INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

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

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

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

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

UMI

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Order Number 9100046

Orthogonal polynomials in the approximation of probability distributions

Oakley, Steven James, Ph.D.

The University of Arizona, 1990

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

ORTHOGONAL POLYNOMIALS IN THE APPROXIMATION OF PROBABILITY DISTRIBUTIONS

by

Steven James Oakley

A Dissertation Submitted to the Faculty of the
DEPARTMENT OF SYSTEMS AND INDUSTRIAL ENGINEERING
In Partial Fulfillment of the Requirements
For the Degree of
DOCTOR OF PHILOSOPHY
In the Graduate College
THE UNIVERSITY OF ARIZONA

1 9 9 0

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Final Examination Committee, we certify that we have read the dissertation prepared by Steven James Oakley

entitled Orthogonal Polynomials in the Approximation of Probability Distributions

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Julia L. Higle	6/13/90 Date
Ronald G. Askin	6/13/90 Date
Suvrajeet Sen	6/13/90 Date
James M. Cushing	6/13/90 Date
Robert S. Maier	6/13/90 Date

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copy of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

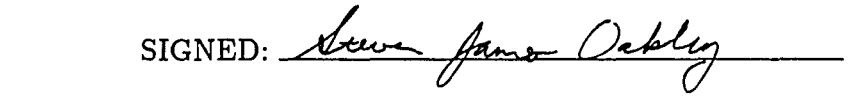
Julia L. Higle	(JH) ~~6/13/90~~ 7/10/90 Date

Dissertation Director

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the copyright holder.

SIGNED: 

ACKNOWLEDGEMENTS

I would like to thank Dr. Julia Higle for suggesting the topic of this dissertation and spending countless hours in discussion of its content and organization. The patient guidance and constant encouragement Dr. Higle provided were invaluable in developing the professional skills required for scientific research. My thanks also to committee members Ronald Askin and Suvrajeet Sen from the Systems and Industrial Engineering Department for carefully reading the dissertation and offering helpful comments, and to Robert Maier and James Cushing from the Department of Mathematics for serving on the defense committee. Finally, I thank my parents for allowing me the opportunity to pursue graduate studies for several years. Their unwavering support provided an ideal environment for finishing my degree.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

An approach to the problem of approximating a continuous probability distribution with a series in orthogonal polynomials is presented. The approach is first motivated with a discussion of theoretical distributions which are inherently difficult to evaluate. Additionally, a practical application which involves such a distribution is developed. The three families of orthogonal polynomials that pertain to the methodology - the Hermite, Laguerre, and Jacobi - are then introduced. Important properties and characterizations of these polynomials are given to lay the mathematical framework for the orthogonal polynomial series representation of the probability density function of a continuous random variable. This representation leads to a similar series for the cumulative distribution function, which is of more practical use for computing probabilities associated with the random variable. It is demonstrated that the representations require only the moments and the domain of the random variable to be known. Relationships of the Hermite, Laguerre, and Jacobi series approximations to the normal, gamma, and beta probability distributions, respectively, are also formally established. Examples and applications of the series are given with appropriate analyses to validate the accuracy of the approximation.

CHAPTER 1

INTRODUCTION

1.0 Introduction

The subject of probability encompasses a multitude of topics ranging from the simple combinatorial analysis of coin tossing and lottery prizes to the complex philosophical arguments between the different factions of frequentists and Bayesians over the definition of probability itself. Within this vast framework of probability, the concept of a random variable is a cornerstone in both theory and applications.

The ability to evaluate the probability density function or cumulative distribution function of a random variable allows us to compute desired probabilities and draw quantitative conclusions about the matter at hand. This has led to widespread use of random variables which have probability density and cumulative distribution functions which are mathematically convenient. Unfortunately, many common random variables do not have convenient formulas, especially in the case of the cumulative distribution function, which is typically of more practical value than is the probability density function or the probability mass function.

In the past, tables have been published for some of these distributions. For example, the literature contains extensive tables of the incomplete gamma and beta functions, which can be used to compute probabilities associated with gamma and beta random variables. The availability of modern computers has made the use of tables obsolete in many cases, but even today elementary probability and statistics textbooks contain tables for the standard normal, student-t, chi-square, and F distributions.

Nevertheless, there remains a large class of random variables which are not convenient to work with and have no available tables. Various techniques have been developed for these cases. It is the aim of this dissertation to discuss and apply one such technique. A method of approximating probability density and cumulative distribution functions with orthogonal polynomial series is presented. We develop the existing techniques more fully, explaining the relationships of the series to common probability distributions. To motivate the need for the technique, we discuss

two examples. We restrict the discussion to continuous random variables. Hereafter any reference to random variables assumes the random variable is continuous unless otherwise stated.

1.1 Motivational Example

It is not difficult to construct random variables whose probability density function or cumulative distribution function is computationally intractable. Such variables can arise from simple algebraic functions of other random variables. For example, given two independent random variables X and Y with respective probability density functions $f_1(x)$ and $f_2(y)$, it can be shown (see, e.g., Hoel, Port, and Stone [1971]) that the probability density function $h(z)$ of the random variable defined by

$$Z = \frac{Y}{X} \qquad [1.1]$$

is given by

$$h(z) = \int_{-\infty}^{\infty} |x| f_1(x) f_2(xz)\, dx, \quad -\infty < z < \infty. \qquad [1.2]$$

While special cases of this integral are known, in general the integral is difficult to evaluate. Consider the case where $X \sim \text{gamma}(\alpha, \beta)$ and $Y \sim \text{beta}(c, d)$. Using the known density functions

$$f_1(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}, \quad x > 0, \qquad f_2(y) = \frac{y^{c-1}(1-y)^{d-1}}{B(c,d)}, \quad 0 < y < 1, \qquad [1.3]$$

the expression becomes

$$h(z) = \frac{\beta^\alpha z^{c-1}}{\Gamma(\alpha) B(c,d)} \int_0^\infty x^{c+\alpha-2} e^{-\beta x} (1-xz)^{d-1} dx, \quad 0 < z < \infty, \qquad [1.4]$$

an unwieldy form to evaluate. Although X and Y both have well-studied distributions, the probability density function of the random variable defined by their quotient is neither commonly known nor easily found. Similar random variables are easily constructed using combinations of sums, differences, quotients, products, and powers of common random variables. While theoretical results exist to find the probability density function, in practice the required computations are often impractical. Further difficulties are encountered when attempting to determine the cumulative distribution function.

Probability density functions from Bayesian probability models also can be inordinately cumbersome to evaluate. Berger [1985] states that in general the posterior probability density function is not easily calculable and gives an example requiring numerical integration. He notes that a large part of the Bayesian literature is devoted to finding conjugate priors because their posterior distributions have convenient forms. The next section introduces a Bayesian traffic model which incorporates distributions from conjugate families, but requires evaluating the distribution of their product, which proves to be as difficult as the case of quotients given above. This model will be described in further detail in Chapter 5, where a specific example is solved using a series in orthogonal polynomials.

1.2 A Bayesian Traffic Conflicts Model

The probabilistic model we describe is designed to approach the general problem of evaluating the accident rate at a given traffic site or intersection. Specifically, it addresses the problem of attempting to evaluate a change in the accident rate after modifications have been made to the intersection in order to improve its safety. Thus, it is assumed throughout that the original accident rate is known to be dangerously high. To determine whether or not the improvements were successful, we wish to quantify the degree of success of the efforts made.

Current techniques for undertaking such a 'before and after' study require a minimum of two years worth of accident data before an accurate assessment of any change in the accident rate can be made. In an effort to decrease this lengthy time requirement, researchers have proposed using *traffic conflict* information. Glauz, Bauer, and Migletz [1985] note that there is no generally accepted definition of a traffic conflict, but give one developed by the Institute of Transport Economics [1977]:

> A traffic conflict is an observable situation in which two or more road users approach each other in space and time to such an extent that there is a risk of collision if their movements remain unchanged.

Traffic conflicts and their potential uses have been studied in the literature for more than twenty years. The potential advantage of using conflict information lies in the frequency of conflicts relative to accidents. While accidents occur only rarely, conflicts can be observed on a daily basis. Once a relationship between conflicts and accidents is established, it is possible to discover information about the accident rate through the study of conflicts.

Such a relationship is implied by the above conflict definition. If every conflict is associated with a 'risk of collision,' then every conflict has a positive probability of becoming an accident. This forms the basis of the model - the link between conflicts and accidents is contained in the assumption that accidents arise from conflicts. Formally, we let $\tilde{P}$ represent the probability that a conflict becomes an accident. Often this is called the accident-to-conflict ratio in the literature. Rather than determine a point estimate of this probability, we model it as a random variable and seek its distribution.

We also require information about the number of conflicts that occur. We let $\tilde{\lambda}_c$ represent the conflict rate at the site, and as with $\tilde{P}$, we model it as a random variable. In the Bayesian structure of the model, we construct distributions for $\tilde{P}$ and $\tilde{\lambda}_c$ before any improvements are made. These are the prior distributions. The method of constructing these priors is given in Chapter 5, where we discuss the theoretical and practical merits of modeling $\tilde{P}$ and $\tilde{\lambda}_c$ with beta and gamma random variables, respectively. One advantage of these choices lies in the fact that beta and gamma distributions belong to conjugate families, which allows easy computation of the posterior distributions from post-improvement conflict and accident data.

To make probabilistic inferences about the post-improvement accident rate $\tilde{\lambda}_a$, we need a formal relationship among the random variables $\tilde{\lambda}_c$, $\tilde{P}$, and $\tilde{\lambda}_a$. Assuming for an instant the degenerate case where $\tilde{\lambda}_c$ and $\tilde{P}$ are constants, say λ_c and P (in the model, a $\tilde{}$ over a letter indicates the variable is a random variable), we know the post-improvement accident rate can be found as the product $P\lambda_c$ of the post-improvement accident-to-conflict ratio and the post-improvement conflict rate. In the more general case where these variables are no longer constants, but rather take on probability distributions, it is clear that for any realization of $\tilde{P}$ and $\tilde{\lambda}_c$ the

product relationship will still hold. Thus, the relationship of the post-improvement accident rate $\tilde{\lambda}_a$ to these random variables is expressed as

$$\tilde{\lambda}_a = \tilde{P}\tilde{\lambda}_c. \qquad [1.5]$$

It is reasonable to assume that the proportion of conflicts that become accidents does not affect the conflict occurrence rate and vice-versa. Formally, this becomes an assumption of independence of the random variables $\tilde{P}$ and $\tilde{\lambda}_c$. The task we face is that of finding the probability density function of the product random variable $\tilde{\lambda}_a$ when the probability density functions of $\tilde{P}$ and $\tilde{\lambda}_c$ are known. Although both of the distributions have the advantage of belonging to conjugate families and have convenient posterior formulas, the evaluation of the distribution of their product is difficult. The problems associated with finding the probability density function of this product random variable are discussed in more detail in the next chapter, which reviews several potential approaches.

1.3 Description of Dissertation

Techniques for evaluating the probability density and cumulative distribution functions of the random variables from the last two sections are presented in a literature review in Chapter 2. The requirements, advantages, and disadvantages of these techniques are discussed in order to give perspective to an orthogonal polynomial series approach, which is developed as the body of the dissertation.

The orthogonal polynomial series approach is outlined in Chapter 3. Definitions and basic properties of orthogonal polynomials are given, and series representations for probability density and cumulative distribution functions of a random variable are provided in a general framework.

The mathematical framework of Chapter 3 is specialized in Chapter 4 for each particular orthogonal family - the Hermite, the Laguerre, and the Jacobi polynomials. Individual orthogonality relations, norms, weighting functions, and series representations are given in each case, and the relationship of each family to a common probability distribution is formally established.

Chapter 5 discusses the applications of the theory presented in Chapter 4. The traffic model introduced earlier is developed in further detail and used as an example for the extended Laguerre series. Examples for Hermite and Jacobi series are also given.

Chapter 6 summarizes the conclusions of the work and discusses future research. Technical material which would disrupt the continuity of the development is contained in appendices.

1.4 Summary of Contribution

Throughout the text of the dissertation, results not prefaced with references from the literature represent original work. These contributions are summarized as follows. The general representation of the cumulative distribution function of a random variable X provided in Chapter 3 in terms of the weighting function and Rodriguez' formula of the classical orthogonal polynomials and the moments of X is new. The specific representations of the probability density and cumulative distribution functions of X given in Chapter 4 for the extended Laguerre and Jacobi polynomials are also original. The Hermite polynomial series expansions are available in the literature, as are expansions in the generalized Laguerre polynomials and the Legendre polynomials, which are special cases of the extended Laguerre and Jacobi polynomials, respectively. The parallel development of the three polynomial families, their series expansions as applied to probability, and their formal relationships to the normal, gamma, and beta distributions represents a more comprehensive and complete view of the role of orthogonal polynomials in the field of applied probability than has previously been presented. Additionally, the traffic conflict model introduced in Chapter 1 and used in Chapter 5 as an application extends models from the current literature of traffic engineering. The consideration of the entire distribution of the mean accident rate generalizes previous work, which considers only the first two moments of the mean accident rate.

CHAPTER 2

LITERATURE REVIEW

2.0 Probability Density Function Evaluation

Many techniques exist for evaluating the probability density function of a random variable. We review the prevalent methods in the literature, with particular emphasis on those applicable to the traffic conflict problem from Chapter 1. Because this problem involves a random variable defined as a function of other random variables, much of the review is focused on techniques pertaining to this special case. Advantages and limitations of the techniques are discussed both to give a broad perspective and to motivate the use of orthogonal polynomials. Additionally, a review of the traffic conflict literature is included as background material for the application in Chapter 5.

2.1 Direct Methods

One method available for the evaluation of the probability density function of a random variable defined by a sum, difference, product, quotient, or power of a given random variable is by direct integration. The most commonly encountered case is the convolution integral for the sum of two random variables. Given two independent random variables X and Y with respective probability density functions $f_1(x)$ and $f_2(y)$, the probability density function $h(z)$ of the random variable defined by $Z = X + Y$ is given by

$$h(z) = \int_{-\infty}^{\infty} f_1(x) f_2(z-x) dx = \int_{-\infty}^{\infty} f_1(z-y) f_2(y) dy, \quad -\infty < z < \infty. \qquad [2.1]$$

It is possible to compute this integral in many cases, especially when X and Y are identically distributed. Results for sums of normal, chi-square, Cauchy, and gamma (with identical scale parameters) random variables can be derived from their convolution integrals.

These cases suggest the possibility of an analogous integral for products of independent random variables. Indeed, such an integral exists. The probability density function $h(z)$ of the random variable $Z = XY$ is

$$h(z) = \int_{-\infty}^{\infty} \frac{f_1(\frac{z}{y})\, f_2(y)}{|y|}\, dy \quad = \int_{-\infty}^{\infty} \frac{f_1(x)\, f_2(\frac{z}{x})}{|x|}\, dx. \qquad [2.2]$$

Unfortunately, as with the quotient case of Chapter 1, convenient forms of this integral do not exist except in trivial cases. For the traffic application, if we let $h(\lambda_a)$, $f_1(\lambda_c)$, and $f_2(P)$ be the probability density functions for, respectively, $\tilde{\lambda}_a$, $\tilde{\lambda}_c$, and $\tilde{P}$, [2.2] becomes

$$h(\lambda_a) = \int_0^1 \frac{f_1(\frac{\lambda_a}{P})\, f_2(P)}{P}\, dP \quad = \int_0^{\infty} \frac{f_1(\lambda_c)\, f_2(\frac{\lambda_a}{\lambda_c})}{\lambda_c}\, d\lambda_c, \qquad [2.3]$$

where f_1 and f_2 are the probability density functions for gamma and beta random variables, respectively. The difficulties of evaluating this integral lead us to consider other approaches.

2.2 Mellin Transforms for Products and Quotients

Epstein [1948] pioneered the use of the Mellin transform for finding probability density functions of products and quotients of independent random variables. Because the Mellin transform is central to the study of products and quotients of random variables in the literature, we discuss it at some length here. The Mellin transform of a function $f(x)$ is defined as

$$M[f(x);s] = \int_0^\infty x^{s-1}\, f(x)\, dx = f^*(s). \qquad [2.4]$$

In general, s is allowed to be complex. Churchill [1972] notes that if $f(x)$ is sectionally continuous over each bounded interval of the positive axis and σ is the real part of s, then the transform exists if $f(x)$ is $O(x^{-m})$ and $0 < \sigma < m$. The inverse Mellin transform exists under these conditions as the complex inversion integral

$$f(x) = \frac{1}{2\pi i} \int_{k-i\infty}^{k+i\infty} x^{-s}\, f^*(s)\, ds. \qquad [2.5]$$

Two properties are needed to derive the Mellin convolution formula for the Mellin transform of the distribution of the product of two independent random variables. The first follows from the definition as

$$M[x^a f(x);s] = \int_0^\infty x^{s-1+a} f(x) dx = f^*(s+a) \qquad [2.6]$$

and the second by letting $x = \frac{u}{k}$ in the defining integral:

$$\int_0^\infty x^{s-1} f(kx) dx = k^{-s} \int_0^\infty u^{s-1} f(u) du, \qquad [2.7]$$

which we write as

$$M[f(kx);s] = k^{-s} f^*(s). \qquad [2.8]$$

Using both properties [2.6] and [2.8], we can write

$$\begin{aligned} M\left[\int_0^\infty y^a f_1\left(\frac{z}{y}\right) f_2(y) dy; s\right] &= \int_0^\infty z^{s-1} \left(\int_0^\infty y^a f_1\left(\frac{z}{y}\right) f_2(y) dy\right) dz \\ &= \int_0^\infty y^a f_2(y) dy \int_0^\infty z^{s-1} f_1\left(\frac{z}{y}\right) dz \\ &= \int_0^\infty y^{a+s} f_2(y) dy \int_0^\infty w^{s-1} f_1(w) dw, \end{aligned} \qquad [2.9]$$

or

$$M\left[\int_0^\infty y^a f_1\left(\frac{z}{y}\right) f_2(y)dy; s\right] = f_2^*(a+s+1)f_1^*(s). \qquad [2.10]$$

For the special case a = -1, this reduces to

$$M\left[\int_0^\infty \frac{f_1(\frac{z}{y})\ f_2(y)}{y}\ dy\right] = f_1^*(s)f_2^*(s). \qquad [2.11]$$

This is usually used in the inverse form

$$M^{-1}[f_1^*(s)f_2^*(s); y] = \int_0^\infty \frac{f_1(\frac{z}{y})\ f_2(y)}{y}\ dy. \qquad [2.12]$$

Comparing this to the integral in [2.2], we see it is equivalent for functions defined on the positive axis and therefore

$$M[h(z); s] = f^*(s) = M[f_1(x); s]\ M[f_2(y); s] = f_1^*(s)\ f_2^*(s). \qquad [2.13]$$

Thus, the Mellin transform of the density function of the product of two independent nonnegative random variables is the product of the Mellin transforms of each component density.

It is possible to derive an analogous expression for the case $Z = X/Y$, where X and Y are both nonnegative, by writing $Z = X(1/Y)$ and applying the product property. Epstein [1948] presents this case, showing that

$$M[h(z); s] = f^*(s) = M[f_1(x); s]M[f_2(y); -s+2] = f_1^*(x)f_2^*(-s+2). \qquad [2.14]$$

As with Laplace transforms, tables are available for transforms of common functions. Unfortunately, Mellin transforms are not as fully tabulated as Laplace transforms and can be difficult to invert in practice. For example, for the traffic conflict model, the Mellin transforms of the probability density functions of $\tilde{\lambda}_c$ and $\tilde{P}$ are found to be, for a gamma(α, β) and a beta(c, d) random variable, respectively,

$$f_{\lambda_c}^*(s) = \frac{\beta^{(1-s)}\ \Gamma(\alpha - 1 + s)}{\Gamma(\alpha)} \qquad [2.15]$$

and

$$f_P^*(s) = \frac{\Gamma(c+d)\ \Gamma(c+s-1)}{\Gamma(c+d+s-1)\ \Gamma(c)}. \qquad [2.16]$$

Applying the product property [2.13], we find the transform of $\tilde{\lambda}_a$ to be

$$f^*_{\tilde{\lambda}_a}(s) = \frac{\beta^{(1-s)}\Gamma(\alpha-1+s)\ \Gamma(c+d)\ \Gamma(c+s-1)}{\Gamma(\alpha)\ \Gamma(c+d+s-1)\ \Gamma(c)}. \qquad [2.17]$$

While each component transform is well-known, this last transform is not directly available in, for example, Oberhettinger [1974], and is difficult to invert using the complex inversion integral. We will discuss another method of inverting this transform in the next section.

Several special cases of products and quotients of independent random variables are available in the literature. Jambunathan [1954] presents examples for the product of n independent beta distributions and relates specialized gamma distributions to a beta distribution as well. Wells, Andersen, and Cell [1962] consider the product of two central or non-central chi-square variates and apply the result to the product of two complex random variables. Springer and Thompson [1964] construct tables of both the density and cumulative functions for the product of two, three, and six independent standard normal random variables. Rider [1965] derives the density for the product and quotient of two independent Cauchy random variables with a direct approach, while Springer and Thompson [1966] use the Mellin transform to confirm that result. Additionally, they produce a formula for the density of the product of n independent specialized beta variables and consider products of up to ten Cauchy factors and standard normal distributions up to seven factors, confirming previous results.

Lomnicki [1967] refines Springer and Thompson's research by simplifying their formulas for standard normal products. He develops the relationship between their work and other distributions, including the exponential, gamma, and Weibull. Finally, he demonstrates how Mellin transforms can be applied to products of random variables with identical distributions but differing scale parameters with specific examples for the exponential and centralized normal distributions.

Springer and Thompson [1970] find that in many cases, the probability density function for products of random variables can be put in the form of a Meijer-G

function (a generalization of the hypergeometric functions ${}_aF_b$) multiplied by a normalizing constant. As the Meijer-G function is not well known, we provide the definition in the form of the following contour integral:

$$G_{pq}^{mn}\left(z \,\middle|\, \begin{matrix} a_1, a_2, \ldots, a_p \\ b_1, b_2, \ldots, b_q \end{matrix}\right) =$$

$$\frac{1}{2\pi i}\int_{\omega-i\infty}^{\omega+i\infty} z^{-s}\frac{\prod_{j=1}^{m}\Gamma(s+b_j)\prod_{j=1}^{n}\Gamma(1-a_j-s)}{\prod_{j=n+1}^{p}\Gamma(s+a_j)\prod_{j=m+1}^{q}\Gamma(1-b_j-s)}ds, \qquad [2.18]$$

where the poles of the integrand must be simple, ω is a real constant defining a Bromwich path separating the poles of $\Gamma(s+b_j)$ from the poles of $\Gamma(1-a_k-s)$, and any empty product $\prod_{j=r}^{r-1}$ is defined to be one. Springer and Thompson use the Meijer-G function to extend their previous work on products of specialized beta variates to the general case, and consider other densities including the centralized normal, specialized gamma, and a mixture of M betas and $M-N$ gammas. A closed form solution resulting from a simplification of the Meijer-G function is presented for the case of iid beta random variables. This confirms an earlier result by Lomnicki.

In general, one disadvantage of the Mellin transform is that by its very definition it is limited to nonnegative random variables. We note that extensions of it to include random variables defined on a negative support are found in the literature (see e.g., Epstein [1948] and Springer and Thompson [1966]), but these methods do not appear to have come into wide use. An advantage of the Mellin transform is derived from the fact that the moments of a random variable are easily found from it. From [2.4], we note that the rth moment of a nonnegative random variable Z with Mellin transform $f^*(\cdot)$ is given by

$$E[Z^r] = f^*(r+1). \qquad [2.19]$$

Thus, the mean and variance of Z can be found from

$$E[Z] = f^*(2)\ ,\ Var[Z] = f^*(3) - (f^*(2))^2. \qquad [2.20]$$

2.3 Laplace Transforms

The use of Laplace transforms for finding distributions is most convenient in the case of sums of independent random variables. From the convolution formula in [2.1], one can easily show that the Laplace transform of the probability density function $h(z)$ of the sum of two independent random variables X and Y is given by the product of the Laplace transforms of the probability density functions of X and Y. We write this property as

$$\mathcal{L}\{h(z)\} = \mathcal{L}\{f_1(x)\}\mathcal{L}\{f_2(y)\}, \qquad [2.21]$$

where the Laplace transform of a function $f(x)$ has the conventional definition

$$\mathcal{L}\{f(x)\} = f(s) = \int_0^\infty e^{-sx} f(x)dx. \qquad [2.22]$$

Because Laplace transforms are extensively tabulated in the literature, the transform of $h(z)$ can often be inverted, yielding the desired result. It is also possible to find probabilities associated with the random variable Z directly from the Laplace transform itself; for recent work in this area, see Platzman, Ammons, and Bartholdi [1988].

The Laplace transform is less applicable for the product case $Z = XY$. However, Lew [1975] shows that under certain conditions, the following relationship exists between Mellin and Laplace transforms:

$$\Gamma(1-s)M[h(z);s] = M[\mathcal{L}\{h(z)\};1-s]. \qquad [2.23]$$

Thus it is theoretically possible to find the Mellin transform of $h(z)$, use the above relation to find the Laplace transform of $h(z)$, and invert using tables of Laplace transforms or the method of residues to find the probability density function $h(z)$ itself. Applying this method to the traffic conflict model, we find the Laplace transform of the probability density function of the random variable $\tilde{\lambda}_a$ to be

$$\mathcal{L}[f(\lambda_a)] = \frac{\Gamma(c+d)\sum_{n=0}^{\infty} \frac{\Gamma(\alpha+n)\ \Gamma(c+n)(\frac{-s}{\beta})^n}{\Gamma(c+d+n)n!}}{\Gamma(\alpha)\ \Gamma(c)} = {}_2F_1(\alpha; c; c+d; \frac{-s}{\beta}), \qquad [2.24]$$

where ${}_2F_1$ represents Gauss' hypergeometric function. To invert this, we use the extensive tables of Roberts and Kaufman [1966] and find the density of $\tilde{\lambda}_a$ is

$$f(\lambda_a) = \beta^{r_1}\ \lambda_a{}^{r_2}\ e^{\frac{-\beta\lambda_a}{2}}\ W_{r_3,r_4}(\beta\lambda_a), \qquad [2.25]$$

where

$$r_1 = \frac{\alpha+c-5}{2},\quad r_2 = \frac{\alpha+c-3}{2},\quad r_3 = \frac{\alpha-c-2d+1}{2},\ and\ \ r_4 = \frac{\alpha-c}{2} \qquad [2.26]$$

and W is the Whittaker function, which is defined in terms of the Kummer function, another hypergeometric function. The Kummer function can be approximated by rational functions, but the requirements are lengthy. As before, the difficulties associated with this undertaking lead us to consider other approaches.

2.4 H-Function Distributions

Fox [1961] introduced the H-function, an extremely generalized function which contains Gauss' hypergeometric, the confluent hypergeometric, and Bessel functions as special cases. By multiplying the function by a normalizing constant, the H-function produces probability distributions. Using Mellin and Laplace transforms along with the H-function, Carter and Springer [1977] generalize most previous work for many of the common distributions. They demonstrate that any product of independent random variables with densities that can be expressed as H-functions has a density that can also be put into the form of an H-function. As with the Meijer-G function, the H-function has a contour integral form. The definition Carter and Springer use is

$$H(z) = H_{pq}^{mn}\left[z \,\middle|\, \begin{matrix}(a_1,\alpha_1),\ldots,(a_p,\alpha_p)\\(b_1,\beta_1),\ldots,(b_q,\beta_q)\end{matrix}\right] =$$

$$\frac{1}{2\pi i}\int_C \frac{\prod_{j=1}^{m}\Gamma(b_j-\beta_j s)\prod_{j=1}^{n}\Gamma(1-a_j+\alpha_j s)}{\prod_{j=m+1}^{q}\Gamma(1-b_j+\beta_j s)\prod_{j=n+1}^{p}\Gamma(a_j-\alpha_j s)} z^s\ ds, \qquad [2.27]$$

where

$$0 \le m \le q,\quad 0 \le n \le p,$$

$$\alpha_j > 0\ for\ j = 1,2,\ldots,p,\quad \beta_j > 0\ for\ j = 1,2,\ldots,q,$$

and a_j $(j = 1, 2, \ldots, p)$ and b_j $(j = 1, 2, \ldots, q)$ are complex numbers such that no pole of $\Gamma(b_j - \beta_j s)$ for $j = 1, 2, \ldots, m$ coincides with any pole of $\Gamma(1 - a_j + \alpha_j s)$ for $j = 1, 2, \ldots, n$. C is a contour in the complex s-plane running from $\omega - i\infty$ to $\omega + i\infty$ such that the points

$$s = (b_j + k)/\beta_j$$

for $j = 1, 2, \ldots, m$ and $k = 0, 1, \ldots$ and the points

$$s = (a_j - 1 - k)/\alpha_j$$

for $j = 1, 2, \ldots, n$ and $k = 0, 1, \ldots$ lie to the right and left of C, respectively. Carter and Springer note that the Meijer-G function is in fact a special case of the H-function (with $\alpha_i = \beta_j = 1, i = 1, 2, ..., p, j = 1, 2, ..., q$). The probability density functions that can be described using an H-function are numerous. They include the exponential, chi-square, gamma, beta, Weibull, half-normal, Rayleigh, half-Cauchy, Maxwell, and the generalized hypergeometric. Carter and Springer further demonstrate that the same closure property of the H-function holds for quotients and rational powers of these densities. Eldred [1979] provides a technique and computer code to evaluate H-function distributions using the method of residues, and Cook [1981] refines the technique. Nevertheless, practical use of the H-function is tedious and involved because of its complexity.

