{n-1}\right)\sigma^2\right]$$

It is apparent that if the sample set is statistically independent then

$$\lim_{n\to\infty} \sigma^2(\hat{\sigma}^2) = 0$$

In this case $\hat{\sigma}^2$ is a consistent estimate for σ^2.

The variance of $\hat{\sigma}^2$, $\sigma^2(\hat{\sigma}^2)$ is a measure of the accuracy in estimating the population variance σ^2 by $\hat{\sigma}^2$. The law of large numbers can be applied to show that a sufficiently large sample size, n, can be found such that the error in estimating the population variance σ^2 can be bounded with a specified probability. Of course, the variance of $\hat{\sigma}^2$ is needed which is found to be a function of the higher-order central moment, μ_4, of the population, or of the probability density function f(x). However, if the basic assumption that the mean and variance of x are sufficient to specify f(x), then the higher-order central moment μ_4 is a function of the lower-order moments. For example, consider the case of a Gaussian random variable. In this case, the higher-order central moments are all expressible in terms of the second central moment, i.e., for f(x) a Gaussian probability density function

$$\mu_{2k} = (2k)! \frac{\sigma^{2k}}{2^k(k!)}$$

where μ_{2k} is the (2k) th central moment. It follows that

$$\mu_2 = \sigma^2$$
$$\mu_4 = 3\sigma^4$$

Using these results $\sigma^2(\hat{\sigma}^2)$ becomes

$$\sigma^2(\hat{\sigma}^2) = \frac{1}{n}\left[3\sigma^4 - \left(\frac{n-3}{n-1}\right)\sigma^4\right]$$

$$= \frac{\sigma^4}{n}\left[3 - \left(\frac{n-3}{n-1}\right)\right]$$

$$= \frac{2}{n-1}\sigma^4$$

$$\sigma^2(\hat{\sigma}^2) = \left(\frac{2}{n-1}\right)(\sigma^2)^2$$

Although σ^2 is unknown, it becomes possible to determine the variance of the ratio of $\hat{\sigma}^2$ to σ^2, i.e.,

$$\sigma^2\left[\frac{\hat{\sigma}^2}{\sigma^2}\right] = \frac{1}{(\sigma^2)^2}\,\sigma^2(\hat{\sigma}^2)$$

Thus,

$$\sigma^2\left[\frac{\hat{\sigma}^2}{\sigma^2}\right] = \frac{2}{n-1}$$

It becomes apparent that by making n sufficiently large $\sigma^2[\hat{\sigma}^2/\sigma^2]$ can be made arbitrarily small. Now, $\sigma^2[\hat{\sigma}^2/\sigma^2]$ is, in turn, a direct measure of the error made in estimating σ^2, i.e., let $\epsilon = \hat{\sigma}^2 - \sigma^2$ and let the relative error be ϵ/σ^2 i.e.,

$$\frac{\epsilon}{\sigma^2} = \left(\frac{\hat{\sigma}^2}{\sigma^2} - 1\right)$$

The variance of the relative error is the same as $\sigma^2[\hat{\sigma}^2/\sigma^2]$, i.e.,

$$\sigma^2\left(\frac{\epsilon}{\sigma^2}\right) = \sigma^2\left[\frac{\hat{\sigma}^2}{\sigma^2} - 1\right]$$

$$= \sigma^2\left(\frac{\hat{\sigma}^2}{\sigma^2}\right)$$

$$\sigma^2\left(\frac{\epsilon}{\sigma^2}\right) = \left(\frac{2}{n-1}\right)$$

Thus, the variation of $\hat{\sigma}^2/\sigma^2$ about 1 is the same as the variation of ϵ/σ^2 about 0.

2.2.7.1.2 Statistical Analysis of Correlation

Let x and y be two random variables with covariance μ_{xy} given as follows:

$$\mu_{xy} = E\left\{[x - E(x)][y - E(y)]\right\}$$

$$= E(xy) - E(x)\,E(y)$$

$$\mu_{xy} = m_{xy} - m_x m_y$$

If x and y are statistically independent, then $m_{xy} = m_x\, m_y$ and $\mu_{xy} = 0$, however, $\mu_{xy} = 0$ does not imply statistical independence in the general case. If $\mu_{xy} = 0$ the random variables x and y are statistically uncorrelated random variables. The correlation coefficient, ρ_{xy}, is defined as

$$\rho_{xy} = \frac{\mu_{xy}}{\sigma_x \sigma_y}$$

It has been shown that

$$-1 \leq \rho_{xy} \leq +1$$

The correlation coefficient is a direct measure of the correlation between x and y, however, ρ_{xy} is not a direct measure of statistical independence of x and y. Nonetheless, ρ_{xy} is often used as a measure of dependence between x and y. This is motivated by the following considerations.

Let $y = \pm ax$, where a is a positive constant, then $\mu_{xy} = \pm a\,\sigma_x^2$, $\sigma_y^2 = a^2\sigma_x^2$ and $\rho_{xy} = \pm 1$. Thus, if one random variable is completely determined by another then their correlation coefficient is ± 1. Moreover, if two random variables are statistically indepependent then their correlation coefficient is zero. It should be noted that the correlation coefficient is a direct measure of statistical correlation and only an indirect measure of statistically independence. However, the correlation coefficient can be considered a direct measure of functional dependence of two random variables.

On the other hand, if the two random variables are Gaussian, then zero correlation and statistical independence are equivalent. That is, if x and y are two Gaussian random variables and if $\mu_{xy} = \rho_{xy} = 0$, then x and y are statistically independent. In the case of Gaussian random variables an analysis of correlation is sufficient to measure both functional dependence and statistical independence. This can be seen from the joint probability density function for two Gaussian random variables.

In addition to the foregoing, if the conditional expectation of y, or x, given x, or y, is independent of x, or y, respectively, then $\mu_{xy} = \rho_{xy} = 0$. This can be shown as follows. Let $E(y/x) = C$, then,

$$C = \int_{-\infty}^{\infty} y \cdot f(y/x)\,dy = \int_{-\infty}^{\infty} y \frac{f(x,y)}{f(x)}\,dy$$

$$= \frac{1}{f(x)} \int_{-\infty}^{\infty} y\, f(x,y)\,dy$$

Thus,

$$C f(x) = \int_{-\infty}^{\infty} y\, f(x,y)\,dy$$

and

$$C\int_{-\infty}^{\infty} f(X)\,dX = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\,f(X,y)\,dy\,dX = E(y)$$

Therefore, C = E(y). Using this result μ_{xy} becomes

$$\mu_{xy} = E\left\{\left[X - E(X)\right]\left[y - E(y)\right]\right\}$$

$$= E(Xy) - CE(X)$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} Xy\,f(X,y)\,dx\,dy - C\,E(X)$$

$$= \int_{-\infty}^{+\infty} X f(X) \int_{-\infty}^{+\infty} y\,\frac{f(X,y)}{f(X)}\,dy\,dx - CE(X)$$

$$= \int_{-\infty}^{+\infty} X f(X) E(y/X)\,dx - CE(X)$$

$$= C\int_{-\infty}^{+\infty} X f(X)\,dX - C\,E(X)$$

$$= CE(X) - CE(X)$$

$$\mu_{xy} \equiv 0$$

The foregoing is used as a basis for correlation analysis. In general, the conditional expectation of y, or x, given x, or y, as a function of x, or y,

respectively, is referred to as the "regression" curve of y, or x, on x, or y, respectively. More explicitly, E(x/y) as a function of y is referred to as the regression curve of x on y. Similarly, E(y/x)as a function of x is referred to as the regression curve of y on x.

A special situation arises if x and y are Gaussian random variables. In this case, the conditional expectations are linear functions of the given random variable and, hence, the regression curves are linear in their arguments. This can be seen from the conditional probability density functions for Gaussian random variables. In Appendix B, the conditional expectations are given for a general Gaussian random vector. For the special case of two random variables, the results become

$$E(x/y) = m_x + \frac{\mu_{xy}}{\sigma_y^2}(y - m_y)$$
$$E(y/x) = m_y + \frac{\mu_{xy}}{\sigma_x^2}(x - m_x)$$

Alternatively,

$$E(x/y) = m_x + \frac{\sigma_x}{\sigma_y}\rho_{xy}(y - m_y)$$
$$E(y/x) = m_y + \frac{\sigma_y}{\sigma_x}\rho_{xy}(x - m_x)$$

It is seen that if $\rho_{xy} = 0$, then $E(x/y) = m_x$ and $E(y/x) = m_y$ which is the case of constant regression curves as noted above. On the other hand, if $\rho_{xy} \neq 0$, then the regression curves for x and y are linear. This is a particular case which is referred to as linear regression.

It is important to note that if two random variables, in general, have linear regression curves, then the coefficients are the same as those for the Gaussian case. That is, let x and y be two random variables such that

$$E(y/x) = Ax + B$$
$$E(x/y) = Cy + D$$

where A and C are referred to as regression coefficients of y on x and x on y, respectively. Then the regression coefficients are given by

$$A = \rho_{xy}\frac{\sigma_y}{\sigma_x}$$
$$B = \rho_{xy}\frac{\sigma_x}{\sigma_y}$$

This follows by simply using the conditional expectations to determine the total expectation of y and x and xy, i.e.,

$$E(y) = \int_{-\infty}^{+\infty} E(y/x)\, f(x)\, dx$$

$$= \int_{-\infty}^{+\infty} (Ax + B)\, f(x)\, dx$$

$$E(y) = AE(x) + B$$

$$E(x) = \int_{-\infty}^{+\infty} E(x/y)\, f(y)\, dy$$

$$= \int_{-\infty}^{+\infty} (Cy + D)\, f(y)\, dy$$

$$E(x) = CE(y) + D$$

$$E(xy) = \int_{-\infty}^{+\infty} E(y/x)\, x\, f(x)\, dx = \int_{-\infty}^{+\infty} E(x/y)\, y\, f(y)\, dy$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy\, f(y/x)\, f(x)\, dx\, dy$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy\, f(x/y)\, f(y)\, dy$$

$$E(xy) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy\, f(x,y)\, dx\, dy$$

Thus

$$E(xy) = \int_{-\infty}^{+\infty} (Ax+B)\, x f(x)\, dx = \int_{-\infty}^{+\infty} (Cy+D)\, y f(y)\, dy$$

$$E(xy) = AE(x^2) + BE(x) = \mu_{xy} + m_x m_y$$

$$E(xy) = CE(y^2) + DE(y) = \mu_{xy} + m_x m_y$$

It follows that

$$m_y = A m_x + B$$

$$\mu_{xy} + m_x m_y = A(\sigma_x^2 + m_x^2) + B m_x$$

$$m_x = C m_y + D$$

$$\mu_{xy} + m_x m_y = C(\sigma_y^2 + m_y^2) + D m_y$$

Solving these equations, it is found that $Bm_x = m_x m_y - A m_x^2$ and, hence, $\mu_{xy} = A\sigma_x^2$; therefore,

$$A = \frac{\mu_{xy}}{\sigma_x^2} = \rho_{xy} \frac{\sigma_y}{\sigma_x}$$

and

$$B = m_y - A m_x = m_y - \rho_{xy} \frac{\sigma_y}{\sigma_x} m_x$$

Similiarly,

$$C = \frac{\mu_{xy}}{\sigma_y^2} = \rho_{xy} \frac{\sigma_x}{\sigma_y}$$

and

$$D = m_x - \rho_{xy} \frac{\sigma_y}{\sigma_x} m_y$$

Thus, if for two random variables x and y their regression curves are linear, then,

$$E(y/x) = m_y + \rho_{xy}\frac{\sigma_y}{\sigma_x}(x - m_x)$$

$$E(x/y) = m_x + \rho_{xy}\frac{\sigma_x}{\sigma_y}(y - m_y)$$

Of primary concern in regression analyses of correlation is the possible deviation between observed values of random variables and their conditional expectations. That is, in a set of samples of y and x, consider the random variables δ_x and δ_y defined as follows:

$$\delta_y = y_x - E(y/x)$$
$$\delta_x = x_y - E(x/y)$$

where y_x and x_y denote corresponding pairs of the random variables x and y. It is easily seen that the expected values of δ_y and δ_x are zero, i.e.,

$$E(\delta_y) = E\left[E(\delta_y/x)\right] = 0$$

$$E(\delta_x) = E\left[E(\delta_x/y)\right] = 0$$

Thus, the variances of δ_y and δ_x are given by

$$\sigma^2(\delta_y) = E(\delta_y^2) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left[y_x - E(y/x)\right]^2 f(x,y)\,dx\,dy$$

$$\sigma^2(\delta_x) = E(\delta_x^2) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left[x_y - E(x/y)\right]^2 f(x,y)\,dx\,dy$$

Therefore,

$$\sigma^2(\delta_y) = E(y^2) - E\left[E^2(y/x)\right]$$
$$\sigma^2(\delta_x) = E(x^2) - E\left[E^2(x/y)\right]$$

For the case of linear regression, the variances of δ_y and δ_x become

$$\sigma^2(\delta y) = \sigma_y^2 (1-\rho_{xy}^2)$$
$$\sigma^2(\delta x) = \sigma_x^2 (1-\rho_{xy}^2)$$

In a regression analysis of correlation, it is required to estimate the regression parameters from a set of samples of the random variables x and y. This can be accomplished by the method of "least-squares" curve fitting in the following manner. Consider the case of linear regression between x and y wherein E(y/x) and E(x/y) are linear in x and y, respectively. In this case y and x can be written as follows:

$$y = E(y/x) + \delta y = \alpha_0 + \alpha_1 x + \delta y$$
$$x = E(x/y) + \delta x = \beta_0 + \beta_1 y + \delta x$$

where

$$\alpha_0 = m_y - \frac{\mu_{xy}}{\sigma_x^2} m_x$$
$$\alpha_1 = \frac{\mu_{xy}}{\sigma_x^2} = \rho_{xy}\left(\frac{\sigma_y}{\sigma_x}\right)$$
$$\beta_0 = m_x - \frac{\mu_{xy}}{\sigma_y^2} m_y$$
$$\beta_1 = \frac{\mu_{xy}}{\sigma_y^2} = \rho_{xy}\left(\frac{\sigma_x}{\sigma_y}\right)$$

For a set of samples $\underline{y}$ and $\underline{x}$, these equations become

$$\underline{y} = [\underline{1}, \underline{x}]\,\underline{\alpha} + \underline{\delta y} = A\,\underline{\alpha} + \underline{e}_1$$
$$\underline{x} = [\underline{1}, \underline{y}]\,\underline{\beta} + \underline{\delta x} = B\,\underline{\beta} + \underline{e}_2$$

where $\underline{\alpha}^T = (\alpha_0, \alpha_1)$ and $\underline{\beta}^T = (\beta_0, \beta_1)$. It is apparent that both $\underline{\alpha}$ and $\underline{\beta}$ can be estimated from the sample sets of y and x. However, these regression parameters are not independent and only $\underline{\alpha}$ or $\underline{\beta}$ are estimated using either equation. Consider the equation $\underline{y} = A\,\underline{\alpha} + \underline{\beta}$. The "least-squares" estimate for $\underline{\alpha}$, $\underline{\hat{\alpha}}$, is given by (see References 5 or 15)

$$\hat{\underline{\alpha}} = (A^T A)^{-1} A^T \underline{y}$$

Therefore,

$$\hat{\underline{\alpha}} = \left\{ \begin{bmatrix} \underline{1}^T \\ \underline{X}^T \end{bmatrix} \begin{bmatrix} \underline{1}, \underline{X} \end{bmatrix} \right\}^{-1} \begin{bmatrix} \underline{1}^T \\ \underline{X}^T \end{bmatrix} \underline{y} = \begin{bmatrix} \underline{1}^T\underline{1} & \underline{1}^T\underline{X} \\ \underline{X}^T\underline{1} & \underline{X}^T\underline{X} \end{bmatrix}^{-1} \begin{bmatrix} \underline{1}^T\underline{y} \\ \underline{X}^T\underline{y} \end{bmatrix}$$

It is easily shown that $\hat{\underline{\alpha}}$ is an unbiased estimate of $\underline{\alpha}$, i.e., $E(\hat{\underline{\alpha}}) = \underline{\alpha}$. This follows from

$$\begin{aligned} E(\hat{\underline{\alpha}}) &= E\left[(A^T A)^{-1} A^T \underline{y} \right] \\ &= (A^T A)^{-1} A^T E(\underline{y}) \\ &= (A^T A)^{-1} A^T E(A\underline{\alpha} + \underline{\epsilon}_1) \\ E(\hat{\underline{\alpha}}) &= \underline{\alpha} \end{aligned}$$

Thus, the error $\underline{\epsilon} = (\hat{\underline{\alpha}} - \underline{\alpha})$ in estimating $\underline{\alpha}$ can be assessed in terms of the covariance matrix of $\underline{\alpha}$, i.e.,

$$\Gamma_\epsilon = E\left[\underline{\epsilon}\, \underline{\epsilon}^T \right] = E\left[\hat{\underline{\alpha}}\, \hat{\underline{\alpha}}^T \right]$$

The covariance matrix for a "least-squares" estimate is given by

$$\Gamma_\epsilon = \Gamma_\alpha = (A^T A)^{-1} (A \Gamma_e A) (A^T A)^{-1}$$

For the case of $\Gamma_\epsilon = \sigma_e^2$ I, Γ_ϵ becomes

$$\Gamma_\epsilon = \Gamma_\alpha = \sigma_e^2 (A^T A)^{-1}$$

$$\Gamma_\epsilon = \begin{bmatrix} \underline{1}^T\underline{1} & \underline{1}^T\underline{X} \\ \underline{X}^T\underline{1} & \underline{X}^T\underline{X} \end{bmatrix}^{-1} \sigma_e^2$$

It should be noted that $\sigma_e^2 = \sigma^2(\delta_y) = \sigma_y^2 (1 - \rho_{xy}^2)$.

The second central moments of the sample sets $\underline{x}$ and $\underline{y}$ can be used to estimate the central moment μ_{xy} and the correlation coefficient ρ_{xy} . Let s_x and s_y be the first sample moments of $\underline{x}$ and $\underline{y}$ defined as follows:

$$s_x = \frac{1}{n} \underline{1}^T \underline{x}$$

$$s_y = \frac{1}{n} \underline{1}^T \underline{y}$$

Similarly, let the second central sample moments Δ_x^2 , Δ_y^2 and Δ_{xy} be defined as follows:

$$\Delta_x^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - s_x)^2$$

$$\Delta_y^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - s_y)^2$$

$$\Delta_{xy} = \frac{1}{n} \sum_{i=1}^{n} (x_i - s_x)(y_i - s_y)$$

The sample correlation coefficient, r, is defined as follows

$$r = \frac{\Delta_{xy}}{\Delta_x \Delta_y}$$

$$= \frac{\sum_{i=1}^{n} (x_i - s_x)(y_i - s_y)}{\sqrt{\left[\sum_{i=1}^{n} (x_i - s_x)^2\right]\left[\sum_{i=1}^{n} (y_i - s_y)^2\right]}}$$

By an analysis similar to that of the previous section, it can be shown that Δ_{xy} is a biased-consistent estimate for μ_{xy} , i.e.,

$$E(\Delta_{xy}) = \left(\frac{n-1}{n}\right) \mu_{xy}$$

$$\lim_{n \to \infty} \sigma^2(\Delta_{xy}) = 0$$

Thus, an unbiased-consistent estimate for μ_{xy} becomes

$$\hat{\mu}_{xy} = \left(\frac{n}{n-1}\right)\Delta_{xy}$$

Based upon the consistency of Δ_x^2, Δ_y^2 and Δ_{xy}, it follows that

$$\lim_{n\to\infty}(r) = \lim_{n\to\infty}\left(\frac{\Delta_{xy}}{\Delta_x \Delta_y}\right)$$

$$= \frac{\lim_{n\to\infty}(\Delta_{xy})}{\left[\lim_{n\to\infty}(\Delta_x)\right]\left[\lim_{n\to\infty}(\Delta_y)\right]}$$

$$\lim_{n\to\infty}(r) = \rho_{xy}$$

Thus, an estimate of ρ_{xy} is the sample correlation coefficient r , i.e.,

$$\hat{\rho}_{xy} = r$$

The use of Δ_{xy} and r as estimates of μ_{xy} and ρ_{xy} should be considered on the basis of their accuracy. This can be accomplished in the case of Gaussian random variable, which is considered in a following section.

The general problem of regression and correlation analysis involves more than two random variables as considered above. However, the methods of analysis are effectively equivalent with appropriate extensions. The methods for more than two random variables are discussed in appropriate detail in References 1, 5, 6, and 11.

2.2.7.1.3. Confidence Intervals

Of primary concern in estimating moments from sample sets is an assessment of the accuracy in the resulting estimate. Alternatively, it is of concern to determine a sufficient sample size in order to assure that an estimate of a statistical moment possesses a required accuracy. In general, an estimate based upon a set of samples of a random variable is also a random variable. Thus, the accuracy of such an estimate must be specified in terms of two entities which are: (1) a region which will bound the estimate or error in the estimate; and (2) the probability that the estimate or error will lie within the stated

region. These two entities are usually stated as a "confidence interval" which contains a stated probability and a description of a region. In general terms, a confidence interval is a bound placed upon the error in an estimate in terms of a region and the probability that the error will be contained within the region.

In general, a particular estimate of a statistical moment which is based upon a sample set is referred to as a "point" estimate of the moment of the population. The point estimate is, in general, not very meaningful unless an assessment of the possible error in the point estimate is made. If a point estimate is to be useful, it should be specified in terms of some interval about the moment being estimated such that the true value of the moment will be within the interval with a specified probability. This is the purpose of a confidence interval. In order to have meaning a confidence interval must have a probability associated with the interval given. It is usually desirable to have a small confidence interval with a high probability that the interval will contain the moment being estimated. This is equivalent to an estimate with a high degree of accuracy. However, it is characteristic of estimates which are functions of random sample sets that the confidence interval and the associated probability cannot be stated arbitrarily. Usually, for a given sample size and population characteristics, the smaller the confidence interval the lower is the probability that the moment being estimated will lie in the interval. There exist two extremes for confidence intervals. One is that the modulus of the error in an estimate will lie somewhere between zero and infinity with probability one. The other is that the error in an estimate will be infinitesimally small with vanishing probability. These two extreme confidence intervals are generally true, but they are rather meaningless since they convey little useful information. It is seen that confidence intervals are not unique and they possess various degrees of meaning depending upon the information conveyed. The most meaningful confidence interval is not explicitly defined for all situations. An implicit definition of a useful or meaningful confidence interval depends upon the particular application that the estimate is used for. In general terms, a useful confidence interval is one which determines the probability that the error in an estimate will be contained within a required bound.

A confidence interval can be given for any unbiased estimate which has a finite variance. This follows directly from the Tchebycheff Inequality which states that

$$P_{ROB}\left[|\epsilon| < K\sigma_\epsilon\right] > 1 - \frac{1}{K^2}$$

where ϵ is the error in an estimate. It is seen that the higher the probability that $|\epsilon|$ will lie within the interval $k\sigma_\epsilon$, the larger the interval. Of course, if σ_ϵ is sufficiently small, then the interval $k\sigma_\epsilon$ can be an adequate assurance of the required accuracy of the estimate. Consider the case of estimating the mean from a set of uncorrelated samples. The variance for the sample mean is $1/n\sigma^2$. By taking $k^2 = 10$, it is found that the probability is at least 0.9, or 90 percent, that

$$\epsilon^2 < \frac{10}{n}\sigma^2$$

Alternatively, it could be stated that the probability that ϵ^2 will exceed $(10/n)\sigma^2$ is less than 0.1, or ten percent. The quantity $(10/n)\sigma^2$ is essentially an estimate error bound, however, this bound can be used to determine an interval in which the true value of the population mean should lie. This is, $\epsilon^2 = (s-m)^2$, where s and m are the sample and population means, respectively, thus if $\epsilon^2 < (k^2/n)\sigma^2$ then $(s-m)^2 < (k^2/n)\sigma^2$, or $|s-m| < (k/\sqrt{n})\sigma$ and $[s-(k/\sqrt{n})\sigma] < m < [s+(k/\sqrt{n})\sigma]$. Therefore,

$$P_{ROB}\left[(s-\frac{k}{\sqrt{n}}\sigma) < m < (s+\frac{k}{\sqrt{n}}\sigma)\right] > 1-\frac{1}{k^2}$$

If $k^2 = 10$, then

$$P_{ROB}\left[(s-\sqrt{10/n}\,\sigma) < m < (s+\sqrt{10/n}\,\sigma)\right] > 0.9$$

The interval from $(s-\sqrt{10/n}\,\sigma)$ to $(s+\sqrt{10/n}\,\sigma)$ should contain the population mean m with a probability of 0.9. Thus, for each particular estimate of m, given by the sample mean s, it is possible to state an interval which contains the true value of m with a probability of 0.9. Of course, other intervals exist for other specified probabilities. In this particular case the interval from $(s-\sqrt{10/n}\,\sigma)$ to $(s+\sqrt{10/n}\,\sigma)$ is the confidence interval and the probability of 0.9 is usually referred to as the confidence coefficient.

In determining a confidence interval all of the available information should be utilized in order to obtain the most meaningful confidence interval possible. The use of the Tchebycheff Inequality essentially ignores any information concerning the probability density function of the population or of the estimate itself, hence, a confidence interval derived therefrom is usually quite conservative. That is, the confidence interval for a stated probability is larger than that which is obtained if information concerning the estimate probability density function is used. Consider the case of estimating the mean of a Gaussian random variable. In this case the sample mean s is also a Gaussian random variable with mean and variance given by

$$E(S) = m$$

$$\sigma_S^2 = \frac{1}{n}\sigma$$

where m and σ^2 are the mean and variance of the population, respectively. Now, for the Gaussian probability density function, it is found that (see Appendix C)

$$0.9 = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{m-1.645\sigma}^{m+1.645\sigma} exp - \frac{1}{2\sigma^2}(x-m)^2 dx$$

where m and σ^2 are the mean and variance of x, respectively. Using this result for the sample mean s, it is found that

$$P_{ROB}\left[|S-m| < \frac{1.645}{\sqrt{n}}\sigma\right] = 0.9$$

Therefore,

$$P_{ROB}\left[\left(S - 1.645\frac{1}{\sqrt{n}}\sigma\right) < m < \left(S + 1.645\frac{1}{\sqrt{n}}\sigma\right)\right] = 0.9$$

The confidence interval for 0.9 probability is seen to be significantly smaller than that which would be obtained using the Tchebycheff Inequality. Thus, the confidence interval obtained using knowledge of the probability density function of the estimate is essentially more meaningful than that obtained from the Tchebycheff Inequality in that it contains more precise information concerning the accuracy of the estimate in terms of specifying the true value of the moment being estimated.

In general, a confidence interval can be regarded as a statement of the degree of certainty which is contained in a statistical inference. It is always desirable to determine the smallest confidence interval for a particular probability since this tends to give the most precise information concerning the uncertainty in the inference. In general, this requires use of the probability density function of the estimate when it is available. The construction and use of confidence intervals is discussed in further detail in References 5, 6, and 11.

2.2.7.1.4 Estimating Moments for Gaussian Random Variables

If x and y are two Gaussian random variables, then there exists two first moments and three second-central moments which specify the joint probability function of x and y, f(x, y). These moments are as follows:

$$E(x) = m_x$$

$$E(y) = m_y$$

$$E(x - m_x)^2 = \sigma_x^2$$

$$E(y - m_y)^2 = \sigma_y^2$$

$$E\left[(x - m_x)(y - m_y)\right] = \mu_{xy} = \rho_{xy}\,\sigma_x\,\sigma_y$$

Two sets of independent sample sets are usually used to estimate the moments which specify f(x, y). Let the vectors $\underline{x}$ and $\underline{y}$ denote independent sample sets of the random variables x and y. The following estimates are used for the moments of f(x, y).

$$\hat{m}_x = S_x = \frac{1}{n}\,(\underline{1}^T \underline{x})$$

$$\hat{m}_y = S_y = \frac{1}{n}\,(\underline{1}^T \underline{y})$$

$$\hat{\sigma}_x^2 = \left(\frac{n}{n-1}\right)\Delta_x^2 = \left(\frac{1}{n-1}\right)(\underline{v}_1^T \underline{v}_1)$$

$$\hat{\sigma}_y^2 = \left(\frac{n}{n-1}\right)\Delta_y^2 = \left(\frac{1}{n-1}\right)(\underline{v}_2^T \underline{v}_2)$$

$$\hat{\mu}_{xy} = \left(\frac{n}{n-1}\right)\Delta_{xy} = \left(\frac{1}{n-1}\right)(\underline{v}_1^T \underline{v}_2)$$

$$\hat{\rho}_{xy} = \frac{\Delta_{xy}}{\Delta_x \Delta_y} = \frac{\underline{v}_1^T \underline{v}_2}{\sqrt{(\underline{v}_1^T \underline{v}_1)(\underline{v}_2^T \underline{v}_2)}}$$

where $\underline{V}_1 = \underline{x} - \underline{1}\, s_x$ and $\underline{V}_2 = \underline{y} - \underline{1}\, s_y$. The accuracy of these estimates can be assessed by considering the probability density functions of the estimates.

