DECISION ANALYSIS
Vol. 8, No. 3, September 2011, pp. 206–219
ISSN 1545-8490 | EISSN 1545-8504 | 11 | 0803 | 0206

http://dx.doi.org/10.1287/deca.1110.0213

Quantile-Parameterized Distributions

Thomas W. Keelin
Keelin Reeds Partners, Menlo Park, California 94025, tomk@keelinreeds.com

Bradford W. Powley
Department of Management Science and Engineering, School of Engineering,
Stanford University, Stanford, California 94305, bpowley@stanford.edu

This paper introduces a new class of continuous probability distributions that are flexible enough to represent a wide range of uncertainties such as those that commonly arise in business, technology, and science. In many such cases, the nature of the uncertainty is more naturally characterized by quantiles than by parameters of familiar continuous probability distributions. In the practice of decision analysis, it is common to fit a hand-drawn curve to quantile outputs from probability elicitations on a continuous uncertain quantity and to then discretize the curve. The resulting discrete probability distribution is an approximation that cuts off the distribution's tails and eliminates intermediate values. Quantile-parameterized distributions address this problem by using quantiles themselves to parameterize a continuous probability distribution. We define quantile-parameterized distributions, illustrate their flexibility and range of applicability, and conclude with practical considerations when parameterizing distributions using inconsistent quantile assessments.

Key words: continuous probability distribution; probability encoding; decision analysis; quantile function; inverse cumulative distribution function; basis function
History: Received on September 17, 2010. Accepted on May 24, 2011, after 1 revision. Published online in *Articles in Advance* August 4, 2011.

1. Motivation

There exists a gap in the professional practice of decision analysis: probability distributions capable of representing a broad range of continuous uncertain quantities, especially those that are not well characterized by a simple underlying process. Probability distributions that describe underlying physical processes are unequipped to effectively represent such uncertainties precisely because these distributions are limited by their process interpretations. In contrast, we introduce a more flexible class of probability distributions that take quantiles as their parameters, are continuously differentiable, and are computationally convenient to simulate.

As one application, when modeling a decision, a decision analyst may elicit an expert's knowledge on a continuous uncertain quantity as a finite set of quantiles over a prescribed set of probabilities (points on a cumulative distribution function (CDF)). One can use a quantile-parameterized distribution (QPD) to provide instant feedback on the shape of a distribution consistent with these quantiles and facilitate rapid convergence to a probabilistic representation that the decision maker declares appropriate.

When many discrete points (e.g., as a result of a probabilistic simulation or other data gathering) are the best information a decision maker has for characterizing a continuous probability distribution, he can use QPDs to represent that discrete information with a smooth probability distribution. In practice, the use of QPDs facilitates probability assessments, enables modeling of a decision maker's probabilistic information with greater fidelity, and provides an improved method for communicating and visualizing probabilistic information.

2. Probability Encoding Methodologies

Howard (1988) specifies the basis of any decision by the decision maker's alternatives, information, and preferences. This article focuses on the informational component of decision making: what the decision maker knows. In this matter, we take a Bayesian approach. The Bayesian view of probability asserts that an individual's knowledge about an uncertainty can be quantified by a probability distribution. Once a decision maker expresses his knowledge about a specific decision numerically (and identifies his alternatives and preferences), a decision analyst can apply a formal process to determine his best alternative.

As decision analysis practice matured, a practical question arose: How can one best elicit an expert's probability distribution on a continuous uncertain quantity? It was during that time period when Tversky and Kahneman (1974) described various cognitive biases that apply to decision making. From this confluence of decision analysis and behavioral decision theory, Spetzler and Staël von Holstein (1975) detailed a process for translating an expert's knowledge about an uncertainty into a CDF by way of a series of gambles. A probability encoder asks these questions in a sequence designed specifically to guide the expert away from cognitive biases. To reduce motivational biases, Matheson and Winkler (1976) and José and Winkler (2009) introduced scoring rules for continuous distributions that incent the expert to truthfully respond to the questions. The output of the Spetzler and Staël von Holstein (1975) method is a finite sequence of quantiles and their associated cumulative probabilities. They note that their probability encoding procedure can result in quantile/probability pairs that are inconsistent. In such a case the probability encoder can ask further questions of the expert until his answers are consistent; alternatively, given a set of inconsistent quantile/probability pairs, one can construct a continuous distribution that is consistent with the given information. For example, Abbas (2005) details a method for maximizing entropy between upper and lower bounds of CDFs.