2.5 Fitted Distributions

The difficulties of evaluating the probability density or cumulative distribution function of a given random variable may lead one to consider an approximation. The general idea is to find a distribution that is mathematically convenient and which closely 'fits' the desired distribution in some manner. Several techniques are described in the literature.

Among the simplest and most often-used such techniques is analogous to the method of moments in statistical parameter estimation. Matching a given number of moments of the approximating random variable and the desired random variable is not difficult and can work well in practice for some purposes. The approximation is generally good for rough estimates, but because it is limited to using only the first few moments, information in the tails of the distribution is rarely captured.

The Pearson and Johnson systems are similar in purpose to the method of moments, using as many as four of the moments of the desired random variable to determine an approximating distribution. Hill [1969] points out that the Pearson system is based on thirteen types of frequency curves, the first seven of which are in Pearson [1895]. The Johnson system is divided into three subsystems, which Kendall [1987] develops, citing Johnson [1949] as its main source. However, these systems are traditionally used to represent observational data, and neither is well-suited to computation of the cumulative distribution function. References and examples for these techniques and their extensions as well as other similar methods can be found in Kendall [1987].

Phase distributions represent another possible approximation. Neuts [1981] notes that the set of phase distributions is dense on the set of all probability distributions on $[0, \infty)$, which indicates the potential for approximation. While phase distributions exhibit a variety of shapes and have the advantage of being computationally tractable, Neuts [1981] states no general approximation results are known.

2.6 Orthogonal Polynomial Series Expansions

The earliest literature concerning orthogonal polynomial series expansions as they relate to probability appears to be centered on the Hermite (sometimes referred to as the Chebyshev-Hermite) polynomials. Kendall [1987] cites Charlier [1931] for investigating the roots of the Hermite polynomials, and mentions Charlier developed (sometime in the early 1900's) a series utilizing these polynomials and the moments of a random variable to represent the probability density function. The series appears in the work of several authors, including, as Kendall [1987] notes, Chebyshev, Opperman, Gram, Thiele, and Charlier. The name now attributed to the series is the Gram-Charlier Type A series. (Charlier later developed a series expansion based on the Poisson-Charlier polynomials which is referred to as the Gram-Charlier Type B series.) Edgeworth [1904] approaches the problem from the viewpoint of elementary errors and derives a series equivalent to the Gram-Charlier Type A series using the cumulants of the random variable, which are easily related to the moments. Kendall [1987] restates Edgeworth's claim that for a finite truncation of the series his version gives a better representation of a probability density function because of the ordering of the terms. Cornish and Fisher [1937] also develop the series using cumulants, and by making the additional assumption that the rth cumulant is of order n^{1-r} proceed to derive a truncation using terms up to order n^{-2}. Fisher and Cornish [1960] extend the series to more terms and apply the method to find percentile points of the χ^2, Student-t, and F distributions. Draper and Tierney [1973] present further terms of the Cornish-Fisher series.

The Laguerre polynomials, which Szëgo [1975] notes appear in Chebyshev [1859], can also be used to represent certain functions with an infinite series. Romanovsky [1924] is cited by Hill [1969] for using the Laguerre polynomials in relation to the subject of probability. Hille [1926a] discusses the summability of Laguerre series, while Hille [1926b] and [1926c] proceeds to consider convergence properties of the series for special cases. Uspensky [1927] produces more general results than Hille. Szëgo [1975] (originally published in 1937) provides many theorems, inequalities, and bounds concerning several different types of orthogonal polynomials, including the Laguerre family. Ralston [1965] suggests a more general

definition of the Laguerre polynomials. Hill [1969] develops a series representation in these generalized Laguerre polynomials, but uses a more restrictive series in applications. Sumita [1984] uses Laguerre transforms (functions related to Laguerre polynomials) to determine the elements of the transition probability matrix of a linear growth birth-death process. He notes several applications of the procedure to applied probability and statistics, including the work of Keilson and Nunn [1979], Sumita [1979], Keilson and Sumita [1981], Sumita [1981], Keilson, Nunn, and Sumita [1981], Sumita [1983], Keilson, Petrondas, Sumita, and Wellner [1983], and Keilson and Sumita [1983].

An early reference to the Jacobi polynomials is given by Szëgo [1975], who cites Jacobi [1859] for considering a special case of the Jacobi polynomials. The Jacobi polynomials have several special cases, the most notable being the Legendre polynomials. Shohat and Tamarkin [1943] present a Legendre series for the cumulative distribution function of a random variable defined on $[0, 1]$ when the moments are known, a result which they attribute to Hausdorff [1923]. Sansone [1959] gives a variety of properties and theorems associated with Legendre polynomials, and Szëgo [1975] develops the theory of the more general Jacobi family, noting their applicability for representation of a function defined on any finite interval.

2.7 Conflict Literature Review

We now turn to the traffic conflict problem presented in §1.2. This section summarizes the literature pertaining to traffic conflicts. Traffic conflicts were first considered by Perkins and Harris [1968], who define a conflict as "any potential accident situation" and suggest using conflicts as measures of accident potentials. They describe five main categories of traffic conflicts (left-turn, weave, cross-traffic, red-light violation, and rear-end) and several subcategories amounting to twenty different conflict situations in all. They outline a systematic method called the Traffic Conflicts Technique (TCT) for collecting conflict information and discuss data from thirty intersections. However, no statistical analysis is performed. They conclude that conflicts give insight into the basic causes of accidents and that 'before and after' conflict counts can be used as a quick test of the effectiveness of engineering improvements. Baker [1972] reviews data from three Federal Highway Administration conflict studies performed in cooperation with the state highway departments of Washington, Ohio, and Virginia. Perkins and Harris' [1968] results are reinforced by Baker's conclusions that the studies support the idea that conflicts and accidents are associated and that conflict analysis can be used to evaluate spot improvements shortly after completion. Another conclusion is that the TCT is a faster and more reliable technique for the identification of safety deficiencies than conventional (accident-based) methods and thus should lead to low-cost improvements.

Hayward [1972] points out one deficiency of previous conflict studies - the lack of a precise, quantitative definition of a conflict. Hayward supplies the needed definition through the "time-measured-to-collision" (TMTC), which is defined as the "time required for two vehicles to collide if they continue at their present speeds and on the same path." Filmed sequences from an intersection in Washington, D.C. are used to compute and study TMTC values. The conclusion is that a situation where the TMTC value is one second or less should be declared as a traffic conflict. Hayward suggests that observers could be trained to count these conflicts so that the extensive video equipment used in this study would be unnecessary.

Glennon et al. [1977] critique several conflict studies. They argue that many have claimed conflict information can be beneficial in connection with accident prediction, but no one has quantitatively shown any strong relationship between the two. As does Hayward [1972], they decry the lack of a precise definition of a conflict and place some of the blame for difficulties with conflict studies on this deficiency. They present three uses of a reliable traffic conflict measure: (1) to identify and rank locations for safety improvements, (2) to diagnose specific safety deficiencies at a location for the purpose of determining specific countermeasures, and (3) to measure the safety effectiveness of implemented countermeasures by using the before-and-after study technique. The first possibility is eliminated as not being a realistic use of conflict information - they claim traffic volume will suffice. Potential sample sizes for all three alternatives are computed and they conclude that none of the applications allow practical sample sizes.

Zegeer and Deen [1978] find, in direct opposition to Glennon et al.'s [1977] claim, that traffic volumes accounted for only thirty percent of the variation in the number of conflicts. In a study of five intersections, the TCT method developed by Perkins and Harris is used in a modified form to evaluate the effectiveness of improvements. They advocate the use of conflict data as a base for choosing what improvements to implement and also as a follow-up measure to evaluate those improvements. Hauer [1978] also refutes Glennon et al.'s [1977] arguments by using data from a Canadian study to cast doubt on some of the assumptions on which Glennon et al. [1977] base their conclusions. This is one of the first papers to formulate a probabilistic model of conflicts. Hauer assumes conflicts occur according to a Poisson process with rate λ, but λ is modeled as a gamma random variable whose value changes daily. Appropriate confidence intervals and error probabilities are developed.

Glauz and Migletz [1980] provide a conflict study for twenty-eight intersections using thirteen basic types of conflicts. Their theoretical model has the true mean accident rate proportional to the true mean conflict rate (as in the model of §1.2). The main conclusions of the study are that the TCT is best suited "for diagnosis, improvement evaluation, and confirmation or denial of the presence of

safety hazards or operational problems at suspect locations," and that unless there are many observer errors, the use of conflict data is superior to the use of accident data when predicting the mean accident rate at a location.

Williams [1981] discusses the inconsistencies among several conflict studies, including Perkins and Harris [1968], Baker [1972], Hayward [1972], Glennon et al. [1977], and Zegeer and Deen [1978], and attributes these inconsistencies to the lack of a consistent definition of a conflict and widely varying methods of collecting conflict data. He suggests the possibility of predicting accidents from conflicts is limited even if standard definitions of accidents and conflicts can be adopted. On a positive note, Williams [1981] does state that the common claim that conflicts can be used to evaluate site improvements may hold some promise if an objective form of the TCT (like Hayward's TMTA approach) is used.

Hauer [1982] attempts to refine the definition of a conflict, as well as definitions for exposure and risk. A probabilistic model where every conflict has a probability P of becoming an accident is introduced, where P is unknown but constant. Thus the expected number of accidents at a site is formulated to be the product of P and the number of conflicts at the site. However, no method for using this relationship in an actual study is developed.

Glauz, Bauer, and Migletz [1985] provide a study of forty-six intersections with twelve types of conflicts and a model in which the conflict rate follows a gamma distribution. Conflicts are compared with accident history to see which is better at predicting future accidents. Out of thirteen data sets, the accident-based measure is more precise in eight cases and the conflict-based measure is better in the remaining five cases. This difference is not statistically significant. The overall conclusion is that certain types of traffic conflicts are a useful source of data when estimating the mean accident rate at an intersection.

Hauer and Garder [1986] consider the validity of the traffic conflicts technique. The probabilistic model of Hauer [1982] is used and a technique for estimating the accident rate is deemed valid if it is unbiased and has "satisfactory" variance. The sources of variance are investigated; the variance of the accident-to-conflict ratio (P, the probability of a conflict becoming an accident) is found to

be the main contributor to the variance in estimating the accident rate. Examples using the method of moments and maximum likelihood are presented and tested with simulation. The method of moments is extremely simple to use but sometimes results in a negative variance which is then set to zero. The maximum likelihood estimator is found to be superior although it requires the use of a computer. The question of the validity of either method is left open as Hauer and Garder claim the "simulation experiments lead to the conclusions that both estimators appear unbiased," but never substantiate the statement analytically.

Much of past work has been statistical in nature, but Hauer [1982] is one of the few to construct a probabilistic traffic conflicts model. In Chapter 5 we present a similar model which is more comprehensive in that the problems to which it can be applied are more general, and potential conclusions from the model are stronger. For example, our model can provide probability statements about the post-improvement accident rate based on its entire distribution. Previous models (such as Hauer [1982]) generally consider only the mean, or at most the mean and variance, of the post-improvement accident rate.

CHAPTER 3

INTRODUCTION TO ORTHOGONAL POLYNOMIALS

3.0 Overview of the Orthogonal Approach

This chapter introduces orthogonal polynomials in a general manner, presenting the basic tools needed for the specific applications developed in the next chapter. The following sections define and explain the notation, properties, and functions associated with orthogonal polynomials and their use in the representation of functions. Ultimately, we will use the polynomials to approximate the probability density function or cumulative distribution function of a given random variable using a finite number of moments of the random variable.

3.1 Orthogonality and Weighting Functions

One of the more well-known techniques for function approximation involves the representation of the function using a Fourier series (i.e., a linear combination of sine and cosine functions). In a similar fashion, one can approximate a function using linear combinations of orthogonal polynomial functions. Such polynomial approximations are attractive because the polynomials retain some of the advantages of Fourier series (because of their orthogonality) and yet are computationally simple.

The orthogonality of functions is based on a generalization of the dot product used in vector algebra. Whereas two vectors are said to be orthogonal if their dot product is zero, two functions are said to be orthogonal if the integral of their product is zero. Formally, we use the following definition.

> **Definition** (Beckmann [1973]) Let $\{Q_0(x), Q_1(x), Q_2(x), \ldots\}$ be a countable set of real-valued functions defined on $[a, b]$. The system $\{Q_i\}_{i=0}^{\infty}$ is said to be orthogonal over the interval $[a, b]$ if
>
> $$\int_a^b Q_m(x)Q_n(x)dx = \begin{cases} 0, & \text{for } m \neq n; \\ t_n^2 \neq 0, & \text{for } m = n\ , \end{cases} \qquad [3.1]$$
>
> where t_n is referred to as the norm of $Q_n(x)$ on $[a, b]$.

Additionally, we offer the following generalized notion of orthogonality, which will prove useful in our development of approximations of probability density functions.

Definition (Beckmann [1973]) Let $\{P_0(x), P_1(x), P_2(x), \ldots\}$ be a countable set of functions defined on $[a, b]$. The system of functions is said to be orthogonal over the interval $[a, b]$ with respect to the weighting function $w(x)$ if

$$\int_a^b w(x)P_m(x)P_n(x)dx = \begin{cases} 0, & \text{for } m \neq n; \\ h_n^2 \neq 0, & \text{for } m = n. \end{cases} \qquad [3.2]$$

Clearly the definitions [3.1] and [3.2] are equivalent since a set of functions satisfying [3.2] also satisfies [3.1] when

$$Q_m(x) = \sqrt{w(x)}P_m(x), \; Q_n(x) = \sqrt{w(x)}P_n(x). \qquad [3.3]$$

Formally, the notion of a weighting function is thus a matter of convenience. Nonetheless, certain restrictions are imposed on the weighting function. For example, since $\{Q_i\}_{i=0}^{\infty}$ are real valued, [3.3] implies that

$$w(x) \geq 0, \text{ for } a \leq x \leq b. \qquad [3.4]$$

Additionally, as Beckmann [1973] notes, for infinite limits it is necessary that

$$\begin{aligned} \lim_{x \to -\infty} w(x) &= 0 \text{ if } a = -\infty \\ \lim_{x \to \infty} w(x) &= 0 \text{ if } b = \infty \\ \lim_{x \to \pm\infty} w(x) &= 0 \text{ if } -a = b = \infty. \end{aligned} \qquad [3.5]$$

Although there is a diversity of orthogonal functions and weighting functions, we consider the case when the elements of the set $\{P_i\}_{i=0}^{\infty}$ are polynomial functions.

3.2 The Classical Orthogonal Polynomials

In our approximation of probability functions, we focus specifically on a few well-studied collections of polynomials known as the "classical polynomials." The classical polynomials have a long and rich history in the literature of mathematics, physics, and probability, and their formulas, properties, and relationships are available in standard references. For the purpose of completeness, we will develop them in some detail. There are two commonly used forms for the classical polynomials. The first, a direct expression, is

$$P_k(x) = r_k \sum_{i=0}^{N} c_i v_i(x), \qquad [3.6]$$

which is found, for example, in Abramowitz and Stegun [1964]. The coefficient r_k is typically a simple function of k, c_i is a combinatorial coefficient, v_i a polynomial function, and N takes on only two values, either k or $\left[\frac{k}{2}\right]$, depending on the specific classical polynomial. We note that by convention, the polynomial $P_0(x)$ is always defined to be unity.

An equivalent expression is called "Rodriguez' formula" after Rodriguez [1816], who as Sansone [1959] notes, derived it for the special case of Legendre polynomials. Rodriguez' formula involves successive differentiation:

$$P_k(x) = \frac{l_k}{w(x)} \frac{d^k}{dx^k} \left[w(x)(q(x))^k \right], \qquad [3.7]$$

where l_k is a function of k, $q(x)$ is $1-x^2$, x, or 1, and $w(x)$ is the weighting function. The form [3.6] is useful for determining the polynomials. However, in §3.5, we will see that Rodriguez' formula is more beneficial to our development. Explicit results based on both representations will be given in the next chapter for the classical orthogonal polynomials. In order to establish the relationships between the classical polynomials and probability density functions, we require a formal definition of the former. The definition below refines that of Beckmann [1973].

Definition For a given interval $[a, b]$, a system of orthogonal polynomials is said to be "classical" if, in addition to satisfying conditions [3.4] and [3.5], the weighting function $w(x)$ satisfies the differential equation

$$\frac{w'(x)}{w(x)} = \frac{\gamma(x)}{\beta(x)} = \frac{\gamma_0 + \gamma_1 x}{\beta_0 + \beta_1 x + \beta_2 x^2} = \frac{\gamma_0 + \gamma_1 x}{(x-a)^{\delta(a)}(b-x)^{\delta(b)}}, \qquad [3.8]$$

where

$$\delta(u) = \begin{cases} 1, & \text{for } |u| < \infty; \\ 0, & \text{for } |u| = \infty. \end{cases} \qquad [3.9]$$

and $\gamma_1 + k\beta_2 \neq 0$ for any positive integer k.

The classical polynomials prove especially useful in approximating probability functions because their weighting functions are directly related to well-known probability density functions. The following theorem, based on Beckmann [1973], suggests these relationships.

Theorem 3.1. *If $\{P_i\}_{i=0}^{\infty}$ is a classical system of orthogonal polynomials, the associated weighting function must have one of the following forms:*

$$w(x) = \begin{cases} A(x-a)^{\mu}(b-x)^{\nu} & \text{with } |a| < \infty, |b| < \infty; \qquad [3.10a] \\ A(x-a)^{\mu}e^{-Bx} & \text{with } |a| < \infty, b = \infty; \qquad [3.10b] \\ Ae^{-B(x-C)^2} & \text{with } -a = b = \infty, \qquad [3.10c] \end{cases}$$

where $[a, b]$ is the interval of orthogonality and μ, ν, A, B, and C are real numbers with the restrictions $\mu, \nu > -1$, $A, B > 0$.

Proof.

From the definition of a "classical" system, we see that there are four cases to be considered, depending on the magnitudes of a and b.

(i) First, consider the case when $|a| < \infty$ and $|b| < \infty$. From [3.8] and [3.9], we have $\delta(a) = \delta(b) = 1$, and thus

$$\frac{w'(x)}{w(x)} = \frac{\gamma_0 + \gamma_1 x}{(x-a)(b-x)} = \frac{\mu}{x-a} - \frac{\nu}{b-x}, \qquad [3.11]$$

where μ and ν are constants. Integrating with respect to x yields

$$\ln w(x) + K = \mu \ln(x-a) + \nu \ln(b-x) \qquad [3.12]$$

or

$$w(x) = A(x-a)^{\mu}(b-x)^{\nu} \quad \text{with} \quad A = e^{-K} > 0. \tag{3.13}$$

The norm

$$h_n^2 = \int_a^b w(x)P_n^2(x)dx \neq 0 \tag{3.14}$$

must exist for all n. When $n = 0$, $P_n = 1$, yielding

$$h_0^2 = A\int_a^b (x-a)^{\mu}(b-x)^{\nu}dx. \tag{3.15}$$

We make the linear transformation to the interval [0,1] with the substitution $y = \frac{x-a}{b-a}$. The integral becomes

$$(b-a)^{\mu+\nu+1}A\int_0^1 y^{\mu}(1-y)^{\nu}dy, \tag{3.16}$$

which is a standard beta integral (see [3.21]) requiring $\mu, \nu > -1$. For higher values of n, the restrictions are less stringent.

(ii) Second, consider the case $|a| < \infty$ and $b = \infty$. From [3.8] and [3.9] it follows that $\delta(a) = 1$ and $\delta(b) = 0$, and thus

$$\frac{w'(x)}{w(x)} = \frac{\gamma_0 + \gamma_1 x}{(x-a)}. \tag{3.17}$$

Solving for $w(x)$ in the same manner as above, we find

$$w(x) = A(x-a)^{\mu}e^{-Bx}. \tag{3.18}$$

It is again necessary that $A > 0, \mu > -1$, and from [3.5] it follows that $B > 0$.

(iii) The case in which $a = -\infty$ and $|b| < \infty$ follows the argument of (ii) with the substitution of $-x$ for x.

(iv) Finally, if $a = -\infty$ and $b = \infty$, $\delta(a) = \delta(b) = 0$ and thus

$$\frac{w'(x)}{w(x)} = \gamma_0 + \gamma_1 x, \tag{3.19}$$

which leads to

$$w(x) = Ae^{-B(x-C)^2} \tag{3.20}$$

where A, B, and C are constants with $A, B > 0$. ■

It is interesting to note that the weighting functions specified in Theorem 3.1 agree with the functional forms of certain well-known probability density functions. To see this, suppose the weighting function in [3.10a] is determined by $a = 0$, $b = 1$, $\mu = c - 1$, $\nu = d - 1$, and the constant $A = \frac{1}{B(c,d)}$, where $B(c,d)$ is the beta function defined by the integral

$$B(r,s) = \int_0^1 t^{r-1}(1-t)^{s-1}dt, \quad \text{for } r,s > 0. \tag{3.21}$$

Then we have

$$w(x) = \frac{x^{c-1}(1-x)^{d-1}}{B(c,d)}, \quad 0 < x < 1, \quad c,d > 0, \tag{3.22}$$

which is the probability density function for a beta random variable with parameters c and d. Similarly, if in [3.10b] we let $a = 0$, $\mu = \alpha - 1$, $B = \beta$, and $A = \frac{\beta^\alpha}{\Gamma(\alpha)}$, where

$$\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt, \quad z > 0 \tag{3.23}$$

is the generalized factorial or gamma function, then the weighting function is the density function for a gamma random variable with parameters α and β. Finally, in [3.10c] let $A = \frac{1}{\sqrt{2\pi}}$, $B = \frac{1}{2}$, and $C = 0$. Then $w(x)$ is the density function for a standard normal random variable.

The above result suggests a convenient classification of orthogonal polynomials based on their interval of orthogonality. The families of classical polynomials are naturally subdivided into three categories: the Jacobi polynomials, which are orthogonal over the finite interval $[-1,1]$, the Laguerre polynomials, orthogonal on the semi-infinite interval $[0,\infty]$, and the Hermite polynomials, orthogonal on the infinite interval $[-\infty,\infty]$. Thus, we have three sets of orthogonal polynomials that can be paired with common probability density functions as weighting functions. We will make extensive use of these particular polynomials when approximating the distribution of a random variable; Chapter 4 presents a detailed development for each of these polynomials. The correspondence of the weighting functions of the Jacobi, Laguerre, and Hermite polynomials to the probability density functions of random variables defined over finite, semi-infinite, and infinite intervals will play a key role in the choice of the appropriate orthogonal polynomial with which to approximate.

3.3 Series Representation

Before we construct orthogonal polynomial series expansions to represent probability density or cumulative distribution functions of random variables, we must introduce the basic concepts of square integrability, orthonormality, and completeness. These properties will be essential when using an orthogonal system to represent a function $f(x)$ with a series of the type

$$f(x) = \sum_{k=0}^{\infty} d_k P_k(x) w(x). \qquad [3.24]$$

Not all functions $f(x)$ can be represented by such a series, and the restrictions we must place on $f(x)$ require formal definitions. Our development follows Sansone [1959].

> **Definition** If $f(x)$ is a function defined almost everywhere in the interval $[a, b]$ and $\int_a^b f^2(x)dx < \infty$, then $f(x)$ is said to be square integrable over $[a, b]$. When it is clear what the interval $[a, b]$ is, we will say $f(x)$ is square integrable.
>
> **Definition** A function $f(x)$ that is square integrable over an interval $[a, b]$ is said to be normal over $[a, b]$ if $\int_a^b f^2(x)dx = 1$, i.e., if $f(x)$ has unit norm.
>
> **Definition** A system of functions $\{P_k\}_{k=0}^{\infty}$ which are orthogonal and normal over $[a, b]$ is said to be an orthonormal system. The functions themselves are called orthonormal.

Completeness is now defined using the above definitions.

> **Definition** A system of functions $\{P_k\}_{k=0}^{\infty}$ orthonormal over $[a, b]$ is said to be complete over $[a, b]$ when the only square integrable function which is orthogonal over $[a, b]$ to all functions of the system is a function which is zero almost everywhere with respect to Lebesgue measure.

One of the desirable characteristics of the classical polynomials is their completeness, as established by the following theorem.

Theorem 3.2. *The orthonormal versions of the Jacobi, Laguerre, and Hermite polynomials are complete over the intervals* $[-1,1]$, $[0,\infty)$, *and* $(-\infty,\infty)$, *respectively.*

Proof. For the Jacobi polynomials, see Szegö [1975]. The Laguerre and Hermite cases can be found in Sansone [1959].

In Chapter 4 we introduce a generalization of the Laguerre polynomials called the extended Laguerre polynomials. Proof of the completeness of the extended Laguerre polynomials over the interval $[0,\infty)$ is given in Appendix A. Completeness of an orthogonal system is important when we wish to represent a function with a series such as [3.24]. The properties of a complete system are contained in the following theorems.

Theorem 3.3. *If* $\{P_k\}_{k=0}^{\infty}$ *is a complete system of orthonormal functions and* $f(x)$ *is any square integrable function over* $[a,b]$ *then the sequence* $f_n(x)$, *where*

$$f_n(x) = d_1P_1(x) + d_2P_2(x) + \cdots + d_nP_n(x) \qquad [3.25]$$

with

$$d_k = \int_a^b f(x)P_k(x)dx \qquad [3.26]$$

converges in the mean to $f(x)$, *i.e.,*

$$\lim_{n\to\infty} \int_a^b [f(x) - f_n(x)]^2\, dx = 0. \qquad [3.27]$$

Proof. See Sansone [1959].

Theorem 3.4. *If* $\{P_k\}_{k=0}^{\infty}$ *is a complete system of orthonormal functions over* $[a,b]$ *and* $f(x)$ *is any square integrable function over* $[a,b]$ *then there exist partial sums of the series*

$$\sum_{k=1}^{s_1} d_kP_k(x),\ \sum_{k=1}^{s_2} d_kP_k(x),\ \ldots,\ \sum_{k=1}^{s_n} d_kP_k(x) \qquad [3.28]$$

which converge almost everywhere to $f(x)$ *as* $n \to \infty$.

Proof. See Sansone [1959].

Together, the completeness and orthogonality of the classical polynomials combine to provide us with a reasonable and computationally feasible technique for function approximation. The theorem below defines the coefficients $\{d_k\}_{k=0}^{\infty}$ that appear in the representation [3.24]. The corollary specializes the result for probability density functions.

Theorem 3.5. *If $f : [a,b] \to \Re$ is a square integrable function over $[a,b]$, $\{P_k\}_{k=0}^{\infty}$ is a sequence of polynomials orthogonal over $[a,b]$ with respect to the weighting function $w(x)$, $f(x) = \sum_{k=0}^{\infty} d_k P_k(x) w(x)$ for all $x \in [a,b]$, and $h_k = \left(\int_a^b [P_k(x)]^2 w(x) dx\right)^{\frac{1}{2}}$, then for all k*

$$d_k = \frac{1}{h_k^2} \int_a^b f(x) P_k(x) dx. \qquad [3.29]$$

Proof.

Multiplying [3.24] by $P_k(x)$ and integrating over the interval of orthogonality, we have

$$\begin{aligned} \int_a^b f(x) P_k(x) dx &= \int_a^b \sum_{m=0}^{\infty} d_m P_m(x) P_k(x) w(x) dx \\ &= \sum_{m=0}^{\infty} d_m \int_a^b P_m(x) P_k(x) w(x) dx. \end{aligned} \qquad [3.30]$$

From [3.2], the orthogonality of $\{P_k\}_{k=0}^{\infty}$ with respect to $w(x)$ on $[a,b]$ ensures that the integrals on the right hand side vanish whenever $m \neq k$. Thus,

$$\int_a^b f(x) P_k(x) dx = d_k h_k^2, \qquad [3.31]$$

and the result follows. ∎

Because $P_k(x)$ is a polynomial function for all k, Theorem 3.5 immediately leads to the following corollary for probability density functions.