Although the variances σ_x^2 and σ_y^2 are unknown, it is possible to obtain confidence intervals for the mean values of x and y as a function of the sample data sets $\underline{x}$ and $\underline{y}$. This is accomplished by showing that for a Gaussian random variable the sample mean and variance for an independent sample set are statistically independent and the ratio of (s-m) to $\sqrt{\underline{V}^T\underline{V}/n(n-1)}$ has a Student's probability density function (see Section 2.2.4.5.3), which can be used to determine a confidence interval for m. That is, if

$$t = \frac{s-m}{\sqrt{(\underline{V}^T\underline{V})/(n(n-1))}} = \frac{\sqrt{n}\,(s-m)}{\sqrt{\left(\frac{1}{n-1}\right)\underline{V}^T\underline{V}}}$$

then

$$P_{ROB}\left[s - t_1\sqrt{\frac{\underline{V}^T\underline{V}}{n(n-1)}} < m < s + t_2\sqrt{\frac{\underline{V}^T\underline{V}}{n(n-1)}}\right] =$$

$$P_{ROB}\left[-t_1 \le t \le t_2\right]$$

Now, if the probability density function of t can be determined, then a confidence interval for m can be determined in terms of the sample set $\underline{x}$ only, i.e., the population variance is not needed. It is possible to show that t has the Student's pdf in the following manner. First, t can be written as follows:

$$t = \frac{\frac{\sqrt{n}}{\sigma}(s-m)}{\sqrt{\frac{1}{\sigma^2(n-1)}(\underline{V}^T\underline{V})}} = \frac{U}{\sqrt{\left(\frac{1}{n-1}\right)V^2}}$$

where

$$U = \frac{\sqrt{n}}{\sigma}(s-m)$$

and

$$V^2 = \frac{1}{\sigma^2}\underline{V}^T\underline{V}$$

Second, it is apparent that U is a normal random variable. Third, it is necessary to show that the pdf of V^2 is χ^2 (see Section 2.2.4.5.2) and is independent of U. This step is accomplished by considering an orthorgonal transformation of the sample set $\underline{x}$, i.e., let $\underline{z} = C(\underline{x} - m\underline{1})$ where

$$C = \frac{1}{\sigma}\begin{bmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ \frac{1}{\sqrt{i(i-1)}} & \cdots & \frac{1-i}{\sqrt{i(i-1)}} & & 0 \\ \vdots & & & & \vdots \\ \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \cdots & & \frac{1-n}{\sqrt{n(n-1)}} \end{bmatrix}$$

It is not difficult to show that $C^TC = 1/\sigma^2 I$, therefore, $\underline{z}$ is a Normal random vector, i.e., $E(\underline{z}) = \underline{0}$ and $\Gamma_z = E(\underline{z}\,\underline{z}^T) = I$. It is apparent that $z_1 = U$, i.e.,

$$z_1 = \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n}(x_i - m) = \frac{\sqrt{n}}{\sigma}\left[\frac{1}{n}\right]\sum_{i=1}^{n}(x_i - m)$$

$$z_1 = \frac{\sqrt{n}}{\sigma}(s - m) = U$$

Now, it can be shown that $V^2 = 1/\sigma^2\ \underline{V}^T\underline{V}$ is not a function of $\underline{z}_1$. Consider

$$V^2 = \frac{1}{\sigma^2}\ \underline{V}^T\underline{V} = \frac{1}{\sigma^2}\sum_{i=1}^{n} V_i^2$$

$$= \sum_{i=1}^{n}\left(\frac{x_i - s}{\sigma}\right)^2 = \sum_{i=1}^{n}\left(\frac{x_i - m - s + m}{\sigma}\right)^2$$

$$= \sum_{i=1}^{n}\left(\frac{x_i - m}{\sigma} - \frac{s - m}{\sigma}\right)^2$$

$$= \sum_{i=1}^{n} \left(\frac{x_i - m}{\sigma} - \frac{1}{\sqrt{n}} U \right)^2$$

$$= \sum_{i=1}^{n} \left(\frac{x_i - m}{\sigma} \right)^2 - \frac{2U}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{x_i - m}{\sigma} \right) + \frac{1}{n} U^2$$

$$= \frac{1}{\sigma^2} (\underline{X} - m\underline{1})^T (\underline{X} - m\underline{1}) - z_1^2$$

$$= \underline{z}^T \underline{z} - z_1^2$$

$$V^2 = \sum_{i=2}^{n} z_i^2$$

Therefore, V^2 is only a function of $z_2, z_3, \ldots, z_n$. It now becomes apparent that U and V^2 are statistically independent and that V^2 is a Chi-square random variable with (n - 1) degrees of freedom. This follows from the fact that z_i is a normal random variable for i = 1, 2, ..., n.

Based on the foregoing, it is possible to determine a confidence interval for the mean of a Gaussian random variable without knowledge of the variance of the random variable. The confidence interval is

$$s - t_1 \sqrt{\frac{\underline{v}^T \underline{v}}{n(n-1)}} < m < s + t_2 \sqrt{\frac{\underline{v}^T \underline{v}}{n(n-1)}}$$

where t_1 and t_2 are obtained from the pdf of "t" for (n - 1) degrees of freedom. The values of t_1 and t_2 are selected for the particular confidence coefficient or probability desired. Thus, if the sample mean of an independent Gaussian random variable sample set $\underline{x}$ is used to estimate the population mean, then a confidence interval can be determined in terms of only the sample set $\underline{x}$.

In a similar manner a confidence interval for the variance of a Gaussian random variable can be determined in terms of the sample set $\underline{x}$. This follows directly from the fact that $V^2 = 1/\sigma^2 \, \underline{V}^T\underline{V}$ has a Chi-square pdf with (n - 1) degrees of freedom, therefore,

$$P_{ROB}\left[\chi_1^2 < \frac{1}{\sigma^2}\,\underline{V}^T\underline{V} < \chi_2^2\right] = P_{ROB}\left[\chi_1^2 < \chi^2 < \chi_2^2\right]$$

Alternatively,

$$P_{ROB}\left[\frac{\underline{V}^T\underline{V}}{\chi_2^2} < \sigma^2 < \frac{\underline{V}^T\underline{V}}{\chi_1^2}\right] = P_{ROB}\left[\chi_1^2 < \chi^2 < \chi_2^2\right]$$

It is apparent that a confidence interval can be determined for a probability PROB $[\chi_1^2 < \chi^2 < \chi_2^2]$ by simply finding two values of χ_1^2 and χ_2^2 which bound the probability for a Chi-square random variable with (n - 1) degrees of freedom.

It is possible to determine the joint probability density function of the sample moments Δ_x^2, Δ_y^2, and Δ_{xy} (see Reference 1). The result is

$$f(\Delta_x^2, \Delta_y^2, \Delta_{xy}) = \frac{n^{n-1}\left[\Delta_x^2\Delta_y^2 - \Delta_{xy}^2\right]^{n/2-2}}{4\pi(n-3)!\,|M|^{\frac{n-1}{2}}} \exp -\frac{n}{2|M|}\left[\sigma_x^2\Delta_x^2 - 2\mu_{xy}\Delta_{xy} + \sigma_y^2\Delta_y^2\right]$$

where

$$\Delta_x^2, \Delta_y^2 > 0 \quad AND \quad \Delta_{xy}^2 < \Delta_x^2\Delta_y^2 \quad AND \quad |M| = \sigma_x^2\sigma_y^2(1-\rho_{xy}^2)$$

The joint probability density function for the sample moments involves the moments of x and y, which means that $f(\Delta_x^2, \Delta_y^2, \Delta_{xy})$ cannot be directly used to determine confidence intervals for the central moments of x and y. However, using $f(\Delta_x^2, \Delta_y^2, \Delta_{xy})$, it is possible to determine the probability density functions of certain functions of the sample moments which result in confidence intervals. Consider the following functions of the sample moment Δ_x^2, Δ_y^2 and Δ_{xy}.

$$u = \Delta_x^2$$

$$v = \frac{\Delta_x^2 \Delta_y^2 - \Delta_{xy}^2}{\Delta_x^2}$$

$$w = \Delta_x \left(\frac{\Delta_{xy}}{\Delta_x^2} - \frac{\mu_{xy}}{\sigma_x^2} \right)$$

Now, the joint pdf of u, v, and w can be found by a transformation of variables (see Reference 6). The results are

$$f(u,v,w) = K \left[u^{\frac{n-3}{2}} e^{-\frac{nu}{2\sigma_x^2}} \right] \left[v^{\frac{n-4}{2}} e^{-\frac{nv}{2\sigma_y^2(1-\rho_{xy}^2)}} \right] \left[e^{-\frac{nw^2}{2\sigma_y^2(1-\rho_{xy}^2)}} \right]$$

where $u, v > 0$ and

$$K = \frac{n^{n-1}}{4\pi (n-3)! \left[\sigma_x^2 \sigma_y^2 (1-\rho_{xy}^2)\right]^{\frac{n-1}{2}}}$$

Thus, u, v, and w are statistically independent, i.e., f(u, v, w) can be written as follows

$$f(u,v,w) = f(u)\, f(v)\, f(w)$$

where

$$f(u) = K_1 u^{\frac{n-3}{2}} e^{-\frac{nu}{2\sigma_x^2}} U(u)$$

$$f(v) = K_2 v^{\frac{n-4}{2}} e^{-\frac{nv}{2\sigma_y^2(1-\rho_{xy})}} U(v)$$

$$f(w) = K_3 \; e^{-\frac{n w^2}{2\sigma_y^2 (1-\rho_{xy}^2)}}$$

$$K_1 \; K_2 \; K_3 = K$$

By taking $K_3 = [2\Pi \; \sigma_w^2]^{-1/2}$ where $\sigma_w^2 = 1/n \; \sigma_y^2(1 - \rho_{xy}^2)$, it is seen that w is a Gaussian random variable with variance σ_w^2 and mean value of zero. Similiarly, by an appropriate selection of K_2 it is seen that $nv[\sigma_y^2 (1 - \rho_{xy}^2)]^{-1}$ is a Chi-square random variable with (n-2) degrees of freedom. Also, nu/σ_x^2 is found to have a Chi-square pdf with (n-1) degrees of freedom. This follows from the following factoring of K.

$$K = \frac{n^{n-1}}{4\pi (n-3)! [\sigma_x^2 \sigma_y^2 (1-\rho_{xy}^2)]^{\frac{n-1}{2}}}$$

$$= \frac{\left(n^{\frac{n-1}{2}}\right)\left(n^{\frac{n-1}{2}}\right)\sqrt{\sigma_w^2}}{4\pi (n-3)! (\sigma_x^2)^{\frac{n-1}{2}} [\sigma_y^2 (1-\rho_{xy}^2)]^{\frac{n-1}{2}} \sqrt{\sigma_w^2}}$$

$$= \frac{\left(n^{\frac{n-1}{2}}\right)\left(n^{\frac{n-2}{2}}\right)}{2^{3/2}\sqrt{\pi}\,(n-3)! (\sigma_x^2)^{\frac{n-1}{2}} [\sigma_y^2 (1-\rho_{xy}^2)]^{\frac{n-2}{2}}} \left(\frac{1}{\sqrt{2\pi\,\sigma_w^2}}\right)$$

$$K = \frac{\left(n^{\frac{n-1}{2}}\right)\left(n^{\frac{n-2}{2}}\right)}{2^{3/2}\sqrt{\pi}\,(n-3)! (\sigma_x^2)^{\frac{n-1}{2}} [\sigma_y^2 (1-\rho_{xy}^2)]^{\frac{n-2}{2}}} \; K_3$$

Thus,

$$K_3 = \frac{1}{\sqrt{\frac{2\pi}{n}\,\sigma_y^2 (1-\rho_{xy}^2)}}$$

and

$$K_1 K_2 = \frac{\left(n\,\frac{n-1}{2}\right)\left(n\,\frac{n-2}{2}\right)}{2^{3/2}\sqrt{\pi}\,(n-3)!\,(\sigma_x^2)^{\frac{n-1}{2}}\left[\sigma_y^2(1-\rho_{xy}^2)\right]^{\frac{n-2}{2}}}$$

Since $(n-3)! = \Gamma(n-2)$ and $\sqrt{\pi}\,\Gamma(2n) = 2^{2n-1}\,\Gamma(n)\,\Gamma(n+\frac{1}{2})$ (see Section 2.2.3.2), it follows that

$$2^{3/2}\sqrt{\pi}\,(n-3)! = 2^{3/2}\sqrt{\pi}\,\Gamma(n-2) = (\sqrt{2})^{2n-3}\,\Gamma\left(\frac{n}{2}-1\right)\Gamma\left(\frac{n-1}{2}\right)$$

Therefore,

$$K_1 K_2 = \frac{1}{(\sqrt{2})^{2n-3}\,\Gamma\left(\frac{n-2}{2}\right)\Gamma\left(\frac{n-1}{2}\right)\left(\frac{1}{n}\sigma_x^2\right)^{\frac{n-1}{2}}\left[\frac{1}{n}\sigma_y^2(1-\rho_{xy}^2)\right]^{\frac{n-2}{2}}}$$

$$K_1 K_2 = \left[\frac{1}{(\sqrt{2})^{n-1}\,\Gamma\left(\frac{n-1}{2}\right)\left(\frac{1}{n}\sigma_x^2\right)^{\frac{n-1}{2}}}\right]\left[\frac{1}{(\sqrt{2})^{n-2}\,\Gamma\left(\frac{n-2}{2}\right)\left[\frac{1}{n}\sigma_y^2(1-\rho_{xy}^2)\right]^{\frac{n-2}{2}}}\right]$$

and

$$K_1 = \frac{1}{(\sqrt{2})^{n-1}\,\Gamma\left(\frac{n-1}{2}\right)\left(\frac{1}{n}\sigma_x^2\right)^{\frac{n-1}{2}}}$$

$$K_2 = \frac{1}{(\sqrt{2})^{n-2}\,\Gamma\left(\frac{n-2}{n}\right)\left[\frac{1}{n}\sigma_y^2(1-\rho_{xy}^2)\right]^{\frac{n-2}{2}}}$$

Thus, the pdfs for u, v, and w become

$$f(u) = \frac{u^{\frac{n-3}{2}}}{\Gamma\left(\frac{n-1}{2}\right)(\sqrt{2})^{n-1}\left(\frac{1}{n}\sigma_x^2\right)^{\frac{n-1}{2}}}\,e^{-\frac{n\,u}{2\sigma_x^2}}\,U(u)$$

$$f(v) = \frac{v^{\frac{n-4}{2}}}{\Gamma\left(\frac{n-2}{2}\right)\left(\sqrt{2}\right)^{n-2}\left[\frac{1}{n}\sigma_y^2(1-\rho_{xy}^2)\right]^{\frac{n-2}{2}}} e^{-\frac{nv}{2\sigma_y^2(1-\rho_{xy}^2)}} U(v)$$

$$f(w) = \frac{1}{\sqrt{2\pi\,\frac{1}{n}\sigma_y^2(1-\rho_{xy}^2)}} e^{-\frac{nw^2}{2\sigma_y^2(1-\rho_{xy}^2)}}$$

Now, with the following simple changes of variables, the following probability density functions are determined:

$$V^2 = \frac{nu}{\sigma_x^2}$$

$$R^2 = \frac{nv}{\sigma_y^2(1-\rho_{xy}^2)}$$

$$W = \frac{w}{\sqrt{\frac{1}{n}\sigma_y^2(1-\rho_{xy}^2)}}$$

$$f(V^2) = \frac{(V^2)^{\frac{n-1}{2}-1}}{\Gamma\left(\frac{n-1}{2}\right)\left(\sqrt{2}\right)^{n-1}} e^{-1/2\,V^2}$$

$$f(R^2) = \frac{(R^2)^{\frac{n-2}{2}-1}}{\Gamma\left(\frac{n-2}{2}\right)\left(\sqrt{2}\right)^{n-2}} e^{-1/2\,R^2}$$

$$f(W) = \frac{1}{\sqrt{2\pi}} e^{-1/2\,W^2}$$

Thus, it is seen that V^2 and R^2 are Chi-square random variables with (n-1) and (n-2) degrees of freedom, respectively, and W is a Normal random variable. Of course, this result for V^2 agrees with that previously obtained. Now, since W and R^2 are independent with W being Normal and R^2 being Chi-square with (n-2) degrees of freedom, it follows that the pdf of $W/\sqrt{R^2\ 1/n-2}$ is the Student's pdf, i.e., the pdf of t is a Student's pdf with (n-2) degrees of freedom where

$$t = \frac{w}{\sqrt{\left(\frac{1}{n-2}\right) R^2}}$$

$$= \frac{w}{\sqrt{\frac{1}{n}\sigma_y^2 (1-\rho_{xy}^2)}} \Big/ \sqrt{\left(\frac{1}{n-2}\right) \frac{n v}{\sigma_y^2 (1-\rho_{xy}^2)}}$$

$$= \frac{w}{\sqrt{\left(\frac{1}{n-2}\right) v}}$$

$$= \frac{\Delta_x \left(\frac{\Delta_{xy}}{\Delta_x^2} - \frac{\mu_{xy}}{\sigma_x^2}\right)}{\sqrt{\left(\frac{1}{n-2}\right) \frac{1}{\Delta_x^2}\left(\Delta_x^2 \Delta_y^2 - \Delta_{xy}^2\right)}}$$

$$= \sqrt{\frac{\Delta_x^4 (n-2)}{(\Delta_x^2 \Delta_y^2 - \Delta_{xy}^2)}} \left(\frac{\Delta_{xy}}{\Delta_x^2} - \frac{\mu_{xy}}{\sigma_x^2}\right)$$

$$t = \sqrt{\left(\frac{\Delta_x^2}{\Delta_y^2}\right) \cdot \frac{(n-2)}{(1-r^2)}} \left(\frac{\Delta_{xy}}{\Delta_x^2} - \frac{\mu_{xy}}{\sigma_x^2}\right)$$

where

$$r = \frac{\Delta_{xy}}{\Delta_x \Delta_y}$$

Using the Student's pdf, it is possible to determine a confidence interval for μ_{xy}/σ_x^2 in terms of the sample moments only. The procedure is similar to that used before for the mean m. The resulting confidence interval is as follows:

$$\frac{\Delta_{xy}}{\Delta_x^2} - t_1\sqrt{\left(\frac{\Delta_y^2}{\Delta_x^2}\right)\frac{(1-r^2)}{(n-2)}} < \frac{\mu_{xy}}{\sigma_x^2} < \frac{\Delta_{xy}}{\Delta_x^2} + t_2\sqrt{\left(\frac{\Delta_y^2}{\Delta_x^2}\right)\frac{(1-r^2)}{(n-2)}}$$

The confidence coefficient is, of course, dependent upon the values of t_1 and t_2 and it is equal to PROB $[-t_1 \le t \le t_2]$.

The probability density function for the sample correlation coefficient r can be determined from the joint pdf of Δ_x^2, Δ_y^2, Δ_{xy}. The procedure is to first determine the joint pdf of Δ_x^2, Δ_y^2, and r by a change of variable and then determine the marginal pdf for r. (see Reference 6). The result is as follows:

$$f(r) = \frac{2^{n-3}(1-\rho_{xy}^2)^{\frac{n-1}{2}}}{\pi(n-3)!}(1-r^2)^{\frac{n-4}{2}}\sum_{k=0}^{\infty}\frac{1}{k!}(2\rho r)^k\Gamma^2\left(\frac{n+k-1}{2}\right)$$

The probability density function for r is seen to be only a function of the correlation coefficient ρ_{xy} and is independent of the other moments of x and y. Unfortunately, f(r) does not give a direct measure of the accuracy in r as an estimate for ρ_{xy} since f(r) is a function of ρ_{xy}. However, an important use for f(r) is to test for statistical independence of x and y which is considered in a latter section. Thus, if $\rho_{xy} = 0$, then f(r) becomes

$$f(r) = \frac{2^{n-3}\Gamma^2\left(\frac{n-1}{2}\right)}{\pi(n-3)!}(1-r^2)^{\frac{n-4}{2}}$$

2.2.7.2 Hypothesis Testing

In determining statistical properties of random variables, two basic assumptions are often made concerning statistical independence of random variables and the type of probability density function that a random variable possesses. In general, these assumptions can have a significant effect upon results and the validity of such assumptions should be assessed. It is possible to make an assessment of the validity of such assumptions by methods of hypothesis testing. In such methods assumptions are treated as hypotheses which are

accepted or rejected depending upon the outcome of certain tests which are made on sets of sample data. It is apparent that the tests must be designed to yield information concerning the hypotheses being tested and criteria of acceptance or rejection must be defined. In general, the tests are functions of samples of random variables and the results of such tests are also random variables; thus, there always exists some degree of uncertainty concerning the acceptance or rejection of a hypothesis based upon criteria of tests on samples of random variables. Therefore, it is necessary to specify a measure of accuracy in testing hypotheses. This measure of accuracy in hypotheses testing is usually referred to as the "level-of-significance" and it is, in general terms, the probability of being wrong in the decision of rejecting or accepting the hypothesis being tested. That is, the level of significance is either the probability of accepting a hypothesis which is false or rejecting a hypothesis which is true. Usually, the lower the level-of-significance, the better is the test of the hypothesis.

The method of hypothesis testing can be described in the following manner: There exists a hypothesis, denoted by H, concerning a random variable x, e.g., the hypothesis could be "the expected value of x is zero," which is usually denoted by

$$H: E(x) = 0$$

Let $\underline{x}$ denote a set of samples of the random variable x. Now certain properties of the sample set should be dependent upon the hypothesis H. Thus, it should be possible to design a test on $\underline{x}$, denoted by $T(\underline{x})$, which should be dependent upon the validity of H and certain results of the test would indicate that H should be accepted and certain results would indicate that H should be rejected. For example, if the hypothesis concerns the expected value of the random variable x, then the test could simply be the sample mean, i.e.,

$$T(x) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Obviously, if $E(x) = 0$ it is not expected that $T(\underline{x}) = 0$, i.e., $T(\underline{x})$ is a random variable with expected value equal to that of x and variance of $1/n\,\sigma_x^2$ for an uncorrelated sample set. It is apparent that the result of $T(\underline{x})$ is dependent upon the expected value of x, and if the hypothesis that $E(x) = 0$ is true, then certain results are expected for $T(\underline{X})$, whereas if H is false, then other results are expected for $T(\underline{x})$. For example, if x is a Gaussian random variable with $\sigma_x^2 = 1$ and $E(x) = 0$ and if $n = 9$, then $T(\underline{x})$ should lie within $\pm$ 1.0 with a probability of 0.9974, or 99.74 percent of the time (see Appendix C). Also, $T(\underline{x})$ should lie within $\pm$ 0.6533 with a probability of 0.95, or 95 percent of the time. That is, if it is true that $E(x) = 0$, then $|T(\underline{x})|$ will exceed 0.6533 with only a probability of 0.05. Thus, if it is found that $T(\underline{x}) > 0.6533$, then the hypothesis that $E(x) = 0$ would be rejected with a probability of 0.05 of being wrong. In this case, the level of significance is 0.05. Also, the interval of $\pm$ 0.6533 is referred to as the acceptance region for the hypothesis.

In general, the method of hypothesis testing requires that the statistical variation in the test, $T(\underline{x})$, be known or at least an adequate amount of information be available to determine the acceptance or rejection region for a desired level of significance. Usually the probability density function of $T(\underline{x})$ is required and the acceptance region for a hypothesis at the level of significance, denoted by α, is an interval or region which contains $T(\underline{x})$ with probability $1 - \alpha$. In general, if H is true, then $T(\underline{x})$ lies within the acceptance region with a probability of $1 - \alpha$, where α is the level of significance, i.e.,

$$P_{ROB}\left\{T(\underline{x}) \in R/H\right\} = 1 - \alpha$$

In a particular test, if $T(\underline{x})$ is found to lie within R, then H is accepted and if $T(\underline{x})$ does not lie in R, then H is rejected. Generally, R varies with α and, thus, acceptance or rejection of an hypothesis is dependent upon the level of significance used. Also, for a particular level of significance, the acceptance region is not unique since several regions can be found such that the probability of $T(\underline{x})$ lying within the regions is $1 - \alpha$. Usually, the smallest region is used.

Hypothesis testing can be based upon confidence intervals. Consider a confidence interval with a confidence coefficient of $1 - \alpha$. The true value of a parameter lies within the confidence interval with a probability of $1 - \alpha$; hence, the probability that the true value lies outside of the confidence interval is α. The confidence interval is a function of the sample data set which can be considered as the test for the hypothesis being considered. In this manner, the confidence interval is an acceptance region for the hypothesis and α is the level of significance. For example, the confidence interval for the mean value of a Gaussian random variable can be written as

$$P_{ROB} = \left[s - 1.645\sqrt{\frac{1}{n}\sigma^2} < m < s + 1.645\sqrt{\frac{1}{n}\sigma^2}\right] = 0.9 = 1 - 0.1$$

where s is the sample mean of an independent sample set of size n and σ^2 is the variance of the random variable. The hypothesis "H: $m = m_o$" would be accepted if the sample set mean s yields a confidence interval which contains m_o. The level of significance is 0.1.

It should be pointed out that rejecting a true hypothesis is not the only error that can be made in testing hypotheses. It is also possible to accept a false hypothesis, thus, the probability of making an error in hypothesis testing is the probability of rejecting a true hypothesis or accepting a false hypothesis. Generally, the probability of rejecting a true hypothesis is referred to as the probability of "Type I" error, and the probability of accepting a false hypothesis is referred to as the probability of "Type II" error. Also, the probability of rejecting a hypothesis when it is actually false is often referred to as the "power of the test." In general, the probabilities of Type I and Type II errors can be determined in testing alternative hypotheses, i.e., there exist

two hypothesis, H_1 and H_2, and it is required to make a decision concerning the validity of H_1 and H_2. The case is usually referred to as a simply hypothesis and a simple alternative. The best test to be used in hypothesis testing is usually dependent upon the consequences of being wrong in terms of either Type I or Type II error. The general problem of hypothesis testing is discussed in detail in References 1, 5, and 11. Some particular cases of present interest are discussed below.

2.2.7.2.1 Statistical Independence of Gaussian Random Variables

Let x and y be two Gaussian random variables and let $\underline{x}$ and $\underline{y}$ be two sample sets of x and y. A test of the hypothesis that x and y are statistically independent can be made using the probability density function for the sample correlation coefficient, r. That is, if $\rho_{xy} = 0$, then the probability density function of r is as follows: (See Section 2.2.7.1)

$$f(r) = \frac{2^{n-3}\,\Gamma^2\left(\frac{n-1}{2}\right)}{\pi\,(n-3)!}\,(1-r)^{\frac{n-4}{2}}$$

where

$$r = \frac{\Delta_{xy}}{\Delta_x \Delta_y} = \frac{\sum_{i=1}^{n}\left[(x_i - s_x)(y_i - s_y)\right]}{\sqrt{\left[\sum_{i=1}^{n}(x_i - s_x)^2\right]\left[\sum_{i=1}^{n}(y_i - s_y)^2\right]}}$$

Now, consider the following function of the sample correlation coefficient.

$$v = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2}$$

The inverse transformation becomes

$$r = \frac{v}{\sqrt{v^2 + (n-2)}}$$

The pdf for v is found as follows

$$f(v) = \frac{2^{n-3}\,\Gamma^2\left(\frac{n-1}{2}\right)}{\pi(n-3)!}\left[1 - \frac{v^2}{(v^2+n-2)}\right]^{\frac{n-4}{2}} \frac{\left[1 - \frac{v^2}{(v^2+n-2)}\right]}{\sqrt{v^2 + (n-2)}}$$

$$f(v) = \frac{2^{n-3}\,\Gamma^2\left(\frac{n-1}{2}\right)}{\sqrt{n-2}\,\pi\,(n-3)!}\left[1+\frac{v^2}{n-2}\right]^{-\left(\frac{n-1}{2}\right)}$$

$$= \frac{2^{n-3}\,\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{\pi(n-2)}\,\sqrt{\pi}\,\Gamma(n-2)}\left[1+\frac{v^2}{n-2}\right]^{-\left(\frac{n-1}{2}\right)}$$

$$f(v) = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{(n-2)\pi}\,\Gamma\left(\frac{n}{2}-1\right)}\left[1+\frac{v^2}{n-2}\right]^{-\left(\frac{n-1}{2}\right)}$$

Thus, it is seen that the pdf of v is the Student's "t" pdf for (n-2) degrees of freedom. Using this result a confidence interval can be found for v using the "t" pdf. That is, for a given α, two values t_1 and t_2 can be found such that

$$PROB\,[-t_1 < t < t_2] = 1-\alpha$$

where $0<\alpha<1$. Alternatively,

$$PROB\,[-t_1 < v < t_2/H] = 1-\alpha$$

where H is the hypothesis that x and y are statistically independent, i.e.,

$$H:\ \rho_{xy} = 0$$

The values of t_1 and t_2 are determined from the "t" pdf for n-2 degrees of freedom. If for two sample sets $\underline{x}$ and $\underline{y}$, v as determined by r is not in the interval $-t_1 < v < t_2$ for a particular α, then the hypothesis $\rho_{xy} = 0$ is rejected at the level of significance of α.

