The Dissertation Committee for Robert Kincaid Hammond Certifies that this is the approved version of the following dissertation:

Discrete Approximations to Continuous Distributions in Decision Analysis

Committee:

J. Eric Bickel, Supervisor

James Dyer

John Hasenbein

Larry Lake

David Morton

Discrete Approximations to Continuous Distributions in Decision Analysis

by

Robert Kincaid Hammond, B.S.E.E., M.S.E.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May, 2014

Dedication

To my parents.

Acknowledgements

This work would not have been possible without the mentorship, support, and guidance of my advisor, J. Eric Bickel. He taught me decision analysis and how to do research. I would like to thank David Morton, John Hasenbein, Larry Lake, and Jim Dyer for serving on my doctoral committee and for their many helpful conversations and guidance. The work on PERT formulae was a collaborative effort with Seong Dae Kim. I was fortunate to work with him on a topic in which we could combine our respective expertise to create a new contribution to the PERT literature.

This research was funded by the Center for Petroleum Asset Risk Management (CPARM), but the support of this organization extends beyond funding. Participation in the CPARM events and annual meetings, organized by Emilio Nunez and Larry Lake, gave excellent opportunities to interact with other professors and industry professionals from the sponsor companies, and to hear their feedback on my work. I am extremely grateful for both the funding and the opportunities presented by CPARM.

For excellent and valuable experiences in industry, as well as for their support of this work, I also thank Victor Koosh, Larry Neal, and Bryan Cotner.

The support I received from friends and family cannot be overstated. I thank my fiancé, Christy Treat, for standing by my side with incredible love and support. Long-time family friend Loring “Red” White introduced me to operations research and decision analysis, and has given me advice and feedback throughout my graduate journey.

Discrete Approximations to Continuous Distributions in Decision Analysis

Robert Kincaid Hammond, Ph.D.
The University of Texas at Austin, 2014

Supervisor: J. Eric Bickel

In decision analysis, continuous uncertainties (i.e., the volume of oil in a reservoir) must be approximated by discrete distributions for use in decision trees, for example. Many methods of this process, called discretization, have been proposed and used for decades in practice. To the author's knowledge, few studies of the methods' accuracies exist, and were of only limited scope. This work presents a broad and systematic analysis of the accuracies of various discretization methods across large sets of distributions. The results indicate the best methods to use for approximating the moments of different types and shapes of distributions. New, more accurate, methods are also presented for a variety of distributional and practical assumptions. This first part of the work assumes perfect knowledge of the continuous distribution, which might not be the case in practice. The distributions are often elicited from subject matter experts, and because of issues such as cognitive biases, may have assessment errors. The second part of this work examines the implications of this error, and shows that differences between some discretization methods' approximations are negligible under assessment error, whereas other methods' errors are significantly larger than those because of imperfect assessments. The final part of this work extends the analysis of previous sections to applications to the Project Evaluation and Review Technique (PERT). The accuracies of several PERT formulae for approximating the mean and variance are analyzed, and several new formulae presented. The new formulae provide significant accuracy improvements over existing formulae.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

BACKGROUND

Many decisions in both personal and business environments must be made without complete knowledge of the situation. This incompleteness of knowledge about a past event, fact, or the outcome of a future event is called *uncertainty*. The prospects, or potential future states, of the decisions depend on the actual values of these unknown factors. The presence of uncertainty is often the factor separating the difficult decisions from the trivial (Howard, 1989). Decision analysis theory and methods provide a philosophy and logical framework for making sound decisions under the conditions of uncertainty, scarce resources, and competing objectives. Knowledge about unknown parameters or *uncertain quantities* can be quantitatively described using probability distributions, and this knowledge can be applied to reasoning using probability theory.

The possible values uncertain quantities can realize may be continuous, such as the volume of hydrocarbons in a reservoir or fraction of market share, or discrete, such as the presence of hydrocarbons or the winner of an election. Continuous uncertainties are modeled with a probability density function (pdf), and discrete uncertainties with a probability mass function (pmf). Decision trees are a popular tool for modeling the various uncertainties and decisions of a decision problem, but can represent only discrete scenarios. In order to represent continuous uncertainties in a discrete decision tree, it is common for decision analysts to represent, or approximate, pdfs with properly designed pmfs, a process known as *discretization*.

Monte Carlo methods and more generally numerical integration methods approximate the values of complicated functions by a weighted sum of discrete values. Algorithms such as these utilize discrete representations of continuous values to make intractable calculations tractable and difficult ones easier. Decision trees are one such algorithm, and continuous uncertainties (and decision variables) are discretized to simplify the tree-rollback procedure. Additionally, the computational time requirements associated with complex models, such as reservoir simulations or large portfolio optimizations, often limit the number of model evaluations. In these cases, it is more economical to carefully choose a small set of input values, rather than generate a large number of random scenarios via Monte Carlo simulation.

In addition to easing the computational burden, discretization in decision analysis aids in understanding of the problem by defining distinct scenarios, and can simplify the work of assessing a probability distribution supplied by a decision-maker or subject-matter expert.

Discretization methods are designed to preserve certain properties of the continuous distribution, such as moments or percentiles. This goal is a means to the ultimate end of accurately representing the important characteristics of the output distribution of a decision model. This output is generally a distribution on a value measure, such as net present value (NPV) or utility. A common decision metric is the expectation of this value measure, such as expected NPV (ENPV), or the certainty equivalent (CE) of expected utility (Howard, 1971). The next chapter discusses the importance of matching input distribution moments to the output distribution expectation.

Although our focus lies in discretizing probability distributions, the space of alternatives for a decision variable may also be continuous. This raises similar issues to that of probability discretization, of choosing a discrete set of alternatives to represent a continuous set. Merkhofer (1975) considered these issues in detail.

MOTIVATION

The existence of many discretization methods naturally raises the question as to which method is best in general or for a particular situation. Several prior studies, such as Keefer and Bodily (1983), Keefer (1994), and Smith (1993), indicated that some methods are consistently superior to others, but were limited in the methods considered, the metrics used for evaluation, and the scope of distributions on which they were tested. Other work focused on separately estimating the means and variances of continuous distributions (Perry and Greig, 1975; Keefer and Verdini, 1993; Johnson, 1998; Lau and Lau, 1998; Lau et al., 1998; Shankar et al., 2010). However, this work is often not directly applicable to discretization because separate formulas are typically used for mean and variance estimation and are not always in a form consistent with a probability distribution. This work is, however, related, and is considered in this work in Chapter 6.

Several methods have been proposed, but few have been systematically analyzed, and none in the scope considered here. Many of these can preserve the location and scale (mean and variance, respectively) of many distributions quite well but are quite poor in preserving higher moments, such as the skewness and kurtosis, that define a distribution's shape. Some methods were designed with a specific distribution in mind and applied indiscriminately, while others make no assumptions about the underlying distribution a priori. We examine the questions of which method is best, when, and why.

Despite varying degrees of accuracy, the magnitudes of differences in accuracy, particularly for the mean of a distribution, can be small for the best methods. This fact raises the question of whether these differences are significant when the underlying continuous distribution may not be precisely known. In practice, the continuous distributions that are discretized are assessed from the subjective judgment of subject-matter experts. Due to disagreement between experts, inconsistencies in assessments, and cognitive biases, these assessments may not be accurate representations of the expert's knowledge, a phenomenon we call *assessment error*. Although an expert gives a specific estimate for a distribution percentile, he or she may accept any value within a small neighborhood around their assessment as consistent with their information.

Clearly, accurate approximations to continuous distributions are important for accurate decision models, but they are not the ultimate goal. As mentioned above, the goal of decision analysis is to achieve the clarity required to make a good decision. The decision analysis cycle in Figure 1 is a structured approach to achieving this clarity through iterative model structuring, analysis, and refinement (Matheson and Howard, 1968). Simple, less accurate, discrete approximations may be sufficient in the early stages of analysis, but their flaws might be shown by insights from the model. Uncertainties with more impact on the value function should be approximated in the model with enough accuracy to represent the characteristics that are important to the decision.

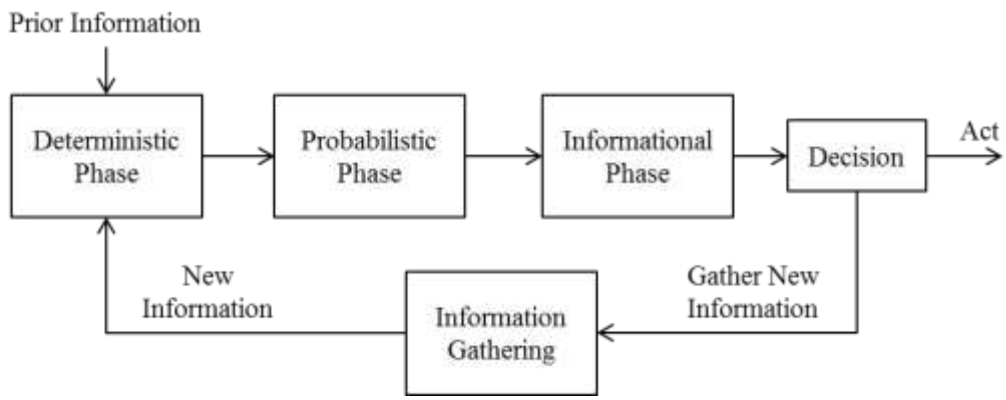

Figure 1. The Decision Analysis Cycle.

CONTRIBUTIONS

This dissertation makes several contributions to the literature on probability discretization in decision analysis applications. We extend previous research by considering a wide range of distributions, through the use of the Pearson and Johnson distribution systems, which cover several commonly-used distribution families. We analyze discretization methods that have not been previously considered but that are now in common use, and we suggest several new discretizations that are tailored to specific distribution families. Error analysis is performed over a wide range of distributions from the two distribution systems. The author is not aware of any systematic discretization error analysis of as wide a scope as performed here.

The results show that even some quite simple methods can closely approximate the means and variances of a wide range of distributions. These same methods, however, provide very poor estimates of higher moments, such as skewness and kurtosis, that depend much more on the distribution tails. Most discretization methods we consider, and all of the methods typically in use, poorly represent these tails, hence the higher moments. Simple discretizations can accurately preserve the *location* and *scale* of a distribution, but they are generally poor approximations of the *shape*. Our results give insight into the distributions for which different methods perform best, and the reasons for this. The results provide a guide for choosing discretization methods in practice. We construct new methods by optimizing the average performance in matching the mean over wide sets of distributions having distinct qualities. In some cases, our new methods provide distinct improvements over existing methods, and in others demonstrate that the existing methods are already very close to optimal according to mean-squared error.

To analyze the effects of imperfectly assessed distributions, we start by presenting a model of probability assessment error based on the notion of a truth set. Although a discretization method may clearly outperform another in terms of discretization error, in many cases the presence of assessment error overwhelms this difference in performance. However, some methods' discretization errors are large enough to be significant under large assessment error. Our approach to assessment error is new and addresses practical issues with subjectively assessed distributions. We compare a selection of discretization methods, encompassing different types of methods with widely varying accuracies. Our model of assessment error in the percentile assessments is consistent with the errors reported in much of the probability assessment literature.

Other than decision analysis, the Program Evaluation and Review Technique (PERT) frequently utilizes approximations to continuous distributions. Rather than forming discrete approximations to distributions, PERT approximations typically estimate the mean and variance using separate formulae that are functions of distribution characteristics, such as percentiles, endpoints, or the mode. We extend the formulation we use to find new discretization shortcuts to construct new mean and variance approximation formulae that are more accurate than previously proposed formulae. We also demonstrate that the mode is a poor characteristic to use in these formulae.

ORGANIZATION

This work is organized as follows. The next chapter reviews the main discretization methods discussed in the literature, give a brief history of their development, and presents our new methods. Chapter 3 reviews the Pearson and Johnson distribution systems we use for analysis. Chapter 4 presents the analysis of discretization error. Chapter 5 presents the extension to include assessment error. Chapter 6 gives an application of our results to the form of methods used by PERT methods. Chapter 7 concludes and presents further questions and directions for future work.

Chapter 2: Discretization Methods

INTRODUCTION

This chapter reviews the main discretization methods discussed in the literature and gives a brief history of their development. We present several new discretization methods. We briefly review the use of discretization in other fields. We also review probability elicitation.

We broadly categorize the methods as either *distribution-specific*, which require the full distribution, or as a *shortcut*, which only requires a few predefined percentiles. Shortcuts use (often three) fixed percentiles such as the 10^{th} (P10), 50^{th} (P50), and 90^{th} (P90). Distribution-specific methods tailor the percentiles to the shape of the underlying pdf, and they typically give the analyst freedom to choose the number of discrete points. An advantage of shortcut methods is that the full pdf need not be known, requiring only the percentile assessments. In contrast, distribution-specific methods require the full pdf, but this gains modeling flexibility.

Specifically, suppose we are constructing a decision model that takes as an input the continuous random variable X with support $S \subseteq \mathbb{R}$. For a realization $x \in S$ of this uncertainty and a vector of decision variables $\boldsymbol{w}$, the model has value output $v(\boldsymbol{w},x)$. Discretization approximates the pdf $f(x)$, or the cumulative distribution function (cdf) $F(x)$, with a set of values $x_i \in S$, $i = 1, 2, \ldots, N$, and associated probabilities $p_i \equiv p(x_i)$. In most discretization methods, N is equal to three, but five is not uncommon. We denote the discretization operation as $D_d(f)$ for continuous pdf f and discretization method d, and the resulting pmf as $g = D_d(f)$ with corresponding cdf G.

The values and probabilities are chosen so that g preserves certain properties of f. The moments of f are the most common, but aspects of the cdf curve F may also be used. Various measures of accuracy are discussed later in this chapter.

DISTRIBUTION-SPECIFIC METHODS

Distribution-specific methods require that the full distribution be known. The methods we consider can be divided into *bracket methods* and *quadrature methods*.

Bracket Methods

Bracket methods partition the distribution into mutually exclusive intervals, or brackets, $B_i \in S, \bigcup_{i=1,\ldots N} B_i = S, B_i \cap B_j = \emptyset, i, j = 1,\ldots N, i \neq j$, representing each interval with a single value-probability pair, $\{x_i, p_i\}$. The intervals can be defined for discretizing a random variable X by partitioning either the support of X or partitioning the cumulative probability scale, which the cdf F of X transforms into a partition on the support. The second convention is generally easier to work with, since it is independent of the support of *X*, and it is more common in the decision analysis literature.

The bracket methods Bracket Mean (BMn) and Bracket Median (BMd) represent intervals by their means and medians, respectively. The analyst is not restricted to using only means or medians, but these methods are straightforward to apply without requiring any calculation.

Bracket Mean

BMn, also called the "equal areas" method was originally developed by Jim Matheson and his colleagues at the Stanford Research Institute from the late 1960s to the early 1970s (Bickel et al., 2011). BMn represents brackets by their conditional means,

$$x_i = E\left[X \mid x \in B_i\right], \tag{1}$$

and weights them by the probability represented by the bracket,

$$p_i = \int_{x \in B_i} f\left(x\right) dx\,. \tag{2}$$

The method is more adaptable to the underlying distribution than are shortcuts and is also easier to apply than the Gaussian quadrature method described below. BMn perfectly matches the mean of the underlying distribution, but it always underestimates the variance and higher even-moments (Miller and Rice, 1983). The next chapter will show that BMn's error in the variance, for three- and five-point bracket examples, tends to be higher than some three-point discretization shortcuts.

Part of BMn's appeal stems from the fact that it can be graphically explained and applied to a distribution for which the cdf, but not necessarily the functional form, is available (Merkhofer, 1975). For example, Figure 2 shows the cdf of a standard-normal distribution, with the cumulative scale divided into three intervals: [0.0, 0.25], (0.25, 0.75], and (0.75, 1.0]. To find the conditional mean of the first bracket, which we will use as the discretization point x_1, find the location of the vertical line that makes the shaded areas A_1 and A_2 equal. The X value at this point is conditional mean. Repeat for each interval, assigning the probability of an interval to the corresponding point as the discretization. Here, the points -1.24, 0, and 1.24 are respectively assigned probabilities 0.25, 0.50, and 0.25. For a proof of this method, see (Merkhofer, 1975) or McNamee and Celona (1991).

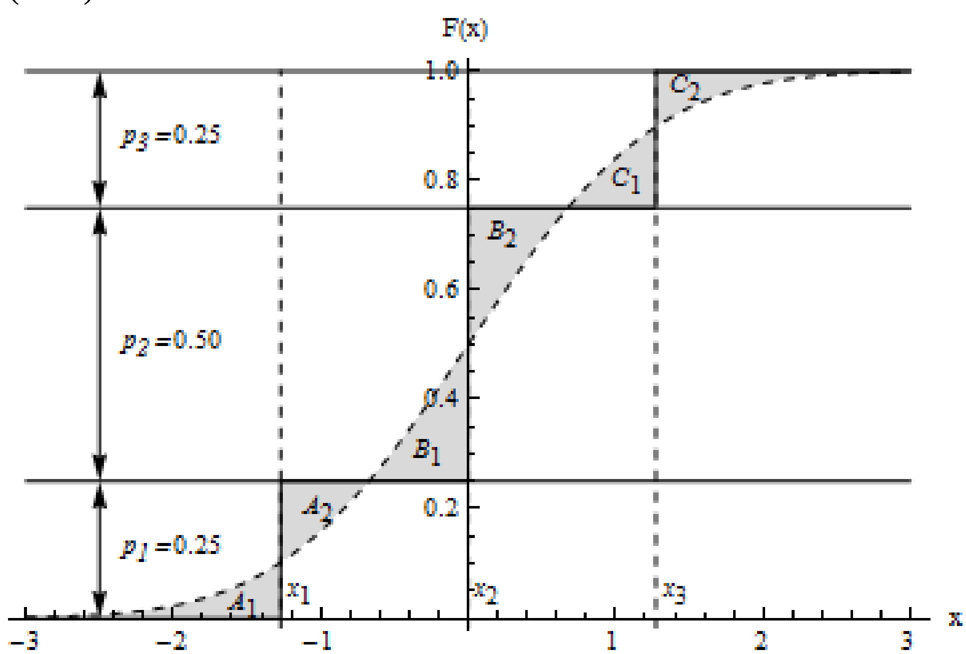

Figure 2. An example application of Bracket Mean with intervals [0, 0.25], (0.25, 0.75], (0.75, 1] on the standard normal distribution.

Bracket Median

BMd uses the conditional *medians* of the partitions of the support. If partitioning on the cumulative probability axis, the median of a bracket can be determined simply by the midpoint of that interval on the cumulative probability scale (a percentile). Using the brackets of Figure 2, the median of the bracket [0, 0.25] is the 0.125 percentile, or the P12.5. Because the conditional distributions are generally skewed, the median and the mean differ and BMd will regularly result in different discretizations from BMn

BMd is frequently recommended in decision analysis texts (Schlaifer, 1969; Holloway, 1979; Clemen, 1991). Although the method is simple to apply, it typically incurs large errors in the moments. The next chapter will show that three- and five-point BMd discretizations using equal-sized brackets tend to perform far worse than three- and five-point BMn in matching the mean and variance of a distribution.

Quadrature Methods[1]

These methods originate from numerical integration methods. The most widely used of these in the decision analysis literature is Gaussian quadrature (GQ), which approximates an integral with a sum of discrete points and weights. For a function $\Omega(x)$ and pdf f(x), the values x_i and probabilities p_i are found for the approximation

$$\int_a^b f(x)\Omega(x)dx \approx \sum_{i=1}^{n} p_i \Omega(x_i), \tag{3}$$

which is exact for functions $\Omega(x)$ that are polynomials of degree *2n-1* or less. An *n*-point GQ preserves the 0^{th} through the $(2n\text{-}1)^{st}$ moments, which Stroud (1974) showed to be the highest degree achievable by an *n*-point approximation. Gauss in 1816 studied the special case of [*a*, *b*] = [-1, 1], and integration rules of this form became known as *Gaussian quadrature formulae*. Formulae for many common weighting functions are tabulated; see for example, Stroud and Secrest (1966). Miller and Rice (1983) first described GQ as a

[1] This section borrows heavily from Smith (1990).

method for discretizing continuous probability distributions to reduce the complexity of decision problems. Smith (1990, 1993) gave an alternative method for constructing GQ discretizations, which utilizes the theory of orthogonal polynomials described in Stroud and Secrest (1966). We review both methods below.

A GQ formula is constructed in two steps: finding p_i and x_i. First, a polynomial is constructed whose roots are the points of the discrete approximation. Second, the probabilities are found by solving the linear system of equations formed by setting the first *n* moments (including the 0th moment) of the discrete approximation equal to the corresponding moments of *f*.

Miller and Rice's Method

The basic approach described by Miller and Rice (1983) is to solve for p_i and x_i in the *2n* equations

$$\int_{-\infty}^{\infty} x^k f(x)dx = m_k = \sum_{i=1}^{n} p_i x_i^k \quad k = 0,\ldots,2N-1. \tag{4}$$

This is accomplished by finding x_i as the roots of a polynomial

$$\pi(x) = \prod_{i=1}^{n}(x - x_i) = \sum_{k=0}^{n} c_k x^k. \tag{5}$$

The coefficients c_k are determined from a set of *n* linear equations formed from (4):

$$\sum_{i=1}^{n} p_i x_i^j \pi(x_i) = 0 = \sum_{k=0}^{n} c_k m_{k+j} \quad j = 0,\ldots,n-1. \tag{6}$$

Having found the points x_i, the weights p_i are determined from the linear system of equations formed by the first *n* equations of Equation (4).

Smith's Method

The methodological simplification described by Smith (1990) utilizes the theory of orthogonal polynomials. Two polynomials *P(x)* and *Q(x)* are *orthogonal* with respect to *f* if

$$\int_{-\infty}^{\infty} P(x)Q(x)f(x)dx = 0. \tag{7}$$

Denote by $P_n(x)$ a degree n polynomial that is orthogonal to all polynomials of degree less than n:

$$\int_{-\infty}^{\infty} P_n(x)x^k f(x)dx = 0 \quad k = 0,...,n-1, \tag{8}$$

which is unique up to a scaling constant. Denote by $P_n^*(x)$ a degree n *orthonormal* polynomial, which is an orthogonal polynomial scaled so that

$$\int_{-\infty}^{\infty} P_n^*(x)P_k^*(x)f(x)dx = \begin{cases} 0, & k \neq n, \\ 1, & k = n. \end{cases} \tag{9}$$

An orthogonal polynomial is exactly the polynomial for which we computed the constants in (6). The orthonormal polynomials $P_0^*,...,P_n^*$ with respect to f are related by the recurrence relation (using Smith (1993)'s notation):

$$\begin{aligned} & P_{-1}(x) = 0; P_0(x) = 1, \\ & P_{i+1}(x) = \left(x - E\left[xP_i^{*2}(x)\right]\right)P_i^*(x) - E\left[P_i^2(x)\right]^{1/2} P_{i-1}^*(x) \\ & \text{for} \quad 1 \leq i \leq n-1 \\ & \text{where} \quad P_i^*(x) = \frac{P_i(x)}{E\left[P_i^2(x)\right]^{1/2}}. \end{aligned} \tag{10}$$

E denotes the expectation with respect to F. The n roots of P_n^*, $y_1,..., y_n$ are distinct and provide the n discretization values. The discretization points y_i are the roots of $P_n^*(x)$, which can be computed using the Newton-Raphson method, for example. The probability weight associated with each point is given by

$$p_i = \frac{k_n}{k_{n-1}P_n^{*\prime}(x_i)P_{n-1}^*(x_i)}, \tag{11}$$

where k_n and k_{n-1} are the leading coefficients of $P_n^*(x)$ and $P_{n-1}^*(x)$, respectively, and $P_n^{*\prime}(x)$ denotes the derivative of $P_n^*(x)$.

These approaches require only the moments of *f*. The theory of orthogonal polynomials guarantees that the roots are contained within the support of *f* and that $\sum_{i=1}^{n} p_1 = 1$, which thus assures that the result is a valid pmf representation of *f*.

GQ has the highest possible moment accuracy of any discrete approximation method, but it has several drawbacks. First, it is by far the most complex method we consider, and its calculations necessitate software implementation. Smith's approach avoids solving the system of linear equations in the Miller and Rice method, but it still requires finding the roots of an *n*-degree polynomial. Second, GQ is limited to distributions for which the 2*n*-1 moments exist. Finally, GQ tends to use points far out in the tails of a distribution, in order to match higher moments. These points correspond to extreme percentiles, such as the P99.9, which are often unreliable for subjectively assessed distributions (Lichtenstein et al., 1982).

SHORTCUT METHODS

Shortcut methods use only a few points from the cdf, making them often the easiest to use in practice, since only these points must be elicited or determined and then appropriately weighted in the decision tree. A shortcut is defined by a set of percentiles $\{q_i\}$ from the pdf and a corresponding set of probabilities $\{p_i\}$. The q•100% percentile for a cdf *F* is the value $F^{-1}(q)$. Shortcut values and probabilities will be summarized in the form, (q_1, q_2, q_3, p_1, p_2, p_3). Shortcuts are typically symmetric, i.e., for a three-point discretization, $p_1 = p_3$ and $F^{-1}(q_1) = 1 - F^{-1}(q_3)$.

Some of the earliest work on shortcut-like methods was that of Pearson and Tukey (1965), who tested many approximation formulae that preserve either the mean or the variance of a set of distributions. They recommended a symmetric three-point approximation for the mean by weighting the P5, P50, and P95 with 0.185, 0.630, and 0.185, respectively. Keefer and Bodily (1983) suggested treating Pearson and Tukey's approximation as a full pmf and referred to it as Extended Pearson-Tukey (EPT).

Examples in the literature include Keeney (1987) in selecting sites for nuclear waste repositories and Brooks and Kirkwood (1988) in selecting a strategy for microcomputer networking. Clemen and Reilly (1999) and Wang and Dyer (2012) used it to discretize Gaussian copulas.

Another mean approximation formula, called "Swanson's Mean," weights the P10, P50 and P90 percentiles of $F(x)$ by 0.300, 0.400, and 0.300, respectively. Roy Swanson proposed, in a 1972 internal Exxon memo, that the mean of a lognormal distribution can be approximated by weighting the P10, P50, and P90 by 0.300, 0.400, and 0.300, respectively (Hurst et al., 2000). Keefer and Bodily (1983) also suggested treating this approximation as a full pmf and called it Extended Swanson-Megill (ESM), to be used with many distribution families in addition to the lognormal. However, Megill (1984) stressed that it should not be used for highly skewed lognormal distributions. Example applications of ESM in the literature include valuation of a refinery (Keefer, 1995), evaluating strategies for meeting environmental regulations (Bailey et al., 2000), decision-making in pharmaceuticals (Stonebraker, 2002; Stonebraker and Keefer, 2009), and bidding (Bailey et al., 2011). Willigers (2009) compared EPT as an alternative to ESM in a portfolio modeling application. ESM is heavily used within the oil & gas industry (Megill, 1984; Hurst et al., 2000; Rose, 2001b; Bickel et al., 2011).

McNamee and Celona (1990), SDG consultants at the time, described another shortcut, which has come to be known as the McNamee-Celona Shortcut (MCS) or the "25-50-25" shortcut. MCS uses (P10, P50, P90, 0.250, 0.500, 0.250). It is based on both MRO and the application of the BMn method. BMn applied to the normal distribution, using the brackets [0, 0.25], [0.25, 0.75], and [0.75, 1], yields the P10.2, P50, and P89.8 percentiles. McNamee and Celona cautioned that MCS was only a first approximation in analyzing a decision problem and that the distributions should be encoded and discretized more carefully as the analysis progressed. MCS has received little attention in the literature, but it is common among practitioners, particularly in the oil & gas industry (Bickel et al., 2011).

Miller and Rice (1983) introduced several shortcuts based on an application of Gaussian quadrature, for situations when the full pdf is not known, as might be the case when one assesses a cdf directly from an expert. Their three-point method (P8.5, P50, P91.5, 0.248, 0.504, 0.248) is known as the Miller and Rice One-Step (MRO).

D'Errico and Zaino (1988) proposed, and Zaino and D'Errico (1989) analyzed, two approximations based on the Taguchi (1978) method. The first (ZDT) uses equal weights (P11, P50, P89, 0.333, 0.333, 0.333). The second (ZDI) is a three-point Gaussian quadrature for a normal distribution, (P4.2, P50, P95.8, 0.167, 0.667, 0.167). Zaino and D'Errico (1989) called ZDI an improvement over ZDT and found the former to be more accurate in their experience.

These shortcuts are summarized in Table 1 below. Each shortcut uses the P50 (the *median*) and is symmetric about the median. They differ only in the distance from the symmetric outer percentiles to the median and in the probability weight on the median. Of these, EPT generally best preserves a distribution's mean and variance. ESM is generally best if the P10, P50, and P90 are used. Although EPT is more accurate, ESM's less extreme percentiles may be easier for experts to assess. MCS, though widely used, is generally inferior to ESM.

Shortcut	Discretization Values	Probability Weights
Extended Swanson-Megill (ESM)	P10, P50, P90	0.300, 0.400, 0.300
McNamee-Celona Shortcut (MCS)	P10, P50, P90	0.250, 0.500, 0.250
Extended Pearson-Tukey (EPT)	P5, P50, P95	0.185, 0.630, 0.185
Zaino-D'Errico "Improved" (ZDI)	P4.2, P50, P95.8	0.167, 0.667, 0.167
Zaino-D'Errico-Tagichi (ZDT)	P11, P50, P89	0.333, 0.333, 0.333
Miller-Rice One-Step (MRO)	P8.5, P50, P91.5	0.248, 0.504, 0.248

Table 1. Summary of discretization shortcut methods.

NEW SHORTCUT METHODS

The shortcuts previously discussed do not typically have broad foundations. For example, ESM was based on the lognormal distribution, and MCS on the normal. EPT, on the other hand, was based on a wide but sparse set of distributions, covering many shapes. The Pearson and Johnson systems can be used to systematically incorporate basic information about a distribution to refine the discretization.

We follow the approach of Lau et al. (1998), who found three-, five-, and seven-point mean-approximation formulae by finding, for a given set of percentiles, the probabilities that minimize the average squared error over the entire set of distributions (i.e., all the Pearson types). Applying a similar approach to *individual* regions of the Pearson and Johnson systems gives mean-approximation formulae which, like ESM and EPT, can be used as full pmfs.

The Pearson system, described in more detail later, subsumes many common distributions (e.g., normal, beta, uniform, exponential) and is divided into several types. Three main types make up the majority of Pearson distributions, the rest being special and limiting cases. The primary distinction among the three major regions (i.e., I, VI, and IV) of the Pearson system is their range of support: the type I has bounded support, $[a, b]$, the type VI support is unbounded in one direction, $[a, \infty]$ or $[-\infty, b]$, and the type IV has unbounded support, $[-\infty, \infty]$. Type I can be further divided into three distinct shapes: U-shapes, J-shapes, and ∩-shapes. The Johnson system also has three types: the bounded S_B, similar to the Pearson type I, the S_L, or generalized lognormal distribution, unbounded in one direction, and the unbounded S_U.

This allows the analyst to tailor a shortcut to a distribution using only knowledge of the support boundedness. This basic characteristic of uncertainty is often easy to determine, allowing focus on a particular Pearson region and thereby choice of the best shortcut.

Lau et al. (1998) found their three-point approximations by solving the following regression problem over *n* distributions:

$$\begin{aligned} &\min_{p_1, p_2} \frac{1}{n} \sum_{i=1}^{n} (\delta_i^1)^2 \\ &\text{s.t.} \quad \hat{m}_i^1 = p_1 F_i^{-1}(\alpha_1) + p_2 F_i^{-1}(0.5) + p_3 F_i^{-1}(\alpha_3) \quad i = 1, \ldots, n \\ &\qquad p_1 + p_2 + p_3 = 1 \\ &\qquad p_3 = p_1 \\ &\qquad \alpha_3 = 1 - \alpha_1 \\ &\qquad 0 \le p_1, p_2, p_3 \le 1. \end{aligned} \tag{12}$$

This procedure yields a symmetric discretization, i.e., where $p_3 = p_1$ and $\alpha_3 = 1 - \alpha_1$. The objective function minimizes the average squared error (ASE) over the *n* distributions. Lau et al. (1998) found that for $\alpha_1 = 0.05$, the best three-point mean-approximation formula is (P5, P50, P95, 0.179, 0.642, 0.179). This shortcut has very similar probabilities to EPT (P5, P50, P95, 0.185, 0.630, 0.185), which also was constructed using points in similar areas of the (β_1, β_2) plane. However, Lau et al. (1998) more densely covered this area, particularly the type IV region.

For each distribution system, we construct three sets of shortcuts. One set of shortcuts, which we call EPT+, determines the best percentiles and probabilities for symmetric shortcuts. The second, EPT++, relaxes the symmetry requirement. The third set, the standard percentile (SP) methods, fixes the points to be the P10, P50, and P90, since these are commonly used in practice, and fits the probabilities.

EPT Extensions

Both Pearson and Tukey (1965) and Lau et al. (1998) considered a limited set of symmetric discretizations. Lau et al. (1998) considered four sets of percentiles $(\alpha_1 = 0.01, 0.05, 0.10, 0.25)$, based on the probability elicitation literature. Pearson and Tukey (1965) investigated a different set of percentiles $(\alpha_1 = 0.005, 0.01, 0.025, 0.05)$. In this section, we vary α_1 over a larger, more complete set of percentiles, from P1 to P20 in increments of 1%, first maintaining that the discretization must be symmetric, and then relaxing this requirement. Since these shortcuts extend Pearson and Tukey's work, we refer to them as "EPT+" discretizations and add an identifier that specifies the Pearson or Johnson region for which a specific discretization is optimized.

Symmetric Type-Specific Shortcuts

For each set of percentiles, we find the probabilities that minimize the ASE. The error-minimizing shortcut discretizations for the type I, IV, and VI distributions are given in Table 2. For some of these shortcuts, as well as those in the tables presented later, the fitted probabilities did not sum to one after rounding to three decimal places. They were then adjusted to sum to one by successively adding 0.001 to the rounded probabilities in the descending order of the amount lost to rounding (if the probability was rounded down). We performed the same procedure on the three Johnson types, yielding the shortcuts in Table 3. Here and for the other two sets of new shortcuts, the Johnson S_B distributions do not include their U-shapes.

Shortcut	Pearson Distribution Type	Discretization Values	Respective Probabilities
EPT1U+	I-U (Beta)	P15, P50, P85	0.296, 0.408, 0.296
EPT1J+	I-J (Beta)	P6, P50, P94	0.203, 0.594, 0.203
EPT1∩+	I-∩ (Beta)	P5, P50, P95	0.184, 0.632, 0.184
EPT3+	III (Gamma)	P5, P50, P95	0.184, 0.632, 0.184
EPT6+	VI (Beta Prime)	P4, P50, P96	0.164, 0.672, 0.164
EPT5+	V (Inverse Gamma)	P4, P50, P96	0.163, 0.674, 0.163
EPT4+	IV	P6, P50, P94	0.212, 0.576, 0.212

Table 2. Symmetric type-specific discretization shortcuts for the Pearson system (EPT+ methods).

Shortcut	Johnson Distribution Type	Discretization Values	Respective Probabilities
EPTS_B+	S_B (unimodal only)	P5, P50, P95	0.182, 0.636, 0.182
EPTS_L+	S_L	P4, P50, P96	0.163, 0.674, 0.163
EPTS_U+	S_U	P4, P50, P96	0.163, 0.674, 0.163

Table 3. Symmetric type-specific discretization shortcuts for the Johnson system (EPT+ methods).

In Table 2, the EPT1∩+ and EPT3+ shortcuts uses the same percentiles and probabilities (rounded to three decimal places) as EPT. This is partly due to Pearson-Tukey's heavy sampling of this region. EPT6+ and EPT5+ use the same percentiles and almost exactly the same probabilities as ZDI, perhaps because these distributions are unbounded above and ZDI is the Gaussian quadrature for the normal distribution. The EPT1U+ and EPT4+ percentiles do not resemble the percentiles of any of the pre-existing shortcuts that we consider, although EPT1U+ uses almost exactly the same probabilities as ESM and the EPT4+ probabilities are similar to MCS.

Asymmetric Type-Specific Shortcuts

We now relax the constraint that the discretizations must be symmetric, but we still require that one point be the P50. We again consider values for the lower (upper) percentiles from P1 (P99) to P20 (P80) in increments of 1%. The resulting EPT++ shortcut methods are shown in Table 4 for the Pearson system and Table 5 for the

Johnson system. The type I-∩ and VI shortcuts have nearly the same percentiles and probabilities as the symmetric shortcuts in Table 2, implying that three points (that include P50) would not better approximate these pdfs. The type I-J and IV shortcuts are similar to their symmetric counterparts in their percentiles and probabilities. However, EPT1U++ and the method for the two transition types, EPT3++ and EPT5++, are significantly altered by allowing for asymmetry. In the Johnson system, $EPTS_L$+ and $EPTS_L$++ are identical. $EPTS_B$++ has very similar percentiles, though different probabilities, from $EPTS_B$+. $EPTS_U$++ differs even more from $EPTS_U$+, in both percentiles and probabilities.

Shortcut	Distribution Type	Discretization Values	Respective Probabilities
EPT1U++	I-U (Beta)	P1, P50, P85	0.216, 0.491, 0.293
EPT1J++	I-J (Beta)	P2, P50, P94	0.184, 0.615, 0.201
EPT1∩++	I-∩ (Beta)	P5, P50, P95	0.184, 0.632, 0.184
EPT3++	III (Gamma)	P2, P50, P99	0.110, 0.793, 0.097
EPT6++	VI (Beta Prime)	P4, P50, P96	0.164, 0.672, 0.164
EPT5++	V (Inverse Gamma)	P9, P50, P95	0.239, 0.565, 0.196
EPT4++	IV	P7, P50, P94	0.231, 0.551, 0.218

Table 4. Non-symmetric type-specific discretization shortcuts for the Pearson system (EPT++ methods).

Shortcut	Johnson Distribution Type	Discretization Values	Respective Probabilities
$EPTS_B$++	S_B (unimodal only)	P6, P50, P97	0.192, 0.648, 0.160
$EPTS_L$++	S_L	P4, P50, P96	0.163, 0.674, 0.163
$EPTS_U$++	S_U	P9, P50, P92	0.283, 0.450, 0.267

Table 5. Non-symmetric type-specific discretization shortcuts for the Johnson system (EPT++ methods).

The skewed distributions considered in this section all have positive skewness, which results in discretizations with more extreme upper percentiles for some distribution types (i.e., $\alpha_3 - 0.5 > 0.5 - \alpha_1$). The upper percentile of the type IV shortcut, for example, is slightly more extreme than the lower percentile and corresponds specifically to the "thicker" upper tail. If the distribution is left-skewed, then the lower percentile should be more extreme to match that tail. Therefore, the shortcuts will need to be accordingly reflected for left-skewed distributions. For example, the shortcut for a right-skewed type IV is (P7, P50, P94, 0.231, 0.551, 0.218), but for a left-skewed type IV, the shortcut would become (P6, P50, P93, 0.218, 0.551, 0.231). These shortcuts are on average more accurate than those in Table 1 in matching the mean (and often the variance as well) of the distributions to which they were tailored.

Standard Percentile Discretizations

The P10, P50, and P90 percentiles have become very common in practice, through the use of ESM and MCS and in other applications. Therefore, in this section, we find the ASE-minimizing three-point discretizations using these percentiles. We call these the Standard Percentile (SP) methods, and they are given in Table 6 for the Pearson system and Table 7 for the Johnson system.

The probabilities for SP1∩, SP3, and SP6 are similar to those of ESM (0.300, 0.400, 0.300). SPS_L also has similar probabilities to these, which is to be expected, as estimating the mean of lognormal distributions was ESM's origin. That is, if one wants to use the P10, P50, and P90 percentiles and is dealing with a unimodal distribution bounded at one or both ends, ESM is close to the discretization that minimizes the error in the mean. The probabilities for SP1J, SP1U, and SPS_B are similar to those of MCS. The probabilities assigned to the P10 and P90 points are largest for the type IV distributions because of their higher kurtosis, and thus thicker tails, being unbounded in both directions. The SP4 discretization is almost an equal weighting of the P10, P50, and P90, which is similar to ZDT.

Shortcut	Distribution Type	Discretization Values	Respective Probabilities
SP1U	I-U (Beta)	P10, P50, P90	0.228, 0.544, 0.228
SP1J	I-J (Beta)	P10, P50, P90	0.273, 0.454, 0.273
SP1∩	I-∩ (Beta)	P10, P50, P90	0.296, 0.408, 0.296
SP3	III (Gamma)	P10, P50, P90	0.302, 0.396, 0.302
SP6	VI (Beta Prime)	P10, P50, P90	0.308, 0.384, 0.308
SP5	V (Inverse Gamma)	P10, P50, P90	0.314, 0.372, 0.314
SP4	IV	P10, P50, P90	0.322, 0.356, 0.322

Table 6. Type-specific P10-P50-P90 discretization shortcuts for the Pearson system (SP methods).

Shortcut	Johnson Distribution Type	Discretization Values	Respective Probabilities
SPS_B	S_B (unimodal only)	P10, P50, P90	0.279, 0.442, 0.279
SPS_L	S_L	P10, P50, P90	0.311, 0.378, 0.311
SPS_U	S_U	P10, P50, P90	0.316, 0.368, 0.316

Table 7. Type-specific P10-P50-P90 discretization shortcuts for the Johnson System (SP methods).

Shortcuts Utilizing the Distribution Mode

The mode (most likely value) of a distribution is a possible alternative to the P50 for the middle discretization point. Several methods in the PERT literature (Keefer and Bodily, 1983; Johnson, 1998) use the mode in conjunction with two percentiles to approximate the mean of a distribution. However, the mode is not a good choice for matching the mean or as a general approximation of distribution shape, as demonstrated below.

We briefly consider shortcuts using the mode for the Johnson system, using the same procedure as before to fit weights of the P10, mode, and P90 to minimize ASE in the mean for each of the three types. The resulting discretizations are given in Table 8.

Johnson Distribution Type	Discretization Values	Probabilities
S_B (unimodal only)	P10, mode, P90	0.384, 0.232, 0.384
S_L	P10, mode, P90	0.409, 0.182, 0.409
S_U	P10, mode, P90	0.480, 0.040, 0.480

Table 8. Best mean-preserving discretization shortcuts using the P10, mode, and P90 for the three Johnson distribution types.

A few observations show the inferiority of the mode as a discretization point. First, all three shortcuts are bimodal: each weights the individual percentiles more than the mode, and thus does not have the same modal-shape of the approximated distribution (which are all unimodal here). Second, these weights indicate that the percentiles are more important than the mode for matching the mean. These shortcuts have average absolute errors in the mean of 4.54%, 0.09%, and 4.96% for the S_B (unimodal only), S_L, and S_U types, respectively, compared to 1.56%, 0.12%, and 0.11% for the MP shortcuts. Although the mode shortcut for the S_L type is slightly more accurate in the mean than the corresponding SP shortcut, it has 21.16% average absolute error in the variance, compared to 10.10% for SPS_L. The mode shortcuts tend to drastically overestimate the variance, due to the large weighting of the P10 and P90. For these reasons, three-point mode-based discretizations poorly estimate both a distribution's shape and moments, and should not be used.

DISCRETE APPROXIMATIONS IN PERT

PERT is a popular method for estimating project completion time. It began as a U.S. Navy project to develop a quantitative evaluation methodology for project management (Malcolm et al., 1959). This section briefly reviews discrete approximations in PERT and presents several new approximations. Our regression procedure used in the previous section to construct new discretization shortcuts is easily extended to construct new PERT formulae.

MacCrimmon and Ryavec (1964) explained and analyzed four assumptions of the Classical PERT method:

1. The beta distribution models activity time.
2. The mode activity time *m* is known, as well as the optimistic time (lower bound) and pessimistic time (upper bound).
3. The mean activity time is estimated as
$$\hat{\mu} = \frac{a + 4m + b}{6}. \tag{13}$$
4. The variance of the activity time is estimated as
$$\hat{\sigma}^2 = \frac{(b-a)^2}{36}. \tag{14}$$

The mean estimation formula implies a 1/6, 2/3, and 1/6 weighting of the P0, mode, and P100, respectively. The variance estimation formula implies an equal weighting with opposite signs of only the P0 and P100. Later work, which we briefly review, challenged these assumptions and proposed new approximation methods for the mean and variance. PERT methods typically use separate estimation formulae for the mean and variance that don't necessarily form discrete probability distributions.

Approximation Formulae

Moder and Rodgers (1968) suggested using the P5 for *a* and P95 for *b* instead of the P0 and P100, respectively, in Equations (13) and (14) (MR). They argued that the P0 and P100 are unlikely to appear in data or experience, and thus are difficult to estimate. Perry and Greig (1975) suggested using these same percentiles, but with different weightings given in Table 9 (PG). Keefer and Bodily (1983) compared several mean and variance approximation formulae, including the two they termed ESM and EPT, over a set of beta distributions, which we discussed earlier. They found EPT to be superior in estimating both the mean and variance. Farnum and Stanton (1987) showed that the classical PERT formulae fail when the mode falls outside the range $\left(a+0.13(b-a), a+0.87(b-a)\right)$.

They proposed different formulae (FS) based on the position of the mode, which we generalize as x_2 in Table 9. Classical PERT is used in the center of the range, but variations on the formulae are used near the support endpoints. Golenko-Ginzburg (1988) proposed another weighting of the classical PERT mean formula and a more complex modification to the variance formula (GG).

Lau et al. (1998) proposed approximation formulae using more than three points by fitting the weights over a large set of distribution from Pearson's, and other, distributions systems, similar to our approach. Johnson (1998) examined the ability of several three-point approximations to match the mean and variance of a selection of beta, gamma, lognormal, and F distributions. He found that using the P4, P50, and P96, instead of the P0, mode, and P100, respectively, in the classical PERT formula gives better estimations of the mean across all four of these distributions. Shankar and Sireesha (2009) and Shankar et al. (2010) proposed modifications to the classical PERT method, but these did not show any improvement.

Table 9 summarizes the methods we consider, which are the classical and modified PERT formulae, methods discussed above that showed improvements over classical PERT, and EPT. Note that the EPT variance formula is not from Pearson and Tukey (1965), but the simplified form of the variance approximation formula in Keefer and Verdini (1993). PERT Mod is the same as classical PERT, but uses the P1 and P99 instead of the P0 and P100, respectively, to avoid some of the issues that arise using the endpoints of the distribution.

Approximation method	Discretization Values $(x_1, x_2, x_3,)$	Formula for mean	Formula for variance
Classical PERT (PERT)	P0, mode, P100	$(x_1 + 4x_2 + x_3)/6$	$\left(x_3 - x_1\right)^2 / 36$
PERT modified (PERT mod)	P1, mode, P99	$(x_1 + 4x_2 + x_3)/6$	$\left(x_3 - x_1\right)^2 / 36$
Moder and Rodgers (MR)	P5, mode, P95	$(x_1 + 4x_2 + x_3)/6$	$\left(x_3 - x_1\right)^2 / 10.2$
Perry and Greig (PG)	P5, mode, P95	$(x_1 + 0.95x_2 + x_3)/2.95$	$(x_3 - x_1)^2 / 3.25^2$
Farnum and Stanton (FS)	P0, mode, P100	If $x_1 + 0.13\left(x_3 - x_1\right) \le x_2 \le x_1 + 0.87(x_3 - x_1)$, $(x_1 + 4x_2 + x_3)/6$ If $x_2 < x_1 + 0.13(x_3 - x_1)$, $x_1 + \frac{2(x_2 - x_1)(x_3 - x_1)}{x_3 - 3x_1 + 2x_2}$ If $x_2 > x_1 + 0.87(x_3 - x_1)$, $x_1 + \frac{(x_3 - x_1)^2}{3x_3 - x_1 - 2x_2}$	$\left(x_3 - x_1\right)^2 / 36$ $\frac{\left(x_2 - x_1\right)^2 \left(x_3 - x_2\right)}{x_3 - 2x_1 + x_2}$ $\frac{(x_2 - x_1)\left(x_3 - x_2\right)^2}{2x_3 - x_1 - x_2}$
Golenko-Ginzburg (GG)	P0, mode, P100	$(2x_1 + 9x_2 + 2x_3)/13$	$\frac{\left(x_3 - x_1\right)^2}{1268}\left[22 + \frac{81\left(x_2 - x_1\right)}{x_3 - x_1} - 81\left(\frac{x_2 - x_1}{x_3 - x_1}\right)^2\right]$
Extended Pearson-Tukey (EPT)	P5, P50, P95	$0.185x_1 + 0.630x_2 + 0.185x_3$	$0.185x_1^2 + 0.630x_2^2 + 0.185x_3^2 - (0.185x_1 + 0.630x_2 + 0.185x_3)^2$

Table 9. Summary of PERT and its variants we consider.

New Formulae

We use the method described above to find new PERT approximation formulae. The goal of estimating moments directly, rather than approximating a full distribution, reduces the constraints on our fitting procedure.

In this section, we propose new formulae to estimate the mean and variance of an activity time given two percentiles and either the median or mode. We us a similar approach as above, minimizing the ASE in these moments (individually) over a large set of distributions. We perform the same optimization as in Equation (12), without requiring that the weights be symmetric. We require that they sum to one, even though these are not meant to be probability distributions. Through experimentation, we found that this constraint produced more robust formulae, and that in some cases the weights naturally summed to one. Equation (15) gives our mean approximation formula using three points, where x_2 is either the median or the mode. The mean is calculated as a weighted average of the three points using probability weights. Equations (16) and (17) are used to estimate the variance. These new approaches result as solutions to two different optimization problems. One problem solves for probability weights (p_1, p_2, p_3) in Equation (15) that minimize ASE in the mean. The other solves for a separate set of weights (p_1, p_2, p_3) in Equation (17) that minimize ASE in the variance. Equation (17) is the equation for the variance of a discrete distribution with three possibilities (x_1, x_2, x_3) with probability weights (p_1, p_2, p_3), respectively. However, these methods are not intended to be used as distributions, and to distinguish them from our new shortcut methods, we call these the "fitted probability" (FP) methods.

$$\mu_1 = p_1 x_1 + p_2 x_2 + p_3 x_3 \tag{15}$$

$$\mu_2 = p_1 x_1 + p_2 x_2 + p_3 x_3 \tag{16}$$

$$\begin{aligned} \sigma^2 &= p_1 (x_1 - \mu_2)^2 + p_2 (x_2 - \mu_2)^2 + p_3 (x_3 - \mu_2)^2 \\ &= p_1 x_1^2 + p_2 x_2^2 + p_3 x_3^2 - (p_1 x_1 + p_2 x_2 + p_3 x_3)^2 \end{aligned} \tag{17}$$

We consider the P0, P5, P10, P50, P90, P95, and P100 percentiles as well as the mode for both the Pearson type I-J and I-∩. Formulae for the mean are given in Table 10. The formulae using the mode place almost no weight on the lower percentile, because the lower percentile is too close to the mode of right skewed distributions. The mode may be less than the P5 and P10for some distributions, which appears to lead to poorer performance for methods using these percentiles with the mode rather than the median. FP 5-50-95 is almost exactly EPT, and FP 10-50-90 is almost exactly ESM, for the I-∩ distributions. FP 5-50-95 is also similar to EPT for I-J, while FP 10-50-90 is more similar to MCS than ESM for this type. The other mean approximation FP formulae are asymmetric, and unlike any of the shortcut methods.

Approximation formulae for the variance are given in Table 11. FP 0-50-100 places no weight on the P100 for either Pearson type, and places more weight on the P0 than the P50. FP 5-50-95 is again similar to EPT for the I-∩ distributions, but is bimodal, placing no weight on the P50 for type I-J. The mean approximation weights in Table 10 tend to bear little resemblance to their corresponding variance approximation weights in Table 11. The only exceptions are the FP 5-50-95 mean and variance approximation formulae for type I-∩, which we show in Chapter 6 are some of the most accurate.

FP method	Discretization Values	Probability Weights	
		I-∩	I-J
FP 0-50-100	P0, P50, P100	0.059, 0.935, 0.005	0.001, 0.986, 0.013
FP 5-50-95	P5, P50, P95	0.184, 0.632, 0.184	0.188, 0.624, 0.187
FP 10-50-90	P10, P50, P90	0.298, 0.406, 0.296	0.238, 0.497, 0.266
FP 0-m-100	P0, mode, P100	0.002, 0.985, 0.013	0.966, 0.002, 0.032
FP 5-m-95	P5, mode, P95	0.000, 0.647, 0.353	0.001, 0.687, 0.312
FP 10-m-90	P10, mode, P90	0.004, 0.563, 0.433	0.001, 0.601, 0.399

Table 10. Summary of FP methods for mean estimation.

FP method	Discretization Values	Probability Weights	
		I-∩	I-J
FP 0-50-100	P0, P50, P100	0.698, 0.302, 0.000	0.502, 0.498, 0.000
FP 5-50-95	P5, P50, P95	0.179, 0.633, 0.188	0.875, 0.000, 0.125
FP 10-50-90	P10, P50, P90	0.178, 0.301, 0.520	0.181, 0.142, 0.677
FP 0-m-100	P0, mode, P100	0.886, 0.114, 0.000	0.496, 0.504, 0.000
FP 5-m-95	P5, mode, P95	0.172, 0.702, 0.126	0.665, 0.214, 0.122
FP 10-m-90	P10, mode, P90	0.000, 0.221, 0.779	0.000, 0.789, 0.211

Table 11. Summary of FP methods for variance estimation.