Corollary 3.6. *Let $f : [a,b] \to \Re$ be a probability density function for a random variable X that is square integrable over $[a,b]$, and let $\{P_k\}_{k=0}^{\infty}$ be a sequence of polynomials orthogonal on $[a,b]$ with respect to a weighting function $w(x)$. If $f(x) = \sum_{k=0}^{\infty} d_k P_k(x) w(x)$ for all $x \in [a,b]$ and $h_k = \left(\int_a^b [P_k(x)]^2 w(x) dx\right)^{\frac{1}{2}}$, then d_k can be expressed as a linear combination of the moments of X, for all k.*

3.4 Weighted Mean Square Error Minimization

Any practical implementation of the polynomial representation in [3.24] will require a truncation of the infinite sum, which induces an error in the approximation. If n polynomials are used, the series may be written as

$$f(x) = \sum_{k=0}^{n} d_k P_k(x) w(x) + E(n,x), \qquad [3.32]$$

where $E(n,x)$ represents the resulting error in the approximation. The following theorem shows that the coefficents d_k as determined by equation [3.29] minimize the weighted mean square error of the truncated series over the interval. An equivalent result is given by Jackson [1941]; however, we use a method of proof found in Beckmann [1973].

Theorem 3.7. *Let $f : [a,b] \to \Re$ be a probability density function for a random variable X that is square integrable over $[a,b]$, $\{P_k\}_{k=0}^{\infty}$ be a system of polynomials orthogonal on $[a,b]$ with respect to the weighting function $w(x)$, and $h_k = \left[\int_a^b [P_k(x)]^2 w(x)dx\right]^{\frac{1}{2}}$.*

Then the coefficients $d_k = \frac{1}{h_k^2} E[P_k(X)]$ minimize the quantity

$$\psi(d_1, d_2, \ldots, d_n) = \int_a^b \left[\frac{f(x)}{\sqrt{w(x)}} - \sqrt{w(x)} \sum_{j=0}^{n} d_j P_j(x)\right]^2 dx. \qquad [3.33]$$

Proof.

By setting the first partial derivatives equal to zero, we obtain

$$\frac{\partial \psi}{\partial d_k} = -2 \int_a^b \left[\frac{f(x)}{\sqrt{w(x)}} - \sqrt{w(x)} \sum_{j=0}^{n} d_j P_j(x)\right] \sqrt{w(x)}\, P_k(x)dx = 0. \qquad [3.34]$$

which yields

$$\begin{aligned} \int_a^b f(x)P_k(x)dx &= \int_a^b w(x) \sum_{j=0}^{n} d_j P_j(x) P_k(x)dx \\ \Rightarrow E[P_k(X)] &= \sum_{j=0}^{n} d_j \int_a^b w(x) P_j(x) P_k(x)dx = d_k h_k^2 \qquad [3.35] \\ \Rightarrow d_k &= \frac{1}{h_k^2} E[P_k(X)]. \end{aligned}$$

We calculate the second partial derivatives from [3.34] to be

$$\frac{\partial^2 \psi}{\partial d_k^2} = 2 \int_a^b w(x)[P_k(x)]^2 dx = 2h_k^2 > 0, \qquad [3.36]$$

where the last term follows from [3.2]. Orthogonality of the polynomials ensures that all off-diagonal elements of the Hessian matrix are zero (corresponding to the mixed partial derivatives), and thus the Hessian is positive definite. This implies that ψ is convex, and thus the coefficients from [3.35] must minimize [3.33]. ■

Theorem 3.7 shows that for orthogonal polynomials the d_k can be calculated independent of n, the number of polynomials used in the truncated series. Thus the coefficients minimizing the weighted mean square error for the finite series [3.32] are identical to those determined by the infinite series [3.24]. This property does not hold in general for non-orthogonal functions, as Beckmann [1973] notes.

Some insight as to why the minimization property of Theorem 3.7 might be advantageous in the representation of probability density functions may be gained by writing the integral in [3.33] as

$$\int_a^b \left[\frac{f(x)}{\sqrt{w(x)}} - \sqrt{w(x)} \sum_{k=0}^{n} d_k P_k(x) \right]^2 dx = \int_a^b \frac{1}{w(x)} \left[f(x) - w(x) \sum_{k=0}^{n} d_k P_k(x) \right]^2 dx$$
$$= \int_a^b \frac{[E(n,x)]^2}{w(x)} dx, \qquad [3.37]$$

where $E(n,x)$ is the error of the truncated series from [3.32]. As discussed in §3.2, the weighting functions for the classical orthogonal polynomials can be written as the probability density functions for gamma, beta, and standard normal random variables. In the case of a standard normal weighting function, the 'bell' shape of the standard normal probability density function suggests the integral in [3.37] above will be more heavily 'weighted' for large $|x|$ since it is inversely proportional to $w(x)$. Because the function $f(x)$ that we wish to represent is itself a probability density function, this further suggests that the truncated series approximation to $f(x)$ may have small errors in the tails (large $|x|$), a desirable attribute. For gamma and beta weighting functions, there is no fixed 'bell' shape, and the matter is unclear. However, in these cases, it is possible to select the weighting function parameters

and thus determine the shape of the weighting function. This is discussed further in the implementation of the approximating series in Chapter 5.

3.5 Cumulative Distribution Function

Knowledge of the probability density function, cumulative distribution function, or characteristic function of a continuous random variable X is sufficient to describe X. However, in many applications, such as those described in Chapter 1, a tractable form for the cumulative distribution function is of the utmost importance. For example, evaluations of certain tail probabilities are most easily obtained from the cumulative distribution function. In the following theorem, we show that the orthogonal representation of the probability density function offers a convenient representation of the cumulative distribution function as well.

Theorem 3.8. *Let X be a random variable defined on $[a, b]$ with probability density function $f(x)$, and let $\{P_k\}_{k=0}^{\infty}$ be orthogonal polynomials with respect to the weighting function $w(x)$ on $[a, b]$ satisfying Rodriguez' formula [3.7]. If $f(x)$ is represented by the series expansion*

$$f(x) = \sum_{k=0}^{\infty} d_k P_k(x) w(x) \qquad [3.38]$$

then the cumulative distribution function of X is given by

$$F(x) = P\{X \leq x\} = d_0 l_0 \int_{-\infty}^{x} w(t)dt + \sum_{k=1}^{\infty} d_k l_k \left[\frac{d^{k-1}}{dt^{k-1}} \left[w(t)(q(t))^k \right] \Big|_{-\infty}^{x} \right], \qquad [3.39]$$

where l_k and $q(t)$ are determined from [3.7] for the specific family of polynomials.

Proof.

By definition,

$$F(x) = \int_{-\infty}^{x} f(t)dt. \qquad [3.40]$$

From the representation [3.38],

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} \sum_{k=0}^{\infty} d_k P_k(t) w(t) dt \\ &= \sum_{k=0}^{\infty} d_k \int_{-\infty}^{x} P_k(t) w(t) dt \end{aligned} \qquad [3.41]$$

Using Rodriguez' formula for $P_k(t)$, we write

$$\begin{aligned} F(x) &= \sum_{k=0}^{\infty} d_k \int_{-\infty}^{x} \frac{l_k}{w(t)} \left[\frac{d^k}{dt^k} \left[w(t)(q(t))^k \right] \right] w(t)dt \\ &= \sum_{k=0}^{\infty} d_k l_k \int_{-\infty}^{x} \frac{d^k}{dt^k} \left[w(t)(q(t))^k \right] dt. \end{aligned} \qquad [3.42]$$

Finally, by separating the summation into two parts and integrating the last sum we obtain [3.39]. ■

In Chapter 4 we show that for a given random variable X, weighting function $w(x)$, interval $[a, b]$, and set of polynomials orthogonal with respect to the weighting function over the interval, the computation of the series [3.42] for $F(x)$ (as in the case of the probability density function) requires that only the moments of X be known. Further simplifications are possible once the specific family of orthogonal polynomials has been determined.

CHAPTER 4

THE THREE ORTHOGONAL POLYNOMIAL FAMILIES

4.0 Introduction

The use of the Hermite, Laguerre, and Jacobi polynomials in the approximation of probability distributions is presented in this chapter. Based on the general theory developed in Chapter 3, series expansions for the probability density function and cumulative distribution function of a continuous random variable are specialized for these polynomials, and the relationships of the polynomials to the normal, gamma, and beta distributions are formalized. In addition, we suggest a method for selecting the appropriate orthogonal system to apply in approximating a continuous random variable.

4.1 The Hermite Polynomials

Historically, the use of orthogonal polynomials in the approximation of random variables originated with the Hermite polynomials. The series representation developed here is called the Gram-Charlier Type A series after J.P. Gram and C.V.L. Charlier, although Kendall [1987] notes that others, including Chebyshev, had previously derived the series.

4.1.1 Definition and Rodriguez' Formula

The Hermite polynomials $\{H_k\}_{k=0}^{\infty}$ have two different definitions in the literature, one used primarily in mathematics and one used in probability theory. We require the probabilistic definition and use the following characterization from Abramowitz and Stegun [1964].

Definition The kth degree Hermite polynomial is defined by

$$H_k(x) = k! \sum_{m=0}^{\lfloor \frac{k}{2} \rfloor} \frac{(-1)^m}{m! 2^m (k-2m)!} x^{k-2m}, \quad k = 0, 1, \ldots \qquad [4.1]$$

where $\lfloor \cdot \rfloor$ is the greatest integer function. The conventional weighting function for the Hermite polynomials is $w(x) = e^{\frac{-x^2}{2}}$. However, to emphasize the relationship of the Hermite polynomials to the standard normal density function, we use the weighting function $w(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$. The introduction of the constant $\frac{1}{\sqrt{2\pi}}$ is for convenience only; it does not affect the definition of the polynomials or the property of orthogonality.

Rodriguez' formula [3.7] for Hermite polynomials is found from Abramowitz and Stegun [1964] to be

$$H_k(x) = \frac{l_k}{w(x)} \frac{d^k}{dx^k} \left[w(x) (q(x))^k \right] = \frac{(-1)^k}{\frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}} \frac{d^k}{dx^k} \left[\frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \right], \quad k = 0, 1, \ldots \qquad [4.2]$$

Although [4.2] can be used as an alternate definition, its practical use in defining the polynomials is limited. We will use Rodriguez' formula when finding an expression for the cumulative distribution function of a random variable. We note for future reference that $l_k = (-1)^k$ and $q(x) = 1$ for all x.

4.1.2 Orthogonality and Norm

As previously noted, the interval of orthogonality for Hermite polynomials is $(-\infty, \infty)$. The orthogonality relation is relatively simple and can be found in several references (see, e.g., Kendall [1987] or Jackson [1941]).

Theorem 4.1. *If* $w(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$ *and* $H_n(x)$ *is defined as in [4.2], then*

$$h_{m,n}^2 = \int_{-\infty}^{\infty} w(x) H_m(x) H_n(x) dx \; = \; \begin{cases} 0, & \text{for } m \neq n; \\ n!, & \text{for } m = n. \end{cases} \qquad [4.3]$$

Proof. See Kendall [1987].

Comparing [3.2] and [4.3], we see that $h_{n,n}^2 = n!$ for the Hermite polynomials. The construction of the series representation for a probability density function can now be undertaken.

4.1.3 Probability Density Function Series Representation

From Corollary 3.6, we know for a given random variable X whose density function $f(x)$ is square integrable over an interval [a,b], the sequence of coefficients required to define the series representation of the probability density function can be constructed from the moments of X. For a representation based on the Hermite polynomials, the series is

$$f(x) = \sum_{k=0}^{\infty} d_k H_k(x) w(x). \qquad [4.4]$$

Theorem [3.5] and [4.1] provide the coefficients

$$\begin{aligned} d_k &= \frac{1}{k!} E[H_k(X)] \\ &= \sum_{m=0}^{\lfloor \frac{k}{2} \rfloor} \frac{(-1)^m}{m! 2^m (k-2m)!} \mu_{k-2m}, \quad k = 0, 1, \ldots \end{aligned} \qquad [4.5]$$

where $\mu_i = E[X^i]$ is the ith moment of X. As an example, the first seven coefficients are tabulated below:

k	d_k
0	1
1	μ_1
2	$\frac{1}{2!}(\mu_2 - 1)$
3	$\frac{1}{3!}(\mu_3 - 3\mu_1)$
4	$\frac{1}{4!}(\mu_4 - 6\mu_2 + 3)$
5	$\frac{1}{5!}(\mu_5 - 10\mu_3 + 15\mu_1)$
6	$\frac{1}{6!}(\mu_6 - 15\mu_4 + 45\mu_2 - 15)$

[4.6]

Note that for odd k, d_k is a function of the odd moments of X, and correspondingly for even k, d_k is a function of the even moments of X. Thus, for a distribution symmetric about zero, if the odd moments exist, they are zero and the odd coefficients d_k are zero also. This is one representation of the Type A Gram-Charlier series. Alternative representations express the coefficients d_k with the moments about the mean $\mu'_i = E[(X - \mu_1)^i]$, or using the cumulants of X, which can be related to either set of moments, or for a standardized random variable with zero mean and unit variance. See Kendall [1987] for these forms.

4.1.4 Cumulative Distribution Function

In §3.5 we provide the general expression for the cumulative distribution function of X. The theorem below specializes that form for the Hermite polynomials and is equivalent to the representation provided in Kendall [1987].

Theorem 4.2. *If the probability density function $f(x)$ of a continuous random variable X has the Hermite polynomial representation*

$$f(x) = w(x)\sum_{k=0}^{\infty} d_k H_k(x), \qquad [4.7]$$

where $w(x)$ is the probability density function for a standard normal random variable and

$$d_k = \frac{1}{k!}E\left[H_k(X)\right], \qquad [4.8]$$

then the cumulative distribution function of X is determined by

$$F(x) = \Phi(x) - w(x)\sum_{k=1}^{\infty} d_k H_{k-1}(x), \qquad [4.9]$$

where

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{\frac{-t^2}{2}}dt, \qquad [4.10]$$

the cumulative distribution function for a standard normal random variable.

Proof.

From Theorem 3.8,

$$F(x) = l_0 d_0 \int_{-\infty}^{x} w(t)dt + \sum_{k=1}^{\infty} d_k l_k \left[\frac{d^{k-1}}{dt^{k-1}}[w(t)(q(t))^k]\Big|_{-\infty}^{x}\right]. \qquad [4.11]$$

We previously noted $l_k = (-1)^k$, $q(t) = 1$ for all t, and $d_0 = 1$. Since $w(t)$ is the standard normal density function, it follows that

$$F(x) = \Phi(x) + \sum_{k=1}^{\infty} d_k(-1)^k \left[\frac{d^{k-1}}{dt^{k-1}}[w(t)]\Big|_{-\infty}^{x}\right]. \qquad [4.12]$$

Thus, from [4.2]

$$H_k(t) = \frac{(-1)^k}{w(t)}\frac{d^k}{dt^k}[w(t)], \qquad [4.13]$$

and it follows that

$$F(x) = \Phi(x) - \sum_{k=1}^{\infty} d_k \left[H_{k-1}(t)w(t)\right]\Big|_{-\infty}^{x}. \qquad [4.14]$$

Evaluating the argument in brackets yields [4.9]. ■

4.1.5 Relationship to Standard Normal Distribution

From its form alone, [4.14] is clearly related to the standard normal distribution. The following theorem and corollary describe this relationship.

Theorem 4.3. *If X is a standard normal random variable, then each coefficient d_k, $k = 1, 2, \ldots, n$ in the representation*

$$F(x) \simeq \Phi(x) - w(x) \sum_{k=1}^{n} d_k H_{k-1}(x) \qquad [4.15]$$

is zero, and the approximation is exact for all x.

Proof.

From [4.5], we have

$$d_k = \sum_{m=0}^{\lfloor \frac{k}{2} \rfloor} \frac{(-1)^m}{m! 2^m (k-2m)!} \mu_{k-2m}, \quad k = 0, 1, \ldots \qquad [4.16]$$

Since the standard normal distribution is symmetric about zero, its odd moments are zero and $d_k = 0$ for odd k. Thus we need only consider even k. It can be shown (see Kendall [1987], for example) the even moments of a $N(0,1)$ distribution are

$$\mu_{2r} = \frac{(2r)!}{2^r r!}, \quad r = 0, 1, \ldots, \qquad [4.17]$$

which, when rewritten in the form

$$\mu_{k-2m} = \frac{(k-2m)!}{2^{\frac{k}{2}-m} (\frac{k}{2} - m)!}, \quad k - 2m = 0, 2, \ldots, \qquad [4.18]$$

yields

$$d_k = \sum_{m=0}^{k/2} \frac{(-1)^m}{m!\, 2^{\frac{k}{2}} (\frac{k}{2} - m)!}, \quad k = 2, 4, \ldots \qquad [4.19]$$

We make the substitution $j = \frac{k}{2}$ and use the identity $\binom{j}{m} = \frac{j!}{m!(j-m)!}$, which yields

$$d_{2j} = \frac{1}{2^j j!} \sum_{m=0}^{j} (-1)^m \binom{j}{m}, \quad j = 1, 2, \ldots \qquad [4.20]$$

The binomial summation is zero, establishing $d_k = 0$ for all k. ■

The normal distribution has the convenient property that if $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$. This leads immediately to the following result.

Corollary 4.4. *If X is a normally distributed random variable with mean μ and variance σ^2, then the approximation*

$$F(z) = \Phi(z) - w(z) \sum_{k=1}^{n} d_k H_{k-1}(z) \qquad [4.21]$$

to the cumulative distribution function of the random variable $Z = \frac{X-\mu}{\sigma}$ is exact for all z.

4.2 The Laguerre Polynomials

The Hermite polynomial series representation of a probability density function or cumulative distribution function is the one commonly found in the literature. However, similar representations using the Laguerre or Jacobi polynomials can also be developed. This section describes such a development for the Laguerre polynomials.

Although Chebyshev [1859] is the earliest appearance of the Laguerre polynomials in the literature, they are named after Edmond Laguerre, who, as Sansone [1959] notes, discusses them in Laguerre [1879]. We develop an extension of Hill [1969], who uses the conventional generalized Laguerre polynomials.

4.2.1 Definition and Rodriguez' Formula

Two systems of orthogonal polynomials from the Laguerre family are found in the literature. The first, usually called the Laguerre polynomials, is a special case of the second, which are the generalized Laguerre polynomials. In order to fully understand the relationship of the Laguerre family to the gamma distribution, we further generalize the polynomials with the addition of a parameter. We call these polynomials the extended Laguerre polynomials. Ralston [1965] and Hill [1969] both allude to the extended Laguerre polynomials, but no full treatment is given. We define them as follows.

Definition. The kth degree extended Laguerre polynomial with parameters $\alpha > -1$ and $\beta > 0$ is given by

$$L_k^{(\alpha,\beta)}(x) = \sum_{m=0}^{k} \binom{k}{m} \frac{(-1)^m \Gamma(\alpha+1+k)}{\Gamma(\alpha+1+k-m)} (\beta x)^{k-m} \quad , \quad k = 0, 1, 2, \ldots \qquad [4.22]$$

If $\beta = 1$, the extended Laguerre polynomials collapse to the generalized Laguerre polynomials, and if $\alpha = 0$ and $\beta = 1$ they collapse to the Laguerre polynomials. From [3.10b], the weighting function for the extended Laguerre polynomials may be written as

$$w(x) = \frac{\beta^{\alpha+1} x^{\alpha} e^{-\beta x}}{\Gamma(\alpha+1)}, \qquad [4.23]$$

where α and β are the parameters in [4.22]. Qualitatively, the extended Laguerre polynomials differ from the Hermite polynomials in that the Hermite have no parameters. The relationship of the weighting function to a probability density function aids in understanding this difference. The Hermite polynomials are related to the standard normal density function, which has fixed parameters, while, as is clear from [4.23], the extended Laguerre polynomials are related to a gamma distribution with shape parameter $\alpha + 1$ and scale parameter β. This relationship is formalized in §4.2.5.

Although Rodriguez' formula for the extended Laguerre polynomials is a natural extension of the form for the generalized Laguerre polynomials, it is not readily available in the literature. We therefore present it formally with the following theorem.

Theorem 4.5. *Rodriguez' formula for the extended Laguerre polynomials (defined by [4.22]) is given by*

$$L_k^{(\alpha,\beta)}(x) = \frac{l_k}{w(x)}\frac{d^k}{dx^k}[w(x)(q(x))^k] = \frac{(-1)^k\Gamma(\alpha+1)}{\beta^{\alpha+1}x^\alpha e^{-\beta x}}\frac{d^k}{dx^k}\left[\frac{\beta^{\alpha+1}x^\alpha e^{-\beta x}}{\Gamma(\alpha+1)}x^k\right] = \frac{(-1)^k}{x^\alpha e^{-\beta x}}\frac{d^k}{dx^k}\left[x^{\alpha+k}e^{-\beta x}\right]. \quad [4.24]$$

Proof. (method of Sansone [1959])

From [4.22] it follows that

$$(-1)^k x^\alpha e^{-\beta x}L_k^{(\alpha,\beta)}(x) = \sum_{m=0}^{k}\binom{k}{m}\frac{(-1)^{k-m}\Gamma(\alpha+1+k)}{\Gamma(\alpha+1+k-m)}\beta^{k-m}x^{\alpha+k-m}e^{-\beta x} = \sum_{m=0}^{k}\binom{k}{m}\left[(\alpha+1+k-m)_m\ x^{\alpha+k-m}\right]\left[(-\beta)^{k-m}e^{-\beta x}\right] \quad [4.25]$$

where the Pochhammer notation $(\alpha)_n$ is defined for $\alpha > 0$ as

$$(\alpha)_n = \begin{cases} 1, & \text{for } n = 0; \\ (\alpha)(\alpha+1)\cdots(\alpha+n-1) = \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)}, & \text{for } n = 1,2,\ldots \end{cases} \quad [4.26]$$

and we have used the fact that $(-1)^m = (-1)^{-m}$. Writing the terms in brackets as derivatives yields

$$(-1)^k x^\alpha e^{-\beta x}L_k^{(\alpha,\beta)}(x) = \sum_{m=0}^{k}\binom{k}{m}\frac{d^m}{dx^m}\left[x^{\alpha+k}\right]\frac{d^{k-m}}{dx^{k-m}}\left[e^{-\beta x}\right]. \quad [4.27]$$

Applying Leibniz' rule for differentiation of products (see Appendix C) gives the last part of [4.24]. ∎

Thus, from the general form [3.7] of Rodriguez' formula, we have $l_k = (-1)^k$ and $q(x) = x$ for the extended Laguerre polynomials.

4.2.2 Orthogonality and Norm

The orthogonality relations for the Laguerre and generalized Laguerre polynomials are available in standard references such as Abramowitz and Stegun [1964] and Beyer [1987]. However, as this is not true for the extended Laguerre polynomials, we provide the following.

Theorem 4.6. *If* $w(x) = \frac{\beta^{\alpha+1}x^{\alpha}e^{-\beta x}}{\Gamma(\alpha+1)}$, *then the extended Laguerre polynomials satisfy*

$$h_{m,n}^2 = \int_0^{\infty} w(x)L_m^{(\alpha,\beta)}(x)L_n^{(\alpha,\beta)}(x)dx = \begin{cases} 0, & \text{for } m \neq n; \\ n!(\alpha+1)_n\ , & \text{for } m = n\ . \end{cases} \qquad [4.28]$$

Proof. (method of Hill [1969])

We assume without loss of generality that $m \leq n$ and let $\gamma(x) = w(x)x^n$. Rodriguez' formula [4.24] is then

$$L_n^{(\alpha,\beta)}(x) = \frac{(-1)^n}{w(x)}\frac{d^n}{dx^n}[\gamma(x)]. \qquad [4.29]$$

We thus have

$$h_{m,n}^2 = (-1)^n \int_0^{\infty} L_m^{(\alpha,\beta)}(x)\ \gamma^{(n)}(x)dx. \qquad [4.30]$$

To evaluate [4.30] we successively integrate by parts, letting

$$u = \frac{d^{k-1}}{dx^{k-1}}L_m^{(\alpha,\beta)}(x) = \left[L_m^{(\alpha,\beta)}(x)\right]^{(k-1)},\ dv = \gamma^{(n-k+1)}(x)dx \qquad [4.31]$$

for the kth step. For example, in the first step, let

$$u = L_m^{(\alpha,\beta)}(x),\quad dv = \gamma^{(n)}(x)dx. \qquad [4.32]$$

Then the uv term is zero at both infinity and zero, leaving

$$h_{m,n}^2 = (-1)^n\left[-\int_0^{\infty}\gamma^{(n-1)}(x)\left[L_m^{(\alpha,\beta)}(x)\right]^{(1)}dx\right]. \qquad [4.33]$$

Continuing this process leads to

$$h_{m,n}^2 = (-1)^{n+2}\int_0^{\infty}\gamma^{(n-2)}(x)\left[L_m^{(\alpha,\beta)}(x)\right]^{(2)}dx \qquad [4.34]$$

after the second step, and at the mth step,

$$h_{m,n}^2 = (-1)^{n+m} \int_0^\infty \gamma^{(n-m)}(x) \left[L_m^{(\alpha,\beta)}(x) \right]^{(m)} dx. \qquad [4.35]$$

For the case $m < n$, one more integration will show $h_{m,n} = 0$ since the $(m+1)$st derivative of the mth order polynomial $L_m^{(\alpha,\beta)}(x)$ must be zero. For the case $m = n$, [4.35] simplifies to

$$h_{m,n}^2 = \int_0^\infty \gamma(x) \left[L_n^{(\alpha,\beta)}(x) \right]^{(n)} dx. \qquad [4.36]$$

The coefficient of x^n in the nth degree extended Laguerre polynomial is found from [4.22] to be β^n, so

$$\left[L_n^{(\alpha,\beta)}(x) \right]^{(n)} = n!\beta^n, \qquad [4.37]$$

which yields

$$\begin{aligned} h_{m,n}^2 &= n!\beta^n \int_0^\infty \gamma(x) dx \\ &= n!\beta^n \int_0^\infty \frac{\beta^{\alpha+1} x^{\alpha+n} e^{-\beta x}}{\Gamma(\alpha+1)} dx \\ &= \frac{n!}{\Gamma(\alpha+1)} \int_0^\infty \beta^{n+\alpha+1} x^{\alpha+n} e^{-\beta x} dx. \end{aligned} \qquad [4.38]$$

For $\beta > 0$ it is known (see Burington[1973], p. 105) that

$$\int_0^\infty \beta^{n+\alpha+1} x^{\alpha+n} e^{-\beta x} dx = \Gamma(\alpha+1+n), \qquad [4.39]$$

and thus,

$$h_{m,n}^2 = \frac{n!\Gamma(\alpha+1+n)}{\Gamma(\alpha+1)}. \qquad [4.40]$$

Using the Pochhammer notation [4.26] yields the required expression. ■

4.2.3 Probability Density Function Series Representation

For random variables defined on $[0, \infty)$, a series expansion based on the gamma probability density function may be more appropriate than one based on the normal density. We develop such an expansion as an extension to Hill [1969], who uses the gamma density function, but limits the parameters. For a nonnegative random variable X with probability density function $f(x)$, the series is

$$f(x) = \sum_{k=0}^{\infty} d_k L_k^{(\alpha,\beta)}(x) w(x). \qquad [4.41]$$

From Theorem 3.5,

$$d_k = \frac{1}{k!(\alpha+1)_k} E[L_k^{(\alpha,\beta)}(X)], \qquad [4.42]$$

which, upon using [4.22], becomes

$$d_k = \frac{1}{k!(\alpha+1)_k} \sum_{m=0}^{k} \binom{k}{m} \frac{(-1)^m \Gamma(\alpha+1+k)}{\Gamma(\alpha+1+k-m)} \beta^{k-m} \mu_{k-m} \qquad [4.43]$$

The first few coefficients are

$$\begin{aligned} d_0 &= 1 \\ d_1 &= \frac{\beta\mu_1 - \alpha - 1}{\alpha+1} \\ d_2 &= \frac{\beta^2\mu_2 - 2(\alpha+2)\beta\mu_1 + (\alpha+1)(\alpha+2)}{2(\alpha+1)(\alpha+2)}. \end{aligned} \qquad [4.44]$$

When $\beta = 1$ and $\alpha = 0$ or 1, these correspond to the Laguerre series of Hill[1969].

4.2.4 Cumulative Distribution Function

From §3.5 we have the following for the extended Laguerre polynomials.

Theorem 4.7. *If the probability density function $f(x)$ of a nonnegative continuous random variable X has the extended Laguerre polynomial representation*

$$f(x) = w(x)\sum_{k=0}^{\infty} d_k L_k^{(\alpha,\beta)}(x) \qquad [4.45]$$

almost everywhere in $[0,\infty)$ where $w(x)$ is the probability density function for a gamma $(\alpha+1,\beta)$ random variable and

$$d_k = \frac{1}{k!(\alpha+1)_k} E[L_k^{(\alpha,\beta)}(X)], \qquad [4.46]$$

then the cumulative distribution function of X is determined by

$$F(x) = P(\alpha+1,\beta x) - \frac{(\beta x)^{\alpha+1}e^{-\beta x}}{\Gamma(\alpha+1)}\sum_{k=1}^{\infty} d_k L_{k-1}^{(\alpha+1,\beta)}(x), \qquad [4.47]$$

where the incomplete gamma function $P(\alpha+1,\beta x)$ is defined for $\alpha+1,\beta x>0$ as

$$P(\alpha+1,\beta x) = \frac{1}{\Gamma(\alpha+1)}\int_0^{\beta x} w^{\alpha}e^{-w}dw. \qquad [4.48]$$

Proof.