2.2.7.2.2 Goodness-of-Fit Test

In determining statistical properties it is common practice to assume that the type of probability density function is known and only a set of parameters

which specify the probability density function need to be estimated from a sample set. For example, if it is known that a random variable x has a Gaussian probability density function, it is sufficient to estimate the mean and variance from a sample set $\underline{x}$. However, there remains an uncertainty concerning the type of probability density function assumed. Fortunately, it is possible to assess the validity of assumptions concerning types of probability density function by a rather general method which is referred to as the "Goodness-of-Fit Test." This method is a method of hypothesis testing wherein the approximate density of the sample set is "compared" with the hypothesized density function and a decision is made to accept or reject the hypothesis. The "comparison" which is made is, in general terms, the actual deviations between the sample set density and the assumed probability density function. The actual test is based upon a particular measure of the observed deviations between the sample set density and the hypothesized probability density function. This measure of deviations can be expressed as follows:

$$T = \sum_{i=1}^{m} \frac{(O_i - E_i)^2}{E_i} = \sum_{i=1}^{m} \frac{D_i^2}{E_i}$$

In the measure, or test T, the term O_i denotes the number of observed occurrences, E_i is the expected number of occurrences, and D_i is the deviation between the observed and expected number of occurrences. The occurrences are simply the number of observations, or sample set points, which fall within an interval I_i. That is, for a total of n samples there can be constructed a set of m intervals, I_i, each of which will contain a certain number, say k_i, of the total number of observations. The expected number of occurrences within I_i is determined from the assumed probability density function, i.e., E_i is the number of occurrences within I_i given that the hypothesis is true. Of course, if the hypothesis is true and a sufficiently large sample size is used, then it is expected that T should be "relatively" small. However, some explicit measure of T is required. This measure is provided through the limit behavior of T for a general probability density function. It can be shown that under rather general conditions, the pdf of T approaches the Chi-square pdf. Thus, a confidence interval for T can be determined using the x^2 pdf, i.e.,

$$P_{ROB}\left[T_1 < T < T_2\right] = 1 - \alpha$$

approaches

$$P_{ROB}\left[\chi_1^2 < \chi^2 < \chi_2^2\right]$$

where

$$\chi_1^2 = T_1 \text{ AND } \chi_2^2 = T_2$$

The proof of this result is based upon the local limit theorem (Section 2.2.6.2). The significant steps in the proof are discussed below.

Using the results of the local limit theorem it is found that

$$p(n, \underline{k}) \simeq K e^{-\frac{1}{2}T}$$

where

$$K = \frac{\sqrt{n}}{(2\pi)^{\frac{m-1}{2}} \prod_{i=1}^{m} \sqrt{n p_i}}$$

$$T = \sum_{i=1}^{m} \frac{(k_i - np_i)^2}{np_i} = \sum_{i=1}^{m} \frac{(O_i - E_i)^2}{E_i}$$

and m is the number of intervals I_i, k_i is the number of outcomes within I_i, and n is the total number of outcomes. It should be apparent that $O_i = k_i$ and $E_i = np_i$.

It is seen that the maximum probability occurs for T = 0. Therefore, T as defined is a reasonable measure of the deviation from a set of expected results. In general, as $|O_i - E_i|$ increases $p(n,\underline{k})$ decreases.

The probability density function for T for large n approaches that for Chi-square for (n-1) degrees of freedom. This can be shown as follows. Consider T as a function of x_i where

$$\chi_i = \frac{k_i - np_i}{\sqrt{np_i}}$$

and

$$T = \sum_{i=1}^{m} \frac{\chi_i^2}{p_i}$$

Now, since the k_i are not independent, the x_i are not independent, i.e.,

$$\sum_{i=1}^{m} k_i = n \quad AND \sum_{i=1}^{m} x_i = 0$$

Therefore,

$$T = \sum_{i=1}^{m-1} \frac{x_i^2}{p_i} + \frac{x_m^2}{p_m}$$

$$= \sum_{i=1}^{m-1} \frac{x_i^2}{p_i} + \frac{1}{p_m} \left(\sum_{i=1}^{m-1} x_i \right)^2$$

$$= \sum_{i=1}^{m-1} \frac{x_i^2}{p_i} + \frac{1}{p_m} \sum_{i=1}^{m-1} x_i^2 + \frac{1}{p_m} \sum_{i \neq j}^{m-1} \sum_{j=1}^{m-1} x_i x_j$$

$$= \sum_{i=1}^{m-1} \left(\frac{1}{p_m} + \frac{1}{p_i} \right) x_i^2 + \frac{1}{p_m} \sum_{i \neq j}^{m-1} \sum_{i=1}^{m-1} x_i x_j$$

$$T = \underline{X}^T A \underline{X}$$

where A is a symmetrical positive definite matrix of order m-1 with diagonal terms $(1/ p_m + 1/p_i)$ and all off-diagonal terms $1/p_m$, and $\underline{x}$ is a vector of x_i for i = 1, 2, ..., m-1 . The moment generating function for T becomes

$$m_T(s) = E\left[e^{sT}\right] = \sum_{\underline{k}} e^{sT} p(n, \underline{k})$$

where the summation is taken over all $\underline{k}$ such that $\underline{1}^T\underline{k} = n$ and, of course, $0 \le k_i$. However, for sufficiently large n, the summation can be replaced by continuous integration over $\underline{x}$ with a change of variable of $\Delta x_i = 1/\sqrt{n}\,\Delta k_i$; hence, as $n \to \infty$, $m_T(s)$ becomes

$$\lim_{n\to\infty} [m_T(s)] = (2\pi)^{\frac{1-m}{2}} \left(\prod_{i=1}^{m} p_i^{-1/2} \right) \int_{D(x)} \left\{ \exp\left[(s - 1/2)\, \underline{x}^T A\, \underline{x} \right] \right\} d\underline{x}$$

Using $I_2(\underline{s})$ of Appendix A and noting that $|A| = \prod_{i=1}^{m} (p_i^{-1})$, it is found that

$$\lim_{n\to\infty} \left[m_T(s) \right] = \frac{1}{(1-2s)^{\frac{m-1}{2}}}$$

Therefore, as $n \longrightarrow \infty$, T has a Chi-square pdf distribution for m-1 degrees of freedom.

With the foregoing, a test of "goodness" of fit can be constructed which assesses the validity of assumptions concerning types of probability density functions which random variables possess. A test of the hypothesis that a set of sample data is generated from a random variable with a particular probability density function is made by simply computing T and comparing the result with the $1 - \alpha$ confidence interval. The hypothesis is accepted or rejected at the level of significance α .

It should be noted that the foregoing test does not consider properties of the assumed probability density function which are estimated from the sample set $\underline{x}$. That is, usually the type of probability density function is assumed with the first and second moments equal to those of the sample set $\underline{x}$. However, in such cases the method of the Goodness-of-Fit remains essentially the same; only the number of degrees of freedom change. Generally, the number of degrees of freedom is simply reduced by the number of statistical moments which are estimated from the sample set $\underline{x}$ and the test is the same. The method of Goodness-of-Fit is discussed in detail in References 1, 4, and 6.

2.2.8 IMU Error Model and Analysis

In the evaluation of the performance of a G&N system the analysis of the Inertial Measurement Unit (IMU) is of prime importance. The function of the IMU is to provide a self-contained reference coordinate system and a means of measuring accelerations of the vehicle. The measured accelerations determine gyro torquing signals to maintain the orientation of the coordinate system and are used to determine the trajectory of the vehicle. The objective of an IMU error analysis is to evaluate the effect of errors inherent in the manufacture and installation of inertial platform components on the measurement of accelerations and the uncertainty in orientation of the coordinate system. Generally, IMU inaccuracies result in measured acceleration errors and their effect is ultimately related to trajectory computation errors.

The development in this section employs the inherent assumptions of the state variable linear systems approach to the solutions of the perturbed equations of motion of a space vehicle. Equations are derived which relate the effect of IMU gyro and accelerometer errors to position and velocity errors. The equations are presented in a form which is independent of a particular platform and may be applied to a variety of Inertial Measurement Units. A discussion of platforms using single degree-of-freedom or two degrees-of-freedom gyros and the corresponding error equations is presented.

2.2.8.1 Perturbation Equations

Inertial Measurement Unit errors will be in the form of acceleration measurement errors from the accelerometers and acceleration errors due to misalignment of the platform gyros. It is desired to relate these errors in the sensed acceleration of the platform to errors in position and velocity of the vehicle. The motion of the vehicle in the influence of a gravity field and an applied thrust is given by equation (8.1.1)

$$\ddot{\bar{r}}_0 + \bar{g}_0(\bar{r}_0) = \bar{A}_0 \tag{8.1.1}$$

where $\ddot{\bar{r}}_0$ is the second time derivative of the position of the vehicle, $\bar{g}(r_0)$ is the total gravity vector, and $\bar{A}_0$ is the applied thrust acceleration as measured by a perfect IMU. A similar equation may be written in which the quantities of equation (8.1.1) are interpreted as the acceleration sensed by an IMU with sensor errors present, $\bar{A}$, and the resulting "error-corrupted" computed value of the second time derivative of position, $\ddot{\bar{r}}$.

$$\ddot{\bar{r}} + \bar{g}(\bar{r}) = \bar{A} \tag{8.1.2}$$

The variational, or perturbed equation, is the difference of the quantities in (8.1.1) and (8.1.2).

$$\ddot{\bar{r}} - \ddot{\bar{r}}_o + \left[\bar{g}(\bar{r}) - \bar{g}_o(\bar{r}_o)\right] = \bar{A} - \bar{A}_o$$

or

$$\delta\ddot{\bar{r}} + \delta\left[\bar{g}(\bar{r})\right] = \delta\bar{A} \tag{8.1.3}$$

Equation (8.1.3) gives a relationship between the IMU measurement errors, $\delta\bar{A}$, and their affect on the computed value of $\ddot{\bar{r}}$, $\delta\ddot{\bar{r}}$.

Before the solution of the equation may be completed the form of $\delta[\bar{g}(\bar{r})]$ must be determined. As the term appears in equation (8.1.3), it represents a general expression for the gravity field and may represent the influence of more than one attracting body. In most cases of interest the vehicle is in the influence of a single attracting body and a simple form for the gravity term may be derived. To solve (8.1.3) it is necessary to express the gravity term as a function of variations in $\delta\bar{r}$.

For a gravity field of more than one attracting body an approximation of the term is made by a Taylor series expansion about $\bar{g}_o(\bar{r}_o)$ in which terms involving derivatives higher than first are neglected (Reference (18)). For the central force field of a spherical homogenous attracting body the evaluation of the gravity term as a function of $\delta\bar{r}$ assumes a simple analytical form.

In any case the quantity $\bar{g}(\bar{r})$ may be written as

$$\bar{g}(\bar{r}) = \bar{g}\left[g_1(x,y,z,t), g_2(x,y,z,t), g_3(x,y,z,t)\right] \tag{8.1.4}$$

where x, y, and z are the components of $\bar{r}$ expressed in an inertial three-dimensional Cartesian coordinate system.

The differential of $\bar{g}(\bar{r})$ is then

$$\delta\bar{g}(\bar{r}) = \nabla g_1 \cdot \delta\bar{r}\,\dot{s} + \nabla g_2 \cdot \delta\bar{r}\,j + \nabla g_3 \cdot \delta\bar{r}\,k\,\dot{s}$$

$$= \left(\frac{\partial g_1}{\partial x}\dot{s} + \frac{\partial g_1}{\partial y}j + \frac{\partial g_1}{\partial z}k\right) \cdot \delta\bar{r}\,\dot{s}$$

$$+ \left(\frac{\partial g_2}{\partial x}\dot{s} + \frac{\partial g_2}{\partial y}j + \frac{\partial g_2}{\partial z}k\right) \cdot \delta\bar{r}\,j$$

$$+ \left(\frac{\partial g_3}{\partial x}\dot{s} + \frac{\partial g_3}{\partial y}j + \frac{\partial g_3}{\partial z}k\right) \cdot \delta\bar{r}\,k$$

and in matrix form

$$\delta \bar{g}(\bar{r}) = \begin{bmatrix} \frac{\partial g_1}{\partial x} & \frac{\partial g_1}{\partial y} & \frac{\partial g_1}{\partial z} \\ \frac{\partial g_2}{\partial x} & \frac{\partial g_2}{\partial y} & \frac{\partial g_2}{\partial z} \\ \frac{\partial g_3}{\partial x} & \frac{\partial g_3}{\partial y} & \frac{\partial g_3}{\partial z} \end{bmatrix} \begin{bmatrix} \delta x \\ \delta y \\ \delta z \end{bmatrix} \tag{8.1.5}$$

and

$$\delta \bar{g}(\bar{r}) = [G]\, \delta \bar{r} \tag{8.1.6}$$

The form of the elements of the $[G]$ matrix in equation (8.1.5) is the same if the truncated Taylor series expansion of Reference (18) or the method of Reference (19) is used to evaluate $\delta[\bar{g}(\bar{r})]$. Using vector·operational symbolism, equation (8.1.5) may be expressed as

$$\delta \bar{g}(\bar{r}) = \frac{\partial \bar{g}(\bar{r})}{\partial \bar{r}}\, \delta \bar{r} = [G]\, \delta \bar{r} \tag{8.1.7}$$

Substituting equation (8.1.7) in equation (8.1.3) gives

$$\delta \ddot{\bar{r}} + [G]\, \delta \bar{r} = \delta \bar{A} \tag{8.1.8}$$

To specify the time history of the affect of IMU errors on the computed trajectory the errors must be related to the position and velocity errors. Errors in the computed position are $\delta \bar{r} = \bar{r} - \bar{r}_0$ and the first time derivative of position errors is

$$\delta \bar{v} = \delta \dot{\bar{r}} = \dot{\bar{r}} - \dot{\bar{r}}_0 \tag{8.1.9}$$

and time derivative of velocity is

$$\delta \dot{\bar{v}} = \delta \ddot{\bar{r}} = \ddot{\bar{r}} - \ddot{\bar{r}}_0 \tag{8.1.10}$$

Using equations (8.1.8), (8.1.9), and (8.1.10), the linear differential equation, is state vector form, relating position and velocity errors is

$$\begin{bmatrix} \delta\dot{\bar{r}} \\ \hline \delta\dot{\bar{v}} \end{bmatrix}_{6\times1} = \begin{bmatrix} 0 & I \\ \hline -G & 0 \end{bmatrix}_{6\times6} \begin{bmatrix} \delta\bar{r} \\ \hline \delta\bar{v} \end{bmatrix}_{6\times1} + \begin{bmatrix} 0 \\ \hline \delta\bar{A} \end{bmatrix}_{6\times1} \tag{8.1.11}$$

or

$$\delta\dot{\bar{X}}(t) = B(t)\,\delta\bar{X}(t) + \bar{a}(t) \tag{8.1.12}$$

The matrix B is a function of the gravitational forces acting on the vehicle and the matrix $\bar{a}$, the forcing function of the differential equation, represents the errors in the control vector. Subsequent sections will give the solution of equation (8.1.12) and specifications of $\bar{a}$ in terms of the IMU model.

The G matrix of equation (8.1.11) assumes a simple form when the gravity field may be represented by

$$\bar{g}(\bar{r}) = \frac{\mu\bar{r}}{r^3} \tag{8.1.13}$$

Using the operational vector identities

$$\frac{\delta\bar{r}}{\delta\bar{r}} = [I] \quad (\text{IDENTITY MATRIX}) \tag{8.1.14}$$

and

$$\frac{\delta r^n}{\delta\bar{r}} = (n r^{n-2})\,\bar{r}^T \tag{8.1.15}$$

the $\frac{\partial\bar{g}(\bar{r})}{\partial\bar{r}}$ may be evaluated as follows:

$$-\frac{\partial\bar{g}(\bar{r})}{\partial\bar{r}} = \left[\frac{\partial}{\partial\bar{r}}\,\mu r^{-3}\bar{r}\right]$$

$$= (\mu r^{-3})\frac{\partial\bar{r}}{\partial\bar{r}} + (\mu\bar{r})\frac{\partial(r^{-3})}{\partial\bar{r}}$$

$$= \mu r^{-3}[I] + \mu(-3)r^{-5}\bar{r}\bar{r}^T$$

$$= \frac{\mu}{r^3} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \frac{3\mu}{r^5} \begin{bmatrix} x^2 & xy & xz \\ yx & y^2 & yz \\ zx & xy & z^2 \end{bmatrix}$$

or

$$\frac{\partial \bar{g}(\bar{r})}{\partial \bar{r}} = [G] = \frac{-\mu}{r^5} \begin{bmatrix} (r^2 - 3x^2) & -3xy & -3xz \\ -3xy & (r^2 - 3y^2) & -3yz \\ -3xz & -3yz & (r^2 - 3z^2) \end{bmatrix} \tag{8.1.15}$$

For the derivation above it was assumed that $\bar{r} = (x, y, z)$ in an inertial Cartesian coordinate system with the origin at the center of the attracting body. Derivations resulting in equation (8.1.15) by different methods may be found in References (15) and (19).

2.2.8.2 Solution of the Perturbed Equation

The solution of equation (8.1.12) may be accomplished by a variety of methods of which two are most frequently used - the adjoint technique and variation of parameters. The variation of parameters method is the more direct method and is used here. In each case the solution of the homogenous part of equation (8.1.12) is the "Fundamental Solution Matrix" or "State Transition Matrix"; and the solution of the differential equation is determined in terms of a transition matrix, the initial conditions of the state variables, and the forcing function.

2.2.8.2.1 Variation of Parameter Solution

The homogenous part of equation (8.1.12) is

$$\delta \dot{\bar{X}}(t) - B(t)\, \delta \bar{X}(t) = 0 \tag{8.2.1}$$

The solution of (8.2.1) [References (15) and (20)] is given in terms of the state transition matrix $\phi(t, t_o)$ relating the state at time t to the state at time t_o as

$$\underset{6\times 1}{\delta X(t)} = \underset{6\times 6}{\phi(t, t_o)}\, \delta \bar{X}(t_o) \tag{8.2.2}$$

$$\phi(t_o, t_o) = [I]$$

The variation of parameters solution proceeds from equation (8.2.1) by assuming a solution of the form

$$\delta\bar{x}(t) = \phi(t,t_o)\,\bar{y}(t) \tag{8.2.3}$$

where $\phi(t,t_o)$ is the fundamental solution matrix and $\bar{y}(t)$ is a function to be determined. The first time derivative of equation (8.2.3) is

$$\dot{\delta\bar{x}}(t) = \dot{\phi}(t,t_o)\,\bar{y}(t) + \phi(t,t_o)\,\dot{\bar{y}}(t) \tag{8.2.4}$$

Substituting equation (8.2.4) into equation (8.1.12) gives

$$\phi(t,t_o)\,\dot{\bar{y}}(t) + \left\{\dot{\phi}(t,t_o) - B(t)\,\phi(t,t_o)\right\} = \bar{a}(t) \tag{8.2.5}$$

Since $\phi(t,t_o)$ is the fundamental solution matrix of the homogenous part of equation (8.1.12), $\phi(t,t_o)$ must satisfy the homogenous equation.

$$\dot{\phi}(t,t_o) = B(t)\,\phi(t,t_o) \tag{8.2.6}$$

This relationship shows the quantity in braces to be identically zero and

$$\phi(t,t_o)\;\;\dot{\bar{y}}(t) = \bar{a}(t) \tag{8.2.7}$$

or

$$\dot{\bar{y}}(t) = \phi^{-1}(t,t_o)\,\bar{a}(t) \tag{8.2.8}$$

The integral of equation (8.2.8) is

$$\begin{aligned}\bar{y}(t) &= \int_{-\infty}^{t} \phi^{-1}(\tau,t_o)\,\bar{a}(\tau)\,d\tau \\ &= \int_{-\infty}^{t_o} \phi^{-1}(\tau,t_o)\,\bar{a}(\tau)\,d\tau + \int_{t_o}^{t} \phi^{-1}(\tau,t_o)\,\bar{a}(\tau)\,d\tau\end{aligned} \tag{8.2.9}$$

Substituting (8.2.9) into (8.2.3)

$$\delta\bar{x}(t) = \phi(t,t_o)\int_{-\infty}^{t_o}\phi^{-1}(\tau,t_o)\,\bar{a}(\tau)\,d\tau + \phi(t,t_o)\int_{t_o}^{t}\phi^{-1}(\tau,t_o)\,\bar{a}(\tau)\,d\tau \tag{8.2.10}$$

and equation (8.2.10) evaluated at $t = t_o$ gives

$$\delta\bar{x}(t_o) = \int_{-\infty}^{t_o}\phi^{-1}(\tau,t_o)\,\bar{a}(\tau)\,d\tau \tag{8.2.11}$$

and

$$\delta\bar{X}(t) = \phi(t,t_o)\,\delta\bar{X}(t_o) + \int_{t_o}^{t} \phi(t,t_o)\,\phi^{-1}(T,t_o)\,\bar{a}(T)\,dT \qquad (8.2.12)$$

A further simplification in the integrand of equation (8.2.12) may be accomplished through the use of the properties of the transition matrix. However, the simplification of the form of the equation may not give a corresponding simplification in the evaluation of the integral. The properties of the transition to be used are indicated in Figure 8.1. In the figure the time T is shown to fall between times t and t_o although the relationship is valid for the case in which T lies outside the interval indicated.

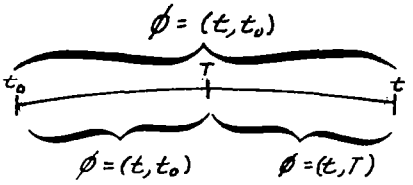

Figure 8.1

$$\phi(t,t_o) = \phi(t,T)\,\phi(T,t_o) \qquad (8.2.13)$$

and

$$\phi(t,T) = \phi(t,t_o)\,\phi^{-1}(T,t_o) \qquad (8.2.14)$$

Substitution of equation (8.2.14) into (8.2.12) gives

$$\delta\bar{X}(t) = \phi(t,t_o)\,\delta\bar{X}(t_o) + \int_{t_o}^{t} \phi(t,T)\,\bar{a}(T)\,dT \qquad (8.2.15)$$

Equation (8.2.15) or equation (3.2.12) is the desired solution of the perturbed equation.

2.2.8.2.2 Statistical Evaluation of the Perturbed Equations

In order to make a statistical study of the errors $\delta\bar{x}(t)$ the first and second moments are completed. For these equations the symbol $(\widetilde{\quad})$ represents the expected value of the quantity involved. The first moment of $\delta x(t)$ is

$$E\left[\delta X(t)\right] = m_X(t) = \phi(t,t_o)\,\widetilde{\delta\bar{X}(t_o)} + \int_{t_o}^{t} \phi(t,T)\,\widetilde{\bar{a}(t)}\,dT \qquad (8.2.16)$$

and the covariance matrix is

$$\Gamma_X(t) = \overbrace{\left[\delta\bar{X}(t) - m_X(t)\right]\left[\delta\bar{X}(t) - m_X(t)\right]^T} \tag{8.2.17}$$

For a zero mean

$$\Gamma_X(t) = \phi(t,t_o)\Gamma_X(t_o)\phi^T(t,t_o) + \int_{t_o}^{t} d\lambda \int_{t_o}^{t} d\lambda \left[\phi(t,t_o)\,\overbrace{\bar{a}(\tau)a(\lambda)^T}\,\phi(t,\tau)\right] d\tau$$
$$+ \int_{t_o}^{t} \left[\phi(t,t_o)\,\overbrace{\delta\bar{X}(t_o)\,\bar{a}^T(t)}\,\phi(t,\tau)\right] d\tau \tag{8.2.18}$$
$$+ \int_{t_o}^{t} \left[\phi(t,\tau)\,\bar{a}(t)\,\delta\bar{X}(t_o)^T(t,t_o)\right] d\tau$$

The last two terms of equation (8.2.18) are ordinarily zero.

2.2.8.3 Platform Error Equations

Presented in this section are the gyroscope and accelerometer error equations used for defining the IMU error model. The equations are presented in general form applicable to inertial units using either single degree-of-freedom or two degree-of-freedom gyros. The platform errors are related to measured acceleration errors thereby specifying the form of the forcing function of the perturbed equations of motion.

2.2.8.3.1 Inertial Measurement Unit and Transformation

The inertial platform of an inertial measurement unit is a sensor which provides acceleration signals resolved along known coordinates. Conventionally, the platform is supported by a set of gimbals, the inner gimbal serving as a stable member supporting the gyroscopes and accelerometers. The function of the gyroscopes is to maintain the orientation of the platform, and the function of the accelerometers is to measure the total acceleration. The physical orientation of the gyros input axis and accelerometer alignment on the platform is determined by the specific application for which the IMU is designed. Generally, a three-dimensional orthogonal coordinate system is constructed through the use of three single-degree-of-freedom gyros or two degree-of-freedom gyros. A gyro input axis and an accelerometer may be aligned along each coordinate axis.

The derivation of the perturbed equation of motion for the trajectory computation coordinate frame assumed an inertial coordinate system. Further, the expression for the [G] matrix for the central force field assumed the coordinate system to be centered at the center of the attracting body. The alignment of the platform coordinate axis with respect to this inertial coordinate system is also dependent on the specific application of the measurement unit. A discussion of the advantages and disadvantages of a particular

platform alignment is given in Reference (21). For the different platform alignments a transformation between the trajectory computation coordinates and platform coordinates may be defined and assumes the form

$$\left[T\right]_{TC} = \begin{bmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{bmatrix} \tag{8.3.1}$$

The subscript "TC" indicates the transformation from the trajectory to the platform coordinates. The elements T_{ij} will be constants if the initial platform orientation is maintained or functions of time if the platform coordinates are changing with respect to the trajectory frame.

2.2.8.3.2 Gyroscope and Accelerometer Error Equations

The acceleration measurement errors from the platform gyros, in platform coordinates, are determined by use of equations of the general form

$$\epsilon_i = i K_0 + i K_1 a_x + i K_2 a_y + i K_3 a_z + i K_4 a_x^2 + i K_5 a_y^2 + i K_6 a_x^2 + i K_7 a_x a_y + i K_8 a_x a_z + i K_9 a_y a_z \tag{8.3.2}$$

$(i = x, y, z)$

where ϵ_i represents the total drift rate of the gyro during the acceleration of the vehicle about the "i" platform coordinate axis. The terms iK_j are the drift rate coefficients and are dependent on the construction of the gyros. It has been tacitly assumed that a gyro input axis has been aligned along each platform coordinate axis. In general, the drift rate coefficients may be functions of time but are usually expressed as constants representing average values. The terms a_x, a_y, and a_z are the components of the applied accelerations along the platform coordinate axis. The particular error sources represented by the iK coefficents are:

K_0 = bias coefficient (non-g sensitive)

$K_{1,2,3}$ = mass unbalance coefficients (g-sensitive)

$K_{4,5,6,7,8,9}$ = anisoelasticity coefficients (g^2-sensitive)

Analysis of a specific inertial unit in which the orientation of the spin, input, and output axis is specified, and the drift-rate coefficients given, will allow a reduction of number of terms used in equation (8.3.2). Terms may be eliminated by comparison of the relative magnitudes of the coefficients and by proper orientation of the platform coordinates with respect to the trajectory plane. In such cases, the number of terms retained may be reduced to six or seven. The particular terms which may be neglected can be determined only through an analysis of the gyros in conjunction with platform orientation and the trajectory.

Although the gyro characteristics are specified by the drift rate characteristics, the acceleration errors due to the gyros arise from the angular misalignments of the gyros with the platform axis. The total misalignment along the platform axis may be found by integrating the drift rates.

$$\bar{\psi}(t) = \int_{t_0}^{t} \bar{\epsilon}(T)\, dT + \bar{\psi}(t_0) \tag{8.3.3}$$

where $\bar{\psi}(t)$ is the angular misalignment vector about the platform axis and $\bar{\epsilon}(T)$ are the drift rates in vector form. The acceleration error in platform coordinates is then given by

$$\delta \bar{A}\,(gyros) = -\bar{\psi}(t) \times \bar{A} \tag{8.3.4}$$

where $\bar{A}$ is the acceleration in platform coordinates.