DISCRETIZATION IN OTHER FIELDS

Discretization has wide application and use in fields other than decision analysis. Tauchen and Hussey (1991) used GQ in asset pricing with dynamic nonlinear models, and Sullivan (2000) used it to value American put options. Lind (1982) discussed discretization for modeling uncertainty in dynamical systems. Kotsiantis and Kanellopoulos (2006) reviewed the literature on discretization for machine learning applications. They discussed variations of bracket methods for summarizing and categorizing data statically (considering one uncertainty at a time) and dynamically (considering multiple uncertainties simultaneously to incorporate dependence). Their focus was classifying data sampled from continuous ranges, rather than discretizing continuous distributions explicitly.

Multistage stochastic programming is another field that often requires approximations of uncertainties to make the problems tractable. Kaut and Wallace (2007) reviewed several methods used in the stochastic programming literature, including moment matching using GQ. Kall et al. (1988) discussed discretization using partitions of the distributions' support (bracket methods), particularly a method corresponding to BMn, which can be used with the Jensen and Edmundson-Madansky inequalities to bound the optimal value (Huang et al., 1977). Other methods include Monte Carlo sampling (Birge and Louveaux, 1997) and nonlinear optimization (Høyland and Wallace, 2001; Pflug, 2001). Discretization is also sometimes called scenario generation, since the result is a discrete set of uncertainty realizations.

Histograms are commonly used to graphically represent frequency data and are discrete by definition. Histograms represent the frequencies or counts of observations falling within a particular range (also known as a *bin*) as the height of the bar in the graph. The discretization methods discussed in this work are typically used as simplifying approximations to continuous distributions, whereas histograms are used to represent the distributions implied by empirical data (Scott, 2011). The parameters for constructing a histogram include the number of bins, bin widths, and orientation of the bins. Two common heuristic rules for selecting the parameters for histograms with homogenous bin widths are Sturges' Rule, which determines the number of bins, and Scott's Rule, which determines the bin width for a given number of bins (Scott, 1979, 2010, 2011). The flexibility of histogram derivation rules are limited by the quality and format of the empirical data, such as when the data are collected into bins of unequal width (Scott and Scott, 2008).

EVALUATING DISCRETIZATION METHODS

Several characteristics of a distribution might be desirable to preserve, and several corresponding measures of error. Distribution moments are an obvious characteristic, which much of the literature has considered. Moments not only are summary statistics of the entire distribution but are important to matching the moments of the output distribution.

Prior Work

Keefer and Bodily (1983) made the first systematic study of the performance of several discretization methods, including mean- and variance-approximation formulas, shortcuts (EPT and ESM), and a distribution-specific method (Bracket Median). They studied how well the methods preserved the mean and variances of 78 bell-shaped (∩-shaped) beta distributions and found EPT to be the best. ESM had slightly inferior performance but was also found to be better than most of the other methods they examined. Keefer and Verdini (1993) performed a similar study with the same set of beta distributions and included some approximations developed after 1983, which they found inferior to EPT and ESM in matching means and variances. They noted that EPT's improvement over ESM may be outweighed by the evidence that EPT's more extreme percentiles are more difficult to subjectively assess. Keefer (1994) examined the performance of all six shortcuts described above in preserving certainty equivalents, which depend upon all the moments of the underlying pdf, over a set of 169 right-skewed, left-skewed, ∩-shaped, and J-shaped beta distributions, as well as four lognormal distributions, and again found EPT to perform the best. This analysis showed that EPT had the lowest error across several levels of risk aversion, including risk neutrality, in which case the certain equivalent is simply the mean.

The distributions in Figure 3 that were considered by both Keefer and Bodily (1983) and Keefer (1994) are denoted by grey circles. Distributions considered by Keefer (1994), but not by Keefer and Bodily (1983), are denoted by black circles. Both Keefer and Bodily (1983) and Keefer (1994) considered a relatively small sample of distributions within the Pearson system, being confined to ∩-shaped beta (type I) distributions with low skew. Keefer (1994) expanded this somewhat but limited the analysis to beta distributions along the transition between the ∩-shaped and J-shaped beta regions and four lognormal distributions.

The distributions Pearson and Tukey (1965) used to construct their mean- and variance- matching formulae are shown as black diamonds in Figure 3. Although 11 of the 96 points they used fall outside the area of Figure 3, these points are sparsely distributed over a range of kurtoses from 10 to 20 and are inconsistent with the denser spacing of the rest of their grid. Pearson and Tukey's analysis did not fully explore the Pearson system, as it was limited to the tables of Pearson distributions that were available at the time.

Bickel et al. (2011) showed that ESM can perform very poorly on moderate-to-highly-skewed lognormal distributions, significantly underestimating the mean, variance, and skewness of the underlying pdf, and instead recommended the use of GQ. Yet, as discussed above, GQ can be difficult to use in practice. Analysts may therefore want to use one of the existing shortcut methods, such as ESM or MCS.

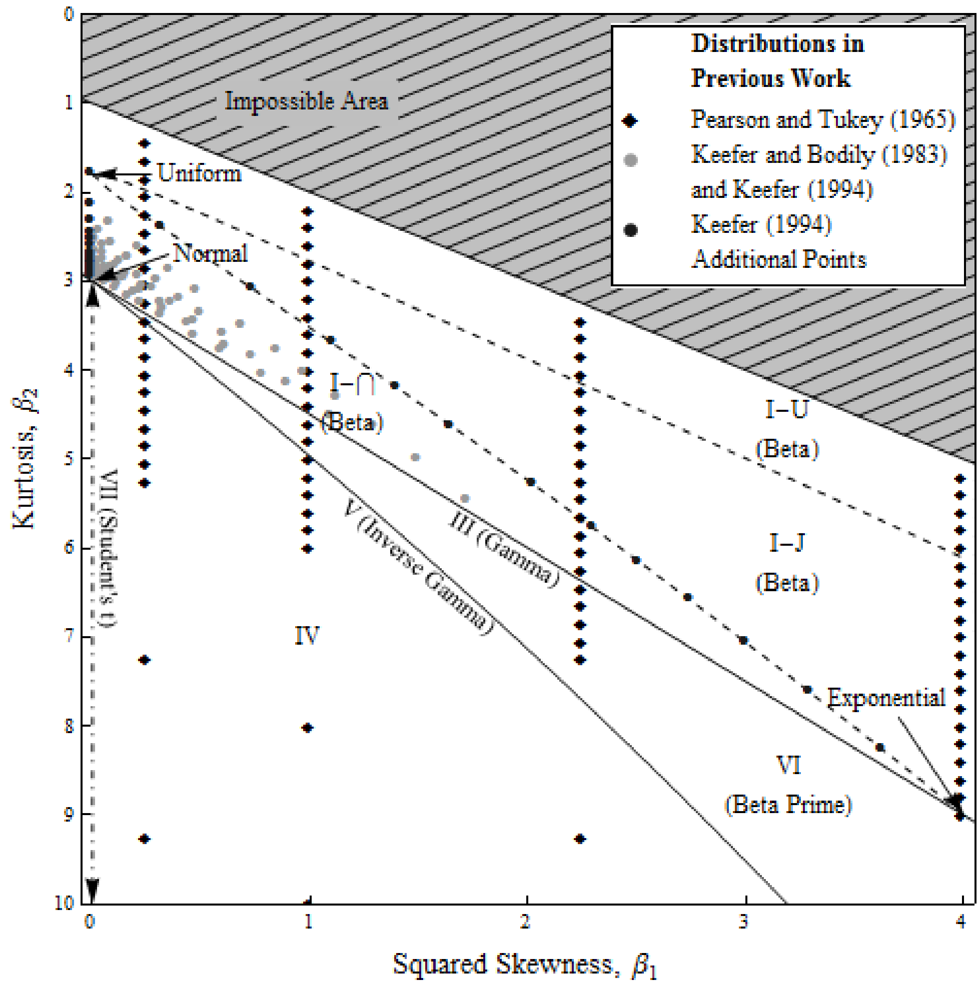

Figure 3. The Pearson distribution system (β_1, β_2) plot showing distributions used in previous work.

Moments

Common properties of interest are the moments of *X* (e.g., the mean, variance, skewness, and kurtosis). Smith (1993) provided the following argument in support of matching moments: In decision and risk analysis, we are often interested in computing the expectation of some value function (e.g., net present value or utility) $v(x)$, that is a function of the uncertain quantity *X*. If $v(x)$ can be closely approximated by a differentiable function, it can be written as a Taylor series expansion about *a*:

$$v(x) = v(a) + \frac{v'(a)}{1!}(x-a) + \frac{v''(a)}{2!}(x-a)^2 + \frac{v'''(a)}{3!}(x-a)^3 + ... \quad (18)$$

Taking the expectation $E[v(X)]$ yields a weighted sum of the raw moments of X. For example, if $a = 0$, we have

$$E[v(X)] = v(0) + \frac{v'(0)}{1!}E[X] + \frac{v''(0)}{2!}E[X^2] + \frac{v'''(0)}{3!}E[X^3] + ... \quad (19)$$

Thus, a good approximation of the mean of our output variable, $E[v(X)]$, must closely approximate the moments of the input variable X,. For a thorough discussion of polynomial approximation, see Hamming (1973).

Certain Equivalents

Rather than explicitly consider risk aversion and matching certain equivalents as Keefer (1994) did, we apply the same reasoning as before for matching moments. The certain equivalent can be considered part of the value function $v(x)$, so that to match the certain equivalent, we would generally need to also match several moments.

For example, Howard (1971) demonstrated a formula from Pratt (1964) for approximating a certain equivalent using only the mean and variance of the value lottery.

$$CE\left(v(X)\right) \approx E\left[v(X)\right] - \frac{\text{var}(v(X))}{2\rho(v(X))}. \quad (20)$$

Summary Statistics

In many cases, it will be convenient to summarize the error measures over a set of distributions. Keefer and Bodily (1983) used the four summary statistics given in Table 12 to summarize their results for the beta distribution. We define these error measures in our notation as follows: Let H_r denote the indexed set of distributions corresponding to region $r \in \{\text{I-}\cap\text{,I-J,I-U,III,IV,V,VI}\}$ in the Pearson system or $r \in \{S_B, S_L, S_U\}$ in the Johnson system. The number of points that we sample in this region is s_r. For a distribution index $i \in D_r$, the true k-th moment is m_i^k, the estimate from a discretization is $\hat{m}_i^k$, and the *error* in the k-th moment is $\delta_i^k = (\hat{m}_i^k - m_i^k)$. ME and MPE give the error with the greatest magnitude, regardless of sign. For these two measures, positive error indicates that the discretized moment is larger than the actual moment.

Error Statistic	Formula
Average Absolute Error (AAE)	$\frac{1}{s_r}\sum_{i\in D_r}\left\lvert\delta_i^k\right\rvert$
Average Absolute Percentage Error (AAPE)	$\frac{1}{s_r}\sum_{i\in D_r}\left\lvert\frac{\delta_i^k}{m_i^k}\right\rvert\cdot 100\%$
Maximum Error (ME)	$\delta_j^k, \quad j = \arg\max_{i\in D_r}\left\lvert\delta_i^k\right\rvert$
Maximum Percentage Error (MPE)	$\frac{\delta_j^k}{m_j^k}\cdot 100\%, \quad j = \arg\max_{i\in D_r}\left\lvert\frac{\delta_i^k}{m_i^k}\right\rvert$

Table 12. Keefer and Bodily (1983) summary statistics.

Because the errors, ε_i^k, and percentage errors, ε_i^k / m_i^k, are the same for the standardized distributions, we typically use percentage error in moments, unless otherwise specified. We use the ME statistic from Table 12, along with two other statistics: the Average Error (AE) and the Average Squared Error (ASE), which are given in Table 13.

Error Statistic	Formula
Average Error (AE)	$\frac{1}{s_r}\sum_{i\in D_r}\delta_i^k$
Average Squared Error (ASE)	$\frac{1}{s_r}\sum_{i\in D_r}(\delta_i^k)^2$

Table 13. Additional Error Measures.

EXAMPLES

This section illustrates the use of SP, EPT++, GQ, and BMn on a lognormal distribution with parameters μ = 1, σ = 0.5. We use the SPS_L and $EPTS_L$++ shortcuts, since we are discretizing a lognormal distribution.

SPS_L Shortcut

The lognormal cdf and the resulting SP6 discretization cdf are shown in Figure 4. The values and probabilities used by the discretization are given in Table 14. The SPS_L shortcut is similar to ESM but assigns slightly higher probabilities to the P10 and P90.

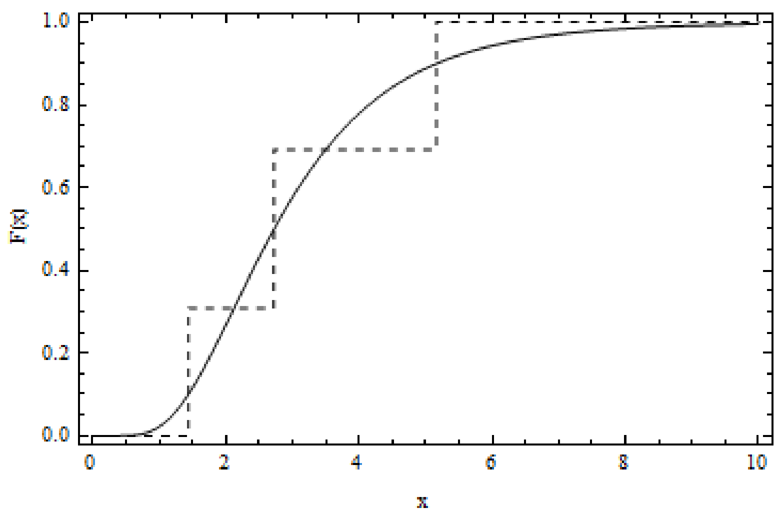

Figure 4. SPS_L (dashed line) and lognormal (solid line) cdfs.

Quantity	x_1	x_2	x_3
Values	1.4322	2.7183	5.1592
Probabilities	0.311	0.378	0.311
Percentiles	10.00%	50.00%	90.00%

Table 14. SPS_L values and probabilities for Lognormal(1, 0.5).

$EPTS_L$++ Shortcut

The lognormal cdf and resulting discretization cdf are shown in Figure 5. The values and probabilities for the discretization are given in Table 15.

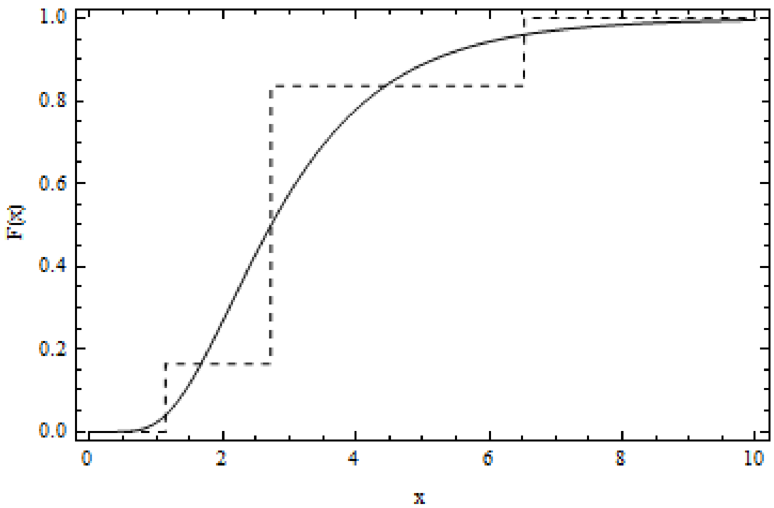

Figure 5. $EPTS_L$++ (dashed line) and lognormal (solid line) cdfs.

Quantity	x_1	x_2	x_3
Values	1.1328	2.7183	6.5231
Probabilities	0.163	0.674	0.163
Percentiles	4.00%	50.00%	96.00%

Table 15. $EPTS_L$ ++ values and probabilities for Lognormal(1, 0.5).

Gaussian Quadrature

Here we demonstrate three- and five-point GQs (GQ3 and GQ5, respectively). The cdfs are shown in Figure 6, and the values, probabilities, and equivalent percentiles are given for GQ3 and GQ5 in Table 16 and Table 17, respectively. Each method uses the P99.9 or a more extreme percentile. The orthonormal polynomials for GQ3 in this example are

$$P_{-1}^*(x) = 0,$$
$$P_0^*(x) = 1,$$
$$P_1^*(x) = -1.876 + 0.609x,$$
$$P_2^*(x) = 2.640 - 1.524x + 0.169x^2,$$
$$P_3^*(x) = -3.207 + 2.484x - 0.489x^2 + 0.024x^3.$$

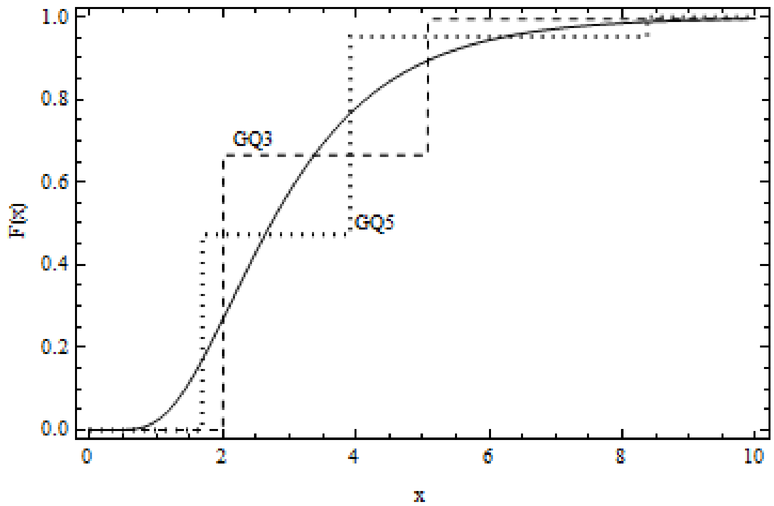

Figure 6. GQ3 (dotted line), GQ5 (dashed line), and lognormal (solid line) cdfs.

Quantity	x_1	x_2	x_3
Values	2.0003	5.0784	12.8934
Probabilities	0.6652	0.3285	0.0063
Percentiles	26.98%	89.44%	99.91%

Table 16. Three-point GQ of Lognormal(1, 0.5).

Quantity	x_1	x_2	x_3	x_4	x_5
Values	1.6885	3.9121	8.3729	17.9201	41.5203
Probabilities	0.4729	0.4789	0.0477	0.0005	2.6081E-07
Percentiles	17.05%	76.67%	98.78%	99.99%	≈100.00%

Table 17. Five-point GQ of Lognormal(1, 0.5).

Bracket Mean

Figure 7 shows the cdfs for a three-point Bracket Mean discretization (BMn3) using the brackets [0, 0.25], [0.25, 0.75], and [0.75, 1] (alternatively referred to as 25-50-25) and a five-point Bracket Mean discretization (BMn5) using equal brackets of 0.2. The values and probabilities for BMn3 and BMn5 are given in Table 18 and Table 19, respectively.

The discretization points are more tightly clustered about the mode of the distribution than GQ, causing BMn to fail to fully capture the distribution tails, as did the shortcuts. Whether BMn or a shortcut is preferred depends on the importance of the different moments.

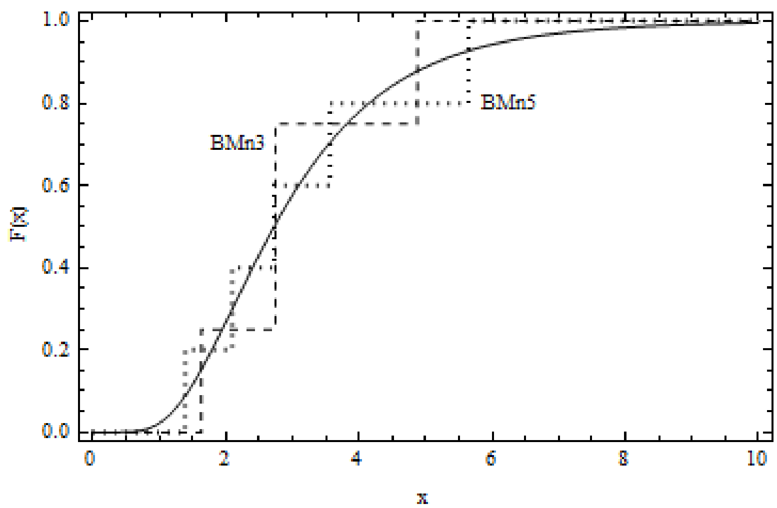

Figure 7. BMn3 (dotted line), BMn5 (dashed line), and lognormal (solid line) cdfs.

Quantity	x_1	x_2	x_3
Values	1.4797	2.7670	5.3071
Probabilities	0.2500	0.5000	0.2500
Percentiles	11.19%	51.42%	90.96%

Table 18. Three-point BMn values and probabilities for lognormal(1, 0.5) using cumulative brackets [0, 0.25], [0.25, 0.75], and [0.75, 1].

Quantity	x_1	x_2	x_3	x_4	x_5
Values	1.3839	2.0909	2.7255	3.5591	5.6417
Probabilities	0.2000	0.2000	0.2000	0.2000	0.2000
Percentiles	8.85%	29.98%	50.21%	70.51%	92.79%

Table 19. Five-point BMn values and probabilities for lognormal(1, 0.5) using equal cumulative brackets of 0.2.

The percentage errors in the moments and the three L-norms are given for each method in Table 20. The high SPS_L errors in the skewness and kurtosis estimations are -75.57% and -81.57%, respectively. $EPTS_L$++ closely matches the mean and variance, with errors of 0.07% and -1.52%, respectively, but significantly underestimates the skewness and kurtosis, with errors of -28.20% and -59.52%, respectively. This poor performance in matching the skewness and kurtosis is typical of shortcut methods, especially for unbounded distributions. The distribution tails, which shortcuts do not capture well, are more important to these higher moments than to the mean and variance. $EPTS_L$++ has lower errors in each of the first four moments than does SPS_L, because its more extreme percentiles better capture the tails than does SPS_L. However, EPT6++'s more extreme percentiles may be more difficult to assess than the P10 and P90 used by SP6.

GQ3 matches moments 0 through 5, and GQ5 matches moments 0 through 9. Their errors in the first four moments are all 0%. While any Bracket Mean discretization will theoretically match the mean of the distribution, they are often inferior to some shortcuts in matching higher moments. The moment errors for BMn3 and BMn5 are larger than those for $EPTS_L$++ in the variance, skewness, and kurtosis. BMn3 and BMn5 have the lowest L_p distances, despite having significantly higher moment errors than the other four methods.

	Mean	Variance	Skewness	Kurtosis
SPS_L	-0.20%	-17.69%	-75.57%	-81.57%
$EPTS_L$++	-0.07%	-1.52%	-28.20%	-59.52%
GQ3	0.00%	0.00%	0.00%	0.00%
GQ5	0.00%	0.00%	0.00%	0.00%
BMn3	0.00%	-28.41%	-63.31%	-76.47%
BMn5	0.00%	-20.05%	-60.17%	-74.65%

Table 20. Error measures for each discretization in the example.

Probability Assessment

Decision analysis problems often require subjective assessment of probability distributions, especially in cases where relevant historical data is limited or nonexistant. One of the most common assessment methods uses percentiles elicited from a subject matter expert to construct a probability distribution. These assessments may be used directly in a discrete approximation to the distribution or to fit the parameters of a continuous distribution.

When using subjectively assessed probabilities, the quality of the assessment, as a representation of an expert's knowledge and beliefs, is of great importance to the quality of the decision model. Winkler and Murphy (1968) defined the quality of probability assessments as having a normative component, that the assessments form coherent probabilities that are consistent with the expert's beliefs, and a substantive component, or the extent of the expert's knowledge about the uncertain quantity. Wallsten and Budescu (1983) similarly defined assessment quality in terms of reliability (repeatability/consistency of assessments) and validity (representativeness of the expert's actual opinion). Lichtenstein et al. (1982), in a survey of the early work on probability elicitation and calibration, described an expert as *well calibrated* "if over the long run, for all propositions assigned a given probability, the proportion that is true is equal to the probability assigned."

The assessments made by an expert may not be representative of their actual knowledge and beliefs, due to cognitive biases such as those identified by Tversky and Kahneman (1974). For example, experts and decision-makers can often be overconfident, assigning distribution that are too narrow (Capen, 1976; Alpert and Raiffa, 1982). To reduce the effects of biases in elicitation, several methods were developed. Hampton et al. (1973), Spetzler and Holstein (1975), and Wallsten and Budescu (1983) described and reviewed several such methods. Wallsten and Budescu (1983) reported that the results of several studies “suggest high validity and coherence, given outcome feedback and experience at encoding probabilities.” Merkhofer (1987), in a retrospective on the elicitation process described by Spetzler and Holstein (1975), claimed that despite the lack of formal studies at the time, there was enough experiential evidence from practicing decision analysts to conclude that the elicitation process greatly increases the consistency and the experts’ confidence in the validity of probability assessments. These findings indicate that the biases evident in assessments from un-calibrated experts can be greatly mitigated with sufficient and appropriate training.

Two commonly-used assessment methods are “fixed-probability,” in which the cumulative probability, $F(x)$, is fixed and the corresponding value, x, is elicited, and “fixed-value,” in which the value, x, is fixed and the corresponding cumulative probability, $F(x)$, is elicited (Abbas et al., 2008). Spetzler and Holstein (1975) described several encoding techniques applicable to both methods and additionally described a third hybrid method, in which neither probabilities nor values are fixed. We focus on the fixed-probability, or percentile (also referred to as the “quantile” or “fractile” method by some authors), method for the 10^{th} (P10), 50^{th} (P50), and 90^{th} (P90) percentiles.

It is common practice in the oil & gas industry to assess the P10, P50, and P90 (Rose, 2001a; Jahn et al., 2008), which has become an accepted standard for reporting reserves uncertainty (Ross, 2011). This provided the basis for Swanson’s Rule (Megill, 1984; Hurst et al., 2000) for estimating the mean of a distribution by weighting the P10, P50, and P90 by 0.300, 0.400, and 0.300, respectively. Other schemes are used; Schuenemeyer (2002), for example, described an application in which five percentiles and the minimum and maximum values are assessed.

The issues of calibration and assessment accuracy are important to our discussion of discretization accuracy, in the quality of the methods’ representations of assessed distributions. Intuitively, there is little justification for using an extremely accurate discretization if the continuous distribution is not believed to be an accurate representation of the uncertainty. It was not previously clear to what levels of discretization and assessment accuracy this applies. Chapter 5 addresses this question by comparing the relative importance of these errors for different methods and levels of error. The moment estimates by some methods, such as ESM and EPT, are similar enough that even a small degree of assessment error makes the methods' differences insignificant. MCS's moment estimates, as compared to ESM or EPT, can in many cases be more significant than the assessment errors of reasonably well-calibrated experts seen in practice.

SUMMARY

This chapter reviewed several discretization methods and illustrated their use in an example. Chapter 4 will analyze and compare their accuracies in matching the moments of large sets of distributions by using the evaluation metrics discussed in this chapter. PERT approximation formulae for the mean and variance were also reviewed, and will be analyzed in Chapter 6. Finally, the chapter discussed practical issues with probability assessment, which Chapter 5 will incorporate into our analysis as assessment error.

Chapter 3: The Pearson and Johnson Distribution Systems

DISTRIBUTION SYSTEMS

A distribution system defines a set of continuous distributions. The best-known distribution system is the Pearson system (Pearson, 1895, 1901, 1916). We use this, along with the Johnson (1949) system, for analysis. Each system has advantages and disadvantages, but contain the same general distribution shapes. The Pearson system contains many common distributions as special cases and three ranges of support boundedness. The Johnson system contains only two common distributions, the normal and lognormal, which are treated as special cases, and only two ranges of support boundedness. The Pearson system includes the normal distribution, but not the lognormal. Although distributions in the two systems with similar moments have similar shapes, some moments do not exist for certain Pearson system distributions, but always exist for Johnson system distributions. The Pearson system has convenient relationships between its parameters and distribution moments, but has cdfs that must be numerically evaluated, making it convenient for working with moments, but not percentiles. The Johnson system, on the other hand, has cdfs that are simple to evaluate as transformations of the normal distribution, but its moments must be evaluated numerically for the main distribution types, making it convenient for working with percentiles, but not moments.

Other distributions systems include Burr (1973), Ramberg and Schmeiser (1974), Schmeiser and Deutsch (1974), and Butterworth (1987). We do not use these systems for a variety of reasons. The Burr (1973) system and the Ramberg and Schmeiser (1974) system each have only a single range of support, versus the three of Pearson's system. The Ramberg and Schmeiser (1974) system covers a smaller range of shapes than Pearson's or Johnson's. The Butterworth (1987) system only approximates several of the named distributions included in the Pearson system.

Kolari et al. (1989) identified distribution shapes for 11 financial ratios throughout the Pearson and Johnson systems. Lau et al. (1995) found the moments of the empirical distributions of two financial ratios to lie in the Pearson I∩, IV, and VI types.

PEARSON DISTRIBUTION SYSTEM

A distribution $f(x)$ in the Pearson system is a solution of the differential equation

$$\frac{1}{f}\frac{df}{dx} = \frac{a+x}{b_0 + b_1 x + b_2 x^2}. \tag{21}$$

Multiplying each side of Equation (21) by x^r and rearranging yields

$$x^r\left(b_0 + b_1 x + b_2 x^2\right)\frac{df}{dx} = x^r f\left(a+x\right). \tag{22}$$

Integrating each side with respect to x, using integration by parts on the left side, gives

$$\begin{aligned} &x^r\left(b_0 + b_1 x + b_2 x^2\right) f - \int\left(rb_0 x^{r-1} + (r+1)b_1 x^r + (r+2)b_2 x^{r+1}\right) f dx \\ &= \int fx^{r-1}dx + \int fax^r dx. \end{aligned} \tag{23}$$

It is assumed that the term $x^r\left(b_0 + b_1 x + b_2 x^2\right) f$ vanishes at the boundaries of the distribution f's support, leaving the recursive moment relation

$$-rb_0\mu_{r-1} - \left(r+1\right)b_1\mu_r - \left(r+2\right)b_2\mu_{r+1} = \mu_{r+1} + a\mu_r. \tag{24}$$

The requirement that a probability distribution pdf integrate to one provides the boundary condition $\mu_0 = 1$ for this recursion. Setting r = 0, 1, 2, 3 yields the system of equations

$$\begin{aligned} b_1 + 2b_2\mu_1 + a_1\mu_1 &= -a_0, \\ b_0 + 2b_1\mu_1 + 3b_2\mu_2 + a_1\mu_2 &= -a_0\mu_1, \\ 2b_0\mu_1 + 3b_1\mu_2 + 4b_2\mu_3 + a_1\mu_3 &= -a_0\mu_2, \\ 3b_0\mu_2 + 4b_1\mu_3 + 5b_2\mu_4 + a_1\mu_4 &= -a_0\mu_3. \end{aligned} \tag{25}$$

These four equations in the four unknowns, a, b_0, b_1, and b_2, define the mean, variance, skewness, and kurtosis of a distribution, but the distribution must still be normalized so that it integrates to one. The form of the normalizing factor depends on the distribution type and is given for each of the types discussed below in their pdfs.

Together, within the Pearson system, the third and fourth moments determine a unique location-scale distribution (Elderton and Johnson, 1969). This fact allows distributions in the Pearson system to be conveniently characterized by their shape, which is defined by their skewness γ_1 and kurtosis β_2 (the third and fourth central moments, respectively). Since skewness can be positive or negative but is symmetric under reflection, it is convenient to consider squared-skewness, $\beta_1 = \gamma_1^2$. Figure 8 shows a portion of the Pearson system, denoting several regions, or classes of distributions. The β_1 and β_2 axes are not bounded above. The vertical axis denoting kurtosis is inverted following the convention of previous work (Rhind, 1909; Pearson, 1916; Craig, 1936; Draper, 1952; Ord, 1972; Lau et al., 1995).

The Pearson system includes all possible combinations of skewness and kurtosis, which Pearson (1916) showed must have

$$\beta_2 > \beta_1 + 1. \tag{26}$$

The region above the line $\beta_2 = \beta_1 + 1$ is shaded and labeled as the "Impossible Area" in Figure 8. Although the Pearson system covers all possible (β_1, β_2) pairs, it does not include all possible pdfs, most notably the lognormal (which is included in the Johnson system).

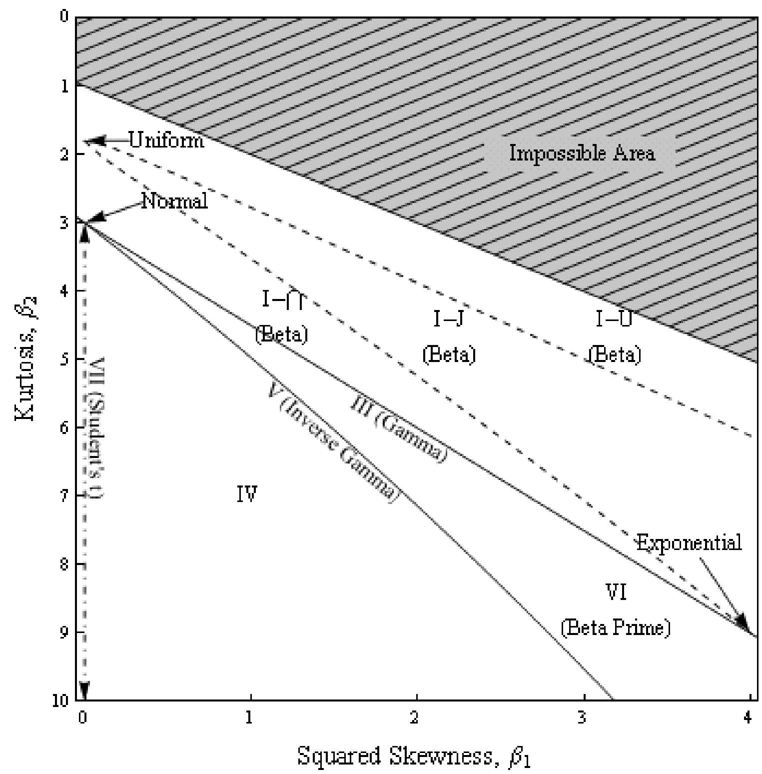

Figure 8. Pearson Distribution System.

The Pearson system is a convenient and natural choice because of its direct relation to named distributions over much of the feasible (β_1, β_2) region, its variety of distribution shapes and support ranges, and general ease of use as compared to other systems. Three main types of distributions cover the feasible region, which Pearson designated type I, type IV, and type VI, as highlighted in Figure 8. Pearson defined nine additional types, which are special cases of the main three or transition boundaries between them. For example, types III and V are the gamma and inverse gamma distributions, respectively. The normal distribution is a special case, where types I, II (not-shown), III, IV, V, and VI intersect.

Pearson Distribution Types

We now briefly describe the three main Pearson distributions and two transition distributions, top-to-bottom as they appear in Figure 8. The distributions in the Pearson system are location-scale generalizations, but we give the standard forms of the distributions, which are equivalent under an appropriate shifting by μ_1 and scaling by $\sigma = \sqrt{\mu_2 - \mu_1^2}$ of *x*,

$$y = \frac{x - \mu_1}{\sigma} \tag{27}$$

Type I (Beta Distribution)

Type I corresponds to the beta distribution, with pdf

$$f_I(y) = \frac{1}{B(\alpha, \beta)} y^{\alpha-1}(1-y)^{\beta-1}, \quad 0 < y < 1, \tag{28}$$

where α and β are parameters and $B(\alpha, \beta)$ is the beta function. Pearson (1895) characterized this type as having skewness and limited range. This type arises when the denominator of Equation (21) has roots with opposite signs. The beta can be ∩-shaped ($\alpha \ge 1, \beta \ge 1$), J-shaped ($\alpha \ge 1, \beta < 1$ or $\alpha < 1, \beta \ge 1$), or U-shaped ($\alpha < 1, \beta < 1$). We denote these shapes as types I-∩, I-J, and I-U, respectively. When a beta distribution is J-shaped, the value of *f* approaches infinity as *x* approaches 0 (when $\alpha \ge 1, \beta < 1$) or 1 (when $\alpha < 1, \beta \ge 1$). When it is U-shaped, *f* goes to infinity at both 0 and 1. Type I-U is the only Pearson type that is not unimodal. The symmetric type I is called type II and lies along the β_2 axis (not shown) between kurtoses 1 and 3. The uniform distribution is the point $(\beta_1 = 0, \beta_2 = 1.8)$, which is also the point where the three regions of type I meet.

Type III (Gamma Distribution)

Type III, or the gamma distribution, is a transition distribution that forms the boundary between type I and type VI in Figure 8. It has pdf

$$f_{III}(y) = \frac{1}{\theta^k \Gamma(k)} y^{k-1} e^{-y/\theta}, \quad 0 \le y < \infty, \tag{29}$$

where $\Gamma(k)$ is the gamma function, and k and θ are parameters. This type occurs when $b_2 = 0$ in Equation (21). At the point where this type intersects with the line that divides the ∩-shape and J-shape type I regions, is the exponential distribution, or type X.

Type VI (Beta Prime Distribution)

Type VI corresponds to the beta prime distribution, also called the inverted beta distribution or the beta distribution of the second kind. For parameters α and β, the pdf is

$$f_{VI}(y) = \frac{y^{\alpha-1}(1+y)^{-\alpha-\beta}}{B(\alpha,\beta)}, \quad 0 < y < \infty. \tag{30}$$

Pearson (1901) characterized this distribution as being unbounded in one direction. As seen in Figure 8, type VI covers the region between the gamma and inverse gamma distributions, each of which also has this property. Type VI distributions are the solution to Equation (21) when its denominator has roots of the same sign.

Type V (Inverse Gamma Distribution)

Type V is the second transition type, which separates the regions of type IV and type VI and is known as the inverse gamma distribution. For parameters α and β, it has pdf

$$f_V(y) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{-\alpha-1} e^{-\beta/y}, \quad 0 < y < \infty. \tag{31}$$

This type occurs when the roots of Equation (21) are real and equal.

Type IV

Type IV does not correspond to any single common distribution. For parameters *m* and *v*, the pdf is

$$f_{IV}(y) = \frac{\Gamma(m)}{\sqrt{\pi}a\Gamma(\mathrm{m}-1/2)} \left|\frac{\Gamma(m+iv/2)}{\Gamma(m)}\right|^2 \left(1+y^2\right)^{-m} e^{-v\tan^{-1}(y)}, \quad -\infty < y < \infty. \tag{32}$$

Pearson (1895) characterized this type as being unbounded in both directions and possibly having skewness. It is the solution to Equation (21) that arises when the denominator has complex roots. A special case when type IV is symmetric $(\beta_1 = 0, \beta_2 > 3)$ is Student's *t*-distribution, which Pearson (1916) called type VII.

Fitting Pearson Distributions to Moments

The Pearson system lends itself well to fitting to moments. Elderton and Johnson (1969) provided straightforward procedures for fitting each of the Pearson types discussed here.

JOHNSON DISTRIBUTION SYSTEM

The distribution system developed by Johnson (1949) takes a different approach from the Pearson system, by generating distributions through transformations of a normal random variable. A distribution in the Johnson system is defined by the four parameters γ, δ, ξ, and λ in the general equation

$$z = \gamma + \delta g\left(\frac{x-\xi}{\lambda}\right), \tag{33}$$

where *z* is a standard normal random variable, and *g* is a monotonic function. As with the Pearson system, the four parameters of the Johnson distribution uniquely define the first four moments of the distribution. The Johnson system also includes all possible combinations of skewness and kurtosis with $\beta_2 > \beta_1 + 1$.

Figure 9 displays a portion of the Johnson system, plotted as a function of β_1 and β_2, which is composed of three types denoted as regions in Figure 9. These types depend on the transform *g*: the lognormal (S_L), support bounded at both ends (S_B), and unbounded range at both ends (S_U). The S_L region, which is a transition type, is a generalized lognormal distribution, but the other two types do not correspond to common distributions. The normal distribution is a special case where all three types meet. Distributions in the upper part of the S_B region, between the dashed line and the impossible area, have bimodal "U" shapes, whereas distributions in the remainder of the system are unimodal.

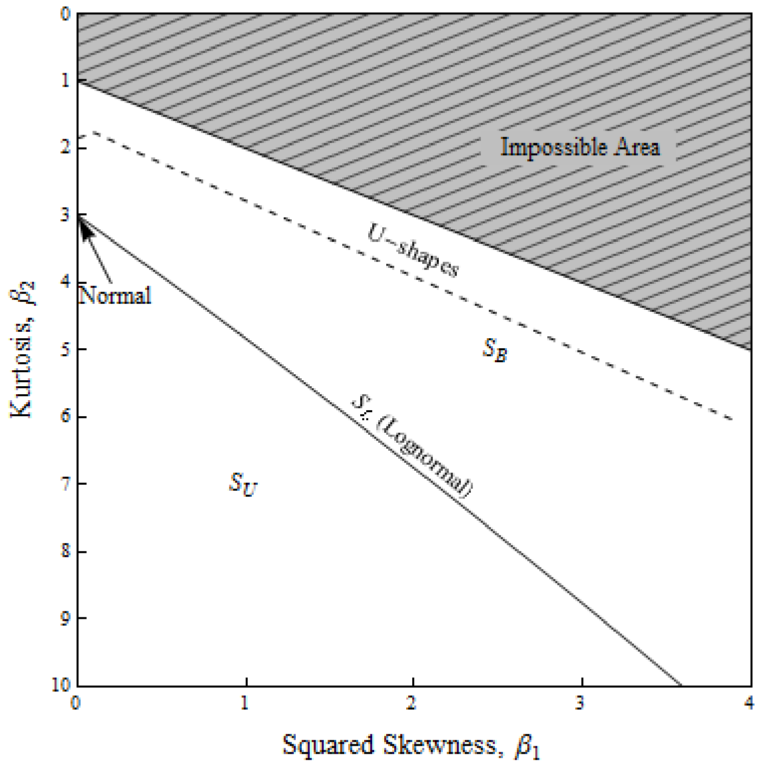

Figure 9. Johnson Distribution System.

Johnson Distribution Types

Letting $y = (x-\xi)/\lambda$, the transformations and distributions for each of the three types are given below. The moments of type S_L and S_U distributions can be calculated in closed form, and the S_B distribution moments by numerical integration. Equations for each are given below.

Bounded S_B

The S_B type is given by the transformation $g(x) = \ln\left(\frac{x}{1-x}\right), \quad 0 < x < 1.$

$$f_{S_B}(y) = \frac{\delta}{\sqrt{2\pi}\, y(1-y)} \exp\left\{-\tfrac{1}{2}\left(\gamma + \delta \ln\left(\frac{y}{1-y}\right)\right)^2\right\}, \quad 0 < y < 1. \qquad (34)$$

The S_B moments have very complex forms, but can be found by numerical integration using the pdf.

Lognormal S_L

The S_L type is given by the transformation $g(x) = \ln(x)$:

$$f_{S_L}(x) = \frac{\delta}{\sqrt{2\pi}y} \exp\left\{-\tfrac{1}{2}\left(\gamma + \delta \ln\left(y\right)\right)^2\right\}, \quad y > 0. \tag{35}$$

This can be expressed as a three-parameter distribution by re-parameterizing with $\gamma' = \gamma - \delta \ln\left(\lambda\right)$ yielding

$$\gamma + \delta \ln\left(\frac{x-\xi}{\lambda}\right) = \gamma' + \delta \ln\left(x - \xi\right), \tag{36}$$

(Elderton and Johnson, 1969). The r^{th} moment of the S_L distribution is given by

$$\mu_{S_L}(r) = \exp\left\{\tfrac{1}{2} r^2 \gamma^{-2} - r\gamma^{-1}\right\}. \tag{37}$$

Unbounded S_U

The S_U type is given by the transformation $g(x) = \sinh^{-1}(x) = \ln\left(x + \sqrt{x^2+1}\right)$;

$$f_{S_U}(y) = \frac{\delta}{\sqrt{2\pi}\sqrt{y^2+1}} \exp\left\{-\tfrac{1}{2}\left(\gamma + \delta \ln\left(y + \sqrt{y^2+1}\right)\right)^2\right\}, \quad -\infty < y < \infty. \tag{38}$$

The r^{th} even-moment of the S_U distribution is given by

$$\mu_{S_U}(r) = 2^{-(r-1)}\left(\sum_{s=0}^{r/2-1}(-1)^s \binom{r}{s} e^{\frac{1}{2}(r-2s)^2\delta^{-2}} \cosh\left[(r-2s)(\gamma/\delta)\right] + (-1)^{\frac{1}{2}r}\,\tfrac{1}{2}\binom{r}{r/2}\right), \tag{39}$$

and the r^{th} odd-moment is given by

$$\mu_{S_U}(r) = 2^{-(r-1)}\left(\sum_{s=0}^{(r-1)/2}(-1)^{s+1}\binom{r}{s} e^{\frac{1}{2}(r-2s)^2\delta^{-2}} \sinh\left[(r-2s)(\gamma/\delta)\right]\right). \tag{40}$$

Fitting Johnson Distributions to Moments

Matching Johnson distributions to moments is not as straightforward as for Pearson distributions, but Hill et al. (1976) gave algorithms for matching Johnson parameters to the first four moments for each of the three Johnson Types. Mathematica® code for these algorithms are given in the Appendix.

Comparison of the Pearson and Johnson Systems

Although the Pearson and Johnson systems have different derivations, functional forms, and region types on the (β_1, β_2) plane, distributions in either system having the same moments are very similar in shape. Although our results are similar between these two systems, considering both allows us to expand the scope of named distributions that we consider, namely to include the lognormal distribution and to circumvent the nonexistence of some higher moments in the Pearson system. The similarity can be demonstrated by considering the L_∞ distance measure, or Kolmogorov-Smirnov distance (Darling, 1957),

$$L_\infty\left(F,G\right) = \max_{x \in S}\left|F(x) - G(x)\right|, \qquad (41)$$

which is the maximum absolute error between cdfs F and G.

Figure 10 shows the L_∞ distance between standardized distributions in the Pearson and Johnson systems. In the figure, the regions and labels for both systems are overlaid. The L_∞ distance is less than 0.1 over most of the figure, with the largest differences between the U-shaped distributions. The thin region of higher (between 0.1 and 0.2) L_∞ distance in the middle of the figure is where the semi-bounded Pearson Type VI region overlaps with the unbounded Johnson S_U region.

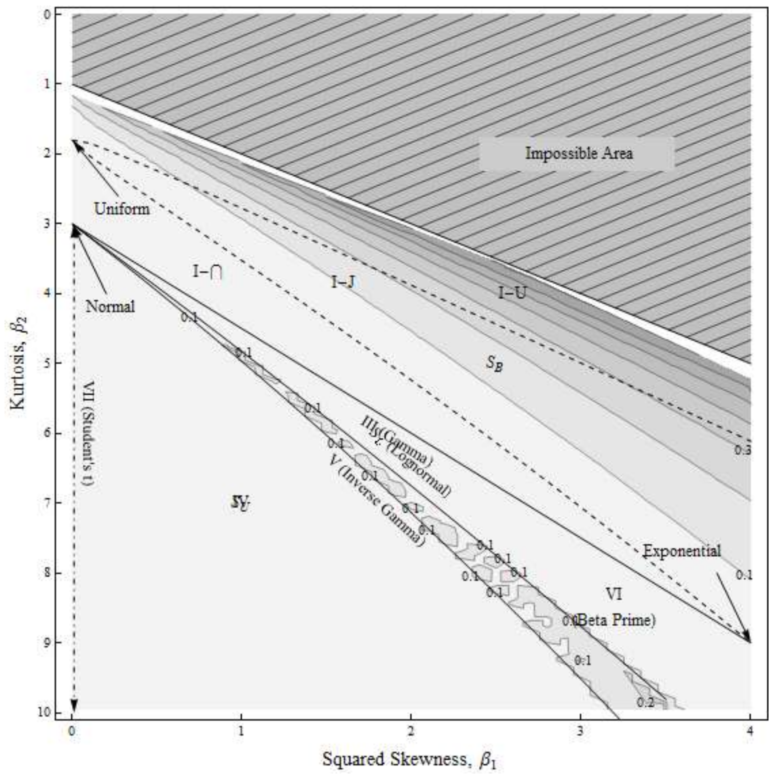

Figure 10. KS-distance between standardized Pearson and Johnson distributions.

Figure 11 shows two examples of bounded ∩-shape Pearson and Johnson distributions with the same first four moments, displaying almost identical pdfs. Figure 12 shows two examples of U-shaped distributions and indicates that the higher L_∞ distance for these shapes is due to slight boundary differences, amplified by the asymptotic boundary behavior for these distributions.

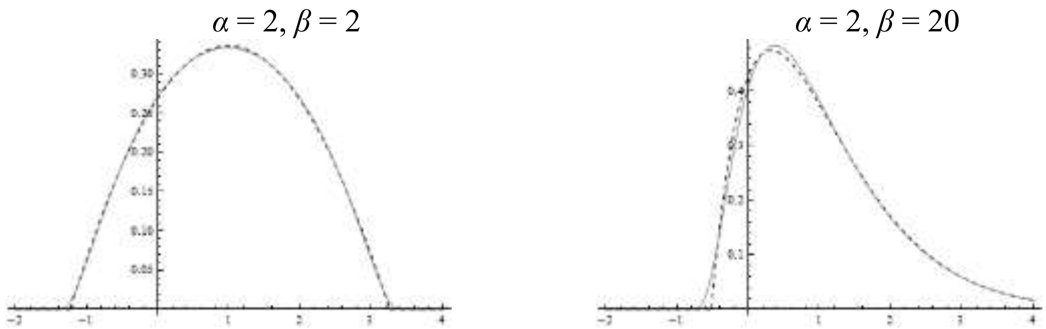

Figure 11. Comparison of two standardized ($\mu = \sigma = 1$) ∩-shape beta distributions (solid grey) to S_B distributions (dashed black) having the same first four moments.

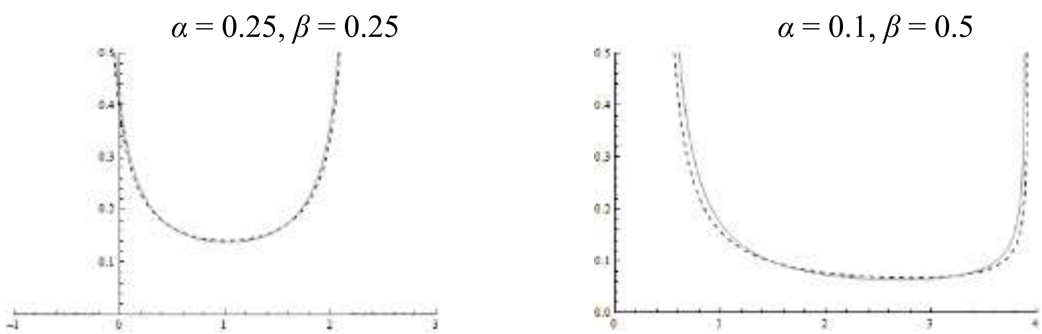

Figure 12. Comparison of two standardized ($\mu = \sigma = 1$) U-shape beta distributions (solid grey) to S_B distributions (dashed black) having the same first four moments.

SUMMARY

This chapter described the Pearson and Johnson distributions systems, which we use for our analyses in the following chapters. The combination of these systems includes many commonly used distributions as special cases. Although the shapes of distributions in each system are very similar, the systems each have unique qualities that make them convenient and useful for our analysis.

Chapter 4: Discretization Error[2]

This chapter explores the effects of discretization on different types and shapes of perfectly assessed distributions. As discussed in Chapter 2, comparison we will evaluate and compare discretization methods using their accuracy in estimating moments of distributions in the Pearson and Johnson systems. The overall results are similar between the two systems, because they produce very similar pdfs, as discussed in Chapter 3. However, the analyses have slightly different focuses. The Pearson system's division into more types gives more granularity to the analysis by shape, and the Johnson system allows greater focus on the lognormal distribution and its generalization. GQ is not included in the moment results because a three-point GQ will match the first five moments, which includes the moments of interest.

The next section describes our analysis method. We then analyze the two distribution systems separately, first the Pearson, then the Johnson. This analysis includes our new shortcut methods for each distribution system. This chapter concludes with a summary of the results and discussion of practical implications.

METHOD

We compute the discretizations and their errors for a large set of distributions in the Pearson and Johnson systems. These distributions are parameterized by (β_1, β_2), defining unique location-scale distributions in each system. The set of distributions is defined in (β_1, β_2) space by an evenly distributed grid of points. We standardize the distributions to have unit mean and variance, preserving the skewness, $\gamma_1 = \mu_3 / \sigma^3$, and kurtosis, $\beta_2 = \mu_4 / \sigma^4$, where μ_k is the k^{th} central moment, and σ is the standard deviation. This

---

2 Summaries of the results of this chapter are published with my advisor, Eric Bickel, in the following:

Hammond, Robert and J. Eric Bickel. 2013a. Reexamining discrete approximations to continuous distributions. *Decision Analysis* **10**(1) 6-25.

Hammond, Robert and J. Eric Bickel. 2013b. Approximating continuous probability distributions using the 10th, 50th, and 90th percentiles. *The Engineering Economist* **58**(3) 189-208.

allows flexibility in distribution shape, while still enabling comparison of location and scale. Because skewness can be positive or negative but is symmetric under reflection, it is convenient to consider squared-skewness, $\beta_1 = \gamma_1^2$. We use a range of $\beta_1 < 4, \beta_2 < 10$, slightly larger than that of Lau et al. (1998), to increase coverage of kurtosis and include more of the distributions considered by Pearson and Tukey (1965). This range is arbitrary, but our experimentation determined that increasing this range by 50% in each dimension does not change our conclusions.