Since X is nonnegative, Theorem 3.8 yields

$$F(x) = l_0 d_0 \int_0^x w(t)dt + \sum_{k=1}^{\infty} d_k l_k \left[\frac{d^{k-1}}{dt^{k-1}}[w(t)(q(t))^k]\Big|_0^x\right]. \qquad [4.49]$$

We have $d_0 = 1$ from [4.44] and from Rodriguez' formula we know $l_k = (-1)^k$ and $q(t) = t$. Substituting the full expression for $w(t)$ then leads to

$$F(x) = \frac{\beta^{\alpha+1}}{\Gamma(\alpha+1)}\left[\int_0^x t^{\alpha}e^{-\beta t}dt + \sum_{k=1}^{\infty} d_k(-1)^k \frac{d^{k-1}}{dt^{k-1}}\left[t^{\alpha+k}e^{-\beta t}\right]\Big|_0^x\right]. \qquad [4.50]$$

It is apparent from Rodriguez' formula [4.24] that

$$L_{k-1}^{(\alpha+1,\beta)}(t)(-1)^{k-1}t^{\alpha+1}e^{-\beta t} = \frac{d^{k-1}}{dt^{k-1}}\left[t^{\alpha+k}e^{-\beta t}\right]. \qquad [4.51]$$

Using this in [4.50] gives

$$F(x) = \frac{\beta^{\alpha+1}}{\Gamma(\alpha+1)}\left[\int_0^x t^\alpha e^{-\beta t}dt - x^{\alpha+1}e^{-\beta x}\sum_{k=1}^{\infty} d_k L_{k-1}^{(\alpha+1,\beta)}(x)\right]. \qquad [4.52]$$

We make the substitution $w = \beta t$ in the first integral:

$$\begin{aligned}\int_0^x t^\alpha e^{-\beta t}dt &= \int_0^{\beta x}\left(\frac{w}{\beta}\right)^\alpha e^{-w}\frac{dw}{\beta} \\ &= \frac{1}{\beta^{\alpha+1}}\int_0^{\beta x} w^\alpha e^{-w}dw \\ &= \frac{\Gamma(\alpha+1)\ P(\alpha+1,\beta x)}{\beta^{\alpha+1}}.\end{aligned} \qquad [4.53]$$

The required formula [4.47] then follows. ■

4.2.5 Relationship to Gamma Distribution

In Theorem 4.3, we show that the Hermite polynomial series is exact for a standard normal random variable. The analogous property for the extended Laguerre polynomial series is given below.

Theorem 4.8. *If X is a gamma $(\alpha+1,\beta)$ random variable, then each coefficient*

$$d_k = \frac{1}{k!(\alpha+1)_k}\sum_{m=0}^{k}\binom{k}{m}\frac{(-1)^m\Gamma(\alpha+1+k)}{\Gamma(\alpha+1+k-m)}\beta^{k-m}\mu_{k-m}, \quad k = 1,2,\ldots,n \qquad [4.54]$$

in the representation

$$F(x) \simeq P(\alpha+1,\beta x) - w(x)\sum_{k=1}^{n} d_k L_{k-1}^{(\alpha+1,\beta)}(x) \qquad [4.55]$$

is zero, and the approximation is exact for all x.

Proof.

The moments of a gamma $(\alpha+1,\beta)$ random variable are found from Larsen and Marx [1985] to be

$$\mu_{k-m} = \frac{\Gamma(\alpha+1+k-m)}{\beta^{k-m}\Gamma(\alpha+1)} = \frac{(\alpha+1)_{k-m}}{\beta^{k-m}}. \qquad [4.56]$$

Using this in [4.54] gives

$$d_k = \frac{1}{k!(\alpha+1)_k} \sum_{m=0}^{k} \binom{k}{m} (-1)^m \frac{\Gamma(\alpha+1+k)}{\Gamma(\alpha+1+k-m)} (\alpha+1)_{k-m}. \qquad [4.57]$$

Converting the Pochhammer notation into gamma functions and taking the resulting constant out of the summation yields

$$d_k = \frac{1}{k!(\alpha+1)_k} \sum_{m=0}^{k} \binom{k}{m} (-1)^m \frac{\Gamma(\alpha+1+k)\Gamma(\alpha+1+k-m)}{\Gamma(\alpha+1+k-m)\Gamma(\alpha+1)}$$

$$d_k = \frac{1}{k!(\alpha+1)_k} \sum_{m=0}^{k} \binom{k}{m} (-1)^m \frac{\Gamma(\alpha+1+k)}{\Gamma(\alpha+1)} \qquad [4.58]$$

$$= \frac{1}{k!} \sum_{m=0}^{k} \binom{k}{m} (-1)^m.$$

The binomial sum is zero, the sum in [4.55] vanishes, and the approximation is exact. ■

4.3 The Jacobi Polynomials

The Jacobi polynomials are a generalization of many better-known orthogonal polynomials, including the Legendre, Chebyshev Types I and II, and the ultraspherical or Gegenbauer polynomials. Although the Jacobi polynomials are orthogonal on $[-1, 1]$, they are often used to represent functions on any finite interval through the use of a linear transformation. We use such a transformation to the interval $[0, 1]$, an important special case for probabilistic applications. Jacobi [1859] appears to be the earliest publication concerning this case.

4.3.1 Definition and Rodriguez' Formula

The Jacobi polynomials are defined in several references; we use a definition that differs from that of Abramowitz and Stegun [1964] only by a constant factor.

Definition. The kth degree Jacobi polynomial with parameters β and α is given by

$$J_k^{(\beta,\alpha)}(x) = \sum_{m=0}^{k} \binom{k+\beta}{m}\binom{k+\alpha}{k-m}(x-1)^{k-m}(x+1)^m, \quad -1 < x < 1. \qquad [4.59]$$

Although the conventional weighting function for the Jacobi polynomials is

$$w(x) = (1-x)^\beta(1+x)^\alpha, \qquad [4.60]$$

a weighting function for the transformed Jacobi polynomials is of more use to our development. The parameters in [4.59] and [4.60] are identical, and from Theorem 3.1 we have the constraints $\alpha, \beta > -1$.

While most of the classical results for the Jacobi polynomials are derived for the interval $[-1,1]$, our development favors the interval $[0,1]$. Consequently, Rodriguez' formula for the interval $[-1,1]$ is presented in Theorem 4.9 and our adaptation for the interval $[0,1]$ follows as Corollary 4.10.

Theorem 4.9. *Rodriguez' formula for the Jacobi polynomials on the interval* $[-1,1]$ *(see [4.59]) is given by*

$$\begin{aligned} J_k^{(\beta,\alpha)}(x) &= \frac{l_k}{w(x)}\frac{d^k}{dx^k}[w(x)(q(x))^k] \\ &= \frac{(-1)^k/k!}{(1-x)^\beta(1+x)^\alpha}\frac{d^k}{dx^k}\left[(1-x)^\beta(1+x)^\alpha(1-x^2)^k\right] \qquad [4.61] \\ &= \frac{(-1)^k/k!}{(1-x)^\beta(1+x)^\alpha}\frac{d^k}{dx^k}\left[(1-x)^{k+\beta}(1+x)^{k+\alpha}\right]. \end{aligned}$$

Proof.

Using the Pochhammer notation [4.26], it follows from [4.59] that

$$
\begin{aligned}
&(1-x)^{\beta}(1+x)^{\alpha}J_k^{(\beta,\alpha)}(x) \\
&= \sum_{m=0}^{k}\binom{k+\beta}{m}\binom{k+\alpha}{k-m}(-1)^{k-m}(1-x)^{k-m+\beta}(1+x)^{m+\alpha} \\
&= (-1)^k\sum_{m=0}^{k}\frac{\Gamma(k+\beta+1)\Gamma(k+\alpha+1)(-1)^m(1-x)^{k-m+\beta}(1+x)^{m+\alpha}}{\Gamma(m+1)\Gamma(k+\beta-m+1)\Gamma(k-m+1)\Gamma(m+\alpha+1)} \\
&= \frac{(-1)^k}{k!}\sum_{m=0}^{k}\binom{k}{m}(k+\beta-m+1)_m(1-x)^{k-m+\beta}(-1)^m(m+\alpha+1)_{k-m}(1+x)^{m+\alpha} \\
&= \frac{(-1)^k}{k!}\sum_{m=0}^{k}\binom{k}{m}\frac{d^m}{dx^m}\left[(1-x)^{k+\beta}\right]\frac{d^{k-m}}{dx^{k-m}}\left[(1+x)^{k+\alpha}\right].
\end{aligned}
\qquad [4.62]
$$

Applying Leibniz' rule for differentiation of a product (see Appendix C) leads to the last equation of [4.61]. ■

To adapt this result to the interval $[0,1]$, we use the linear transformation

$$y = \frac{x+1}{2}. \qquad [4.63]$$

This change of variable leads to the following corollary.

Corollary 4.10. *If* $0 \le y \le 1$, *the Jacobi polynomials satisfy*

$$J_k^{(\beta,\alpha)}(2y-1) = \frac{(-2)^k/k!}{(1-y)^{\beta}y^{\alpha}}\frac{d^k}{dy^k}\left[(1-y)^{k+\beta}y^{k+\alpha}\right]. \qquad [4.64]$$

Proof.

With the change of variable given by [4.63], we have $\frac{dy}{dx} = \frac{1}{2}$. By repeatedly applying the chain rule to [4.61],

$$J_k^{(\beta,\alpha)}(2y-1) = \frac{(-1)^k/k!}{(1-y)^{\beta}y^{\alpha}2^{\alpha+\beta}}\frac{d^k}{dy^k}\left[(1-y)^{k+\beta}y^{k+\alpha}2^{2k+\alpha+\beta}\right]\left(\frac{1}{2}\right)^k. \qquad [4.65]$$

Collecting the factors of two then yields [4.64]. ■

4.3.2 Orthogonality and Norm

As discussed in Szegö [1975], the orthogonality relation for the Jacobi polynomials is

$$h_{m,n}^2 = \int_{-1}^{1} (1-x)^\beta (1+x)^\alpha J_m^{(\beta,\alpha)}(x) J_n^{(\beta,\alpha)}(x)\, dx$$

$$= \begin{cases} 0, & \text{for } m \neq n; \\ \dfrac{2^{\alpha+\beta+1}\Gamma(\alpha+1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+2)}, & \text{for } m = n = 0; \\ \dfrac{2^{2n+\alpha+\beta+1}\Gamma(n+\beta+1)\Gamma(n+\alpha+1)}{(2n+\alpha+\beta+1)n!\Gamma(n+\alpha+\beta+1)}, & \text{for } m = n \geq 1. \end{cases} \quad [4.66]$$

To relate the Jacobi polynomials to the beta distribution requires the previously discussed transformation to the interval $[0,1]$. The theorem below supplies the corresponding orthogonality relation.

Theorem 4.11. *Upon transformation from* $[-1,1]$ *to* $[0,1]$, *the Jacobi polynomials satisfy*

$$k_{m,n}^2 = \int_0^1 \frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+1)\Gamma(\beta+1)} y^\alpha (1-y)^\beta J_m^{(\beta,\alpha)}(2y-1) J_n^{(\beta,\alpha)}(2y-1)\, dy$$

$$= \begin{cases} 0, & \text{for } m \neq n; \\ 1, & \text{for } m = n = 0; \\ \dfrac{2^{2n}(\alpha+1)_n(\beta+1)_n}{(2n+\alpha+\beta+1)n!(\alpha+\beta+2)_{n-1}}, & \text{for } m = n \geq 1. \end{cases} \quad [4.67]$$

Proof.

We again make the linear transformation from $[-1,1]$ to $[0,1]$ with the substitution defined by [4.63]. The integral in [4.66] then becomes

$$h_{m,n}^2 = \int_0^1 2^{\alpha+\beta+1} y^\alpha (1-y)^\beta J_m^{(\beta,\alpha)}(2y-1) J_n^{(\beta,\alpha)}(2y-1) dy$$

$$= \begin{cases} 0, & \text{for } m \neq n; \\ \dfrac{2^{\alpha+\beta+1}\Gamma(\alpha+1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+2)}, & \text{for } m = n = 0; \\ \dfrac{2^{2n+\alpha+\beta+1}\Gamma(n+\beta+1)\Gamma(n+\alpha+1)}{(2n+\alpha+\beta+1)n!\Gamma(n+\alpha+\beta+1)}, & \text{for } m = n \geq 1. \end{cases} \quad [4.68]$$

It is clear from the above integral that $k_{m,n}^2 = \frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+1)\Gamma(\beta+1)} \left(\frac{1}{2}\right)^{\alpha+\beta+1} h_{m,n}^2$, and the first two cases of [4.67] follow easily from those of [4.68]. Thus, we consider the case $m = n \geq 1$. We introduce the beta function constant and divide by $2^{\alpha+\beta+1}$ to yield

$$\begin{aligned} &\int_0^1 \frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+1)\Gamma(\beta+1)} y^\alpha (1-y)^\beta J_m^{(\beta,\alpha)}(2y-1) J_n^{(\beta,\alpha)}(2y-1) dy \\ &= \frac{2^{2n}\Gamma(n+\beta+1)\Gamma(n+\alpha+1)\Gamma(\alpha+\beta+2)}{(2n+\alpha+\beta+1)n!\Gamma(n+\alpha+\beta+1)\Gamma(\alpha+1)\Gamma(\beta+1)}, \\ &= \frac{2^{2n}(\alpha+1)_n(\beta+1)_n}{(2n+\alpha+\beta+1)n!(\alpha+\beta+2)_{n-1}}, \end{aligned} \qquad [4.69]$$

where the last part follows from three applications of the Pochhammer identity [4.26]. ■

From examining [4.69], we see that the weighting function for the Jacobi polynomials transformed to $[0,1]$ is given by

$$w(y) = \frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+1)\Gamma(\beta+1)} y^\alpha (1-y)^\beta = \frac{y^\alpha (1-y)^\beta}{B(\alpha+1,\beta+1)}, \qquad [4.70]$$

which we recognize as the probability density function of a beta$(\alpha+1,\beta+1)$ random variable.

4.3.3 Probability Density Function Series Representation

The Jacobi polynomial series expansion for a probability density function is more complicated than the Hermite or Laguerre cases because the transformation to the interval $[0, 1]$ is required in order to retain the beta probability density function as the weighting function. That is, to use the Jacobi polynomials in a series expansion with the beta weighting function for a continuous random variable Z defined on a finite interval $[a, b]$, we must transform Z to the interval $[0, 1]$. This is accomplished by defining the random variable Y as

$$Y = \frac{Z - a}{b - a}. \qquad [4.71]$$

Using the beta weighting function given above, the series representation of a probability density function $f(y)$ is

$$f(y) = \sum_{k=0}^{\infty} d_k J_k^{(\beta,\alpha)}(2y - 1)w(y). \qquad [4.72]$$

From Theorem [3.5], [4.67], and the transformation to $[0, 1]$,

$$d_k = \frac{(2k + \alpha + \beta + 1)k!(\alpha + \beta + 2)_{k-1}}{2^{2k}(\alpha + 1)_k(\beta + 1)_k} E[J_k^{(\beta,\alpha)}(2Y - 1)], \quad k = 0, 1, \ldots, \qquad [4.73]$$

which involves the moments of Y, and consequently those of Z. From a practical standpoint, this formula is inconvenient to evaluate when used within the definition [4.59] of the Jacobi polynomials because of the presence of a product of powers of Y and $Y - 1$ after the transformation. Fortunately, there is an alternate form of the Jacobi polynomials based on powers of $x - 1$ which, when transformed, is in powers of $Y - 1$. The following equivalent form of the Jacobi polynomials is from Abramowitz and Stegun [1964]:

$$J_k^{(\beta,\alpha)}(x) = \frac{2^k\Gamma(\beta + k + 1)}{k!\Gamma(\alpha + \beta + k + 1)} \sum_{m=0}^{k} \binom{k}{m} \frac{\Gamma(\alpha + \beta + k + m + 1)}{2^m\Gamma(\beta + m + 1)}(x - 1)^m. \qquad [4.74]$$

The transformation to $[0, 1]$ with the substitution [4.63] yields

$$J_k^{(\beta,\alpha)}(2y - 1) = \frac{2^k\Gamma(\beta + k + 1)}{k!\Gamma(\alpha + \beta + k + 1)} \sum_{m=0}^{k} \binom{k}{m} \frac{\Gamma(\alpha + \beta + k + m + 1)}{\Gamma(\beta + m + 1)}(y - 1)^m. \qquad [4.75]$$

Using the Pochhammer notation, we find from [4.73] and [4.75] an expression for the coefficients d_k in terms of the moments of $Y-1$, where $\mu_m^* = E[(Y-1)^m]$:

$$\begin{aligned} d_k &= \frac{(2k+\alpha+\beta+1)(\alpha+\beta+2)_{k-1}\Gamma(\beta+k+1)}{2^k(\alpha+1)_k(\beta+1)_k\Gamma(\alpha+\beta+k+1)} \sum_{m=0}^{k} \binom{k}{m} \frac{\Gamma(\alpha+\beta+k+m+1)}{\Gamma(\beta+m+1)} \mu_m^* \\ &= \frac{(2k+\alpha+\beta+1)(\alpha+\beta+2)_{k-1}}{2^k(\alpha+1)_k} \sum_{m=0}^{k} \binom{k}{m} \frac{\Gamma(\alpha+\beta+k+m+1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+k+1)\Gamma(\beta+m+1)} \mu_m^* \\ &= \frac{(2k+\alpha+\beta+1)(\alpha+\beta+2)_{k-1}}{2^k(\alpha+1)_k} \sum_{m=0}^{k} \binom{k}{m} \frac{(\alpha+\beta+k+1)_m}{(\beta+1)_m} \mu_m^*. \end{aligned} \tag{4.76}$$

The moments of $Y-1$ can be determined from the moments of Y, and the series [4.72] can be computed for as many terms as desired.

4.3.4 Cumulative Distribution Function

The determination of the coefficients d_k and the transformed weighting function $w(y)$ given by [4.70] allow us to find an expression for the cumulative distribution function of a continuous random variable defined on $[0,1]$. This is formalized in the following theorem.

Theorem 4.12. *If the probability density function $f(y)$ of a continuous random variable Y defined almost everywhere in $[0,1]$ has the Jacobi polynomial series representation*

$$f(y) = w(y) \sum_{k=0}^{\infty} d_k J_k^{(\beta,\alpha)}(2y-1), \tag{4.77}$$

where $w(y)$ is the probability density function for a beta $(\alpha+1, \beta+1)$ random variable and

$$d_k = \frac{(2k+\alpha+\beta+1)k!(\alpha+\beta+2)_{k-1}}{2^{2k}(\alpha+1)_k(\beta+1)_k} E[J_k^{(\beta,\alpha)}(2Y-1)], \quad k=0,1,\ldots, \tag{4.78}$$

then the cumulative distribution function of Y is determined by

$$F(y) = I_y(\alpha+1,\beta+1) - \frac{2(1-y)^{\beta+1}y^{\alpha+1}}{B(\alpha+1,\beta+1)} \sum_{k=1}^{\infty} \frac{d_k J_{k-1}^{(\beta+1,\alpha+1)}(2y-1)}{k}, \tag{4.79}$$

where the standard definition of the incomplete beta function $I_y(\alpha+1,\beta+1)$ *is given by Abramowitz and Stegun [1964] for* $\alpha,\beta > -1$ *as*

$$I_y(\alpha+1,\beta+1) = \frac{1}{B(\alpha+1,\beta+1)}\int_0^y (1-t)^\beta t^\alpha dt, \quad 0 \le y \le 1. \qquad [4.80]$$

Proof.

Since Y is defined on $[0,1]$, Theorem 3.8 yields

$$F(y) = l_0 d_0 \int_0^y w(t)dt + \sum_{k=1}^{\infty} d_k l_k \left[\frac{d^{k-1}}{dt^{k-1}}[w(t)(q(t))^k]\Big|_0^y\right]. \qquad [4.81]$$

We have $d_0 = 1$ from [4.76] and from the transformed Rodriguez' formula [4.64] we have $l_k = \frac{(-2)^k}{k!}$ and $q(t) = t(1-t)$. Using the beta weighting function we thus have

$$\begin{aligned} F(y) = &\int_0^y \frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+1)\Gamma(\beta+1)} y^\alpha (1-y)^\beta dy \\ &+ \sum_{k=1}^{\infty} \frac{d_k(-2)^k}{k!}\left[\frac{d^{k-1}}{dt^{k-1}}\frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+1)\Gamma(\beta+1)} t^\alpha (1-t)^\beta t^k (1-t)^k\Big|_0^y\right]. \end{aligned} \qquad [4.82]$$

Again referring to Rodriguez' formula [4.64] for the transformed system, it is apparent that

$$\frac{d^{k-1}}{dt^{k-1}}\left[(1-t)^{k+\beta}t^{k+\alpha}\right] = \frac{J_{k-1}^{(\beta+1,\alpha+1)}(2t-1)(k-1)!(1-t)^{\beta+1}t^{\alpha+1}}{(-2)^{k-1}}. \qquad [4.83]$$

Using this in [4.82] along with the incomplete beta notation yields

$$F(y) = I_y(\alpha+1,\beta+1) - \frac{2}{B(\alpha+1,\beta+1)}\sum_{k=1}^{\infty}\frac{d_k J_{k-1}^{(\beta+1,\alpha+1)}(2t-1)(1-t)^{\beta+1}t^{\alpha+1}}{k}\Big|_0^y, \qquad [4.84]$$

which, when evaluated, agrees with [4.79]. ■

4.3.5 Relationship to Beta Distribution

We have seen that the weighting function for the transformed Jacobi polynomials is the beta probability density function. Thus, the relationships of the Hermite polynomials to the normal distribution and the extended Laguerre polynomials to the gamma distribution suggest the Jacobi polynomials may have a corresponding relationship to the beta distribution. The following theorem verifies this conjecture.

Theorem 4.13. *If Y is a beta $(\alpha+1,\beta+1)$ random variable with probability density function $f(y)$, then each coefficient*

$$d_k = \frac{(2k+\alpha+\beta+1)k!(\alpha+\beta+2)_{k-1}}{2^{2k}(\alpha+1)_k(\beta+1)_k} E[J_k^{(\beta,\alpha)}(2Y-1)], \quad k=1,2,\ldots,n \qquad [4.85]$$

in the representation

$$F(y) \simeq I_y(\alpha+1,\beta+1) - \frac{2(1-y)^{\beta+1}y^{\alpha+1}}{B(\alpha+1,\beta+1)} \sum_{k=1}^{n} \frac{d_k J_{k-1}^{(\beta+1,\alpha+1)}(2y-1)}{k}, \qquad [4.86]$$

is zero, and the approximation is exact for all y.

Proof.

From [4.85], we have

$$d_k = \frac{(2k+\alpha+\beta+1)k!(\alpha+\beta+2)_{k-1}}{2^{2k}(\alpha+1)_k(\beta+1)_k} \int_0^1 f(y) J_k^{(\beta,\alpha)}(2y-1)dy. \qquad [4.87]$$

Using the transformation [4.63], the definition [4.59] of Jacobi polynomials becomes

$$J_k^{(\beta,\alpha)}(2y-1) = \sum_{m=0}^{k} \binom{k+\beta}{m}\binom{k+\alpha}{k-m} 2^k(-1)^{k-m}(1-y)^{k-m}y^m. \qquad [4.88]$$

Letting C_k be the constant factor from [4.87],

$$\begin{aligned} d_k &= C_k \int_0^1 f(y) \left[\sum_{m=0}^{k} \binom{k+\beta}{m}\binom{k+\alpha}{k-m} 2^k(-1)^{k-m}(1-y)^{k-m}y^m\right] dy \\ &= C_k 2^k \sum_{m=0}^{k} \binom{k+\beta}{m}\binom{k+\alpha}{k-m}(-1)^{k-m} \int_0^1 f(y)(1-y)^{k-m}y^m dy. \end{aligned} \qquad [4.89]$$

Because Y has a beta$(\alpha+1,\beta+1)$ distribution,

$$f(y) = \frac{1}{B(\alpha+1,\beta+1)}(1-y)^{\beta}y^{\alpha}, \qquad [4.90]$$

and the integral from [4.89] is a beta function:

$$\begin{aligned}\int_0^1 \frac{1}{B(\alpha+1,\beta+1)}(1-y)^{k+\beta-m}y^{m+\alpha}dy &= \frac{B(m+\alpha+1,k+\beta-m+1)}{B(\alpha+1,\beta+1)} \\ &= \frac{\Gamma(m+\alpha+1)\Gamma(k+\beta-m+1)}{B(\alpha+1,\beta+1)\Gamma(\alpha+k+\beta+2)},\end{aligned} \qquad [4.91]$$

where the last expression comes from the identity

$$B(\alpha+1,\beta+1) = \frac{\Gamma(\alpha+1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+2)}. \qquad [4.92]$$

From the combinatorial identity

$$\binom{n}{r} = \frac{\Gamma(n+1)}{\Gamma(r+1)\Gamma(n-r+1)}, \qquad [4.93]$$

we convert the binomial coefficient terms from [4.89] into gamma functions, yielding

$$d_k = \frac{C_k(-2)^k\Gamma(k+\beta+1)\Gamma(k+\alpha+1)}{B(\alpha+1,\beta+1)\Gamma(\alpha+k+\beta+2)}\sum_{m=0}^{k}\frac{(-1)^{-m}}{\Gamma(m+1)\Gamma(k-m+1)}. \qquad [4.94]$$

By absorbing the constant factors and changing the inner gamma terms back to a binomial coefficient,

$$d_k = C_k'\sum_{m=0}^{k}\frac{(-1)^{-m}}{\Gamma(k+1)}\binom{k}{m} = \frac{C_k'}{\Gamma(k+1)}\sum_{m=0}^{k}(-1)^m\binom{k}{m}. \qquad [4.95]$$

The binomial sum is zero, and hence the result. ∎

CHAPTER 5

IMPLEMENTATION

5.0 Introduction

In this chapter, we illustrate the mechanics of the proposed method and present evidence that the approximations can work well in practice, while providing practical insight into the general accuracy of the approximations. We make no claim that this is an exhaustive study of the orthogonal polynomial series approximations, but rather present the applications as empirical support for their use.

The random variables we consider are simple functions of independent random variables. While the individual component random variables are simple, well-known, and often used in probabilistic models, the resulting random variable is difficult to evaluate. We begin with the presentation of a procedure that can be used to estimate the cumulative distribution function of a general random variable using orthogonal polynomials. Following this presentation, we return to the traffic conflict model of Chapter 1. The model is fully described, and the extended Laguerre series is used to develop a procedure for estimating the effectiveness of an improvement. Following the traffic conflict results, we use our method to estimate the distributions of functions of various random variables.

5.1 Procedure for Applying an Orthogonal Polynomial Series

In order to use the proposed methodology, one must decide which set of orthogonal polynomials to use and which parameter values to use in the weighting function. In the selection of the polynomials, we suggest the use of a simple and natural correspondence between the domain of the random variable and the interval of orthogonality of the orthogonal polynomial family. In what follows, we assume that the domain of the random variable is known and that its moments exist and can be computed.

For a random variable which can assume values in the interval $(-\infty, \infty)$, we are constrained to choose the Hermite polynomials because the coefficients d_k that define the series approximation are determined by the integral in [3.29], which will not be known in general, but can be evaluated (by Corollary 3.6) from the moments if the interval of orthogonality is $(-\infty, \infty)$, as is the case for the Hermite polynomials.

When the random variable has the domain $[a, \infty)$ or $(-\infty, a]$, it can be easily transformed to a random variable on the interval $[0, \infty)$. Although the coefficients in [3.29] can then be determined for either the Hermite or the extended Laguerre polynomials, we suggest the choice of the extended Laguerre polynomials, since their interval of orthogonality is also $[0, \infty)$. As supporting evidence for this case, consider the special case of approximating the cumulative distribution function of a gamma random variable. From Theorem 4.8, we know the extended Laguerre series approximation is exact, while the Hermite series will in general have some error.

Finally, a random variable defined on a finite interval $[a, b]$ can be approximated using any of the three orthogonal polynomial families. Of course the Jacobi polynomials are orthogonal on a finite interval and thus yield a choice consistent with the above discussion.

After the polynomial family has been chosen, the parameters for the corresponding weighting function must be determined. This applies only to the extended Laguerre and Jacobi polynomials, which, as can be seen from their respective definitions [4.22] and [4.59], each require two parameters, while the Hermite polynomials (see [4.1]) do not require any.