The error model for the accelerometers mounted on the platform is of the form

$$\begin{aligned} \nabla_i = {} & i k_0 + i k_1 a_x + i k_2 a_y + i k_3 a_z + i k_4 a_x^2 \\ & + i k_5 a_x a_y + i k_6 a_x a_z \; i k_7 a_x^3 \end{aligned} \tag{8.3.5}$$

$$(i = x\, y\, z$$

The term $\nabla_i\,(i = x\, y\, z)$ is the acceleration measurement error along the indicated platform axis, assuming an accelerometer aligned along each axis. The terms a_x, a_y, and a_z are the thrust acceleration components along each platform coordinate axis. The coefficients $i k_j$ are defined below and are dependent on the construction and accuracy of alignment of the accelerometers on the platform.

ik_0 = bias coefficient

ik_1 = linear scale factor coefficient

$ik_{2,3}$ = bias sensitivity to cross-axis acceleration

ik_4 = 2nd-order nonlinearity coefficient

$ik_{5,6}$ = scale factor sensitivity to cross-axis acceleration

ik_7 = 3rd-order nonlinearity coefficient

Equations (8.3.4) and (8.3.5) may be combined to give the total acceleration measurement errors from gyro misalignment and accelerometer errors, $\bar{A}_p(t)$.

$$\bar{A}_p(t) = \nabla - \bar{\psi} \times \bar{A} \tag{8.3.5}$$

If the initial alignment of the platform remains fixed with respect to the inertial coordinate system, elements of the transformation, equation (8.3.1), will be constants. This platform coordinate system is referred to as either "platform inertial" or "launch point fixed." The initial alignment of the platform determines the transformation for all points along the trajectory, and the acceleration measurement errors in inertial coordinates are given by

$$\bar{a}(t) = \left[T\right]_{TC}^{T}\left[\bar{\nabla} - \bar{\psi} \times \bar{A}\right] \tag{8.3.6}$$

$\bar{a}(t)$ as determined above are the acceleration errors required for the evaluation of equations (8.2.15) and (8.2.25).

Determination of $\bar{a}(t)$ is more complicated if the platform orientation changes as a function of time. The elements of the transformation matrix become functions of time, and terms involving the coordinate system angular rate and angular rate derivative, with the inertial position and velocity, appear in equation (8.3.6). Letting $\bar{\omega}$ represent the vector rate of change of the platform coordinates with respect to trajectory coordinates, the position and velocity in the two frames are related by

$$\bar{r}_T = \left[T\right]_{TC}^{T} \bar{r}_C \tag{8.3.7}$$

$$\bar{V}_T = \bar{V}_C + \bar{\omega} \times \bar{r}_T \tag{8.3.8}$$

The time derivative of equation (8.3.8) gives

$$\bar{A}_T = \bar{A}_C + \bar{\omega} \times \bar{V}_T + \dot{\bar{\omega}} \times \bar{r}_T \tag{8.3.9}$$

The subscript "T" symbolizes the trajectory inertial frame, "C" the platform rotating frame, and the "·" time derivative. Equation (8.3.9) may then be solved for a_c .

An alternative to the use of the time varying transformation matrix is the solution of the perturbed equations of motion in a rotating coordinate frame. The solution of the perturbed equation proceeds in the same manner as in Section 2.2.8.1. The angular rate terms in this case are explicit in the formulation of the "B" matrix of equation (8.1.12), and are a part of the fundamental solution matrix. As such, they again appear in the integral in equation (8.2.12).

3.0 RECOMMENDED PROCEDURES

In Section 2.1 the general problem of systems performance analysis was defined in terms of two parameter sets and a known functional dependence. That is, the general problem involves a vector function of random variables denoted by

$$\underline{y} = \underline{G}(\underline{x})$$

where $\underline{y}$ is a vector of performance parameters, $\underline{x}$ is a random vector of causal parameters and $\underline{G}(\ )$ is a known vector function. Of course, $\underline{y}$ is a random vector since it is a function of the random vector $\underline{x}$. Now, for each mission or mission phase there exists a region in the space of the vector $\underline{y}$ which is conducive to mission success. This region can be defined as the "region-of-success" or success region for $\underline{y}$ and is denoted by R_S. That is, if $\underline{y}$ lies in R_S then the mission or mission phase is successful, hence, $\underline{y} \in R_S$ is equivalent to mission success. Unfortunately, due to the random or uncertain nature of $\underline{y}$, as caused by $\underline{x}$, it cannot be stated with certainty that $\underline{y} \in R_S$ or that mission success will be achieved. On the other hand, if the probability density function of $\underline{y}$ is known, then it is theoretically possible to determine the probability that $\underline{y}$ will lie in any region in the space of $\underline{y}$. In particular, the probability that $\underline{y}$ will lie in the success region R_S could theoretically be determined. This probability can be considered as the probability of success for the mission, denoted by P_S, i.e., generally

$$P_S = P[\underline{y} \in R_S] = \int_{R_S} f(\underline{y})\, d\underline{y}$$

where $f(\underline{y})$ is the probability density function of the performance parameter set $\underline{y}$. It is characteristic of space flight missions that a relatively "high" probability of success is required which reflects a high loss in the event of mission failure. Thus, in order to maintain a low risk it is necessary to reduce the probability of failure to negligible proportions.

Nonetheless, it is the general purpose of system performance analysis to determine or assess P_S and to ultimately determine the system configuration and system function requirements which will fulfill a specified lower bound constraint on P_S or a minimum requirement for P_S. The general tasks involved in this effort consist of (1) determining the statistical properties of the causal parameter set $\underline{x}$, i.e., specifying the probability density function of the random vector $\underline{x}$; (2) transforming the pdf of $\underline{x}$ into the probability density function of the performance parameter set $\underline{y}$, which is dependent upon $\underline{G}(\ )$; and (3) determining P_S as required. However, it is usually required that these tasks be accomplished such that the dependence of P_S upon $\underline{G}(\ )$ and the statistical moments of $\underline{x}$ is known. This is required in order to facilitate the definition of the optimum system configuration and the requirements of system

functions. Theoretically, the problem of system performance analysis is readily solved, i.e., the tasks involved are easily stated. Unfortunately, the tasks are generally not as easily accomplished. First, even if the probability density function of $\underline{y}$ is determined it is not usually easy to determine an explicit evaluation of P_s as a function of the statistical properties of $\underline{x}$. Second, even if the probability of $\underline{x}$ is known it is not always easy to determine an explicity form for the probability density function of $\underline{y}$. And, third, the explicit form of the probability density function of $\underline{x}$ is not generally known, rather, only estimates of the statistical moments of $\underline{x}$ are known and a limiting form of the probability density function is assumed. Notwithstanding this, the general objectives of system performance analysis can be accomplished through appropriate use of the statistical methodology discussed in the previous sections. Some general applications of the procedures are discussed below.

It is generally possible and often convenient to define the success region R_s with respect to a "point" $\underline{y}_s$ in the space of $\underline{y}$ which assures mission success. This point is often referred to as "nominal" conditions which are usually directly representative of mission objectives in terms of system state quantities. In this manner, it is understood that both $\underline{x}$ and $\underline{y} = \underline{G}(\underline{x})$ are variations about nominal conditions. This, in turn, implies a nominal system configuration or system design and requirements which usually represent gross requirements. However, final and/or complete system requirements must be determined such that $\underline{y}$ lies in R_s with the required probability of success. The use of nominal conditions often provides a linear relationship between the parameter sets $\underline{y}$ and $\underline{x}$, i.e., $\underline{y} = \underline{G}(\underline{x}) = A\,\underline{x}$, where A is a constant matrix. If the relationship between $\underline{y}$ and $\underline{x}$ is linear, then the statistical analyses involved in system performance analysis are greatly simplified. The use of nominal conditions is generally useful and particularly convenient if a linear relationship between the parameter sets $\underline{y}$ and $\underline{x}$ is obtained. However, linear relationships obtained in this manner are usually first order approximations and the effects of inaccuracies of such approximations upon the results of system performance analysis must be assessed.

It should be apparent that there exists a particular, and perhaps hypothetical, situation where the tasks of system performance analysis can be easily accomplished. This situation is characterized by (1) a linear relationship between the parameter sets $\underline{y}$ and $\underline{x}$, i.e., $\underline{y} = A\,\underline{x}$ which can often be obtained using nominal conditions as discussed above; (2) a Gaussian joint probability density function for the causal parameter set $\underline{x}$, with possibly statistically independent subsets; and (3) a relatively convenient region of success R_s defined in the space of $\underline{y}$, i.e., an R_s for which $P_s = [\underline{y} \in R_s]$ can be determined. In this particular situation the joint probability density function of the performance parameter set $\underline{y}$ is also Gaussian with its statistical moments easily related to those of the causal parameter set $\underline{x}$, as discussed in Sections 2.2.2.13 and 2.4.4. The marginal and conditional probability density functions for the performance parameter set can be easily determined as discussed in Appendix B. The evaluation of probabilities for the parameter set $\underline{y}$ for certain regions can be made as discussed in Appendix C.

Frequently, the performance of a navigation and guidance system is evaluated in terms of a quadratic form of vehicle state quantities. That is, optimization criteria and performance indices are often quadratic forms of system state parameters. In particular, loss functions are often quadratic forms and an optimization criterion is often the minimization of the expectation of a quadratic loss function. In this case, there exist two questions which concern, first, the actual minimum expected loss and, second, the relationship between the minimum loss and system design parameters. Generally, the loss, L, can be written as $L = \underline{y}^T Q \underline{y}$ where $\underline{y}$ is a set of vehicle state parameters. The loss is a scalar random variable which is specified by its probability density function. Usually, the explicit form of the pdf of L is not easily determined, i.e., a closed form expression is generally not possible. However, if $\underline{y}$ is a linear function of $\underline{x}$, $\underline{y} = A \underline{x}$, and $\underline{x}$ is a Gaussian random vector then an explicit form can be obtained for the first and second moments of the loss L which represent the actual minimum expected loss and a measure of its variation. In Section 2.2.4.5.5 these moments are shown to be an explicit function of the elements of the matrix Q and the co-variance matrix of $\underline{y}$, which is easily related to the co-variance matrix of $\underline{x}$, since $\underline{y} = A \underline{x}$. In particular the expected value of the loss is equal to the trace of the matrix product of Q and the co-variance matrix of $\underline{y}$. It is noted that the selection of system design parameters which minimize the trace of this product leads to an optimum system configuration.

As noted above, it is not always possible to obtain an explicit form for the probability density function of the parameter set $\underline{y}$ or functions of $\underline{y}$ which are used in the evaluation of performance. For example, as noted above, even if $\underline{y}$ has a known Gaussian probability density function, the probability density function for a quadratic form of $\underline{y}$ is not easily determined. On the other hand, the lower order statistical moments can generally be determined as discussed in Section 2.2.4. These moments, in turn, can be used to determine probability bounds as discussed in Section 2.2.5. Thus, in cases where the explicit form of the probability density functions of performance functions cannot be obtained and, hence, an explicit evaluation of the probability of success cannot be made, it is possible to bound this probability by using only lower order statistical moments which can usually be obtained.

In the foregoing it is tacitly assumed that the probability density function of the causal parameter set $\underline{x}$ is known and, hence, that of the performance parameter set $\underline{y}$ can be determined, which is generally possible for a linear relationship $\underline{y} = A \underline{x}$ and a Gaussian probability density function for $\underline{x}$. However, a complete statistical description of $\underline{x}$ is usually not explicitly known. That is, only estimates of lower order statistical moments are usually available and assumptions are made concerning the type or form of the probability density function of the set $\underline{x}$. The accuracy of such estimates and the validity of such assumptions directly affect the accuracy and validity of statements concerning system performance. It must be recognized that statements concerning system performance are, at best, statistical inferences which must be based upon the available information of the statistical properties of the parameter sets $\underline{y}$ and $\underline{x}$, which is usually not complete and/or explicit. The methods of estimating statistical moments are discussed in Sections 2.2.7.1.1 and 2.2.7.1.2, wherein methods of assessing the accuracy of the estimates are considered. A somewhat "universal" assumption concerning the form or type of

probability density function when definite information is not available is that it is Gaussian. The general validity of this assumption can be based upon the limiting theorems discussed in Section 2.2.6. These theorems provide a rather general basis for the validity of the assumption of Gaussian probability density functions, however, there always exists some question concerning the convergence of the limiting form and the existence of the proper conditions for convergence to the Gaussian form. Most experience indicates that the convergence is rather rapid, but, each situation encountered should be considered upon its own basis. In general, the validity of assumptions concerning probability density functions can be assessed by the means of hypothesis testing as discussed in Section 2.2.7.2.

It becomes apparent that the particular procedures to be utilized in systems performance analysis depends upon several particular aspects of the problem involved which concern (1) functional dependence of performance parameters and causal parameters, (2) functional forms used in the evaluation of system performance, (3) type of probability density functions involved, and (4) available information of statistical properties of the parameters. Generally, no particular set of procedures applies to all problems involved and the particular procedures utilized are dictated by the nature of the aspects stated. The procedures discussed in the previous sections comprise a set of methods which are usually adequate to treat most problems of navigation and guidance systems performance analysis, however, often extensions of the methods are required, which are adequately discussed in the references cited.

4.0 REFERENCES

1. Cramer, H. Mathematical Methods of Statistics, Princeton University Press, 1946.

2. Gredenko, B. V. The Theory of Probability, Chelsea Publishing Company, Inc. (1963).

3. Loive, Michel. Probability Theory, D. Van Nostrand Company, Inc. (1963).

4. Uspensky, J. V. Introduction to Mathematical Probability, McGraw-Hill Book Company, Inc., 1937.

5. Mood, A. M. and F. A. Graybill. Introduction to the Theory of Statistics, McGraw-Hill Book Company, Inc., 1963.

6. Birnbaum, Z. W. Introduction to Probability and Mathematical Statistics, Harper & Brothers, New York, 1962.

7. Papoulis, A. Probability, Random Variables and Stochastic Processes, McGraw-Hill Book Company, Inc., 1965.

8. Churchill, Ruel V. Modern Operational Mathematics in Engineering, Mc-Graw-Hill Book Company, Inc. (1944).

9. Sneddon, Ian N. Fourier Transforms, McGraw-Hill Book Company, Inc. (1951).

10. Campbell, George A. and Ronald M. Foster. Fourier Integrals for Practical Applications, D. Van Nostrand Company, Inc. (1948).

11. Ehrenfeld, S. and S. B. Littauer. Introduction to Statistical Method, McGraw-Hill Book Company, Inc., 1964.

12. Dixon, W. J. and F. J. Massey, Jr. Introduction to Statistical Analysis, McGraw-Hill Book Company, Inc.

13. Middleton, D. Introduction to Statistical Communication Theory, McGraw-Hill Book Company, Inc., 1960.

14. Tables of Normal Probability Function, U.S. Department of Commerce National Bureau of Standard Applied Mathematics, Series Vol. #23.

15. "State Determination and/or Estimation," G. E. Townsend, et. al. SID 65-1200-5, February 28, 1966, prepared under Contract NAS8-11495.

16. Laning, J. H. and R. H. Battin. Random Processes in Automatic Control, McGraw-Hill Book Company, Inc., 1956.

17. DeRusso, P. M., R. J. Roy and C. M. Close. State Variables for Engineers, John Wiley and Sons, Inc., 1956.

18. Some Explanetory Notes Supplementing Portions of MIT/IL R-341, R. J. Ruggiero, IL 211-120-63-102, Space and Information Systems Division of NAA, February 1, 1963.

19. "Observation Theory and Sensors," Dr. B. J. Miller, A. S. Abbott, and G. E. Townsend, December 29, 1965, prepared under Contract No. NAS8-11495.

20. Schwartz, Ralph J. and Friedland, Bernard, Linear Systems, McGraw-Hill Book Company, Inc., 1965.

21. Pitman, George R., et al., Inertial Guidance, John Wiley & Sons, Inc.

BIBLIOGRAPHY

1. Fisher, R. A. Statistical Methods and Scientific Inference, Hafner Publishing Co., New York (1956).

2. Good, I. J. Probability and the Weighting of Evidence, Hafner Publishing Co., New York (1950).

3. Jeffreys, Harold. Scientific Inference, Cambridge University Press, London (1957).

4. Kolmogoroff, A. N. Foundations of the Theory of Probability, Chelsea Publishing Company, New York (1950).

5. Lindley, D. V. "Statistical Inference," Journal of the Royal Statistical Society, Series B, Vol. 15, pp. 30-76 (1953).

6. Savage, Leonard J. The Foundations of Statistics, John Wiley & Sons, Inc., New York (1954).

7. Hogg, R., and A. Craig. Introduction to Mathematical Statistics, The MacMillan Co., New York (1959).

8. Rao, C. R. Advanced Statistical Methods in Biometric Research, John Wiley & Sons, Inc., New York (1952).

9. Bross, I. D. J. Design for Decision, The MacMillan Company, New York (1953).

10. Cochran, W. Sampling Techniques, John Wiley & Sons, Inc., New York (1953).

11. Dixon, W. J., and F. J. Massey, Jr. Introduction to Statistical Analysis, McGraw-Hill Book Co., Inc., New York (1957).

12. Neyman, J. First Course in Probability & Statistics, Holt, Rinehart & Winston, New York (1950).

13. Wilks, S. S. Elementary Statistical Analysis, Princeton University Press, Princeton, New Jersey (1948).

14. Chernoff, H. and L. E. Moses. Elementary Decision Theory, John Wiley & Sons, Inc., New York (1959).

15. Fisher, R. A. "On the Mathematical Foundations of Theoretical Statistics," Philosophical Transactions of the Royal Society of London, Series A, Vol. 222 (1922).

16. Fisher, R. A. "Theory of Statistical Estimation," Proceedings of the Cambridge Philosophical Society, Vol. 22 (1925).

17. Lehmann, E. L. Testing Statistical Hypotheses, John Wiley & Sons, Inc., New York (1959).

18. Neyman, J. "Contributions to the Theory of the X^2 Test," Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, California, pp. 239-273 (1949).

19. Rao, C. R. "Information and Accuracy Attainable in the Estimation of Statistical Parameters," Bulletin of the Calcutta Mathematical Society, Vol. 37, p. 81 (1945).

20. Rao, C. R. "Sufficient Statistics and Minimum Variance Estimates," Proceedings of the Cambridge Philosophical Society, Vol. 45, p. 213 (1948).

21. Thrall, Robert M., Clyde H. Coombs, and Robert L. Doris (Eds.). Decision Processes, John Wiley & Sons, Inc., New York (1954).

22. Wald, Abraham. On the Principles of Statistical Inference, University of Notre Dame Press, Notre Dame, Indiana (1942).

23. Wald, Abraham. Statistical Decision Functions, John Wiley & Sons, Inc., New York (1950).

24. Weiss, Lionel. Statistical Decision Theory, McGraw-Hill Book Co., Inc., New York (1961).

25. Anderson, T. W. An Introduction to Multivariate Statistical Analysis, John Wiley and Sons, Inc., New York (1958).

26. Graybill, F. A. An Introduction to Linear Statistical Models, Vol. I, McGraw-Hill Book Co., Inc., New York (1961).

27. Hald, A. Statistical Theory with Engineering Applications, John Wiley & Sons, Inc., New York (1952).

28. Anderson, R. L. and T. A. Bancroft. Statistical Theory in Research, McGraw-Hill Book Co., Inc., New York (1952).

29. Kendall, M. G. The Advanced Theory of Statistics, Vols. I and II, Charles Griffin & Co., Ltd., London (1946).

30. Parzen, E. Modern Probability Theory and Its Applications, John Wiley & Sons, Inc., New York (1960).

31. Blackwell, D. and M. A. Girochick. Theory of Games and Statistical Decisions, John Wiley & Sons, Inc., New York (1954).

32. Lehmann, E. L. "Some Principles of the Theory of Hypothesis Testing," Annals of Mathematical Statistics, Vol. 2, pp. 1-26 (1950).

33. Lehmann, E. L. Testing Statistical Hypotheses, John Wiley & Sons, Inc., New York (1959).

34. Neyman, J. and E. S. Pearson. "On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference," Biometrika, Vol. 20A (1928), pp. 175-240, 263-294.

35. Stein, C. M. "A Property of Some Tests of Composite Hypotheses," Annals of Mathematical Statistics, Vol. 22, pp. 475-476 (1951).

36. Wald, Abraham. "Contributions to the Theory of Statistical Estimation and Testing Hypotheses," Annals of Mathematical Statistics, Vol. 10, pp. 299-326 (1939).

37. Fisher, R. A. "The Goodness of Fit and Regression Formulae, and the Distribution of Regression Coefficients," Journal of The Royal Statistical Society, Vol. 85, Part IV (1922).

38. Kolodziejczyk, S. "On an Important Class of Statistical Hypotheses," Biometrika, Vol. 27 (1935).

39. Kullbock, S. Information Theory and Statistics, John Wiley & Sons, New York (1959).

40. Cochran, W. G. "Some Methods for Strengthening the Common Chi-Square Tests," Biometrika, Vol. 10, p. 417 (1954).

41. Croxton, F. E. and D. J. Cowden. Applied General Statistics, Second Ed., Prentice-Hall, Inc., New York (1955).

42. Chernoff, H. and E. L. Lehmann. "The Use of Maximum Likelihood Estimates in X^2 Tests for Goodness of Fit," Annals of Mathematical Statistics, Vol. 25, p. 579 (1954).

43. Edwards, A. L. Statistical Analysis, Rinehart & Co., Inc., New York (1946).

44. Eisenhart, C., M. W. Hastay, and W. A. Wallis. Techniques of Statistical Analysis, McGraw-Hill Book Co., Inc., New York (1947).

45. Ferris, C. P., F. E. Grubbs, and C. L. Weaver. "Operating Characteristics for the Common Tests of Significance," Annals of Mathematical Statistics, Vol. 17, p. 178 (1946).

46. Fry, T. C. "The X^2 Test of Significance," Journal of the American Statistical Association, Vol. 33, p. 513 (1938).

47. Goulden, C. H. Methods of Statistical Analysis, Second Edition, John Wiley & Sons, Inc., New York (1952).

48. Hald, A. Statistical Tables and Formulas, John Wiley & Sons, Inc., New York (1952).

49. Leland, O. M. Practical Least Squares, McGraw-Hill Book Co., Inc., New York (1921).

50. Massey, F. J. "The Kolmogorov-Smirnov Test for Goodness of Fit," Journal of the American Statistical Association, Vol. 46, p. 68 (1951).

51. National Bureau of Standards, "Tables of Probability Functions," Vol. I and II, Mathematical Tables 8, Washington, D. C. (1941).

52. Neyman, J. and E. S. Pearson. "On the Problem of the Most Efficient Tests of Statistical Hypotheses," Philosophical Transactions of the Royal Society (London), Series A, Vol. 231, p. 289 (1933).

53. Neyman, J. and B. Zokarska. "Errors of the Second Kind in Testing 'Student's' Hypothesis," Journal of the American Statistical Association, Vol. 31, p. 318 (1936).

54. Wald, A. Sequential Analysis, John Wiley & Sons, Inc., New York (1947).

55. Walker, H. M. and J. Lev. Statistical Inference, Henry Holt & Co., New York (1953).

56. Wallis, W. A. "Tolerance Intervals for Linear Regression," Second Berkeley Symposium on Mathematical Statistics and Probability, p. 43, J. Neyman (Ed.), University of California Press, Berkeley, California (1951).

57. Orley, N. and K. R. Buch. Introduction to the Theory of Probability and Statistics, John Wiley & Sons, New York (1950).

58. Doob, J. L. Stochastic Processes, John Wiley & Sons, New York (1950).

59. Barlett, M. S. An Introduction to Stochastic Processes, Cambridge Press, New York (1955).

APPENDIX A
SOME USEFUL MULTIPLE INTEGRALS

In statistical analyses the following multiple integrals often arise.

$$I_1(\underline{s}) = \int_{D(\underline{x})} exp\left[\underline{s}^T\underline{x} - \underline{x}^T A\,\underline{x}\right] d\underline{x}$$

$$I_2(\underline{s}) = \int_{D(\underline{x})} G(\underline{s}^T\underline{x})\, exp - \left[\underline{x}^T A\,\underline{x}\right] d\underline{x}$$

$$I_3(\underline{\omega}) = \int_{D(\underline{x})} exp\left[J\,\underline{\omega}^T\underline{x} - \underline{x}^T A\,\underline{x}\right] d\underline{x}$$

where $\int_{-\infty}^{\infty} G(u)\,du$ exists, A is an nxn symmetrical positive definite matrix, $J = \mp\sqrt{-1}$, $\underline{x}$ is an arbitrary vector of dimension n, and $\underline{s}$ is a constant vector of dimension n. It is understood that the integrals are multiple integrals over the domain of $\underline{x}$. It is noted that the integrals are always encountered in statistical analyses which involve Gaussian random vectors.

The evaluation of $I_1(\underline{s})$ is facilitated by a linear transformation of coordinates such that the quadratic form $\underline{x}^T A\,\underline{x}$ is diagonalized, thus, let $\underline{x} = M\underline{Y}$ where M is the modal matrix for A, i.e., $M^TAM = \Lambda$ where Λ is the diagonal matrix of the eigenvalues of A, $M^TM = I$ and $|M| = 1$. In this manner, $I_1(\underline{s})$ becomes

$$I_1(\underline{s}) = \int_{D(\underline{Y})} exp\left[\underline{s}^T M\,\underline{Y} - \underline{Y}^T \Lambda\,\underline{Y}\right] d\underline{Y}$$

A second linear transformation can be made such that $\underline{Y}^T \Lambda \underline{Y} = \underline{Z}^T \underline{Z}$, which is essentially a scaling of coordinates, i.e., let $\underline{Y} = D\underline{Z}$ where $D = \Lambda^{-\frac{1}{2}}$, or D is a diagonal matrix with elements equal to the reciprocal of the positive square root of the corresponding elements of Λ. Thus,

$$I_1(\underline{S}) = \int_{D(\underline{Z})} exp\left[\underline{S}^T M D \underline{Z} - \underline{Z}^T \underline{Z}\right] |\Lambda|^{-1/2} d\underline{Z}$$

$$= |\Lambda|^{-1/2} \int_{D(\underline{Z})} exp\left[\underline{S}^T M D \underline{Z} - \underline{Z}^T \underline{Z}\right] d\underline{Z}$$

$$I_1(\underline{S}) = |\Lambda|^{-1/2} \int_{D(\underline{Z})} exp\left[\underline{Z}^T D M^T \underline{S} - \underline{Z}^T \underline{Z}\right] d\underline{Z}$$

By a translation of coordinates the exponent in $I_1(\underline{s})$ can be expressed in terms of a "square", i.e., if $\underline{Z} = \underline{\omega} + \underline{c}$, where $\underline{c}$ is a constant vector, then the exponent becomes

$$(\underline{\omega} + \underline{c})^T D M^T \underline{S} - (\underline{\omega} + \underline{c})^T (\underline{\omega} + \underline{c}) = \underline{\omega}^T D M^T \underline{S} + \underline{c}^T D M^T \underline{S} - \underline{\omega}^T \underline{\omega} - 2\,\underline{\omega}^T \underline{c} - \underline{c}^T \underline{c}$$

Now, if $\underline{c} = \frac{1}{2} D M^T \underline{s}$ then $2\underline{\omega}^T \underline{c} = \underline{\omega}^T D M^T \underline{s}$, $\underline{c}^T D M^T \underline{s} = \frac{1}{2} \underline{s}^T M D D M^T \underline{s}$, $\underline{c}^T \underline{c} = \frac{1}{4} \underline{s}^T M D D M^T \underline{s}$ and the exponent becomes

$$-\underline{\omega}^T \underline{\omega} + \tfrac{1}{4} \underline{S}^T M D D M^T \underline{S} =$$

$$\tfrac{1}{4} \underline{S}^T M D^2 M^T \underline{S} - \underline{\omega}^T \underline{\omega} =$$

$$\tfrac{1}{4} \underline{S}^T M \Lambda^{-1} M^T \underline{S} - \underline{\omega}^T \underline{\omega}$$

Thus, $I_1(\underline{s})$ can be written as

$$I_1(\underline{s}) = |\Lambda|^{1/2} \exp\left[\tfrac{1}{4}\underline{s}^T(M\Lambda^{-1}M^T)\underline{s}\right]\int_{D(\underline{w})} \exp\left[-\underline{w}^T\underline{w}\right] d\underline{w}$$

$$= |\Lambda|^{-1/2} \exp\left[\tfrac{1}{4}\underline{s}^T(A^{-1})\underline{s}\right]\int_{D(\underline{w})} \exp\left[-\sum_{i=1}^{n} w_i^2\right] d\underline{w}$$

$$= |\Lambda|^{-1/2} \exp\left[\tfrac{1}{4}\underline{s}^T A^{-1}\underline{s}\right]\left[\int_{-\infty}^{\infty} e^{-w^2} dw\right]^n$$

$$I_1(\underline{s}) = \sqrt{\frac{\pi^n}{|A|}} \exp\left[\tfrac{1}{4}\underline{s}^T A^{-1}\underline{s}\right]$$

The last step follows from $\int_{-\infty}^{\infty} e^{-\omega^2} d\omega = \sqrt{\pi}$ and that if $M^TAM = \Lambda$ and $|M| = 1$ then $\Lambda^{-1} = M^{-1}A^{-1}M^{-T}$ or $M\Lambda^{-1}M^T = A^{-1}$ and $|\Lambda| = |A|$.