Denoting the set of all Pearson distributions by Q_P, the (β_1, β_2) grid is defined as

$$\Gamma = \left\{ \left(\beta_1, \beta_2 \right) \mid \beta_1 = 0.1i, i = 0, ..., 40, \beta_2 = 0.1j, j = 10, ..., 100, \beta_1 + 1 \geq \beta_2 \right\} \tag{42}$$

This results in a grid of approximately 2,900 (β_1, β_2) points in the feasible area shown in Figure 3, spaced 0.1 in both dimensions. The corresponding set of Pearson distributions is

$$H_P = \left\{ f \mid \left(\mu'_3(f)^2, \mu'_4(f) \right) \in \Gamma, f \in Q_P \right\}. \tag{43}$$

The Johnson system has the same range of feasible (β_1, β_2) points, and we use the same grid to define the set of Johnson distributions,

$$H_J = \left\{ f \mid \left(\mu'_3(f)^2, \mu'_4(f) \right) \in \Gamma, f \in Q_J \right\}, \tag{44}$$

where Q_J denotes the set of all Johnson distributions.

Each (β_1, β_2) point corresponds to a unique standardized distribution. We discretize each of these distributions and calculate the absolute error in the mean, variance, skewness, and kurtosis. The approximations of the true mean (μ), variance (σ^2), skewness (γ_1), kurtosis (β_2), respectively, are

$$\hat{\mu} = p_1 F^{-1}(\pi_1) + p_2 F^{-1}(\pi_2) + p_3 F^{-1}(\pi_3), \tag{45}$$

$$\hat{\sigma} = p_1 \left(F^{-1}(\pi_1) \right)^2 + p_2 \left(F^{-1}(\pi_2) \right)^2 + p_3 \left(F^{-1}(\pi_3) \right)^2 - \hat{\mu}^2, \tag{46}$$

$$\hat{\gamma}_1 = \left(p_1 \left(F^{-1}(\pi_1) \right)^3 + p_2 \left(F^{-1}(\pi_2) \right)^3 + p_3 \left(F^{-1}(\pi_3) \right)^3 \right) / \hat{\sigma}^3 , \tag{47}$$

$$\hat{\beta}_2 = \left(p_1 \left(F^{-1}(\pi_1) \right)^4 + p_2 \left(F^{-1}(\pi_2) \right)^4 + p_3 \left(F^{-1}(\pi_3) \right)^4 \right) / \hat{\sigma}^4 . \tag{48}$$

We define the *discretization error* in the k^{th} as the difference between the true moment of distribution *f* and the approximate moment using discretization *d* as

$$\delta_d^k(f) = \mu_{D_d(f)}^k - \mu_f^k . \tag{49}$$

The *percentage error* is the discretization error divided by the true moment

$$\pi_d^k(f) = \frac{\mu_{D_d(f)}^k - \mu_f^k}{\mu_f^k}, \tag{50}$$

and we define 0/0 = 0 for this measure.

There is another reason for our distribution standardization. The beta distribution has the property that $F(\alpha,\beta) = 1 - F(\beta,\alpha)$. However, percentage error does not follow this relation for non-standardized beta distributions, and introduces bias, as pointed out by Lau et al. (1998). They standardized each distribution to have unit mean and variance. This practice, which we follow, eliminates the bias and allows for consistent comparison of errors between distributions with different support ranges. This standardization makes the errors in these moments equivalent to their respective percentage errors. For example,

$$\left|\mu_{D_{ESM}(f)}^1 - \hat{\mu}_f^1\right| = \frac{\left|\mu_{D_{ESM}(f)}^1 - \hat{\mu}_f^1\right|}{\left|\hat{\mu}_f^1\right|} . \tag{51}$$

We performed all of our numerical calculations using Mathematica® software. Numerical errors produced in these calculations are generally negligible, but they do appear in some of our results below. These errors are identified and discussed.

PEARSON SYSTEM

We first consider the Pearson distribution system. Our approach is based in large part on the results of Keefer and Bodily (1983).

Keefer and Bodily's Analysis

Table 21 gives the results for the I-∩ distributions, in the same manner as Keefer and Bodily (1983) (hereafter, KB). In this table and in each of the tables of results that follow, the lowest errors, to three decimal places, are highlighted. A selection of KB's results is repeated in Table 22 for comparison. Although each of EPT, ZDI, BMn3, and BMn5 has at least one measure of "0.000" for the tables, none is exactly equal to zero. The grey-shaded cells show the most accurate shortcut (top-portion of Table 21) and the most accurate distribution-specific method (bottom-portion of Table 21) for each measure. As noted by KB, EPT performs very well. ZDI has comparable performance to EPT and is better than ESM. Although MRO is similar to MCS, it is distinctly more accurate by all measures for both the mean and variance. MRO and ZDT perform similarly although they use very different weights. Both BMd methods perform rather poorly. The BMn3 and BMn5 methods perfectly match the mean (with negligible numerical error), but they are not as accurate as EPT in matching the variance.

All of the error measures are increased relative to KB’s results because of our expanded distribution set. For example, EPT’s AAPE in the mean and the variance are about three times larger in our case than reported by KB (0.066% compared to 0.020%). In addition, our expanded analysis demonstrates that EPT outperforms ESM more than was found by KB. For example, in our case, ESM’s AAPE is about five times as large as EPT, whereas KB found that it was only about two and one-half times as large.

BMn, in theory, matches the means of all of the distributions considered, but in practice the integrals involved in computing the conditional means often can be evaluated only using numerical integration methods, which introduces numerical error. However, good numerical integration software will generally produce negligible errors, which, in our case, are several orders of magnitude smaller than the smallest discretization errors. Also, in practical application, GQ will often have small numerical errors.

		Mean				Variance			
		AAE	AAPE	ME	MPE	AAE	AAPE	ME	MPE
Shortcuts Methods	EPT	0.000	0.066	-0.001	-0.178	0.000	1.096	-0.008	-10.090
	ZDI	0.000	0.152	-0.002	-0.589	0.000	1.189	-0.013	-16.094
	ESM	0.001	0.331	0.004	1.419	0.001	7.776	0.013	-19.960
	MCS	0.002	2.255	-0.005	-4.852	0.002	20.384	-0.005	-32.051
	MRO	0.000	0.621	-0.001	-1.647	0.000	9.601	0.002	-20.721
	ZDT	0.001	0.653	0.006	1.919	0.001	8.280	0.016	-21.253
Distributio n-Specific Methods	BMd3	0.005	5.124	-0.012	-10.494	0.004	40.390	-0.012	-53.752
	BMd5	0.003	3.027	-0.006	-6.390	0.002	25.961	-0.007	-38.402
	BMn3	0.000	0.000	0.000	0.000	0.002	22.201	-0.013	-29.199
	BMn5	0.000	0.000	0.000	0.000	0.001	13.042	-0.004	-20.526

Table 21. Errors in the mean and variance for expanded set of type I-∩ (beta) distributions.

	Mean				Variance			
	AAE	AAPE	ME	MPE	AAE	AAPE	ME	MPE
EPT	0.000	0.020	0.000	0.070	0.000	0.460	-0.001	-1.600
ESM	0.000	0.050	0.001	0.330	0.000	2.700	0.006	11.100
BMd5	0.001	0.750	-0.004	-3.350	0.002	21.500	-0.006	-30.200

Table 22. Selected results from KB (discretized – actual mean and variance).

General Shortcut Moment Errors

We turn now to the set of distributions H_P from the Pearson system, first considering the shortcut methods from the literature. The following tables give the AE, ASE, and ME statistics. We use these instead of KB's measures for various reasons. AE displays a method's tendency to under- or over-estimate a moment, which AAE does not (AAE and AAPE are equal for errors in the mean and variance as well). ASE describes magnitude, giving more emphasis to larger errors, and which we used to fit our new methods, as described in Chapter 2. We use ME to display the largest magnitude error, regardless of sign (again, ME is equal to MPE for the mean and variance).

Table 23 and Table 24 summarize our results for each of the main Pearson types (I, VI, and IV) and two transition types (III and V) over H_P. As before, the best measures are highlighted for each distribution type. In these tables, ASE's results are shown in scientific notation rather than rounded to three decimal places, since many would round to 0.000. These results identify the method that performs best, when performance is averaged over a given distribution region. Although different methods may have lower errors for specific distributions, these tables show which method has the lowest errors on average.

All six shortcuts tend to underestimate the moments of types III-VI, with the only exception being EPTs estimates of the type III mean. These four distributions all have at least one unbounded tail. EPT and ZDI tend to underestimate the mean of types I-U and I-J, but tend to overestimate the variance of both, along with the higher moments of type I-U. The shortcuts' performance is highly varied among type I-U and type I-J, whereas either EPT or ZDI performs best for type I-∩ and type III-VI distributions in all four moments. The only exception is type I-∩ skewness, for which MRO has the lowest ME. EPT and ZDI poorly estimate the skewness for distributions close to the uniform at $(\beta_1 = 0, \beta_2 = 1.8)$. Either EPT or ZDI is also best or second best for many measures in the I-J moments, although MRO has the lowest AE and ASE, and ESM the lowest ME, in the variance. ESM, MCS, ZDT, and MRO have larger errors in the skewness and kurtosis for all of the Pearson types, often larger than 0.500.

ZDI is distinctly the best in the skewness and kurtosis of types III-VI, due to its having the most extreme percentiles of all of these methods, although EPT generally has similar errors. The I-U distribution is the only type for which MCS performs best by any measure. Each method has its highest errors in the mean in the I-U distributions, and with the exception of MCS, their highest errors in the variance. ESM, MCS, ZDT, and MRO on average perform the best for these distributions, despite also having their largest MEs.

This performance may be due to those methods' having more probability weight on their upper and lower percentiles, making them more similar in shape to U-shaped distributions than are EPT and ZDI. The fact that ESM, MCS, ZDT, and MRO use less extreme percentiles than EPT and ZDI does not matter as much for these distributions, since their tails are bounded.

MRO performs better than MCS on all types except the U-shape beta, for which it performs slightly worse. ZDT performs slightly worse than ESM for types I and III, and nearly identically to ESM for types IV, V, and VI. ESM has lower error than MRO by all measures in the mean and variance for types III–VI, but higher errors in the skewness and kurtosis for these types. ESM has slightly higher errors in the mean and variance than does ZDT for types IV–VI, but slightly lower errors in skewness and kurtosis.

These results indicate that the method that is best at matching the mean is not necessarily the best at matching the variance or higher moments of a distribution. For example, EPT has the lowest error statistics for types III and VI for the mean, but not for the variance, although EPT's errors are only slightly larger than the best methods'. EPT was designed to approximate only the mean, but it preserves the variance better than, or nearly as well as, any of the shortcut methods. ZDI was designed to match the first five moments of the normal distribution (being the result of a three-point GQ) and often has the lowest errors in the moments for many other distributions as well.

Table 23 and Table 24 indicate that ESM more accurately matches the mean and variance for many more distributions than does MCS, but it has slightly higher error in skewness and kurtosis than does MCS. This is due to MCS's poorer performance in matching the variance, resulting in smaller denominators in Equations (47) and (48). Both ESM and MCS underestimate skewness and kurtosis, but MCS's greater underestimation of the variance lowers the effective error in skewness and kurtosis. Of the four shortcuts using percentiles close to the P10, P50, and P90, MRO has the lowest errors in skewness and kurtosis, since it uses the most extreme percentiles of these methods, the P8.5 and P91.5.

This analysis yields the conclusion that ZDI is generally the best shortcut method for matching moments, but EPT has nearly identical performance using slightly less extreme percentiles. Even the best methods for type I-U distributions have significant errors, which suggests that none of these shortcuts should be used for this type.

		Mean			Variance		
		AE	ASE	ME	AE	ASE	ME
I-U (Beta)	ESM	0.155	3.83E-02	0.443	0.303	1.31E-01	0.800
	MCS	0.041	1.24E-02	-0.351	0.154	5.73E-02	0.607
	EPT	-0.045	1.28E-02	-0.485	0.199	1.02E-01	-0.602
	ZDI	-0.086	1.78E-02	-0.523	0.139	8.68E-02	-0.641
	ZDT	0.203	5.76E-02	0.539	0.296	1.33E-01	0.845
	MRO	0.065	1.58E-02	-0.355	0.258	1.03E-01	0.658
I-J (Beta)	ESM	0.034	1.67E-03	0.094	-0.065	1.37E-02	-0.213
	MCS	-0.035	1.35E-03	-0.051	-0.195	4.54E-02	-0.331
	EPT	-0.005	9.44E-05	-0.041	0.099	2.00E-02	0.394
	ZDI	-0.017	4.90E-04	-0.070	0.119	2.57E-02	0.434
	ZDT	0.048	3.10E-03	0.118	-0.080	1.78E-02	-0.222
	MRO	0.002	1.39E-04	0.045	-0.064	1.19E-02	-0.214
I-∩ (Beta)	ESM	0.004	3.81E-05	0.018	-0.055	9.10E-03	-0.204
	MCS	-0.030	1.07E-03	-0.050	-0.206	4.62E-02	-0.324
	EPT	0.001	9.75E-07	-0.002	-0.003	2.12E-04	-0.085
	ZDI	-0.002	5.76E-06	-0.008	0.001	3.79E-04	-0.142
	ZDT	0.009	1.09E-04	0.025	-0.044	1.00E-02	-0.213
	MRO	-0.008	8.86E-05	-0.017	-0.094	1.21E-02	-0.207
III (Gamma)	ESM	-0.001	1.71E-06	-0.002	-0.118	1.74E-02	-0.213
	MCS	-0.036	1.46E-03	-0.051	-0.258	6.87E-02	-0.331
	EPT	0.000	1.84E-07	0.001	-0.020	5.22E-04	-0.037
	ZDI	-0.001	1.95E-06	-0.002	-0.003	1.26E-05	-0.005
	ZDT	0.003	1.04E-05	0.006	-0.114	1.76E-02	-0.223
	MRO	-0.013	1.74E-04	-0.017	-0.144	2.26E-02	-0.215
VI (Beta Prime)	ESM	-0.006	4.46E-05	-0.012	-0.166	2.93E-02	-0.221
	MCS	-0.042	1.84E-03	-0.051	-0.298	9.01E-02	-0.341
	EPT	-0.001	8.92E-07	-0.002	-0.045	2.32E-03	-0.076
	ZDI	-0.001	2.03E-06	-0.002	-0.021	5.89E-04	-0.048
	ZDT	-0.002	1.67E-05	-0.009	-0.166	2.96E-02	-0.226
	MRO	-0.017	2.99E-04	-0.021	-0.185	3.52E-02	-0.228
V (Inverse Gamma)	ESM	-0.007	5.67E-05	-0.012	-0.130	1.94E-02	-0.204
	MCS	-0.033	1.14E-03	-0.044	-0.272	7.54E-02	-0.331
	EPT	-0.001	1.67E-06	-0.002	-0.040	2.04E-03	-0.075
	ZDI	-0.001	1.40E-06	-0.002	-0.024	7.48E-04	-0.047
	ZDT	-0.005	2.83E-05	-0.009	-0.125	1.86E-02	-0.205
	MRO	-0.014	2.22E-04	-0.020	-0.158	2.68E-02	-0.220
IV	ESM	-0.007	6.01E-05	-0.012	-0.173	3.12E-02	-0.224
	MCS	-0.022	6.10E-04	-0.044	-0.309	9.66E-02	-0.353
	EPT	-0.002	3.37E-06	-0.003	-0.068	5.07E-03	-0.104
	ZDI	-0.001	1.57E-06	-0.002	-0.047	2.51E-03	-0.078
	ZDT	-0.006	4.30E-05	-0.010	-0.170	3.03E-02	-0.223
	MRO	-0.011	1.47E-04	-0.020	-0.197	3.98E-02	-0.243

Table 23. Shortcut method errors in the mean and variance for the main Pearson types.

		Skewness			Kurtosis		
		AE	ASE	ME	AE	ASE	ME
I-U (Beta)	ESM	-0.294	2.21E-01	2.787	-0.432	2.53E-01	-0.709
	MCS	-0.063	2.49E-01	4.159	-0.256	1.72E-01	1.029
	EPT	0.333	6.37E-01	6.507	0.146	2.56E-01	2.158
	ZDI	0.473	8.75E-01	7.354	0.321	4.10E-01	2.652
	ZDT	-0.431	2.71E-01	2.004	-0.513	3.14E-01	-0.752
	MRO	-0.050	2.56E-01	4.220	-0.247	1.69E-01	1.054
I-J (Beta)	ESM	-0.529	2.91E-01	-0.684	-0.655	4.45E-01	-0.809
	MCS	-0.369	1.55E-01	-0.574	-0.558	3.35E-01	-0.758
	EPT	-0.038	5.14E-02	1.032	-0.320	1.54E-01	-0.627
	ZDI	0.075	7.01E-02	1.315	-0.215	1.13E-01	0.719
	ZDT	-0.628	4.01E-01	-0.755	-0.701	5.03E-01	-0.832
	MRO	-0.348	1.42E-01	-0.553	-0.551	3.28E-01	-0.753
I-∩ (Beta)	ESM	-0.607	3.74E-01	-0.683	-0.594	3.74E-01	-0.802
	MCS	-0.466	2.27E-01	-0.572	-0.502	2.82E-01	-0.750
	EPT	-0.069	3.95E-02	0.882	-0.281	1.30E-01	-0.616
	ZDI	0.069	5.29E-02	1.166	-0.183	9.44E-02	0.630
	ZDT	-0.699	4.92E-01	-0.754	-0.638	4.25E-01	-0.827
	MRO	-0.430	1.96E-01	-0.550	-0.496	2.76E-01	-0.746
III (Gamma)	ESM	-0.660	4.46E-01	-0.687	-0.690	4.86E-01	-0.810
	MCS	-0.546	3.06E-01	-0.577	-0.619	3.97E-01	-0.759
	EPT	-0.229	5.60E-02	-0.315	-0.448	2.22E-01	-0.630
	ZDI	-0.119	1.86E-02	-0.228	-0.371	1.63E-01	-0.572
	ZDT	-0.733	5.51E-01	-0.756	-0.724	5.32E-01	-0.833
	MRO	-0.517	2.75E-01	-0.556	-0.614	3.91E-01	-0.755
IV (Beta Prime)	ESM	-0.715	5.12E-01	-0.764	-0.771	5.97E-01	-0.830
	MCS	-0.613	3.77E-01	-0.679	-0.717	5.19E-01	-0.791
	EPT	-0.327	1.10E-01	-0.432	-0.584	3.49E-01	-0.695
	ZDI	-0.229	5.64E-02	-0.346	-0.524	2.84E-01	-0.651
	ZDT	-0.782	6.11E-01	-0.820	-0.796	6.37E-01	-0.849
	MRO	-0.587	3.46E-01	-0.656	-0.713	5.13E-01	-0.789
V (Inverse Gamma)	ESM	-0.728	5.31E-01	-0.764	-0.713	5.17E-01	-0.827
	MCS	-0.629	3.97E-01	-0.679	-0.651	4.35E-01	-0.788
	EPT	-0.328	1.13E-01	-0.431	-0.505	2.76E-01	-0.690
	ZDI	-0.221	5.65E-02	-0.345	-0.440	2.19E-01	-0.646
	ZDT	-0.793	6.29E-01	-0.820	-0.743	5.59E-01	-0.847
	MRO	-0.601	3.62E-01	-0.656	-0.647	4.30E-01	-0.785
IV	ESM	-0.794	6.32E-01	-0.855	-0.765	5.90E-01	-0.833
	MCS	-0.719	5.19E-01	-0.802	-0.716	5.19E-01	-0.799
	EPT	-0.478	2.36E-01	-0.625	-0.607	3.80E-01	-0.728
	ZDI	-0.392	1.63E-01	-0.560	-0.560	3.28E-01	-0.698
	ZDT	-0.844	7.13E-01	-0.890	-0.789	6.26E-01	-0.850
	MRO	-0.696	4.87E-01	-0.785	-0.713	5.15E-01	-0.797

Table 24. Shortcut method errors in the skewness and kurtosis for the main Pearson types.

Figure 13 through Figure 18 show plots for the absolute percentage error $\left|\pi_d^k(f)\right|$ in the moments for the six shortcuts. These plots quantify the performance of the various methods as a function of distribution shape. The contours and error magnitude ranges are standardized separately for each moment throughout the error plots. Absolute error in the mean is shown to vary from 0 to 0.15, and in the variance from 0 to 0.5 (with our $\mu = \sigma = 1$ normalization, these are equivalent to 0% to 15% and 0% to 50% error ranges, respectively), with darker shading indicating higher absolute error. Skewness and kurtosis errors are both shown between 0 and 100%. Black areas in the plots indicate where the errors exceed the upper bound.

For most of the distributions in H_P, the errors of each method behave similarly, as functions of β_1 and β_2 for their respective moments. It is readily apparent that the highest errors are in the type I-U region. The error manifests a transition from high, to low, and back to high, in this type as β_1 increases. This is due to large overestimation error near the point $\left(\beta_1 = 0, \beta_2 = 1\right)$, which decreases to zero, and then to large underestimation error for larger β_1. Errors in the mean typically increase with both β_1 and β_2, and errors in variance and kurtosis increase primarily with β_2 (kurtosis), i.e., as the tails of the distributions get "fatter."

The EPT and ZDI errors are less sensitive to changes in skewness than are those of either ESM or MCS. With the exception of type I-U, EPT and ZDI generally match the variance better than either ESM or MCS. This undoubtedly stems from the fact that both EPT's and ZDI's percentiles capture more of the tail effects than do ESM's and MCS's P10 and P90. Both ESM and MCS display significantly higher error sensitivity to distribution shape than do either EPT or ZDI, and error generally increases with β_1 and β_2. However, MCS is much more sensitive to skewness than is ESM, because MCS places less weight on the P10 and P90. ESM clearly outperforms MCS for type I-∩, III, IV, V, and VI distributions. All of the shortcut methods produce large errors within the I-

U region, which is a strong indication that they should not be used for this Pearson type. Although ESM's performance is rather poor, it is superior to that of MCS. For example, although ESM misestimates the variance of type I-∩ in the middle of our plots ($\beta_1 = 2$ and $\beta_2 = 5$) by about 10%, MCS's error rate is 25%.

MRO is generally more accurate in the mean and variance than is MCS, but less accurate than EPT, ZDI, or ESM, and ZDT's performance on the variance is similar to that of ESM. ZDT actually performs quite well outside of the beta region, having absolute errors of less than 0.01 over all of the type VI and most of the type IV regions in the plot, and generally performing similar to ESM. ZDT's less extreme percentiles and higher weighting of these points (0.333) might help it better account for the tails.

With the exception of type I-U, and I-J in the case of ZDI, EPT and ZDI perform well over most of the plot area. ESM displays greater errors than EPT and ZDI over type I-U and I-J, and portions of the type IV, V, and VI distributions.

MCS results in the highest errors among the six shortcut methods, indicating that its justification for use on the mean may be weak and does not extend to other distributions.. We noted in Chapter 2 that MCS originated from an application of BMn with 25-50-25 brackets to the normal distribution. Figure 14 shows that although MCS matches the mean of the normal distribution (and all symmetric distributions, since it is a symmetric discretization), it poorly estimates the mean of most other distributions, even those having a shape similar to that of the normal distribution. It also underestimates the variance by almost 20% and the skewness and kurtosis by approximately 55% and 35%, respectively.

EPT and ZDI perform similarly over the type I region, but ZDI more accurately matches the variance for type IV, V, and VI distributions. The similarity in the EPT and ZDI errors is expected, considering the similarity of these methods' percentiles and probabilities. ZDI performs better over the type IV distributions, perhaps because it is a

GQ for the normal distribution, which is unbounded, as are the distributions in this region. Additionally, the formula used to derive EPT was designed by Pearson and Tukey (1965) based on performance over shapes mostly located in the type I and VI regions (as shown in Figure 3).

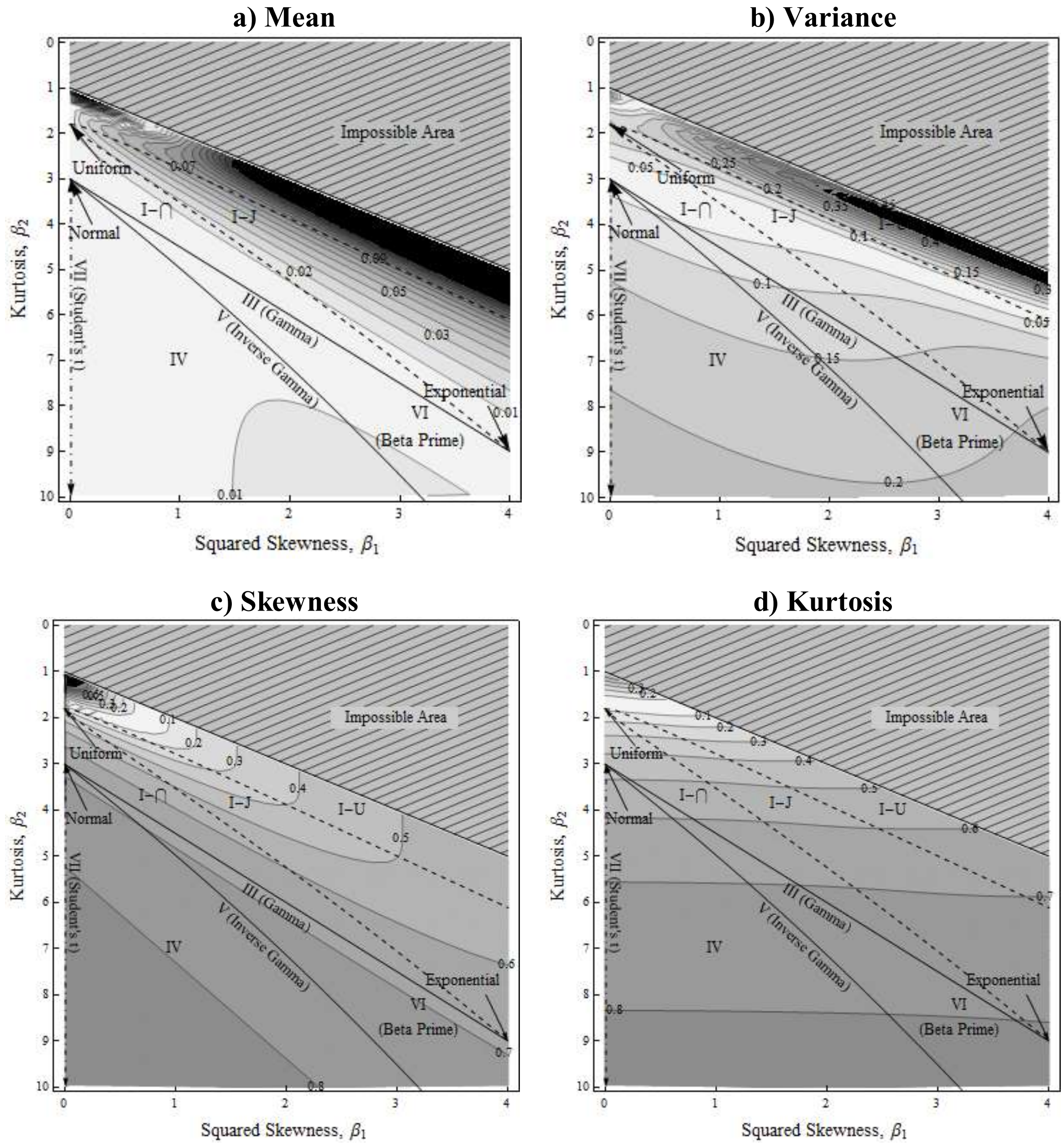

Figure 13. ESM errors in matching moments in the Pearson system.

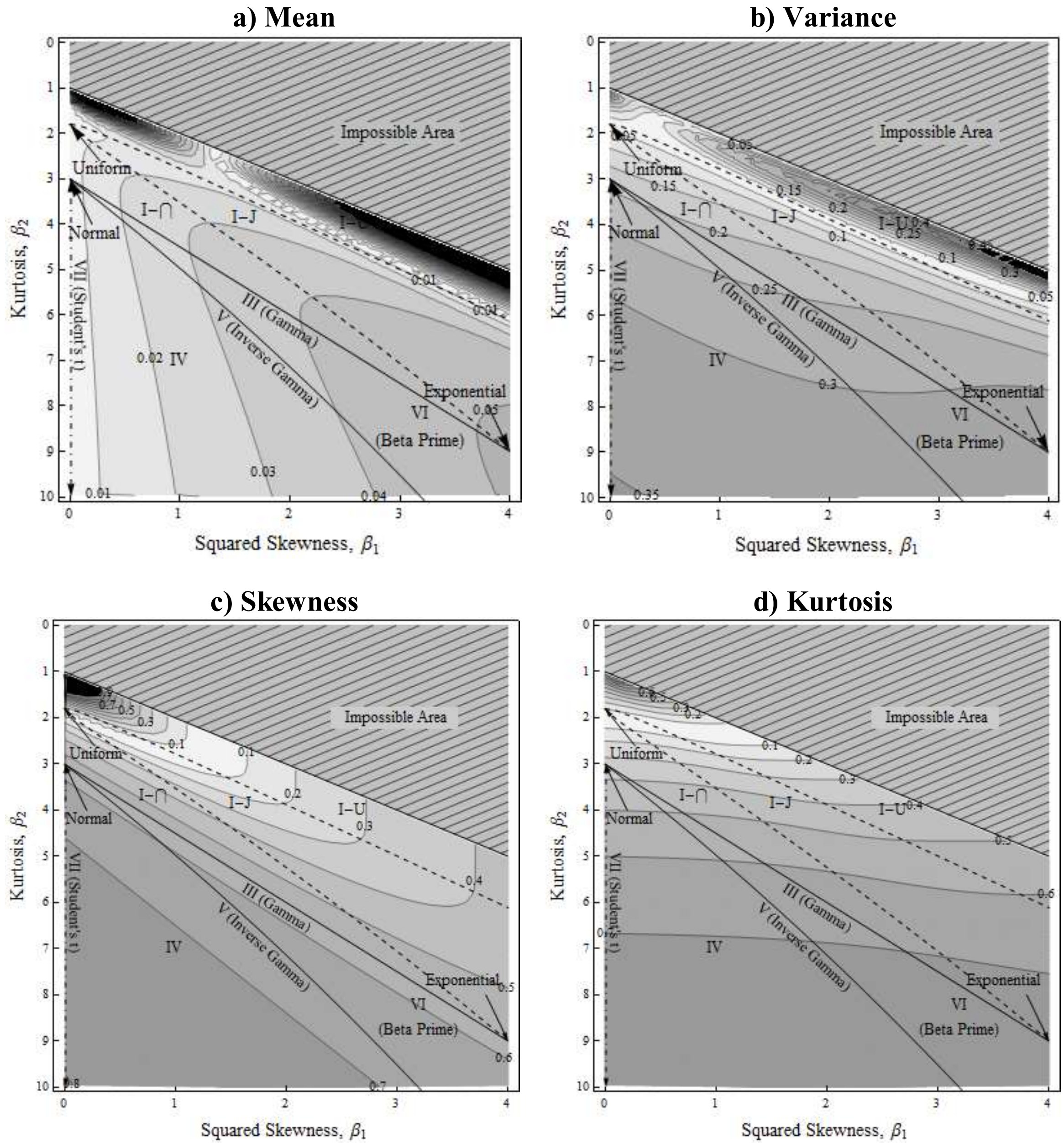

Figure 14. MCS errors in matching moments in the Pearson system.

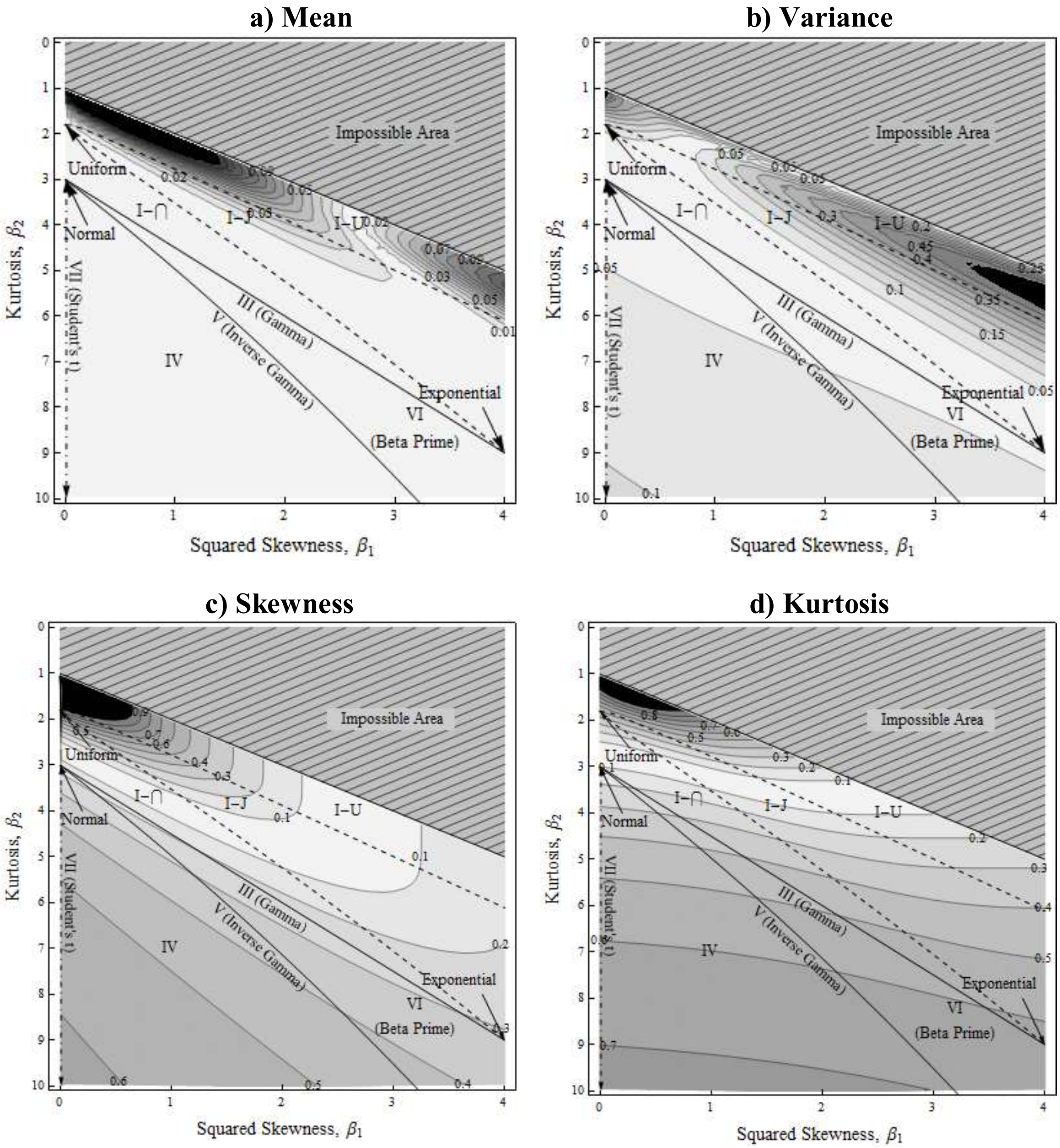

Figure 15. EPT errors in matching moments in the Pearson system.

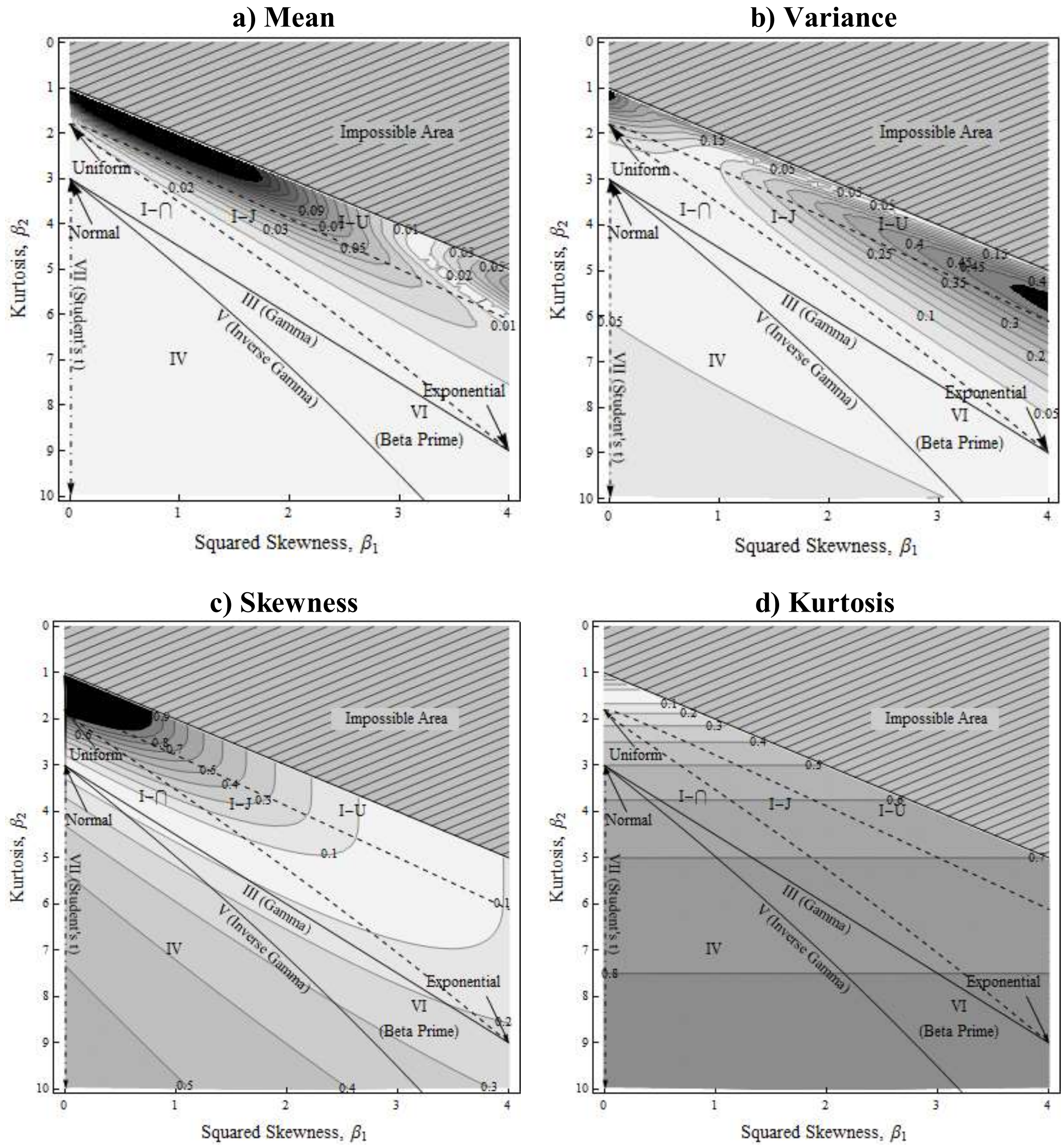

Figure 16. ZDI errors in matching moments in the Pearson system.

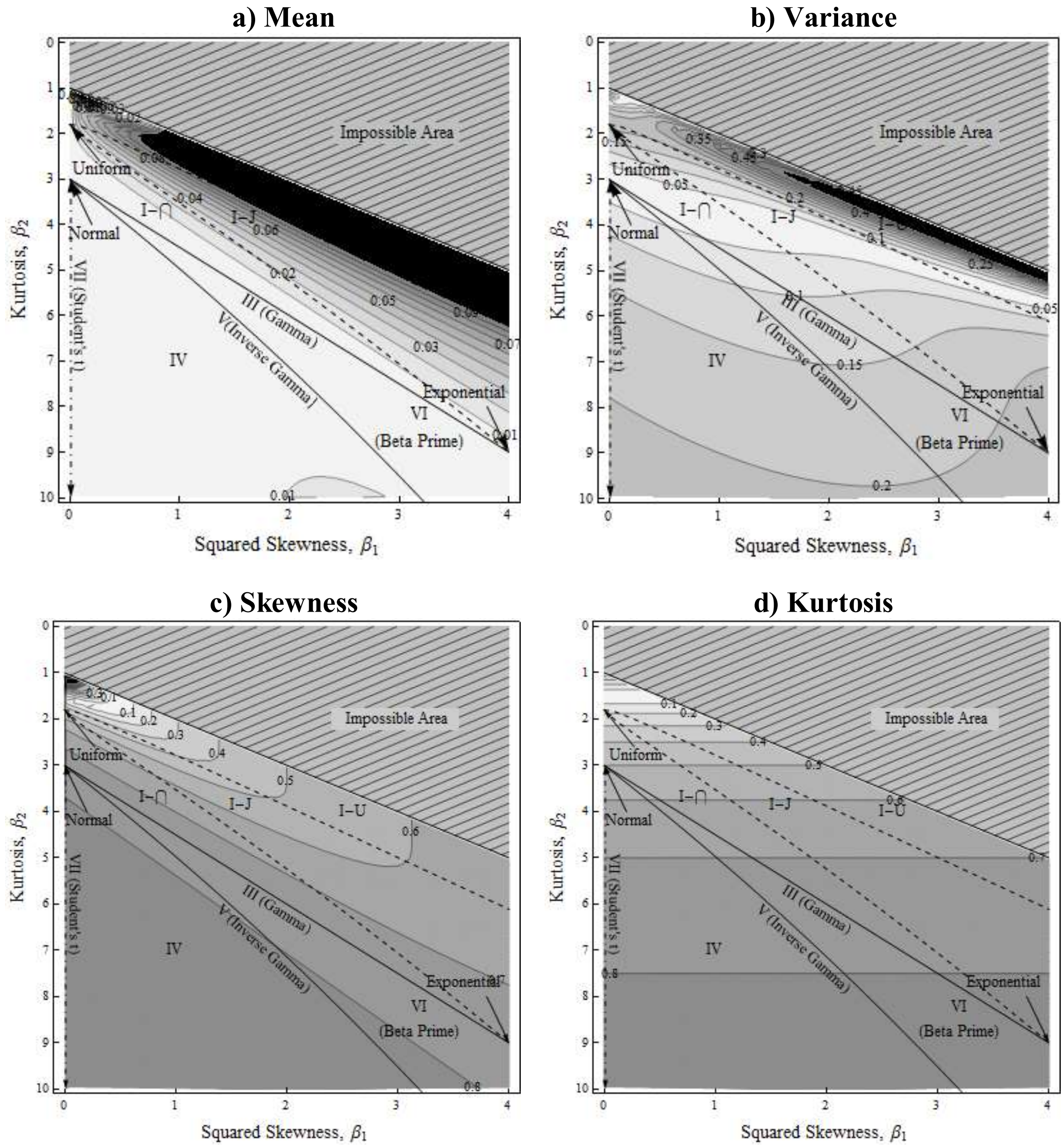

Figure 17. ZDT errors in matching moments in the Pearson system.

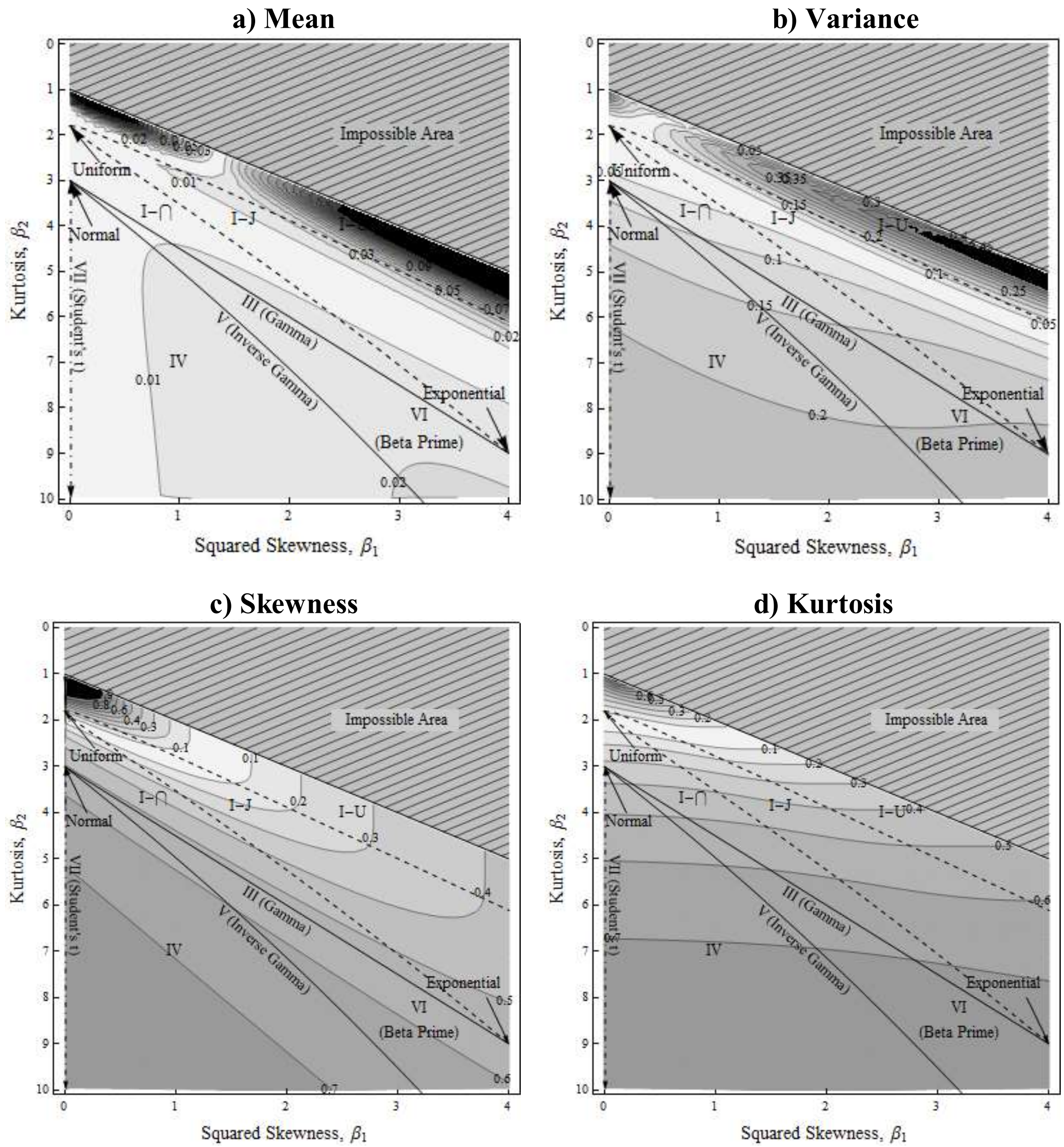

Figure 18. MRO errors in matching moments in the Pearson system.

New Shortcut Moment Errors

Here we consider the new shortcut methods described in Chapter 2. Summary statistics for the mean and variance are given in Table 25, and for the skewness and kurtosis in Table 26. Symmetric distributions are not included in the results for EPT++, because some of these methods are asymmetric and the percentage error measurements in the skewness would result in division by zero. All of these methods were optimized with respect to ASE in the mean. EPT+ has similar performance to that of EPT++, often outperforming EPT++. Both of these methods have lower ASEs in the mean than do any of the six shortcuts discussed in the previous section, except for type III and VI, where EPT has lower ASE. However, these improvements are not large. They also typically have lower AEs in the mean. EPT also has lower error by each measure in the variance for type I-∩. This similarity in performance indicates that EPT and ZDI are near-optimal in matching the moments among three-point shortcuts using the P50.

Although the EPT++ methods had the most flexibility to fit to the distributions in each type, rounding the probabilities to three decimal places increased the ASEs for these types to where the ASE for EPT++ exceeded that for EPT+. This indicates that the EPT++ methods may be over-fit and highly sensitive to the probability weights.

The similar performance of EPT+ and EPT++, which are identical for types I-∩ and VI, indicates that the extra freedom of allowing asymmetry for EPT++ does not significantly improve upon the shortcuts' estimates. In some cases, it has increased their estimates' sensitivity to the probability weights to the point that roundoff at three decimal places degrades their performance.

In contrast, the SP methods show distinct improvements in the mean and variance over ESM, MCS, ZDT, and MRO, often having error measures an order of magnitude lower than these methods. Although SP methods tend to have higher errors than do the other shortcuts in the skewness and kurtosis, they are usually within an order of magnitude.

The improvement in the mean, but in some cases degraded performance in the variance, is a result of our procedure's considering error only in matching the means of a set of distributions, without consideration of the variance. One could consider some weighting of the mean and the variance to develop other approximations. Indeed, Keefer (1994) considered all the moments via the computation of a certain equivalent. However, developing a shortcut method that preserved certain equivalents would require knowledge of the decision maker's utility function, which is likely to differ widely across decision makers and decision situations.

		Mean			Variance		
		AE	ASE	ME	AE	ASE	ME
I-U (Beta)	EPT1U+	-0.004	9.11E-03	0.319	-0.167	1.00E-01	-0.832
	EPT1U++*	-0.001	8.59E-03	0.312	-0.183	1.04E-01	-0.833
	SP1U	-0.009	9.68E-03	-0.396	0.080	3.69E-02	-0.509
I-J (Beta)	EPT1J+	-0.001	3.44E-05	0.034	0.050	1.13E-02	0.312
	EPT1J++	0.001	2.55E-05	0.030	0.036	9.98E-03	0.300
	SP1J	-0.003	2.32E-04	0.037	-0.134	2.60E-02	-0.276
I-∩ (Beta)	EPT1∩+	0.000	2.70E-07	-0.003	-0.008	2.69E-04	-0.089
	EPT1∩++	0.000	2.70E-07	-0.003	-0.008	2.69E-04	-0.089
	SP1∩	0.001	2.18E-05	0.016	-0.067	1.04E-02	-0.213
III (Gamma)	EPT3+	-0.001	6.27E-07	-0.001	-0.025	7.42E-04	-0.042
	EPT3++	-0.001	1.25E-06	-0.001	0.249	7.76E-02	0.441
	SP3	0.000	6.53E-07	0.002	-0.112	1.62E-02	-0.209
VI (Beta Prime)	EPT6+	0.001	1.05E-06	0.001	-0.004	1.34E-04	-0.029
	EPT6++	0.001	1.05E-06	0.001	-0.004	1.34E-04	-0.029
	SP6	0.000	9.79E-06	0.007	-0.145	2.29E-02	-0.202
V (Inv. Gam.)	EPT5+	0.000	7.05E-08	-0.001	-0.014	3.19E-04	-0.034
	EPT5++	0.000	5.56E-09	0.000	-0.012	2.85E-04	-0.036
	SP5	0.000	2.69E-06	-0.003	-0.091	1.10E-02	-0.168
IV	EPT4+	0.000	1.08E-06	0.002	-0.073	5.98E-03	-0.114
	EPT4++*	0.001	1.31E-06	0.003	-0.063	4.80E-03	-0.121
	SP4	0.000	3.53E-06	0.005	-0.113	1.43E-02	-0.167

*Symmetric distributions not included in these results.

Table 25. New shortcut method errors in the mean and variance for the main Pearson types.

		Skewness			Kurtosis		
		AE	ASE	ME	AE	ASE	ME
I-U (Beta)	EPT1U+	-0.286	2.17E-01	2.887	-0.421	2.46E-01	-0.703
	EPT1U++*	-0.309	2.81E-01	-3.829	-0.406	2.43E-01	0.894
	SP1U	0.054	3.17E-01	4.865	-0.149	1.57E-01	1.331
I-J (Beta)	EPT1J+	-0.138	5.78E-02	0.780	-0.403	2.03E-01	-0.673
	EPT1J++	-0.134	5.51E-02	0.715	-0.394	1.98E-01	-0.670
	SP1J	-0.446	2.14E-01	-0.627	-0.608	3.89E-01	-0.784
I-∩ (Beta)	EPT1∩+	-0.063	3.93E-02	0.893	-0.277	1.28E-01	-0.613
	EPT1∩++	-0.063	3.93E-02	0.893	-0.277	1.28E-01	-0.613
	SP1∩	-0.597	3.62E-01	-0.675	-0.588	3.67E-01	-0.799
III (Gamma)	EPT3+	-0.225	5.40E-02	-0.312	-0.444	2.19E-01	-0.627
	EPT3++	0.662	5.50E-01	1.712	0.242	1.18E-01	0.668
	SP3	-0.664	4.51E-01	-0.691	-0.692	4.89E-01	-0.811
VI (Beta Prime)	EPT6+	-0.210	4.86E-02	-0.330	-0.513	2.74E-01	-0.643
	EPT6++	-0.210	4.86E-02	-0.330	-0.513	2.74E-01	-0.643
	SP6	-0.730	5.33E-01	-0.776	-0.777	6.07E-01	-0.835
V (Inv. Gam.)	EPT5+	-0.195	4.65E-02	-0.324	-0.425	2.07E-01	-0.635
	EPT5++	-0.221	7.91E-02	-0.407	-0.537	3.06E-01	-0.709
	SP5	-0.752	5.67E-01	-0.785	-0.726	5.36E-01	-0.836
IV	EPT4+	-0.577	3.38E-01	-0.698	-0.661	4.46E-01	-0.763
	EPT4++*	-0.462	2.53E-01	-0.600	-0.674	4.63E-01	-0.775
	SP4	-0.823	6.78E-01	-0.876	-0.782	6.15E-01	-0.845

*Symmetric distributions not included in these results.