There is a choice for these parameters that is consistent with the literature when appropriately interpreted. When applying the Cornish-Fisher series (see Appendix B for a derivation of this variant of the Gram-Charlier Type A series) to a random variable Y with mean μ and variance σ^2, it is convenient to standardize Y by defining the random variable $X = \frac{Y-\mu}{\sigma}$ and apply the series approximation to X, which has a mean of zero and a variance of one. These correspond to the first two moments of the random variable associated with the weighting function for the Hermite polynomials, which is a standard normal random variable with a mean of zero and a variance of one. When approximating the distribution of a random variable X with mean μ and variance σ^2 with the extended Laguerre polynomial series, an analogous procedure is to choose the parameters α and β so that the gamma random variable associated with the extended Laguerre weighting function has mean μ and variance σ^2. Similarly, for the Jacobi polynomial series, we choose the Jacobi parameters α and β so that the beta random variable associated with the Jacobi weighting function has the same mean and variance as the random variable X which we wish to approximate.

There is an intuitive appeal to this procedure. Referring to the cumulative distribution function representations [4.9], [4.47], and [4.79] of the Hermite, extended Laguerre, and Jacobi polynomials, we note that each is the difference of two terms. The first term is the cumulative distribution function of an associated random variable as discussed above, and the second is a sum of the terms of a series involving the product of the coefficients $\{d_k\}$ and the orthogonal polynomials. Thus, we can think of our approximation of the cumulative distribution function of the random variable X as a series of 'adjustments' (based on the moments of X, since the d_k are determined by the moments) from the cumulative distribution function of a standard normal, gamma, or beta random variable in the Hermite, extended Laguerre, or Jacobi cases, respectively, which has mean and variance identical to those of X. This is reinforced by Theorems 4.3, 4.8, 4.13, and Corollary 4.4, which reveal that no adjustments are required ($d_k = 0$ for all k) if X is a normal, gamma, or beta random variable.

Once the polynomial family and weighting function parameter values have been determined, the approximation of the cumulative distribution function of X requires the evaluation of the first term of [4.9], [4.47], or [4.79], and a truncation of the corresponding series. The computation of the first term is easily accomplished using either existing code or algorithms. The cumulative distribution function for a standard normal random variable can be computed to nine decimal places for $x \leq 7$ with a simple formula provided by Kendall [1987], who attributes the result to Moran [1980]. Accurate FORTRAN routines for computing the incomplete gamma and incomplete beta functions are given in Press et al. [1989].

Determining the number of moments to use in the summation term is a difficult matter. The literature gives few clues as to this choice. Hill [1969] uses eight moments for the Laguerre series expansion, but no formal justification is provided. Kendall [1987] demonstrates the use of the Gram-Charlier Type A series with four, five, and six moments, and derives the Cornish-Fisher series up to the sixth cumulant. Additionally, Kendall [1987] (citing Barton and Dennis [1952] and Draper and Tierney [1973]) indicates that the Gram-Charlier Type A series and the Edgeworth form of the Type A series (see Appendix B) can result in negative values when the fifth and higher moments are ignored. Hill [1969] provides an example in which the Laguerre series (using eight moments) yields values greater than one. These problems can also occur with the Cornish-Fisher and extended Laguerre series. Such a case can be seen graphically in Figure 2 of §5.3.1.

5.2 The Traffic Conflicts Model

Recalling the traffic conflict model of Chapter 1, we now turn to developing a procedure for estimating probabilities associated with the accident rate at an intersection both before and after improvements have been made to the intersection. The evaluation of the effectiveness of an improvement program based on such a 'before and after' scenario necessarily requires two sets of information. The first set of information provides an understanding of the accident rate before the improvement, and the second set provides an understanding of the accident rate after the improvement. The evaluation technique described here is built around the procedure for identifying hazardous locations proposed by Higle and Witkowski [1988]. In this paper, the authors describe a procedure for identifying a so called 'refined estimate' of the distribution of the accident rate before the improvements are implemented. This distribution is then used to assess the safety of the intersection by calculating the probability that the pre-improvement accident rate exceeds some bound deemed appropriate for the particular intersection. For details on two such bounds, see Higle and Witkowski [1988]. If this probability is greater than some specified value δ (typical values of δ are 0.90 or 0.95), then the site is identified as hazardous.

For the purposes of the example to be presented, a complete conflict model for a region of sites is unnecessary. We therefore concentrate on a one-site case, where the goal is to determine whether improvements made to increase the safety of the intersection are effective. We assume the pre-improvement accident rate distribution is known (e.g., by the method of Higle and Witkowski [1988]) and that the intersection was originally identified as hazardous. We also assume that by use of a similar technique using conflict data collected before the improvements are made, the conflict rate distribution and the distribution of the probability $\tilde{P}$ are known. These distributions are the prior distributions of the model. After the improvements are completed, posterior distributions will be found from these priors using conflict and accident data from the site. This process is described below.

We model the conflict rate $\tilde{\lambda}_c$ with a gamma distribution. From a theoretical standpoint, there are precedents with accident models (as well as the conflict model of Hauer [1978]) to support this choice. From a practical viewpoint, the gamma

distribution is convenient because it belongs to a conjugate family. In the Bayesian framework this allows us to easily compute the posterior distribution. For example, if conflicts occur according to a Poisson distribution with mean $\tilde{\lambda}_c V_1$, where $\tilde{\lambda}_c$ has a gamma (a, b) prior distribution and V_1 is the traffic volume of the intersection, then it is known that the posterior distribution of $\tilde{\lambda}_c$ is also a gamma distribution with parameters a' and b' determined by

$$\begin{aligned} a' &= a + C \\ b' &= b + V_2, \end{aligned} \qquad [5.1]$$

where C is the number of post-improvement conflicts and V_2 is the new traffic volume.

An analogous result is applicable to the random variable $\tilde{P}$, which represents the (unknown) probability that a conflict results in an accident. If $\tilde{P}$ has a beta prior distribution with parameters c and d and the number of post-improvement accidents is N, the posterior distribution of $\tilde{P}$ is a beta distribution with parameters c' and d' given by

$$\begin{aligned} c' &= c + N \\ d' &= d + C - N. \end{aligned} \qquad [5.2]$$

We have the following quantities in the model:

$\tilde{\lambda}_a$ = accident rate (a gamma random variable)

$\tilde{\lambda}_c$ = conflict rate (a gamma random variable)

$\tilde{P}$ = the probability that a conflict becomes an accident (a beta random variable)

N = the number of accidents occurring after improvements are made

C = the number of conflicts occurring after improvements are made

V_1 = the traffic volume before improvements are made

V_2 = the traffic volume after improvements are made

$\bar{\lambda}$ = the upper bound on the 'safe' accident rate

To assess the post-improvement accident rate of the site, we use the equation $\tilde{\lambda}_a = \tilde{P}\tilde{\lambda}_c$ which relates conflicts to accidents. Using Mellin transforms, we can find the moments of $\tilde{\lambda}_a$ and then apply an orthogonal polynomial series to approximate

$$P\{\tilde{\lambda}_a > \bar{\lambda}\}. \qquad [5.3]$$

If this probability falls below δ the site can be deemed nonhazardous and if it exceeds δ the site is still identified as hazardous.

5.2.1 Application of the Model

We now consider a hypothetical application of the conflict model. We consider an intersection whose pre-improvement accident rate follows a gamma(α, β) distribution with $\alpha = 56.8$ and $\beta = 44.0$. The expected value of this random variable is 1.29, which represents the mean accident rate per million vehicles. The value of $\bar{\lambda}$ is 1.0 accidents per million vehicles entering the intersection, and using the above gamma distribution, we compute

$$P\{\tilde{\lambda}_a > \bar{\lambda} = 1.0\} = 0.9641, \qquad [5.4]$$

which identifies the intersection as hazardous for the commonly used values of $\delta = 0.90$ and $\delta = 0.95$.

From a conflict study and the accident history of the site, we presume that the prior distributions for $\tilde{\lambda}_c$ and $\tilde{P}$ are known to be gamma$(400.0, 0.210)$ and beta$(46.0, 43000.0)$, respectively. After the modifications are implemented and all transient effects to the intersection have decayed, the intersection is again studied. Suppose in a ten-day study traffic volume is 52,000 vehicles daily, no accidents occur, and daily conflict counts for the site are 25, 32, 19, 37, 34, 19, 41, 22, 30, and 38. Using this data, we update the distribution of $\tilde{\lambda}_c$ by applying [5.1]:

$$\begin{aligned} a' &= 400.0 + (25 + 32 + \cdots + 38) = 697.0 \\ b' &= 0.210 + (52000)(10)(1.0e-6) = 0.730, \end{aligned} \qquad [5.5]$$

where the factor of 1.0e-6 normalizes the ten-day volume in terms of millions of vehicles. Similarly, we find from [5.2] that the posterior distribution of $\tilde{P}$ is determined by

$$\begin{aligned} c' &= 46.0 + 0 = 46.0 \\ d' &= 43000.0 + (25 + 32 + \cdots + 38) - 0 = 43297.0 \end{aligned} \qquad [5.6]$$

To approximate the post-improvement accident rate distribution, we must choose an orthogonal system. We choose the extended Laguerre polynomials for the orthogonal family since the domain of $\tilde{\lambda}_a$ is the semi-infinite interval $[0, \infty)$. To find

the moments of $\tilde{\lambda}_a$ needed for the series, we rewrite the Mellin transform [2.15] of $\tilde{\lambda}_c$ and the corresponding transform [2.16] of $\tilde{P}$ in the Pochhammer notation. This results in

$$f^*_{\lambda_c}(s) = b'^{1-s}(a')_{s-1} \qquad [5.7]$$

and

$$f^*_P(s) = \frac{(c')_{s-1}}{(c'+d')_{s-1}}. \qquad [5.8]$$

To find the Mellin transform of $\tilde{\lambda}_a = \tilde{P}\tilde{\lambda}_c$, we apply the Mellin transform product property [2.13]. This yields

$$f^*_{\lambda_a}(s) = \frac{b'^{1-s}(a')_{s-1}(c')_{s-1}}{(c'+d')_{s-1}}. \qquad [5.9]$$

The moments of $\tilde{\lambda}_a$ are then determined from the transform using the Mellin moment formula [2.19]:

$$E[\tilde{\lambda}_a^n] = \mu_n = \frac{(a')_n(c')_n}{b'^n(c'+d')_n}. \qquad [5.10]$$

Actual computation of the moments can be simplified using the property

$$(a')_n = (a'+n-1)(a')_{n-1}, \qquad [5.11]$$

which leads to the recursive formula

$$\mu_n = \frac{(a'+n-1)(c'+n-1)}{b'(c'+d'+n-1)}\,\mu_{n-1}. \qquad [5.12]$$

From [5.12], the first seven moments of $\tilde{\lambda}_a$ are calculated and given in Table 5.1 below.

Table 5.1 - Moments of $\tilde{\lambda}_a$

n	μ_n
1	1.0133
2	1.0506
3	1.1140
4	1.2076
5	1.3376
6	1.5134
7	1.7482

Employing these moments in the truncated version of the extended Laguerre series [4.47] for the cumulative distribution function, we estimate the post-improvement probability to be

$$P\{\tilde{\lambda}_a = \tilde{P}\tilde{\lambda}_c > \bar{\lambda} = 1.0\} = 0.5127, \qquad [5.13]$$

which, when compared to the prior probability of 0.9641 (see [5.4]), indicates the improvements have been successful. This probability is smaller than any of the commonly used values of δ, and thus the intersection is no longer identified as hazardous under this criterion. Other probability statements are possible. For example, the Bayesian analog to the classical confidence interval is a credible set, which represents an actual probability rather than a statement of confidence. The computation of a credible set can be accomplished using the identical extended Laguerre series from above. Some experimentation is needed to find two points such that the interval is symmetric, but this is not difficult. From the extended Laguerre series, we find a 95 % credible set on the post-improvement accident rate:

$$P\{0.736 < \tilde{\lambda}_a = \tilde{P}\tilde{\lambda}_c < 1.338\} = 0.95 \qquad [5.14]$$

Compare this to the pre-improvement statement using the prior distribution for $\tilde{\lambda}_a$:

$$P\{0.977 < \tilde{\lambda}_a < 1.647\} = 0.95 \qquad [5.15]$$

Again, the clear indication of a probable drop in the accident rate is evident.

5.2.2 Validation of the Approximation

Because the probabilities found above were calculated from a truncated extended Laguerre series for the cumulative distribution function, there is necessarily some error included in them. Our main concern is the maximum error of the approximation, and we concentrate on its quantification. An empirical distribution function obtained from simulated observations forms the basis of this effort.

If we let $x_1, x_2, \ldots, x_n$ be the order statistics for n independent realizations of the random variable $\tilde{\lambda}_a$ (i.e., $x_1 \leq x_2 \leq \cdots \leq x_n$), then the empirical distribution function $S_n(x)$ is defined as

$$S_n(x) = \begin{cases} 0, & x < x_1; \\ \frac{i}{n}, & x_i \leq x < x_{i+1}; \\ 1, & x_n < x. \end{cases} \qquad [5.16]$$

We expect that as n increases, the empirical distribution function will tend to more closely approximate the true distribution function, say $F_0(x)$. To formally quantify the difference between $S_n(x)$ and $F_0(x)$, we introduce the Kolmogorov-Smirnov statistic D_n, defined as

$$D_n = \sup_x |S_n(x) - F_0(x)|. \qquad [5.17]$$

The Kolmogorov-Smirnov statistic is intuitively appealing because it represents the largest absolute difference between the empirical distribution function and the true distribution function. It also has the property that its distribution is independent of the distribution of $\tilde{\lambda}_a$, and the Glivenko-Cantelli theorem (see Billingsley [1979]) assures us that $D_n \to 0$ with probability 1 as $n \to \infty$. Kendall and Stuart [1979] describe how it is possible, using critical values of D_n, to make a confidence statement about the entire cumulative distribution function $F_0(x)$. If d_ϵ is the critical value of D_n for a test of significance ϵ, i.e.,

$$P\{D_n = \sup_x |S_n(x) - F_0(x)| > d_\epsilon\} = \epsilon, \qquad [5.18]$$

then it follows that the confidence statement for $F_0(x)$ is given by

$$P\{S_n(x) - d_\epsilon \leq F_0(x) \leq S_n(x) + d_\epsilon, \text{ for all } x\} = 1 - \epsilon. \qquad [5.19]$$

Using the known critical value $d_{0.05} = 1.3581/\sqrt{n}$ from Kendall and Stuart [1979], we find for a sample of $n = 100{,}000$ variates that we can be 95 % confident that the resulting empirical distribution function $S_n(x)$ is within 4.295 x 10^{-3} of the true distribution function $F_0(x)$.

To describe the error of the cumulative distribution function approximation from the extended Laguerre series, which we denote by $F_L(x)$, we use the following procedure. First we generate 100,000 observations of $\tilde{\lambda}_a = \tilde{P}\tilde{\lambda}_c$ using the known distributions for $\tilde{P}$ and $\tilde{\lambda}_c$. Comparing the empirical distribution function thus created with the cumulative distribution function approximation $F_L(x)$ over the observations, we find the maximum absolute deviation between the two, say Δ_n. Formally, we have

$$|S_n(x) - F_L(x)| \leq \Delta_n, \qquad [5.20]$$

which combines with the confidence statement [5.19] to yield

$$P\{F_L(x) - (d_{0.05} + \Delta_n) \leq F_0(x) \leq F_L(x) + (d_{0.05} + \Delta_n), \text{ for all } x\} = 0.95. \qquad [5.21]$$

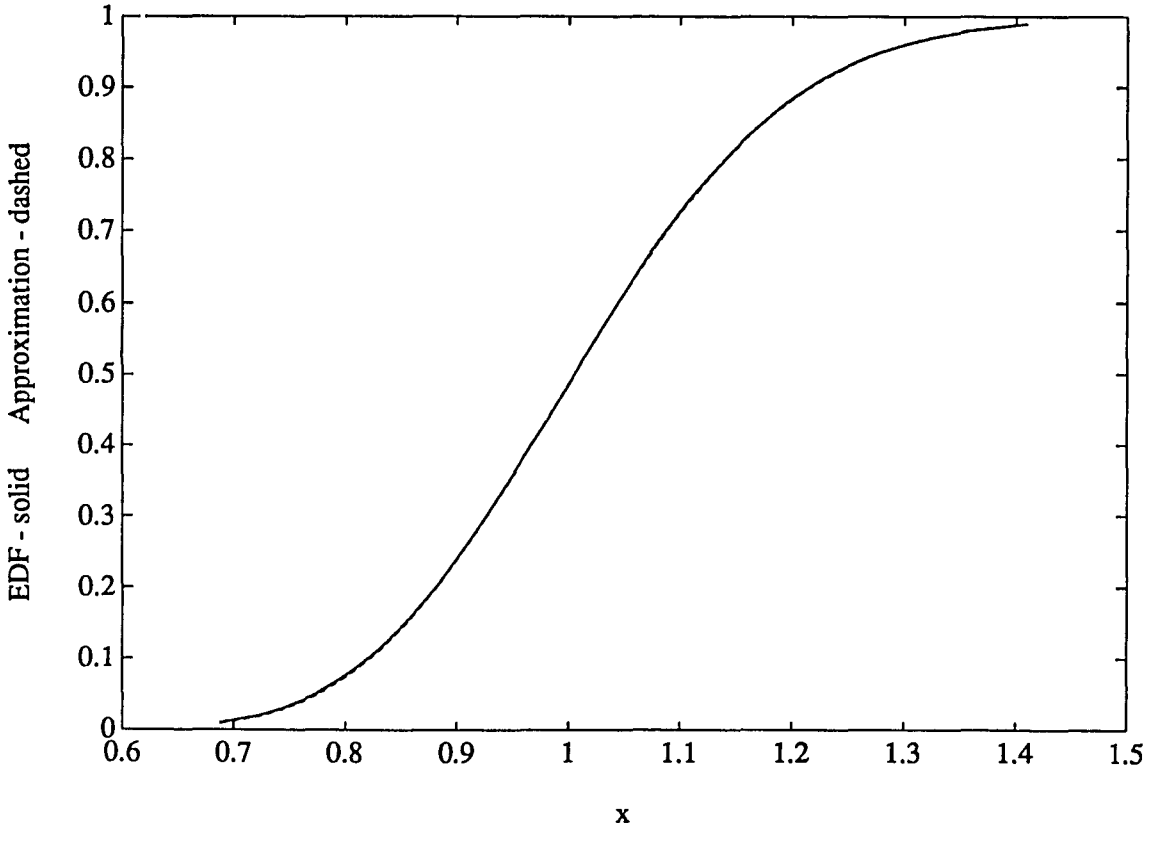
Thus we can say with 95 % confidence that our extended Laguerre polynomial series approximation does not deviate from $F_0(x)$ by more than 4.295 x $10^{-3} + \Delta_n$.

The experimental result for the example in §5.2.2 for 100,000 simulated variates is $\Delta_n = 0.0026$, which gives a probable *maximum* absolute difference from the true cumulative distribution function of 0.0069. In applied problems such as the traffic conflicts model, this is generally a reasonable error, although for probabilities close to δ, an error less than 0.005 would be desirable. A check on the accuracy of the approximation in the upper tail (95-100%) reveals a maximum difference from the empirical distribution function of 6.245 x 10^{-4}, which, when added to 4.295 x 10^{-3} gives 0.0049, just under the desirable limit. It is prudent to keep in mind that this is a worst case analysis and the actual error is often smaller. We can compute an estimate of the mean error of the extended Laguerre approximation by calculating the average absolute deviation between $S_n(x)$ and $F_L(x)$. We define this average error as

$$\bar{d}_n = \frac{1}{n}\sum_{i=1}^{n} |S_n(x_i) - F_L(x_i)|. \qquad [5.22]$$

For the experimental run described above, this average is 1.069 x 10^{-3}. We note this is less than half of the maximum absolute deviation given by the Kolmogorov-Smirnov statistic. A fuller appreciation of the accuracy of the approximation is provided in Figure 5.1 below with a graph presenting both the empirical distribution function $S_n(x)$ and the extended Laguerre series approximation $F_L(x)$. As can be seen from the figure, there is little discernible difference between the two functions within the limits of resolution.

Figure 5.1 - Graphical Comparison of $S_n(x)$ and $F_L(x)$



5.3 Empirical Results

The following subsections contain empirical results from the application of the orthogonal polynomial series expansions to functions of commonly used random variables. These examples further demonstrate the implementation of the three families of orthogonal polynomials, and provide insights as to the nature of the approximations. Error analyses are given in the form of tables and graphs.

5.3.1 The Difference of Two IID Exponential Random Variables

We first consider the task of approximating the distribution of $Z = X - Y$, where X and Y are independent and identically distributed exponential random variables. Since the domain of Z is the entire real line, we choose the Hermite polynomials and apply the series in the Cornish-Fisher form (see Appendix B) for the approximation. The error in the approximation can be quantified easily in this case because an exact formula is available for the cumulative distribution function of Z. If $X \sim \text{expo}(\beta)$ and $Y \sim \text{expo}(\beta)$, then Z is known to have a Laplace or double exponential (see Kendall [1987]) distribution with parameter β. The probability density function for Z is

$$f(z) = \tfrac{1}{2}\beta e^{-\beta|z|}, \quad -\infty < z < \infty, \qquad [5.23]$$

and the cumulative distribution function is

$$F_0(z) = \begin{cases} \tfrac{1}{2}e^{\beta z}, & z < 0; \\ 1 - \tfrac{1}{2}e^{-\beta z}, & z \geq 0. \end{cases} \qquad [5.24]$$

Thus, the need for simulation to validate the accuracy of the approximation can be eliminated for this special case.

Table 5.2 summarizes the results of an experiment with β varying from 0.1 to 10,000. In each case, the Cornish-Fisher series approximation to $F_0(z)$ using the first eight moments of Z is calculated. As in the previous example, we consider both the maximum error and the average error of the approximation. The maximum absolute deviation is defined by

$$D = \max_i \{|F_0(z_i) - F_H(z_i)|\}, \qquad [5.25]$$

where $F_H(z)$ is the Cornish-Fisher approximation, and the average absolute deviation is defined by

$$\delta_n = \frac{1}{n}\sum_{i=1}^{n} |F_0(z_i) - F_H(z_i)|. \qquad [5.26]$$

The values given in Table 5.2 were computed over $n = 120$ values of z for uniformly spaced intervals from $\mu - 3\sigma$ to $\mu + 3\sigma$.

Table 5.2 - Performance of Cornish-Fisher Approximation

β	D	δ_n
0.1	0.0411	0.0195
1.0	0.0411	0.0195
10.0	0.0411	0.0195
100.0	0.0411	0.0195
1000.0	0.0411	0.0195
10000.0	0.0411	0.0195

From the table, we see that the series seems to approximate every Laplace distribution with equal precision. This insensitivity to the choice of β can be explained through an analysis of the structure of the series approximation for the Laplace distribution. The Cornish-Fisher series is based on the standardized cumulants of the random variable in question (see Appendix B), which can be computed from the moments. For the Laplace distribution we calculate the appropriate moments using the probability density function from [5.23] and find:

$$\mu_i = E[Z^i] = \begin{cases} 0, & i \text{ odd;} \\ \dfrac{i!}{\beta^i}, & i \text{ even.} \end{cases} \qquad [5.27]$$

The first eight cumulants κ_i of the Laplace distribution are then found using the

following formulas from Kendall [1987]:

$$\begin{aligned}
\kappa_1 &= \mu_1 = 0 \\
\kappa_2 &= \mu_2 \\
\kappa_3 &= \mu_3 = 0 \\
\kappa_4 &= \mu_4 - 3\mu_2^2 \\
\kappa_5 &= \mu_5 - 10\mu_3\mu_2 = 0 \\
\kappa_6 &= \mu_6 - 15\mu_4\mu_2 - 10\mu_3^2 + 30\mu_2^3 \\
\kappa_7 &= \mu_7 - 21\mu_5\mu_2 - 35\mu_4\mu_3 + 210\mu_3\mu_2^2 = 0 \\
\kappa_8 &= \mu_8 - 28\mu_6\mu_2 - 56\mu_5\mu_3 - 35\mu_4^2 + 420\mu_4\mu_2^2 + 560\mu_3^2\mu_2 - 630\mu_2^4
\end{aligned} \qquad [5.28]$$

The results are summarized as

$$\kappa_i = \begin{cases} 0, & i = 1,3,5,7; \\ \dfrac{2(i-1)!}{\beta^i}, & i = 2,4,6,8\ . \end{cases} \qquad [5.29]$$

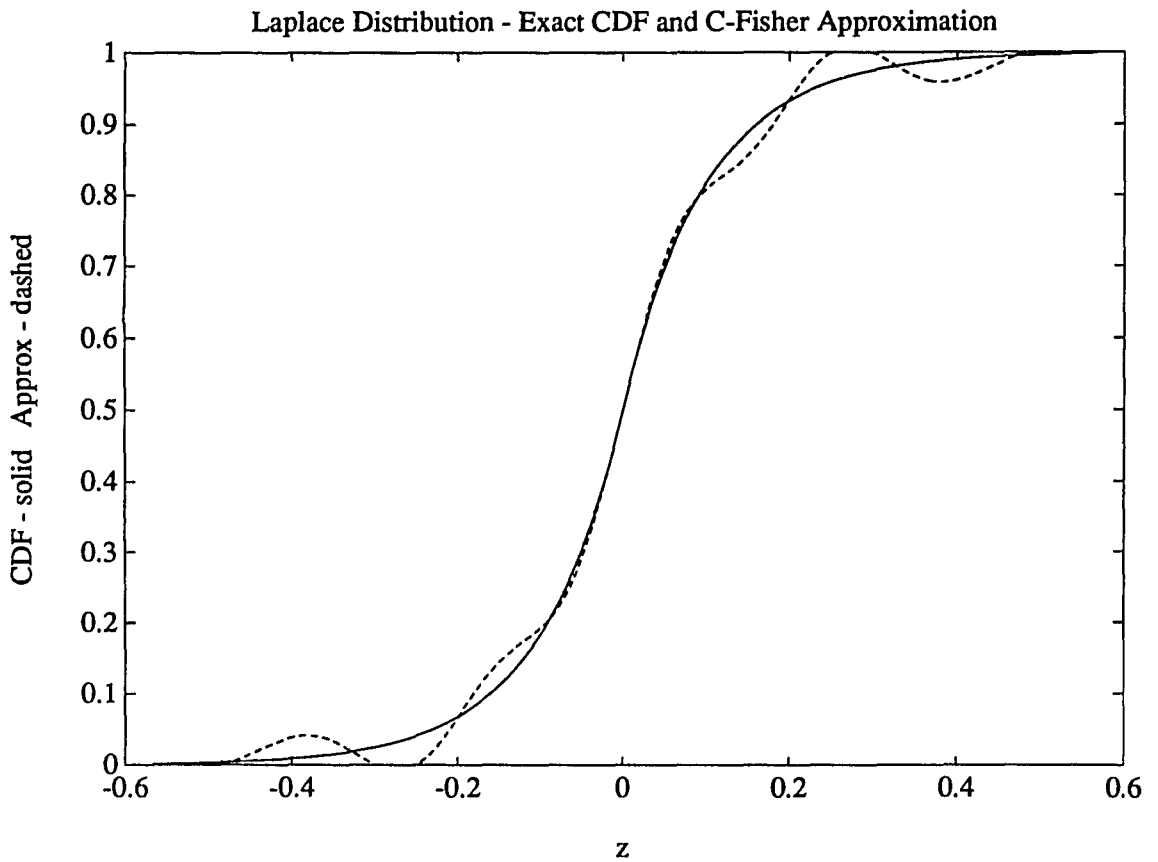
The ith standardized cumulant κ_{si} is defined to be $\kappa_{si} = \frac{\kappa_i}{\sigma^i}$. For the Laplace distribution, $\sigma = \frac{\sqrt{2}}{\beta}$, and thus

$$\kappa_{si} = \begin{cases} 0, & i = 1,3,5,7; \\ 2^{1-\frac{i}{2}}(i-1)!\ , & i = 2,4,6,8\ , \end{cases} \qquad [5.30]$$

which are independent of β. Since the κ_{si} are the coefficients for the Cornish-Fisher series, the series will always produce the results given in the table for any Laplace distribution. A graphical display of the Cornish-Fisher series approximation is displayed below in Figure 5.2 for $\beta = 10.0$. Note the symmetry of the approximation about zero - this is not unexpected in light of the nature of the Hermite polynomials and the cumulants of the Laplace distribution. No odd Hermite polynomials contribute to the approximation since the odd cumulants are zero, and the Hermite polynomials of even degree are even functions since they contain only terms of even powers (correspondingly, the Hermite polynomials of odd degree are odd functions). Thus, the Cornish-Fisher series approximation returns the same value

for x and $-x$. In fact, this property of the Cornish-Fisher series will hold whenever the random variable to be approximated has a symmetric probability density function. Note also that the series approximation can result in negative values and values greater than one. These inadequacies have been artificially adjusted here, and will be further discussed in the next chapter.