The integral $I_2(\underline{s})$ can be evaluated by using the two linear transformation $\underline{x} = M\underline{Y}$ and $\underline{Y} = D\underline{Z}$ as defined above. In this manner $I_2(\underline{s})$ becomes

$$I_2(\underline{S}) = \int_{D(\underline{X})} G(\underline{S}^T \underline{X}) \exp\left[-\underline{X}^T A \underline{X}\right] d\underline{X}$$

$$= \int_{D(\underline{Y})} G(\underline{S}^T M \underline{Y}) \exp\left[-\underline{Y}^T \Lambda \underline{Y}\right] d\underline{Y}$$

$$I_2(\underline{S}) = |\Lambda|^{-1/2} \int_{D(\underline{Z})} G(\underline{Z}^T D M^T \underline{S}) \exp\left[-\underline{Z}^T \underline{Z}\right] d\underline{Z}$$

Now, let $\underline{Z} = Q\underline{W}$ where Q is an orthogonal matrix which rotates the coordinates such that one axis is co-linear with the vector $DM^T \underline{S}$, i. e., Q is a set of orthogonal vectors $\underline{g}_i$ which span the n dimensional space with one vector, say $\underline{g}_1$, co-linear with $DM^T \underline{S}$; therefore,

$$Q^T D M^T \underline{S} = (\alpha, 0, 0, \cdots, 0)$$

and

$$\underline{Z}^T D M^T \underline{S} = \underline{W}^T Q D M^T \underline{S} = \alpha W_1$$

where

$$\alpha^2 = \underline{S}^T M D Q Q^T D M^T \underline{S} = \underline{S}^T M \Lambda^{-1} M^T \underline{S} = \underline{S}^T A^{-1} \underline{S}$$

Thus, $I_2(\underline{S})$ becomes

$$I_2(\underline{S}) = |\Lambda|^{-1/2} \int_{D(\underline{W})} G(\underline{W}^T Q^T D M^T \underline{S}) \exp\left[-\underline{W}^T \underline{W}\right] d\underline{W}$$

$$= |\Lambda|^{-1/2} \int_{-\infty}^{\infty} G(\alpha w_1) e^{-w_1^2} dw_1 \int_{D(W-w_1)} \exp\left[-\sum_{i=2}^{n} w_i^2\right] \frac{d\underline{X}}{dw_1}$$

$$= |\Lambda|^{-1/2} \left[\int_{-\infty}^{\infty} exp\left[-w^2\right] dw \right]^{n-1} \int_{-\infty}^{\infty} G(\alpha u)\, e^{-u^2}\, du$$

$$= |\Lambda|^{-1/2} \sqrt{\pi^{n-1}} \int_{-\infty}^{\infty} G(\alpha u)\, e^{-u^2}\, du$$

$$I_2(\underline{s}) = \sqrt{\frac{\pi^{n-1}}{|A|}} \int_{-\infty}^{\infty} G\left[\left(\sqrt{\underline{s}^T A^{-1} \underline{s}}\right) u\right] e^{-u^2}\, du$$

The integral $I_3(\underline{s})$ can be evaluated as a special case of $I_1(\underline{s})$, i.e., simply let $\underline{s} = J\underline{\omega}$ in $I_1(\underline{s})$ given above. The result is

$$I_3(\underline{\omega}) = \sqrt{\frac{\pi^n}{|A|}}\, exp\left[-\tfrac{1}{4}\, \underline{\omega}^T A^{-1} \underline{\omega}\right]$$

Thus, the integrals $I_1(\underline{s})$, $I_2(\underline{s})$ and $I_3(\underline{\omega})$ become

$$I_1(\underline{S}) = \sqrt{\frac{\pi^n}{|A|}}\, exp\left[\tfrac{1}{4}\, \underline{S}^T A^{-1} \underline{S}\right]$$

$$I_2(\underline{S}) = \sqrt{\frac{\pi^{n-1}}{|A|}} \int_{-\infty}^{\infty} G_1(\alpha u)\, e^{-u^2}\, du$$

$$I_3(\underline{\omega}) = \sqrt{\frac{\pi^n}{|A|}}\, exp\, -\left[\tfrac{1}{4}\, \underline{\omega}^T A^{-1} \underline{\omega}\right]$$

where

$$\alpha^2 = \underline{S}^T A^{-1} \underline{S}$$

APPENDIX B
MULTIVARIATE GAUSSIAN PROBABILITY DENSITY FUNCTION

Joint Probability Density Function

Let $f(\underline{x})$ be a pdf for the random vector $\underline{x}$ which has the following general form

$$f(\underline{x}) = K \exp -(\underline{x}^T A \underline{x})$$

Where A is a symmetrical positive definite matrix. If

$$\int_{D(\underline{x})} f(\underline{x})\, d\underline{x} = 1$$

then $f(\underline{x})$ has the basic property of a pdf for $\underline{x}$. Now

$$\int_{D(\underline{x})} f(\underline{x})\, d\underline{x} = K \int_{D(\underline{x})} \exp -(\underline{x}^T A \underline{x})\, d\underline{x} = K I_1 (\underline{s} = 0) = K \sqrt{\frac{\pi^n}{|A|}}$$

where $I_1(\underline{s})$ is given in Appendix A. Thus, if $K = \left[\pi^n |A|^{-1}\right]^{1/2}$ then

$$f(\underline{x}) = \sqrt{\frac{|A|}{\pi^n}} \exp -(\underline{x}^T A \underline{x})$$

is a pdf for $\underline{x}$. The moment generating function for $\underline{x}$ becomes

$$mgf_x(\underline{s}) = E\left[\exp(\underline{s}^T \underline{x})\right]$$

$$= \int_{D(\underline{x})} \exp(\underline{s}^T \underline{x})\, f(\underline{x})\, d\underline{x}$$

$$= \sqrt{\frac{|A|}{\pi^n}} \int_{D(\underline{x})} exp\left[\underline{s}^T\underline{x} - x^T A\,\underline{x}\right] d\underline{x}$$

$$= \sqrt{\frac{|A|}{\pi^n}}\; I_1(\underline{s})$$

$$mgf_x(\underline{s}) = exp\left[1/4\; \underline{s}^T A^{-1}\underline{s}\right]$$

The first moments of $\underline{x}$ are found to be zero, i.e.,

$$E(\underline{x}) = \frac{\partial}{\partial \underline{s}}\; mgf_x(\underline{s})\Big|_{s=0} = \underline{0}$$

Since $E(\underline{x}) = \underline{0}$ the covariance matrix, Γ_x, for $\underline{x}$ can be found as follows.

$$\Gamma_x = \frac{\partial^2}{\partial \underline{s}\,\partial \underline{s}^T}\; mgf_x(\underline{s})\Big|_{\underline{s}=0} = 1/2\, A^{-1}$$

Thus, noting that $A = 2^{-1}\,\Gamma_x^{-1}$ AND $|A| = 2^{-n}/|\Gamma_x|^{-1}$ $f(\underline{x})$ can be written as follows.

$$f(\underline{x}) = \frac{1}{\sqrt{(2\pi)^n\,|\Gamma_x|}}\; exp - 1/2\,(\underline{x}^T \Gamma_x^{-1}\underline{x})$$

If the random vector $\underline{x}$ has the pdf $f(\underline{x})$ then $E(\underline{x}) = \underline{0}$, however, if $z = \underline{x} + \underline{m}$ then $E(\underline{z}) = \underline{m}$ and $\Gamma_z = \Gamma_x$. The Jacobian of the transformation $\underline{z} = \underline{x} + \underline{m}$ is simply "one" and; hence, the pdf of z is simply $f(\underline{z}) = f(\underline{x} = \underline{z}-\underline{m})$, i.e.,

$$f(\underline{z}) = \frac{1}{\sqrt{(2\pi)^n\,|\Gamma_z|}}\; exp - 1/2\,(\underline{z} - \underline{m})^T \Gamma_z^{-1}(\underline{z} - \underline{m})$$

The moment generating function for $\underline{z}$ is given by

$$mgf_z(\underline{s}) = E\left[exp\,(\underline{s}^T\underline{z})\right]$$

$$= \frac{1}{\sqrt{(2\pi)^n\,|\Gamma_z|}} \int_{D(\underline{z})} exp\left[\underline{s}^T(\underline{z} - \underline{m}) - 1/2(\underline{z}-m)^T\Gamma_z^{-1}(\underline{z}-\underline{m})\right] d\underline{z}$$

$$= \frac{exp(\underline{s}^T \underline{m})}{\sqrt{(2\pi)^n |\Gamma_{\underline{z}}|}} \int_{D(\underline{z})} exp\left[\underline{s}^T(\underline{z} - \underline{m}) - \tfrac{1}{2}(\underline{z}-\underline{m})^T \Gamma_{\underline{z}}^{-1}(\underline{z}-\underline{m})\right] d\underline{z}$$

$$= \frac{exp(\underline{s}^T \underline{m})}{\sqrt{(2\pi)^n |\Gamma_{\underline{z}}|}} \int_{D(\underline{x})} exp\left[\underline{s}^T \underline{x}^T \Gamma_{\underline{z}}^{-1} \underline{x}\right] d\underline{x}$$

$$mgf_{\underline{z}}(\underline{s}) = exp\left[\underline{s}^T \underline{m} + \tfrac{1}{2} \underline{s}^T \Gamma_{\underline{z}} \underline{s}\right]$$

It is easily seen that

$$E(\underline{z}) = \frac{\partial}{\partial \underline{s}^T} mgf_{\underline{z}}(\underline{s}) \Big|_{\underline{s}=0} = \underline{m}$$

Therefore, the pdf $f(\underline{x})$ can be written for a random vector $\underline{x}$ with $E(\underline{x}) = \underline{m} \neq \underline{0}$.

Thus, if $\underline{x}$ is a random vector with a pdf given by

$$f(\underline{x}) = \frac{1}{\sqrt{(2\pi)^n |\Gamma_x|}} exp - \tfrac{1}{2} (\underline{x} - \underline{m})^T \Gamma_x^{-1} (\underline{x} - \underline{m})$$

then $E(\underline{x}) = \underline{m}$ and the covariance matrix for $\underline{x}$ is $\Gamma_{\underline{x}}$. This pdf is defined as the "multivariate" Gaussian pdf and $f(\underline{x})$ is the joint pdf for the components of $\underline{x}$. The moment generating function for $\underline{x}$ is given by

$$mgf_x(\underline{s}) = exp\left[\underline{s}^T \underline{m} + \tfrac{1}{2} \underline{s}^T \Gamma_x \underline{s}\right]$$

Marginal Probability Density Functions

Let $\underline{x}$ be composed of two subvectors $\underline{x}_1$ and $\underline{x}_2$ such that

$$E(\underline{x}_1) = \underline{m}_1 \qquad E(\underline{x}_2) = \underline{m}_2$$

$$\Gamma_{11} = E\left[(\underline{x}_1 - \underline{m}_1)(\underline{x}_1 - \underline{m}_1)^T\right]$$

$$\Gamma_{22} = E\left[(\underline{x}_2 - \underline{m}_2)(\underline{x}_2 - \underline{m}_2)^T\right]$$

$$\Gamma_{12} = E\left[(\underline{x}_1 - \underline{m}_1)(\underline{x}_2 - \underline{m}_2)^T\right]$$

$$\Gamma_{21} = E\left[(\underline{x}_2 - \underline{m}_2)(\underline{x}_1 - \underline{m}_1)^T\right] = \Gamma_{12}^T$$

It is easily seen that

$$\Gamma_x = E\left[(\underline{x} - \underline{m})(\underline{x} - \underline{m})^T\right] = \left[\begin{array}{c|c} \Gamma_{11} & \Gamma_{12} \\ \hline \Gamma_{21} & \Gamma_{22} \end{array}\right]$$

where

$$\underline{x} = \left[\begin{array}{c} \underline{x}_1 \\ \hline \underline{x}_2 \end{array}\right]$$

Now, if $f(\underline{x})$ is a multivariate Gaussian pdf, as given above, then the marginal pdfs of $\underline{x}_1$ and $\underline{x}_2$ are also Gaussian as given below:

$$f(\underline{x}_1) = \frac{1}{\sqrt{(2\pi)^{n_1}|\Gamma_{11}|}} \exp -1/2\,(\underline{x}_1 - \underline{m}_1)^T \Gamma_{11}^{-1} (\underline{x}_1 - \underline{m}_1)$$

$$f(\underline{x}_2) = \frac{1}{\sqrt{(2\pi)^{n_2}|\Gamma_{22}|}} \exp -1/2\,(\underline{x}_2 - \underline{m}_2)\,\Gamma_{22}^{-1} (\underline{x}_2 - \underline{m}_2)$$

where n_1 and n_2 are the dimensions of $\underline{x}_1$ and $\underline{x}_2$, respectively. This can be established as follows. Let $M = \Gamma_x^{-1}$ where

$$M = \left[\begin{array}{c|c} M_{11} & M_{12} \\ \hline M_{21} & M_{22} \end{array}\right]$$

and where $M_{21} = M_{12}^T$ since Γ_x is symmetrical. Now Γ_x M becomes

$$\Gamma_x M = \left[\begin{array}{c|c} \Gamma_{11} & \Gamma_{12} \\ \hline \Gamma_{21} & \Gamma_{22} \end{array}\right] \left[\begin{array}{c|c} M_{11} & M_{12} \\ \hline M_{21} & M_{22} \end{array}\right] = \left[\begin{array}{c|c} \Gamma_{11} M_{11} + \Gamma_{12} M_{21} & \Gamma_{11} M_{12} + \Gamma_{12} M_{22} \\ \hline \Gamma_{21} M_{11} + \Gamma_{22} M_{21} & \Gamma_{21} M_{12} + \Gamma_{22} M_{22} \end{array}\right]$$

Since $\Gamma_x M = \Gamma_x \Gamma_x^{-1} = I$ it follows that

$$\begin{aligned} \Gamma_{11} M_{11} + \Gamma_{12} M_{21} &= I \\ \Gamma_{11} M_{12} + \Gamma_{12} M_{22} &= 0 \\ \Gamma_{21} M_{11} + \Gamma_{22} M_{21} &= 0 \\ \Gamma_{21} M_{12} + \Gamma_{22} M_{22} &= I \end{aligned}$$

thus,

$$\begin{aligned} M_{11} &= (\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21})^{-1} \\ M_{22} &= (\Gamma_{22} - \Gamma_{12} \Gamma_{11}^{-1} \Gamma_{12})^{-1} \\ M_{12} &= -\Gamma_{11}^{-1} \Gamma_{12} M_{22} \\ M_{21} &= -\Gamma_{22}^{-1} \Gamma_{21} M_{11} \end{aligned}$$

The joint pdf of $\underline{x}$ can be written as follows

$$f(\underline{x}) = \frac{|M|^{1/2}}{\sqrt{(2\pi)^n}} \exp - 1/2 \, (\underline{x} - \underline{m})^T M (\underline{x} - \underline{m})$$

$$f(\underline{x}) = \frac{|M|^{1/2}}{\sqrt{(2\pi)^n}} \exp - 1/2 \, \underline{z}^T M \underline{z}$$

where, for convenience, $\underline{z} = \underline{x} - \underline{m}$. Now, $\underline{z}^T M \underline{z}$ can be written in terms of $\underline{z}_1 = \underline{x}_1 - \underline{m}_1$ and $\underline{z}_2 = \underline{x}_2 - \underline{m}_2$, i.e.,

$$\underline{z}^T M \underline{z} = \left[\begin{array}{c} \underline{z}_1 \\ \hline \underline{z}_2 \end{array}\right]^T \left[\begin{array}{c|c} M_{11} & M_{12} \\ \hline M_{21} & M_{22} \end{array}\right] \left[\begin{array}{c} \underline{z}_1 \\ \hline \underline{z}_2 \end{array}\right]$$

$$= \left[\underline{z}_1^T \;\vdots\; \underline{z}_2^T\right] \left[\begin{array}{c} M_{11}\underline{z}_1 + M_{12}\underline{z}_2 \\ \hline M_{21}\underline{z}_1 + M_{22}\underline{z}_2 \end{array}\right]$$

$$= \underline{z}_1^T M_{11}\underline{z}_1 + \underline{z}_1^T M_{12}\underline{z}_2 + \underline{z}_2^T M_{21}\underline{z}_1 + \underline{z}_2^T M_{22}\underline{z}_2$$

$$= \underline{z}_1^T M_{11}\underline{z}_1 + \underline{z}_1^T (M_{12} + M_{21}^T)\underline{z}_2 + \underline{z}_2^T M_{22}\underline{z}_2$$

$$\underline{z}^T M \underline{z} = \underline{z}_1^T M_{11}\underline{z}_1 + 2\,\underline{z}_1^T M_{12}\underline{z}_2 + \underline{z}_2^T M_{22}\underline{z}_2$$

The marginal pdfs of $\underline{x}_1$ and $\underline{x}_2$ are given by

$$f(\underline{x}_1) = \int_{D(\underline{x}_2)} f(\underline{x}_1, \underline{x}_2)\, d\underline{x}_2$$

$$f(\underline{x}_2) = \int_{D(\underline{x}_1)} f(\underline{x}_1, \underline{x}_2)\, d\underline{x}_1$$

Thus,

$$f(\underline{x}_1) = \frac{|M|^{1/2}}{(2\pi)^{n/2}} \exp -\tfrac{1}{2}(\underline{z}_1^T M_{11}\underline{z}_1) \int_{D(\underline{z}_2)} \exp -\tfrac{1}{2}(2\underline{z}_1^T M_{12}\underline{z}_2 + \underline{z}_2^T M_{22}\underline{z}_2)\, d\underline{z}_2$$

$$f(\underline{x}_2) = \frac{|M|^{1/2}}{(2\pi)^{n/2}} \exp -1/2\,(\underline{z}_2^T M_{22}\,\underline{z}_2) \int_{D(\underline{z})} \exp -1/2\,(2\underline{z}_2^T M_{21}\,\underline{z}_1 + \underline{z}_1^T M_{11}\,\underline{z}_1)\, d\underline{z}_1$$

The two integrals can be evaluated using $I_1(\underline{s})$ of Appendix A with $\underline{s} = -M_{21}$ and $-M_{12}$ $_2$ and $A = \frac{1}{2}M_{22}$ and $\frac{1}{2}M_{11}$, respectively. The results are as follows:

$$f(\underline{x}_1) = \frac{|M|^{1/2}}{|M_{22}|^{1/2}\,(2\pi)^{\frac{n_1}{2}}} \exp -1/2\,(\underline{x}_1 - \underline{m}_1)^T (M_{11} - M_{12}\,M_{22}^{-1}\,M_{21})\,(\underline{x}_1 - \underline{m}_1)$$

$$f(\underline{x}_2) = \frac{|M|^{1/2}}{|M_{11}|^{1/2}\,(2\pi)^{n_2/2}} \exp -1/2\,(\underline{x}_2 - \underline{m}_2)^T (M_{22} - M_{21}\,M_{11}^{-1}\,M_{12})(\underline{x}_2 - \underline{m}_2)$$

Using the above equations, it is found that

$$\begin{aligned}(M_{11} - M_{12}\,M_{22}^{-1}\,M_{21}) &= (I + M_{12}\,M_{22}^{-1}\,\Gamma_{22}^{-1}\,\Gamma_{21})\,M_{11} \\ &= (I - \Gamma_{11}^{-1}\,\Gamma_{12}\,\Gamma_{22}^{-1}\,\Gamma_{21})\,M_{11} \\ &= \Gamma_{11}^{-1}\,(\Gamma_{11} - \Gamma_{12}\,\Gamma_{22}^{-1}\,\Gamma_{21})\,M_{11} \\ (M_{11} - M_{12}\,M_{22}^{-1}\,M_{21}) &= \Gamma_{11}^{-1}\end{aligned}$$

Similarly,

$$M_{22} - M_{21}\,M_{11}^{-1}\,M_{12} = \Gamma_{22}^{-1}$$

Therefore, $f(\underline{x}_1)$ and $f(\underline{x}_2)$ become

$$f(\underline{x}_1) = \frac{|M|^{1/2}}{|M_{22}|^{1/2}\,(2\pi)^{n/2}} \exp -1/2\,(\underline{x}_1 - \underline{m}_1)^T\,\Gamma_{11}^{-1}\,(\underline{x}_1 - \underline{m}_1)$$

$$f(\underline{x}_2)=\frac{|M|^{1/2}}{|M_{11}|^{1/2}\,(2\pi)^{n_2/2}}\exp -1/2\,(\underline{x}_2-\underline{m}_2)^T\,\Gamma_{22}^{-1}\,(\underline{x}_2-\underline{m}_2)$$

Now, consider the following matrix product.

$$C = M\begin{bmatrix} I & 0 \\ 0 & M_{22}^{-1} \end{bmatrix}\begin{bmatrix} I & 0 \\ 0 & M_{21} \end{bmatrix}\begin{bmatrix} I & 0 \\ -I & I \end{bmatrix}$$

$$C = \begin{bmatrix} M_{11} - M_{12}\,M_{22}^{-1}\,M_{21} & M_{12}\,M_{22}^{-1}\,M_{21} \\ 0 & M_{21} \end{bmatrix}$$

Using the rules of determinants it is found that

$$|C| = |M|\,|M_{22}|^{-1}\,|M_{21}|\cdot 1 = |M_{21}|\,|M_{11} - M_{12}\,M_{22}^{-1}\,M_{21}|$$

or

$$\frac{|M|}{|M_{22}|} = |M_{11} - M_{12}\,M_{22}^{-1}\,M_{21}| = |\Gamma_{11}|^{-1}$$

Similarly,

$$\frac{|M|}{|M_{11}|} = |\Gamma_{22}|^{-1}$$

Therefore, $f(\underline{x}_1)$ and $f(\underline{x}_2)$ become

$$f(\underline{x}_1) = \frac{1}{\sqrt{(2\pi)^{n_2}\,|\Gamma_{11}|}}\exp -\frac{1}{2}(\underline{x}_1 - \underline{m}_1)^T\,\Gamma_{11}^{-1}(\underline{x}_1 - \underline{m}_1)$$

$$f(\underline{x}_2) = \frac{1}{\sqrt{(2\pi)^{n_2}\,|\Gamma_{22}|}}\, exp - 1/2\,(\underline{x}_2 - \underline{m}_2)^T\, \Gamma_{22}^{-1}(\underline{x}_2 - \underline{m}_2)$$

In particular, the marginal pdf of any component x_i of $\underline{x}$ is as follows

$$f(x_i) = \frac{1}{\sqrt{2\pi\,\sigma_i^2}}\, exp - 1/2\, \frac{(x_i - m_i)^2}{\sigma_i^2}$$

for i = 1, 2, . . . , n.

It is important to note that the marginal pdf for any subset of a set of Gaussian random variables is also a Gaussian pdf, i.e., if the random vector $\underline{x}$ has a joint Gaussian probability density function then any subvector of $\underline{x}$ has a marginal probability density function which is Gaussian. The elements which specify the marginal probability density function are simply the corresponding elements of the vector $E(\underline{x}) = \underline{m}$ and the covariance matrix Γ_x. Thus, the marginal probability density function for any subset of $\underline{x}$ can be specified directly from the joint probability density function of $\underline{x}$.

Conditional Probability Density Functions

Let $\underline{x}$ be composed of two subvectors $\underline{x}_1$ and $\underline{x}_2$ with $\underline{m}_1$, $\underline{m}_2$, Γ_{11}, Γ_{22}, Γ_{12} and Γ_{21} as defined above. If $f(\underline{x})$ is a multivariate Gaussian pdf as defined above, then the conditional pdfs $f(\underline{x}_1/\underline{x}_2)$ and $f(\underline{x}_2/\underline{x}_1)$ are Gaussian as given below.

$$f(\underline{x}_1/\underline{x}_2) = \frac{|M_{11}|^{1/2}}{(2\pi)^{n_1/2}}\, exp - 1/2\left[\underline{z}_1 - \Gamma_{12}\,\Gamma_{22}^{-1}\,\underline{z}_2\right]^T M_{11}\left[\underline{z}_1 - \Gamma_{12}\,\Gamma_{22}^{-1}\,\underline{z}_2\right]$$

$$f(\underline{x}_2/\underline{x}_1) = \frac{|M_{22}|^{1/2}}{(2\pi)^{n_2/2}}\, exp - 1/2\left[\underline{z}_2 - \Gamma_{21}\,\Gamma_{11}^{-1}\,\underline{z}_1\right] M_{22}\left[\underline{z}_2 - \Gamma_{21}\,\Gamma_{11}^{-1}\,\underline{z}_1\right]$$

where

$$\underline{z}_1 = \underline{x}_1 - \underline{m}_1$$
$$z_2 = \underline{x}_2 - \underline{m}_2$$
$$M_{11} = (\Gamma_{11} - \Gamma_{12}\,\Gamma_{22}^{-1}\,\Gamma_{21})^{-1}$$
$$M_{22} = (\Gamma_{22} - \Gamma_{21}\,\Gamma_{11}^{-1}\,\Gamma_{12})^{-1}$$

These results can be established in the following manner.

The conditional pdfs $f(\underline{x}_1/\underline{x}_2)$ and $f(\underline{x}_2/\underline{x}_1)$ can be expressed as ratios of $f(\underline{x})$ to $f(\underline{x}_2)$ and $f(\underline{x}_1)$, respectively, i.e.,

$$f(\underline{x}_1/\underline{x}_2) = \frac{f(\underline{x})}{f(\underline{x}_2)} = \frac{f(\underline{x}_1, \underline{x}_2)}{f(\underline{x}_2)}$$

$$f(\underline{x}_2/\underline{x}_1) = \frac{f(\underline{x})}{f(\underline{x}_1)} = \frac{f(\underline{x}_1, \underline{x}_2)}{f(\underline{x}_1)}$$

Substituting the previous results, it is found that

$$f(\underline{x}_1/\underline{x}_2) = \frac{\sqrt{|M||\Gamma_{22}|}}{(2\pi)^{n/2}} \exp - 1/2\,(\underline{z}^T M \underline{z} - \underline{z}_2^T \Gamma_{22}^{-1} \underline{z}_2)$$

$$f(\underline{x}_2/\underline{x}_1) = \frac{\sqrt{|M||\Gamma_{11}|}}{(2\pi)^{n/2}} \exp - 1/2\,(\underline{z}^T M \underline{z} - \underline{z}_1^T \Gamma_{11}^{-1} \underline{z}_1)$$

It was shown that $|M| \cdot |\Gamma_{11}| = |M_{22}|$ and $|M| \cdot |\Gamma_{22}| = |M_{11}|$
The exponents can be expanded as follows.

$$\underline{z}^T M \underline{z} - \underline{z}_2^T \Gamma_{22}^{-1} \underline{z}_2 = \underline{z}_1^T M_{11} \underline{z}_1 + 2 \underline{z}_1^T M_{12} \underline{z}_2 + \underline{z}_2^T M_{22} \underline{z}_2 + - \underline{z}_2^T (M_{22} - M_{21} M_{11}^{-1} M_{12}) \underline{z}_2$$

$$= \underline{z}_1^T M_{11} \underline{z}_1 + 2 \underline{z}_1^T M_{12} \underline{z}_2 + \underline{z}_2^T M_{21} M_{11}^{-1} M_{12} \underline{z}_2$$

$$\underline{z}^T M \underline{z} - \underline{z}_2^T \Gamma_{22}^{-1} \underline{z}_2 = (\underline{z}_1 + M_{11}^{-1} M_{12} \underline{z}_2)^T M_{11} (\underline{z}_1 + M_{11}^{-1} M_{12} \underline{z}_2)$$

Similarly,

$$\underline{z}^T M \underline{z} - \underline{z}_1 \Gamma_{11}^{-1} \underline{z}_1 = (\underline{z}_2 + M_{22}^{-1} M_{21} \underline{z}_1)^T M_{22} (\underline{z}_2 + M_{22}^{-1} M_{21} \underline{z}_1)$$

Since $M_{12} = -\Gamma_{11}^{-1} \Gamma_{12} M_{22}$ and $M_{21} = -\Gamma_{22}^{-1} \Gamma_{21} M_{11}$, it follows that $M_{22}^{-1} M_{21} = -\Gamma_{21} \Gamma_{11}^{-1}$ and $M_{11}^{-1} M_{12} = -\Gamma_{12} \Gamma_{22}^{-1}$. Thus, $f(\underline{x}_1/\underline{x}_2)$ and $f(\underline{x}_2/\underline{x}_1)$ become

$$f(\underline{x}_1/\underline{x}_2) = \frac{|M_{11}|^{1/2}}{(2\pi)^{n/2}} \exp -1/2 \left[\underline{z}_1 - \Gamma_{12} \Gamma_{22}^{-1} \underline{z}_2\right]^T M_{11} \left[\underline{z}_1 - \Gamma_{12} \Gamma_{22}^{-1} \underline{z}_2\right]$$

$$f(\underline{x}_2/\underline{x}_1) = \frac{|M_{22}|^{1/2}}{(2\pi)^{n/2}} \exp -1/2 \left[\underline{z}_2 - \Gamma_{21} \Gamma_{11}^{-1} \underline{z}_1\right]^T M_{22} \left[\underline{z}_2 - \Gamma_{21} \Gamma_{11}^{-1} \underline{z}_1\right]$$

where the terms are defined above.