Table 26. New shortcut method errors in the skewness and kurtosis for the main Pearson types.

The absolute percentage errors in the first four moments are plotted for EPT+ and EPT++ in Figure 19 and Figure 20, respectively. The error in each region is shown only for the shortcut corresponding to that region. The contour levels for each moment are consistent with the plots in the previous section. The scalloping along the regional borders is a result of our discrete sampling of these regions, and of our transitioning to a new discretization along these borders.

Comparing Figure 19 and Figure 20 to Figure 15 shows that EPT+ and EPT++'s errors in the moments are nearly identical to those for EPT. Although the EPT++ methods for type I-U, I-J, and IV distributions differ from those for EPT+, they bring negligible improvement in accuracy.

The absolute error in the distribution mean in panel a of Figure 19 and Figure 20 is less than 0.01 over almost the entire plot, with the exception of the type I-U region. For this region, the tailored EPT1U+ shortcut does not provide much improvement over the errors seen in Figure 2. In terms of the mean, the primary benefit of EPT1J+ over EPT is a reduction in error within the type I-J region. EPT+ reduces the error in the variance, compared to EPT, for the type VI and most of the type I-J region. There is, however, an increase in error for type I-J distributions with relatively high skew and kurtosis. In addition, error in the variance is increased slightly within the type IV region. Errors in the skewness and kurtosis tend to be slightly increased as well.

Plots of the absolute percentage error in the moments for the SP shortcuts are shown in Figure 21. Comparing these to ESM and MCS in Figure 2 and Figure 3, respectively, shows the significant improvement made by the SP shortcuts, especially over MCS. There is less contrast between SP and ESM, being that the SP shortcuts use similar probabilities as ESM, but the tailoring to individual regions better preserves the mean in each region. The same is true for the variance in all but the type I-J region, where SP1J performs slightly worse than ESM. As noted earlier, SP tends to have higher errors in skewness and kurtosis.

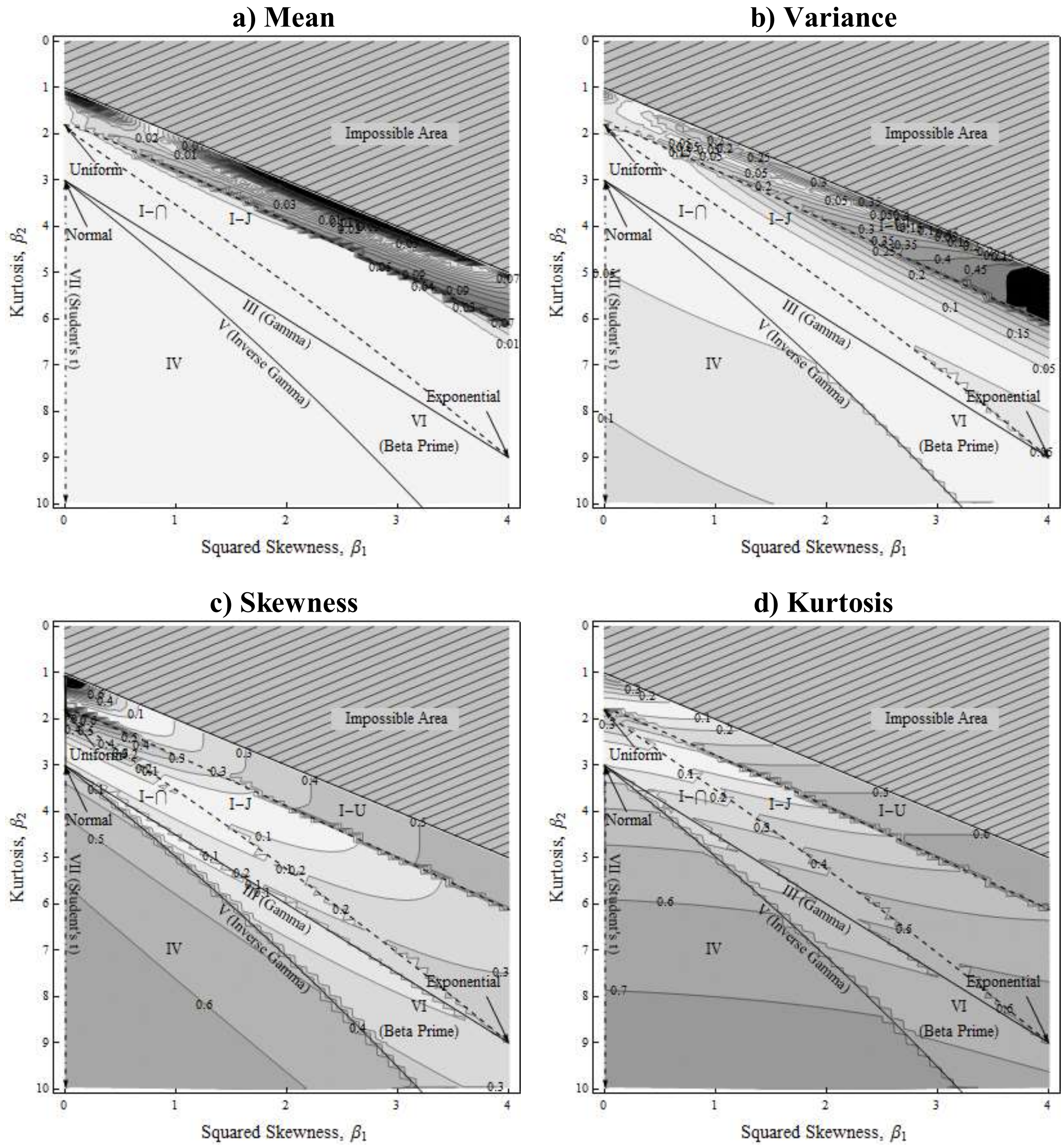

Figure 19. EPT+ errors in matching moments in the Pearson system.

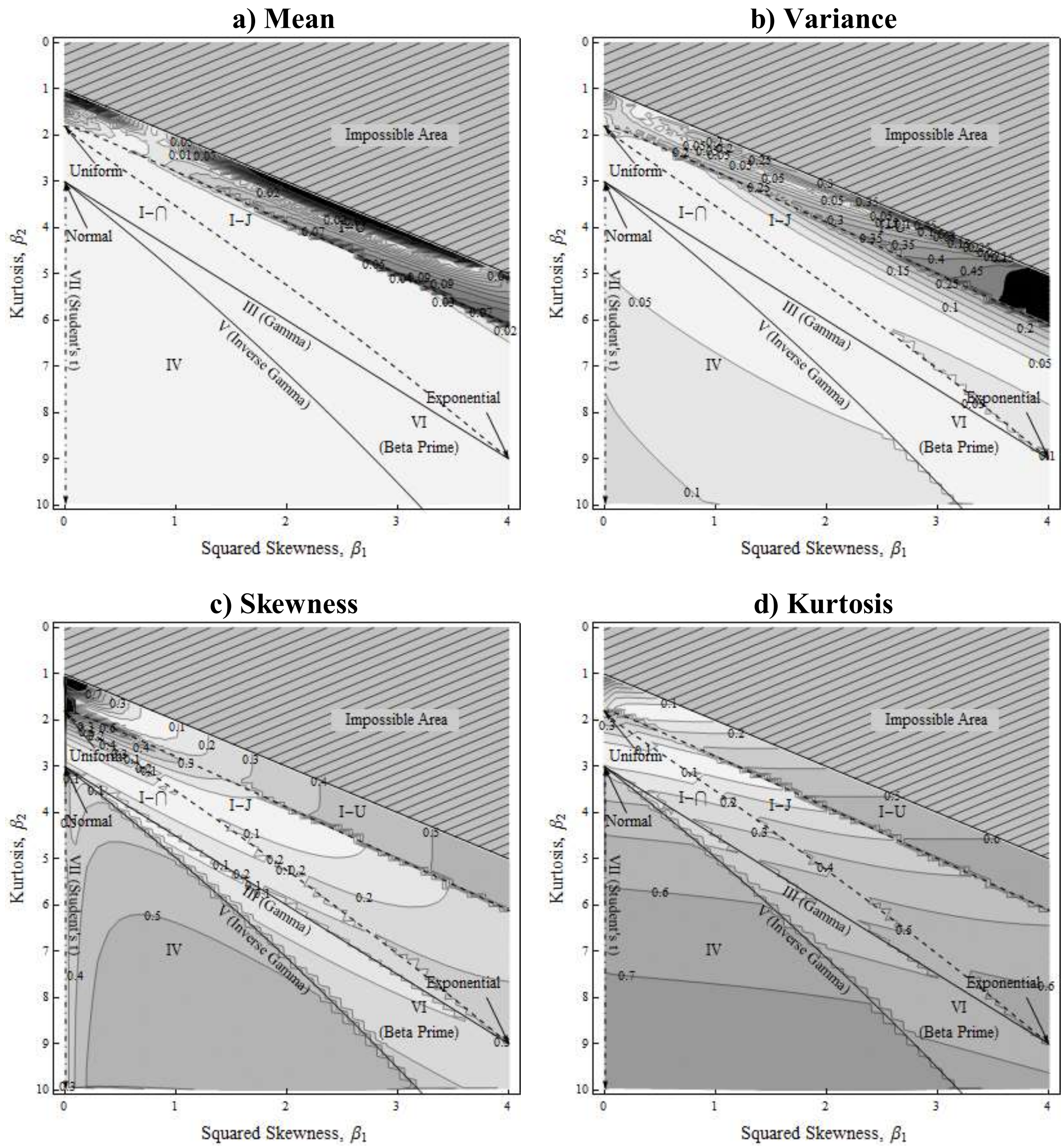

Figure 20. EPT++ errors in matching moments in the Pearson system.

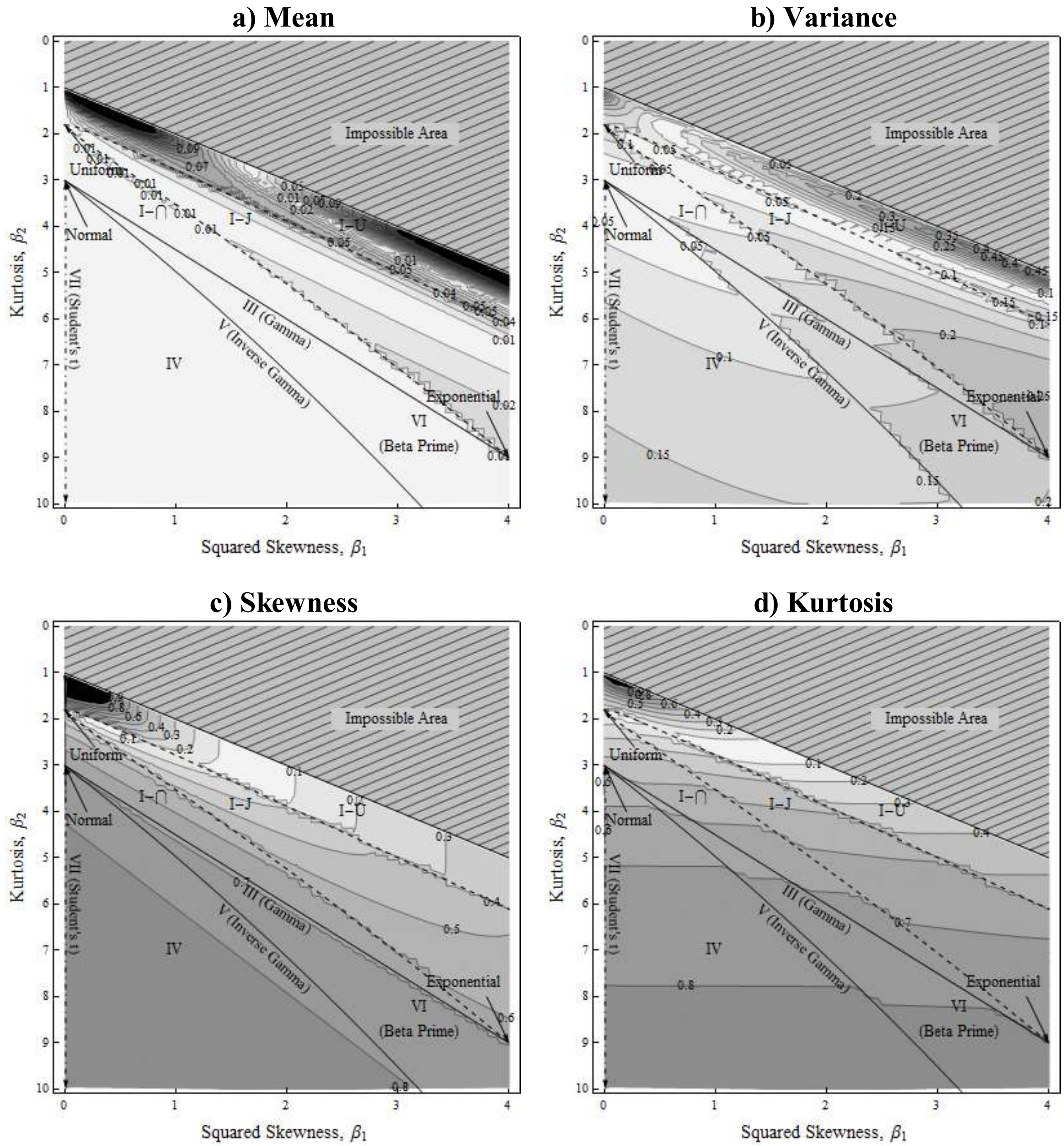

Figure 21. SP errors in matching moments in the Pearson system.

Comparing Individual Distributions

We saw previously that fitting the new shortcuts to minimize ASE did not also minimize other error statistics. Nor does it necessarily minimize error for individual distributions, as we discuss here. We compare individual distributions from the Pearson system, showing for exactly which distributions the errors are the lowest. Figure 22 shows which of EPT++, EPT+, or EPT best matches the mean or variance for specific distributions by the white, light gray, or dark grey regions, respectively. Tailoring EPT+ and EPT++ to specific distribution types in most cases improves performance in matching the mean. No one shortcut dominates the others over any entire region, and the shortcut most accurate in the mean of a distribution is often not the most accurate for that distribution's variance.

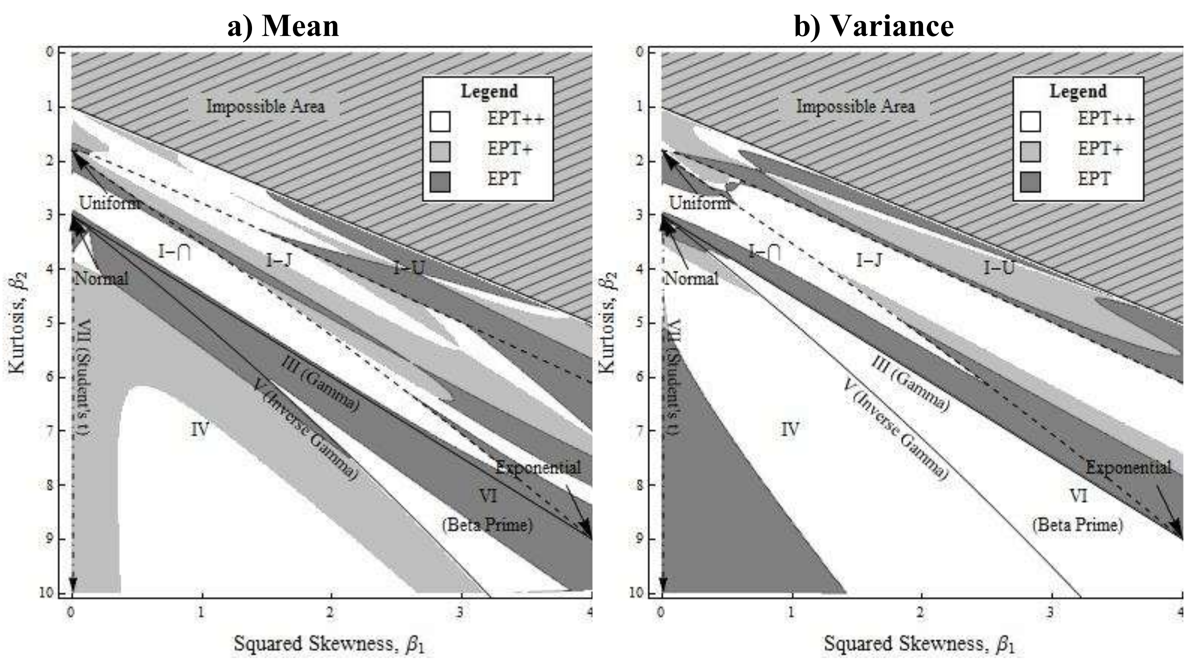

Figure 22. Regions where EPT, EPT+, or EPT++ is the most accurate.

Figure 23 shows which of SP, ESM, and MCS, all using the same percentiles, is most accurate in estimating the mean or variance for a given distribution as indicated by white, light grey, or dark grey regions, respectively. The SP discretizations are the best of the three over most of the region considered, and are the most accurate for both the mean and the variance over nearly the entire type IV and VI regions. SP is a distinct improvement over ESM and MCS. MCS is better than the other two only for parts of the type I-U region, where it still displays significant error. ESM is most accurate in the mean only for small portions of the type I-J and I-∩ regions, but is most accurate in the variance over most of these same regions.

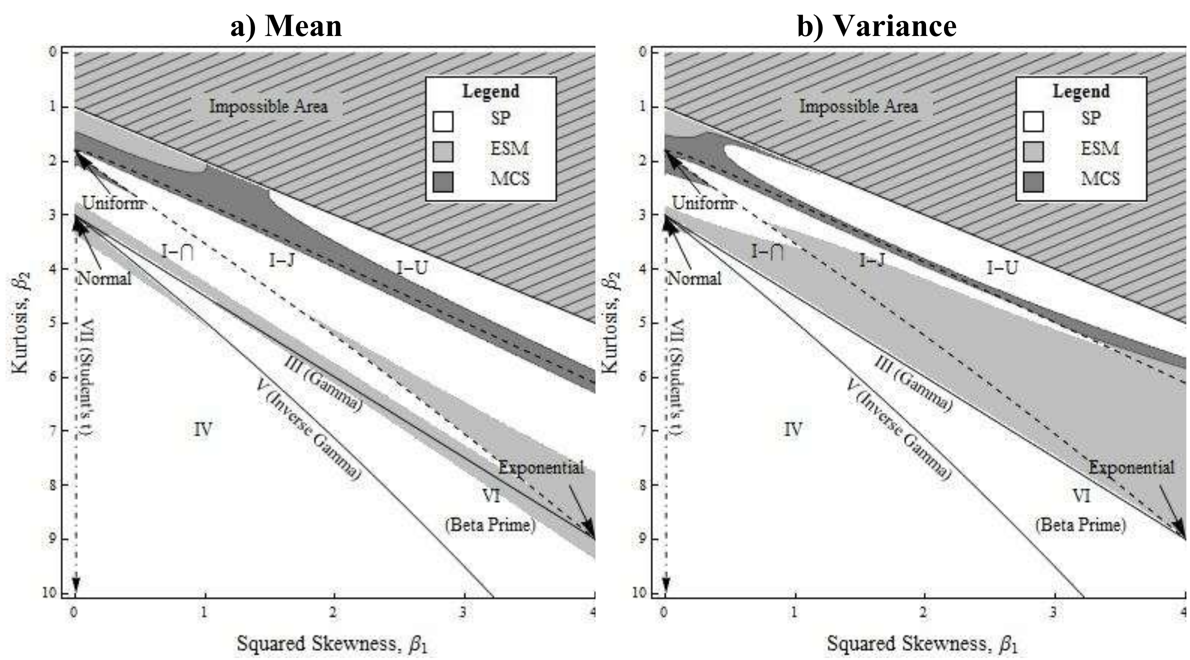

Figure 23. Regions where ESM, MCS, or SP is the Most Accurate.

Distribution-Specific Method Moment Errors

Finally, we turn our attention to the distribution specific methods. The AE, ASE, and ME statistics are shown in Table 27 for mean and variance, and in Table 28 for skewness and kurtosis.

As expected, the BMn methods better match the mean than do BMd or the shortcuts. The mean error statistics for both BMn methods are highlighted, since both theoretically match the mean, and any discrepancy is purely numerical error. Table 27 indicates that all of the numerical integration errors for the BMn methods average less than 10^{-3} and are generally several orders of magnitude smaller than the discretization errors for the BMd and shortcut methods. Miller and Rice (1983) proved that BMn will systematically underestimate variance and kurtosis, which is evident in Table 27 and Table 28, respectively. Table 28 also shows that BMn underestimates the skewness. The tables also show that five points are generally better than three. In every statistic, for each distribution type and moment, BMd5 has lower error than BMd3, and BMn5 has lower error than BMn3.

The best of these methods in matching the avriance, BMn5, has errors two to three orders of magnitude higher than those of EPT and ZDI in Table 23 and Table 24, respectively. BMn5 also has larger errors than EPT and ZDI in skewness and kurtosis, except for the type I-U distributions. ESM, ZDT, and MRO's errors in variance are also typically lower than those of BMn5, but are higher in skewness and kurtosis. BMn provides improved accuracy in the mean at the cost of reduced accuracy in the variance and higher moments, as compared to the best shortcuts.

		Mean			Variance		
		AE	ASE	ME	AE	ASE	ME
I-U (Beta)	BMd3	-0.020	9.23E-03	-0.335	-0.288	1.78E-01	-0.982
	BMd5	0.004	1.45E-03	0.156	-0.007	1.82E-02	0.371
	BMn3	0.000	1.18E-10	0.000	-0.207	4.90E-02	-0.446
	BMn5	0.000	1.19E-10	0.000	-0.103	1.29E-02	-0.262
I-J (Beta)	BMd3	-0.097	9.84E-03	-0.121	-0.466	2.27E-01	-0.616
	BMd5	-0.051	2.74E-03	-0.068	-0.284	8.73E-02	-0.395
	BMn3	0.000	8.63E-12	0.000	-0.248	6.27E-02	-0.299
	BMn5	0.000	7.62E-12	0.000	-0.156	2.58E-02	-0.212
I-∩ (Beta)	BMd3	-0.068	5.35E-03	-0.108	-0.404	1.69E-01	-0.538
	BMd5	-0.040	1.88E-03	-0.066	-0.260	7.21E-02	-0.384
	BMn3	0.000	8.69E-14	0.000	-0.222	5.04E-02	-0.292
	BMn5	0.000	4.48E-14	0.000	-0.130	1.84E-02	-0.205
III (Gamma)	BMd3	-0.077	6.71E-03	-0.111	-0.466	2.19E-01	-0.549
	BMd5	-0.047	2.52E-03	-0.068	-0.317	1.03E-01	-0.395
	BMn3	0.000	2.80E-13	0.000	-0.248	6.26E-02	-0.299
	BMn5	0.000	1.24E-13	0.000	-0.160	2.67E-02	-0.213
VI (Beta Prime)	BMd3	-0.087	7.74E-03	-0.111	-0.504	2.55E-01	-0.549
	BMd5	-0.055	3.06E-03	-0.068	-0.357	1.29E-01	-0.401
	BMn3	0.000	4.01E-14	0.000	-0.271	7.37E-02	-0.309
	BMn5	0.000	1.69E-14	0.000	-0.185	3.48E-02	-0.230
V (Inverse Gamma)	BMd3	-0.065	4.46E-03	-0.085	-0.473	2.26E-01	-0.532
	BMd5	-0.042	1.84E-03	-0.055	-0.329	1.10E-01	-0.389
	BMn3	0.000	4.41E-20	0.000	-0.248	6.23E-02	-0.286
	BMn5	0.000	4.14E-21	0.000	-0.163	2.73E-02	-0.204
IV	BMd3	-0.042	2.20E-03	0.084	-0.508	2.59E-01	-0.548
	BMd5	-0.028	9.54E-04	0.055	-0.366	1.35E-01	-0.409
	BMn3	0.000	6.34E-09	0.000	-0.262	6.87E-02	-0.287
	BMn5	0.000	3.47E-12	0.000	-0.180	3.28E-02	-0.205

Table 27. Distribution-specific method errors in the mean and variance for the main Pearson types.

		Skewness			Kurtosis		
		AE	ASE	ME	AE	ASE	ME
Beta (U)	BMd3	-0.439	2.74E-01	2.004	-0.513	3.14E-01	-0.752
	BMd5	-0.097	2.94E-02	0.639	-0.170	6.73E-02	0.552
	BMn3	-0.141	1.07E-01	1.051	-0.266	1.67E-01	0.905
	BMn5	-0.125	2.65E-02	-0.312	-0.185	6.59E-02	-0.494
Beta (J)	BMd3	-0.664	4.47E-01	-0.792	-0.701	5.03E-01	-0.832
	BMd5	-0.416	1.84E-01	-0.609	-0.533	3.03E-01	-0.748
	BMn3	-0.399	1.72E-01	-0.575	-0.562	3.39E-01	-0.758
	BMn5	-0.376	1.50E-01	-0.542	-0.517	2.86E-01	-0.728
Beta (N)	BMd3	-0.754	5.71E-01	-0.798	-0.638	4.25E-01	-0.827
	BMd5	-0.542	2.98E-01	-0.609	-0.521	2.94E-01	-0.742
	BMn3	-0.467	2.26E-01	-0.571	-0.502	2.82E-01	-0.750
	BMn5	-0.455	2.11E-01	-0.540	-0.498	2.70E-01	-0.721
III (Gamma)	BMd3	-0.778	6.20E-01	-0.802	-0.724	5.32E-01	-0.833
	BMd5	-0.596	3.64E-01	-0.612	-0.626	4.03E-01	-0.750
	BMn3	-0.536	2.95E-01	-0.577	-0.618	3.96E-01	-0.759
	BMn5	-0.510	2.67E-01	-0.545	-0.603	3.75E-01	-0.730
VI (Beta Prime)	BMd3	-0.822	6.76E-01	-0.854	-0.796	6.37E-01	-0.849
	BMd5	-0.654	4.29E-01	-0.710	-0.719	5.20E-01	-0.792
	BMn3	-0.599	3.59E-01	-0.659	-0.716	5.17E-01	-0.790
	BMn5	-0.570	3.25E-01	-0.628	-0.698	4.91E-01	-0.775
V (Inverse Gamma)	BMd3	-0.834	6.96E-01	-0.854	-0.743	5.59E-01	-0.847
	BMd5	-0.672	4.53E-01	-0.711	-0.660	4.46E-01	-0.789
	BMn3	-0.607	3.69E-01	-0.658	-0.650	4.34E-01	-0.787
	BMn5	-0.579	3.36E-01	-0.628	-0.640	4.20E-01	-0.772
IV	BMd3	-0.876	7.68E-01	-0.914	-0.789	6.26E-01	-0.850
	BMd5	-0.749	5.62E-01	-0.820	-0.724	5.29E-01	-0.804
	BMn3	-0.692	4.81E-01	-0.776	-0.716	5.19E-01	-0.799
	BMn5	-0.662	4.41E-01	-0.748	-0.707	5.06E-01	-0.793

Table 28. Distribution-specific method errors in the skewness and kurtosis for the main Pearson types.

Figure 24 through Figure 27 show the absolute percentage errors in the moments for BMd3, Bmd5, BMn3, and Bm5, respectively. The contour levels are the same as those for the shortcuts in the previous sections.

BMd3 and BMd5 produced significant errors over most of H_P. BMd3 underestimated the variance by more than 50% in approximately half of the area of the plot in Figure 24b. This occurs because the conditional distributions are skewed, and therefore the conditional median is not equal to the conditional mean. Below the type I-J region, error increased with skewness. Adding more points clearly improved the performance of BMd, although it was still inferior to the shortcut methods. For example, BMd5 still performed worse than does MCS, albeit slightly.

Figure 26b and Figure 27b imply that the BMn errors in the variance are far less sensitive to distribution shape than are those for the shortcut methods. Error in the variance is primarily a function of kurtosis for both BMd and BMn. BMd's performance was especially poor, although error was reduced by adding more discretization points. Yet, even a five-point BMn can underestimate the variance of low-skew type I-∩ distributions by more than 10%. This suggests that a different discretization approach may be needed if preserving the variance is important. Even though BMd and BMn are tailored to the underlying distribution, the errors in the variance for each of these methods, over most of H_P, were significantly larger than those for EPT and ZDI. However, the BMn methods exhibited more gradual increases in error in and around the type I-U region than did the shortcut methods.

Although the BMn3 method provided the foundation for MCS, the errors in the variance, skewness, and kurtosis in Figure 26b through d are significant, even though BMn3 produced a discretization that is specific to a given distribution. MCS, in contrast, applied a single set of percentiles over this whole region, producing generally higher errors in all of the moments, as seen by comparing Figure 26 to Figure 14.

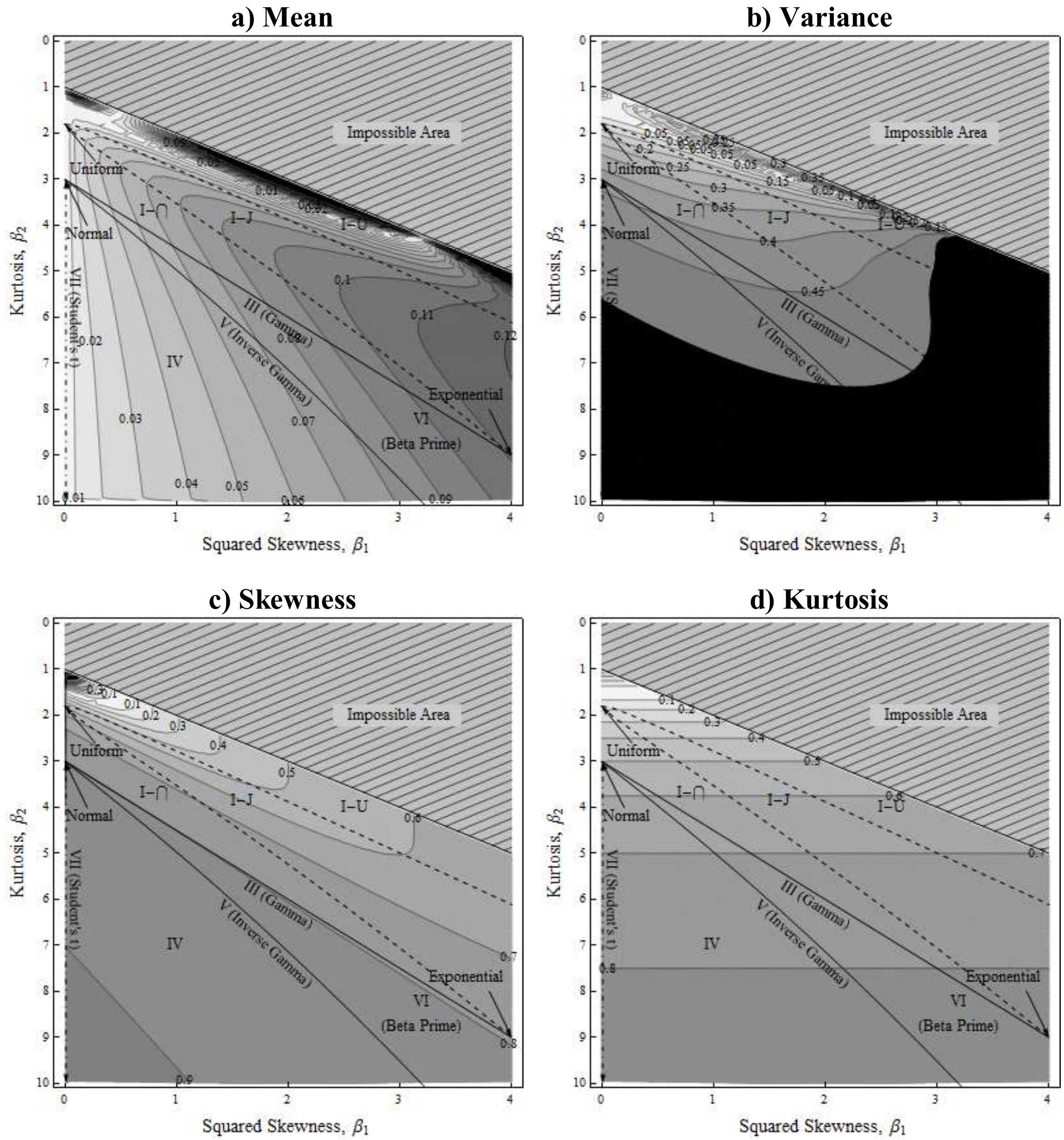

Figure 24. BMd3 errors in matching moments in the Pearson system.

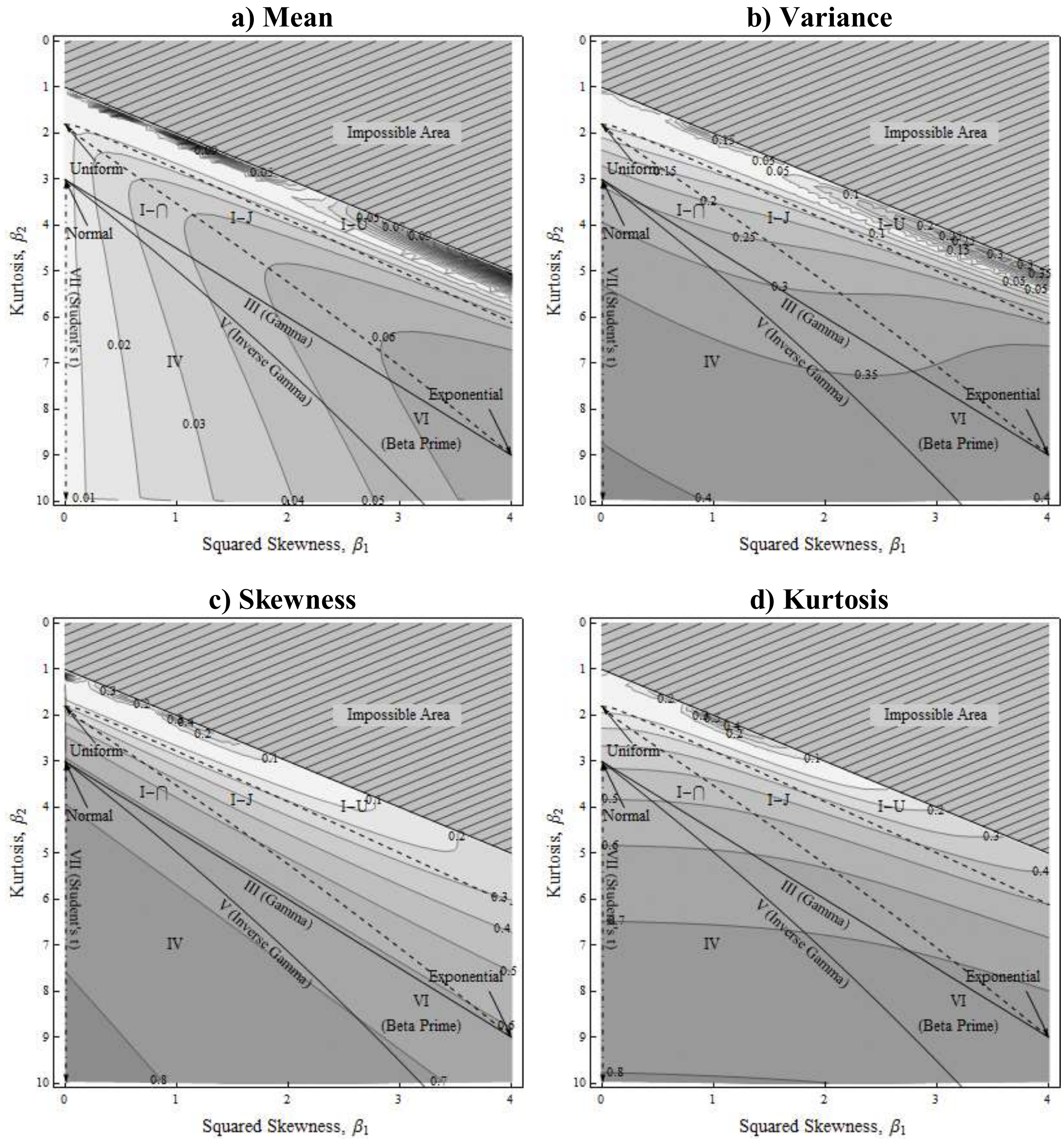

Figure 25. BMd5 errors in matching moments in the Pearson system.

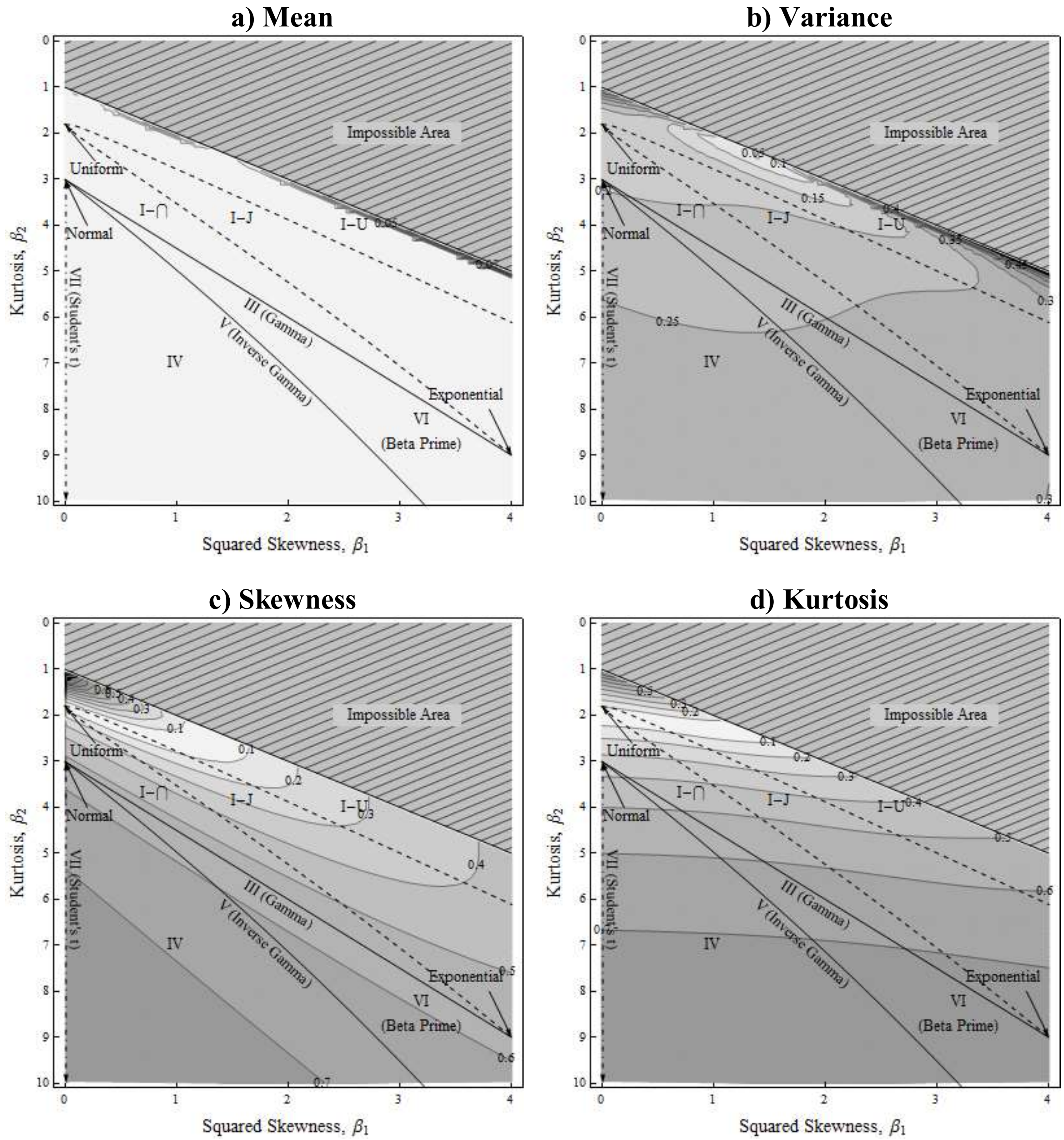

Figure 26. BMn3 errors in matching moments in the Pearson system.

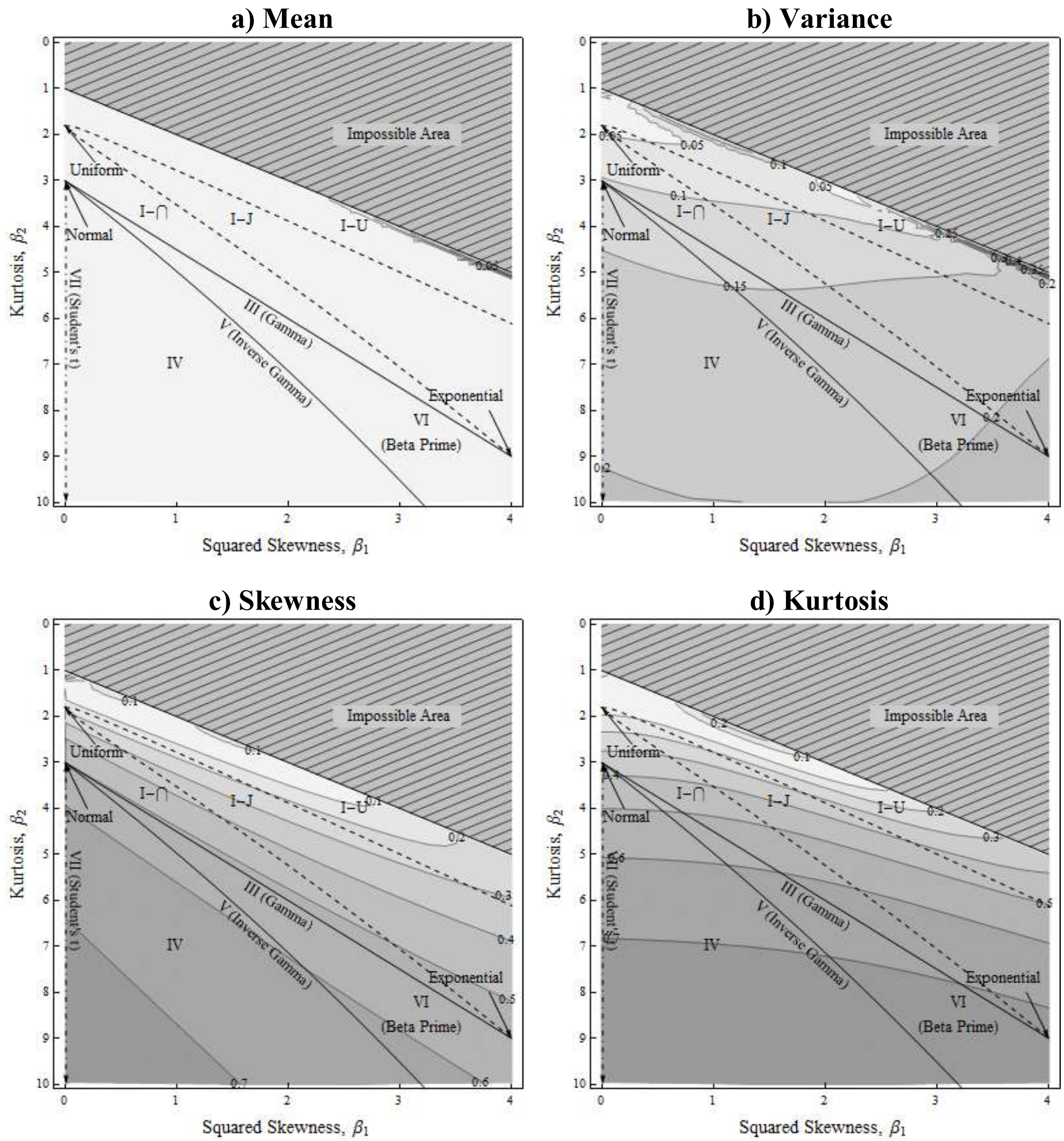

Figure 27. BMn5 errors in matching moments in the Pearson system.

JOHNSON SYSTEM

We used the same framework for the Johnson system as for the Pearson system. Many of the conclusions from the Pearson system extend to the Johnson system as well. We give all of the corresponding results for discretization error, but focus on the generalized lognormal distribution, S_L, given its importance in many applications.

General Shortcut Moment Errors

The errors in the mean and variance for the shortcuts are given in Table 29, and the errors in the skewness and kurtosis in Table 30. The results in these tables for the S_B type do not include U-shaped distributions. As seen previously with the Pearson distribution, shortcuts generally performed very poorly on U-shapes, which would skew the results for this type.

EPT and ZDI generally had the lowest average errors in all four moments for each of the three Johnson types. These two shortcuts had the two lowest errors by each measure for S_L and S_U, and had the two best ASEs for S_B. In the variance for S_B, ESM had a lower ME and ZDT a lower AE. MRO had a lower ME in the skewness, where ZDI actually had the highest ME of these methods. Figure 31c shows that this is due to ZDI's high error in the skewness near $(\beta_1 = 0, \beta_2 = 1)$ and that the error over most of the S_B region was much lower. EPT displays similar performance in Figure 30c.

		Mean			Variance		
		AE	ASE	ME	AE	ASE	ME
S_B	ESM	0.016	8.41E-04	0.095	-0.097	1.96E-02	-0.234
	MCS	-0.034	1.31E-03	-0.051	-0.233	6.22E-02	-0.353
	EPT	0.003	9.25E-05	0.046	0.048	1.21E-02	0.399
	ZDI	-0.003	1.26E-04	-0.066	0.068	1.52E-02	0.443
	ZDT	0.049	3.68E-03	0.160	-0.009	1.22E-02	0.289
	MRO	-0.003	2.27E-04	0.051	-0.107	1.99E-02	-0.239
S_L	ESM	-0.006	4.39E-05	-0.010	-0.131	2.03E-02	-0.215
	MCS	-0.034	1.25E-03	-0.047	-0.271	7.57E-02	-0.339
	EPT	-0.001	7.06E-07	-0.002	-0.036	1.77E-03	-0.071
	ZDI	-0.001	1.04E-06	-0.001	-0.020	5.08E-04	-0.040
	ZDT	0.013	1.64E-04	0.014	-0.038	5.56E-03	-0.133
	MRO	-0.014	2.24E-04	-0.021	-0.158	2.70E-02	-0.226
S_U	ESM	0.006	3.88E-05	0.011	-0.202	4.36E-02	-0.294
	MCS	0.023	6.40E-04	0.046	-0.333	1.13E-01	-0.411
	EPT	0.001	1.16E-06	0.002	-0.073	6.13E-03	-0.130
	ZDI	0.001	7.15E-07	0.001	-0.046	2.51E-03	-0.091
	ZDT	-0.006	4.68E-05	-0.014	-0.115	1.66E-02	-0.215
	MRO	0.010	1.32E-04	0.021	-0.220	5.02E-02	-0.300

Note: S_B U-shapes are not included.

Table 29. Shortcut method errors in the mean and variance for the Johnson types.

		Skewness			Kurtosis		
		AE	ASE	ME	AE	ASE	ME
S_B	ESM	-0.600	3.76E-01	-0.748	-0.672	4.69E-01	-0.829
	MCS	-0.464	2.38E-01	-0.658	-0.588	3.72E-01	-0.788
	EPT	-0.139	6.45E-02	0.946	-0.385	2.01E-01	-0.685
	ZDI	-0.028	5.68E-02	1.203	-0.295	1.54E-01	0.665
	ZDT	-0.679	4.73E-01	-0.800	-0.711	5.20E-01	-0.848
	MRO	-0.438	2.16E-01	-0.635	-0.582	3.66E-01	-0.785
S_L	ESM	-0.716	5.14E-01	-0.749	-0.709	5.12E-01	-0.827
	MCS	-0.613	3.77E-01	-0.659	-0.645	4.29E-01	-0.787
	EPT	-0.306	9.89E-02	-0.409	-0.494	2.67E-01	-0.684
	ZDI	-0.198	4.70E-02	-0.323	-0.427	2.10E-01	-0.637
	ZDT	-0.775	6.01E-01	-0.801	-0.740	5.55E-01	-0.847
	MRO	-0.585	3.43E-01	-0.637	-0.641	4.24E-01	-0.783
S_U	ESM	-0.735	5.72E-01	-0.833	-0.767	5.92E-01	-0.833
	MCS	-0.657	4.58E-01	-0.772	-0.718	5.22E-01	-0.800
	EPT	-0.426	1.98E-01	-0.590	-0.608	3.81E-01	-0.730
	ZDI	-0.344	1.33E-01	-0.525	-0.560	3.28E-01	-0.700
	ZDT	-0.779	6.42E-01	-0.868	-0.791	6.29E-01	-0.850
	MRO	-0.636	4.29E-01	-0.755	-0.715	5.18E-01	-0.798

Note: S_B U-shapes are not included.

Table 30. Shortcut method errors in the skewness and kurtosis for the Johnson types.

The absolute percentage errors for the shortcuts in Figure 28 through Figure 33 show very similar results for the Johnson system as those seen for the Pearson system in Figure 13 through Figure 18. The methods performed poorly over the U-shaped distributions, near the impossible area, and the error contours typically show very similar behaviors as functions of β_1 and β_2.

Figure 28a indicates that ESM quite accurately preserved the mean of the lognormal distribution throughout the range of shapes that we considered, with errors generally less than 1% (the lognormal distribution on which Hurst et al. (2000) demonstrated ESM corresponds to the point $(\beta_1 = 0.377, \beta_2 = 3.678)$). However, the lower right corner of Figure 28b shows that ESM's error in the variance exceeds 20% and is less than 5% only near the normal distribution when squared skewness is less than about 0.5. Figure 28c and Figure 28d indicate that, even for lognormal distributions close to the normal distribution, ESM’s skewness and kurtosis errors exceeded 60% and 40%, respectively. This indicates that ESM accurately matched the mean and variance for lognormal distributions with very low skewness, but did not accurately represent the shape of skewed distributions. As with the Pearson system, Figure 29 shows that MCS incurred significant errors of about 17% in the variance and 30% in the kurtosis on the normal distribution. MCS’s errors in these moments rise quickly as skewness and kurtosis increase along the S_L line.

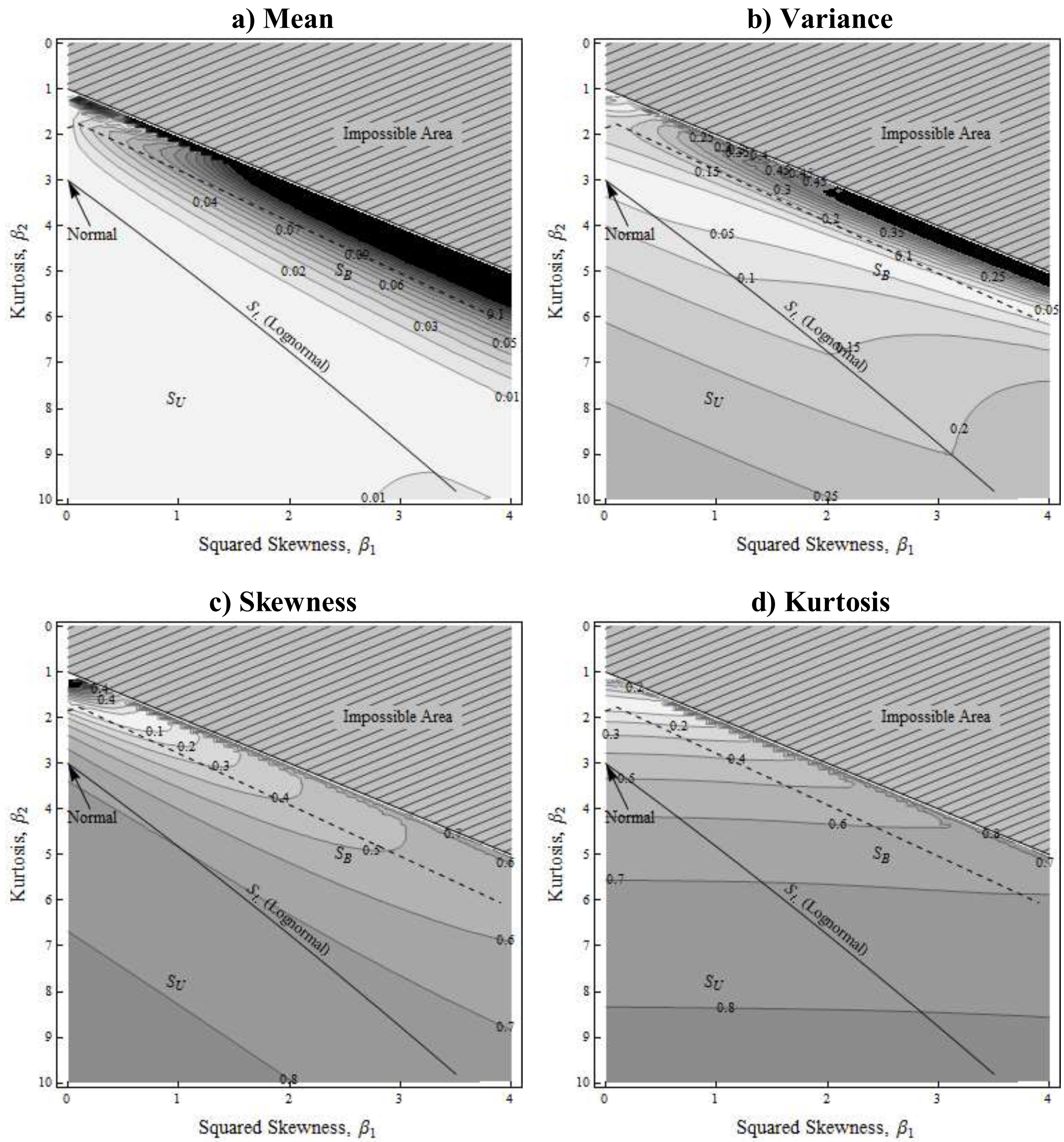

Figure 28. ESM errors in matching moments in the Johnson system.