Figure 5.2 Cornish-Fisher Approximation and Exact CDF of Laplace Distribution



5.3.2 Difference of Two IID Gamma Random Variables

In this section we generalize the above example to the case of gamma(α, β) random variables. No general analytic result for the probability density function is available, so we rely on simulation to check the accuracy of the approximation.

It is known that a gamma random variable approaches a normal random variable as the shape parameter, α, increases indefinitely. Since the difference of two normal random variables is also a normal random variable, the distribution of $Z = X - Y$ where X and Y are independent and identically distributed gamma random variables can be expected to approach normality as α is increased. From Theorem 4.3, we know the Hermite series is exact for a normal random variable. Therefore, we expect the approximation of the cumulative distribution function of Z to improve as α increases. Furthermore, as with the special case in §5.3.1 above, the cumulants of the random variable Z can be shown to be independent of the parameter β. Therefore, for this study we fix $\beta = 1$. We vary α from 1.0 to 5.0 in increments of 0.5. Because the exponential distribution is a gamma distribution with $\alpha = 1.0$, we expect the results for this first case to be similar to those presented in Table 5.2. Table 5.3 provides the statistics Δ_n and $\bar{d}_n$, which are defined in an analogous manner to those in §5.2.3 as

$$\Delta_n = \sup_x |S_n(x) - F_H(x)|, \quad \bar{d}_n = \frac{1}{n}\sum_{i=1}^{n} |S_n(z_i) - F_H(z_i)|. \qquad [5.31]$$

As with the traffic conflict example, $n = 100,000$ observations of Z are generated. The Cornish-Fisher series approximation utilizes the first eight moments.

Table 5.3 - Deviations of Cornish-Fisher Approximation from $S_n(z)$

α	Δ_n	$\bar{d}_n$
1.0	0.04217	0.01179
1.5	0.01044	0.00346
2.0	0.00516	0.00148
2.5	0.00329	0.00152
3.0	0.00295	0.00102
3.5	0.00269	0.00070
4.0	0.00238	0.00075
4.5	0.00237	0.00065
5.0	0.00185	0.00048

The results agree with the theory, as the maximum and average differences do follow the general trend of decreasing as α increases. The deviations for the Laplace case $\alpha = 1$ differ slightly from those of Table 5.2 due to the use of simulated observations in this experiment. This also explains the aberrations in the average deviation column, which are of relatively small magnitude. Note that α does not need to be very large before the approximation becomes extremely accurate. Graphical comparisons of the approximation and the empirical distribution function are provided in Figure 5.3 for $\alpha = 1.5$ and Figure 5.4 for $\alpha = 3.0$.

Figure 5.3 - Comparison of Cornish-Fisher Approximation and $S_n(z)$ for $\alpha = 1.5$

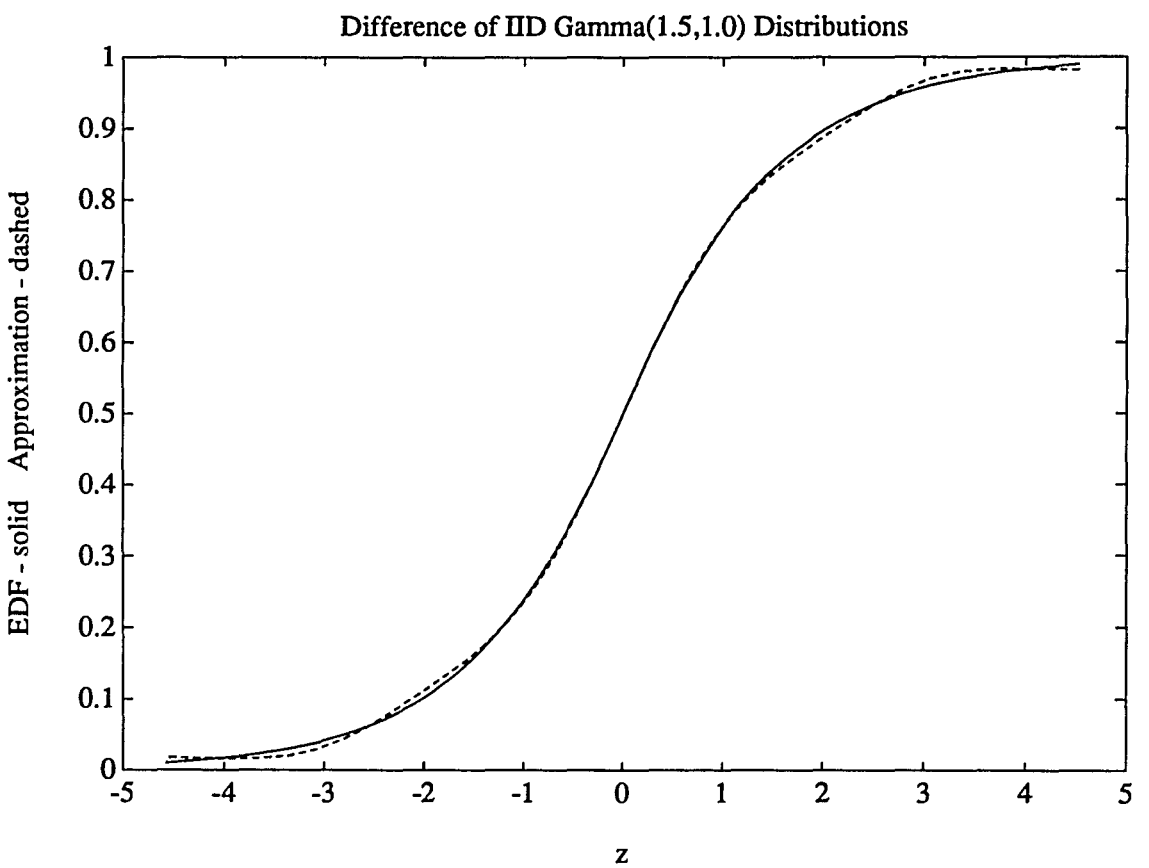
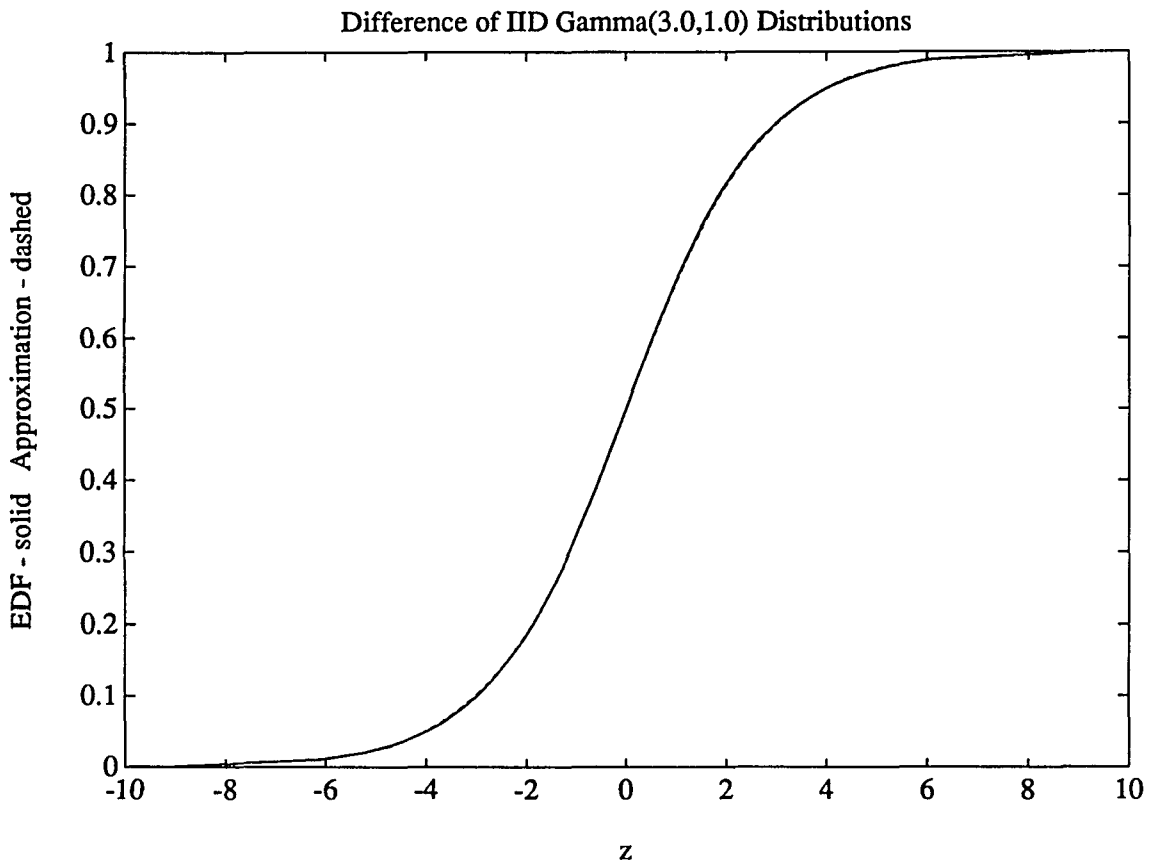


Figure 5.4 - Comparison of Cornish-Fisher Approximation and $S_n(z)$ for $\alpha = 3.0$



As expected, the approximation is symmetric about zero. Figure 5.3 shows the accuracy to be much better for $\alpha = 1.5$ as compared to the Laplace case $\alpha = 1.0$ of §5.3.1. When α reaches 3.0, it is apparent from Figure 5.4 the approximation 'hugs' the empirical distribution function so closely that the two curves are barely discernible from each other.

5.3.3 Product of a Gamma and a Beta using H-Function for Verification

As alluded to previously, the H-Function distribution can be used to evaluate many distributions which are defined using functions of two or more random variables. Cook [1981] provides FORTRAN code for the evaluation of the probability density function and cumulative distribution function of H-function variates. Using this program, it is possible to validate the extended Laguerre series approximation to the product of a gamma(α, β) random variable and an independent beta(c, d) random variable. Unfortunately, overflow errors were encountered for relatively small values of the parameters and the study is thus limited.

The parameters β, c, and d were tested at two values, and α was varied over 3 values. Table 5.4 presents results of comparing the cumulative distribution function approximation $F_L(x)$ using the extended Laguerre series (with seven moments) to the cumulative distribution function approximation $F_C(x)$ as computed by Cook's program. For comparative statistics, we define E_n and $\bar{e}_n$ to correspond with our previous definitions of Δ_n and $\bar{d}_n$, i.e.,

$$E_n = \max_x |F_C(x) - F_L(x)|, \quad \bar{e}_n = \frac{1}{n}\sum_{i=1}^{n} |F_C(x) - F_L(x)|. \qquad [5.32]$$

The results of Table 5.4 were computed over 100 points evenly spaced between $\mu - 3\sigma$ and $\mu + 3\sigma$ (or 100 points evenly spaced between 0 and $\mu + 3\sigma$ if $\mu - 3\sigma < 0$).

Table 5.4 - Performance of Laguerre Approximation versus H-Function Evaluation

α	β	c	d	E_n	$\bar{e}_n$
1	1	1	1	0.0219	0.0029
1	1	1	2	0.0388	0.0049
1	1	2	1	0.0116	0.0021
1	1	2	2	0.0199	0.0033
1	2	1	1	0.0219	0.0029
1	2	1	2	0.0388	0.0049
1	2	2	1	0.0116	0.0021
1	2	2	2	0.0199	0.0033
2	1	1	1	0.0000	0.0000
2	1	1	2	0.0116	0.0021
2	1	2	1	0.0034	0.0010
2	1	2	2	0.0074	0.0021
2	2	1	1	0.0000	0.0000
2	2	1	2	0.0116	0.0021
2	2	2	1	0.0034	0.0010
2	2	2	2	0.0074	0.0021
3	1	1	1	0.0089	0.0023
3	1	1	2	0.0000	0.0000
3	1	2	1	0.0002	0.0000
3	1	2	2	0.0025	0.0009
3	2	1	1	0.0089	0.0023
3	2	1	2	0.0000	0.0000
3	2	2	1	0.0002	0.0000
3	2	2	2	0.0025	0.0009

From the table, we conclude the extended Laguerre series approximation performs well for $\alpha = 2$ or 3 (exact to 4 decimal places in 4 of these 16 cases), and reasonably well with some errors of large magnitude when $\alpha = 1$. Both the maximum absolute deviation and the average absolute deviation are strongly correlated with the parameters c and d of the beta distribution. The recurrent cycles of identical errors (both maximum and average) over sets of four combinations of c and d persist over all 24 cases. The interpretation of this effect is unclear, but it is possible (as occurred with Cornish-Fisher approximation in §5.3.2) that the extended Laguerre approximation is independent of the parameter β. Figure 5.5 depicts the extended Laguerre and Cook's approximation in one of the poorest cases in the study, when $\alpha = \beta = c = 1.0$ and $d = 2.0$. The maximum difference of $E_n = 0.0388$ occurs at $z = 0.02$. Figure 5.6 provides a graph of the more typical case $\alpha = 2.0$, $\beta = 1.0$,

$c = 1.0$, and $d = 2.0$, where $E_n = 0.0116$ and the average absolute deviation is $\bar{e}_n = 0.0021$.

Figure 5.5 - Extended Laguerre and Cook Approximations (1)

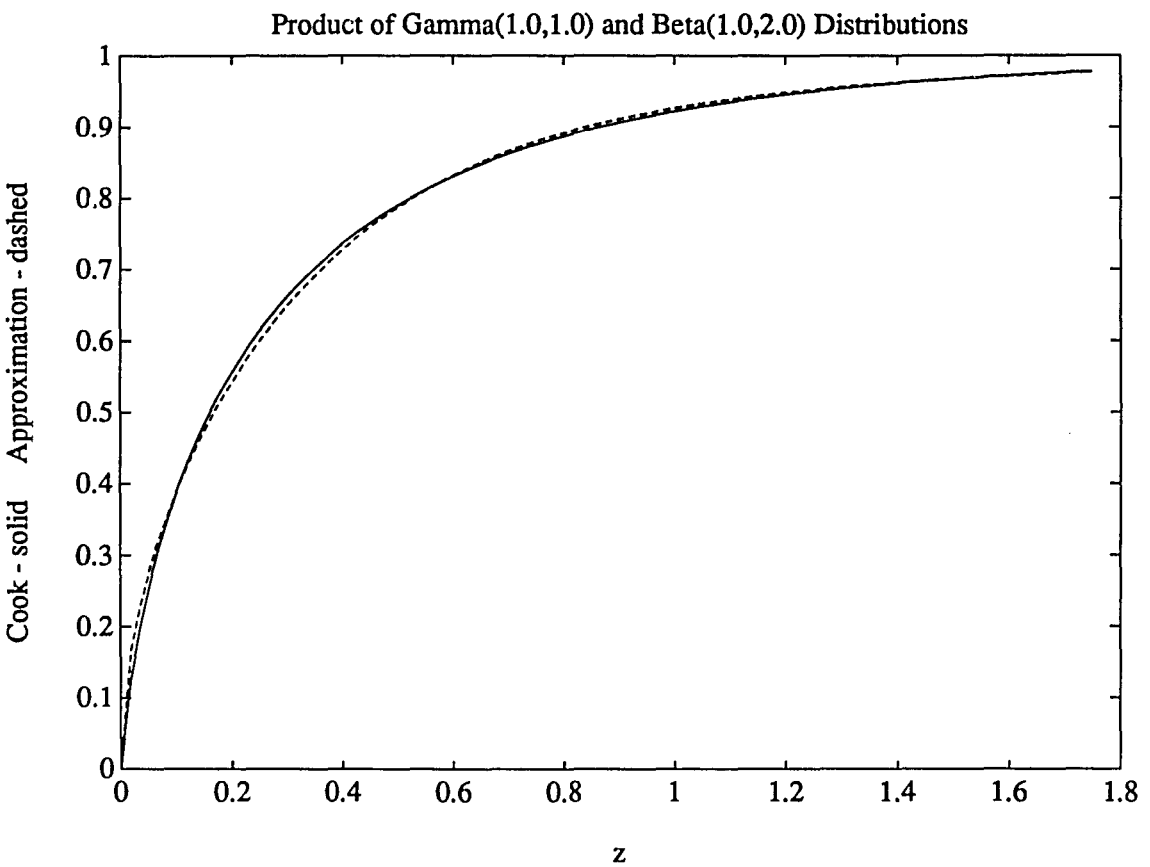
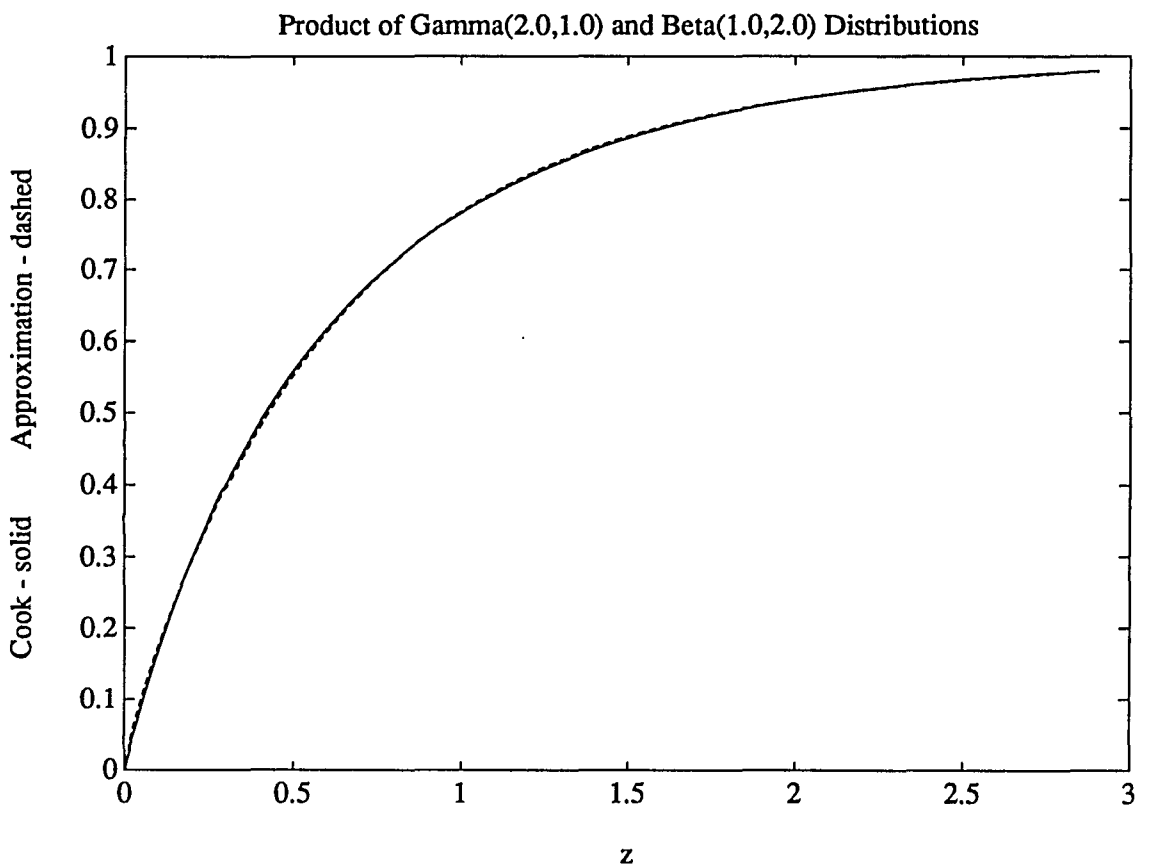


Figure 5.6 - Extended Laguerre and Cook Approximations (2)



5.3.4 Difference of Two IID Beta Random Variables

Let $X \sim \text{beta}(a, b)$, $Y \sim \text{beta}(a, b)$ and independent of X. Define $Z = X - Y$. In this case we have a random variable whose domain is finite. The Jacobi polynomial series can be used to approximate the cumulative distribution function. We let the parameters a and b vary over 5 values from the set $\{1, 2, 3, 4, 5\}$ and use the performance statistics of §5.3.2 for 10,000 observations. Additionally, to investigate the accuracy of the series in the tails of the distributions, we define the statistics

$$D^{0.05} = \max_{n \leq 500} |S_n(x) - F_J(x)|, \quad D^{0.95} = \max_{n \geq 9500} |S_n(x) - F_J(x)|, \qquad [5.33]$$

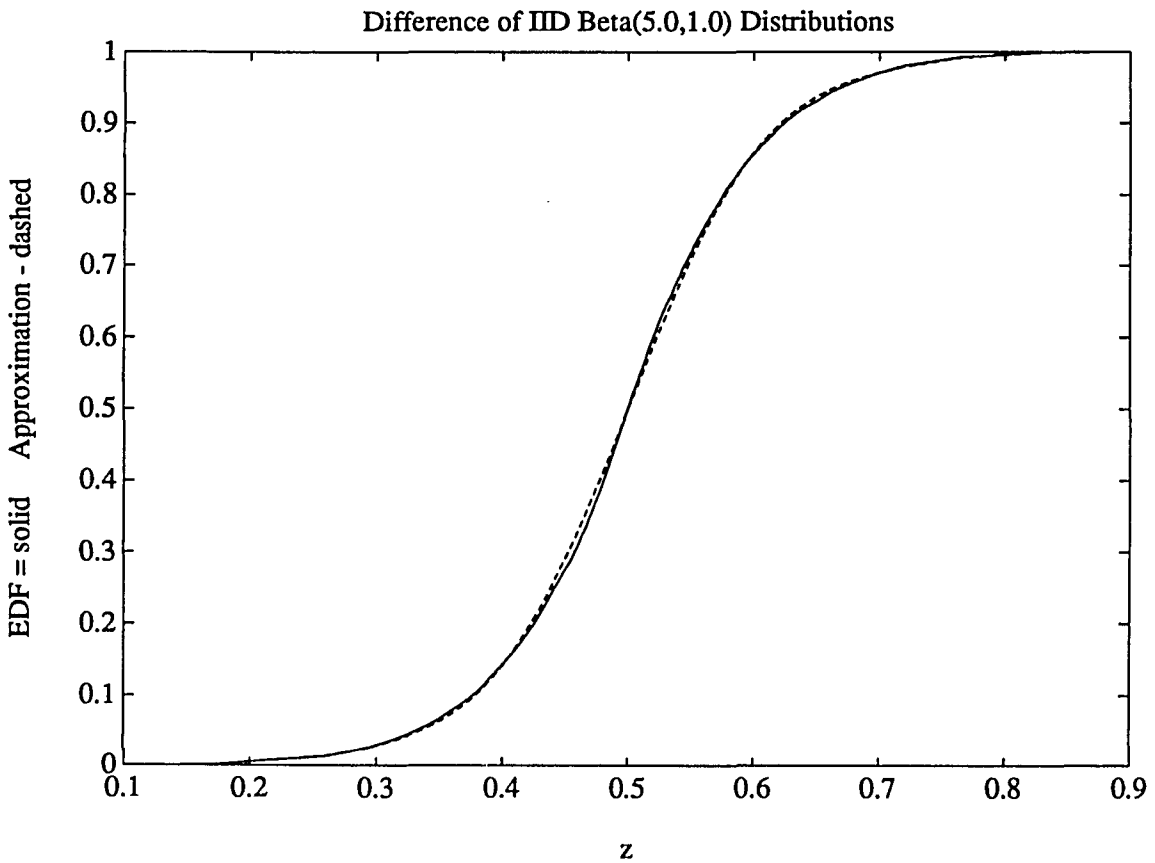
which represent the maximum absolute deviations of the Jacobi series approximation $F_J(x)$ from the empirical distribution function $S_n(x)$ in the lower and upper 5 % tails of $S_n(x)$, respectively. The results are given below in Table 5.5 for a Jacobi series using the first six moments of Z.

Table 5.5 - Jacobi Approximation for Difference of Beta Distributions

a	b	Δ_n	$\bar{d}_n$	$D^{0.05}$	$D^{0.95}$
1	1	0.0096	0.0033	0.0041	0.0016
1	2	0.0144	0.0046	0.0028	0.0041
1	3	0.0108	0.0041	0.0027	0.0024
1	4	0.0157	0.0069	0.0027	0.0024
1	5	0.0169	0.0077	0.0038	0.0043
2	1	0.0086	0.0029	0.0035	0.0043
2	2	0.0122	0.0054	0.0013	0.0026
2	3	0.0086	0.0023	0.0044	0.0013
2	4	0.0140	0.0052	0.0025	0.0019
2	5	0.0092	0.0028	0.0027	0.0037
3	1	0.0107	0.0034	0.0024	0.0022
3	2	0.0060	0.0022	0.0034	0.0018
3	3	0.0062	0.0021	0.0017	0.0028
3	4	0.0077	0.0029	0.0031	0.0022
3	5	0.0080	0.0033	0.0026	0.0016
4	1	0.0140	0.0051	0.0021	0.0027
4	2	0.0061	0.0017	0.0045	0.0020
4	3	0.0100	0.0038	0.0049	0.0021
4	4	0.0066	0.0017	0.0017	0.0015
4	5	0.0047	0.0015	0.0013	0.0019
5	1	0.0218	0.0074	0.0028	0.0044
5	2	0.0071	0.0019	0.0028	0.0026
5	3	0.0069	0.0024	0.0044	0.0021
5	4	0.0125	0.0052	0.0028	0.0020
5	5	0.0081	0.0024	0.0026	0.0020

While the statistic Δ_n demonstrates the approximation has relatively large errors ($>$ 0.01) in some cases, the statistic $\bar{d}_n$, representing the average error, indicates much better performance. The accuracy in the tails is excellent, especially considering this is the area of crucial importance for many tests and a troublesome one for many approximation methods. The *maximum* observed error in both the upper and lower tails over all 25 test cases never exceeds 0.0050. Figure 5.7 presents a comparison of the Jacobi approximation to the empirical distribution function for the case $a = 5$, $b = 1$.

Figure 5.7 - Jacobi Approximation for Difference of Beta(5.0,1.0) Distributions



5.3.5 Product of n IID Uniform$[0,1]$ Random Variables

Let $X_i \sim \mathrm{U}[0,1]$, where the X_i are mutually independent, and let

$$Y = \prod_{i=1}^{n} X_i. \qquad [5.34]$$

Using Mellin transforms, Springer [1979] finds the probability density function of Y to be

$$f(y) = \begin{cases} \dfrac{(\ln 1/y)^{n-1}}{(n-1)!} & \text{for } 0 < y \leq 1, \\ 0, & \text{otherwise.} \end{cases} \qquad [5.35]$$

To find the cumulative distribution function from this, we start from the definition

$$F(y) = \frac{1}{(n-1)!} \int_0^y (\ln 1/t)^{n-1}\, dt. \qquad [5.36]$$

By use of the substitution $x = \ln 1/t$ and conversion of the factorial to the more general gamma function, the integral becomes

$$F(y) = \frac{1}{\Gamma(n)} \int_{\ln 1/y}^{\infty} x^{n-1} e^{-x} dx, \qquad [5.37]$$

which is now recognizable as the complement of the incomplete gamma function. We have previously noted that the incomplete gamma function can be accurately computed. In the notation of §4.2.5, we have

$$F(y) = 1 - P(n, -\ln y). \qquad [5.38]$$

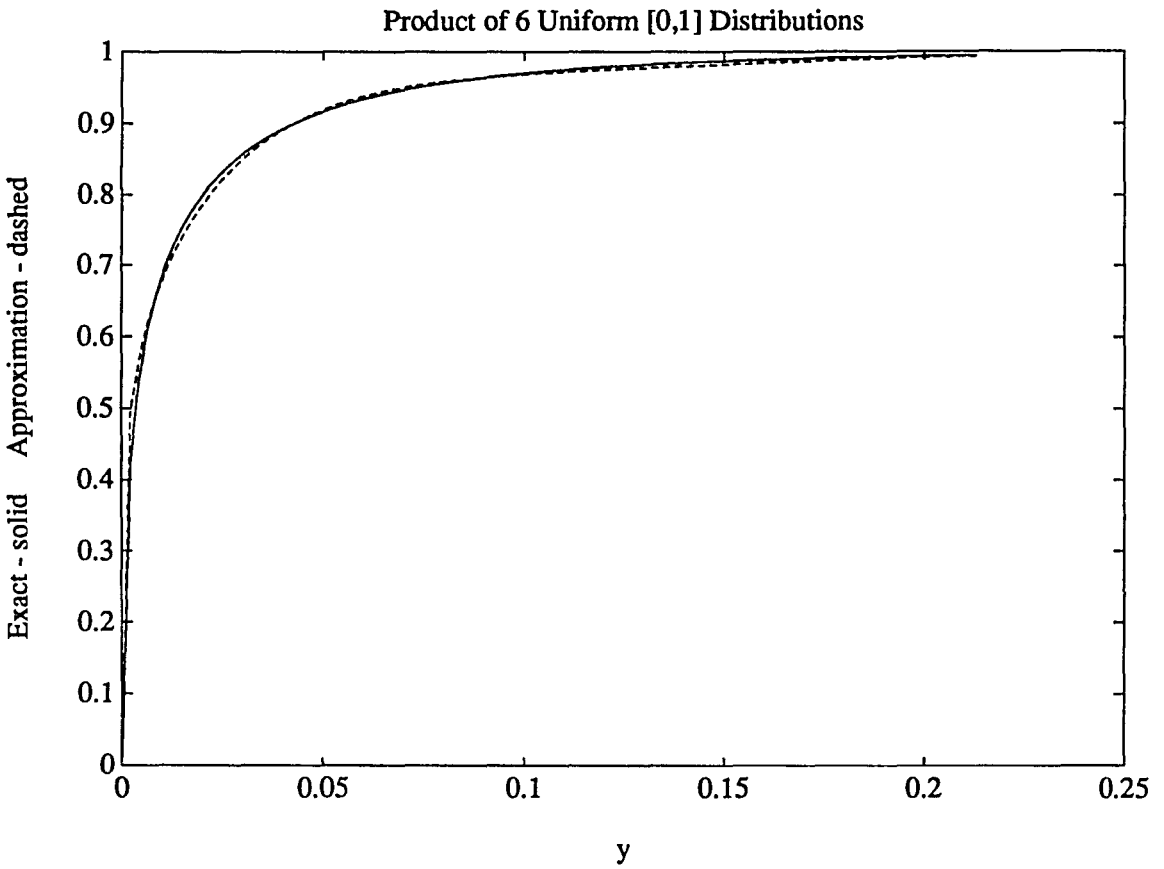
We use the Jacobi series expansion to approximate this cumulative distribution function for $n = 1, 2, \ldots, 10$, and compile the statistics D and δ_n defined in §5.3.1. The results for 100 values of y evenly spaced in the interval from $\mu - 3\sigma$ to $\mu + 3\sigma$ are provided in Table 5.6. below.