It is seen that if the joint pdf for a random vector is Gaussian then the conditional pdfs $f(\underline{x}_1/\underline{x}_2)$ and $f(\underline{x}_2/\underline{x}_1)$ are also Gaussian with covariance matrices M_{11}^{-1} and M_{22}^{-1}, respectively.

It should be noted that the first moments of the conditional pdfs are not the expected values of $\underline{x}_1$ and $\underline{x}_2$. Rather; these moments are the conditional expectations of $\underline{x}_1$, given $\underline{x}_2$, and $\underline{x}_2$, given $\underline{x}_1$; i.e.,

$$E(\underline{X}_1/\underline{X}_2) = \int_{D(\underline{x}_1)} f(\underline{x}_1/\underline{x}_2)\, d\underline{x}_1$$

$$E(\underline{X}_2/\underline{X}_1) = \int_{D(\underline{x}_2)} f(\underline{x}_2/\underline{x}_1)\, d\underline{x}_2$$

Using $I_1(s)$ of Appendix A for $\underline{s} = \underline{0}$, it is found that

$$E(\underline{X}_1/\underline{X}_2) = \underline{m}_1 + \Gamma_{12}\Gamma_{22}^{-1}(\underline{X}_2 - \underline{m}_2)$$

$$E(\underline{X}_2/\underline{X}_1) = \underline{m}_2 + \Gamma_{21}\Gamma_{11}^{-1}(\underline{X}_1 - \underline{m}_1)$$

Note that

$$E(\underline{x}_1/\underline{x}_2 = \underline{m}_2) = \underline{m}_1$$

$$E(\underline{x}_2/\underline{x}_1 = \underline{m}_1) = \underline{m}_2$$

Also, it should be noted that M_{11}^{-1} and M_{22}^{-1} are the conditional covariance matrices for x_1, given x_2, and x_2, given x_1, respectively.

APPENDIX C
SPECIAL GAUSSIAN RANDOM VECTORS

Gaussian random vectors are invaribly encountered in statistical analyses. In system performance analyses the two cases which often arise are those of two and three dimensions; i.e., two or three random variables have a known joint Gaussian pdf and it is desired to determine the probability that the vectors will lie in some specified region. Also, the special case of a single Gaussian random variable often arises, even in the cases of higher order Gaussian random vectors. The three special cases of dimensions one, two and three are considered below.

One Dimensional Case

Let x be a Gaussian random variable with mean value and variance of m and σ^2, respectively. The pdf of x is as follows:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

The probability that x lies in the interval $a \le x \le b$ is given by

$$P[a \le x \le b] = \int_a^b f(x)\,dx = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx$$

An explicit expressed for $P[a \le x \le b]$ cannot be determined since the integral cannot be evaluated in closed form for arbitrary a and b. Therefore, $P[a \le x \le b]$ must be determined by numerical integration of the integral. The integral has been tabulated extensively for the Normal random variable, i.e., for the case of $m = 0$ and $\sigma^2 = 1$. Thus, $P[\alpha \le y \le \beta]$ can be determined from a table for a Normal random variable y. Now y can be expressed in terms of the Gaussian random variable x as follows:

$$y = \frac{x-m}{\sigma}$$

Therefore, $\sigma y = x-m$ or $x = \sigma y + m$ and

$$P[a \le x \le b] = P[\alpha \le y \le \beta]$$

where

$$\alpha = \frac{a-m}{\sigma}, \quad \beta = \frac{b-m}{\sigma}$$

Thus, the method for determining the probability that a Gaussian random variable will lie within some interval is to simply translate and scale the random variable and use a table for a Normal random variable with zero mean value and unity variance. A few values of $P[|y| \le k]$ are given in Table C-1 below. It is noted that $P[|y| \le k] = P[|x-m| \le k\sigma]$. An extensive tabulation of the Normal random variable probability is given in Reference 14 and useful tables are given in References 1, 2, 5 and 6.

k	$P[\|y\| \le k]$
0.500	0.383
1.000	0.683
1.500	0.866
1.645	0.900
1.960	0.950
2.000	0.955
2.576	0.990
3.000	0.997
3.291	0.999

Table C-1: Probabilities for a Normal Random Variable

Two Dimensional Case

Let $\underline{x}$ be a Gaussian random vector with two components x_1 and x_2 such that

$$E(x_1) = m_1$$
$$E(x_2) = m_2$$
$$E(x_1 - m_1)^2 = \sigma_1^2$$
$$E(x_2 - m_2)^2 = \sigma_2^2$$
$$E\left[(x_1 - m_1)(x_2 - m_2)\right] = \mu$$

The joint pdf of $\underline{x}$ is as follows:

$$f(\underline{x}) = \frac{1}{2\pi\sqrt{|\Gamma_x|}} \exp -\tfrac{1}{2} (\underline{x} - \underline{m})^T \Gamma_x^{-1} (\underline{x} - \underline{m})$$

where

$$\Gamma_x = \begin{bmatrix} \sigma_1^2 & \mu \\ \mu & \sigma_1^2 \end{bmatrix} \quad \text{and} \quad \underline{m} = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}$$

It follows that $|\Gamma_x| = \sigma_1^2 \sigma_2^2 - \mu^2$ and

$$\Gamma_x^{-1} = \frac{1}{|\Gamma_x|} \begin{bmatrix} \sigma_2^2 & -\mu \\ -\mu & \sigma_1^2 \end{bmatrix}$$

The covariance matrix Γ_x has two eigenvectors $\underline{\phi}_1$ and $\underline{\phi}_2$ such that $\Gamma_x \underline{\phi}_1 = \lambda_1 \underline{\phi}_1$ and $\Gamma_x \underline{\phi}_2 = \lambda_2 \underline{\phi}_2$ where λ_1 and λ_2 are the eigenvalues of Γ_x which are determined by $(\Gamma_x - \lambda I)\,\underline{\phi} = \underline{0}$ or, equivalently, by

$$|\Gamma_x - \lambda I| = 0$$

Thus, the eigenvalues are roots of the following quadratic equation.

$$(\sigma_1^2 - \lambda)(\sigma_2^2 - \lambda) - \mu^2 = 0$$
$$\lambda^2 - (\sigma_1^2 + \sigma_2^2)\lambda + \sigma_1^2 \sigma_2^2 - \mu^2 = 0$$

It is easily shown that the eigenvalues for Γ_x are given by

$$\lambda_1 = \frac{\sigma_1^2 + \sigma_2^2}{2} + \left[\left(\frac{\sigma_x^2 - \sigma_y^2}{2}\right)^2 + \mu^2\right]^{1/2}$$

$$\lambda_2 = \frac{\sigma_1^2 + \sigma_2^2}{2} - \left[\left(\frac{\sigma_x^2 - \sigma_y^2}{2}\right)^2 + \mu^2\right]^{1/2}$$

Consider the random vector $\underline{y}$ such that $M\underline{y} = (\underline{x} - \underline{m})$ where M is the modal matrix for Γ_x, i.e., $M = [\underline{\phi}_1, \underline{\phi}_2]$ where $\underline{\phi}_1$ and $\underline{\phi}_2$ are the eigenvectors of Γ_x. Thus, $E(\underline{y}) = \underline{0}$ and $\Gamma_y = M^T \Gamma_x M = \Lambda$ where

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

The pdf for $\underline{y}$ becomes

$$f(\underline{y}) = \frac{1}{2\pi\sqrt{|\Lambda|}} \exp -1/2\ (\underline{y}^T \Lambda^{-1} \underline{y})$$

$$= \frac{1}{2\pi\sqrt{\lambda_1 \lambda_2}} \exp -1/2 \left[y_1^2/\lambda_1 + y_2^2/\lambda_2 \right]$$

$$f(\underline{y}) = f(y_1)\, f(y_2)$$

where

$$f(y_1) = \frac{1}{\sqrt{2\pi\lambda_1}} e^{-\frac{1}{2\lambda_1} y_1^2}$$

$$f(y_2) = \frac{1}{\sqrt{2\pi\lambda_2}} e^{-\frac{1}{2\lambda_2} y_2^2}$$

Thus, the random variables y_1 and y_2 are statistically independent Gaussian random variables whose probability of occurrence for certain regions can be readily determined. It will be shown that $P[\underline{y} \in R(\underline{y})]$ can be readily determined for rectangular and elliptical regions.

It should be apparent that the determination of the probability of occurrence for $\underline{y}$ essentially determines the probability of occurrence for $\underline{x}$. That is,

$$P[\underline{y} \in R(\underline{y})] = P[\underline{x} \in R(\underline{x})]$$

where the regions $R(\underline{y})$ and $R(\underline{x})$ are related by the transformation $\underline{y} = M^T(\underline{x} - \underline{m})$. This transformation is simply a translation and rotation of coordinates as depicted in Figure C-1. The translation is simply along the mean vector $\underline{m}$ as shown. The rotation is determined by the eigenvectors of Γ_x or, equivalently, the modal matrix M; however, it is possible to define the angle of rotation, α, directly in terms with the elements of Γ_x. That is, let

$$\underline{\phi}_1 = \begin{bmatrix} \sin\alpha \\ \cos\alpha \end{bmatrix}$$

where $\Gamma_x \underline{\phi}_1 = \lambda_1 \underline{\phi}_1$. Thus,

$$\sigma_1^2 \sin\alpha + \mu \cos\alpha = \eta_1 \sin\alpha$$
$$\mu \sin\alpha + \sigma_2^2 \cos\alpha = \eta_1 \cos\alpha$$

and

$$\tan\alpha = \frac{\sigma_1^2 \sin\alpha + \mu \cos\alpha}{\mu \sin\alpha + \sigma_2^2 \cos\alpha}$$

$$\tan\alpha = \frac{\sigma_1^2 \tan\alpha + \mu}{\mu \tan\alpha + \sigma_2^2}$$

Solving, it is found that

$$\frac{\tan\alpha}{1-\tan^2\alpha} = \frac{\mu}{\sigma_2^2 - \sigma_1^2} = \frac{1}{2}\tan 2\alpha$$

Thus,

$$\alpha = \frac{1}{2}\tan^{-1}\left(\frac{2\mu}{\sigma_2^2 - \sigma_1^2}\right)$$

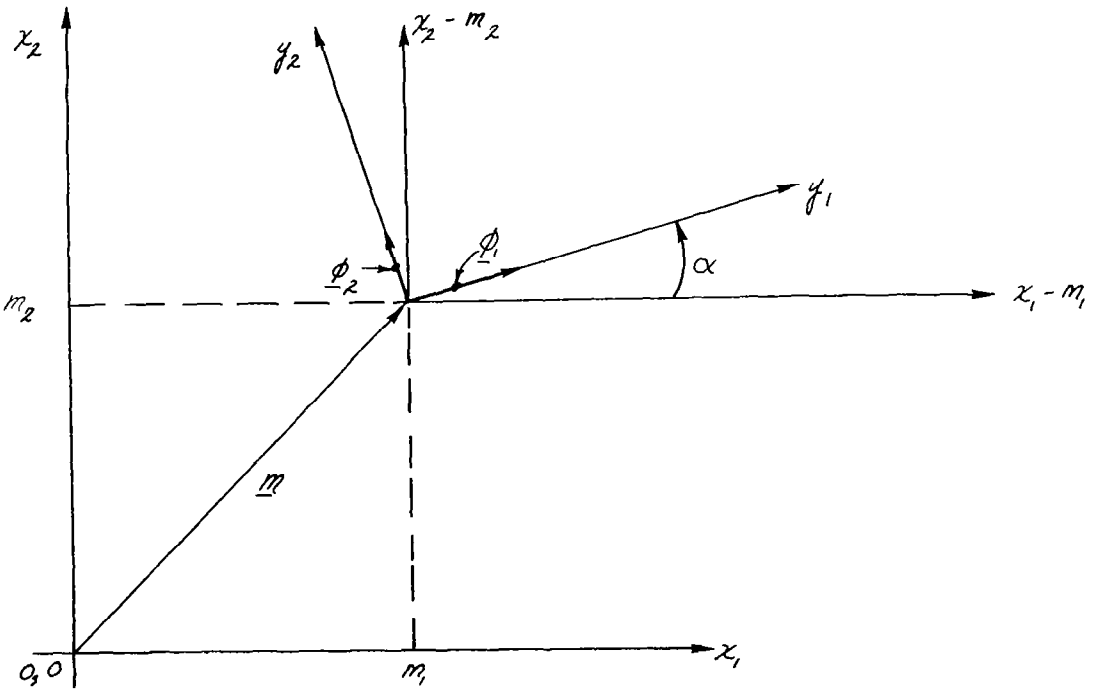

Figure C-1: Transformation of $\underline{y} = M^T(\underline{x} - \underline{m})$

Consider, the rectangular region R shown in Figure C-2, which is defined by

$$a_1 \le y_1 \le a_2$$
$$b_1 \le y_2 \le b_2$$

The probability that $\underline{y}$ lies in R, $P[\underline{y} \in R]$, is given by

$$P[\underline{y} \in R] = \left[\int_{a_1}^{a_2} f(y_1)\,dy_1\right]\left[\int_{b_1}^{b_2} f(y_2)\,dy_2\right]$$

Now, if the limits are expressed in terms of $\sqrt{\lambda_1}$ and $\sqrt{\lambda_2}$ the integrals can be determined from a table for a Normal random variable and this determines $P[\underline{y} \in R]$.

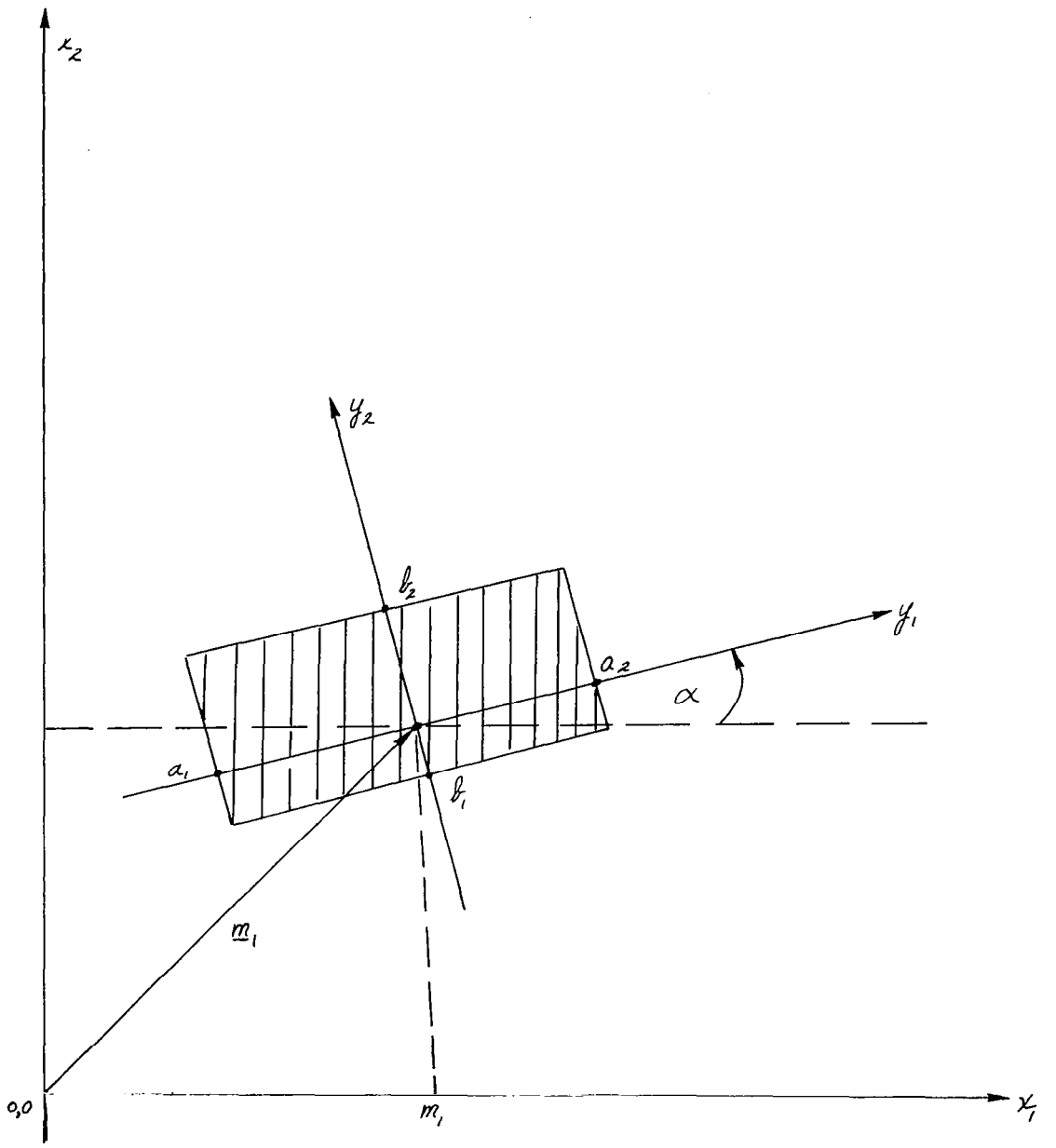

Figure C-2: Rectangular Region

If an elliptical region Δ is used, then, it is possible to obtain a closed form solution for P $[\underline{y} \in \Delta]$. Let Δ denote the elliptical region enclosed by an ellipse which is defined by $\underline{y}^T \Lambda^{-1} \underline{y} = 2\ell_1$, as shown in Figure C-3. It is noted that the vectors $\underline{r}_1$ and $\underline{r}_2$ are the semi-principal axes of the ellipse. The lengths of $\underline{r}_1$ and $\underline{r}_2$ are related to ℓ_1 and are found by setting y_2 and y_1 equal to zero, respectively; to wit,

$$a = \sqrt{2\ell_1 \lambda_1}$$

$$b = \sqrt{2\ell_1 \lambda_2}$$

where a and b are the lengths of $\underline{r}_1$ and $\underline{r}_2$, respectively. It is also noted that the area, A, of the ellipse is given by

$$A = \pi a b = 2\pi \sqrt{\lambda_1 \lambda_2}\, \ell_1$$

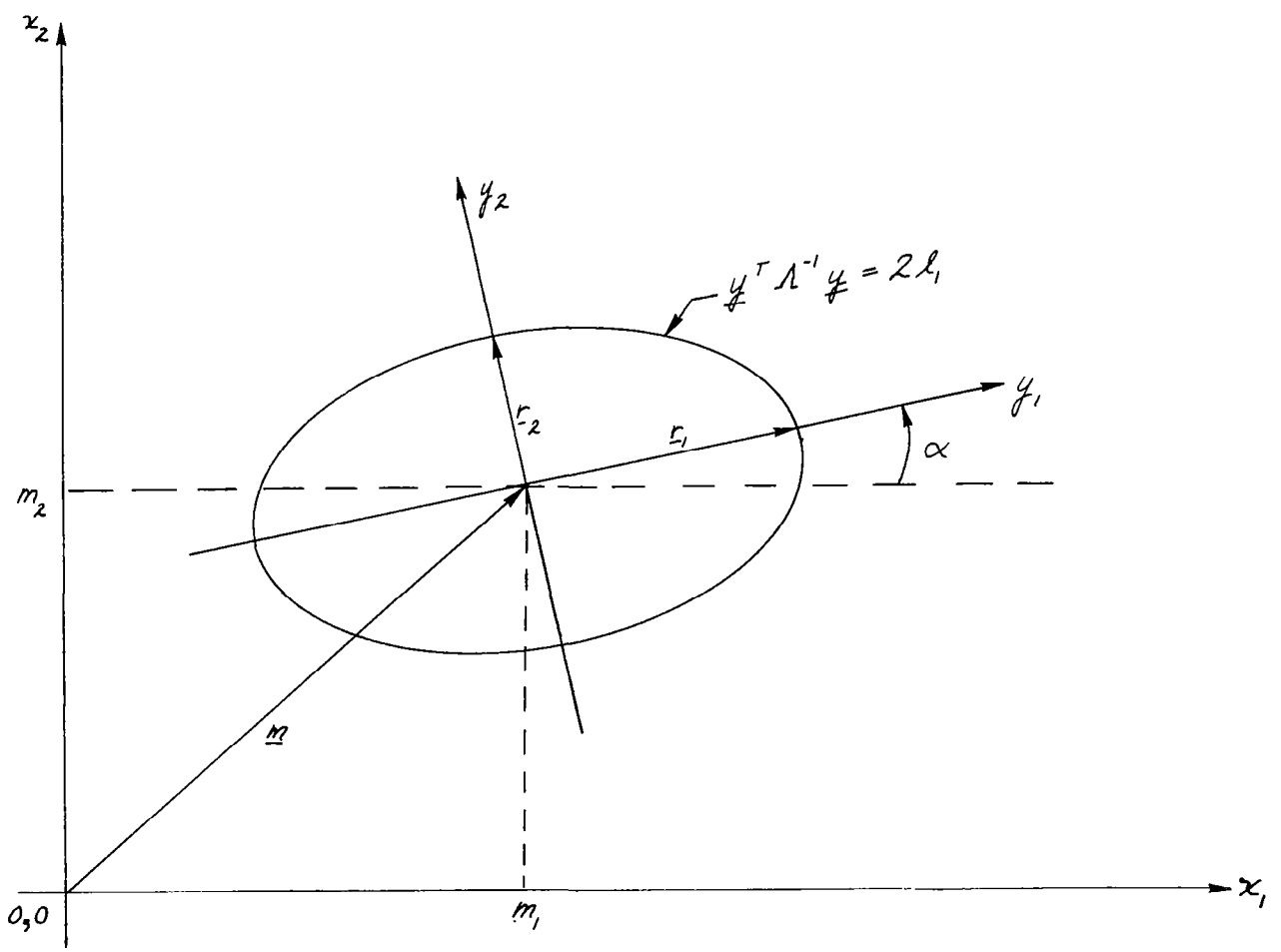

Figure C-3: Elliptical Region

The probability that $\underline{y}$ will lie in Δ , $P\left[\underline{y} \in \Delta\right]$, is given by

$$P\left[\underline{y} \in \Delta\right] = \int_{\Delta} f(\underline{y})\, d\underline{y}$$

$$= \frac{1}{2\pi\sqrt{\lambda_1 \lambda_2}} \iint_{\Delta} \exp - 1/2\, (\underline{y}^T \Lambda^{-1} \underline{y})\, dy_1\, dy_2$$

$$P\left[\underline{y} \in \Delta\right] = \frac{1}{2\pi\sqrt{\lambda_1 \lambda_2}} \iint_{\Delta} e^{-\ell}\, dy_1\, dy_2$$

It is apparent that the probability density is constant along an ellipse defined by $\underline{y}^T \Lambda^{-1} \underline{y} = 2\ell$, thus, it is convenient to change the infinitesimal area $dy_1\, dy_2$ to $dA = 2\pi\sqrt{\lambda_1 \lambda_2}\, d\ell$, where dA is the infinitesimal area lying between two ellipses defined by ℓ and $\ell + d\ell$.

In this manner,

$$P\left[\underline{y} \in \Delta\right] = \frac{1}{2\pi\sqrt{\lambda_1 \lambda_2}} \int_0^{A_1} e^{-\ell}\, dA$$

$$= \int_0^{\ell_1} e^{-\ell}\, d\ell$$

$$P\left[\underline{y} \in \Delta\right] = 1 - e^{-\ell_1}$$

Thus, the probability of occurrence for the elliptical region enclosed by $\underline{y}^T \Lambda^{-1} \underline{y} = 2\ell_1$, is readily determined in terms of ℓ_1. Several values are listed in Table C-2 below, where .

$$P\left[\underline{y} \in \Delta\right] = P(\Delta)$$

ℓ_1	$P(\Delta)$
0.5	0.3935
0.7	0.5034
0.9	0.5934
1.2	0.6988
1.5	0.7769
1.8	0.8347

Table C-2: Probability of Occurrence : $P(\Delta)$

It should be noted that for each value of ℓ_1 the lengths of $\underline{r}_1$ and $\underline{r}_2$ are given in terms of λ_1 and λ_2, e.g., for $\ell_1 = 1/2$, $a = \sqrt{\lambda_1}$ and $b = \sqrt{\lambda_2}$. The probability that $\underline{y}$ will lie in the corresponding ellipse is 0.3935 as noted in Table C-2.

Three Dimensional Case

Let $\underline{x}$ be a Gaussian random vector with three components x_1, x_2 and x_3 with pdf as follows:

$$f(\underline{x}) = \frac{1}{(\sqrt{2\pi})^3 \sqrt{|\Gamma_x|}} \exp -\tfrac{1}{2}(\underline{x} - \underline{m})^T \Gamma_x^{-1} (\underline{x} - \underline{m})$$

where

$$\underline{m} = E(\underline{x})$$

$$\Gamma_x = E\left[(\underline{x} - \underline{m})(\underline{x} - \underline{m})^T \right]$$

The covariance matrix Γ_x has three eigenvectors $\underline{\phi}_i$ corresponding to three eigenvalues λ_i, for $i = 1, 2, 3$; i.e.,

$$\Gamma_x \underline{\phi}_i = \lambda_i \underline{\phi}_i \qquad i = 1, 2, 3$$

The eigenvalues are solutions of $|\Gamma_x - \lambda I|$ = 0 where

$$|\Gamma_x - \lambda I| = \begin{vmatrix} \sigma_1^2 - \lambda & \mu_{12} & \mu_{13} \\ \mu_{12} & \sigma_2^2 - \lambda & \mu_{23} \\ \mu_{13} & \mu_{23} & \sigma_3^2 - \lambda \end{vmatrix}$$

$$= \left[(\sigma_1^2 - \lambda)(\sigma_2^2 - \lambda)(\sigma_3^2 - \lambda) + 2\mu_{12}\mu_{23}\mu_{13}\right] +$$

$$- \left[(\sigma_1^2 - \lambda)\mu_{23}^2 + (\sigma_2^2 - \lambda)\mu_{13}^2 + (\sigma_3^2 - \lambda)\mu_{12}^2\right]$$

Setting $|\Gamma_x - \lambda I|$ equal to zero the following cubic equation is obtained.

$$\lambda^3 + K_1\lambda^2 + K_2\lambda + K_3 = 0$$

where

$$K_1 = -(\sigma_1^2 + \sigma_2^2 + \sigma_3^2)$$

$$K_2 = \sigma_1^2\sigma_2^2 + \sigma_2^2\sigma_3^2 + \sigma_1^2\sigma_3^2 - (\mu_{12}^2 + \mu_{23}^2 + \mu_{13}^2)$$

$$K_3 = \sigma_1^2\mu_{23}^2 + \sigma_2^2\mu_{13}^2 + \sigma_3^2\mu_{12}^2 +$$

$$- (\sigma_1^2\sigma_2^2\sigma_3^2 + 2\mu_{12}\mu_{23}\mu_{13})$$

The roots of a cubic equation can be solved by a change of variable $\lambda = \psi - 1/3\, K_1$ which yields the following reduced cubic or normal form in ψ .

$$\psi^3 + K_4\, \psi + K_5 = 0$$

where

$$K_4 = 1/3\,(3K_2 - K_1^2)$$
$$K_5 = 1/27\,(2K_1^3 - 9K_1K_2 + 27K_3)$$

In general, the reduced cubic has three roots which can be real or complex, positive or negative; however, since Γ_x is a real symmetrical matrix, its eigenvalues are real and, hence, only real roots of the reduced cubic need be considered. These roots are as follows:

$$\psi_i = 2\sqrt{-\frac{K_4}{3}}\cos(\sigma_i)$$

$$\psi_i = 2/3\sqrt{K_1^2 - 3K_2}\cos(\sigma_i)$$

for i = 1, 2, 3 where

$$\sigma_1 = 1/3\,\alpha$$
$$\sigma_2 = 1/3\,(\alpha + 2\pi)$$
$$\sigma_3 = 1/3\,(\alpha + 4\pi)$$

where

$$\alpha = \cos^{-1}\left[-\sqrt{\frac{K_5^2}{4}\cdot\left(-\frac{27}{K_4^3}\right)}\right]$$

if $K_5 > 0$ and

$$\alpha = \cos^{-1}\left[+\sqrt{\frac{K_5^2}{4}\cdot\left(-\frac{27}{K_4^3}\right)}\right]$$

if $K_5 < 0$. Thus, the eigenvalues of Γ_x become

$$\lambda_i = 2/3\sqrt{K_1^2 - 3K_2}\,\cos(\sigma_i)$$

for i = 1, 2, 3.