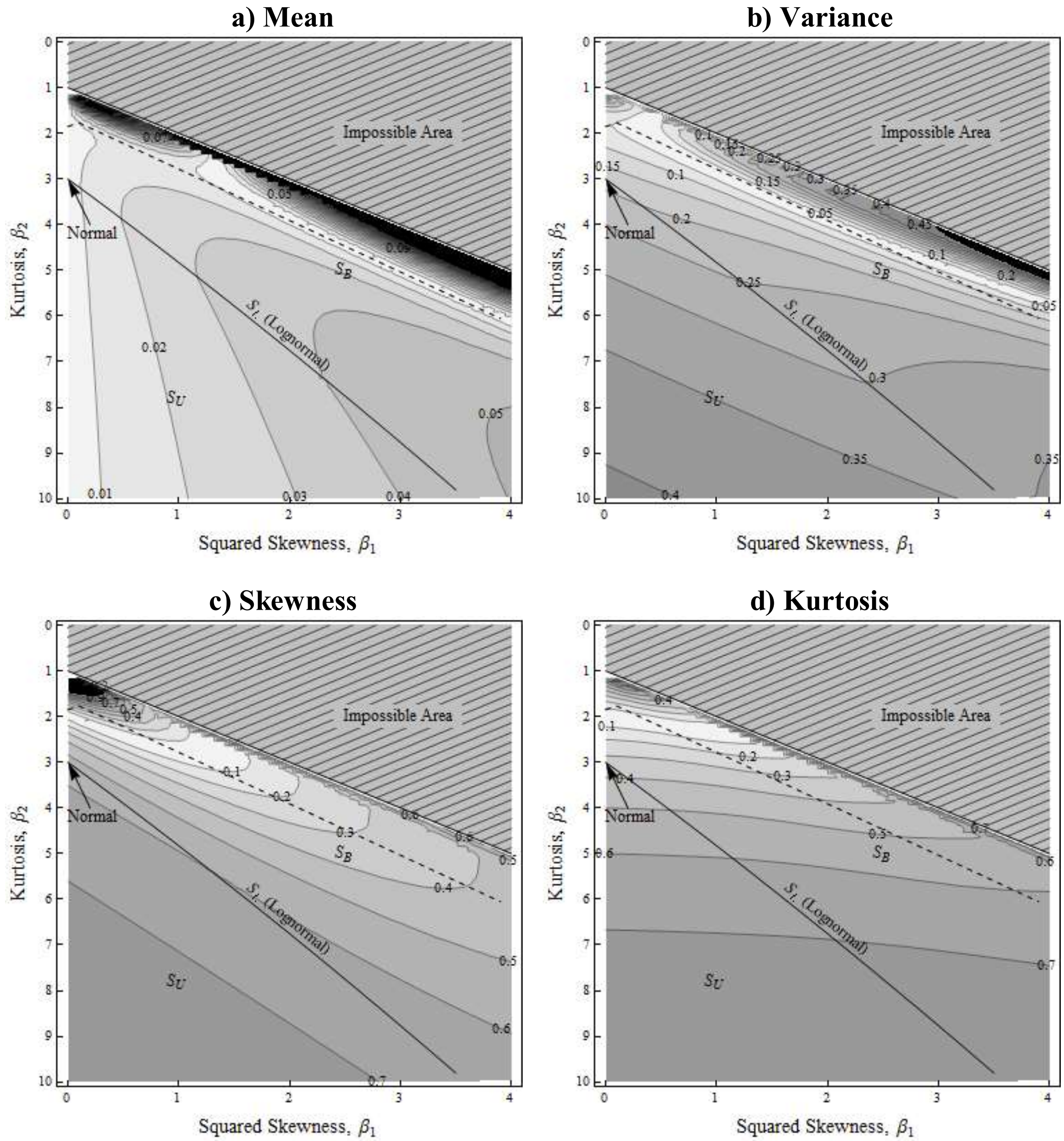

Figure 29. MCS errors in matching moments in the Johnson system.

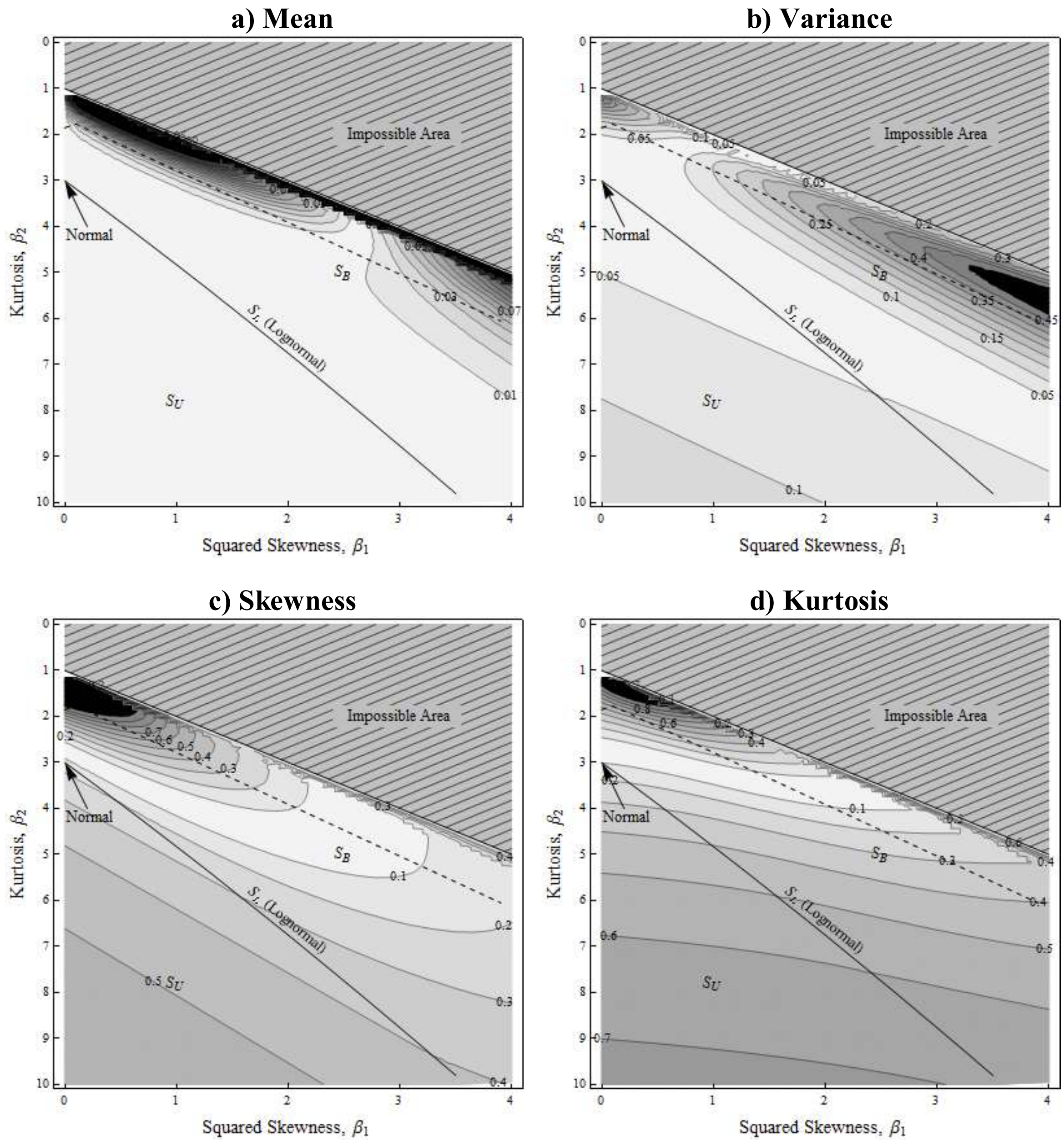

Figure 30. EPT errors in matching moments in the Johnson system.

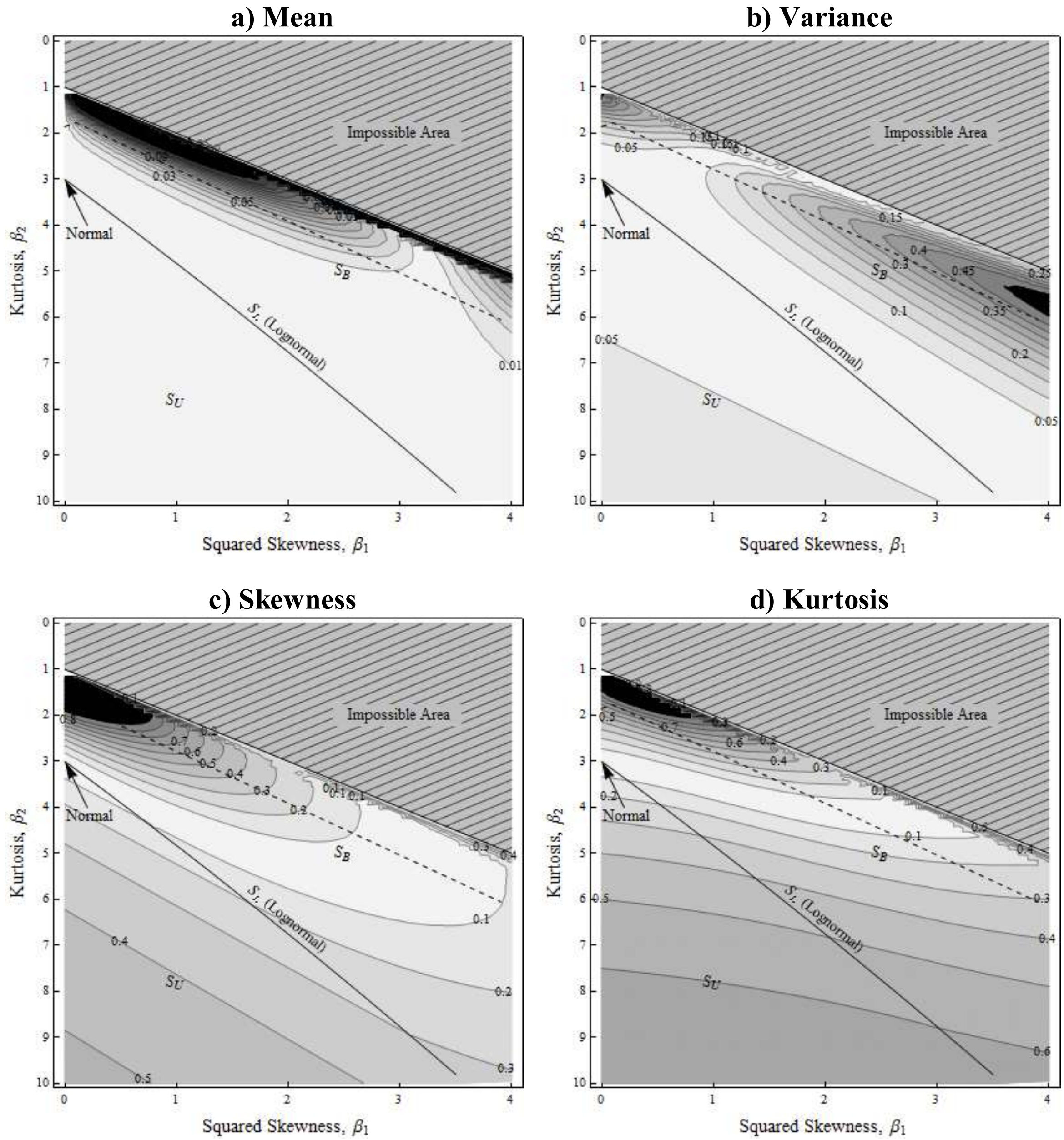

Figure 31. ZDI errors in matching moments in the Johnson system.

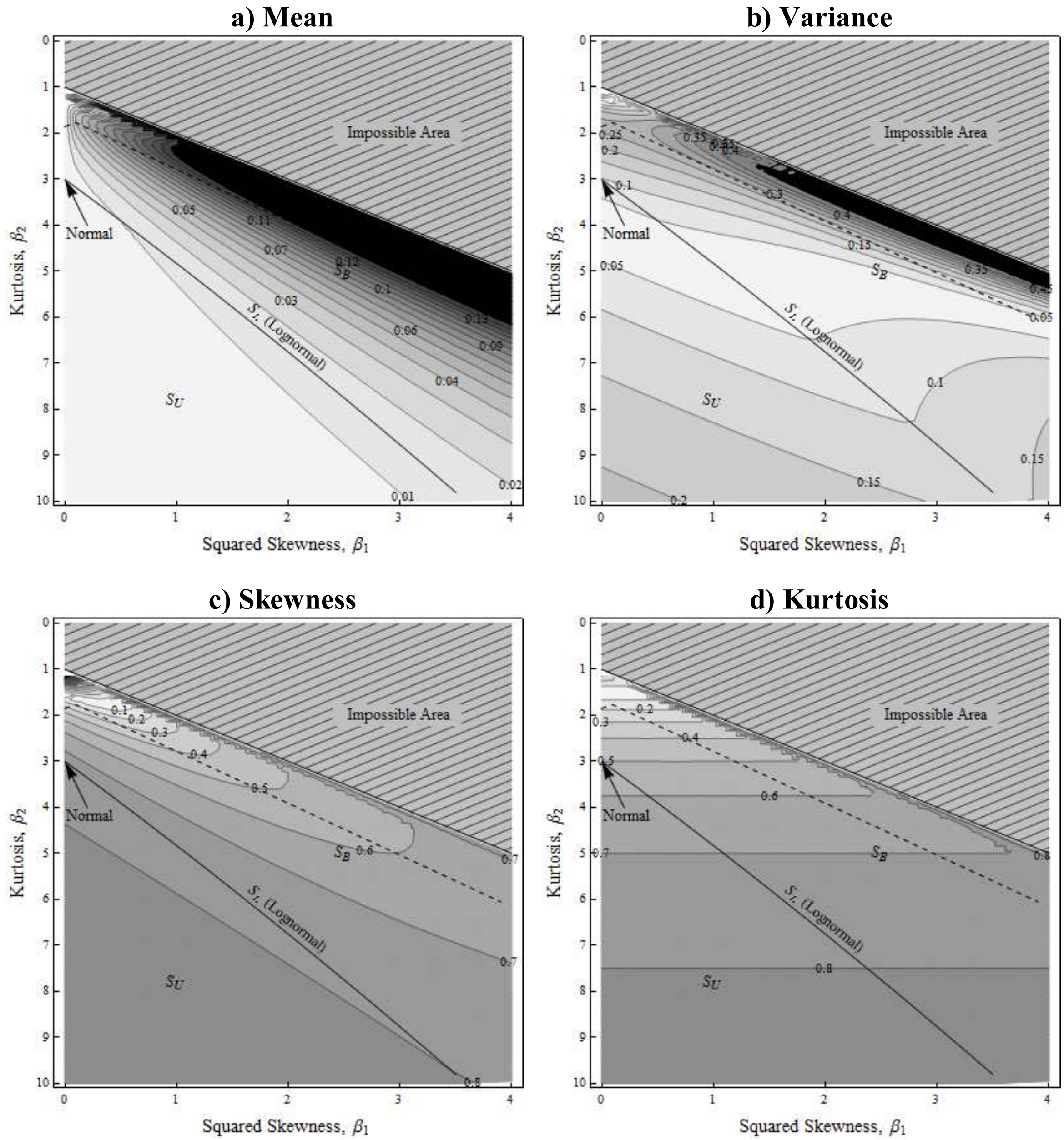

Figure 32. ZDT errors in matching moments in the Johnson system.

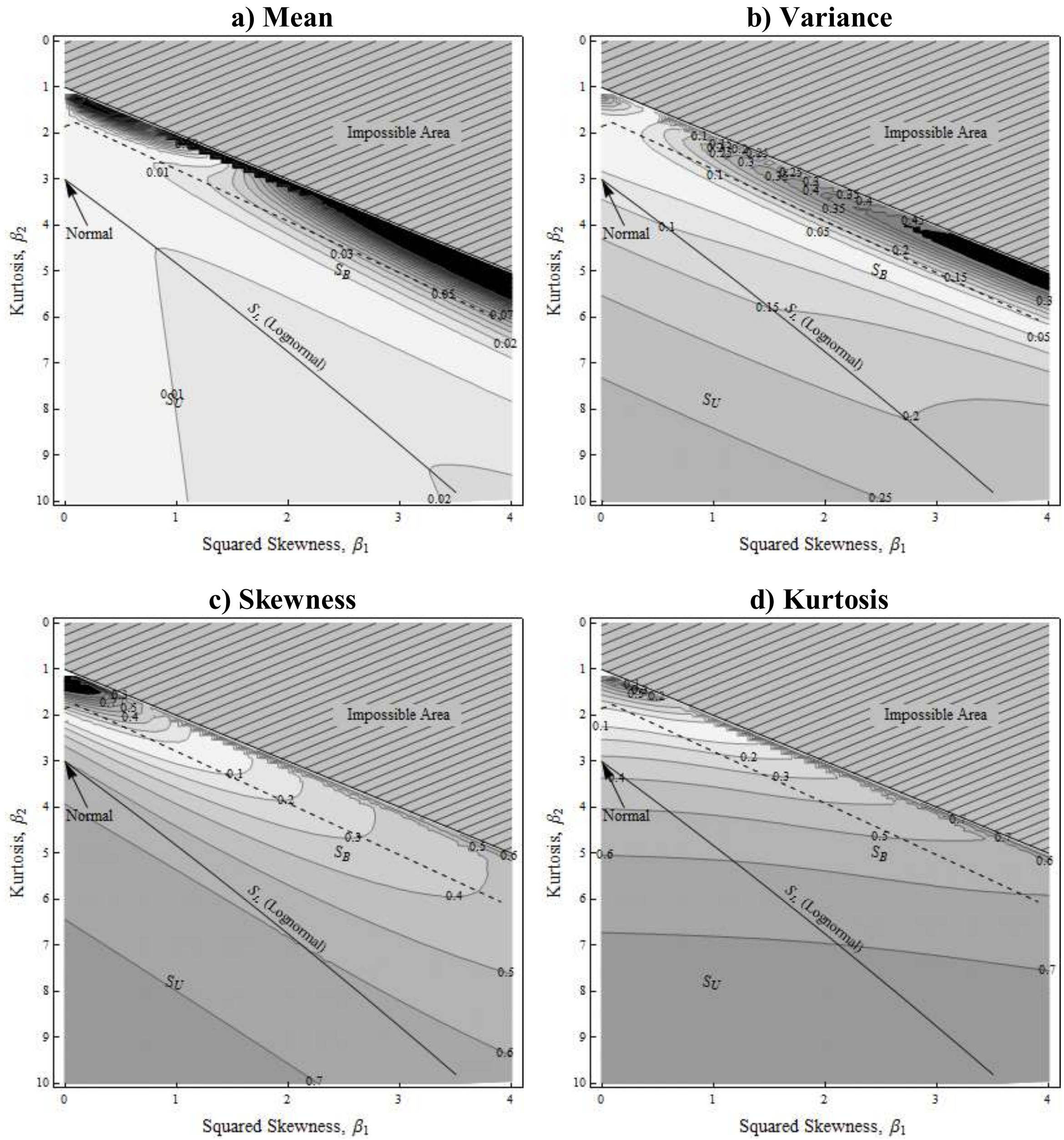

Figure 33. MRO errors in matching moments in the Johnson system.

New Shortcut Moment Errors

This section discusses EPT+, EPT++, and SP methods for the Johnson distribution types. Their AE, ASE, and ME for each type are given in Table 31 for the mean and variance, and in Table 32 for the skewness and kurtosis. Results for symmetric distributions are excluded for EPT++. The EPTS$_B$+ and EPTS$_B$++ are identical because the methods use the same percentiles and probabilities. Both tables for EPTS$_U$++ show that, as with the Pearson system EPT++ methods, EPTS$_U$+ had lower errors, despite having the more constrained regression fit. This indicates that EPTS$_U$++ is very sensitive to the probability weights.

SP distinctly outperformed ESM and MCS in the mean of each Johnson type by each measure (except for S$_B$ ME, which is lower for MCS). SP was also better than ESM and MCS by each measure in the variance for S$_L$ and S$_U$, but inferior to ESM in the S$_B$ variance. This shows that ESM, while similar to SPS$_L$, can be improved for estimating the mean and variance of lognormal distributions. This issue will be further discussed later in this chapter. Finally, MCS again had lower errors in the skewness and kurtosis than did either SP or ESM, despite typically having higher errors in the mean and variance than did those methods.

		Mean			Variance		
		AE	ASE	ME	AE	ASE	ME
S_B	$EPTS_B$+	-0.002	7.35E-05	-0.046	0.032	1.08E-02	0.381
	$EPTS_B$++*	-0.001	3.77E-05	0.029	-0.055	1.09E-02	0.247
	SPS_B	-0.004	3.49E-04	0.056	-0.151	3.19E-02	-0.282
S_L	$EPTS_L$+	0.000	8.08E-09	0.000	-0.010	1.61E-04	-0.026
	$EPTS_L$++	0.000	8.08E-09	0.000	-0.010	1.61E-04	-0.026
	SPS_L	0.000	1.84E-06	-0.002	-0.100	1.35E-02	-0.188
S_U	$EPTS_U$+	0.000	1.73E-08	0.000	-0.033	1.40E-03	-0.075
	$EPTS_U$++*	0.000	2.87E-07	0.002	-0.122	1.72E-02	-0.218
	SPS_U	0.000	1.84E-06	-0.004	-0.160	2.87E-02	-0.256

Note: S_B U-shapes are not included.
*Symmetric distributions are not included for EPT++.

Table 31. New shortcut method errors in the mean and variance for the Johnson types.

		Skewness			Kurtosis		
		AE	ASE	ME	AE	ASE	ME
S_B	$EPTS_B$+	-0.125	6.22E-02	0.980	-0.372	1.93E-01	-0.679
	$EPTS_B$++*	-0.236	1.02E-01	-1.030	-0.416	2.22E-01	-0.701
	SPS_B	-0.549	3.19E-01	-0.714	-0.643	4.33E-01	-0.815
S_L	$EPTS_L$+	-0.172	3.82E-02	-0.303	-0.411	1.98E-01	-0.626
	$EPTS_L$++	-0.172	3.82E-02	-0.303	-0.411	1.98E-01	-0.626
	SPS_L	-0.736	5.43E-01	-0.767	-0.720	5.27E-01	-0.834
S_U	$EPTS_U$+	-0.325	1.20E-01	-0.510	-0.549	3.16E-01	-0.693
	$EPTS_U$++*	-0.634	4.05E-01	-0.703	-0.743	5.56E-01	-0.817
	SPS_U	-0.757	6.06E-01	-0.851	-0.779	6.11E-01	-0.842

Note: S_B U-shapes are not included.
*Symmetric distributions are not included for EPT++.

Table 32. New shortcut method errors in the skewness and kurtosis for the Johnson types.

Figure 34 and Figure 35 show the absolute percentage errors in the moments for EPT+ and EPT++. The scalloping along the lognormal line is due to the grid of discrete samples used and that the discretization method changes at this border. The errors over the Johnson system for the SP shortcuts are shown in Figure 36. In these figures, the errors given in each region are only for the shortcut corresponding to that region.

The contours in the plots denoting different error levels display distinct discontinuities in several places, most notably in the variance for each method. These discontinuities are even more distinct than in the Pearson system, because the new Johnson system shortcuts must minimize ASE over larger areas of the system. Unless there is a specific reason to use the Johnson system, the new Pearson system shortcuts should be used, since they were fit to smaller regions of that system and hence will be more accurate.

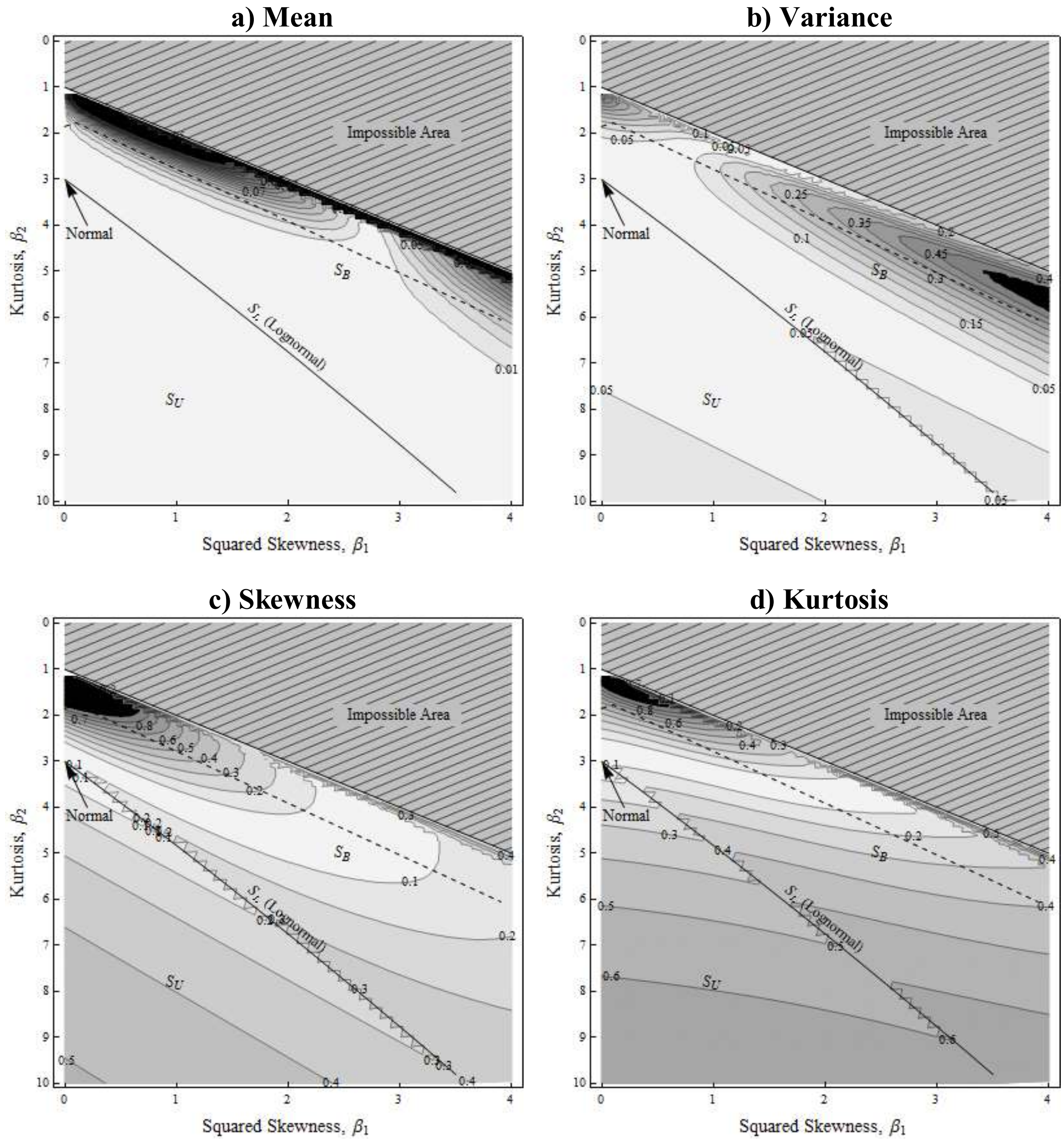

Figure 34. EPT+ errors in matching moments in the Johnson system.

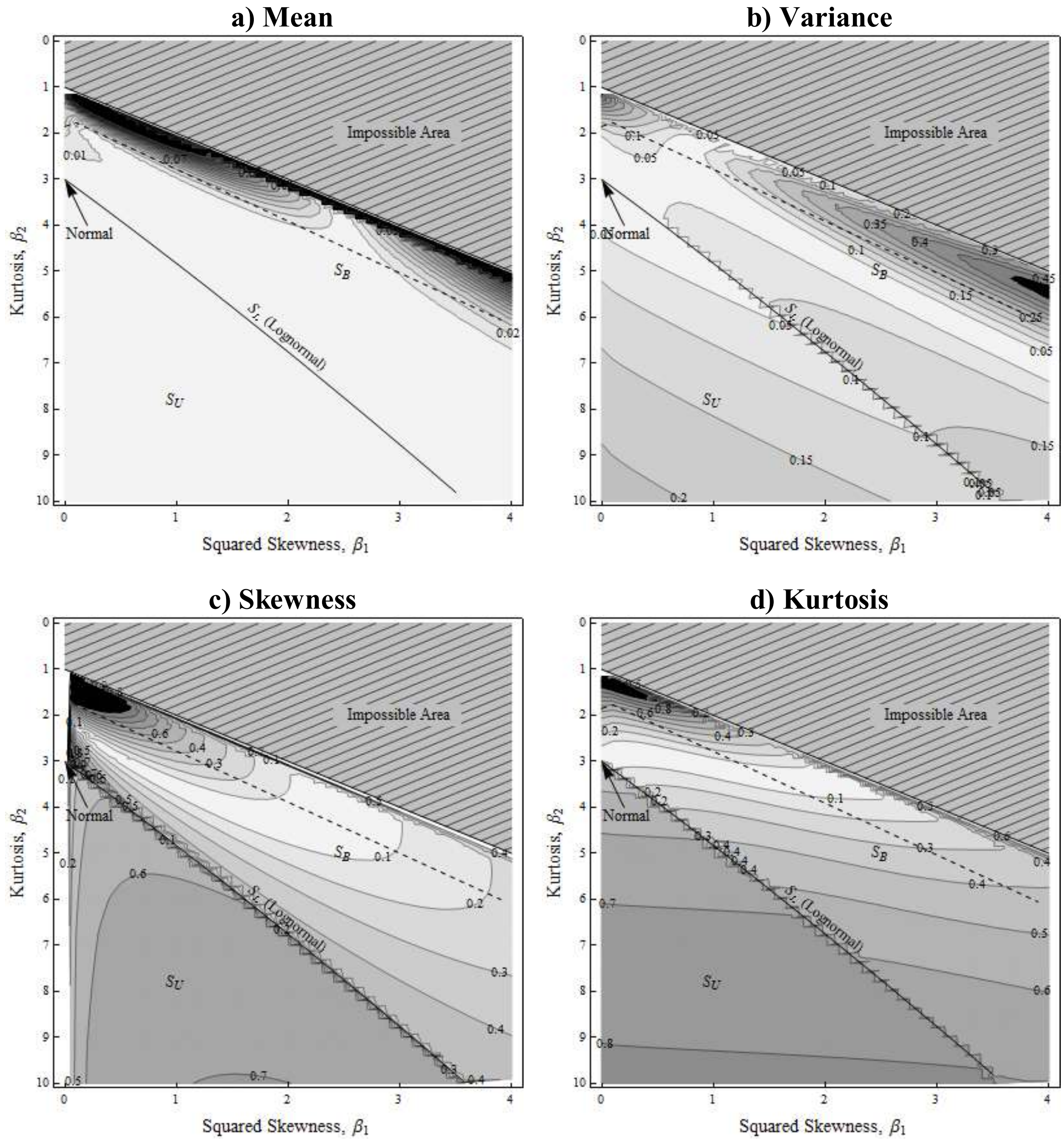

Figure 35. EPT++ errors in matching moments in the Johnson system.

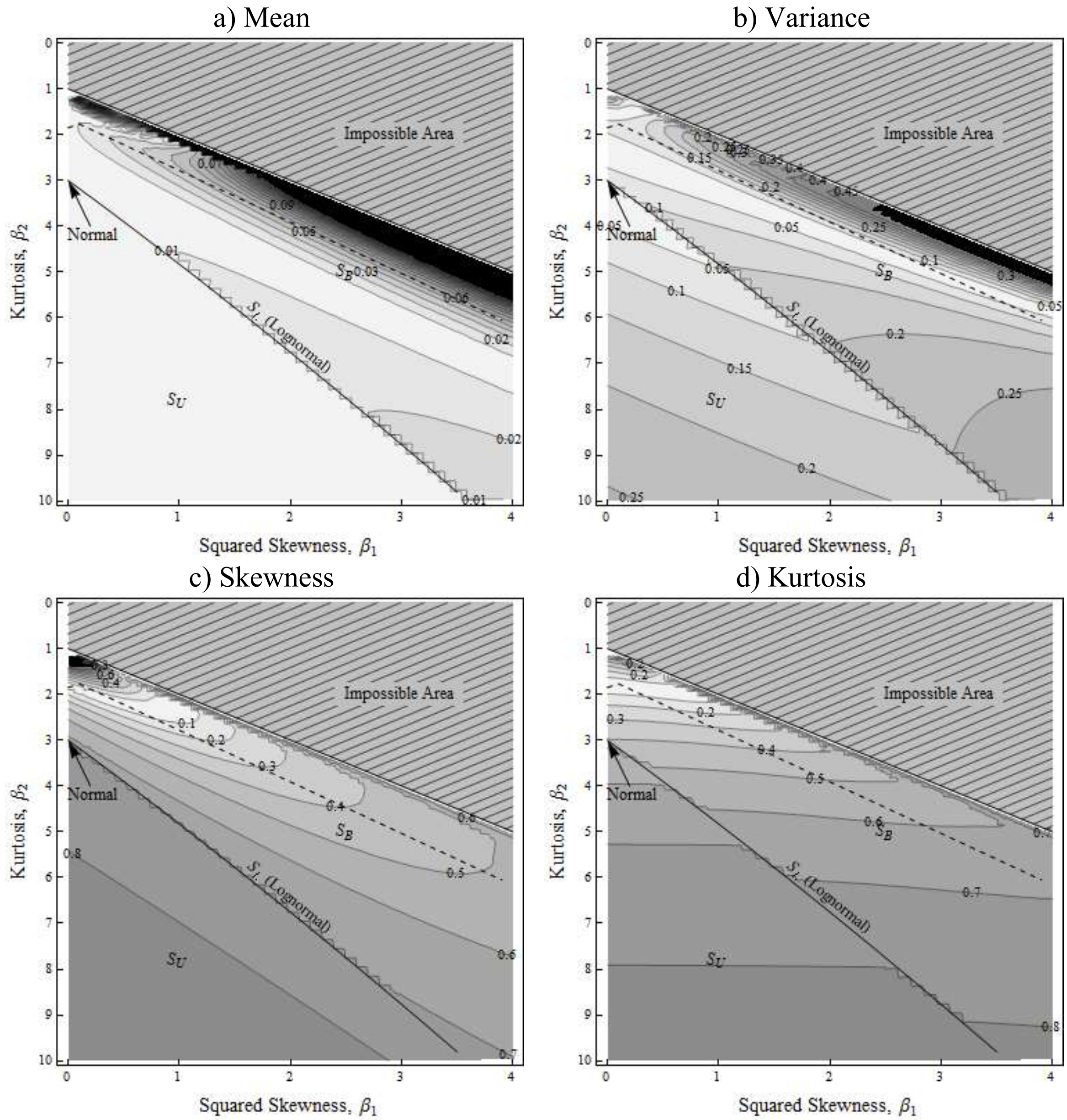

Figure 36. SPS shortcut errors in matching moments in the Johnson system.

Comparing Individual Distributions

Again, EPT+ and EPT++ can be compared to EPT for individual distributions. Figure 37 shows for which distributions each method had the lowest error in the mean and variance. EPT++ performed best in the white regions, EPT+ in the light grey, and EPT in the dark grey. Much like in Figure 22 for the Pearson system shortcuts, there was no clear dominance in the S_B region, but EPT+ appeared best in both the mean and variance over nearly the entire S_U region.

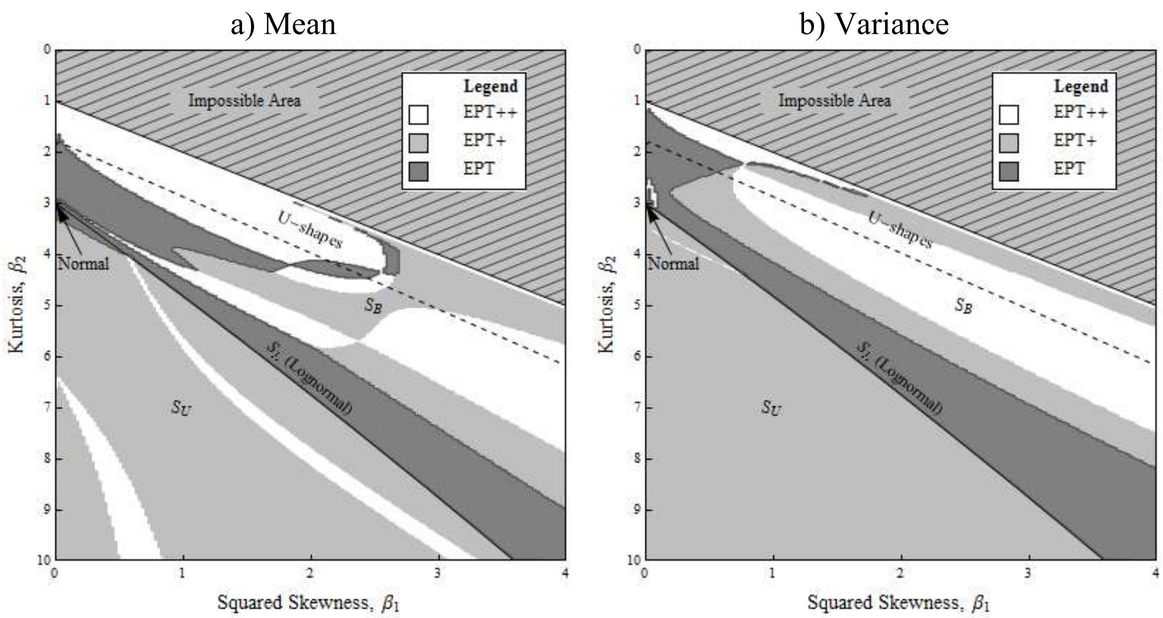

Figure 37. Distributions for which EPT++, EPT+, or EPT had the lowest absolute error in the distribution mean and variance.

Figure 38 presents the regions in which each of SP, ESM, or MCS best matched the mean. Areas colored white indicate distributions for which SP had the lowest absolute error in the mean. Light grey and dark grey indicate that ESM and MCS, respectively, had the lowest absolute error. Our new SP shortcut had the lowest error for nearly the entire S_U region, most of the S_L line, and a portion of the S_B region. Either ESM and MCS had the lowest error for other portions of the S_B region. The order of the regions where each method is best for the S_B type reflects the ordering of the weights they each place on the P10 and P90. MCS, which weights the P10 and P90 each at 0.250, performed best for distributions at the top of the feasible region. The largest region where SP performed best for this type is just below MCS's region, and this shortcut weights the P10 and P90 with 0.278. Finally, ESM was best at the bottom of the S_B region.

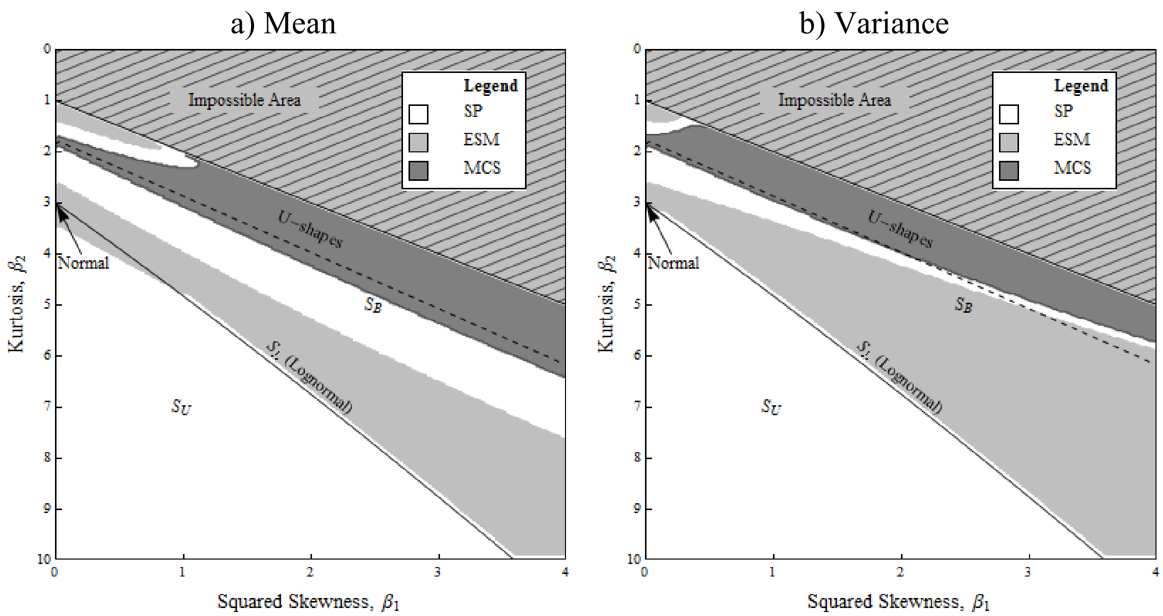

Figure 38. Distributions for which SP, ESM, or MCS has the lowest absolute error in the distribution mean and variance.

Distribution-Specific Method Moment Errors

The BMd and BMn summary statistics are shown in Table 33 for the mean and variance and in Table 34 for the skewness and kurtosis. These results, like those for the Pearson system in Table 27 and Table 28, show BMn5 to dominate the other three methods, and BMd5 to dominate BMd3. ESM had lower AE and ASE in the variance of S_L distributions than did BMn5, and both SP and ESM had lower AE in the variance for each type than did BMn5. SPS_L with AE of 0.000 in the mean for S_L and less error in the variance than BMn5, may be a better choice for discretizing lognormal distributions, even though it has fewer points and is a shortcut.

Figure 39 through Figure 42 show the absolute percentage errors in the four moments for the four distribution-specific methods. These plots are nearly identical to those for the Pearson system in Figure 24 through Figure 27.

		Mean			Variance		
		AE	ASE	ME	AE	ASE	ME
S_B	BMd3	-0.087	8.28E-03	-0.131	-0.478	2.37E-01	-0.640
	BMd5	-0.049	2.59E-03	-0.069	-0.307	1.01E-01	-0.414
	BMn3	0.000	3.80E-09	0.001	-0.250	6.42E-02	-0.305
	BMn5	0.000	2.74E-09	0.000	-0.161	2.78E-02	-0.222
S_L	BMd3	-0.069	5.08E-03	-0.093	-0.475	2.28E-01	-0.543
	BMd5	-0.043	2.05E-03	-0.060	-0.329	1.10E-01	-0.398
	BMn3	0.000	1.09E-12	0.000	-0.250	6.33E-02	-0.294
	BMn5	0.000	1.09E-12	0.000	-0.164	2.79E-02	-0.211
S_U	BMd3	0.044	2.47E-03	0.092	-0.536	2.89E-01	-0.608
	BMd5	0.029	1.03E-03	0.059	-0.391	1.55E-01	-0.468
	BMn3	0.000	2.17E-12	0.000	-0.275	7.60E-02	-0.314
	BMn5	0.000	2.37E-12	0.000	-0.195	3.86E-02	-0.240

Note: S_B U-shapes are not included.

Table 33. Distribution-specific method errors in the mean and variance for the Johnson types.

		Skewness			Kurtosis		
		AE	ASE	ME	AE	ASE	ME
S_B	BMd3	-0.728	5.42E-01	-0.842	-0.711	5.20E-01	-0.848
	BMd5	-0.507	2.76E-01	-0.689	-0.578	3.56E-01	-0.786
	BMn3	-0.476	2.43E-01	-0.642	-0.590	3.74E-01	-0.787
	BMn5	-0.446	2.12E-01	-0.610	-0.559	3.34E-01	-0.768
S_L	BMd3	-0.826	6.83E-01	-0.843	-0.740	5.55E-01	-0.847
	BMd5	-0.658	4.34E-01	-0.691	-0.654	4.39E-01	-0.785
	BMn3	-0.594	3.54E-01	-0.643	-0.644	4.28E-01	-0.786
	BMn5	-0.566	3.21E-01	-0.611	-0.633	4.12E-01	-0.767
S_U	BMd3	-0.817	7.05E-01	-0.897	-0.791	6.29E-01	-0.850
	BMd5	-0.685	4.97E-01	-0.788	-0.722	5.27E-01	-0.801
	BMn3	-0.642	4.37E-01	-0.758	-0.718	5.21E-01	-0.800
	BMn5	-0.610	3.95E-01	-0.724	-0.705	5.03E-01	-0.788

Note: S_B U-shapes are not included.

Table 34. Distribution-specific method errors in the skewness and kurtosis for the Johnson types.

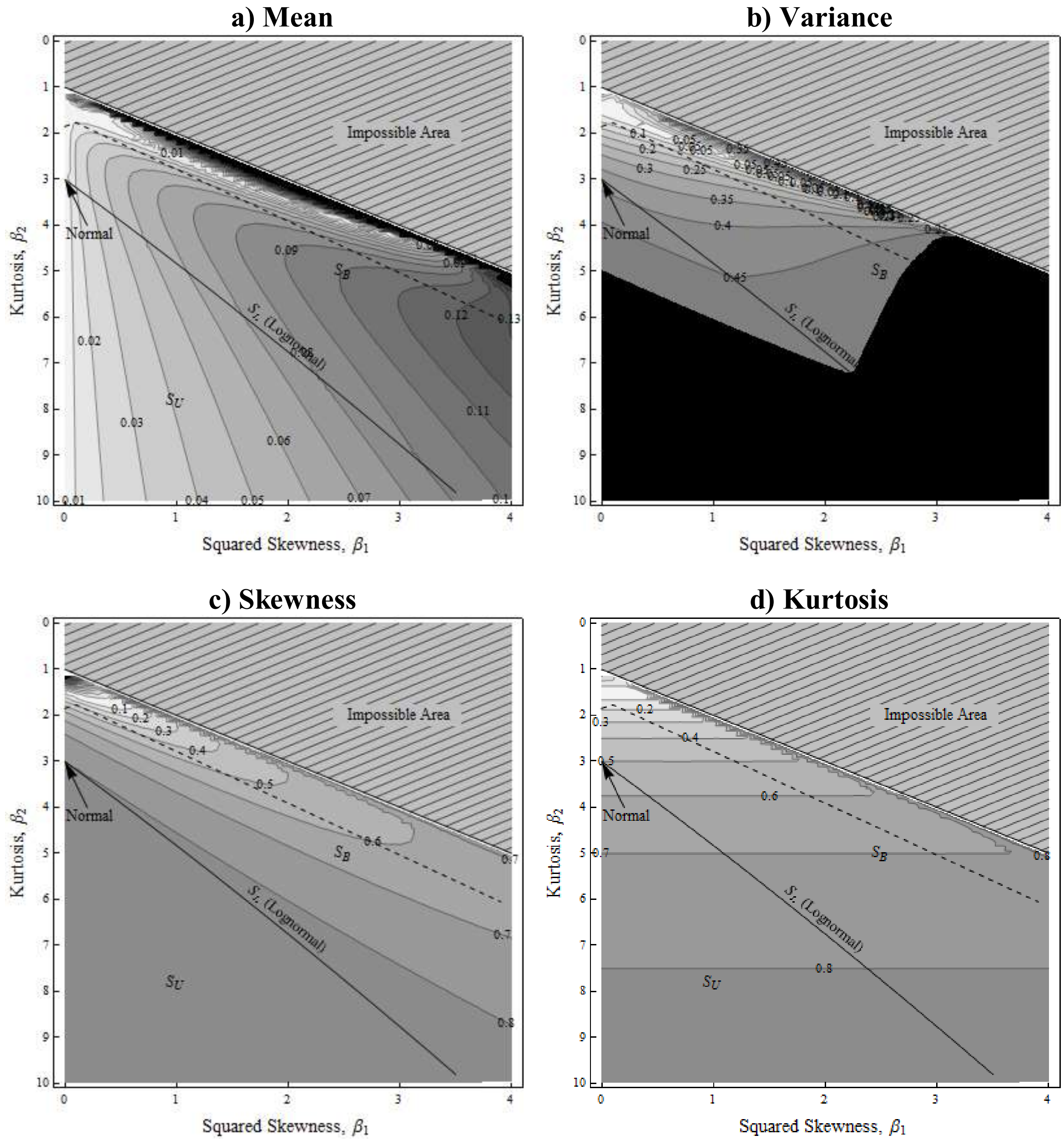

Figure 39. BMd3 errors in matching moments in the Johnson system.

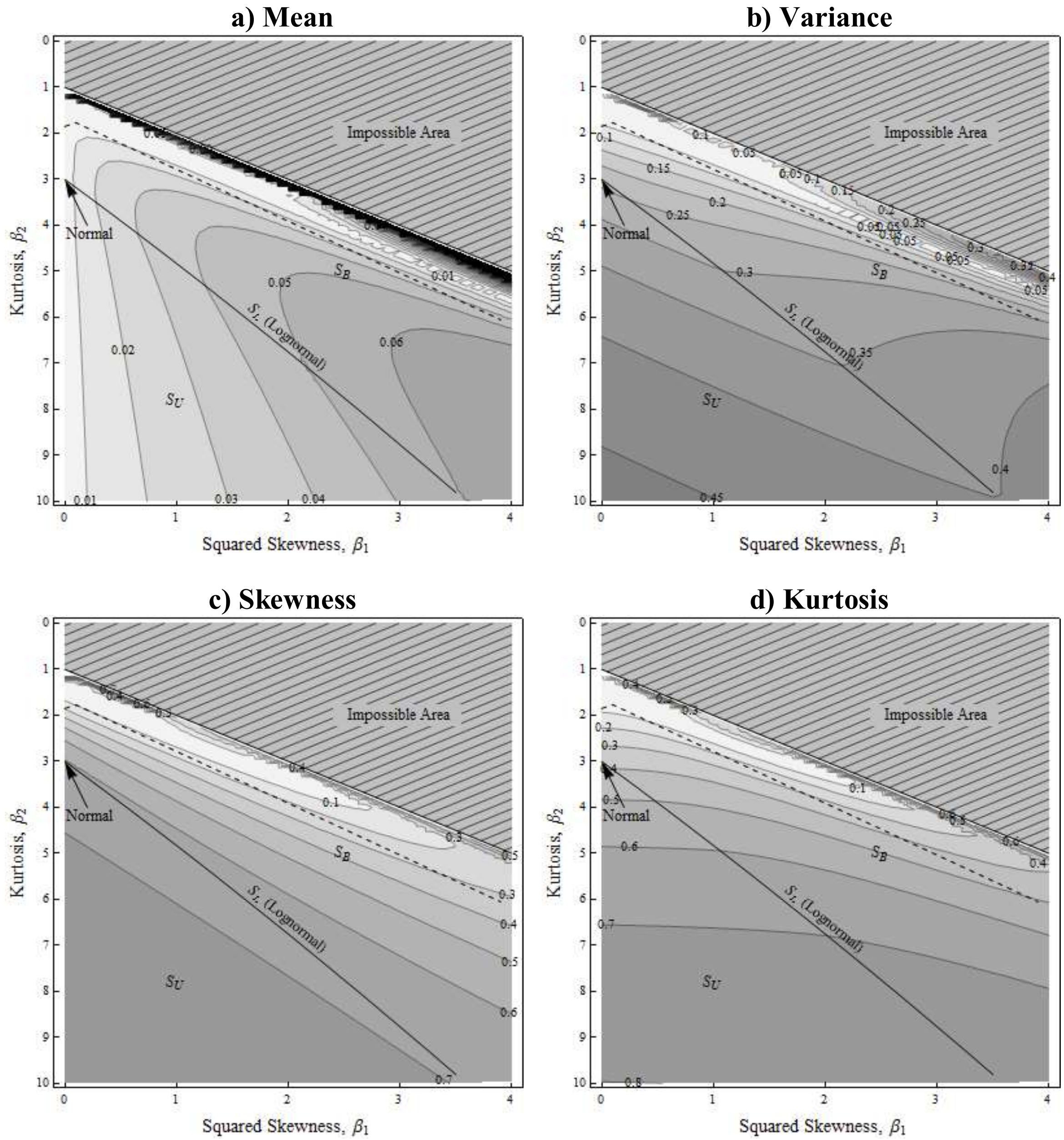

Figure 40. BMd5 errors in matching moments in the Johnson system.