Table 5.6 Jacobi Approximation for Product of n Uniform Distributions

n	D	δ_n	$D^{0.05}$	$D^{0.95}$
1	0.0000	0.0000	0.0000	0.0000
2	0.0035	0.0008	0.0028	0.0007
3	0.0140	0.0019	0.0102	0.0011
4	0.0313	0.0034	0.0257	0.0050
5	0.0530	0.0045	0.0530	0.0056
6	0.0952	0.0056	0.0952	0.0049
7	0.1530	0.0064	0.1530	0.0038
8	0.2253	0.0071	0.2253	0.0045
9	0.3093	0.0088	0.3093	0.0116
10	0.4037	0.0135	0.4037	0.0143

The approximation is initially exact, as is expected since a uniform distribution is a special case of the beta distribution, and we know from Theorem 4.13 the Jacobi series is exact for a beta distribution. However, the accuracy deteriorates rapidly, becoming unacceptable in the lower tail as early as $n = 4$, while remaining fairly good in the upper tail as far as $n = 8$. The difficulties in approximating lower tail probabilities can be easily understood by examining the graph in Figure 5.8 for $n = 6$. While the perpendicular distance between the Jacobi series approximation and the exact cumulative distribution function is relatively small, the extreme slope of the cumulative distribution function causes a relatively large difference in the vertical distance between the two curves.

Figure 5.8 - Jacobi Approximation for Product of 6 Uniform [0,1] Distributions



5.3.6 Quotient of a Weibull and Uniform[1,2] Random Variable

We consider the quotient $Z = X/Y$, where X has a two parameter Weibull distribution with probability density function

$$f(x) = \begin{cases} \alpha\beta^{-\alpha}x^{\alpha-1}e^{-\left(\frac{x}{\beta}\right)^{\alpha}} & \text{for } x > 0; \\ 0, & \text{otherwise,} \end{cases} \qquad [5.39]$$

where $\alpha > 0$ is the shape parameter and $\beta > 0$ is the scale parameter, and $Y \sim$ U[1,2], independent of X. As with the product functions of earlier sections, Mellin transforms are of considerable use in computing the moments of Z. The Mellin transform of the probability density function of a Weibull random variable is known to be

$$f_X^*(s) = \beta^{s-1}\Gamma\left(\frac{s+\alpha-1}{\alpha}\right). \qquad [5.40]$$

From the definition [2.4] of the Mellin transform, we find the transform of the probability density function of Y:

$$f_Y^*(s) = \frac{2^s - 1}{s}. \qquad [5.41]$$

Applying the Mellin transform quotient property [2.14] and simplifying yields the transform of Z:

$$f_Z^*(s) = \frac{\beta^{s-1}\Gamma\left(\frac{s+\alpha-1}{\alpha}\right)\left(1 - \left(\frac{1}{2}\right)^{s-2}\right)}{s-2}. \qquad [5.42]$$

The moments of Z are provided by [2.19], which leads to

$$E[Z^n] = \begin{cases} \frac{\beta}{\alpha}\Gamma\left(\frac{1}{\alpha}\right)\ln 2, & n = 1; \\ \left(\frac{n}{n-1}\right)\left(\frac{2^{n-1}-1}{2^{n-1}}\right)\frac{\beta^n\Gamma\left(\frac{n}{\alpha}\right)}{\alpha}, & n = 2, 3, \ldots, \end{cases} \qquad [5.43]$$

where the case $n = 1$ is found from an application of L'Hopital's rule.

Using the first seven moments in an extended Laguerre series, we consider choosing α and β from the set $\{0.5, 1.0, 5.0, 10.0, 50.0\}$. Table 5.7 summarizes the statistics $\Delta_n, \bar{d}_n, D^{0.05}$, and $D^{0.95}$ over all twenty-five test cases.

Table 5.7 - Performance of Extended Laguerre Approximation in Quotient Case

α	β	Δ_n	$\bar{d}_n$	$D^{0.05}$	$D^{0.95}$
0.5	0.5	0.0954	0.0558	0.0917	0.0490
0.5	1.0	0.0908	0.0540	0.0905	0.0491
0.5	5.0	0.0901	0.0537	0.0893	0.0493
0.5	10.0	0.0948	0.0552	0.0884	0.0493
0.5	50.0	0.0942	0.0554	0.0885	0.0489
1.0	0.5	0.0064	0.0029	0.0064	0.0033
1.0	1.0	0.0093	0.0036	0.0048	0.0027
1.0	5.0	0.0079	0.0030	0.0039	0.0018
1.0	10.0	0.0106	0.0027	0.0040	0.0014
1.0	50.0	0.0062	0.0029	0.0029	0.0034
5.0	0.5	0.0063	0.0019	0.0029	0.0013
5.0	1.0	0.0108	0.0047	0.0026	0.0032
5.0	5.0	0.0052	0.0018	0.0033	0.0027
5.0	10.0	0.0080	0.0026	0.0028	0.0023
5.0	50.0	0.0049	0.0014	0.0027	0.0013
10.0	0.5	0.0074	0.0029	0.0059	0.0026
10.0	1.0	0.0164	0.0058	0.0023	0.0045
10.0	5.0	0.0087	0.0029	0.0038	0.0023
10.0	10.0	0.0113	0.0044	0.0048	0.0039
10.0	50.0	0.0104	0.0036	0.0044	0.0019
50.0	0.5	0.0334	0.0102	0.0334	0.0122
50.0	1.0	0.0357	0.0118	0.0261	0.0078
50.0	5.0	0.0295	0.0109	0.0295	0.0074
50.0	10.0	0.0279	0.0096	0.0279	0.0077
50.0	50.0	0.0307	0.0110	0.0303	0.0091

It is evident from the table that the quality of the approximation is closely tied to the α parameter of the Weibull distribution. The approximation performs well for the intermediate values of α, but poorly for $\alpha = 0.5$ and $\alpha = 50.0$. Some general remarks concerning the problems encountered with the orthogonal polynomial approximations which are typified by this example are discussed in the next chapter. Figure 5.9 displays the case $\alpha = 0.5$, $\beta = 0.5$, which is the poorest result of this experiment. Figure 5.10 presents the case $\alpha = 1.0$, $\beta = 0.5$, which shows a marked improvement in the approximation.

Figure 5.9 - Extended Laguerre Approximation for Weibull(0.5,0.5) Quotient Case

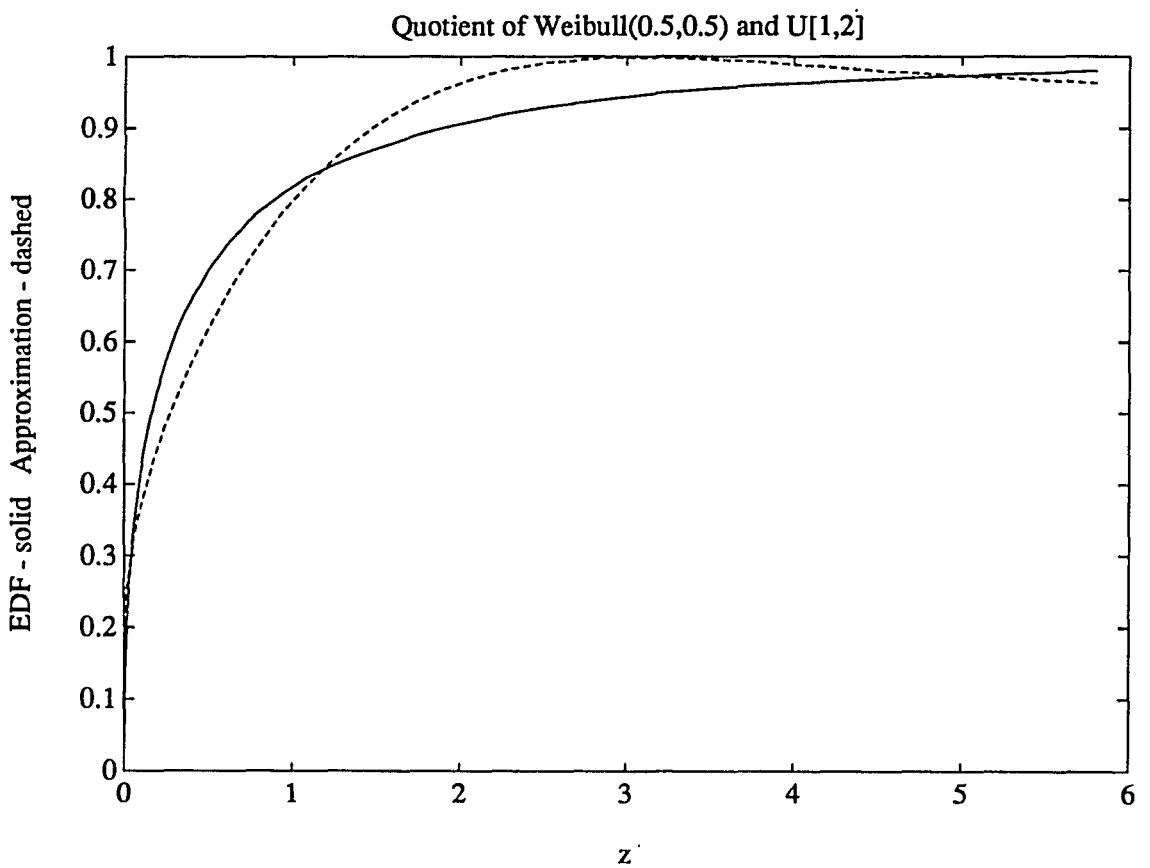
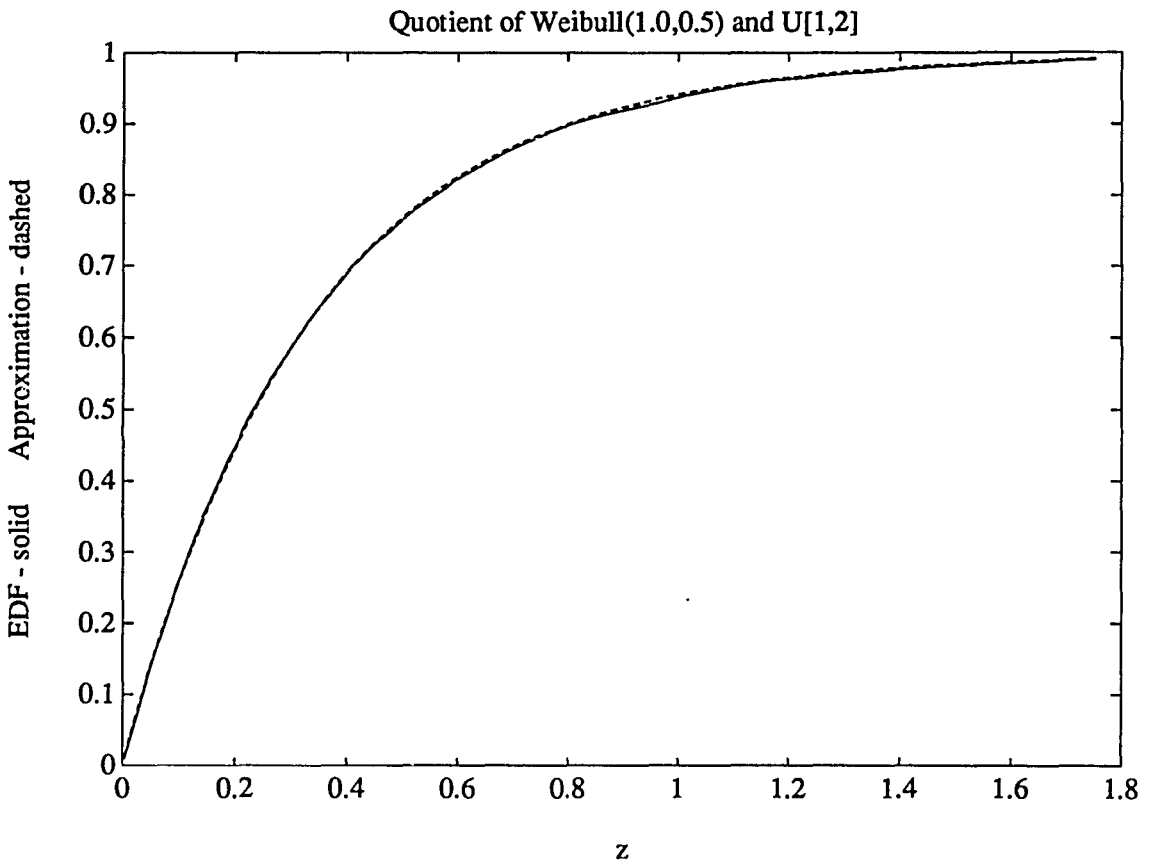


Figure 5.10 - Extended Laguerre Approximation for Weibull(1.0,0.5) Quotient Case



CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

6.0 Discussion

The technique developed in this dissertation offers practitioners a methodology for approximating the probability density functions and cumulative distribution functions of random variables which previously were difficult to evaluate. In this chapter we discuss the advantages and disadvantages of the technique as well as suggestions for future research. Theoretical and empirical aspects are considered.

Theoretically, the inverse weighted mean square error minimization property of Theorem 3.7 provides support for use of the orthogonal polynomial series expansions. When the weighting function has no poles and takes on a typical 'bell-shaped' form, the theorem suggests the approximation may be accurate in the tails, which is important in many applications. However, it is also possible for this property to be a disadvantage. For example, suppose the probability density function $f(x)$ of a random variable X defined on $(0, \infty)$ has a pole at $x = 0$, i.e., $\lim_{x \to 0} f(x) = \infty$. If the gamma weighting function associated with the extended Laguerre series for $f(x)$ has shape parameter $\alpha < 1$, then this associated weighting function also has a pole at zero. Considering the inverse weighting relationship, this suggests almost no 'weight' will be given to the approximation in the region near $x = 0$. The result could be large approximation errors in the lower tail. Empirically, such a case occurs in §5.3.5 for the Jacobi series. From [5.35], we see the probability density function $f(y)$ for the random variable Y is unbounded as $y \to 0$. For the associated beta(α, β) weighting function, the method suggested in Chapter 5 for choosing the parameters results in $\alpha = 0.1975$ and $\beta = 12.44$. It is known that for $\alpha < 1$, the probability density function of the beta is unbounded at zero. Thus, the relatively large error observed for this case (see Table 5.6 or Figure 5.8) is partially explained by this phenomenon. The possibility arises of determining the parameters of a weighting function that avoids this problem. Certainly this is an area worthy of future investigation.

Another advantage of the orthogonal polynomial series expansions is the requirement that only the moments and domain of the random variable be known. For many applications this is desirable, as the moments can be computed and the domain is generally known. Even functions of dependent random variables can be approximated if the moments are known. However, if the moments of the random variable of interest do not exist, the polynomial series cannot be applied. For the functions considered in Chapter 5, only the quotient case of §5.3.6 presents any difficulties (e.g., we could not approximate the quotient of a Weibull and a U[0, 1] random variable because the moments of the reciprocal uniform do not exist). This existence problem is mainly of theoretical importance, as applications do not tend to produce pathological cases. As noted in the literature review and some of the examples, actual computation of the moments can be greatly simplified by the use of Mellin transforms for product and quotient functions.

Empirically, the results of Chapter 5 generally lend reinforcement to the orthogonal polynomial series expansion approach. The formerly cumbersome product case $\tilde{\lambda}_a = \tilde{P}\tilde{\lambda}_c$ of the traffic conflict application is handled with ease, even though some of the parameters of the component distributions are relatively large. The accuracy of the approximation is maintained over the entire distribution, with a *maximum* difference from the empirical distribution function of 100,000 observations of $\Delta = 0.0026$ (using just seven moments). It is thus possible to make precise inferences and probability statements using the model. The graphs and tables provided in Chapter 5 attest to the ability of the series expansions to closely approximate various types of distributions. A problem that occasionally occurs can be seen in Figure 5.2. The approximations are not always monotonic, and values of the approximation to the cumulative distribution function can be negative or greater than one. The works of Barton and Dennis [1952] and Draper and Tierney [1972] cited by Kendall [1987] provide some insights for these anomalies, but general results for any of the polynomial series expansions do not seem to be available in the literature. The specification of necessary or sufficient conditions for monotonicity or requirements on the range should be considered for future research.

There are several other potential topics for future research. Bivariate and multivariate generalizations of the Hermite polynomials, Gram-Charlier and Edgeworth series exist, as Kendall [1987] mentions. Analogous expressions for the Jacobi and extended Laguerre polynomial series expansions should be investigated also. Although we have presented a natural method to determine the parameters in the Jacobi and extended Laguerre weighting functions, as mentioned above there is some flexibility in the choice of these parameters. Perhaps it is possible to determine the parameters in such a manner that the approximation is accurate in, for example, the upper tail only. If knowledge of the upper tail is all that is needed, this could be useful. A related question is the general accuracy of the approximation. Some theoretical analysis relating the number of moments used in the truncated series to the error of the approximation would be a great aid to the practitioner. The potential for reducing the order of the error of the extended Laguerre and Jacobi series expansions (as Cornish and Fisher did with the Hermite series) indicates this should also be investigated.

The approximation of discrete distributions with orthogonal polynomials has not been discussed, but this could be attempted by introducing Lebesgue-Stieltjes integrals into the derivations of the series expansions. Coefficients of the approximating series could then be determined from the moments, as in the continuous case. While the prevalence of discrete random variables in stochastic modeling may not be as widespread as that of continuous random variables, the method could be of some use.

In conclusion, we feel the orthogonal polynomial series expansions provide a straightforward and often extremely accurate approximation to the cumulative distribution functions of random variables which previously required laborious efforts to evaluate. Theoretical concerns have been addressed, a natural methodology for applying the expansions has been described, empirical results which support the technique have been presented, and potential problems of the approximations have been discussed. The relationships of the orthogonal polynomial series expansions to common distributions have clarified and complemented the existing literature. We

feel the methodology provides a solid and more complete understanding of the uses of orthogonal polynomials in the approximation of probability distributions.

Appendix A

Completeness of the Extended Laguerre Polynomials

In Chapter 3 the three families of the Hermite, Jacobi, and Laguerre orthogonal polynomials were introduced. It is well known that each family is complete. Here we demonstrate the completeness of the more general extended Laguerre polynomials defined in Chapter 4 and used in some applications in Chapter 5. From the definition of completeness given in Chapter 3, the system $\{P_n\}_{n=0}^{\infty}$ of functions in question is required to be orthonormal. Therefore, we first establish the orthonormal version of the extended Laguerre polynomials with the following lemma.

Lemma A.1. *The system of functions*

$$\{P_n\}_{n=0}^{\infty} = \left[\sqrt{\frac{\beta^{\alpha+1}x^{\alpha}e^{-\beta x}}{(\alpha+1)_n n!\Gamma(\alpha+1)}}L_n^{(\alpha,\beta)}(x)\right]_{n=0}^{\infty}, \qquad [A.1]$$

where $L_n^{(\alpha,\beta)}(x)$ represents the extended Laguerre polynomials as defined by [4.22], is orthonormal.

Proof.

Orthonormality requires orthogonality and normality. The system $\{P_n\}_{n=0}^{\infty}$ is easily shown to be orthogonal over $[0,\infty)$:

$$\int_0^{\infty}\sqrt{\frac{\beta^{\alpha+1}x^{\alpha}e^{-\beta x}}{(\alpha+1)_m m!\Gamma(\alpha+1)}}L_m^{(\alpha,\beta)}(x)\sqrt{\frac{\beta^{\alpha+1}x^{\alpha}e^{-\beta x}}{(\alpha+1)_n n!\Gamma(\alpha+1)}}L_n^{(\alpha,\beta)}(x)dx$$

$$= \frac{1}{\sqrt{(\alpha+1)_m m!(\alpha+1)_n n!}}\int_0^{\infty}\frac{\beta^{\alpha+1}x^{\alpha}e^{-\beta x}}{\Gamma(\alpha+1)}L_m^{(\alpha,\beta)}(x)L_n^{(\alpha,\beta)}(x)dx \qquad [A.2]$$

$$= \begin{cases} 0, & \text{if } m \neq n; \\ 1, & \text{if } m = n, \end{cases}$$

where the last equation follows from the extended Laguerre polynomial orthogonality relation [4.28]. Since the norm of $\{P_n\}_{n=0}^{\infty}$ is unity, the orthonormality is apparent. ■

With this result, we now demonstrate completeness, following the method of Sansone [1959].

Theorem A.2. *The system* $\{P_n\}_{n=0}^{\infty}$ *is complete.*

Proof. According to the definition of completeness, we must show that the only square integrable function which is orthogonal over $[0,\infty)$ to all functions of the system $\{P_n\}_{n=0}^{\infty}$ is the function which is zero almost everywhere. Assume such a square integrable function f exists. By hypothesis,

$$\sqrt{\frac{\beta^{\alpha+1}}{(\alpha+1)_n n!\Gamma(\alpha+1)}}\int_0^{\infty} f(x)e^{\frac{-\beta x}{2}}x^{\frac{\alpha}{2}}L_n^{(\alpha,\beta)}(x)dx = 0, \quad n = 0,1,2,\ldots \qquad [A.3]$$

Since the $L_n^{(\alpha,\beta)}(x)$ are polynomials, this implies

$$\int_0^{\infty} f(x)e^{\frac{-\beta x}{2}}x^{\frac{\alpha}{2}}x^n dx = 0, \quad n = 0,1,2,\ldots \qquad [A.4]$$

We make the substitution $x = \frac{2t}{\beta}$, yielding

$$\int_0^{\infty} f(\tfrac{2t}{\beta})e^{-t}t^{\frac{\alpha}{2}}t^n dt = 0, \quad n = 0,1,2,\ldots, \qquad [A.5]$$

and let the function $F(t)$ be defined by

$$F(t) = f(\tfrac{2t}{\beta})t^{1+\frac{\alpha}{2}}e^{\frac{-t}{2}}. \qquad [A.6]$$

Then clearly

$$\int_0^{\infty} F(t)e^{\frac{-t}{2}}t^{n-1}dt = 0, \quad n = 1,2,\ldots, \qquad [A.7]$$

and

$$\int_0^{\infty} F(t)\left[e^{\frac{-t}{2}}L_n^{(0,1)}(t)\right]dt = 0, \quad n = 0,1,2,\ldots \qquad [A.8]$$

But $F(t)$ is square integrable over $[0,\infty)$ since $t^{1+\frac{\alpha}{2}}e^{\frac{-t}{2}}$ is bounded in $[0,\infty)$. Since the $L_n^{(0,1)}(t)$ polynomials are known to be complete (see Sansone [1959]), $F(t)$ and therefore $f(t)$ is zero almost everywhere in $[0,\infty)$. ■

Appendix B

Edgeworth and Cornish-Fisher Series Expansions

B.0 Introduction

The Gram-Charlier Type A series expansion presented in Chapter 4 is refined in Edgeworth [1904] and in Cornish and Fisher [1937]. We develop these expansions in detail, following Kendall [1987]. Since both the Edgeworth and Cornish-Fisher forms are developed in terms of the cumulants of the random variable of interest rather than the moments used in the Gram-Charlier version, we first introduce the necessary background matter.