The eigenvectors can be determined by solving $\Gamma_x \underline{\phi}_i - \lambda_i \underline{\phi}_i = \underline{0}$ for each λ_i. Let each eigenvector be given by

$$\underline{\phi}_i = \begin{bmatrix} \phi_{i1} \\ \phi_{i2} \\ \phi_{i3} \end{bmatrix}$$

for i = 1, 2, 3. Substituting into $\Gamma_x \underline{\phi}_i - \lambda_i \underline{\phi}_i = \underline{0}$ it is found that

$$(\sigma_1^2 - \lambda_i)\phi_{i1} + \mu_{12}\phi_{i2} + \mu_{13}\phi_{i3} = 0$$

$$\mu_{12}\phi_{i1} + (\sigma_2^2 - \lambda_i)\phi_{i2} + \mu_{23}\phi_{i3} = 0$$

$$\mu_{13}\phi_{i1} + \mu_{23}\phi_{i2} + (\sigma_3^2 - \lambda_i)\phi_{i3} = 0$$

Any two of these equations can be used to solve ϕ_{i_1} and ϕ_{i_2} in terms of ϕ_{i_3}. Using the first two equations

$$(\sigma_1^2 - \lambda_i)\phi_{i1} + \mu_{12}\phi_{i2} = -\mu_{13}\phi_{i3}$$
$$\mu_{12}\phi_{i1} + (\sigma_2^2 - \lambda_i)\phi_{i2} = \mu_{23}\phi_{i3}$$

Solving for ϕ_{i_1} and ϕ_{i_2},

$$\phi_{i1} = \frac{(\lambda_i - \sigma_2^2)\mu_{13} + \mu_{12}\mu_{23}}{(\sigma_1^2 - \lambda_i)(\sigma_2^2 - \lambda_i) - \mu_{13}^2}\phi_{i3}$$

$$\phi_{i2} = \frac{(\lambda_i - \sigma_1^2)\mu_{23} + \mu_{12}\mu_{23}}{(\sigma_1^2 - \lambda_i)(\sigma_2^2 - \lambda_i) - \mu_{12}^2}\phi_{i3}$$

for i = 1, 2, 3. It is seen that the first two components of each eigenvector are proportional to the third component which is essentially arbitrary; i.e., the directions of the eigenvectors are determined with arbitrary magnitudes. If the eigenvectors are normalized to unity magnitude than an addition equation specifies the normalized eigenvectors, i.e.,

$$\phi_{i1}^2 + \phi_{i2}^2 + \phi_{i3}^2 = 1$$

The foregoing equations define a set of normalized eigenvectors which constitute the modal matrix M for Γ_x. It should be noted that for actual use it is not necessary to normalize the eigenvectors since the directions of the eigenvectors are of primary concern. However, the modal matrix M implies a set of normalized eigenvectors.

Using the above results, it is possible to define a statistically independent random vector $\underline{y}$ by a linear transformation; i.e., consider the transformation

$$\underline{y} = M^T(\underline{x} - \underline{m})$$

where $\underline{m} = E(\underline{x})$ and $M = [\underline{\phi}_1, \underline{\phi}_2, \underline{\phi}_3]$. The transformation is simply a translation and rotation of coordinates as depicted in Figure C-4. It follows that $E(\underline{y}) = \underline{0}$ and $\Gamma_y = M^T \Gamma_x M = \Lambda$ where

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$$

The pdf for $\underline{y}$ becomes

$$f(y) = \frac{1}{(2\pi)^{3/2}\sqrt{|\Lambda|}} \exp -\tfrac{1}{2} y^T \Lambda^{-1} y$$

$$= \frac{1}{(2\pi)^{3/2}\sqrt{\lambda_1 \lambda_2 \lambda_3}} \exp -\tfrac{1}{2}\left[\frac{y_1^2}{\lambda_1} + \frac{y_2^2}{\lambda_2} + \frac{y_3^2}{\lambda_3} \right]$$

$$f(y) = f(y_1)\, f(y_2)\, f(y_3)$$

where

$$f(y_1) = \frac{1}{\sqrt{2\pi\lambda_1}} e^{-\frac{y_1^2}{2\lambda_1}}$$

$$f(y_2) = \frac{1}{\sqrt{2\pi\lambda_3}} e^{-\frac{y_2^2}{2\lambda_2}}$$

$$f(y_3) = \frac{1}{\sqrt{2\pi\lambda_3}} e^{-\frac{y_3^2}{2\lambda_3}}$$

Thus, the components of $\underline{y}$, y_i, are statistically independent with zero mean values are variances λ_i for $i = 1, 2, 3$.

Consider the solid rectangular region, R, defined by

$$a_1 \le y_1 \le a_2$$
$$b_1 \le y_2 \le b_2$$
$$c_1 \le y_3 \le c_2$$

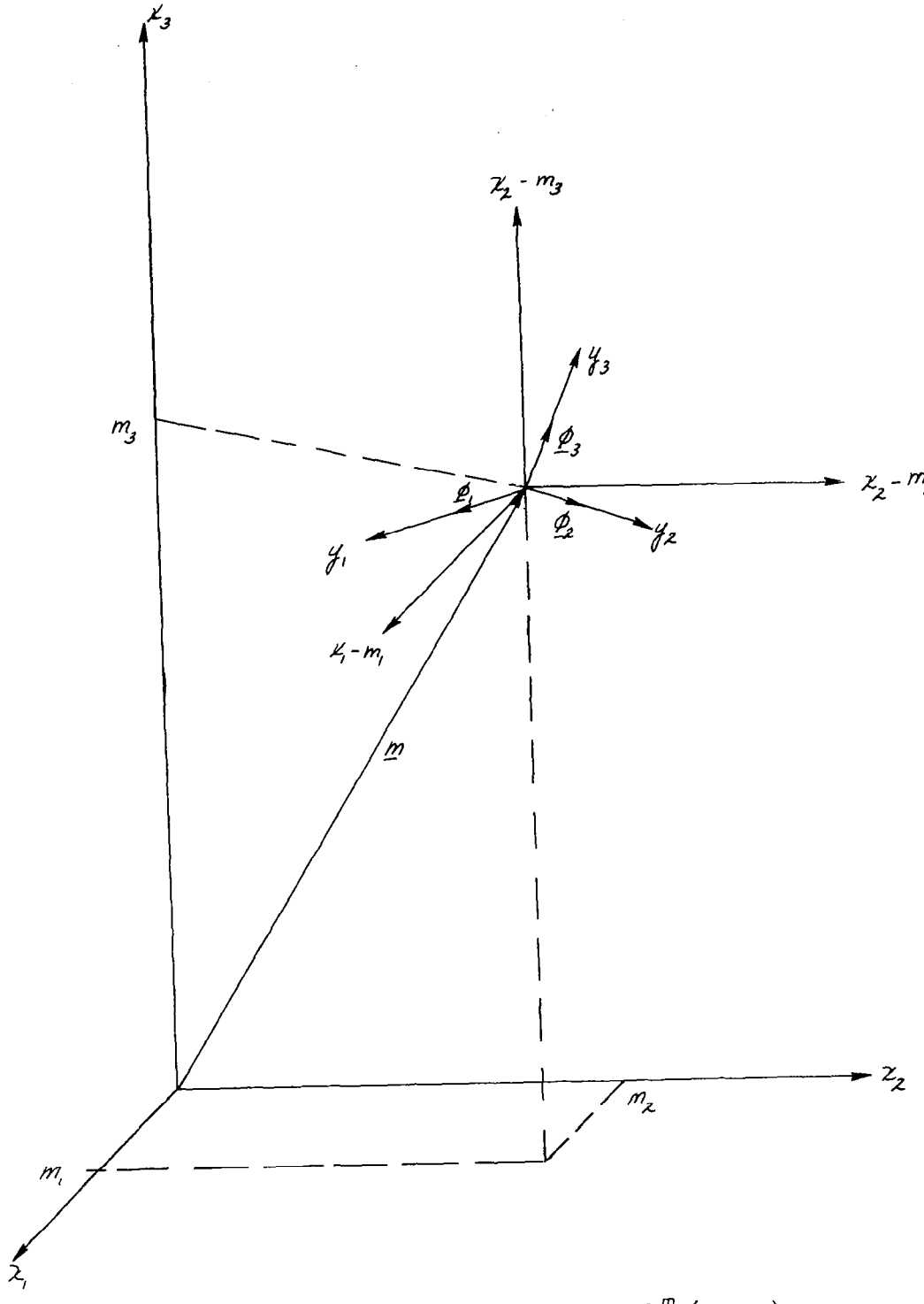

Figure C-4: Transformation of $\underline{y} = M^T (\underline{x} - \underline{m})$

The orientation of R is such that opposite sides are perpendicular to an eigenvector ϕ_i of Γ_x. The probability that $\underline{y}$ lies in R, P $[\underline{y} \in R]$, is given by

$$P[\underline{y} \in R] = \left[\int_{a_1}^{a_2} f(y_1)\,dy_1\right]\left[\int_{b_1}^{b_2} f(y_2)\,dy_2\right]\left[\int_{c_1}^{c_2} f(y_3)\,dy_3\right]$$

Now, if the limits are expressed in terms of $\sqrt{\lambda_1}$, $\sqrt{\lambda_2}$ and $\sqrt{\lambda_3}$ the integrals can be determined from a table for a Normal random variable.

Consider an ellipsoidal region, Δ , which is enclosed by the ellipsoid defined by $\underline{y}^T \Lambda^{-1} \underline{y} = \ell_1^2$, as shown in Figure C-5. It is noted that the vectors $\underline{r}_1$, $\underline{r}_2$ and $\underline{r}_3$ are the semi-principal axes of the ellipsoid. The lengths of these axes are found by setting y_2 and y_3, y_1 and y_3, and y_1 and y_2 equal to zero, respectively; to wit

$$\begin{aligned} a &= \sqrt{\lambda_1}\ \ell_1 \\ b &= \sqrt{\lambda_2}\ \ell_1 \\ c &= \sqrt{\lambda_3}\ \ell \end{aligned}$$

where a, b, and c are the lengths of $\underline{r}_1$, $\underline{r}_2$ and $\underline{r}_3$, respectively. It is noted that the volume of the ellipsoid, V, is given by

$$V = 4/3\,\pi\, a\, b\, c = 4/3\,\pi \sqrt{\lambda_1 \lambda_2 \lambda_3}\ \ell_1^3$$

The probability that $\underline{y}$ will lie in R, $P[\underline{y} \in R]$, is given by

$$P[\underline{y} \in \Delta] = \int_{\Delta} f(y)\,dy$$

$$= \frac{1}{(2\pi)^{3/2}\sqrt{\lambda_1 \lambda_2 \lambda_3}} \iiint_{\Delta} \exp -1/2\,(\underline{y}^T \Lambda^{-1} \underline{y})\,dy_1\,dy_2\,dy_3$$

$$P[\underline{y} \in \Delta] = \frac{1}{(2\pi)^{3/2}\sqrt{\lambda_1 \lambda_2 \lambda_3}} \iiint_{\Delta} e^{-1/2\,\ell^2}\,dy_1\,dy_2\,dy_3$$

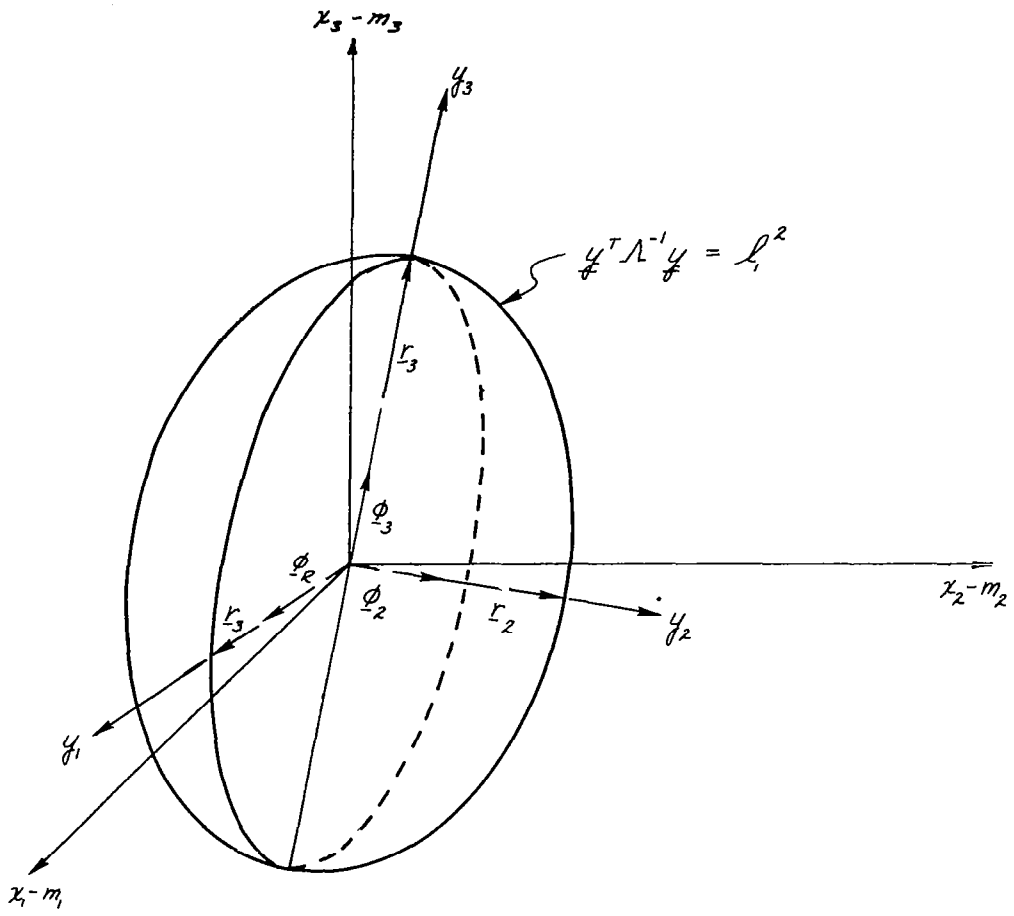

Figure C-5: Ellipsoidal Region $y^T \Lambda^{-1} y = \ell_1^2$

It is apparent that the probability density is constant along an ellipsoid defined by $y^T \Lambda^{-1} y = \ell^2$; thus, it is convenient to change the infinitesimal volume $dy_1\, dy_2\, dy_3$ to $dV = 4\pi \sqrt{\lambda_1 \lambda_2 \lambda_3}\, \ell^2 d\ell$, where dV is the infinitesimal volume between two ellipsoids defined by ℓ and $\ell + d\ell$. In this manner,

$$P[y \in \Delta] = \frac{1}{(2\pi)^{3/2}\sqrt{\lambda_1 \lambda_2 \lambda_3}} \int_0^{V_1} e^{-\frac{1}{2}\ell^2} dV$$

$$= \sqrt{\frac{2}{\pi}} \int_0^{\ell_1} e^{-\frac{1}{2}\ell^2} \ell^2 d\ell$$

$$= \sqrt{\frac{2}{\pi}} \int_0^{\ell_1} \ell \cdot e^{-\frac{1}{2}\ell^2} \ell\, d\ell$$

$$P[y \in \Delta] = \sqrt{\frac{2}{\pi}} \int_0^{\ell_1} u\, dv$$

where

$$u = \ell$$
$$dv = -e^{-\frac{1}{2}\ell^2} \ell\, d\ell$$

Integrating by parts,

$$P[y \in \Delta] = -\sqrt{\frac{2}{\pi}} \left[uv \Big|_0^{\ell_1} - \int_0^{\ell_1} v\, du \right]$$

$$= -\sqrt{\frac{2}{\pi}} \left[\ell_1 e^{-\frac{1}{2}\ell_1^2} - \int_0^{\ell_1} e^{-\frac{1}{2}\ell^2} d\ell \right]$$

$$= 2\left[\left(\frac{1}{\sqrt{2\pi}} \int_0^{\ell_1} e^{-\frac{1}{2}\ell^2} d\ell \right) - \ell_1 \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\ell_1^2} \right) \right]$$

$$P[y \in \Delta] = 2\left[\int_0^{\ell_1} f(\ell)\, d\ell - \ell_1 \cdot f(\ell_1) \right]$$

where

$$f(\ell) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\ell^2}$$

It is seen that $f(\ell)$ is the pdf for a Normal random variable. Thus, $P[y \in \Delta]$ can be evaluated with a table for a Normal random variable which tabulates both the pdf and the area enclosed about the mean value. Several values of $P[y \in \Delta]$ are given in Table C-3 below.

ℓ_2	$P(\Delta)$
1.00	0.1987
1.41	0.4251
1.50	0.4886
2.00	0.7385
2.50	0.9000

Table C-3: Probability of Occurrence $P[y \in \Delta] = P(\Delta)$

It should be noted that the "size" of the ellipsoid is given by the lengths of the semi-principals axes a, b and c, e.g., for $\ell = 1$ $a = \sqrt{\lambda_1}$ $b = \sqrt{\lambda_2}$ and $c = \sqrt{\lambda_3}$ with the corresponding probability of 0.1987. The orientation of the ellipsoid is given by the eigenvectors of Γ_x, i.e., the principal axes of the ellipsoid are co-linear with the eigenvectors given above.

APPENDIX D

SOME EXTREMAL PROPERTIES OF QUADRATIC FORMS

In the design of optimum navigation and guidance procedures criteria of quadratic forms are frequently used. The selection of various design parameters directly affects the resulting performance of the optimum procedures. Thus, in the selection of design parameters and in the analysis of performance the behavior of the extremal properties of quadratic forms is of considerable interest. In this appendix several properties of the extrema of ratio functions of quadratic forms are considered. The results are presented in two parts. Part I presents the basic results in terms of three theorems related to a particular ratio of two quadratic forms. Part II extends these basic results to more general cases.

PART 1. BASIC THEOREMS CONCERNING THE EXTREMAL PROPERTIES OF THE RATIO OF TWO QUADRATIC FORMS

The subject matter of this part is concerned with a real function, denoted by $f(\underline{X})$, of n independent variables which is the ratio of two quadratic forms. The function $f(\underline{X})$ is expressed in matrix form as follows.

$$f(\underline{x}) = \frac{\underline{x}^T A \underline{x}}{\underline{x}^T B \underline{x}} \tag{1.1}$$

where A and B are real symmetrical matrices of order n X n and $\underline{X}$ is an n dimensional column vector. The superscript T is used to denote the transpose of a matrix or a vector. The problem to be considered is that of determining the extremal properties of $f(\underline{X})$, i.e., what, if any, bounds exist on $f(\underline{X})$ as $\underline{X}$ varies throughout the range of all real n dimensional vectors excluding the null vector. The most general case where the matrices A and B are not related in any way is not considered in this part; rather, the less general case where A is equal to B raised to some integer power is considered. In this case, it is shown that if B is a positive definite matrix, then the extremal properties of f(X) are readily expressed in terms of the eigenvalues of B. Actually, the basic theorems are slightly more restrictive in that the degenerate case where the matrix B has a zero eignevalue is not considered, i.e., the eigenvalues of B are also positive definite. This is no severe restriction, it simply precludes the situation where the n space degenerates into a space of lower dimension, where the basic results apply.

The basic theorems are based upon the following three (3) lemmas.

Lemma 1

If $f(\underline{X})$ is given by Eq. (1.1), then the critical points of $f(\underline{X})$ occur for those $\underline{X}$ which satisfy the following vector equation for arbitrary A and B such that $\underline{X}^T B \underline{X}$ is positive definite.

$$(\underline{x}^T A \underline{x})\, B\underline{x} = (\underline{x}^T B \underline{x})\, A\underline{x} \tag{1.2}$$

Lemma 1 is readily established by taking the partial derivative of $f(\underline{X})$ with respect to $\underline{X}^T$ and setting the results equal to zero; to wit,

$$\begin{aligned}\frac{\partial}{\partial \underline{x}^T} f(x) &= \frac{\partial}{\partial \underline{x}^T}\left(\frac{\underline{x}^T A \underline{x}}{\underline{x}^T B \underline{x}}\right) \\ &= \frac{1}{\underline{x}^T B \underline{x}} \cdot \frac{\partial}{\partial \underline{x}^T}(\underline{x}^T A \underline{x}) + (\underline{x}^T A \underline{x}) \cdot \frac{\partial}{\partial \underline{x}^T}(\underline{x}^T B \underline{x})^{-1} \\ &= 2\frac{A\underline{x}}{\underline{x}^T B \underline{x}} - 2\frac{\underline{x}^T A \underline{x}}{(\underline{x}^T B \underline{x})^2}\, B\underline{x}\end{aligned} \tag{1.3}$$

Setting Eq. (1.3) equal to zero and multiplying through by $(\underline{X}^T B \underline{X})^2$, Lemma 1 is established.

It is noted that Lemma 1 is somewhat general in terms of A and B; i.e., no particular relationship between A and B is assumed. However, the quadratic form $\underline{X}^T B \underline{X}$ is assumed positive definite such that the partial derivitive of $f(\underline{X})$ exists. This is satisfied for a real symmetrical positive definite matrix B which is of concern in this discussion.

Lemma 2

If $A = B^N$ in Eq. (1.1), then any eigenvector of B multiplied by any arbitrary scalar constant is a critical point of $f(\underline{X})$. To establish this lemma, let $\underline{V}_i$ represent the eigenvector of the matrix B which corresponds to the ith eigenvalue of B, denoted by λ_i. Letting $A = B^N$ and $\underline{X} = \alpha \underline{V}_i$ in Eq. (1.2), the following equation results.

$$(\alpha \underline{V}_i^T B^N \alpha \underline{V}_i)\, B \alpha \underline{V}_i = (\alpha \underline{V}_i^T B \alpha \underline{V}_i)\, B^N \alpha \underline{V}_i \tag{1.4}$$

Since $\underline{V}_i$ is an eigenvector of B, $B\alpha \underline{V}_i = \alpha\lambda_i \underline{V}_i$ and $B^N \alpha \underline{V}_i = \alpha \lambda_i^N \underline{V}_i$, i.e., the vector transformation represented by the matrix B when applied to an eigenvector of B simply multiplies that eigenvector by the corresponding eigenvalue of B. Hence, Eq. (1.4) reduces to

$$\alpha^3 \lambda_i^{N+1} (\underline{V}_i^T \underline{V}_i) = \alpha^3 \lambda_i^{N+1} (\underline{V}_i^T \underline{V}_i) \tag{1.5}$$

Note that the eigenvectors of B form an orthonormal set of vectors, i.e.,

$$\underline{V}_i^T \underline{V}_j = \delta_{ij} \tag{1.6}$$

where $\delta_{ij} = 1$ for $i = j$ and $\delta_{ij} = 0$ for $i \neq j$. It follows that Eq. (1.5) could be written as

$$\alpha^3 \lambda_i^{N+1} = \alpha^3 \lambda_i^{N+1} \tag{1.7}$$

Of course, either Eq. (1.5) or (1.7) establishes Lemma 2.

Lemma 3

If $A = B^N$ in Eq. (1.1) and if the eigenvalues of B are distinct and positive definite, then the critical points of $f(\underline{X})$ occur only for those $\underline{X}$ which are the eigenvectors of B, multiplied by an arbitrary scalar. The difference between Lemma 3 and Lemma 2 should be pointed out. In effect, Lemma 2 establishes the "if" portion of an "if, and only if" condition. Lemma 3 establishes the "only if" portion of this conditionality. To establish Lemma 3, the eigenvectors of the matrix B are used as a basis for the space; any any arbitrary vector $\underline{X}$ is written in terms of this basis, i.e.,

$$\underline{x} = \alpha_1 \underline{V}_1 + \alpha_2 \underline{V}_2 + \ldots + \alpha_n \underline{V}_n$$

$$\underline{x} = \sum_{j=1}^{n} \alpha_j \underline{V}_j \tag{1.8}$$

The transpose of $\underline{X}$ is given by

$$\underline{x}^T = \sum_{j=1}^{n} \alpha_j \underline{V}_j^T \tag{1.9}$$

By using the eigenvectors of the matrix B as a basis, the product $B^N\underline{X}$ is expressed simply as

$$B\underline{x} = \sum_{j=1}^{n} \alpha_j \lambda_j^N \underline{V}_j \tag{1.10}$$

The product $B\underline{X}$ is the vector obtained by applying the transformation represented by the matrix B to the vector $\underline{X}$. The product $B^N\underline{X}$ is the vector obtained by N such transformations applied in succession; the resulting vector is expressed simply as

$$B^N\underline{x} = \sum_{j=1}^{n} \alpha_j \lambda_j^N \underline{V}_j \tag{1.11}$$

By pre-multiplying Eqs. (1.10) and (1.11) by Eq. (1.9) and recalling that the eigenvectors of B form an orthonormal set of vectors, the following expressions are obtained for the quadratic forms $\underline{X}^T B \underline{X}$ and $\underline{X}^T B^N \underline{X}$.

$$\underline{x}^T B \underline{x} = \sum_{i=1}^{n} \alpha_i^2 \lambda_i \tag{1.12}$$

$$\underline{x}^T B^N \underline{x} = \sum_{i=1}^{n} \alpha_i^2 \lambda_i^N \tag{1.13}$$

Now, by substituting Eqs. (1.10) through (1.13) into Eq. (1.2) for $A = B^N$, the following expression is obtained.

$$\left[\sum_{i=1}^{n} \alpha_i^2 \lambda_i^N\right] \sum_{j=1}^{n} \alpha_j \lambda_j \underline{V}_j = \left[\sum_{i=1}^{n} \alpha_i^2 \lambda_i\right] \sum_{j=1}^{n} \alpha_j \lambda_j^N \underline{V}_j \tag{1.14}$$

By transposing the interchanging the order of summation, Eq. (1.14) becomes

$$\sum_{j=1}^{n} \left[\alpha_j \lambda_j^N \sum_{i=1}^{n} \alpha_i^2 \lambda_i - \alpha_j \lambda_j \sum_{i=1}^{n} \alpha_i \lambda_i^N\right] \underline{V}_j = 0 \tag{1.15}$$

The values of αj which satisfy Eq. (1.15) determine the vectors for which critical points of $f(\underline{X})$ occur. Since the eigenvectors of the matrix B are linearly independent, each multiplier of $\underline{V}j$ in Eq. (1.15) must be zero in order to satisfy Eq. (1.15). Thus, determining the values of αj which yield critical points of $f(\underline{X})$ is equivalent to determining values of αj for which the following equation is satisfied for all j.

$$\alpha_j \lambda_j \left[\lambda_j^{N-1} \sum_{i=1}^{n} \alpha_i^2 \lambda_i - \sum_{i=1}^{n} \alpha_i \lambda_i^N\right] = 0 \tag{1.16}$$

Writing the terms within the parentheses as one summation, Eq. (1.16) becomes

$$\alpha_j \lambda_j \left[\sum_{i=1}^{n} \alpha_i^2 \lambda_i \left(\lambda_j^{N-1} - \lambda_i^{N-1} \right) \right] = 0 \tag{1.17}$$

In order to establish Lemma 3, it is necessary to show that if Eq. (1.17) is satisfied for all j; then only one αj can be selected arbitrarily non-zero. To this end, let the eigenvalues of B be ordered such that

$$\lambda_1 > \lambda_2 > \lambda_3 > \ldots > \lambda_{n-1} > \lambda_n > 0 \tag{1.18}$$

Next, for convenience, let N = 2 and write Eq. (1.17) as the following system of equations. Note that nothing is lost in generality since the following argument applies for all N greater than 2.

$$\alpha_1 \lambda_1 \left[0 + \alpha_2^2 \lambda_2 (\lambda_1 - \lambda_2) + \alpha_3^2 \lambda_3 (\lambda_2 - \lambda_3) + \ldots + \alpha_{n-1}^2 \lambda_{n-1} (\lambda_1 - \lambda_{n-1}) + \alpha_n^2 \lambda_n (\lambda_1 - \lambda_n) \right] = 0$$

$$\alpha_2 \lambda_2 \left[\alpha_1^2 \lambda_1 (\lambda_2 - \lambda_1) + 0 + \alpha_3^2 \lambda_3 (\lambda_2 - \lambda_3) + \ldots + \alpha_{n-1}^2 \lambda_{n-1} (\lambda_2 - \lambda_{n-1}) + \alpha_n^2 \lambda_n (\lambda_2 - \lambda_n) \right] = 0$$

$$\alpha_3 \lambda_3 \left[\alpha_1^2 \lambda_1 (\lambda_3 - \lambda_1) + \alpha_2^2 \lambda_2 (\lambda_3 - \lambda_2) + 0 + \ldots + \alpha_{n-1}^2 \lambda_{n-1} (\lambda_3 - \lambda_{n-1}) + \alpha_n^2 \lambda_n (\lambda_3 - \lambda_n) \right] = 0$$

$$\vdots$$

$$\alpha_j \lambda_j \left[\alpha_1^2 \lambda_1 (\lambda_j - \lambda_1) + \alpha_2^2 \lambda_2 (\lambda_j - \lambda_2) + \ldots + 0 + \ldots + \alpha_{n-1}^2 \lambda_{n-1} (\lambda_j - \lambda_{n-1}) + \alpha_n^2 \lambda_n (\lambda_j - \lambda_n) \right] = 0$$

$$\vdots$$

$$\alpha_{n-1} \lambda_{n-1} \left[\alpha_1^2 \lambda_1 (\lambda_{n-1} - \lambda_1) + \alpha_2^2 \lambda_2 (\lambda_{n-1} - \lambda_2) + \alpha_3^2 \lambda_3 (\lambda_{n-1} - \lambda_3) + \ldots + 0 + \alpha_n^2 \lambda_n (\lambda_{n-1} - \lambda_n) \right] = 0$$

$$\vdots$$

$$\alpha_n \lambda_n \left[\alpha_1^2 \lambda_1 (\lambda_n - \lambda_1) + \alpha_2^2 \lambda_2 (\lambda_n - \lambda_2) + \alpha_3^2 \lambda_3 (\lambda_n - \lambda_3) + \ldots + \alpha_{n-1}^2 \lambda_{n-1} (\lambda_n - \lambda_{n-1}) + 0 \right] = 0$$

[System of Equations (1.19)]

An examination of the above system of equations shows the following to be true. If α_1 is non-zero, then all other αj's ($j = 2, 3, \cdots n$) must be zero in order to satisfy this first equation, which also satisfies the entire system of n equations. This is true since each term within the bracket of the first equation is positive and this sum must be zero if α_1 is non-zero. Therefore, all other αj's must be zero if α_1 is non-zero. Thus, if α_2 is non-zero, then α_1 must be zero; moreover, if α_1 is zero *and* $\propto_2$ is non-zero, then all other $\propto j$'s must be zero to satisfy the second equation of the system which also satisfies the entire system of n equations. This is true since if α_1 is zero, the remaining non-zero terms within the bracket of the second equations are all positive and their sum must be zero. Therefore, if α_2 is non-zero, all other αj's must be zero. The same argument can be applied generally. That is, let any combination of the $\propto j$'s be assumed to be non-zero. An examination of the equation in the above system which corresponds to the first non-zero αj's shows that the remaining αj's cannot be non-zero if the entire system of equations is satisfied. In this manner Lemma 3 is established.