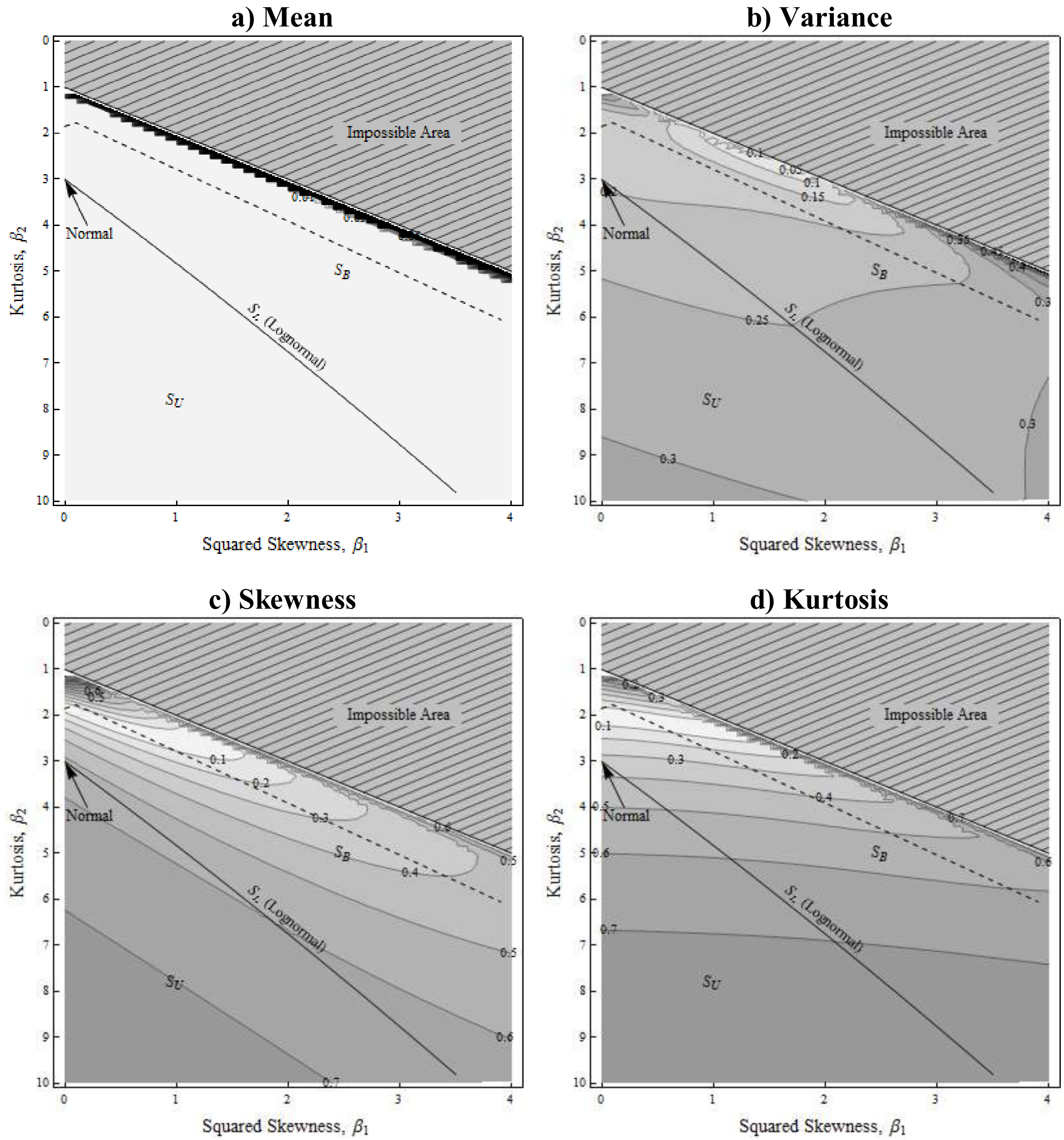

Figure 41. BMn3 errors in matching moments in the Johnson system.

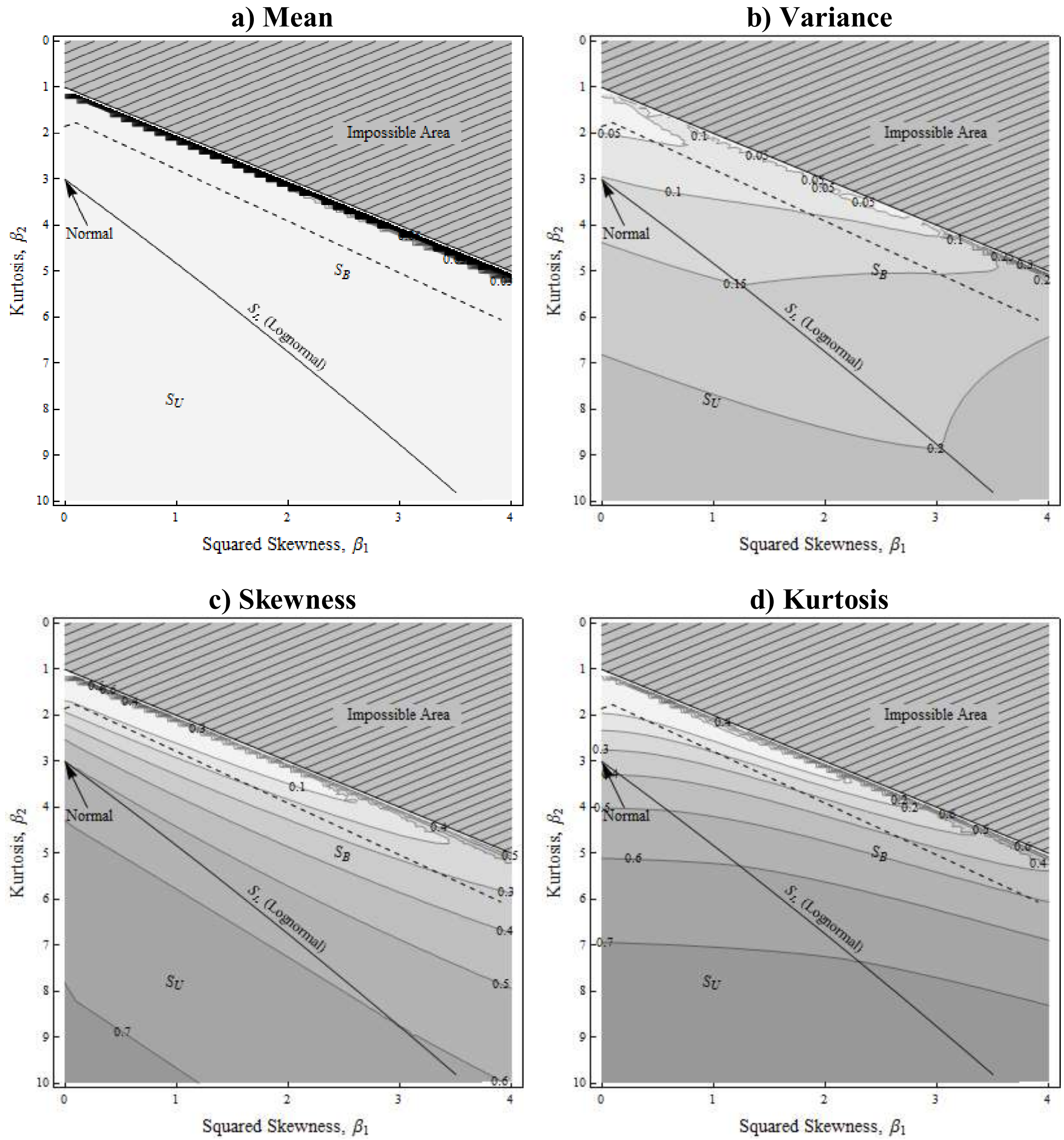

Figure 42. BMn5 errors in matching moments in the Johnson system.

U-shaped S_B Distributions

Here we briefly consider the U-shaped S_B distributions that were excluded from the previous summary statistics in this chapter. Table 35 gives the AE, ASE, and ME for the mean and variance, and Table 36 for the skewness and kurtosis, for all of the discretization methods. Of the original six shortcuts, EPT performed best in the mean, and EPT++ performed even better. Although of the distribution-specific methods, BMn5 performed best in the mean and variance, BMd5 performed best in the skewness and kurtosis and also had lower AE in the variance than BMn5. BMd5 also had the lowest AE in the variance, skewness, and kurtosis for the Pearson type I-U in Table 27 and Table 28. BMn5 appears to be the best overall method for the U-shapes.

	Mean			Variance		
	AE	ASE	ME	AE	ASE	ME
ESM	0.151	3.30E-02	0.473	0.297	1.22E-01	0.768
MCS	0.041	8.55E-03	0.323	0.145	5.08E-02	0.579
EPT	-0.025	8.27E-03	-0.295	0.248	1.17E-01	0.592
ZDI	-0.064	1.16E-02	-0.333	0.190	9.61E-02	0.551
ZDT	0.224	6.34E-02	0.572	0.385	1.83E-01	0.871
MRO	0.072	1.33E-02	0.345	0.273	1.07E-01	0.693
EPTb+	-0.033	8.57E-03	-0.301	0.232	1.08E-01	0.572
EPTb++	-0.016	7.13E-03	-0.275	0.227	9.16E-02	0.543
SPb	0.107	2.02E-02	0.413	0.239	8.97E-02	0.697
BMd3	-0.036	6.25E-03	-0.187	-0.328	1.85E-01	-0.893
BMd5	0.005	1.15E-03	0.174	-0.019	1.72E-02	0.347
BMn3	0.001	2.33E-05	0.038	-0.205	4.63E-02	-0.397
BMn5	0.001	2.25E-05	0.038	-0.102	1.27E-02	-0.208

Table 35. Errors in the mean and variance for U-shaped S_B Distributions

	Skewness			Kurtosis		
	AE	ASE	ME	AE	ASE	ME
ESM	-0.336	1.83E-01	1.234	-0.461	2.66E-01	-0.712
MCS	-0.125	1.39E-01	2.000	-0.296	1.72E-01	0.702
EPT	0.236	3.22E-01	3.288	0.081	1.86E-01	1.553
ZDI	0.364	4.64E-01	3.748	0.245	2.93E-01	1.926
ZDT	-0.459	2.59E-01	0.791	-0.537	3.31E-01	-0.754
MRO	-0.113	1.41E-01	2.034	-0.287	1.69E-01	0.721
EPTb+	0.256	3.41E-01	3.360	0.105	1.98E-01	1.609
EPTb++	0.198	2.97E-01	3.195	0.013	1.57E-01	1.459
SPb	-0.256	1.53E-01	1.524	-0.403	2.27E-01	-0.679
BMd3	-0.469	2.65E-01	0.785	-0.537	3.31E-01	-0.754
BMd5	-0.120	2.77E-02	-0.325	-0.205	7.30E-02	-0.505
BMn3	-0.181	9.36E-02	0.870	-0.305	1.71E-01	0.667
BMn5	-0.147	3.12E-02	-0.328	-0.222	7.63E-02	-0.506

Table 36. Errors in the skewness and kurtosis for U-shaped S_B Distributions

Extended Lognormal Analysis

The lognormal distribution is heavily used in several industries. Shortcut methods are not recommended for highly skewed distributions, but we need to quantify "highly skewed" to determine the applicability of a shortcut. In particular, we look at the best P10, P50, and P90 weights for individual lognormal distributions over a wider range of skewness and kurtosis than was examined in the previous section.

As before, we will require that the probability weightings p_1, p_2, and p_3 of the percentile points x_1, x_2, and x_3 be symmetric $(p_1 = p_3)$ and sum to one. We can then find the probability weighting p_1 of the P10 and P90 that matches the mean for a distribution F with mean μ,

$$p_1 = \frac{\mu - F^{-1}(0.5)}{F^{-1}(0.1) + F^{-1}(0.9) - 2F^{-1}(0.5)} \,. \tag{52}$$

Figure 43 plots p_1 as a function of the skewness for the lognormal. The dashed curve denotes the error in matching the mean for ESM. As the lognormal distributions become more skewed, more probability should be placed on the P10 and P90 to compensate for the thickening tail. The three points are equally weighted, with probability 1/3, when skewness reaches 4.29 (a P90/P50 ratio of approximately 3.0). Beyond a skewness of 4.29, the optimal discrete approximation is bimodal. We take this point as the limit on the applicability of P10-P50-P90 approximations of the lognormal distribution. At this limit, ESM, for example, underestimates the mean by 3%, the variance by 49%, and the skewness by 85%.

The requirement that $0 \leq p_1 \leq 1$ gives Equation (52) the theoretical bounds $0 < \beta_1^{1/2} < 15{,}162$. As skewness increases, the mean moves farther from the median $(F^{-1}(0.5))$, reaching the limit at which *any* symmetric weighting of the P10, P50, and P90 can match the mean when

$$\mu - F^{-1}(0.5) = F^{-1}(0.9) - F^{-1}(0.5) - \left(F^{-1}(0.5) - F^{-1}(0.1)\right). \qquad (53)$$

This limit is reached when the mean is as far from the median as the difference in the distances of the P90 from the median and the P10 from the median. For the lognormal distribution, this limit is reached when skewness is greater than 15,162 (a P90/P50 ratio of approximately 25.7). As skewness decreases, the lognormal distribution approaches the normal distribution as a limiting case, which has a skewness of zero. This occurs when both sides of Equation (53) are zero, meaning that the P10 and P90 are equidistant from the median. The resulting division of zero by zero in Equation (52) implies that any value of p_1 is feasible; i.e., any symmetric weighting will match the mean.

Since any symmetric weighting will match the mean of a symmetric distribution, this gives another degree of freedom, whereby we can attempt to match a second attribute of the distribution, such as the variance, by adjusting the amount of weight placed equally on the P10 and P90. If we choose p_1 such that the discretization matches the variance of a normal distribution, the result is almost exactly the 0.300, 0.400, 0.300 weighting of

ESM. This follows since, as shown by Bickel et al. (2011), ESM is a GQ for the normal distribution when the values are fixed to be the P10, P50, and P90.

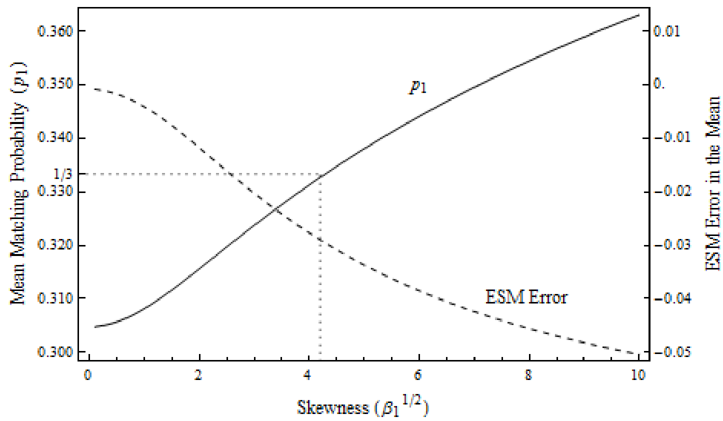

Figure 43. The P10 and P90 probabilities that match the mean of a lognormal distribution defined by skewness, and the corresponding ESM error.

Figure 43 can be used to fine-tune probability weights to a specific lognormal distribution. Accurate estimates of the moments of an assessed or empirical distribution are not often available, but can be determined from the P10, P50, and P90 assessments. For a lognormal distribution X with parameters μ and σ (equivalently, an S_L distribution translated so that its lower bound is at the origin), we have

$$\mu = \ln \text{P50} \tag{54}$$

$$\sigma = \frac{1}{\Phi^{-1}(0.9)} \ln \frac{\text{P90}}{\text{P50}}, \tag{55}$$

With these parameters, we can determine the distribution's skewness and read the appropriate weighting off Figure 43.

Best Mean-Matching Methods for Distributions in the Johnson System

The analysis of optimal lognormal weights can be extended to determine the optimal weighting of the P10, P50, and P90 across the range of distributions in the Johnson system. As in the previous section, we summarize this weighting by specifying only the weight on the P10, since the discretizations are symmetric. The results are shown in Figure 44. The contours corresponding to MCS and ESM are labeled. The ESM weighting runs diagonally from about the normal distribution point to the bottom right of Figure 44. This indicates that ESM can tolerate (in terms of matching the mean) increasing skewness as long as there is an accompanying increase in kurtosis. ESM is also near-optimal for matching the mean of S_L distributions with skewness less than one (a P90/P50 ratio less than approximately 1.5), but SP (near the 0.31 contour) is better for higher skewness. The contour corresponding to MCS lies in the upper portion of the S_B region that contains the U-shaped distributions and also runs diagonally through most of the region, but curves upward near a skewness of one. The region where an equal weighting of the three percentiles is optimal requires a kurtosis greater than 10 and is not shown in Figure 44. Over most of the regions in Figure 44, increasing kurtosis requires that more weight should be placed on the P10 and P90 in order to match the mean.

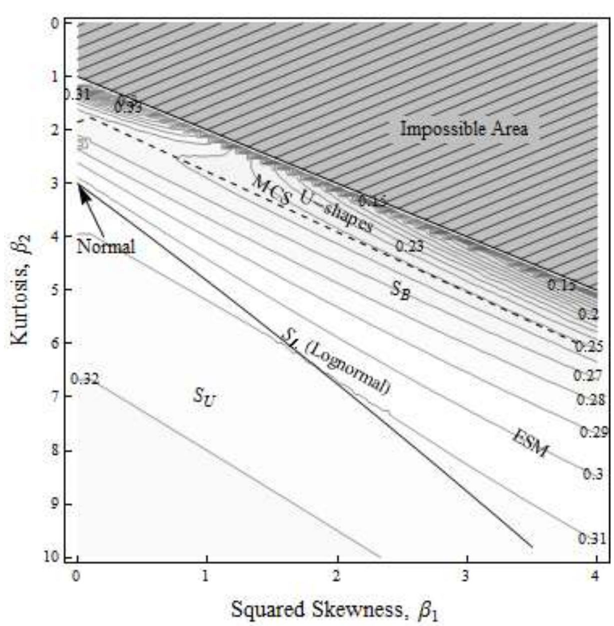

Figure 44. Mean-Matching Symmetric Weightings of the P10, P50, and P90 Percentiles.

SUMMARY AND DISCUSSION

In this chapter, we have tested the accuracy of existing discretization methods over a much wider set of distributions than in previous work. We also examined our new discretization shortcuts, which are more accurate in many cases than previously proposed methods. Our use of the Pearson and Johnson systems, which together include many common distributions as special cases, allowed us to make a thorough and systematic analysis. While considering only a finite portion of each of these infinite systems, we were able to identify performance relations to distribution shape and make broader conclusions.

The discretization methods' (both shortcuts and bracket methods) ability to preserve the mean and variance, but not skewness and kurtosis, indicates that their fidelity is often sufficient for preserving the location and scale, but not the shape or tails of a distribution. The higher moments are highly dependent on the tails, which are typically difficult to capture without using extreme percentiles, especially for unbounded distributions.

Keefer and Bodily (1983) and Keefer (1994) concluded that EPT is a good general discretization method, and our results extended this conclusion to a much wider range of distribution shapes and for different support ranges. EPT's and ZDI's similarities with EPT+ and EPT++ in both systems indicate that the former methods are near-optimal for matching moments in the class of three-point shortcuts using the P50. However, our new SP methods showed distinct improvement over both ESM and MCS in matching the mean and variance, with only somewhat degraded performance in the skewness and kurtosis. Allowing the shortcuts' weights to be asymmetric did not appreciably improve their accuracy. Further, replacing the median with the mode did not improve performance or lead to reasonable weightings. Increasing the number of points used for BMd or BMn decreased their errors, but BMd5 and BMn5 often still had higher errors than some three-point shortcuts. Increasing the number of points used by BMn can decrease the errors, but can also cause the complexity of the decision model to rapidly increase.

We saw that ESM and MCS are special cases of a general symmetric-weighting scheme and that each is appropriate for different kinds of distributions. However, on average, ESM was more accurate than MCS in matching both the mean and variance for most of the distributions in the Johnson system, in particular the lognormal. By tailoring the probability weights in either the Pearson or Johnson system with the SP methods, we can improve the performance of P10, P50, P90 discretizations in matching the mean and variance.

Our new methods' poor performance in preserving the higher moments of a distribution is a broad indication of the limits of shortcut methods, and it reinforces that they are approximations that should be refined during the analysis. Shortcuts are nonetheless valuable modeling tools with appropriate uses. Which method is "best" depends on the decision problem, but in a broad sense we can improve upon both by using basic information about the underlying uncertainty.

Guidelines for Practice

If the goal of a discretization is to match the moments of the underlying pdf, GQ should ideally be used because it can match the most moments using a given number of points of any distribution with finite moments. However, GQ can be difficult to implement and might not always be feasible in practice. Unless one is discretizing a common family (e.g., normal, uniform, triangular) for which GQs have been tabulated (Bickel et al. 2011), use of GQ requires software. In addition, GQ requires that the moments of the underlying distribution be known.

BMn matches the mean and is relatively simple to use, as discussed in Chapter 2. However, BMn can significantly underestimate the variance, and usually has more error than some shortcut methods (e.g., EPT). BMn's accuracy can be increased somewhat by using more points. One drawback, however, of both the BMn and BMd methods is that the entire distribution is required, whereas shortcut methods need only specific percentiles. That aside, BMd performs quite poorly and should be avoided if one's goal is to closely approximate a distribution's moments.

If the distribution is not known, and especially if assessments are time consuming, we are left with the shortcut methods. Of the pre-existing shortcut methods, EPT and ZDI are good general choices, but better performance can generally be obtained using the EPT+ and EPT++ shortcuts for their respective distribution types. If a distribution is thought to have low skew, then EPT+ might be better than EPT++. If one is constrained to use the common P10, P50, and P90 percentiles, then choosing one of the SP methods provides distinct improvements over ESM or MCS. By considering the bounds (or support) of a distribution, we can make an appropriate selection from these new methods. However, even our new shortcuts severely underestimate the skewness and kurtosis of most distributions, and they should be cautiously applied if the distribution shape is important.

If the skewness and kurtosis of a distribution are well-estimated, the symmetric mean-matching weighting of the P10, P50, and P90 can be determined from Figure 44. The practical applicability of shortcuts using the P10, P50, and P90 is limited to the unimodal distributions for which the mean-matching symmetric weighting places more weight on the P50 than on the P10 or the P90 (which for the lognormal is limited to skewness below 4.29; in general, the limit is a function of both skewness and kurtosis). Beyond this limit, the mean-matching weighting is no longer unimodal, and the errors in the moments are large. If the distribution falls outside this limit, other discretization methods using more extreme percentiles (e.g., EPT, ZDI, EPT+) should be used.

An analyst might not be working with a distribution that belongs to the Pearson or Johnson systems. However, Pearson and Tukey (1965) showed that distributions that share the same skewness and kurtosis are often very close in shape, even if they are not contained within the Pearson system. We also showed the similarity between Pearson and Johnson shapes in Chapter 3. With this knowledge, the appropriate EPT+, EPT++, or SP shortcut can be determined by considering the distribution's support (bounded at both ends, bounded at one end, or unbounded). This basic characteristic is determined by the nature of the uncertain quantity and should be apparent (for example, the volume of oil in a reservoir cannot be negative, and uncertainty about it should be moedeled with a distribution that is bounded below). If the distribution is bounded on both ends, then the analyst might be able to use knowledge of its shape (e.g., whether it is ∩-shaped) to further specify the appropriate type I approximation. If there is sufficient knowledge of the distribution to more narrowly specify its location in Figure 1 (if it indeed falls within this region), then Figure 22 and Figure 23 for the Pearson system, or Figure 37 and Figure 38 for the Johnson system, can be used to recommend a particular shortcut.

The Pearson system is comprised of smooth distributions, most of which are unimodal, with the type I-U as the exception. This type, and perhaps other oddly shaped or multi-modal distributions, should be discretized with care. Neither the pre-existing shortcut methods we analyze, nor the new shortcuts we present, perform well over even a quarter of the type I-U region, which is strong evidence that general shortcut methods will not accurately represent them. A method that takes the actual distribution into account, such as BMn, is better for these kinds of distributions.

As a general approach, shortcut methods are useful as a first approximation, which, when aided by sensitivity analysis, will help identify important uncertainties. These uncertainties can then be given more attention when ascertaining the full distribution and using discretization methods such as Gaussian quadrature or BMn with several points.

Decision analysis is iterative. As the analysis evolves, the discretizations that are used can and should be adapted to the importance of specific uncertainties. How a distribution is ultimately treated in a decision problem is a function not only of the distribution itself, but of its relation to other aspects of the problem. Often, analysis arbitrarily does not consider refinement of uncertainty assessments or discretizations (Bickel et al., 2011). Although refinement is not always necessary, its appropriateness should be determined from characteristics of the decision.

Chapter 5: Assessment Error

The previous chapter assumed perfect knowledge of the continuous distribution, which is often an unrealistic assumption. As discussed in Chapter 2, the assessments made by an expert might not precisely reflect their actual beliefs, in which case they would have what we term *assessment error*.

This chapter will examine the effects of incorrect percentile assessments on moment estimates, building upon the work of the previous chapter. There, we determined the best moment-preserving discretizations for different distributions, and we now consider how assessment error affects those choices.

INTRODUCTION

The best shortcut methods can preserve the mean and variance of many distributions quite well, but are quite poor in preserving higher moments, such as the skewness and kurtosis. The methods' varying degrees of accuracy, particularly for approximating the mean of a distribution, can be small for the best methods. In practice, the underlying continuous distribution may not be precisely known, raising the question of whether the differences in methods' accuracies are significant. For example, ESM and MCS use the same percentiles but different probability weights. The different weightings result in distinctly different estimates of the moments of many distributions. However, it is not clear whether these differences are significant when the continuous distribution is not completely known.

In this analysis, we consider only five discretization methods. We consider three shortcuts: ESM and MCS, for their propensity in practice, and EPT, as one of the most accurate shortcuts. The effectiveness of EPT's increased discretization accuracy is questioned, since it uses more extreme, and hence more difficult to assess, percentiles than does ESM. We also consider two distribution-specific methods: BMn5 and BMd3.

Chapter 4 found these to be the best and worst, respectively, of the four distribution-specific methods considered.

We present a model of assessment error in the percentile assessments, which we use to compare the discretization methods over the Pearson type I distribution family. Our results compare the differences between discretization moment-estimates to the assessment error of a single discretization method, to show how differences in percentile assessment error can influence the effects of using different discretization methods.

We first present the assessment error model and formally define the framework we use to investigate the effects of assessment error. The framework is illustrated with a numerical example. Next we give our analysis for selection of discretization methods we examine. We also cast the results of several empirical studies of assessor calibration into our error framework, to give context to our results. The chapter concludes with a summary and discussion of the results.

ASSESSMENT ERROR MODEL

Error in probabilistic assessments can be viewed in multiple ways. Wallsten and Budescu (1983) used an error model in which the variation was described explicitly as a difference between the value of the percentile requested and that given by the expert. Specifically, for a requested percentile $F(t) = q$, where t is the true value, the value $x = t + e$ is given, with error e. We instead model assessment error as a deviation in the percentile assignment $F(t)$ of a given value t as $F(t) = q + \delta$. Rather than assume a specific distribution on δ, we quantify the error as a range:

$$|\delta| \le \Delta . \tag{56}$$

This gives a scale-invariant error model that naturally yields larger error ranges for portions of the distribution that have low density. This corresponds to the observation that extreme percentiles are more difficult to assess, and have higher errors, than percentiles closer to the median (Alpert and Raiffa, 1982).

Suppose an expert is asked for the P10 of a distribution *h* and gives a value of *y*, which is actually the expert's P12, or a deviation of δ = 0.02 in the cumulative probability assigned to *y*. We refer to the bound of absolute deviation as Δ, so that Δ = 0.05 indicates that an assessed P10 could feasibly be any percentile from P5 to P15. We assume identical Δ for each assessed percentile. Then, given a set of percentile assessments from *h*, $(x_i^h, q_i)_{i=1,\dots,n}$ and an assessment error Δ, any distribution *f* that satisfies the following constraints on its cdf *F* is *feasible*.

$$\begin{aligned} F(x_i^h) &\geq q_i - \Delta, \quad i=1,\dots,n, \\ F(x_i^h) &\leq q_i + \Delta, \quad i=1,\dots,n. \end{aligned} \tag{57}$$

We define a *truth set* as the set of distributions satisfying these constraints, which is given by

$$T_\Delta^h = \left\{ F \mid F(x_i^h) \geq \pi_i - \Delta, F(x_i^h) \leq \pi_i + \Delta, i=1,\dots,n \right\}. \tag{58}$$

Because the distributions in the Pearson system have four parameters and four degrees of freedom, constraining on only three percentiles does not generally form a bounded truth set. However, the ∩-shaped Pearson type I is itself bounded in β_1 and β_2, as seen in Chapter 3. Despite these limitations, this distribution family is a generalization of the beta distribution, which can take on a wide range of shapes that occur in practice (Keefer and Bodily, 1983). Chapter 4 showed that the performance of the best methods did not change drastically over much of the Pearson system outside of the ∩-shaped type I. For these reasons, this chapter's numerical analysis will consider only distributions of this type.

An Illustrative Example

We use an example distribution to illustrate concepts throughout this chapter. Suppose we have P10, P50, and P90 assessments, through which a Pearson type I-∩ distribution,

h_{ex}, is fit with moments $\mu = 1, \sigma = 1, \beta_1 = 0.5, \beta_2 = 3.5$ and support [-0.997, 11.517]. Figure 45 shows the cdf (solid curve) and percentile bounds for two levels of assessment error (dashed lines) of this distribution for which the P10, P50, and P90 are assessed with error Δ = 0.02 and 0.05. Figure 46 shows the ranges of error in the percentiles. The full range of error for Δ = 0.02 is 0.04 for the P10, P50, and P90, but reaches approximately 0.08 for the P23. Similarly, the full range of error for Δ = 0.05 is 0.10 for the P10, P50, and P90, but reaches nearly 0.16 in the P33 and P50. The percentiles with the largest error for each level of Δ are not necessarily the same. It is also evident that the percentile errors in the longer right tail taper off more slowly than for the shorter left tail.

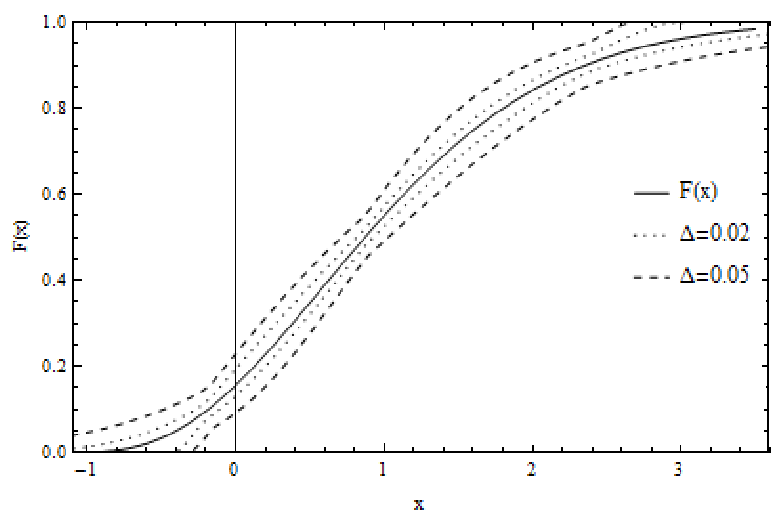

Figure 45. CDF for h_{ex} and bounds on Pearson type I-∩ distributions with assessment errors $\Delta = 0.02$ and $\Delta = 0.05$.

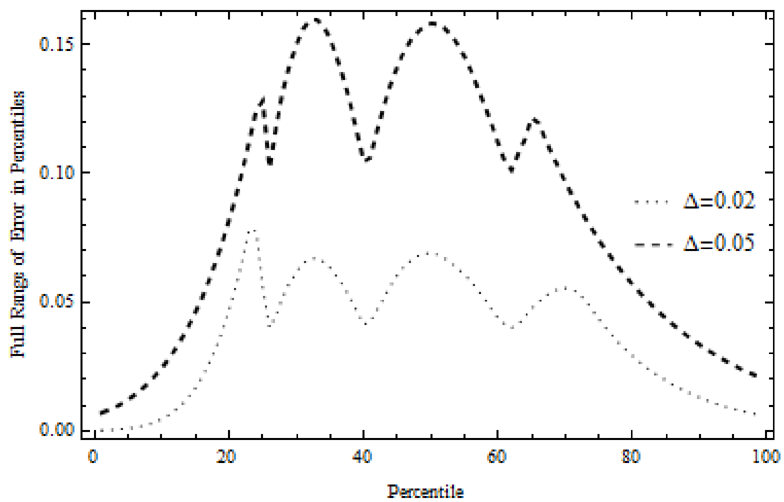

Figure 46. Error ranges for all percentiles using P10, P50, and P90 assessments of h_{ex} .

ANALYSIS METHOD

Comparison Measures

We will consider a subset $H' \subset H_P$ of Pearson type I-∩ distributions as our test set. Each distribution $h \in H'$ will be taken in turn as the distribution actually assessed, with errors in its assessed percentiles. These distributions are shown in (β_1, β_2) space in Figure 48. This plot uses $\beta_2 - \beta_1 - 1$ instead of β_2 on the vertical axis to emphasize the area of interest, following Pearson and Tukey (1965).

We compare two primary quantities of interest: the difference between using two discretization methods d_1 and d_2 on the assessed distribution $h \in H'$, and the difference between using one discretization method d_1 on the assessed distribution *h* and on another Δ-feasible distribution *f*. We define the difference between methods d_1 and d_2 for distribution *h* and moment *k* as

$$\varepsilon^k_{d_1,d_2}(h) = \left| \mu^k_{D_{d_1}(h)} - \mu^k_{D_{d_2}(h)} \right| \ . \tag{59}$$

Consider our example assessment, where h_{ex} is the assessed distribution and we wish to compare ESM and EPT. The ESM discretization of this distribution, $g_{ESM} = D_{ESM}(h_{ex})$, has mean 1.001 and variance 0.979. The EPT discretization, $g_{EPT} = D_{EPT}(h_{ex})$, has mean 1.001 and variance 1.002. The difference in these methods' means is $\varepsilon^1_{ESM,EPT} = 4.4\times10^{-4}$ (which we round to 0.000), and the difference in their variances is $\varepsilon^2_{ESM,EPT} = 0.023$.

The difference between the discretization of distributions h_{ex} and *f* for moment *k* using discretization method *d* is

$$\gamma^k_d(h, f) = \left|\mu^k_{D_d(h)} - \mu^k_{D_d(f)}\right| . \qquad (60)$$

In our example, consider a distribution $f \in T^h_{\Delta=0.05}$ with $\mu = 1.05, \sigma^2 = 1.21, \beta_1 = 0.5, \beta_2 = 3.3$. $\gamma^1_{ESM} = 0.048$ is the difference in the means, and $\gamma^2_{ESM} = 0.206$ the difference in the variances, using ESM on h_{ex} and *f*. This is effectively ESM's transformation of the assessment error in the moments between the two distributions. The actual differences in the mean and variance are 0.050 and 0.210, respectively. Similarly for EPT, $\gamma^1_{EPT} = 0.049$ and $\gamma^2_{EPT} = 0.208$.

Figure 47 illustrates these error measures in (β_1, β_2) space for an assessed distribution *h* and feasible distribution $f \in T^h_\Delta$. The black circles denote these two distributions in (β_1, β_2) space within the truth set, and the black squares denote their corresponding discrete distributions under ESM and MCS. As before, δ^h_d is the discretization error induced by discretization *d*. $\varepsilon^k_{d_1,d_2}(h)$ is the difference between the moment estimates of two discretization methods. $\gamma_d(h, f)$ is the difference between distributions *h* and *f* as represented by discretization *d* in the model. Our concern is whether the difference due to using different discretizations, $\varepsilon^k_{d_1,d_2}(h)$, is significant as compared to the difference between discretizing different feasible distributions, $\gamma_d(h, f)$.

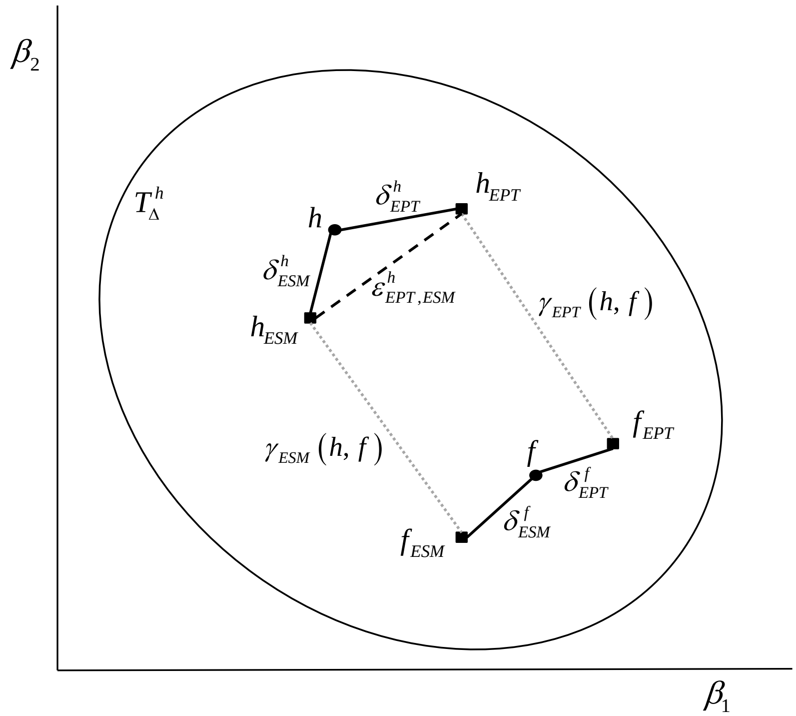

Figure 47. Assessment error measures.

The previous work on discretization accuracy was concerned only with $\varepsilon_{d_1,d_2}^k(h)$ when comparing methods. Assessment error manifests as $\gamma_d^k(h, f)$, where distribution h_{ex} is discretized, but f is the "true" distribution. To compare the effects of assessment error to the effects of using different discretizations, we consider the ratio

$$\phi_{d_1,d_2}^k(h, f) = \frac{\gamma_{d_1}^k(h, f)}{\varepsilon_{d_1,d_2}^k(h)}, \tag{61}$$

which will be abbreviated as ϕ when the context is clear. When $\phi_{d_1,d_2}^k(h, f) > 1$ $\left(\phi_{d_1,d_2}^k(h, f) < 1\right)$, the assessment error for the k^{th} moment of discretization d_1 between h and f is larger (smaller) than the difference between the estimates of the k^{th} moment estimates by d_1 and d_2 for h. In our example, $\phi_{ESM,EPT}^1(h_{ex}, f) = 107.478$, $\phi_{ESM,EPT}^2(h_{ex}, f) = 8.941$, $\phi_{EPT,ESM}^1(h_{ex}, f) = 110.787$, and $\phi_{EPT,ESM}^2(h_{ex}, f) = 9.032$. The assessment errors in the mean and variance preserved by both ESM and EPT are much larger than their differences in the moments of h_{ex}. Considering that f might be the true distribution instead of h_{ex}, this difference in distribution is much more significant than the difference in discretization error. However, this considers only individual distributions, and we are really concerned with the entire truth set.

As an aggregate measure over the truth set, define $\overline{\psi}^k_{d_1,d_2}(h,\Delta)$ as the proportion of distributions $f \in T^h_\Delta$ having $\phi^k_{d_1,d_2}(h,f) > 1$. $\overline{\psi}^k_{d_1,d_2}(h,\Delta)$ measures for how much of T^h_Δ the model assessment error γ is more significant than the difference in the discretizations ε. Let $R^h_\Delta(n) \in T^h_\Delta$ be a set of *n* distributions sampled uniformly with respect to the four-dimensional moment space M^4 of the first four standardized moments. Formally, the proportion is

$$\overline{\psi}^k_{d_1,d_2}(h,\Delta) = \lim_{n\to\infty} \frac{1}{n} \sum_{f \in R^h_\Delta(n)} I_{>1}\left(\phi^k_{d_1,d_2}(h,f)\right), \tag{62}$$

where $I_{>1}$ is the indicator function,

$$I_{>1}(x) = \begin{cases} 1 & x > 1 \\ 0 & x \le 1 \end{cases}. \tag{63}$$

Because the discretization methods can have different magnitudes of discretization error, this measure is not necessarily symmetric, i.e.,

$$\overline{\psi}^k_{d_1,d_2}(h,\Delta) \ne \overline{\psi}^k_{d_2,d_1}(h,\Delta). \tag{64}$$

$\overline{\psi}^k_{d_1,d_2}(h,\Delta)$ is the fraction of the truth set T^h_Δ for which the estimation differences between the two methods for μ^k_h are smaller than the estimation differences for discretization d_1 due to assessment error. This can be interpreted as the proportion of the truth set for which the difference of using method d_2 instead of d_1 is not significant.

Simulation Methodology

We use acceptance-rejection sampling of the truth sets, sampling 1 million uniform points in the four-dimensional moment space of the Pearson system that contains the ∩-shaped type I distributions. The set of samples for a given *h* and Δ is denoted $S^h_\Delta \in T^h_\Delta$. Although the distributions $h \in H'$ are either symmetric or right-skewed, the truth sets (and samples) may include left-skewed or right-skewed distributions. And although the $h \in H'$ are normalized with $\mu = \sigma = 1$, the truth sets include feasible distributions with $\mu \neq 1$ and $\sigma \neq 1$. We estimate $\overline{\psi}^{k}_{d_1,d_2}(h,\Delta)$ with

$$\overline{\psi}^{k}_{d_1,d_2}(h,\Delta) \approx \psi^{k}_{d_1,d_2}(h,\Delta) = \frac{1}{\left|S^{h}_{\Delta}\right|} \sum_{f \in S^{h}_{\Delta}} I_{>1}\left(\phi^{k}_{d_1,d_2}(h,f)\right). \tag{65}$$

We compare discretization methods only on their accuracies in the mean and variance, due to their generally large errors in the skewness and kurtosis. The errors in matching the first four moments over the set of distributions H' for the methods we consider are shown in Table 37. For each distribution $h \in H'$, we construct the discretization and measure the error between the moments of the discretization and the actual moments of the associated distribution. Positive error means that the discretization moment is larger than the true moment. With the exception of MCS and BMd3, the methods tend to perform well on the mean, and to a lesser extent, the variance. However, all of the methods have very wide ranges of error in the skewness and kurtosis, suggesting that methods such as these should not be used if the shape or tails of a distribution are very important to the decision model.

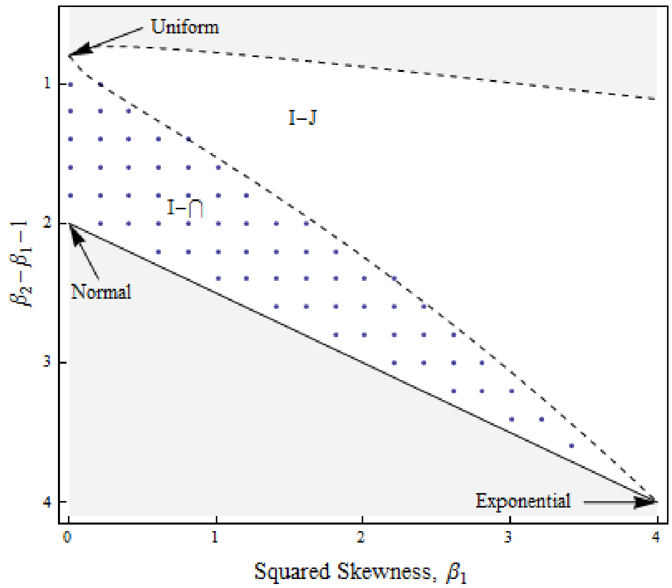

Figure 48. *H'* grid of ∩-shaped type I region of the Pearson distribution system.

a) Mean

	Avg. Err.	Max +	Max -
ESM	0.446	1.819	-0.084
MCS	-2.885	-0.000	-4.845
EPT	0.080	0.148	-0.263
BMd3	-6.584	-0.000	-10.514
BMn5	0.000	0.000	0.000

b) Variance

	Avg. Err.	Max +	Max -
ESM	-4.641	13.289	-18.536
MCS	-19.889	-5.593	-30.922
EPT	-0.208	2.077	-4.001
BMd3	-39.805	-18.365	-52.536
BMn5	-12.749	-5.000	-19.750

c) Skewness

	Avg. Err.	Max +	Max -
ESM	-60.708	-33.682	-67.688
MCS	-46.572	-9.472	-56.345
EPT	-7.141	58.137	-28.300
BMd3	-75.379	-58.538	-79.351
BMn5	-45.415	-24.179	-53.210

d) Kurtosis

	Avg. Err.	Max +	Max -
ESM	-57.852	-16.667	-78.654
MCS	-48.366	-0.000	-73.123
EPT	-25.717	35.135	-58.922
BMd3	-62.448	-25.000	-81.250
BMn5	-48.244	-10.921	-70.312

Table 37. Percentage errors in the moments without assessment error.

Simulation Analysis

We first provide a more detailed analysis of one method, ESM, on our distribution h_{ex} to demonstrate the comparison measures. We then present a summarized analysis over all $h \in H'$.

Detailed Example

This section completes our example and demonstrates the ψ metric that will be used for our primary analysis in the next section. In this section, we perform a detailed analysis of h_{ex}, shown in Figure 45, as the assessed distribution. We compare ESM to each of EPT, MCS, BMd3, and BMn5. Figure 49 shows the distributions of ϕ for the truth set samples $S_{h_{ex}}^{\Delta}$, comparing ESM to the other four methods for $\Delta = 1\%, 3\%, 5\%$. The vertical dashed line in each plot indicates where $\phi = 1$. The fractions of these samples with $\phi > 1$ are given in Table 38, which includes Δ = 2% and 4%. Figure 49a shows that $\phi_{ESM,EPT}^{1,\Delta} > 1$ for the vast majority of samples, even for Δ = 1%, and is quite large for many of these samples; over 400 for $\Delta = 5\%$. This indicates that the difference between ESM's and EPT's estimates of the mean of the assessed distribution is extremely small compared to the differences in ESM's mean estimates for different distributions in the truth set. For this distribution, ESM and EPT do not have an appreciable difference in the mean estimate for imperfect assessments. However, the variance estimates have $\phi_{ESM,EPT}^{2,\Delta} > 1$ for about 61.69% of the samples for Δ = 1%, indicating that the difference in variance estimates between ESM and EPT might not be negligible for small assessment errors. At $\Delta = 3\%$, over 85% of the samples have $\phi_{ESM,EPT}^{2,\Delta} > 1$, greatly diminishing the methods' relative differences.

Comparing ESM to MCS in Figure 49b and Table 38b indicates marked differences in the methods' mean and variance estimates, particularly under low assessment error. The last chapter conveyed significant differences in ESM and MCS estimates of means and variances, and we see here that for $\Delta = 5\%$, errors in the mean estimates are relatively significant for 20.42% of the truth set, and for the variance, 46.11% of the truth set. We also saw significant differences between the errors of ESM and BMd3, which, as seen in Figure 49c and Table 38c, are relatively significant even up to $\Delta = 5\%$, with $\psi^{1}_{ESM,BMd3}(h,5\%) = 56.47\%$ and $\psi^{2}_{ESM,BMd3}(h,5\%) = 12.10\%$. Comparing BMn5 in Figure 49d and Table 38d indicates that assessment error overwhelms discretization differences in the mean, similar to the comparison with EPT. BMn5 perfectly matches the means, and because ESM has very little error in the means of these distributions, as seen in Table 37, the methods' differences are not significant. Unlike EPT, BMn5 provides poor estimates of the variances, and we see that $\psi^{2}_{ESM,BMn5}(h,1\%) = 0.00\%$ for $\Delta = 1\%$, and at $\Delta = 3\%$ is only just over half (52.23%).

a) ESM compared to EPT

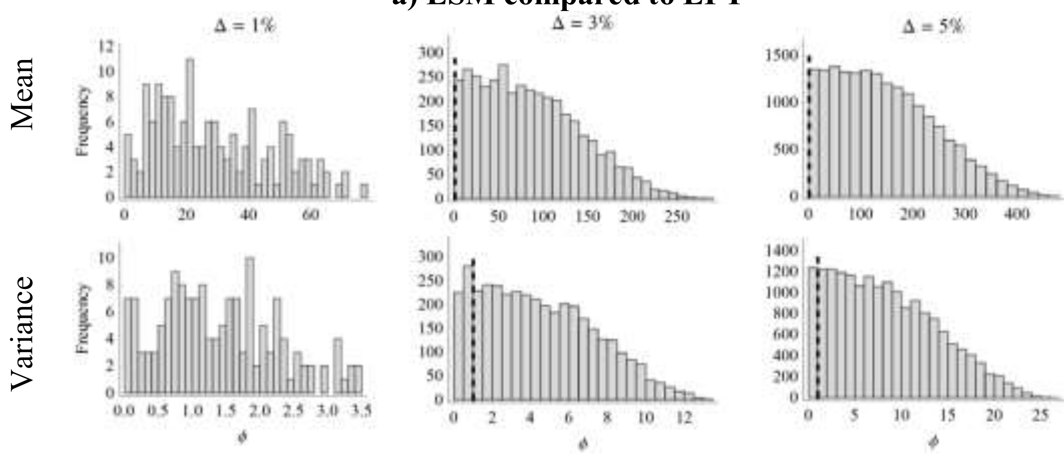

Figure 49. Distributions of ϕ, comparing ESM to EPT, MCS, Bmd3, and BMn5.

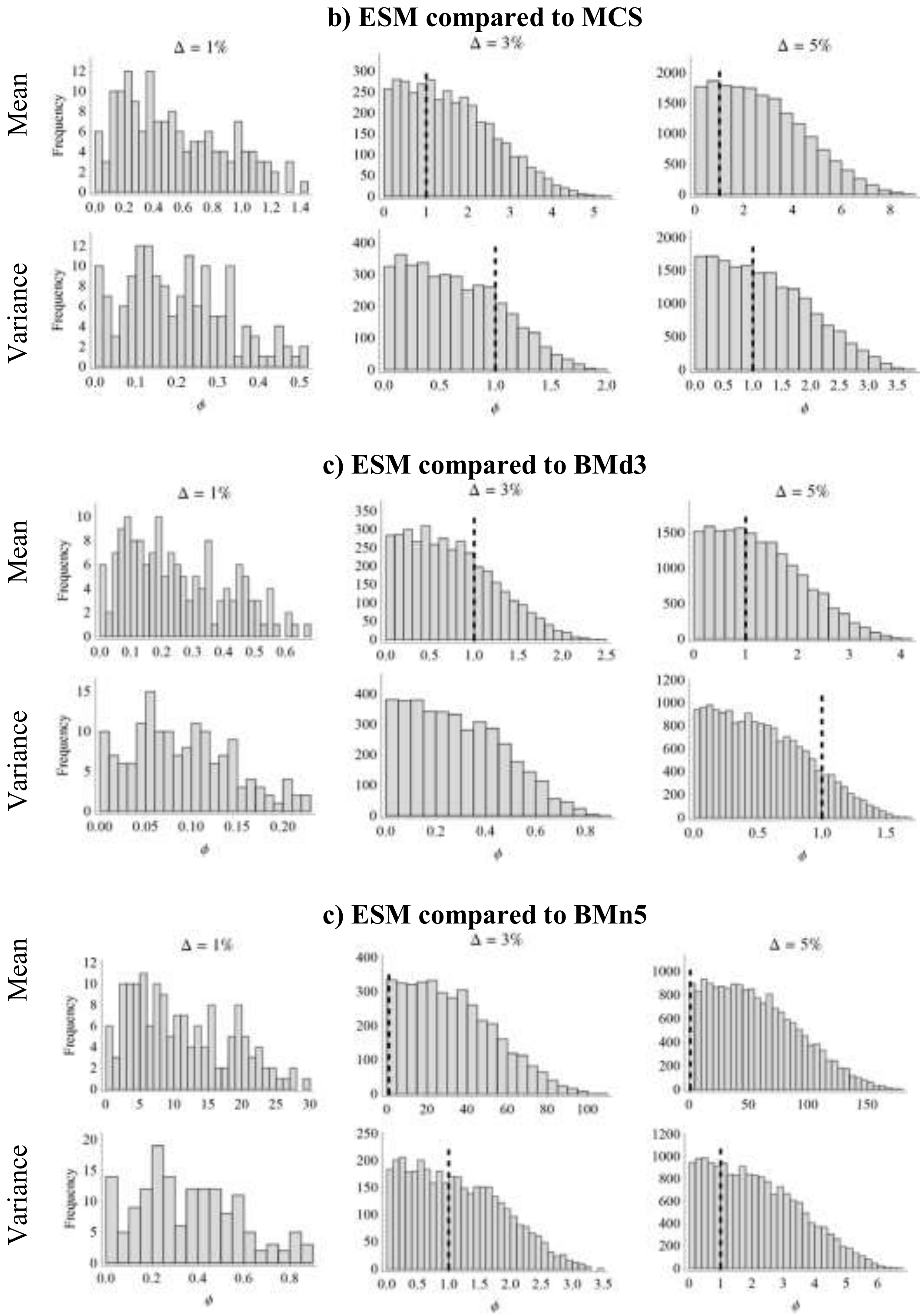

Figure 49. cont.

a) ESM compared to EPT.

	ψ				
Δ	1%	2%	3%	4%	5%
Mean	96.75	99.01	99.37	99.55	99.64
Variance	61.69	80.16	86.81	90.93	93.06

b) ESM compared to MCS.

	ψ				
Δ	1%	2%	3%	4%	5%
Mean	12.99	47.43	65.23	74.20	79.58
Variance	0.00	0.81	21.16	41.02	53.89

c) ESM compared to BMd3.

	ψ				
Δ	1%	2%	3%	4%	5%
Mean	0.00	7.39	28.82	45.09	56.47
Variance	0.00	0.00	0.00	1.50	12.10

d) ESM compared to BMn5.

	ψ				
Δ	1%	2%	3%	4%	5%
Mean	96.10	97.39	98.33	98.68	98.96
Variance	0.00	28.49	52.23	65.69	73.23

Table 38. Comparing ESM to EPT, MCS, BMd3, and BMn5 using ψ, the fraction of samples with $\phi > 1$.