B.1 Cumulants

The concept of cumulants, as Kendall [1987] notes, originates in the literature in Thiele [1903]. While cumulants are not as prevalent as moments in the literature, they possess similarly useful properties. Fortunately for our purposes, they can be computed from the moments. If we define the rth moment of a random variable X with cumulative distribution function $F(x)$ with the standard Stieltjes integral

$$\mu_r = \int_{-\infty}^{\infty} x^r \, dF(x), \qquad [B.1]$$

then the cumulants $\kappa_1, \kappa_2, \ldots, \kappa_r$ are defined with the following identity in t:

$$\exp\left(\sum_{r=1}^{\infty} \frac{\kappa_r t^r}{r!}\right) = \sum_{r=0}^{\infty} \frac{\mu_r t^r}{r!}. \qquad [B.2]$$

If t is complex, we substitute it for t to get

$$\begin{aligned} \exp\left(\sum_{r=1}^{\infty} \frac{\kappa_r (it)^r}{r!}\right) &= \sum_{r=0}^{\infty} \frac{\mu_r (it)^r}{r!} \\ &= \int_{-\infty}^{\infty} e^{itx} dF, \end{aligned} \qquad [B.3]$$

which is immediately recognized as the characteristic function, $\phi(t)$, of X. Thus, κ_r is seen to be the coefficient of $(it)^r/r!$ in $ln\ \phi(t)$, if the power series expansion can be put in this form. We call the function $ln\ \phi(t)$ the *cumulant generating function.* If the random variable X is standardized by subtracting its mean μ_1 and dividing by its standard deviation σ, the effect on the rth cumulant is a simple division by σ^r. Denoting the rth standardized cumulant by κ_{sr}, we have

$$\kappa_{sr} = \frac{\kappa_r}{\sigma^r}. \qquad [B.4]$$

The cumulants of a random variable are not directly calculable unless the series form of the logarithm of the characteristic function can be found. However, as alluded to above, the cumulants can be related to the moments of the same and lesser orders. Differentiating [B.2] with respect to κ_j gives

$$\left(\sum_{r=0}^{\infty} \frac{\mu_r t^r}{r!}\right) \frac{t^j}{j!} = \sum_{r=1}^{\infty} \frac{\partial \mu_r}{\partial \kappa_j} \frac{t^r}{r!}, \quad j \geq 1. \qquad [B.5]$$

By identifying powers of t we get in general

$$\frac{\partial \mu_r}{\partial \kappa_j} = \binom{r}{j} \mu_{r-j}, \qquad [B.6]$$

and specifically for κ_1,

$$\frac{\partial \mu_r}{\partial \kappa_1} = r\mu_{r-1}. \qquad [B.7]$$

Using this relation we can find the moments in terms of the cumulants and then solve explicitly for the cumulants in terms of the moments. The first ten of these equations are given by Kendall et al. [1987]; we include the first five here:

$$\begin{aligned}
\kappa_1 &= \mu_1 \\
\kappa_2 &= \mu_2 - {\mu_1}^2 \\
\kappa_3 &= \mu_3 - 3\mu_2\mu_1 + 2{\mu_1}^3 \\
\kappa_4 &= \mu_4 - 4\mu_3\mu_1 - 3{\mu_2}^2 + 12\mu_2{\mu_1}^2 - 6{\mu_1}^4 \\
\kappa_5 &= \mu_5 - 5\mu_4\mu_1 - 10\mu_3\mu_2 + 20\mu_3{\mu_1}^2 + 30{\mu_2}^2\mu_1 - 60\mu_2{\mu_1}^3 + 24{\mu_1}^5.
\end{aligned} \qquad [B.8]$$

B.2 Edgeworth's version of the Type A Series

The Gram-Charlier Type A series can behave poorly in that approximations of the cumulative distribution function using j terms in the summation may be a worse approximation than the sum of $j-1$ terms. Edgeworth [1904] derives another Type A series as an improvement. In what follows, D is the symbol for differentiation, $H_r(x)$ is the rth degree Hermite polynomial, and $\tau(x)$ is the weighting function (i.e., the probability density function of a standard normal random variable). We start by finding the Fourier transformation of $H_r(x)\tau(x)$. Because

$$\sqrt{2\pi}\,\tau(t) \;=\; e^{\frac{-t^2}{2}} \;=\; \int_{-\infty}^{\infty} e^{itx}\frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}dx, \qquad [B.9]$$

it follows from Rodriguez' formula [4.2] and the differentiation property of Fourier transforms that

$$\sqrt{2\pi}\frac{d^r}{dt^r}\tau(t) \;=\; (-1)^r\sqrt{2\pi}H_r(t)\tau(t) \;=\; \int_{-\infty}^{\infty} i^r x^r \frac{e^{itx}}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}dx. \qquad [B.10]$$

Therefore, the transform of $x^r\tau(x)$ is $i^r\sqrt{2\pi}H_r(t)\tau(t)$, and

$$x^r\tau(x) \;=\; \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ixt}i^r\sqrt{2\pi}H_r(t)\tau(t)dt. \qquad [B.11]$$

Now by interchanging the roles of x and t,

$$\sqrt{2\pi}(-i)^r t^r\tau(t) \;=\; \int_{-\infty}^{\infty} e^{-ixt}H_r(x)\tau(x)dx. \qquad [B.12]$$

Finally, by changing the sign of t, the transform of $H_r(x)\tau(x)$ is seen to be $\sqrt{2\pi}i^r t^r\tau(t)$. Starting with the expression

$$[\exp(\kappa_r D^r)]\,\tau(x), \qquad [B.13]$$

we proceed to find its characteristic function:

$$\begin{aligned}\int_{-\infty}^{\infty} e^{itx}\exp(\kappa_r D^r)\tau(x)dx &= \int_{-\infty}^{\infty} e^{itx}\sum_{j=0}^{\infty}\left(\frac{\kappa_r^j D^{rj}}{j!}\right)\tau(x)dx \\ &= \sum_{j=0}^{\infty}\frac{\kappa_r^j}{j!}\int_{-\infty}^{\infty} e^{itx}D^{rj}\tau(x)dx \\ &= \sum_{j=0}^{\infty}\frac{\kappa_r^j}{j!}\int_{-\infty}^{\infty} e^{itx}(-1)^{rj}H_{rj}(x)\tau(x)dx \\ &= \sum_{j=0}^{\infty}\frac{\kappa_r^j}{j!}\sqrt{2\pi}(-i)^{rj}t^{rj}\tau(t) \\ &= \sqrt{2\pi}\tau(t)\exp[\kappa_r(-it)^r],\end{aligned} \qquad [B.14]$$

where again we have used Rodriguez' formula in passing from the second line to the third. A similar process yields the characteristic function of

$$\exp\left[-\frac{\kappa_1-a}{1!}D+\frac{\kappa_2-b}{2!}D^2-\frac{\kappa_3}{3!}D^3+\frac{\kappa_4}{4!}D^4\ldots\right]\tau(x) \qquad [B.15]$$

to be

$$\sqrt{2\pi}\tau(t)\exp\left[\frac{\kappa_1-a}{1!}it+\frac{\kappa_2-b}{2!}(it)^2+\frac{\kappa_3}{3!}(it)^3+\frac{\kappa_4}{4!}(it)^4+\ldots\right]. \qquad [B.16]$$

A more general expression uses the $\mathrm{N}(\mu,\sigma^2)$ density function

$$\nu(x) \;=\; \frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-m)^2}{2\sigma^2}} \;=\; \tau[(x-m)/\sigma]/\sigma. \qquad [B.17]$$

We then find the characteristic function of

$$\exp\left[-\frac{\kappa_1-a}{1!}D+\frac{\kappa_2-b}{2!}D^2-\frac{\kappa_3}{3!}D^3+\frac{\kappa_4}{4!}D^4\ldots\right]\nu(x) \qquad [B.18]$$

as

$$\sqrt{2\pi}\tau(t\sigma)e^{imt}\exp\left[\frac{\kappa_1-a}{1!}it+\frac{\kappa_2-b}{2!}(it)^2+\frac{\kappa_3}{3!}(it)^3+\frac{\kappa_4}{4!}(it)^4+\ldots\right]. \qquad [B.19]$$

Assuming [B.18] is a density function, we find its cumulant generating function by taking the logarithm of [B.19], which is

$$\frac{(\kappa_1-a+m)it}{1!}+\frac{\kappa_2-b+\sigma^2}{2!}(it)^2+\frac{\kappa_3}{3!}(it)^3+\frac{\kappa_4}{4!}(it)^4+\ldots \qquad [B.20]$$

By definition, we know the cumulants must be $\kappa_1-1+m, \kappa_2-b+\sigma^2, \kappa_3, \kappa_4, \ldots, \kappa_r, \ldots$ If we set $a = m$ and $b = \sigma^2$ we have a distribution whose cumulants are $\kappa_1, \kappa_2, \ldots$ If the series [B.19] converges to a distribution which is uniquely determined by its moments, then the series will have the desired distribution. Standardizing will give $\kappa_{s1} = 0$ and $\kappa_{s2} = 1$, and the Edgeworth series becomes

$$f(x) = \exp\left[-\kappa_{s3}\frac{D^3}{3!} + \kappa_{s4}\frac{D^4}{4!} - \ldots\right]\tau(x), \qquad [B.21]$$

where $\nu(x)$ is replaced by $\tau(x)$ because the mean is zero and the variance is unity. Expanding [B.21] results in the more useful form

$$f(x) = \tau(x)\left[1 + \frac{\kappa_{s3}}{6}H_3 + \frac{\kappa_{s4}}{24}H_4 + \frac{\kappa_{s5}}{120}H_5 + \frac{\kappa_{s6}+10\kappa_{s3}^2}{720}H_6 + \ldots\right]. \qquad [B.22]$$

B.3 Cornish-Fisher Series Expansion

Cornish and Fisher [1937] derive another Hermite polynomial series expansion to approximate the cumulative distribution function of a random variable whose cumulants are known. They apply it to estimate values of the Student-t and F-distributions. We begin with the general Edgeworth expression [B.18] with $a = m$ and $b = \sigma^2$:

$$\exp\left[-\frac{\kappa_1 - m}{1!}D + \frac{\kappa_2 - \sigma^2}{2!}D^2 - \frac{\kappa_3}{3!}D^3 + \frac{\kappa_4}{4!}D^4 \ldots\right]\nu(x). \qquad [B.23]$$

We suppose that a parameter $n \geq 1$ exists such that the rth cumulant is $O(n^{1-r})$. Then we choose $l_1, l_2, \ldots$ such that

$$\begin{aligned} \kappa_1 - m &= l_1\sigma \\ \kappa_2 - \sigma^2 &= l_2\sigma^2 \\ \kappa_3 &= l_3\sigma^3 \\ &\vdots \end{aligned} \qquad [B.24]$$

So since $\sigma^2 = \kappa_2$, it is of order n^{-1}, and then

$$\frac{\kappa_r}{\sigma^r} = O(n^{1-\frac{r}{2}}). \qquad [B.25]$$

Now we write [B.23] as

$$\exp\left[-l_1\sigma D + \tfrac{1}{2}l_2\sigma^2D^2 - \tfrac{1}{6}l_3\sigma^3D^3 + \tfrac{1}{24}l_4\sigma^4D^4 - \tfrac{1}{120}l_5\sigma^5D^5 + \ldots\right]\nu(x), \qquad [B.26]$$

where

$$\begin{aligned} l_1 &= O(n^{-\frac{1}{2}}), \qquad & l_3 &= O(n^{-\frac{1}{2}}) \\ l_2 &= O(n^{-1}), \qquad & l_4 &= O(n^{-1}) \\ l_5 &= O(n^{-\frac{3}{2}}), \qquad & l_6 &= O(n^{-2}). \end{aligned} \qquad [B.27]$$

Expanding the exponential part of [B.26] but including terms only up to order n^{-2}, we get

$$\begin{aligned} &1 - l_1\sigma D + \tfrac{1}{2}l_2\sigma^2D^2 - \tfrac{1}{6}l_3\sigma^3D^3 + \tfrac{1}{24}l_4\sigma^4D^4 - \tfrac{1}{120}l_5\sigma^5D^5 + \tfrac{1}{720}l_6\sigma^6D^6 + \tfrac{1}{2}(l_1^2\sigma^2D^2 \\ &+ \tfrac{1}{4}l_2^2\sigma^4D^4 + \tfrac{1}{36}l_3^2\sigma^6D^6 + \tfrac{1}{576}l_4^2\sigma^8D^8 - l_1l_2\sigma^3D^3 + \tfrac{1}{3}l_1l_3\sigma^4D^4 - \tfrac{1}{12}l_1l_4\sigma^5D^5 \\ &+ \tfrac{1}{60}l_1l_5\sigma^6D^6 - \tfrac{1}{6}l_2l_3\sigma^5D^5 + \tfrac{1}{24}l_2l_4\sigma^6D^6 - \tfrac{1}{72}l_3l_4\sigma^7D^7 + \tfrac{1}{360}l_3l_5\sigma^8D^8) \\ &+ \tfrac{1}{6}(-l_1^3\sigma^3D^3 - \tfrac{1}{216}l_3^3\sigma^9D^9 + \tfrac{3}{2}l_1^2l_2\sigma^4D^4 - \tfrac{1}{2}l_1^2l_3\sigma^5D^5 - \tfrac{1}{8}l_1^2l_4\sigma^6D^6 + \tfrac{1}{288}l_3^2l_4\sigma^{10}D^{10} \\ &- \tfrac{1}{12}l_1l_3^2\sigma^7D^7 + \tfrac{1}{24}l_2l_3^2\sigma^8D^8 + \tfrac{1}{2}l_1l_2l_3\sigma^6D^6 + \tfrac{1}{24}l_1l_3l_4\sigma^8D^8) + \tfrac{1}{24}(l_1^4\sigma^4D^4 \\ &+ \tfrac{1}{1296}l_3^4\sigma^{12}D^{12} + \tfrac{2}{3}l_1^3l_3\sigma^6D^6 + \tfrac{1}{6}l_1^2l_3^2\sigma^8D^8 + \tfrac{1}{54}l_1l_3^3\sigma^{10}D^{10}). \end{aligned} \qquad [B.28]$$

If we apply the above to $\nu(x)$, arrange the result in terms of the Hermite polynomials, and integrate, then we have the following second-order approximation to the cumulative distribution function of a random variable X:

$$\begin{aligned} F(x) \approx \int_{-\infty}^{x} \tau(t)dt + \tau(x)[&-(l_1 + \tfrac{1}{6}l_3H_2) - (\tfrac{1}{2}l_1^2H_1 + \tfrac{1}{2}l_2H_1 + \tfrac{1}{6}l_1l_3H_3 + \tfrac{1}{24}l_4H_3 \\ &+ \tfrac{1}{72}l_3^2H_5) - (\tfrac{1}{6}l_1^3H_2 + \tfrac{1}{2}l_1l_2H_2 + \tfrac{1}{12}l_1^2l_3H_4 + \tfrac{1}{12}l_2l_3H_4 + \tfrac{1}{24}l_1l_4H_4 + \tfrac{1}{120}l_5H_4 \\ &+ \tfrac{1}{72}l_1l_3^2H_6 + \tfrac{1}{144}l_3l_4H_6 + \tfrac{1}{1296}l_3^3H_8) - (\tfrac{1}{24}l_1^4H_3 + \tfrac{1}{8}l_2^2H_3 + \tfrac{1}{4}l_1^2l_2H_3 + \tfrac{1}{36}l_1^3l_3H_5 \\ &+ \tfrac{1}{12}l_1l_2l_3H_5 + \tfrac{1}{48}l_1^2l_4H_5 + \tfrac{1}{48}l_2l_4H_5 + \tfrac{1}{120}l_1l_5H_5 + \tfrac{1}{720}l_6H_5 + \tfrac{1}{144}l_1^2l_3^2H_7 \\ &+ \tfrac{1}{144}l_2l_3^2H_7 + \tfrac{1}{1152}l_4^2H_7 + \tfrac{1}{144}l_1l_3l_4H_7 + \tfrac{1}{720}l_3l_5H_7 + \tfrac{1}{1296}l_1l_3^3H_9 + \tfrac{1}{1728}l_3^2l_4H_9 \\ &+ \tfrac{1}{31104}l_3^4H_{11})] \ . \end{aligned} \qquad [B.29]$$

Appendix C

Leibniz' Rule for Differentiation of a Product

Theorem C.1. *Let u and v be functions of x and n be a positive integer. Then the expression for the nth derivative with respect to x of the product uv is given by*

$$\frac{d^n}{dx^n}(uv) = \sum_{m=0}^{n} \binom{n}{m} \frac{d^{n-m}u}{dx^{n-m}} \frac{d^m v}{dx^m}. \qquad [C.1]$$

Proof. The proof is by induction. For $n = 1$, the expression holds trivially from the familiar product rule of calculus. We assume it holds for n, that is, we assume [C.1]. Then by application of the product rule, we have

$$\frac{d^{n+1}}{dx^{n+1}}(uv) = \sum_{m=0}^{n} \left[\binom{n}{m} \frac{d^{n-m}u}{dx^{n-m}} \frac{d^{m+1}v}{dx^{m+1}} + \binom{n}{m} \frac{d^{n-m+1}u}{dx^{n-m+1}} \frac{d^m v}{dx^m} \right]. \qquad [C.2]$$

We separate the sum into two parts, and to avoid confusion, reindex both:

$$\frac{d^{n+1}}{dx^{n+1}}(uv) = \sum_{r=0}^{n} \binom{n}{r} \frac{d^{n-r}u}{dx^{n-r}} \frac{d^{r+1}v}{dx^{r+1}} + \sum_{i=0}^{n} \binom{n}{i} \frac{d^{n-i+1}u}{dx^{n-i+1}} \frac{d^i v}{dx^i}. \qquad [C.3]$$

Separating out the last term of the first sum and the first term of the second sum,

$$\frac{d^{n+1}}{dx^{n+1}}(uv) = u\frac{d^{n+1}v}{dx^{n+1}} + \sum_{r=0}^{n-1} \binom{n}{r} \frac{d^{n-r}u}{dx^{n-r}} \frac{d^{r+1}v}{dx^{r+1}} + \sum_{i=1}^{n} \binom{n}{i} \frac{d^{n-i+1}u}{dx^{n-i+1}} \frac{d^i v}{dx^i} + \frac{d^{n+1}u}{dx^{n+1}}v. \qquad [C.4]$$

Combining like terms from the sums yields

$$\begin{aligned} \frac{d^{n+1}}{dx^{n+1}}(uv) &= u\frac{d^{n+1}v}{dx^{n+1}} + \sum_{s=0}^{n-1} \left[\binom{n}{s} + \binom{n}{s+1} \right] \frac{d^{n-s}u}{dx^{n-s}} \frac{d^{s+1}v}{dx^{s+1}} + \frac{d^{n+1}u}{dx^{n+1}}v \\ &= u\frac{d^{n+1}v}{dx^{n+1}} + \sum_{s=0}^{n-1} \binom{n+1}{s+1} \frac{d^{n-s}u}{dx^{n-s}} \frac{d^{s+1}v}{dx^{s+1}} + \frac{d^{n+1}u}{dx^{n+1}}v, \end{aligned} \qquad [C.5]$$

where the last equation follows from the binomial coefficient identity

$$\binom{n}{s} + \binom{n}{s+1} = \binom{n+1}{s+1}. \qquad [C.6]$$

Reindexing again with $m = s + 1$ leads to

$$\frac{d^{n+1}}{dx^{n+1}}(uv) = u\frac{d^{n+1}v}{dx^{n+1}} + \sum_{m=1}^{n}\binom{n+1}{m}\frac{d^{n+1-m}u}{dx^{n+1-m}}\frac{d^m v}{dx^m} + \frac{d^{n+1}u}{dx^{n+1}}v, \qquad [C.7]$$

and finally, by placing the two outer terms into the summation,

$$\frac{d^{n+1}}{dx^{n+1}}(uv) = \sum_{m=0}^{n+1}\binom{n+1}{m}\frac{d^{n+1-m}u}{dx^{n+1-m}}\frac{d^m v}{dx^m}, \qquad [C.8]$$

which completes the induction. ■

REFERENCES

Abramowitz, M., and Stegun, I. [1964] *Handbook of Mathematical Functions*, Applied Mathematics Series, vol. 55, Washington: National Bureau of Standards; reprinted (ninth reprinting [1970]) by Dover Publications, New York.

Baker, W.T. [1972] 'An Evaluation of the Traffic Conflicts Technique,' *Highway Research Record 384*, pp. 1-8.

Barton, D.E., and Dennis, K.E. [1952] 'The conditions under which Gram-Charlier and Edgeworth curves are positive definite and unimodal,' *Biometrika*, vol. 39, p. 425.

Beckmann, P. [1973] *Orthogonal Polynomials for Engineers and Physicists*, The Golem Press, Boulder.

Berger, J.O. [1985] *Statistical Decision Theory and Bayesian Analysis*, Second Edition, Springer-Verlag, New York.

Beyer, W.H. [1987] *CRC Standard Mathematical Tables*, CRC Press, Boca Raton.

Billingsley, P. [1979] *Probability and Measure*, John Wiley and Sons, New York.

Burington, R.S. [1973] *Handbook of Mathematical Tables and Formulas*, McGraw-Hill, New York.

Carter, B.D., and Springer, M.D. [1977] 'The Distribution of Products, quotients and powers of independent H-function variates,' *SIAM Journal of Appl. Math.*, vol. 33, No. 4, pp. 542-558.

Charlier, C.V.L. [1931] 'Applications [de la théorie des probabilités] à l'astronomie,' (Part of the *Traité* edited by Borel), Gauthier-Villars, Paris.

Chebyshev, P.L. [1859] 'Sur le développement des fonctions à une seule variable,' *Bulletin de l'Académie Impériale des Sciences de St. Pétersbourg*, vol. 1, pp. 193-200.

Churchill, R.V. [1972] *Operational Mathematics*, McGraw-Hill, New York.

Cook, I.D. [1981] 'The H-function and Probability Density Functions of certain Algebraic Combinations of Independent Random Variables with H-function Probability Distributions,' Ph.D. Dissertation, University of Texas at Austin.

Cornish, E.A., and Fisher, R.A. [1937] 'Moments and cumulants in the specification of distributions,' *Extrait de la Revue l'Institut International de Statistiquc*, 4, pp. 1-14.

Draper, N.R. and Tierney, D.E. [1973] 'Regions of positive and unimodal series expansion of the Edgeworth and Gram-Charlier approximations,' *Biometrika*, vol. 59, p. 463.

Edgeworth, F.Y. [1904] 'The Law of Error,' *Trans. Camb. Phil. Soc.*, **20**, pp. 36 and 113, (with an appendix not printed in the Trans. but included with reprints).

Eldred, B.S. [1979] 'A Methodology for Obtaining the Probability Density Function of the Present Worth of Probabilistic Cash Flow Profiles,' Ph.D. Dissertation, University of Texas at Austin.

Epstein, B. [1948] 'Some applications of the Mellin transform in statistics,' *Ann. Math. Statist.*, vol. 19, pp. 370-379.

Fisher, R.A., and Cornish, E.A. [1960] 'The Percentile Points of Distributions Having Known Cumulants,' *Technometrics*, vol. 2, No. 2, pp. 209-225.

Fox, C. [1961] 'The G and H-Functions as Symmetrical Fourier Kernels,' *Transactions of the American Mathematical Society*, vol. 98, pp. 395-429.

Glauz, W.D., Bauer, K.M., and Migletz, D.J. [1985] 'Expected Traffic Conflict Rates and Their Use in Predicting Accidents,' *Transportation Research Record 1026*, Washington, D.C., pp. 1-12.

Glauz, W.D., and Migletz, D.J. [1980] 'Application of Traffic Conflict Rates and Their Use in Predicting Accidents,' *NCHRP Report 219*, Washington, D.C.

Glennon, J.C. et al. [1977] 'Critique of the Traffic-Conflict Technique,' *Transportation Research Record 630*, pp. 32-38.

Hauer, E. [1978] 'Design Considerations of Traffic Conflicts Surveys,' *Transportation Research Record 667*, pp. 57-66.

Hauer, E. [1982] 'Traffic Conflicts and Exposure,' *Accid. Anal. and Prev.*, vol 14, pp. 359-364.

Hauer, E. and Garder, P. [1986] 'Research into the validity of the traffic conflicts technique,' *Accid. Anal. and Prev.*, vol. 18, pp. 471-481.

Hausdorff, F. [1923] 'Momentprobleme für ein endliches Intervall,' *Mathematische Zeitschrift*, **16**, pp. 220-248.

Hayward, J.C. [1972] 'Near-Miss Determination Through Use of a Scale of Danger,' *Highway Research Record 384*, pp. 24-34.

Higle, J.L., and Witkowski, J.M. [1988] 'Bayesian Identification of Hazardous Locations,' *Transportation Research Record 1185*, pp. 24-36.

Hill, T.W. [1969] 'On Determining a Distribution Function Known only by Its Moments and/or Moment Generating Function,' Ph.D. Dissertation, Arizona State University.

Hille, E. [1926a] 'On Laguerre's Series. First Note,' *Proceedings of the National Academy of Sciences*, vol. 12, pp. 261-265.

Hille, E. [1926b] 'On Laguerre's Series. Second Note,' *Proceedings of the National Academy of Sciences*, vol. 12, pp. 265-269.

Hille, E. [1926c] 'On Laguerre's Series. Third Note,' *Proceedings of the National Academy of Sciences*, vol. 12, pp. 348-352.

Hoel, P.G., Port, S.C., and Stone, C.J. [1971] *Introduction to Probability Theory*, Houghton Mifflin, Boston.

Institute of Transport Economics [1977] 'Proceedings from the First Workshop on Traffic Conflicts,' Oslo, Norway, 142 pp.

Jackson, D. [1941] *Fourier Series and Orthogonal Polynomials*, The Mathematical Association of America, Collegiate Press, Menasha.

Jacobi, C.G.J. [1859] 'Untersuchungen über die Differentialgleichung der hypergeometrischen Reihe,' *Journal für die reine und angewandte Mathematik*, vol. 56, pp. 149-165.

Jambunathan, M.V. [1954] 'Some Properties of Beta and Gamma Distributions,' *Ann. Math. Statist.*, vol. 25, pp. 401-404.

Johnson, N.L. [1949] 'Systems of frequency curves generated by methods of translation,' *Biometrika*, vol. 36.

Keilson, J., and Nunn, W.R. [1979] 'Laguerre Transformation as a Tool for the Numerical Solution of Integral Equations of Convolution Type,' *Appl. Math. Comp.*, vol. 5, pp. 313-359.

Keilson, J., Nunn, W.R., and Sumita, U. [1981] 'The Bilateral Laguerre Transform,' *Appl. Math. Comp.*, vol. 8, pp. 137-174.

Keilson, J., Petrondas, D., Sumita, U., and Wellner, J. [1983] 'Significance Points for Some Tests of Uniformity on the Sphere,' *J. Statist. Comp. Simul.*, vol. 17, No. 3, pp. 195-218.

Keilson, J., and Sumita, U. [1981] 'Waiting Time Distribution Response to Traffic Surges via the Laguerre Transform,' *The Proceedings of the Conference on Applied Probability-Computer Science: The Interface*, Boca Raton, Florida.

Keilson, J., and Sumita, U. [1983] 'The Depletion Time for M/G/1 Systems and a Related Limit Theorem,' *Adv. Appl. Prob.*, vol. 15, pp. 420-443.

Kendall, M., and Stuart, A. [1979] *The Advanced Theory of Statistics*, Volume 2, Fourth Edition, Charles Griffin and Company Limited, London.

Kendall, M., Stuart, A., and Ord, J.K. [1987] *Kendall's Advanced Theory of Statistics*, Volume 1, Fifth Edition, Oxford University Press, New York.

Laguerre, E.N. [1879] 'Sur l'intégrale $\int_x^\infty x^{-1}e^{-x}dx$,' *Bulletin de la Société Mathématique de France*, vol. 7., pp. 72-81.

Larsen, R.J., and Marx, M.L. [1985] *An Introduction to Probability and Its Applications*, Prentice-Hall, New Jersey.

Lew, J.S. [1975] 'On Some Relations between the Laplace and Mellin Transforms,' *IBM Journal of Research and Development*, vol. 19, pp. 582-86.

Lomnicki, Z.A. [1967] 'On the Distribution of Products of Random Variables,' *Royal Statistical Society Journal, Series B*, vol. 29, pp. 513-524.

Moran, P.A.P. [1980] 'Calculation of the normal distribution function,' *Biometrika*, vol. 67, p. 675.

Neuts, M.F. [1981] *Matrix-Geometric Solutions in Stochastic Models*, The Johns Hopkins University Press, Baltimore, Maryland.

Oberhettinger, F. [1974] *Tables of Mellin Transforms*, Springer-Verlag, New York, New York.

Pearson, K. [1895] 'Mathematical Contributions to the Theory of Evolution,' *Philosophical Transactions*, Ser. A, vol. 186, pp. 343-414.

Perkins, S.R., and Harris, J.I. [1968] 'Traffic Conflict Characteristics–Accident Potential at Intersections,' *Highway Research Record 225, HRB*, Washington, D.C., pp. 35-43.

Platzman, L.K., Ammons, J.C., and Bartholdi, J.J. [1988] 'A Simple and Efficient Algorithm to Compute Tail Probabilities from Transforms,' *Operations Research*, **36**, pp. 137-144.

Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T. [1989] *Numerical Recipes*, Cambridge University Press, Cambridge.

Ralston, A. [1965] *A First Course in Numerical Analysis*, McGraw-Hill, New York.

Rider, P.R. [1965] 'Distributions of Product and Quotient of Cauchy Variables,' *Amer. Math. Mon.*, vol. 72, pp. 303-305.

Roberts, G.E., and Kaufman, H. [1966] *Table of Laplace Transforms*, W.B. Saunders Company, Philadelphia.

Rodriguez, O. [1816] 'Mémoire sur l'attraction des spheroïdes,' *Corr. Ec. Roy. Polytech.*, III.

Romanovsky, V. [1924] 'Generalizations of Some Frequency Curves of Professor Pearson,' *Biometrika*, vol. 16, pp. 106-117.

Sansone, G. [1959] *Orthogonal Functions*, Interscience Publishers, Inc., New York.

Shohat, J.A., and Tamarkin, J.D. [1943] *The Problem of Moments*, American Mathematical Society, New York, New York.

Springer, M.D. [1979] *The Algebra of Random Variables*, John Wiley and Sons, New York.

Springer, M.D. and Thompson, W.E. [1964] 'The Distribution of Products of Independent Random Variables,' Tech. Rpt. TR 64-46, General Motors Defense Research Laboratories, Santa Barbara, CA.

Springer, M.D. and Thompson, W.E. [1966] 'The Distribution of Products of Independent Random Variables,' *SIAM Journal of Applied Math.*, vol 14., no. 3, pp. 511-526.

Springer, M.D. and Thompson, W.E. [1970] 'The Distribution of Products of Beta, Gamma, and Gaussian Random Variables,' *SIAM Journal of Applied Math.*, vol. 18., no. 4, pp. 721-737.

Sumita, U. [1979] 'Numerical Evaluation of Multiple Convolutions, Survival Functions for the One-Sided Normal Distribution and the Rayleigh Distribution via the Laguerre Transformation,' Working Paper Series No. 7912, Graduate School of Management, University of Rochester.

Sumita, U. [1981] 'Development of the Laguerre Transform Method for Numerical Exploration of Applied Probability Models,' Ph.D. Dissertation, Graduate School of Management, University of Rochester.

Sumita, U. [1983] 'On Sums of Independent Logistic and Folded Logistic Variants,' *J. Statist. Comp. Simul.*, vol. 17, No. 4, pp. 251-274.

Szegö, G. [1975] *Orthogonal Polynomials*, American Mathematical Society, Providence.

Thiele, T.N. [1903] *Theory of Observations*. Reprint in *Ann. Math. Statist.*, vol. 2, No. 165, of the English version published in 1903; the original (Danish) appeared in 1889 and 1897.

Uspensky, J.V. [1927] 'On the Development of Arbitrary Functions in Series of Hermite's and Laguerre's Polynomials,' *Annals of Mathematics* (2), vol. 28, pp. 593-619.

Wells, W.T., Andersen, R.L., and Cell, J.W. [1962] 'The Distribution of the Product of two central or non-central Chi-square variates,' *Ann. Math. Statist.*, vol. 36, pp. 1016-1020.

Williams, M.J. [1981] 'Validity of the traffic conflicts technique,' *Accid. Anal. and Prev.*, vol. 13, pp. 133-145.

Zegeer, C.V. and Deen, R.C. [1978] 'Traffic Conflicts as a Diagnostic Tool in Highway Safety,' *Transportation Research Record 667*, pp. 48-55.