It is noted that the argument is valid for N greater than 2 since the terms which are positive remain positive for $N > 2$. The case of $N = 1$ is of no interest since $f(\underline{X})$ is a constant. If N is an integer such that $N \leq 0$, the argument remains valid with a simple change of sign in all of the equations. Thus, Lemma 3 is true for all integer N, excluding 0, which is of no concern.

Theorem 1

If B is a real symmetrical positive definite matrix and if $f(\underline{X})$ is defined by

$$f(\underline{x}) = \frac{\underline{x}^T B^N \underline{x}}{\underline{x}^T B \underline{x}}$$

then a critical point of $f(\underline{X})$ occurs if, and only if, $\underline{X}$ is equal to an eigenvector of the matrix B multiplied by an arbitrary scalar constant. The proof of this theorem follows directly from Lemmas 1, 2, and 3 given above.

Theorem 2

If the eigenvalues of the matrix B are distinct and positive definite, then

$$\lambda_1^{N-1} \geq \frac{\underline{x}^T B^N \underline{x}}{\underline{x}^T B \underline{x}} \geq \lambda_n^{N-1} \qquad N > 1 \tag{1.20A}$$

and

$$\lambda_1^{N-1} \leq \frac{\underline{x}^T B^N \underline{x}}{\underline{x}^T B \underline{x}} \leq \lambda_n^{N-1} \qquad N \leq 0 \tag{1.20B}$$

where λ_1 and λ_n denote the largest and smallest eigenvalues of B, respectively. Moreover, the maximum and minimum values occur for $\underline{X}$ equal to the eigenvectors of B, which correspond to the largest and smallest eigenvalues of B, multiplied by an arbitrary scalar constant.

Generally, in order to determine which critical points of a function define the extremal points the second partial derivatives are examined. However, in the present case it is easier to simply evaluate $f(\underline{X})$ for $A = B^N$ at all of its critical points. By Theorem 1 it is known that the critical points of $f(\underline{X})$ occur only at the eigenvectors of B; thus, it is a simple matter to evaluate $f(\underline{X})$ at all of its critical points in general. To wit, let $\underline{X}$ be an eigenvector of B multiplied by an arbitrary scalar constant, i.e.,

$$\underline{x} = \alpha_j \underline{V}_j \tag{1.21}$$

Substituting Eq. (1.20) into Eq. (1.1) for $A = B^N$ it is found that

$$f(\underline{x} = \alpha_j \underline{V}_j) = \lambda_j^{N-1} \tag{1.22}$$

where λ_j is the eigenvalue of B which corresponds to the eigenvector $\underline{V}j$. Note that Eq. (1.21) is actually Eq. (1.13) divided by Eq. (1.12) for the special case of αi equal to zero for $i \neq j$. Eq. (1.21) is all that is necessary to establish Theorem 2. From Eq. (1.20) it is seen that no matter what $\underline{X}$ is taken as in the n space, $f(\underline{X})$ is constrained to the bounds of Eq. (1.20).

It is seen that when $N>1$, the maximum and minimum values of $f(\underline{X})$ occur for $\underline{X}$ equal to $\alpha_1 \underline{V}_1$ and $\alpha_n \underline{V}_n$, respectively, where $\underline{V}_1$ and $\underline{V}_n$ are the eigenvectors of B which correspond to the eigenvalues λ_1 and λ_n, respectively. Of course, when $N \leq 0$, the opposite extremal value occurs. This establishes Theorem 2.

Theorem 2 has been proved for the case in which the eigenvalues of the matrix B are positive definite and distinct. For the case where the eigenvalues of B are positive definite but not necessarily distrinct, the following theorem is established.

Theorem 3

If the eigenvalues of the matrix B are positive definite but not distinct, i.e., multiplicities of various orders exist in the eigenvalues of B, the extremal properties of $f(\underline{X})$ given in Theorem 2 are unaltered. That is, the extremal properties of $f(\underline{X})$ are not affected by multiple eigenvalues of B.

Assume that the eigenvalues of B have a single multiplicity of order k in an eigenvalue λj, i.e., in the array of n eigenvalues of B there are k eigenvalues equal to λj. For this case, the system of equations Eq. (1.19) becomes

$$\alpha_1 \lambda_1 \left[0 + \alpha_2^2 \lambda_2 (\lambda_1 - \lambda_2) + \alpha_3^2 \lambda_3 (\lambda_1 - \lambda_3) + \ldots + \alpha_n^2 \lambda_n (\lambda_1 - \lambda_n) \right] = 0$$

$$\alpha_2 \lambda_2 \left[\alpha_1^2 \lambda_1 (\lambda_2 - \lambda_1) + 0 + \alpha_3^2 \lambda_3 (\lambda_2 - \lambda_3) + \ldots + \alpha_n^2 \lambda_n (\lambda_2 - \lambda_n) \right] = 0$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\alpha_j \lambda_j \left[\alpha_1^2 \lambda_1 (\lambda_j - \lambda_1) + \ldots + \alpha_{j-1}^2 \lambda_{j-1} (\lambda_j - \lambda_{j-1}) + 0 + 0 + \ldots \right.$$
$$\left. \ldots + 0 + \alpha_{j+k}^2 \lambda_{j+k} (\lambda_j - \lambda_{j+k}) + \alpha_n^2 \lambda_n (\lambda_j - \lambda_n) \right] = 0$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\alpha_{j+k-1} \lambda_{j+k-1} \left[\alpha_1^2 \lambda_1 (\lambda_{j+k-1} - \lambda_1) \ldots + \alpha_{j-1}^2 \lambda_{j-1} (\lambda_j - \lambda_{j-1}) + 0 + 0 + \ldots \right.$$
$$\left. + 0 + \alpha_{j+k}^2 \lambda_{j+k} (\lambda_j - \lambda_{j+k}) + \ldots \alpha_n^2 \lambda_n (\lambda_j - \lambda_n) \right] = 0$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\alpha_{j+k} \lambda_{j+k} \left[\alpha_1^2 \lambda_1 (\lambda_{j+k} \lambda_1) + \ldots + \alpha_{j+k-1}^2 \lambda_{j+k-1} (\lambda_{j+k} - \lambda_{j+k-1}) + 0 + \ldots \right.$$
$$\left. + \alpha_{j+k-1}^2 \lambda_{j+k-1} (\lambda_{j+k} - \lambda_{j+k-1}) + \ldots + \alpha_n^2 \lambda_n (\lambda_{j+k} - \lambda_n) \right] = 0$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\alpha_n \lambda_n \left[\alpha_1^2 \lambda_1 (\lambda_n - \lambda_1) + \alpha_2^2 \lambda_2 (\lambda_n - \lambda_2) + \alpha_3^2 \lambda_3 (\lambda_n - \lambda_3) + \ldots + 0 \right] = 0$$

[System of Equations (1.24)]

Comparing the two systems of Eq. (1.19) and (1.24) shows that if all multiplicities of the eigenvalues of B are of order one, i.e., the eigenvalues are distinct, then a single zero occurs in each of the n equations which determines the possible αj's; whereas, if a single multiplicity of order k occurs, then those equations which correspond to the multiple eigenvalues each have a total of k zeros. The effect of having additional zeros in these equations is that more than one non-zero αj is possible in satisfying the n equations. This, in turn, means that for multiple eigenvalues of B the critical points of $f(\underline{X})$ occur for certain $\underline{X}$ other than the eigenvectors of B. More specifically, if the eigenvalues of B have a single multiplicity of order k, then critical points of $f(\underline{X})$ occur for any linear combination of the k eigenvectors of B which correspond to the k multiple values of the eigenvalues of B. This is verified by an examination of the system of equations (1.24). If α_1 through α_{j-1} are selected as zero, then αj through $\alpha j + k - 1$ can be selected arbitrarily non-zero. However, the remaining αj's, $\alpha j + k$ through α_n must be zero.

From the foregoing it is seen that the effect of multiple eigenvalues of B is that critical points of $f(\underline{X})$ exist for $\underline{X}$ other than the eigenvectors of B. However, the additional critical points of $f(\underline{X})$ occur only for linear combinations of those eigenvectors which correspond to the multiple eigenvalues of B. This is important because for this reason the values of $f(\underline{X})$ at these additional critical points are all the same. To show that this is true, let $\underline{V}_j$ through $\underline{V}_{j+k-1}$ represent the eigenvectors of B which correspond to the multiple eigenvalues λj through $\lambda j + k - 1$ of B. Let $\underline{X}$ be any arbitrary linear combination of these eigenvectors, i.e.,

$$\underline{x} = \sum_{i=j}^{j+k-1} \alpha_i \underline{V}_i \tag{1.25}$$

Now, the products $B\underline{X}$ and $B^N\underline{X}$ become

$$B\underline{x} = \sum_{i=j}^{j+k-1} \alpha_i \lambda_i \underline{V}_i \tag{1.26}$$

$$B^N\underline{x} = \sum_{i=j}^{j+k-1} \alpha_i \lambda_i^N \underline{V}_i \tag{1.27}$$

However, since the λi's are equal for $i = j, j + 1, \cdots, j + k - 1$, Eqs. (1.26) and (1.27) become

$$B\underline{x} = \lambda_j \sum_{i=j}^{j+k-1} \alpha_i \underline{V}_i \tag{1.28}$$

$$B^N \underline{z} = \lambda_j^N \sum_{i=j}^{j+k-1} \alpha_i \underline{V}_i \tag{1.29}$$

Hence, the two quadratic forms $\underline{X}^T B \underline{X}$ and $\underline{X}^T B^N \underline{X}$ become

$$\underline{z}^T B \underline{z} = \lambda_j \sum_{i=j}^{j+k-1} \alpha_i^2 \tag{1.30}$$

$$\underline{z}^T B^N \underline{z} = \lambda_j^N \sum_{i=j}^{j+k-1} \alpha_i^2 \tag{1.31}$$

Therefore, $f(\underline{X})$, which is the ratio of Eq. (1.31) to Eq. (1.30), is equal to λ_j^{N-1} at all of the critical points defined by Eq. (1.25). The value of $f(\underline{X})$ at these critical points is independent of the order of the multiplicity of λ_j; and, furthermore, the value of $f(\underline{X})$ at these critical points is the same as its value for a multiplicity of order one. Thus, the effect of a multiplicity in the eigenvalues of B is that additional critical points of $f(\underline{X})$ exist; but the values of $f(\underline{X})$ at these points are all equal to the value of $f(\underline{X})$ for a multiplicity of order one. Therefore, the extremal values of $f(\underline{X})$ are unaffected by a multiplicity in the eigenvalues of B.

Strictly speaking, the foregoing argument has been given for the case of a single multiplicity. In the interest of generality, the argument should be extended to the case of m multiplicities each of order k_m. By extending the above argument it is found that the results are the same for m different multiplicities in the eigenvalues of B. The important factor in this extension is that no more than k_m αj's can be selected arbitrarily non-zero which must correspond to eigenvalues in the mth multiplicity. To verify this, note that for m multiplicities the set of equations (1.19) has m sets of equations which have more than one zero. In fact, the set of equations which corresponds to the mth multiplicity has k_m zeros. However, if k_m αj's are selected arbitrarily non-zero on the basis of the mth set of equations, the remaining αj's must be zero in order to satisfy the entire system. Therefore, no two sets of αj's corresponding to different multiplicities can be taken as non-zero and satisfy the entire n equations of system (1.19).

The foregoing argument shows that the extremal properties of $f(\underline{X})$ are unaffected by multiplicities in the eigenvalues of B. The remarkable result is that in spite of the fact that a multiplicity in the eigenvalues of B greatly increases the number of critical points of $f(\underline{X})$, the values of $f(\underline{X})$

at these critical points are all equal to the value of $f(\underline{X})$ for a multiplicity of order one. This is the essence of the foundation of Theorem 3.

Summary of Part 1

The essense of the above three theorems can be summarized in more precise mathematical terminology as follows. If B is an n x n positive definite matrix whose n eigenvalues have m distinct values, $m \leq n$, yielding a total of m multiplicities of order k_m, respectively, including all multiplicities of order one, and if $f(\underline{X})$ is given by

$$f(\underline{x}) = \frac{\underline{x}^T B^N \underline{x}}{\underline{x}^T B \underline{x}} \tag{1.32}$$

then the vectors $\underline{X}$ for which critical points of $f(\underline{X})$ occur can be collected into m different sets of vectors, denoted by $Sm(\underline{X})$. Each set $Sm(\underline{X})$ is a linear manifold of dimension k_m and is spanned by the k_m eigenvalues of B which correspond to the k_m eigenvalues of B contained in the mth multiplicity of the eigenvalues of B. The sum of the dimensions of these m linear manifolds is equal to n; the dimension of the space. These m manifolds $Sm(\underline{X})$ are invariant with respect to the linear transformation represented by the matrix B. Also, the values of $f(\underline{X})$ for $\underline{X}$ in the manifold $Sm(\underline{X})$ is independent of the dimension of the manifold and is equal to λ_j^{N-1} for all $\underline{X}$ in $Sm(\underline{X})$, where λj is the value of those eigenvalues in the mth multiplicity. Thus, the values of $f(\underline{X})$ for all $\underline{X}$ in the n space lie in the range

$$\begin{aligned} \lambda_1^{N-1} \geq f(\underline{x}) \geq \lambda_n^{N-1} \qquad & N > 1 \\ \lambda_1^{N-1} \leq f(\underline{x}) \leq \lambda_n^{N-1} \qquad & N \leq 0 \end{aligned} \tag{1.33}$$

The maximum value of $f(\underline{X})$ occurs for all vectors in the manifold $S_1(\underline{X})$, where $S_1(\underline{X})$ is the manifold which corresponds to the largest eigenvalue of B. Likewise, the minimum value of $f(\underline{X})$ occurs for all vectors in the manifold $Sn(\underline{X})$, where $Sn(\underline{X})$ is the manifold which corresponds to the smallest eigenvalue of B.

The meaning of the foregoing is that the variation of $f(X)$ is determined solely by N and the largest and smallest values of the eigenvalues of B. If the difference in the smallest and largest eigenvalues of B is small, then the variation of $f(\underline{X})$ is likewise relatively small; and conversely, if the difference in these extremal eigenvalues is large, then the variation in $f(\underline{X})$ can be large. Of course, "relatively" small or large in terms of the variation of $f(\underline{X})$ is a function of N; i.e., the variation of $f(\underline{X})$ can exceed the difference in the largest and smallest eigenvalues of B. One of the most important results of the foregoing is that the extreme values of

$f(\underline{X})$ are not uniquely determined with respect to the vectors $\underline{X}$. That is, in the case of distinct eigenvalues of B the extremal values of $f(\underline{X})$ occur for those $\underline{X}$ equal to the two eigenvectors of B which correspond to the extreme eigenvalues of B, multiplied by an arbitrary constant. In the case where multiplicities occur in the extreme eigenvalues of B, then the extreme values of $f(\underline{X})$ occur for all vectors $\underline{X}$ in the manifolds which correspond to these extreme eigenvalues. As an interesting example of this situation, consider the case where all of the eigenvalues of B are equal, then $f(\underline{X})$ is equal to a constant independent of $\underline{X}$. This constant is λ^{N-1}, where λ is the value of the n eigenvalues of B.

PART 2. EXTENSIONS OF THE BASIC RESULTS

The basic results of Part 1 are extended in the following theorems. An immediate extension concerns the reciprocal of $f(\underline{X})$ as defined previously. Other extensions of important concern include the situation where the matrices A and B are not related by a particular function as considered previously. In extending the previous results, it is convenient to consider the basic results of Part 1 in the following matrix notation.

Let M be the modal matrix for the real symmetrical positive definite matrix B, i.e., M is the matrix of orthonormal eigenvectors of B. The modal matrix M has the following properties.

$$\begin{aligned} M^T M &= I \\ M^{-1} &= M^T \\ M^T B M &= \Lambda \end{aligned} \tag{2.1}$$

where Λ is the diagonal matrix of eigenvalues of B. Let any vector $\underline{X}$ be expressed in terms of the orthonormal vectors of M; i.e., the eigenvectors of B are used as a basis for the n dimensional space. In this manner any vector $\underline{X}$ can be written as

$$\underline{x} = M \underline{\alpha}$$

where $\underline{\alpha}$ are the components of $\underline{X}$ in the basic vector M.

Theorem 4

If B is a real symmetrical positive definite matrix, then

$$\lambda_1^{1-N} \leq \frac{\underline{x}^T B \underline{x}}{\underline{x}^T B^N \underline{x}} \leq \lambda_N^{1-N} \qquad N > 1$$

$$\lambda_1^{1-N} \geq \frac{\underline{x}^T B \underline{x}}{\underline{x}^T B^N \underline{x}} \geq \lambda_n^{1-N} \quad N \leq 0$$

It is noted that Lemma 1 applies to $g(\underline{X}) = f^{-1}(\underline{X})$ for $A = B^N$ where $\underline{X}^T B \underline{X}$ is positive definite. Thus if

$$g(\underline{x}) = \frac{\underline{x}^T B \underline{x}}{\underline{x}^T B^N \underline{x}}$$

then the critical points of $g(\underline{X})$ are determined by

$$(\underline{x}^T B^N \underline{x}) B \underline{x} = (\underline{x}^T B \underline{x}) B \underline{x}$$

Therefore, Lemmas 1 and 2 apply to $g(\underline{X})$, and the critical points of $g(\underline{X})$ are the same as those of $f(\underline{X})$ for $A = B^N$. The arguments of Theorem 2 and 3 also apply to $g(\underline{X})$ and Theorem 4 follows immediately.

Theorem 5

Let B be a positive definite matrix and let $f(\underline{X})$ be defined as follows:

$$f(\underline{x}) = \frac{\underline{x}^T A \underline{x}}{\underline{x}^T B \underline{x}}$$

If the eigenvalues of $C = B^{-1}A$ are distinct, then the critical points of $f(\underline{X})$ occur for $\underline{X}$ equal to the eigenvectors of C multiplied by an arbitrary constant. From Lemma 1, the critical points of $f(\underline{X})$ are determined by

$$(\underline{x}^T A \underline{x}) B \underline{x} = (\underline{x}^T B \underline{x}) A \underline{x} \tag{2.2}$$

Letting $\underline{X} = M\underline{\alpha}$, $(\underline{X}^T A \underline{X})B\underline{X} = (\underline{\alpha}^T M^T A M \underline{\alpha})BM\underline{\alpha}$ and $(\underline{X}^T B \underline{X})A\underline{X} = (\underline{\alpha}^T M^T B M \underline{\alpha})AM\underline{\alpha} = (\underline{\alpha}^T \Lambda \underline{\alpha})AM\underline{\alpha}$; hence, Eq. (2.2) becomes

$$(\underline{\alpha}^T \Lambda \underline{\alpha}) AM\underline{\alpha} = (\underline{\alpha}^T M^T A M \underline{\alpha}) BM\underline{\alpha} \tag{2.3}$$

Therefore, the critical points are determined by

$$(B^{-1}A)M\underline{\alpha} = \frac{(\underline{\alpha}^T M^T A M \underline{\alpha})}{(\underline{\alpha}^T \Lambda \underline{\alpha})} M\underline{\alpha} \tag{2.4}$$

Alternatively,

$$(B^{-1}A)M\underline{\alpha} = H(\underline{\alpha}, A, \Lambda)\, M\underline{\alpha} \tag{2.5}$$

where $H(\underline{\alpha}, A, \Lambda)$ is a scalar function of $\underline{\alpha}$, A and Λ given by

$$H(\underline{\alpha}, A, \Lambda) = \frac{(\underline{\alpha}^T M^T A M \underline{\alpha})}{(\underline{\alpha}^T \Lambda\, \underline{\alpha})} \tag{2.6}$$

It is seen that Eq. (2.5) is the characteristic equation which determines the eigenvalues and eigenvectors for the matrix $C = B^{-1} A$, i.e., the equation

$$C\underline{\phi} = \lambda \underline{\phi} \tag{2.7}$$

determines the eigenvalues λ and eigenvectors $\underline{\phi}$ of the matrix C. Eq. (2.5) is the same as Eq. (2.7) where $C = B^{-1} A$, $M\underline{\alpha} = \underline{X} = \underline{\phi}$ and $H(\underline{\alpha}, A, \Lambda) = \lambda$. It is easily shown that if $M\underline{\alpha} = \underline{\phi}$, then $H(\underline{\alpha}, A, \Lambda) = \lambda$; to wit:

$$H(\underline{\alpha}, A, \Lambda) = \frac{(\underline{\alpha}^T M^T A M \underline{\alpha})}{(\underline{\alpha}^T \Lambda\, \underline{\alpha})}$$

$$= \frac{\underline{\alpha}^T M^T A \underline{\phi}}{\underline{\alpha}^T \Lambda\, \underline{\alpha}}$$

$$= \frac{\underline{\alpha}^T M^T (\lambda B \phi)}{\underline{\alpha}^T \Lambda\, \underline{\alpha}}$$

$$= \lambda \frac{\alpha^T M^T B M \underline{\alpha}}{\underline{\alpha}^T \Lambda\, \underline{\alpha}}$$

$$= \lambda \frac{(\underline{\alpha}^T \Lambda\, \underline{\alpha})}{(\underline{\alpha}^T \Lambda\, \underline{\alpha})}$$

$$H(\underline{\alpha}, A, \Lambda) = \lambda$$

The foregoing can be verified directly from Eq. (2.2); i.e., if $\underline{X} = \alpha\underline{\phi}$ where $A\,\underline{\phi} = \lambda B\,\underline{\phi}$, then Eq. (2.2) becomes

$$\alpha^3 \lambda (\underline{\phi}^T B \underline{\phi})\, B\underline{\phi} = \alpha^3 \lambda (\underline{\phi}^T B \underline{\phi})\, B\underline{\phi}$$

Theorem 6

If B is a positive definite matrix and if the eigenvalues of $C = B^{-1}A$ are distinct, then

$$\lambda_1 \geq \frac{\underline{x}^T A \underline{x}}{\underline{x}^T B \underline{x}} \geq \lambda_n \tag{2.8}$$

where λ_1 and λ_n are the maximum and minimum eigenvalues of $B^{-1}A$, which are given by

$$|A - \lambda B| = 0 \tag{2.9}$$

This follows directly from evaluating $f(\underline{X})$ at its critical points which are $\underline{X} = \alpha \underline{\phi}$ where

$$A\underline{\phi} = \lambda B \underline{\phi} \tag{2.10}$$

Hence, for $\underline{X} = \alpha\underline{\phi}$, $f(\underline{X})$ becomes

$$f(\underline{x}) = \frac{\alpha^2 \underline{\phi}^T A \underline{\phi}}{\alpha^2 \underline{\phi}^T B \underline{\phi}}$$

$$= \frac{\lambda \underline{\phi}^T B \underline{\phi}}{\underline{\phi}^T B \underline{\phi}}$$

$$f(\underline{x}) = \lambda$$

Therefore, at each critical point $\underline{X} = \alpha\underline{\phi}$, $f(\underline{X})$ is simply equal to the eigenvalue corresponding to the eigenvector $\underline{\phi}$ which defines the critical point. The bounds of Eq. (2.8) follow from selecting the critical points for the maximum and minimum eigenvalues.

Theorem 7

If B is a positive definite matrix, then

$$\lambda_1 \geq \frac{\underline{x}^T A \underline{x}}{\underline{x}^T B \underline{x}} \geq \lambda_n \tag{2.11}$$

where λ_1 and λ_n are the maximum and minimum eigenvalues of $C = B^{-1} A$. The bounds of Eq. (2.11) are seen to be the same as those of Eq. (2.8) in Theorem 6. However, it is noted that the present theorem does not require distinct eigenvalues of $C = B^{-1} A$.

The present theorem can be established by an argument similar to that of Theorem 3. It is easily seen that if several eigenvectors of $C = B^{-1} A$ exist for a particular eigenvalue, then there exists a manifold of critical points for $f(\underline{X})$. That is, let $\underline{\phi}_j$ denote a set of k eigenvectors which corresponds to the eigenvalue λ_k. Let $\underline{X}$ be an arbitrary linear combination of the set $\underline{\phi}_j$; i.e.,

$$\underline{x} = \sum_{j=1}^{k} \alpha_j \underline{\phi}_j$$

Since $A \underline{\phi}_j = \lambda_k B \underline{\phi}_j$, it follows that

$$A\underline{x} = A \sum_{j=1}^{k} \alpha_j \underline{\phi}_j$$

$$= \sum_{j=1}^{k} \alpha_j A \underline{\phi}_j$$

$$= \lambda_k \sum_{j=1}^{k} \alpha_j B \underline{\phi}_j$$

$$A\underline{x} = \lambda_k B\underline{x}$$

Thus, $\underline{X}$ as defined satisfies the equation for the critical points of $f(\underline{X})$ since $\underline{X}^T A \underline{X} = \lambda_k \underline{X}^T B \underline{X}$. It is easily seen that $f(\underline{X})$ at each critical point is simply λ_k; i.e.

$$f(\underline{x}) = \frac{\underline{x}^T A \underline{x}}{\underline{x}^T B \underline{x}} = \lambda_i \frac{\underline{x}^T B \underline{x}}{\underline{x}^T B \underline{x}}$$

$$f(\underline{x}) = \lambda_i$$

Thus, the results of Theorem 6 are not changed by multiplying eigenvectors for an eigenvalue of $C = B^{-1} A$.