This example of a single assessed distribution, with varying levels of assessment error, leads to several conclusions. First and foremost, even differences between methods apparent under perfect assessments can become insignificant with only a small degree of assessment error (e.g., ESM’s and EPT’s performance in the variance). Second, methods that do not perform similarly can be significantly different even at higher levels of assessment error (ESM compared to MCS and BMd3). Third, high similarity between methods in estimating one moment does not necessitate high similarity in estimating other moments (ESM's and BMn5's similarity in the mean but larger difference in the variance).

Analysis of the Pearson Type I Distributions

Now we broaden the analysis by considering the distribution of ψ over the set *H'* for each pairwise combination of methods. Table 39 gives the average, minimum, and maximum percentage assessment errors in the mean and variance for S_{Δ}^{h}, each averaged over all $h \in H$. The average assessment errors in the mean and variance display positive biases, due to the $h \in H$ being right-skewed.

	Δ	1%	2%	3%	4%	5%
Mean	Avg.	0.121	0.226	0.353	0.521	0.758
	Min	-3.511	-7.522	-11.626	-15.866	-19.984
	Max	3.899	8.275	12.801	17.243	21.855

	Δ	1%	2%	3%	4%	5%
Variance	Avg.	0.672	0.960	1.454	2.068	2.821
	Min	-7.137	-11.550	-15.354	-19.199	-23.084
	Max	7.751	14.494	21.653	29.062	36.403

Table 39. Average, minimum, and maximum percentage assessment error in S_{Δ}^{h} averaged over all $h \in H$.

We first directly extend the analysis of our example by comparing ESM to each of EPT, MCS, BMd3, and BMn5 over all of the distributions in *H'*. Figure 50 shows histograms of ψ for the mean and variance for each comparison at three levels of assessment error. Figure 50a shows that $\psi_{ESM,\mathrm{EPT}}^{1}(h,1\%) > 50\%$ for almost all $h \in H'$ and that $\psi_{ESM,\mathrm{EPT}}^{1}(h,5\%) > 80\%$ for all $h \in H'$. Comparing these results to those for the single $h \in H'$ in the example, it is clear that ESM and EPT are practically indistinguishable for the mean estimate with low assessment error for most of *H'*, though not always. The variance estimates are still significantly different for low assessment error, as evidenced by $\psi^2_{ESM,\mathrm{EPT}}(h,1\%)$ being close to zero for most of *H'*, but that $\psi^2_{ESM,\mathrm{EPT}}(h,5\%) > 60\%$ for the majority of distributions indicates that assessment error tends to dominate.

These plots show only a one-sided comparison between the difference in discretization errors between ESM and the other methods and the assessment error preserved by ESM. When the difference in discretization accuracy is significant, the more accurate method can be determined by comparing $\psi^1_{d_1,d_2}(h,\Delta)$ to $\psi^1_{d_2,d_1}(h,\Delta)$. Comparing Figure 50a to Figure 51a, that the distributions of $\psi^1_{ESM,\mathrm{EPT}}(h,\Delta)$ and $\psi^1_{EPT,ESM}(h,\Delta)$ are almost identical, but $\psi^2_{ESM,\mathrm{EPT}}(h,\Delta)$ and $\psi^2_{EPT,ESM}(h,\Delta)$ show distinct differences, particularly for $\Delta = 1\%$. Both methods have smaller ranges of discretization errors in the mean (Table 37a) than the ranges of assessment errors over S^h_Δ (Table 39) for all $\Delta \le 5\%$. However, ESM's range of discretization error in the variance (Table 37b) is larger than the average ranges of assessment error in the variance (Table 39) for $\Delta = 1,2\%$, whereas EPT's range of discretization error is smaller than the average range of assessment error for all $\Delta \le 5\%$. That ESM is less accurate in matching the variance than is EPT means that the difference between ESM's and EPT's discretization accuracy, $\varepsilon^2_{ESM,EPT}(h)$, is more significant to ESM's model assessment error than to EPT's.

Comparing ESM to MCS in Figure 50b, the results for all of the distributions in *H'* tend to coincide with those seen in Figure 49. For low assessment error, the estimate differences are significant, as $\psi^1_{ESM,\mathrm{MCS}}(h,1\%)$ and $\psi^2_{ESM,\mathrm{MCS}}(h,1\%)$ are close to zero for almost all *h*. At $\Delta = 5\%$ $\Delta = 5\%$, $\psi^1_{ESM,\mathrm{MCS}}(h,5\%)$ and $\psi^2_{ESM,\mathrm{MCS}}(h,5\%)$ are tightly clustered around 70% and 50%, respectively. The comparison with BMd3 in Figure 50c is starker. Except for the few symmetric distributions in *H*, for which $\psi^1_{ESM,\mathrm{BMd3}}(h,5\%)$ is close to 100%, $\psi^1_{ESM,\mathrm{BMd3}}(h,5\%)$ is less than 80%, with most $h \in H'$ between 20% and 50%. All of the distributions in *H* have $\psi^2_{ESM,\mathrm{BMd3}}(h,5\%) < 30\%$.

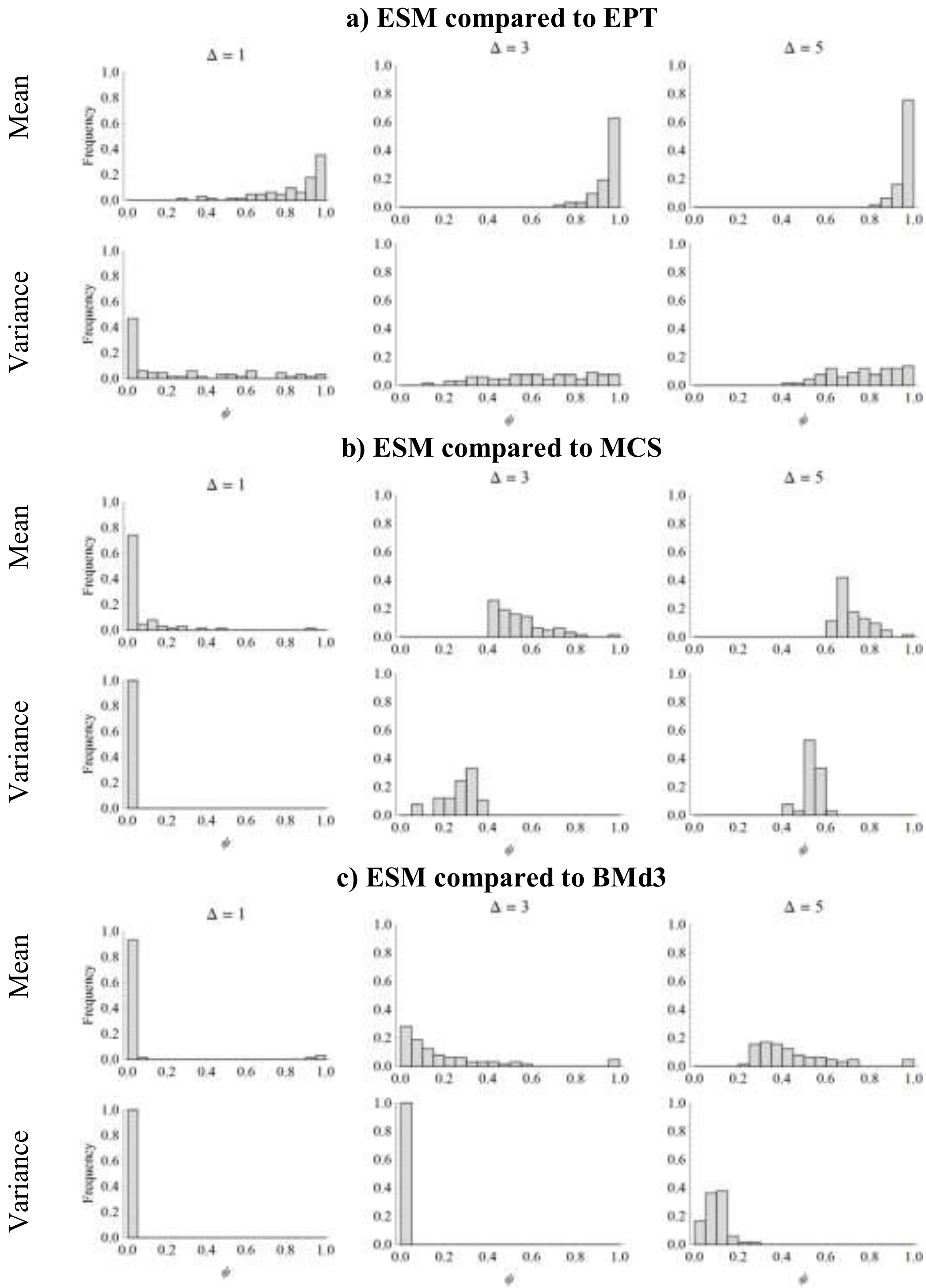

Figure 50. ESM compared to EPT, MCS, BMd3, and BMn5.

d) ESM compared to BMn5

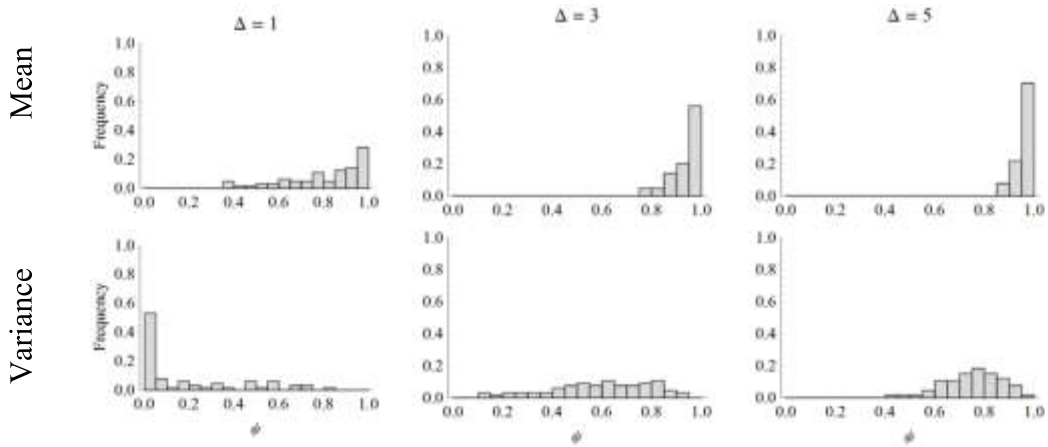

Figure 50. cont.

Figure 51 compares EPT against ESM, MCS, BMd3, and BMn5. As mentioned previously, EPT's greater accuracy than ESM's in matching the variance is evident in comparing Figure 50a to Figure 51a, with $\psi^2_{EPT,ESM}(h,\Delta)$ shifted more to the right than is $\psi^2_{ESM,\mathrm{EPT}}(h,\Delta)$. Figure 51b and c, which compare EPT to MCS and BMd3, respectively, show results that are very similar to the comparison of ESM to MCS and BMd3, which should not be surprising considering ESM's and EPT's similarity. That EPT is more accurate than MCS or BMd3 in both the mean and variance is again apparent in the generally higher $\psi^k_{EPT,d_2}(h,\Delta)$. Figure 51c shows EPT and BMn5 to be virtually indistinguishable in the mean, with $\psi^1_{EPT,BMn5}(h,1\%) > 90\% \;\; \forall h \in H'$. The difference in variance estimates is significant at small Δ, decreasing as Δ increases, but $\psi^2_{EPT,\mathrm{BMn5}}(h,5\%) < 80\%$ for most $h \in H'$. That EPT is more accurate than BMn5 in the variance is again evident in $\psi^2_{EPT,BMn5}(h,\Delta)$ being shifted more to the right than is $\psi^2_{BMn5,EPT}(h,\Delta)$.

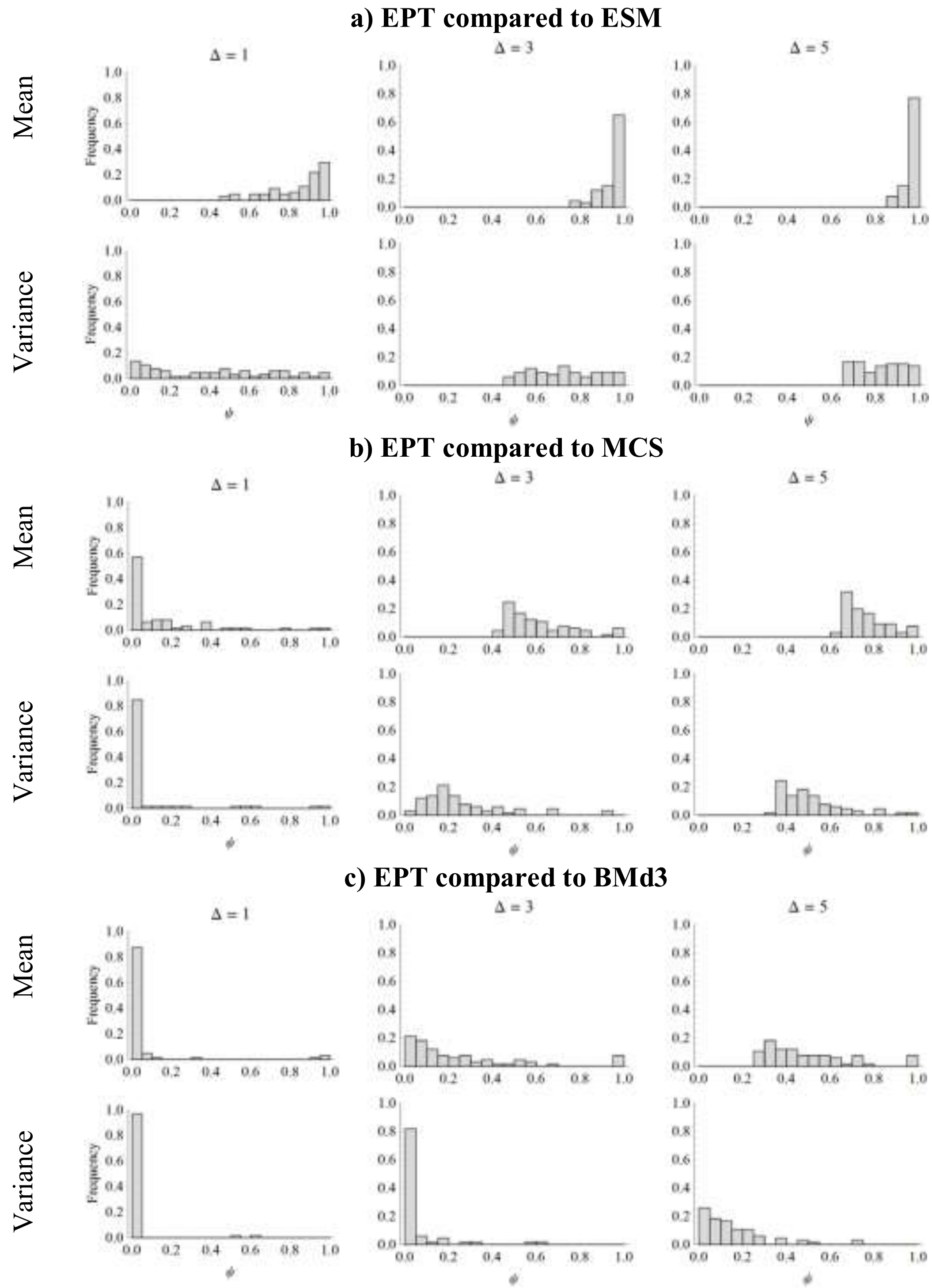

Figure 51. EPT compared to ESM, MCS, BMd3, and BMn5.

d) EPT compared to BMn5

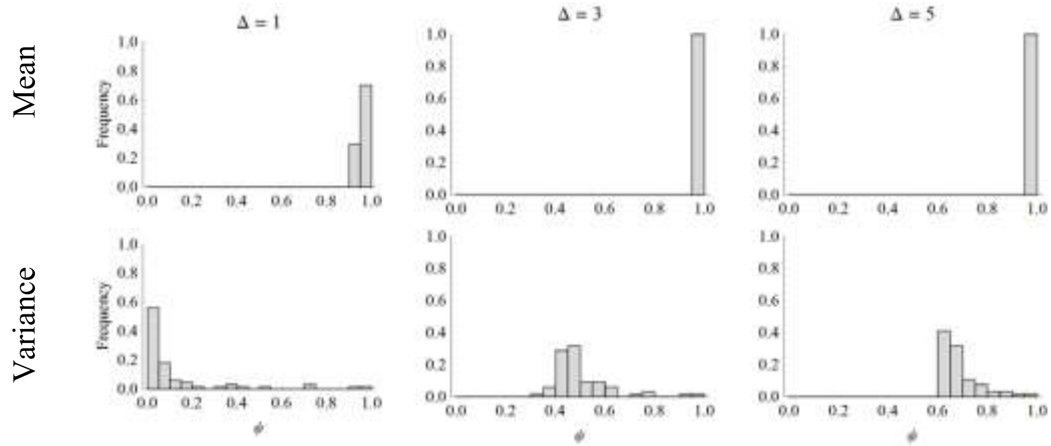

Figure 51. cont.

Figure 52 compares MCS to the other methods. MCS's performance as compared to ESM and EPT has already been noted, but Figure 52 shows the discretization difference between MCS and each of the other methods to be highly significant in both the mean and variance, with $\psi^k_{MCS,d_2}(h,1\%)$ close to 0% for almost all $h \in H$. The comparison to BMd3 in panel (c) shows very low $\psi^1_{\mathrm{MCS,BMd3}}(h,1\%)$ and $\psi^2_{\mathrm{MCS,BMd3}}(h,1\%)$ for most of *H*. These rise to a range of 50–85% (for asymmetric *h*) in the mean, and not greater than 60% in the variance. Although both MCS and BMd3 can have significant errors, their errors are different, and not surmounted by large assessment error. Comparing BMd3 to BMn5 in panel (d) shows that the low similarity for small Δ mostly disappears when Δ = 0.05, where both $\psi^1_{\mathrm{MCS,BMn5}}(h,5\%)$ and $\psi^2_{\mathrm{MCS,BMn5}}(h,5\%)$ are greater than 55% $\forall h \in H$.

Figure 53 compares BMd3 to the other four methods. It is apparent that BMd3's discretization differences with the other methods is large, and very significant even at $\Delta = 5\%$, especially for the variance. Chapter 4 showed BMd3 to be the worst of the methods we considered, which here remains true even for higher assessment error.

Finally, Figure 54 shows the BMn5 comparisons. This method has already been discussed in relation to the other methods, but although its highest similarities in the mean are with ESM and EPT (Figure 54a and b), its highest similarity in the variance tends to be with MCS. BMn5's and MCS's averages and ranges of errors in the variance in Table 37 are also similar, despite their errors for the mean being very different.

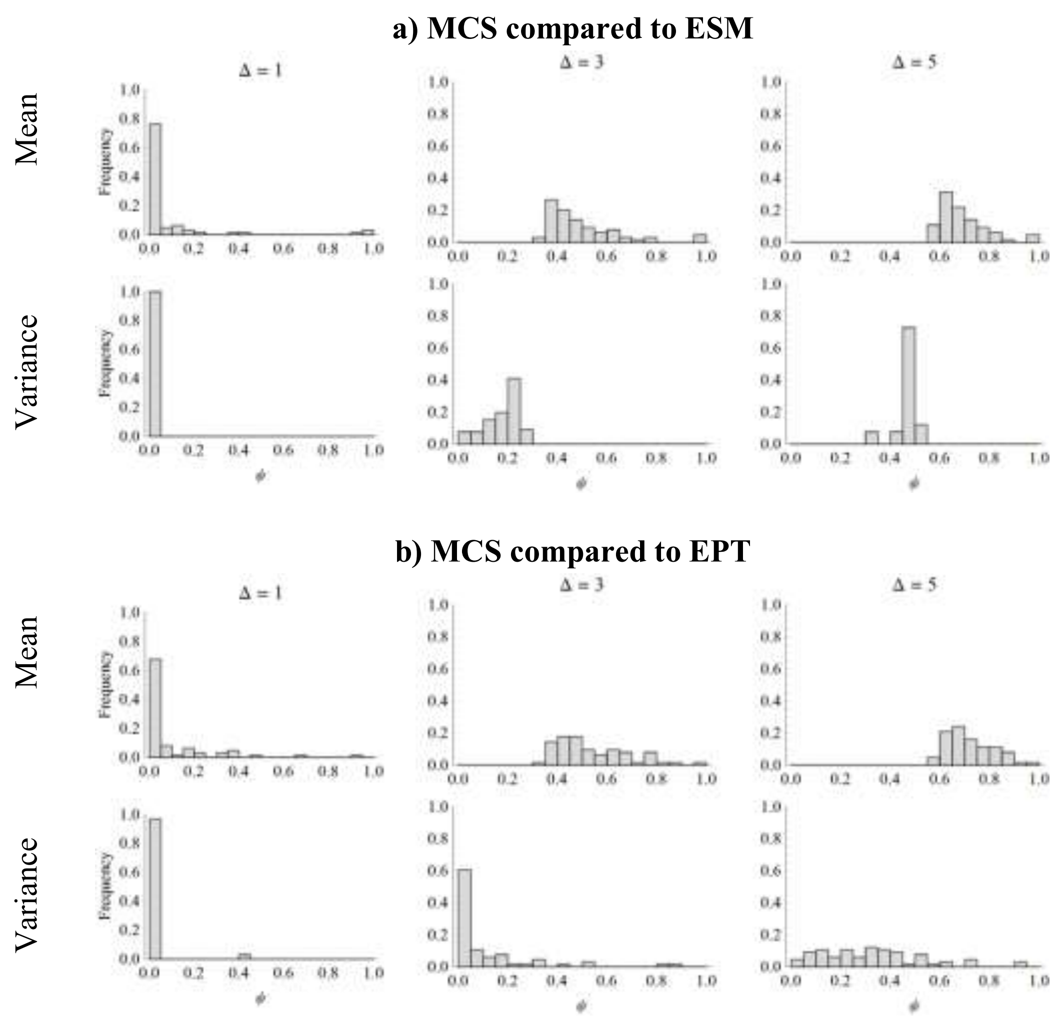

Figure 52. MCS compared to ESM, EPT, BMd3 and BMn5.

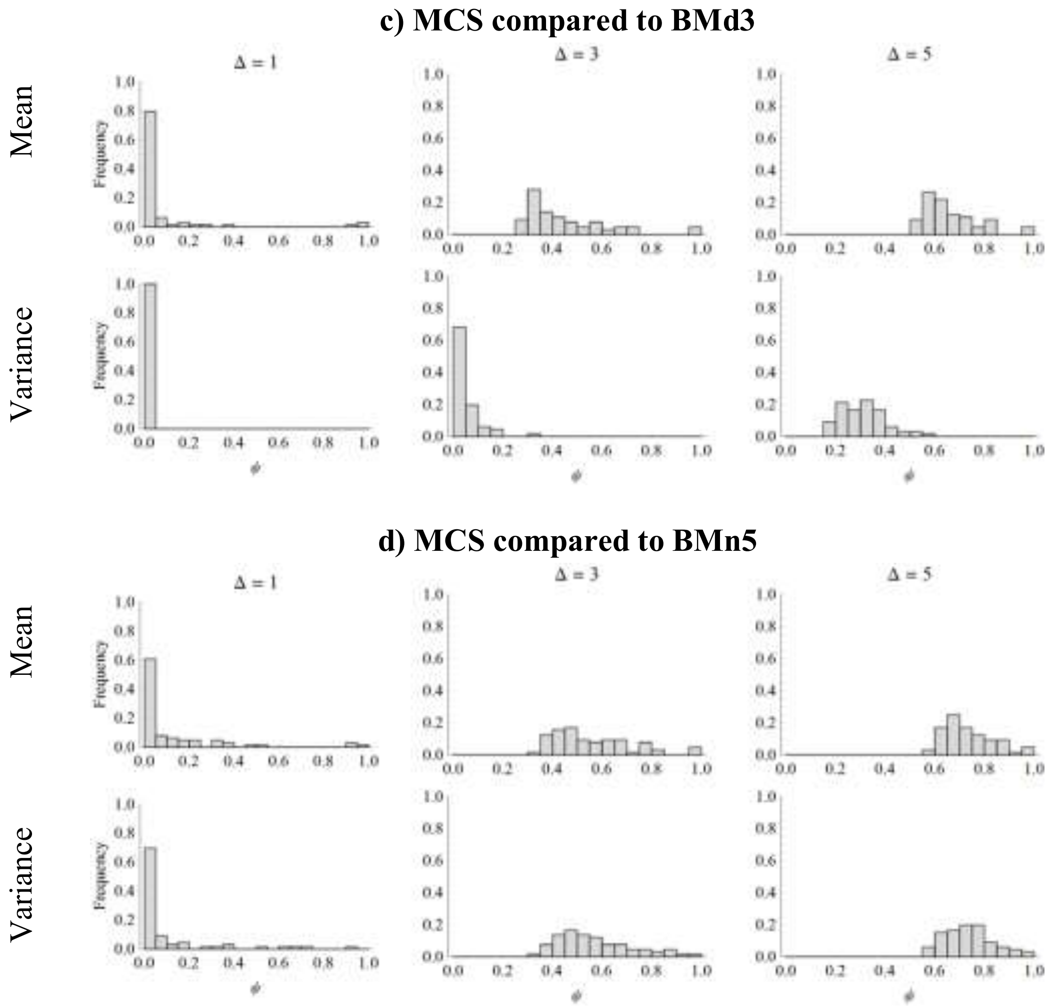

Figure 52. MCS compared to ESM, EPT, BMd3 and BMn5.

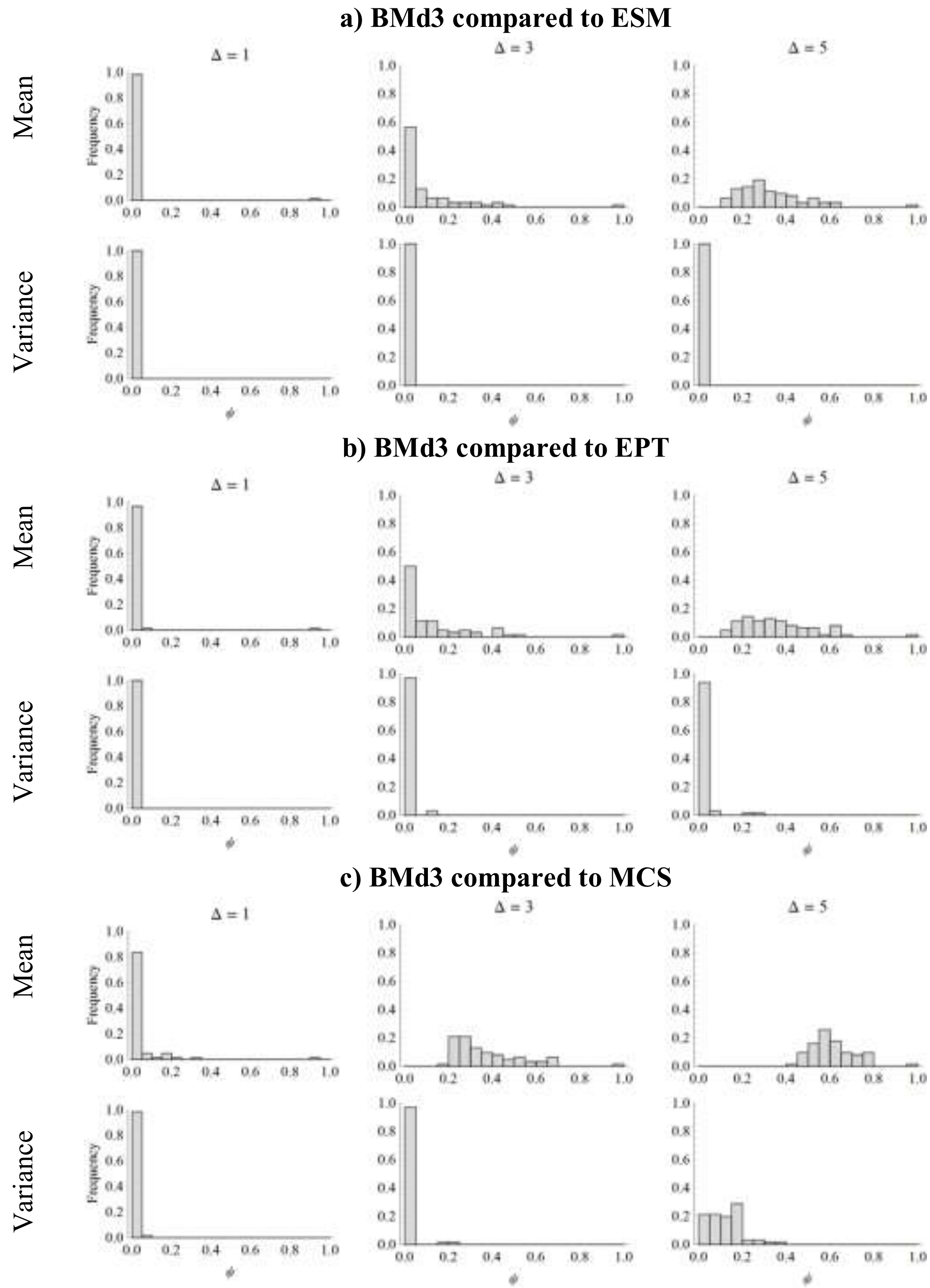

Figure 53. BMd3 compared to ESM, EPT, MCS, and BMn5.

d) BMd3 compared to BMn5

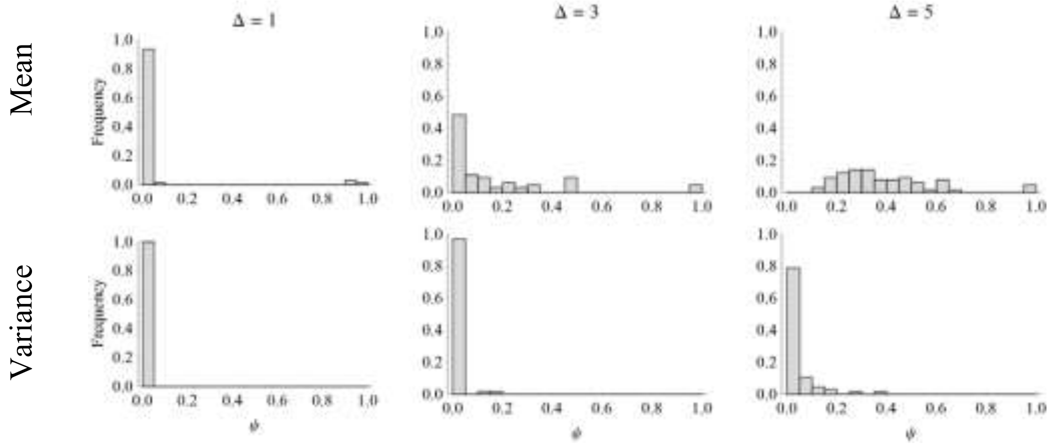

Figure 53. BMd3 compared to ESM, EPT, MCS, and BMn5.

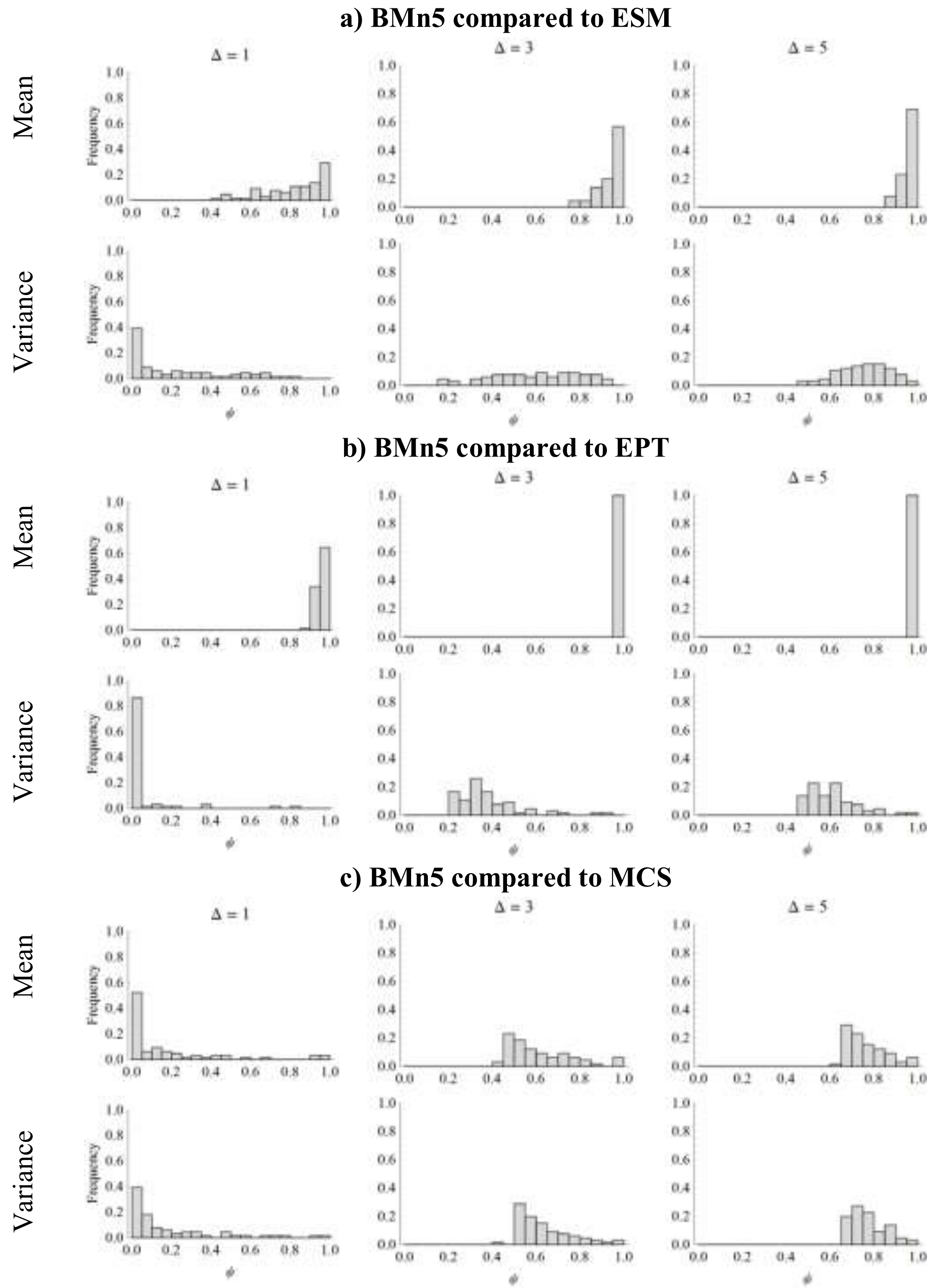

Figure 54. BMn5 compared to ESM, EPT, MCS, and BMd3.

d) BMd3 compared to BMd3

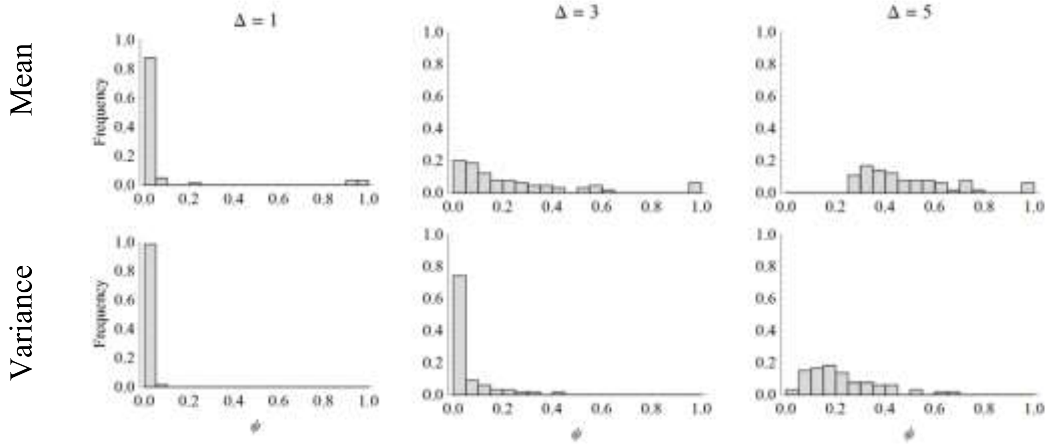

Figure 54. BMn5 compared to ESM, EPT, MCS, and BMd3.

ASSESSMENT ERROR IN PRACTICE

The previous section showed how various levels of assessment error affect the practical impact of using different discretization methods in an experimental setting. However, this section examines empirical studies in probability assessment and calculates the minimum equivalent feasible assessment error Δ in our model. Lichtenstein et al. (1982) gave an excellent overview of work on this topic up to 1980, which they summarized using two measures of calibration: the interquartile index (II) and surprise index (SI). The II is the proportion of actual values that fell between the elicited 25th and 75th percentiles and should be close to 50% for well-calibrated experts. The SI is the proportion of actual values that fell outside the most extreme assessments. An II lower (greater) than 50% and an SI greater (lower) than the amount of probability outside the most extreme percentile assessments indicates overconfidence (underconfidence), or that the assessed distribution is too narrow (wide).

Table 40 summarizes results from six studies on calibration. We consider the results of only those studies that elicited distributions using the percentile method. The first five are a selection from the results surveyed by Lichtenstein et al. (1982), and the sixth is included to broaden the scope of experts. The first three of these (Alpert and Raiffa, 1969; Schaefer and Borcherding, 1972; Selvidge, 1975) used university students as the participants. The last three elicited distributions from meteorologists (Murphy and Winkler, 1974, 1977) and accounting auditors (Tomassini et al., 1982). We report the percentiles elicited in each study, the number of usable distribution assessments N (in most cases multiple subjects each assessed multiple uncertain quantities), the II, the SI and corresponding ideal value, and the minimum feasible assessment error Δ. The minimum feasible Δ was found by calculating the larger deviation of either the II or the SI and dividing by two. The II and SI are both two-sided measures, and the smallest feasible percentile deviation assumes symmetric[3] distribution of errors. For example, the 34% SI in the "Before Training" elicitation of Alpert and Raiffa (1969) means that 34%, rather than 2%, of true values fell outside the assessed P1 and P99. In our error model of uniform Δ in each percentile, the feasible Δ is smallest when the excess 32% of values outside the extreme percentiles are divided equally (or are equally likely to manifest) below the P1 and above the P99. This corresponds to giving the P17 and P83 when asked for the P1 and P99, respectively, or a $\Delta = 0.16$.

The studies in Table 40 show a wide range of assessment errors, from $\Delta = 0.185$ in the first round of elicitations by Schaefer and Borcherding (1972) to $\Delta = 0.02$ of Murphy and Winkler (1977). Based on the results of the previous section, which cover only $0.01 \le \Delta \le 0.05$, choosing a more accurate discretization would be of little use for assessment errors as high as those of the first two studies. However, training and calibration significantly reduce this error. The students in Selvidge (1975), the well-calibrated meteorologists in Murphy and Winkler (1974, 1977), and the second group of accounting auditors in Tomassini et al. (1982) all displayed assessment errors well within the scope of our analysis.

[3] Murphy and Winkler (1974) actually did report that approximately 26% of true values fell below the P25 and 29% above the P75, and 11% fell below the P12.5 and 16% above the P87.5, indicating only minor asymmetry in assessment errors.

Study	Percentiles Elicited	N	Interquartile Index	Surprise Index		Min Δ
				Observed	Ideal	
Alpert and Raiffa (1969)		2270				
Before Training	1,25,50,75,99		34	34	2	16
After Training	1,25,50,75,99		44	19	2	8.5
Schaefer and Borcherding (1972)		396				
First Day	1,12.5,25,50,75,87.5,99		23	39	2	18.5
Fourth Day	1,12.5,25,50,75,87.5,99		38	12	2	6
Selvidge (1975)						
Five Percentiles	1,25,50,75,99	400	56	10	2	4
Seven Percentiles	1,10,25,50,75,90,99	520	50	7	2	2.5
Murphy and Winkler (1974)						
	12.5,25,50,75,87.5	132	45	27	25	2.5
Murphy and Winkler (1977)						
	12.5,25,50,75,87.5	432	54	21	25	2
Tomassini et al. (1982)						
First Group	1,10,25,50,75,90,99	341	71.4	4.7	2	10.7*
				7.8	20	10.7**
Second Group	1,10,25,50,75,90,99	341	54.4	10.8	2	4.4*
				22.1	20	2.2**

*Both the 2% and 20% surprise indices are considered, along with the interquartile index.

**Only the 20% surprise index is considered, along with the interquartile index.

Table 40. Results of selected calibration studies.

Table 41 summarizes results from a fourth study of professionals by Garthwaite and O'Hagan (2000) to elicit assessments of pumping station refurbishment costs and the length of unrecorded S24 sewers in four towns in the UK, using three sets of percentile assessments. The authors gave the frequency of actual values that fell below the lowest quantile and that above the highest quantile. These results indicate a significant tendency to overestimate pumping station refurbishment costs. However, in the discussion of the questionnaires used in the studies, the possibility of anchoring bias is evident: in estimating the actual refurbishment costs of the pumping stations, the experts were given the cost estimates from refurbishment feasibility studies of previous projects, which tended to be distinctly higher than the actual costs realized in those projects. The assessments of the unrecorded length of S24 sewers display less bias, but more overconfidence. The minimum feasible Δ's reflect the asymmetry in the errors and are high in all six cases. The results clearly display some of the biases that careful assessment is designed to avoid.

Uncertain Quantity	Percentiles Elicited	N	Below Quantiles	Between Quantiles	Above Quantiles	Min Δ
Pumping Stations		184				
	33,50,67		65.2	32.6	2.2	32.2
	25,50,75		50	45.7	4.3	25
	17,50,83		40.1	57.3	2.6	23.1
S24 Sewers		252				
	33,50,67		48.1	20.5	31.4	15.1
	25,50,75		46	25.7	28.3	21
	17,50,83		44.8	28.2	27	27.8

Table 41. Garthwaite and O'Hagan (2000) assessment results.

SUMMARY AND DISCUSSION

This chapter extends previous work on discretization by considering the practical issue of error in the assessment of the continuous distribution to be discretized. Using an error model consistent with empirical results from the literature, and a large set of distributions covering a variety of shapes, we have examined the differences in the moment estimates of various discretization methods relative to the errors arising from imprecise assessment.

Earlier chapters confirmed the published finding that EPT is superior to ESM in matching the moments of perfectly assessed distributions. Our results here show that even for small assessment error, such as the level found with well-calibrated meteorologists there is little practical difference in EPT's and ESM's estimates of the mean, in particular, and the variance.

We also concluded that ESM is superior to MCS in matching the moments of perfectly assessed distributions. Unlike EPT, MCS's poorer estimates of the moments are large enough that ESM is still superior under moderate assessment error, such as the levels seen in Alpert and Raiffa (1969) and Schaefer and Borcherding (1972) after training and calibration. BMn, although an exact estimator of the mean, can significantly underestimate the variance, often more so than does EPT. The difference in variance estimates between EPT and BMn5 can also be significant for low degrees of assessment error. BMd3's poor performance, found to be the worst in the previous chapter, continues to be significant throughout our entire range of assessment error, as compared to other methods.

A large body of work exists on probability elicitation, calibration, and training and has been shown both in academic studies and in practice to greatly improve the quality of assessments. Our work indicates that the fidelity of the model (e.g., discretization approximation accuracy in preserving moments) should be matched to, or no lower than, that of the assessment. A carefully elicited assessment from a calibrated expert warrants the most accurate discretization possible, given model complexity constraints (i.e., the set of percentiles or number of discretization points to use). Conversely, a quick, first-pass assessment likely to be of lower quality is not likely to benefit significantly from increases in discretization accuracy, and the simplest method that provides reasonable accuracy is more economical.

Assessment error presents a precision/accuracy tradeoff. More accurate methods will have less discretization error, but will also retain more of the assessment error. In this way, more accurate methods will be more impacted by which distribution is the true one, but might more fully represent the range of possibilities. Fully representing this range is particularly important for sensitivity analysis. An inferior method might not represent the distributions for which the decision would change, and thus mask the necessity for a more accurate discretization and refinement of the assessments.

As noted previously, decision analysis is an iterative process of discovery and refinement of the decision problem and model. Shortcut methods are useful tools for first approximations and initial analysis, but if the assessments of some uncertainties, particularly those to which the decisions are highly sensitive, are refined and improved, the quality of their representation in the model should also be appropriately improved to reflect and preserve that fact.

Chapter 6: Discretization in PERT[4]

This chapter examines several PERT mean and variance estimation formulae. These formulae are related to the discretization methods considered throughout the rest of this work, but the different context yields different constraints. Unlike discretization methods, PERT formulae are not intended for use as discrete probability distributions. Chapter 2 described several new approximation formulae that use common percentiles, which we also examine here. We compare the accuracy of our approach to existing methods by using the Pearson type I-J and I-∩ distributions. The analysis shows that our new method outperforms existing methods when estimating means and variances of most of these distributions.

ANALYSIS METHOD

Our accuracy analysis in this chapter used a different subset of the Pearson distributions than previous chapters, $H'' \subset H_P$. As before, we standardized the distributions to have unit mean and variance. This causes the endpoints of the support to differ from 0 and 1, and because of this, our error results are not directly comparable to those given in the PERT literature, which typically uses the standard (0, 1) beta distribution. Our set H'' of Pearson test distributions contains 274 type I-∩ distributions and 581 type I-J distributions.

For each distribution in H'', we applied the methods under consideration and measured the error between the mean and variance from each method and the distribution's actual mean and variance, respectively.

---

4 The work presented in this chapter was performed in collaboration with Prof. Seong Dae Kim of the University of Alaska Anchorage. Results of this chapter are summarized in

Kim, Seong Dae, Robert Hammond and J. Eric Bickel. 2014. Improved mean and variance estimating formulas for pert analyses. *IEEE Transactions on Engineering Management* **61**(2) 362-369.

.

ANALYSIS

We considered the performance of the mean and variance separately, because many of the methods compared use different formulae for these moments. For ∩-shapes and J-shapes, the AE, ASE, and ME for the mean for each method are given in Table 42, and for the variance in Table 43. The best-performing method for each error measure is highlighted in gray.

Many of the largest errors are seen for PERT variants that use P0 and P100 to estimate the mean. Of these, FS tended to have the lowest errors, because it uses different formulae when the mode is close to a boundary, adapting somewhat to the distribution to reduce error. Both FP methods that use the P0 and P100 had lower errors than the other methods that use these percentiles. These FP methods place almost no weight on the P0 or P100, implying that these percentiles do not add much useful information to that of the median or mode for these distributions.

Classical PERT and GG tend to overestimate the mean for I-∩ or I-J distributions. Most of the PERT variants, including the FP methods that use the P0 and P100, tend to underestimate the mean, although the ME is usually positive. Because all of our test distributions were symmetric or right-skewed, the right tail appears to have caused large errors in some cases.

Our new FP method always estimated the mean more closely when using the P50 instead of the mode in either I-∩ or I-J distributions. For example, in I-J, the ASE of FP 10-50-90 is 0.000, whereas that of FP 10-m-90 is 0.003.

For all methods using the mode, less extreme percentiles led to better accuracy in estimating the mean in either I-∩ or I-J. When using FP 0-m-100 in I-J, the ASE was 0.540, but when using FP 10-m-90, it was reduced to 0.003. When using the P50 instead of the mode, however, using less extreme percentiles did not necessarily lead to better performance. For example, for type I-∩, FP 5-50-95 had the same AE and lower ASE and ME in the mean than did FP 10-50-90.

One of the FP methods had the lowest average error measure in the mean for type I-∩ and I-J, except for ME in the I-∩ mean, which was lowest for EPT. The error table shows that FP performed best among those methods using the P5, P50, and P95. However, among those using the P5, mode, and P95, PG had the lowest ASE and ME, and only a slightly larger AE than did FP. FP was fit to minimize ASE, which was lower for PG, again indicating that rounding the weights affected the performance, although only on the order of 10^{-3}. Among those formulae using the P0, mode, and P100, FS had the lowest AE and ASE, with a ME only slightly larger than that of FP. Although FP was fitted to match these moments, FS has more flexibility to adapt to the distribution. FP generally performs better when it uses the P50 instead of the mode.

Unlike for the I-∩ distributions, the FP formulae had the lowest ASEs for the I-J distributions, as compared to the other methods using the same points. The reason may be that PERT variant methods are typically considered for ∩-shaped beta distributions, and the less regular shape of the I-J distributions left more room for improvement over those methods.

In sum, our new FP methods generally outperformed existing PERT mean-estimation formulue. The FP 5-50-95 or 10-50-90 methods performed the best, having very low errors for ∩- and J-shaped distributions. All of the most accuate measures in Table 42 use the P50 instead of the mode. This is further evidence that discretizations using the mode are generally not as accurate as those using the P50.

The AE, ASE, and ME error measures of variance estimation methods are summarized in Table 43. Both the FP 5-50-95 and FP 5-m-95 methods performed the best on average, for both the I-∩ and I-J distributions. For type I-∩, FP 0-m-100 had the lowest errors in the variance, among methods using the P0 and P100. For type I-J, FP 0-

m-100 performed slightly worse than FP 0-50-100. However, FP 0-50-100 had a large ME of 23.874 for type I-∩. The FP methods using the P10 and P90 also performed well for both distribution types, but those methods did not have the lowest errors.

Method	I-∩			I-J		
	AE	ASE	ME	AE	ASE	ME
PERT	3.581	5.14E+01	37.189	0.637	5.22E+00	39.172
FS	-0.259	4.31E-01	2.186	-0.887	8.12E-01	-1.471
GG	3.248	4.34E+01	34.252	0.520	4.37E+00	36.083
FP 0-m-100	-0.692	7.41E-01	1.978	-0.595	5.40E-01	6.720
FP 0-50-100	-0.155	5.51E-02	0.796	-0.271	1.15E-01	2.814
PERT mod	-0.222	6.33E-02	-0.724	-0.188	6.44E-02	-0.887
MR	-0.378	1.66E-01	-0.764	-0.378	1.65E-01	-0.919
PG	-0.003	1.81E-03	-0.279	0.148	4.17E-02	0.462
FP 5-m-95	0.001	4.81E-03	-0.345	-0.003	3.97E-03	-0.232
EPT	0.001	1.00E-06	-0.002	-0.005	9.40E-05	-0.041
FP 5-50-95	0.000	1.00E-07	-0.003	-0.002	7.50E-05	-0.037
FP 10-m-90	0.001	1.48E-03	-0.254	-0.001	2.53E-03	0.124
FP 10-50-90	0.000	1.90E-05	0.014	0.001	1.61E-04	0.035

Table 42. Errors of mean estimation for different methods

Method	I-∩			I-J		
	AE	ASE	ME	AE	ASE	ME
PERT	58.538	3.81E+04	1457.400	6.202	4.91E+03	1612.140
FS	-0.164	2.40E+00	12.111	-1.000	1.00E+00	-1.000
GG	37.203	1.50E+04	910.298	3.499	1.91E+03	1006.580
FP 0-m-100	-0.901	1.14E+00	7.985	-0.994	9.88E-01	-1.000
FP 0-50-100	-0.410	2.84E+00	23.874	-0.922	8.55E-01	-0.999
PERT mod	-0.458	2.12E-01	-0.663	-0.510	2.65E-01	-0.697
MR	-0.028	3.65E-03	-0.142	-0.104	1.20E-02	-0.163
PG	-0.062	6.46E-03	-0.172	-0.135	1.92E-02	-0.192
FP 5-m-95	0.000	2.98E-04	-0.098	0.000	1.12E-03	0.087
EPT	-0.003	2.09E-04	-0.082	0.099	2.00E-02	0.394
FP 5-50-95	-0.001	1.98E-04	-0.090	0.000	1.43E-03	0.107
FP 10-m-90	0.010	3.90E-03	0.270	0.004	5.19E-03	0.168
FP 10-50-90	-0.002	3.48E-03	0.152	0.004	1.06E-02	0.244

Table 43 Errors of variance estimation for different methods.

Comparing Individual Distributions

The previous section showed that EPT and FP had similar performance, with FP tending to have the lower errors. As discussed in Chapter 4, for our new shortcuts, optimizing FP with respect to ASE does not mean it will have lower error for all distributions. Although FP 5-50-95 performed better on average, EPT had lower ME for I-∩ distributions. This section makes a more granular comparison between EPT and FP methods that use the P5 and P95 for individual distributions in the Pearson distribution system. Figure 55 displays the portion of the Pearson system that encompasses the distributions used in our error comparison of the previous section. We again use the parameterization of $\beta_2 - \beta_1 - 1$ instead of β_2 on the vertical axis in Figure 55 to improve the presentation of the area of interest.

Figure 55 shows which of FP 5-50-95, FP 5-m-95, or EPT had the lowest absolute error for a given distribution. When estimating the mean, FP 5-50-95 had the lowest error for most of the I-∩ region. FP 5-50-95 and EPT were each best for large portions of the I-J region. For the variance, the areas where each method was best divide up the entirety of both regions roughly equally. Either FP 5-50-95 or EPT was best for most of the type I-∩ distributions, whereas each of the three was best for large parts of the I-J distributions.

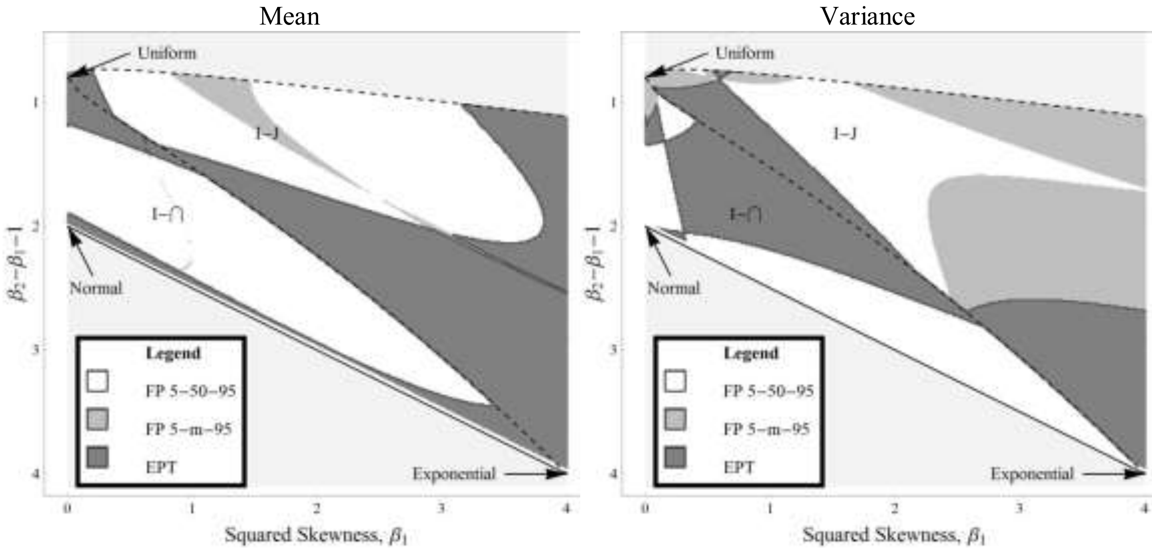

Figure 55. Comparison of FP 5-50-95, FP 5-m-95, and EPT in the type I-∩ and I-J Pearson distributions.

SUMMARY AND DISCUSSION

This chapter compared the performance of several existing PERT methods to each other and to our new methods, for a portion of the Pearson system commonly used to evaluate PERT approximations. The FP methods have two primary merits: accuracy and flexibility. They are more accurate than previous methods for a given set of points, often by significant margins. They are also flexible, in that the appropriate method can be chosen based on the points used, distribution shape, and moment to be estimated.

The results of Chapter 4 showed that more extreme percentiles generally provided better estimates of a distribution's moments. This chapter indicated some limits to this reasoning. PERT methods using the P0 and P100 performed the worst overall. Although two of the EPT++ methods used either the P1 or P99, most of these and the EPT+ methods used less extreme percentiles, despite having the freedom to use any percentiles between the P1 and P20. Our new FP methods, which used the P5 and P95 or the P10 and P90, usually outperformed other methods that use similar percentiles.

For PERT practitioners, the FP methods improved accuracy and easier assessment (for those methods using less extreme percentiles than the P0 and P100) make them excellent substitutes for existing PERT methods in most cases. Because we use different sets of percentiles for the median and mode, the analyst can choose the most appropriate method for the percentiles obtained. The FP methods provided improvements for any of these sets of discrete points, but if any percentiles can be used as input points, FP 5-50-95 performs best and, in most cases, beats the best existing method, EPT, by a small margin. Using the mode with these percentiles is better than the median only when estimating the variance in the I-J region.

Chapter 7: Summary and Future Research

This chapter concludes with a summary of results and conclusions, and some recommendations for future work avenues illustrated by a simple example.

SUMMARY

This work has broadly examined the approximation of continuous probability distributions by discrete distributions, or discretization. Although this idea has long existed in many fields, our particular concern was discretization in decision analysis. We examined the accuracy of both simple shortcut methods and distribution-specific methods of varying degrees of complexity. We compared the methods' accuracies in approximating the moments of a large set of distributions having a variety of shapes and considered the effects of imperfect probability assessments. We also briefly examined extensions of this work as new PERT formulae for approximating the means and variances of task completion time distributions.

This work produced three primary sets of results, which were given in Chapters 4-6. Chapters 2 and 3 laid the foundation for our analysis, by reviewing discretization methods and the Pearson and Johnson distribution systems, respectively. Chapter 2 also presented a procedure for constructing three-point approximations that minimize the average squared error in the mean over sets of distributions. We also used this procedure to construct new PERT mean and variance approximation formulae.

Chapter 4 presented the results of our analysis of error in the moments for a wide set of distributions over large portions of the Pearson and Johnson systems. The results from these two systems were similar because the shape of distributions from each system having the same moments were very similar. Of the previously proposed methods, EPT and ZDI generally had the lowest error for unimodal, ∩-shaped distributions. This confirms the conclusions in earlier work on EPTs performance. ZDI had not been rigorously studied, but uses similar percentiles and probabilities as does EPT. These methods are also close to optimal, because they bear much similarity to several of the EPT+ and EPT++ methods, which are optimal with respect to average squared error. Even when the methods themselves are not similar, their performances in matching the moments often were. These results also indicate that for three-point shortcut methods, the P50 is best complemented by percentiles near the P5 and P95. These methods also show improved performance over a five-point BMn in matching the variance and higher moments.

Of the shortcut methods using the P10, P50, and P90, which are common in practice, ESM typically had smaller errors than did MCS, but still had room for improvement. The SP methods, tailored to the distribution type, provided the best overall performance for these percentiles. ESM has similar probability weights as the SP method for the generalized lognormal distribution, which is close in shape to the distributions for which ESM furnishes the optimal weighting of the P10, P50, and P90 for the mean. However, this improved performance in the mean does not carry over to moments higher than the variance.

Chapter 5 considered the effects of imperfectly assessed distributions. In decision analysis, distributions are often elicited from a subject matter expert who assesses percentiles of the distribution, which may be incorrect because of cognitive biases, for example. Even minimal assessment error would dominate the differences in the accuracies of EPT and ESM in matching the mean and variance. MCS's and BMd3's larger errors, however, were still significant under larger assessment error.

Chapter 6 used the discretization error analysis framework to analyze several discrete approximation formulae of the mean and variance from the PERT literature. Some of the original methods have very large errors in these moments. Our new methods greatly improve upon these and provide the analyst additional modeling flexibility. Some of these methods bear resemblance to discretization methods such as EPT and ESM. These methods tend to be asymmetric and sometimes place no weight on one of the three points.

CONCLUSIONS

Several broad conclusions can be drawn from this work. First and foremost, the number of discrete points used in an approximation is not nearly as important as how those points are chosen. In many instances, three-point shortcut methods had smaller errors in preserving the moments than did five-point distribution-specific methods. Gaussian quadrature can match the highest number of moments out of any *n*-point discretization by carefully choosing both the values and probability weights used. At the opposite extreme, it can take thousands, or tens of thousands, of random Monte Carlo samples to reduce the sampling error to where it is equivalent to the discretization error of the three-point shortcut method EPT (Bickel et al., 2011). Choosing those points wisely can reduce the number of points needed and can increase accuracy.

Second, the fidelity of discretizations as approximation models should be considered in the context of the problem. A highly accurate representation of poor data is no more useful than a simpler one. Chapter 5 showed that as the error in assessed percentiles increased, the effective differences among discretization models' moment estimates decreased. However, some methods required higher assessment error for this to be true. Applying a highly accurate discretization to a distribution that is not assessed with much accuracy (e.g., from an uncalibrated expert not familiar with probability) will hardly affect the quality of the decision. Conversely, a very carefully assessed distribution warrants a highly accurate discretization, to best represent the information contained in the distribution.

Third, discretization methods presented tradeoffs. GQ is the most accurate method for matching moments, but it has several drawbacks. It requires precise knowledge of the full probability distribution, uses extreme percentiles, and generally must be implemented in software. If the distribution is not known with full accuracy, GQ can still be applied to an approximated distribution, but this would raise the issues of fidelity noted in the second point above. BMn perfectly matches the mean of any distribution, but it systematically underestimates the variance, higher even moments, and often the odd moments as well. However, this method can be performed on a graphically fitted distribution without requiring precise knowledge of percentiles. BMd is very poor at matching moments, but it is even easier to apply than MBn, requiring at minimum only a set of percentiles, although to use an arbitrary set of percentiles, the full distribution must be known. Shortcut methods generally require the least information, and some can match the mean and variance of many distributions quite well. However, they can have large errors in the higher moments and do not give the analyst the freedom to choose the number of points. By using basic information about the distribution, such as its support boundedness or the type of distribution shape, the new EPT+, EPT++, or SP shortcuts can be used to better tailor the discretization to the distribution than do general shortcut methods.

Discretization should be treated as a modeling tool, with the method adapted to the distribution as indicated by the analysis of the problem. Decision analysis is an iterative process (although iteration might not be required in a given situation), and the information uncertainty in the problem can change through the analysis, as can the structure of the model itself. A discretizated distribution should not be considered a static probability distribution, but rather one of many possible representations of the uncertainty.

SUGGESTIONS FOR FUTURE RESEARCH

Distribution Characteristics Other than Moments

Moments are the primary concern of this work, but being that they do not always imply a unique distribution, there are several reasons to consider other accuracy measures as well. In practice, the expectation of the value function is not the only metric of interest. Decision-makers might also be interested in percentiles of the value distribution or the probability of a negative-valued outcome. Additionally, for some distributions, such as the Cauchy distribution, no moments exist, and measures of accuracy other than moment preservation must be used.

To compare distributions directly via their cdfs, rather than through summary statistics such as moments, we could consider L_p-norm measures of distribution similarity. The L_p-norm in defined as

$$L_p\left(F,G\right)=\left[\int_{x\in S}\left|F(x)-G(x)\right|^p f(x)dx\right]^{1/p}, \tag{66}$$

for a real number $p \geq 1$. We take F to be the continuous cdf (and f its pdf), with support S, and G to be the discrete approximation cdf. Just as we could consider any of an infinite number of moments, we can select from an infinite set of norms. Different p values emphasize different aspects of the distributions. Three of the most common norms are given below.

The L_1 norm, also called the Kantorovich-Rubinstein distance (Villani, 2009),

$$L_1\left(F,G\right)=\int_{x\in S}\left|F(x)-G(x)\right| f(x)dx, \tag{67}$$

is the expected absolute difference between the cdfs, with respect to F. The L_2 norm, also known as the Cramér–von Mises distance (Anderson, 1962),

$$L_2\left(F,G\right)=\left[\int_{x\in S}\left|F(x)-G(x)\right|^2 f(x)dx\right]^{1/2}, \tag{68}$$

is similar to root-mean-squared error, but it places more emphasis on larger errors. Finally, the L_∞ norm, or Kolmogorov-Smirnov distance (Darling, 1957),

$$L_\infty\left(F,G\right)=\max_{x\in S}\left|F(x)-G(x)\right| , \tag{69}$$

is the maximum absolute error between *F* and *G*.

Multiple Discretizations

Decision problems of any practical interest typically have multiple uncertainties, often with decisions interspersed among the resolution of uncertainties. Throughout this work, we have analyzed discretization performance in matching characteristics of only a single distribution. Although the univariate Taylor expansion argument extends to multiple dimensions, some uncertainties carry more importance in a problem than do others. Because the number of discretization points can quickly increase the complexity of the model, choosing the smallest number of scenarios that adequately represent the problem is an important consideration, and some uncertainties may require better discrete approximations than others.

The following example from Smith (1993) will illustrate some of the issues and questions faced when discretizing multiple, possibly dependent, uncertainties. A wildcatter is valuing a proven but undeveloped oil field. Assuming a flat production-rate model, the field's net present value (NPV) for an oil volume in barrels (bbl) *v*, recovery factor percentage *r*, oil price in \$/bbl *p*, and production cost in \$/bbl *c* is given by

$$\begin{aligned} u(v,r,p,c) &= \int_{t=0}^{T} k(p-c)e^{-\delta t}dt - C \\ &= \frac{1}{\delta}\left(p-c\right)k\left(1-exp\left(-\delta\frac{rv}{k}\right)\right)-C \end{aligned} , \tag{70}$$

where k is the constant production rate of 100,000 bbl/year, C is the initial capital expenditure of \$2.5 million, and δ is the discount rate of 5%. Production begins at time t = 0 and ends at time T. The four parameters oil volume, recovery factor, oil price, and cost are uncertain, and cost is probabilistically dependent on oil price. The influence diagram for this problem is shown in Figure 56.

Recovery has a gamma distribution with parameters $\alpha = 16, \beta = 1$. Reservoir volume has a generalized lognormal, or S_L, distribution with parameters $\gamma = -1.581$, $\delta = 3.162$, $\mu = 3.5$, $\sigma = 1$, giving it a lower bound of 3.5. Oil price has a Pearson type I-∩ distribution with $\mu = 19.362$, $\sigma^2 = 40.643$, $\beta_1 = 0.319$, $\beta_2 = 2.700$. Cost has a normal distribution, conditional on oil price *p*, with parameters $\mu = p/3+3$, $\sigma = \sqrt{p/16}$.

Figure 57 shows a tornado diagram of the uncertainties, using their P50 values as the baseline scenario, and setting each uncertainty to its P10 (black) and P90 (grey) values individually, while holding the rest at their P50s. Oil price by far has the highest impact on NPV, followed by cost, recovery factor, and oil volume, in descending order.

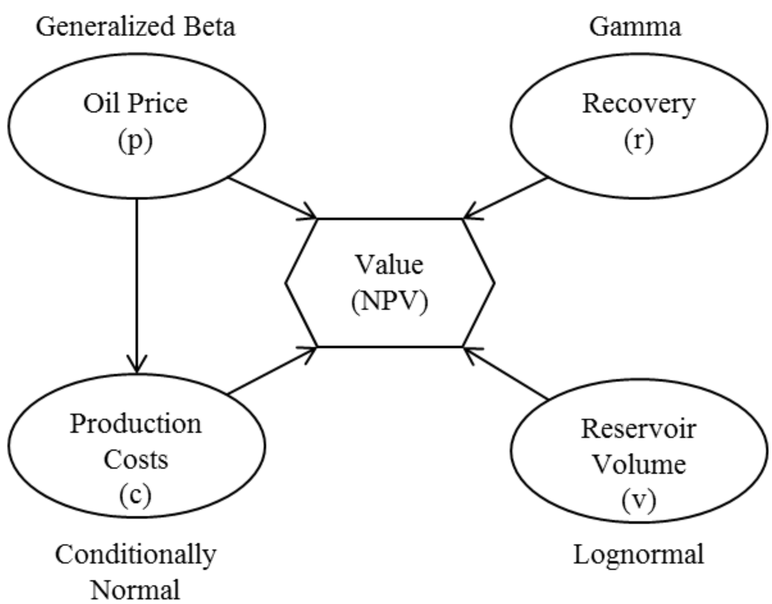

Figure 56. Influence diagram for the Wildcatter Problem.

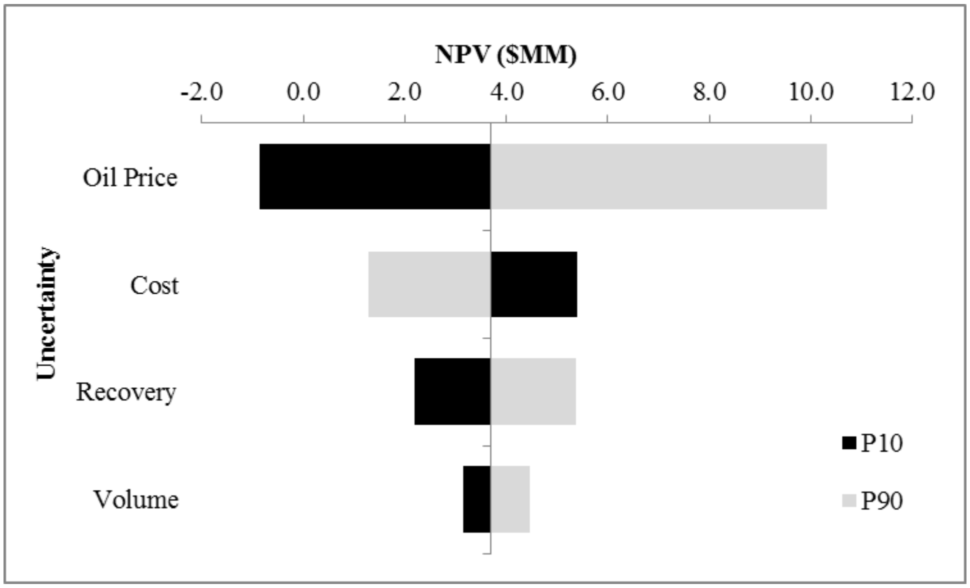

Figure 57. Tornado diagram for the Wildcatter Problem.

Tailoring Discretization to the Problem

An analysis using the same three-point discretization shortcut for each uncertainty will model the uncertainty in oil price with the same fidelity as oil volume, even though the former has a nearly 10-times-larger impact on NPV than does the latter. Rather than use three discretization points for each variable (which would result in nine unique tree endpoint combinations of these two uncertainties), we can create a possibly more accurate model by discretizing the oil price uncertainty with more points, such as six, and treating oil volume as deterministic. This also reduces the computational complexity of the model by using only six points for these two uncertainties.

The tornado diagram indicates the sensitivity of the value function to the uncertain inputs, but it does not say which aspects of their distributions are important. This sensitivity information provides little guidance on how to refine the discrete approximations. Figure 58 shows the KS-distances versus the total number of endpoints of the probabilistic model, for all combinations of one to five GQ points for each of the four uncertainties of the Wildcatter problem. There is a wide range of performance for the same number of total points, meaning that performance is heavily dependent on how points are allocated to specific input distributions. For the optimal allocations of points in this problem, which lie along the bottom and to the left of other points in Figure 58, very large improvements can be made by increasing the total number of points in the value distribution from, say, 10 to 50, but much less improvement from 50 to 500. The best GQ discretization scheme using 60 total points has a KS-distance of 0.045, whereas the scheme using 625, which assigned 5 points to each uncertainty, has a KS-distance of 0.033. This is a small decrease in error, considering the large increase in problem complexity.

The 60-point scheme assigns 3 points to reservoir volume, 2 to recovery, 5 to oil price, and 2 to cost. It should not be surprising that oil price has the most points, because Figure 57 showed it to have the largest impact on NPV. The clusters of points seen in Figure 58, such as between a KS-distance of 0.25 and 0.35 and between zero and 100 points, occur from assigning different numbers of points to the GQ of oil price. To display this more explicitly and to contrast oil price with the other distributions, Table 44 gives the average KS-distances over all discretization schemes having a given number of points for each discretization. The same analyses for BMn and BMd (both used with equally sized brackets) are given in Table 45 and Table 46, respectively. In each table, the average KS-distance is largest when oil price is represented by only one point (which for GQ and BMn is the mean oil price, and for BMd is the median) and is smallest when represented by five points.

Sensitivity analysis guides the selection of uncertainties to model and can also guide the assignment of their approximation methods. More important uncertainties, such as oil price in the Wildcatter problem, warrant more careful assessment and more accurate discretization than do less important uncertainties like reservoir volume. This example motivates the question of how to optimally allocate discretization points and how to improve this aspect of decision analysis practice.

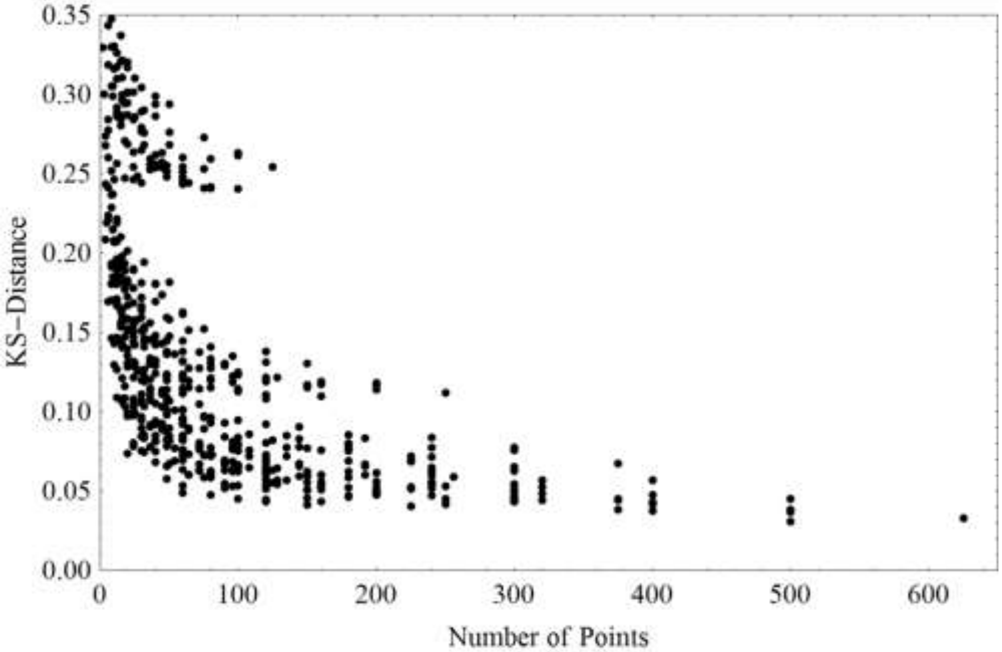

Figure 58. KS-distance vs. number of discretization points for GQ in Wildcatter problem.

	Number of Discretization Points				
	1	2	3	4	5
Volume	0.178	0.154	0.141	0.138	0.135
Recovery	0.218	0.139	0.133	0.130	0.127
Oil Price	0.318	0.155	0.111	0.091	0.073
Cost	0.189	0.143	0.141	0.139	0.135

Table 44. Average KS-Distance for different numbers of GQ points for each uncertainty.

	Number of Discretization Points				
	1	2	3	4	5
Volume	0.138	0.120	0.114	0.110	0.107
Recovery	0.175	0.120	0.105	0.097	0.092
Oil Price	0.320	0.113	0.067	0.049	0.041
Cost	0.149	0.121	0.111	0.106	0.102

Table 45. Average KS-Distance for different numbers of GQ points for each uncertainty.

	Number of Discretization Points				
	1	2	3	4	5
Volume	0.136	0.123	0.117	0.113	0.111
Recovery	0.173	0.124	0.108	0.100	0.095
Oil Price	0.294	0.124	0.078	0.057	0.048
Cost	0.148	0.123	0.114	0.108	0.105

Table 46. Average KS-Distance for different numbers of BMd points for each uncertainty.

Correlation and Discretization

A key question in multivariate discretization is how to treat probabilistic dependence. Independent uncertainties can be discretized separately, by exploiting the separability of this problem as described by Smith (1993). One approach for two variables is to discretize the marginal distribution of one uncertainty and then discretize the conditional distribution of the other, conditioned on each discrete point of the first. However, this procedure considers only one uncertainty at a time, using marginal and conditional distributions, rather than the total joint distribution. Thus, the order of discretization may affect the overall discrete approximation of the joint distribution. Rather than discretizing distributions individually, how can we discretize joint distributions directly, and what performance improvements does this provide?

Evaluating Multivariate Discretization

Because multivariate joint distributions do not have moments, other measures are needed to evaluate multivariate discretizations. For example, a bivariate uniform distribution has a well-defined pdf and cdf, but it has no notion of moments without a function mapping the two variables to one dimension. However, the cdf of the continuous joint distribution can be compared to the discrete joint approximation distribution.

Moments can be used to evaluate the discretization of value distributions, but Smith (1993) warned that certain types of value functions, such as utility functions with small risk tolerances, might not be well approximated by Taylor series expansions. In these cases, moments might still perform well, but preserving other characteristics of the input distributions might give better approximations of the output value function distribution.

Bounds on Distributions

One question raised in comparing distributions is how to bound the moments or cdf of the distribution resulting from a discretization, given limited information. This work has empirically investigated errors resulting from discretization, but can these errors be theoretically bounded? Are there bounds on the percentiles of a value distribution with multiple discretized input uncertainties?

Chebyshev bounds provide limits on expectations of functions of a random variable given some of its moments. Karlin and Studden (1966) gave a detailed treatment of the theory of Chebyshev systems, including Chebyshev inequalities. Zelen (1954) gave specific formulae for Chebyshev bounds that use up to the first four moments. Smith (1990, 1995) discussed applications of Chebyshev bounds in decision analysis. Figure 59 shows the Chebyshev bounds on the NPV distribution of the Wildcatter problem, using its first two or first four moments. These bounds are not particularly narrow, even with four perfectly known moments. The probability of a negative NPV is anywhere between 0% and approximately 25% using four moments, and between 0% and 40% using two moments. These bounds assume perfect knowledge of the required moments. If it were known that the moments were not exact, such as if they were estimated using discretization, the bounds should be made even wider. Further research might investigate variants of Chebyshev-type bounds that can address these issues.

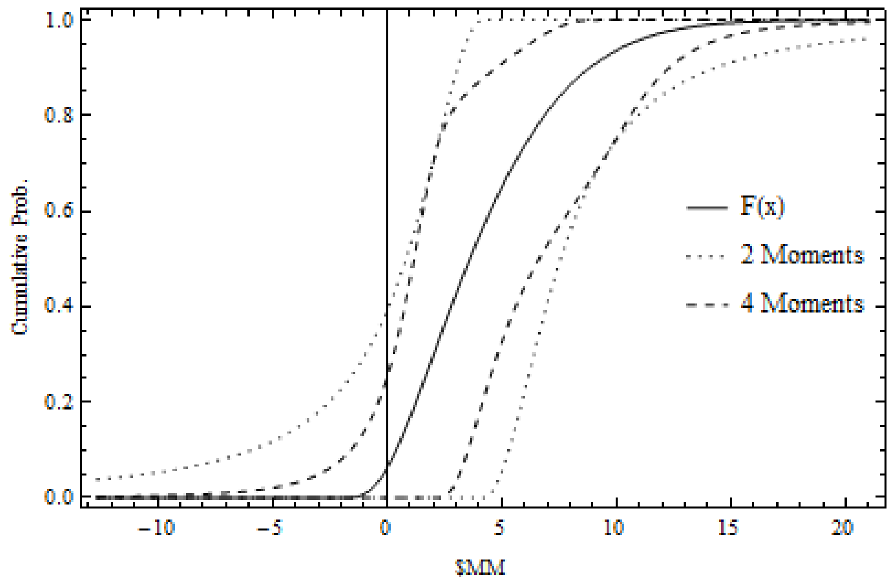

Figure 59. Chebyshev bounds on the value function of our example using up to the first four moments.

Assessment Error

Our assessment error study in Chapter 5 used a limited set of distributions, for the reasons discussed in that chapter. This work can be extended in several ways. Other distributions could be considered, but they would need to address the issues that restricted us to bounded distributions. The P10, P50, and P90 percentiles were used partly because they were easier to assess than more extreme percentiles would have been. However, this considers only the magnitude of assessment errors and not their effects on the problem when approximated by discretization. Finally, our assessment error model assumed a uniform distribution of Δ-feasible distributions in the truth set, but other parameterizations and distributions could be considered. The form and distribution of errors could be chosen to reflect different assumptions of assessment error.

Appendix: Johnson System Moment Fitting Algorithms

This appendix gives algorithms from Hill et al. (1976) to fit each of the three Johnson distribution types to a given set of the first four standardized moments, μ, σ^2, β_1, β_2. The S_L distributions only require three, and the first three standardized moments are used. The code given below can be copied directly into Mathematica® 9.0, and is also available from the author upon request.

S_L DISTRIBUTION

```
MatchParamsSL[\[Mu]1_,\[Mu]2_,\[Beta]1_]:=Module[{y,w,\[Delta],\[Omega],\[Gamma
],\[Mu],\[Sigma]},
(
y=w/.Flatten[Solve[{\[Beta]1==(w-1)*(w+2)^2,w>=0},{w}]]//N;
\[Delta]=Log[y]^(-0.5);
\[Omega]=Exp[\[Delta]^-2];
\[Gamma]=0.5\[Delta]*Log[\[Omega] (\[Omega]-1)/\[Mu]2];
\[Mu]=\[Mu]1-Exp[-\[Gamma]*\[Delta]^-1+0.5\[Delta]^-2];
\[Sigma]=1;
Return[{\[Gamma],\[Delta],\[Mu],\[Sigma]}];
)]
```

S_U DISTRIBUTION

```
MatchParamsSU[\[Mu]1_,\[Mu]2_,\[Beta]1_,\[Beta]2_]:=Module[{\[Delta]1,A0,A1,A2,\
[Beta]1f,\[Delta],\[Omega],\[Gamma],\[Mu],\[Sigma],a,b,c,m,\[Omega]2,\[Delta]2},
(
\[Delta]1=5;
A0[w_]:=w^5+3w^4+6w^3+10w^2+9w+3;
A1[w_]:=8(w^4+3w^3+6w^2+7w+3);
```

```
A2[w_]:=8(w^3+3w^2+6w+6);
\[Beta]1f[m_,w_]:=m(w-1)(4(w+2)m+3(w+1)^2)^2/(2(2m+w+1)^3);
\[Delta]=\[Delta]1;
\[Omega]=Exp[\[Delta]^-2]//N;
a=A2[\[Omega]](\[Omega]-1)-4(2\[Beta]2-6);
b=A1[\[Omega]](\[Omega]-1)-4(\[Omega]+1)(2\[Beta]2-6);
c=(\[Omega]-1)A0[\[Omega]]-(2\[Beta]2-6)(\[Omega]+1)^2;
m=(-b+Sqrt[b^2-4a*c])/(2a)//N;
a=0.5;
b=1;
c=3/2-\[Beta]2+\[Beta]1/\[Beta]1f[m,\[Omega]]*(\[Beta]2-
0.5(\[Omega]^4+2\[Omega]^2+3));
\[Omega]2=Sqrt[(-b+Sqrt[b^2-4a*c])/(2a)];
\[Delta]2=Log[\[Omega]2]^-0.5;
\[Delta]=\[Delta]2;
While[Abs[\[Beta]1f[m,\[Omega]]-\[Beta]1]>=10^-6,
\[Omega]=Exp[\[Delta]^-2]//N;
a=A2[\[Omega]](\[Omega]-1)-4(2\[Beta]2-6);
b=A1[\[Omega]](\[Omega]-1)-4(\[Omega]+1)(2\[Beta]2-6);
c=(\[Omega]-1)A0[\[Omega]]-(2\[Beta]2-6)(\[Omega]+1)^2;
m=(-b+Sqrt[b^2-4a*c])/(2a)//N;
If[\[Beta]1f[m,\[Omega]]==0,Break[],{}];
a=0.5;
b=1;
c=3/2-\[Beta]2+\[Beta]1/\[Beta]1f[m,\[Omega]]*(\[Beta]2-
0.5(\[Omega]^4+2\[Omega]^2+3));
```

```
\[Omega]2=Sqrt[(-b+Sqrt[b^2-4a*c])/(2a)];
\[Delta]2=Log[\[Omega]2]^-0.5;
\[Delta]=\[Delta]2;
];
\[Gamma]=ArcSinh[Sqrt[m/\[Omega]]]*\[Delta];
\[Sigma]=Sqrt[\[Mu]2/(.5*(\[Omega]-1)(\[Omega]*Cosh[2\[Gamma]/\[Delta]]+1))];
\[Mu]=\[Mu]1+\[Sigma]*Sqrt[\[Omega]]Sinh[\[Gamma]/\[Delta]];
{Re[\[Gamma]],Re[\[Delta]],Re[\[Mu]],Re[\[Sigma]]}
)]
```

S_B DISTRIBUTION

```
MatchParamsSB[\[Mu]1_,\[Mu]2_,\[Beta]1_,\[Beta]2_]:=Module[{xbar,sigma,rtb1,b2,H
MU,deriv,dd,tt,tol,limit,rb1,b1,neg,e,u,x,y,w,f,d,g,m,fault,s,h2,t,h2a,h2b,h3,h4,rbet,bet2,d
elta,xlam,xi,gamma,lastd,lastg},
(
(*Inputs*)
xbar=\[Mu]1;
sigma=\[Mu]2;
rtb1=Sqrt[\[Beta]1];
b2=\[Beta]2;
(*Initialize*)
HMU=ConstantArray[0,6];
deriv=ConstantArray[0,4];
dd=ConstantArray[0,4];
tt=10^-4;
tol=0.01;
```

```
limit=50;
rb1=Abs[rtb1];
b1=rb1^2;
neg=rtb1<0;
(*Get D as first estimate of delta*)
e=b1+1;
u=1/3;
x=b1/2+1;
y=rb1*Sqrt[b1/4+1];
w=(x+y)^u+(x-y)^u-1;
f=w^2*(3+w*(2+w))-3;
e=(b2-e)/(f-e);
If[Abs[rb1]>tol,
(
d=1/Sqrt[Log[w]];
If[d<0.64,f=1.25*d,f=2-8.5245/(d*(d*(d-2.163)+11.346))];
),
(f=2;)
];
f=e*f+1;
If[f<1.8,
(d=0.8*(f-1);),
(d=(0.626*f-0.408)*(3-f)^(-0.485);)
];
(*Get g as first estimate of gamma*)
g=0;
```

```
If[b1<tt,{},
(
If[d<=1,g=(0.7466*d^1.7973+0.5955)*b1^0.485;,
(
If[d<=2.5,(u=0.0623;y=0.4043;),(u=0.0124;y=0.5291;)];
g=b1^(u*d+y)*(0.9281+d*(1.0614*d-0.7077));
)];
)];

(*Main iteration starts here*)
m=0;
While[(Abs[u]>tt||Abs[y]>tt)&&m<=limit,
(
m=m+1;
(*fault=m>limit;
If[fault,Return[{"Fault","d,g",d,g,m}]];*)

(*get first 6 moments for latest g and d values*)
HMU=Table[Moment[JohnsonDistribution["SB",g,d,0,1],i],{i,1,6}];
s=HMU[[1]]^2;
h2=HMU[[2]]-s;
If[h2<=0,(Print["nonpositive variance"];g=lastg;d=lastd;Break[])];
t=Sqrt[h2];
h2a=t*h2;
h2b=h2^2;
(*h3=HMU[[3]]-HMU[[1]]*(3*HMU[[2]]-2*s);*)
```

```
h3=HMU[[3]]-3*HMU[[1]]*h2-HMU[[1]]^3;
rbet=h3/h2a;
h4=HMU[[4]]-HMU[[1]]*(4*HMU[[3]]-HMU[[1]]*(6*HMU[[2]]-3*s));
bet2=h4/h2b;
w=g*d;
u=d^2;

(*Get derivatives*)
For[j=1,j<=2,j++,
For[k=1,k<=4,k++,
(
t=k;
If[j==1,s=HMU[[k+1]]-HMU[[k]],s=((w-t)*(HMU[[k]]-
HMU[[k+1]])+(t+1)*(HMU[[k+1]]-HMU[[k+2]]))/u];
dd[[k]]=t*s/d;
)];
t=2*HMU[[1]]*dd[[1]];
s=HMU[[1]]*dd[[2]];
y=dd[[2]]-t;
deriv[[j]]=(dd[[3]]-3*(s+HMU[[2]]*dd[[1]]-t*HMU[[1]])-1.5*h3*y/h2)/h2a;
deriv[[j+2]]=(dd[[4]]-
4*(dd[[3]]*HMU[[1]]+dd[[1]]*HMU[[3]])+6*(HMU[[2]]*t+HMU[[1]]*(s-
t*HMU[[1]]))-2*h4*y/h2)/h2b;
];
t=1/(deriv[[1]]*deriv[[4]]-deriv[[2]]*deriv[[3]]);
u=(deriv[[4]]*(rbet-rb1)-deriv[[2]]*(bet2-b2))*t;
```

```
y=(deriv[[1]]*(bet2-b2)-deriv[[3]]*(rbet-rb1))*t;

(*form new estimates of g and d*)
lastg=g;
lastd=d;
g=g-u(*/Log[4,m+4]*);
If[b1==0||g<0,g=0;];
d=d-y(*/Log[4,m+4]*);
)];
(*end of iteration*)
delta=d;
xlam=sigma/Sqrt[h2];
If[neg,(gamma=-g;HMU[[1]]=1-HMU[[1]];),(gamma=g;)];
xi=xbar-xlam*HMU[[1]];
Return[N[{gamma,delta,xi,xlam}]];
)]
```

References

Abbas, Ali E., David V. Budescu, Hsiu-Ting Yu and Ryan Haggerty. 2008. A comparison of two probability encoding methods: Fixed probability vs. Fixed variable values. *Decision Analysis* **5**(4) 190-202.

Alpert, Marc and Howard Raiffa. 1969, A progress report on the training of probability assessors.

Alpert, Marc and Howard Raiffa. 1982. A progress report on the training of probability assessors. In: Daniel Kahneman, Paul Slovic and Amos Tversky (eds.), *Judgement under uncertainty: Heuristics and biases*. Camgridge University Press, Cambridge, England, pp. 294-305.

Anderson, Theodore W. 1962. On the distribution of the two-sample cramer-von mises criterion. *The Annals of Mathematical Statistics* **33**(3) 847-1226.

Bailey, David E., Jules J. Loos, Elgin S. Perry and Robert J. Wood. 2000. A retrospective evaluation of 316(b) mitigation options using a decision analysis framework. *Environmental Science and Policy* **3** S25-S36.

Bailey, Michael J., John Snapp, Subramani Yetur, Jeffrey S. Stonebreaker, Steven A. Edwards, Aaron Davis and Robert Cox. 2011. Practice summaries: American airlines uses should-cost modeling to assess the uncertainty of bids for its full-truckload shipment routes. *Interfaces* **41**(2) 194-196.

Bickel, J. Eric, Larry W. Lake and John Lehman. 2011. Discretization, simulation, and swanson's (inaccurate) mean. *SPE Economics and Management* **3**(3) 128-140.

Birge, John R. and François Louveaux. 1997. *Introduction to stochastic programming*. Springer-Verlag, New York, New York.

Brooks, Daniel G. and Craig W. Kirkwood. 1988. Decision analysis to select a microcomputer networking strategy: A procedure and a case study. *The Journal of the Operational Research Society* **39**(1) 23-32.

Burr, Irving. 1973. Parameters for a general system of distributions to match a grid of a_3 and a_4. *Communications in Statistics* **2**(1) 1-21.

Butterworth, N. J. 1987. A fast and flexible family of distributions for simulation studies. *IIE Transactions* **19**(4) 439-444.

Capen, E. C. 1976. The difficulty of assessing uncertainty. *Journal of Petroleum Technology* **28**(8) 843-850.

Clemen, Robert T. 1991. *Making hard decisions: An introduction to decision analysis*. PWS-KENT Publishing Company, Boston, Massachusetts.

Clemen, Robert T. and Terence Reilly. 1999. Correlations and copulas for decision and risk analysis. *Management Science* **45**(2) 208-224.

Craig, Cecil C. 1936. A new exposition and chart for the pearson system of frequency curves. *The Annals of Mathematical Statistics* **7**(1) 16-28.

D'Errico, John R. and Nicholas Zaino, A. Jr. 1988. Statistical tolerancing using a modification of taguchi's method. *Technometrics* **30**(4) 397-405.

Darling, Donald A. 1957. The kolmogorov-smirnov, cramer-von mises tests. *The Annals of Mathematical Statistics* **28**(4) 823-1098.

Draper, J. 1952. Properties of distributions resulting from certain simple transformations of the normal distribution. *Biometrika* **39**(3/4) 290-301.

Elderton, William Palin and Norman Lloyd Johnson. 1969. *Systems of frequency curves*. Cambridge University Press, Cambridge, UK.

Farnum, N. R. and L. W. Stanton. 1987. Some results concerning the estimation of beta distribution parameters in pert. *Journal of the Operational Research Society* **38**(3) 287-290.

Garthwaite, Paul H. and Anthony O'Hagan. 2000. Quantifying expert opinion in the uk water industry: An experimental study. *Journal of the Royal Statistical Society. Series D (The Statistician)* **49**(4) 455-477.

Golenko-Ginzburg, D. 1988. On the distribution of activity time in pert. *Journal of the Operational Research Society* **39**(8) 767-771.

Hamming, R. W. 1973. *Numerical methods for scientists and engineers*. McGraw Hill, Inc., New York.

Hammond, Robert and J. Eric Bickel. 2013a. Reexamining discrete approximations to continuous distributions. *Decision Analysis* **10**(1) 6-25.

Hammond, Robert and J. Eric Bickel. 2013b. Approximating continuous probability distributions using the 10th, 50th, and 90th percentiles. *The Engineering Economist* **58**(3) 189-208.

Hampton, J. M., P. G. Moore and H. Thomas. 1973. Subjective probability and its measurement. *Journal of the Royal Statistical Society. Series A (General)* **136**(1) 21-42.

Hill, I.D., R. Hill and R. L. Holder. 1976. Algorithm as 99: Fitting johnson curves by moments. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **25**(2) 180-189.

Holloway, Charles A. 1979. *Decision making under uncertainty: Models and choices*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Howard, Ronald A. 1971. Proximal decision analysis. *Management Science* **17**(9) 507-541.

Howard, Ronald A. 1989. The science of decision making. In: Ronald A. Howard and James E. Matheson (eds.), *Readings on the principles and applications of decision analysis*. Strategic Decisions Group, Menlo Park, California, pp. 159-176.

Høyland, Kjetil and Stein W. Wallace. 2001. Generating scenario trees for multistage decision problems. *Management Science* **47**(2) 295-307.

Huang, C. C., W. T. Ziemba and A. Ben-Tal. 1977. Bounds on the expectation of a convex function of a random variable: With applications to stochastic programming. *Operations Research* **25**(2) 315-325.

Hurst, A., G. C. Brown and R. I. Swanson. 2000. Swanson's 30-40-30 rule. *AAPG Bulletin* **84**(12) 1883-1891.

Jahn, Frank, Mark Cook and Mark Graham. 2008. *Hydrocarbon exploration and production*. Elsevier, Amsterdam, The Netherlands.

Johnson, D. 1998. The robustness of mean and variance approximations in risk analysis. *The Journal of the Operational Research Society* **49**(3) 253-262.

Johnson, N. L. 1949. Systems of frequency curves generated by methods of translation. *Biometrika* **36**(1/2) 149-176.

Kall, Peter, A. Ruszczy'nski and K. Frauendorfer. 1988. Approximation techniques in stochastic programming. In: Yuri Ermoliev and Roger Wets (eds.), *Numerical techniques for stochastic optimization*. Springer-Verlag, Heidelberg, Germany, pp. 33-64.

Karlin, Samuel and William J. Studden. 1966. *Tchebysheff systems: With applications in analysis and statistics*. Interscience, New York.

Kaut, Michal and Stein W. Wallace. 2007. Evaluation of scenario-generation methods for stochastic programming. *Pacific Journal of Optimization* **3**(2) 257-271.

Keefer, Donald L. 1994. Certainty equivalents for three-point discrete-distribution approximations. *Management Science* **40**(6) 760-773.

Keefer, Donald L. 1995. Facilities evaluation under uncertainty: Pricing a refinery. *Interfaces* **25**(6) 57-66.

Keefer, Donald L. and Samuel E. Bodily. 1983. Three-point approximation for continuous random variables. *Management Science* **29**(5) 595-609.

Keefer, Donald L. and William A. Verdini. 1993. Better estimation of pert activity time parameters. *Management Science* **39**(9) 1086-1091.

Keeney, Ralph L. 1987. An analysis of the portfolio of sites to characterize for selecting a nuclear repository. *Risk Analysis* **7**(2) 195-218.

Kim, Seong Dae, Robert Hammond and J. Eric Bickel. 2014. Improved mean and variance estimating formulas for pert analyses. *IEEE Transactions on Engineering Management* **61**(2) 362-369.

Kolari, James, Thomas H. McInish and Erwin M. Saniga. 1989. A note on the distribution types of financial ratios in the commercial banking industry. *Journal of Banking and Finance* **13** 463-471.

Kotsiantis, Sotiris and Dimitris Kanellopoulos. 2006. Discretization techniques: A recent survey. *GESTS International Transactions on Computer Science and Engineering* **32**(1) 47-58.

Lau, Hon-Shiang and Amy Hing-Ling Lau. 1998. An improved pert-type formula for standard deviation. *IIE Transactions* **30**(3) 273-275.

Lau, Hon-Shiang, Amy Hing-Ling Lau and Donald W. Gribbin. 1995. On modeling cross sectional distributions of financial ratios. *Journal of Business Finance and Accounting* **22**(4) 521-549.

Lau, Hon-Shiang, Amy Hing-Ling Lau and Chrwan-Jyh Ho. 1998. Improved moment-estimation formulas using more than three subjective fractiles. *Management Science* **44**(3) 346-351.

Lichtenstein, Sarah, Baruch Fischhoff and L. D. Phillips. 1982. Calibration of probabilities: The state of the art to 1980. In: Daniel Kahneman, Paul Slovic and Amos Tversky (eds.), *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press, Cambridge, UK, pp. 306-334.

Lind, N. C. 1982. Modelling of uncertainty in discrete dynamical systems. *Applied Mathematical Modeling* **7** 146-152.

MacCrimmon, K. R. and C. A. Ryavec. 1964. An analytical study of the pert assumptions. *Operations Research* **12**(1) 16-37.

Malcolm, D. G., J. H. Roseboom, C. E. Clark and W. Fazar. 1959. Application of a technique for research and development program evaluation. *Operations Research* **7**(5) 646-669.

Matheson, James E. and Ronald A. Howard. 1968. An introduction to decision analysis. In: Ronald A. Howard and James E. Matheson (eds.), *Readings on the principles and applications of decision analysis*. Strategic Decisions Group, pp. 17-55.

McNamee, Peter and John Celona. 1990. *Decision analysis with supertree*. The Scientific Press, South San Fransisco, California.

McNamee, Peter and John Celona. 1991. *Decision analysis with supertree: Instructor's manual*. The Scientific Press, South San Francisco, California.

Megill, R. E. 1984. *An introduction to risk analysis*. PennWell, Tulsa, Oklahoma.

Merkhofer, Miley W. 1987. Quantifying judgmental uncertainty: Methodology, experiences, and insights. *IEEE Transactions on Systems, Man, and Cybernetics* **17**(5) 741-752.

Merkhofer, Miley Wesson, 1975. Flexibility and decision analysis, Engineering-Economic Systems, Stanford University.

Miller, Allen C. and Thomas R. Rice. 1983. Discrete approximations of probability distributions. *Management Science* **29**(3) 352-362.

Moder, J. J. and E. G. Rodgers 1968. Judgment estimates of the moments of pert type distributions. *Management Science* **15**(2) B76-B83.

Murphy, Allan H. and Robert L. Winkler. 1974. Subjective probability forecasting experiments in meteorology: Some preliminary results. *Bulletin of the American Meteorological Society* **55**(10) 1206-1216.

Murphy, Allan H. and Robert L. Winkler. 1977. The use of credible intervals in temperature forecasting: Some experimental results. In: Helmut Jungermann and Gerard De Zeeuw (eds.), *Decision making and change in human affairs*. Springer Netherlands.

Ord, J. K. 1972. *Families of frequency distributions*. Charles Griffin & Company Ltd., London.

Pearson, E. S. and J. W. Tukey. 1965. Approximate means and standard deviations based on distances between percentage points of frequency curves. *Biometrika* **52**(3/4) 553-546.

Pearson, Karl. 1895. Contributions to the mathematical theory of evolution. Ii. Skew variation in homogenous material. *Philosophical Transactions of the Royal Society of London* **186** 343-414.

Pearson, Karl. 1901. Mathematical contributions to the theory of evolution. X. Supplement to a memoir on skew variation. *Philosophical Transactions of the Royal Society of London* **197** 443-459.

Pearson, Karl. 1916. Mathematical contributions to the theory of evolution. Xix. Second supplement to a memoir on skew variation. *Philosophical Transactions of the Royal Society of London* **216**(429-457) 429.

Perry, C. and I. D. Greig. 1975. Estimating the mean and variance of subjective distributions in pert and decision analysis. *Management Science* **21**(12) 1477-1480.

Pflug, G. C. 2001. Scenario tree generation for multiperiod financial optimization by optimal discretization. *Mathematical Programming* **89**(2) 251-271.

Pratt, John W. 1964. Risk aversion in the small and in the large. *Econometrica* **32**(1/2) 122-136.

Ramberg, John S. and Bruce W. Schmeiser. 1974. An approximate method for generating asymetric random variables. *Communications of the ACM* **17**(2) 78-82.

Rhind, A. 1909. Tables to facilitate the computation of the probable errors of the chief constants of skew frequency distributions. *Biometrika* **7**(1) 127-147.

Rose, Peter. 2001a. *Risk analysis and management of petroleum exploration ventures*. American Association of Petroleum Geologists, Tulsa, OK.

Rose, Peter R. 2001b. *Risk analysis and management of petroleum exploration ventures*, vol. 12. The American Association of Petroleum Geologists, Tulsa, Oklahoma.

Ross, James G. 2011. Petroleum resources definitions, classification, and categorization guidelines, *Guidelines for application of the petroleum resources management system*, pp. 7-22.

Schaefer, Ralf E. and Katrin Borcherding. 1972. The assessment of subjective probability distributions: A training experiment. *Acta Psychologica* **37** 117-129.

Schlaifer, Robert. 1969. *Analysis of decisions under uncertainty*. McGraw-Hill, New York,.

Schmeiser, Bruce W. and S. J. Deutsch. 1974. A versatile four parameter family of probability distributions suitable for simulation. *AIIE Transactions* **9**(2) 176-181.

Schuenemeyer, John H. 2002. A framework for expert judgment to assess oil and gas resources. *Natural Resources Research* **11**(2) 97-107.

Scott, David W. 1979. On optimal and data-based histograms. *Biometrika* **66** 605-610.

Scott, David W. 2010. Scott's rule. *Wiley Interdisciplinary Reviews: Computational Statistics* **2** 497-502.

Scott, David W. 2011. Sturges' and scott's rules. In: M. Lovric (ed.), *International encyclopedia of statistical science*. Springer-Verlag, Berlin, pp. 1563-1566.

Scott, David W. and Warren R. Scott. 2008. Smoothed histograms for frequency data on irregular intervals. *The American Statistician* **62**(3) 256-261.

Selvidge, J. E. 1975. Experimental comparison of different methods for assessing the extremes of probability distributions by the fractile method, Graduate School of Business Administration, University of Colorado, Boulder, Colorado.

Shankar, N. R., K. S. N. Rao and V. Sireesha 2010. Estimating the mean and variance of activity duration in pert. *International Mathematica Forum* **5**(18) 861-868.

Shankar, N. R. and V. Sireesha. 2009. An approximation for the activity duration distribution, supporting original pert. *Applied Mathematical Sciences* **3**(57) 2823-2834.

Smith, James E., 1990. Moment methods for decision analysis, Engineering-Economic System, Stanford University, Stanford, CA.

Smith, James E. 1993. Moment methods for decision analysis. *Management Science* **39**(3) 340-358.

Smith, James E. 1995. Generalized chebyshev inequalities: Theory and applications in decision analysis. *Operations Research* **43**(5) 807-825.

Spetzler, Carl S. and Carl-Axel S. Stael Von Holstein. 1975. Probability encoding in decision analysis. *Management Science* **22**(3) 340-358.

Stonebraker, Jeffrey S. 2002. How bayer makes decisions to develop new drugs. *Interfaces* **32**(6) 77-90.

Stonebraker, Jeffrey S. and Donald L. Keefer. 2009. Modeling potential demand for supply-constrained drugs: A new hemophilia drug at bayer biological products. *Operations Research* **57**(1) 19-31.

Stroud, A. H. 1974. *Numerical quadrature and solution of ordinary differential equations*. Springer-Verlag, New York, New York.

Stroud, A. H. and Don Secrest. 1966. *Gaussian quadrature formulas*. Prentice-Hall Inc., Englewood Cliffs, New Jersey.

Sullivan, Michael A. 2000. Valuing american put options using gaussian quadrature. *The Review of Financial Studies* **13**(1) 75-94.

Taguchi, G. 1978. Performance analysis design. *International Journal of Production Research* **16**(6) 521-530.

Tauchen, George and Robert Hussey. 1991. Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models. *Econometrica* **59**(2) 371-396.

Tomassini, Lawrence A., Ira Solomon, Marshall B. Romney and Jack L. Krogstad. 1982. Calibration of auditors' probabilistic judgements: Some empirical evidence. *Organizational Behavior and Human Performance* **30** 391-406.

Tversky, Amos and Daniel Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* **185**(4157) 1124-1131.

Villani, Cédric. 2009. *Optimal transport: Old and new*. Springer-Verlag, Berlin.

Wallsten, Thomas S. and David V. Budescu. 1983. Encoding subjective probabilities: A psychological and psychometric review. *Management Science* **29**(2) 151-173.

Wang, Tianyang and James S. Dyer. 2012. A copulas-based approach to modeling dependence in decision trees. *Operations Research* **60**(1) 225-242.

Willigers, B. J. A. 2009, Practical portfolio simulation: Determining the precision of a probability distribution of a large asset portfolio when the underlying project economics are captured by a small number of discrete scenarios, SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers, New Orleans, LA.

Winkler, Robert L. and Allan H. Murphy. 1968. "Good" probability assessors. *Journal of Applied Meteorology* **7** 751-758.

Zaino, Nicholas A. and John D'Errico. 1989. Optimal discrete approximations for continuous outcomes with applications in decision and risk analysis. *The Journal of the Operational Research Society* **40**(4) 379-388.

Zelen, Marvin. 1954. Bounds on a distribution function that are functions of moments to order four. *Journal of Research of the National Bureau of Standards* **53**(6) 377-381.