Decision analysis, as a formal discipline, is more than 40 years old (Howard 1966). In that time, computational power has increased dramatically. The earliest decision analysis software computed certain equivalents of uncertain deals by solving discrete decision trees. This is one possible reason why many methods exist for the discretization of continuous CDFs.[1] Various approaches exist for transforming such sets of assessed coordinates into a usable probability distribution. One common approach is first to apply a hand-drawn smooth curve through the assessed points. The next step is to use an algorithm that chooses discrete points on the value axis based on the smooth curve. Abt et al. (1979) propose the bracket-mean method, a practice used by the decision analysis group at SRI International as far back as the early 1970s.[2] The bracket-mean method discretizes the cumulative probability axis into n brackets. Within each bracket, one chooses a value so that the area to the left of the value and below the CDF equals the area to the right of the value and above the CDF. This method determines conditional means over the support of each bracketed conditional probability. Smith (1993) also mentions the bracket-median method, which is similar in concept except that one discretizes the distribution using the conditional median within each bracket rather than the conditional mean. Keefer and Bodily (1983) introduce the extended Pearson Tukey (eP-T) method. This approach builds on the work of Pearson and Tukey (1965) to estimate the first and second moments of a continuous probability distribution with a probability density function (PDF) having three points of support. It uses the 0.05/0.50/0.95 quantiles and applies probabilities of 0.185/0.630/0.185. Although this method is ad hoc, Reilly (2002) shows it to be empirically robust at approximating the first five moments of some familiar probability distributions.

A second approach is one of maximum entropy distributions. This approach is attractive from a normative sense in that it strives to add little or no information beyond that which the data give. The maximum entropy distribution for a set of quantile/probability data has a piecewise linear CDF and a stair-step PDF. Abbas (2003) calls this type of distribution the fractile maximum entropy distribution (FMED). This distribution adds no information beyond the quantile/probability data themselves. Abbas introduces another maximum entropy distribution that he names as the midpoint maximum entropy distribution. This distribution makes the assumption that the PDF will cross each interval of an FMED at

[1] We note that discrete approximations remain useful for many applications, including the assessment of conditional probability distributions and dynamic programming.

[2] E-mail correspondence with Jim Matheson, former director of the Decision Analysis group at SRI International (June 2010).

its midpoint. The advantage of adding the heuristic midpoint element is that the resulting PDF is continuous (piecewise linear). The class of probability distributions that we propose in this paper takes a further heuristic step in that the distributions are not constructed by maximizing entropy. They retain the advantage of passing through each quantile, having an arbitrary support, and having a smooth PDF.

There are two other methods of note that use neither hand-drawn smooth curves nor entropy maximization. Miller and Rice (1983) introduce a method that strives to approximate the moments of the assessed points by twice applying a Gaussian quadrature procedure. The result is a discrete probability distribution with an arbitrary number of points of support. The other approach is to fit a set of piecewise functions to the quantile data. Many and varied fields apply such piecewise fits, including Boneva et al. (1971) in the field of statistics, Hilger and Ney (2001) in signal processing, and Korn et al. (1999) in data mining. In a direct application for decision analysis, Runde (1997) fits a C^2 Hermite tension spline through the assessed quantiles. This smooth, piecewise curve is a probability distribution; therefore, one can discretize it via one of the aforementioned methods or one can sample from it (as from any piecewise functional fit that satisfies the axioms of probability) via probabilistic simulation. The approach detailed in this paper most closely resembles the last method except that instead of a piecewise function, we introduce a method for constructing a single (nonpiecewise) probability distribution.

3. Introduction to Quantile-Parameterized Distributions

We sought a probability distribution whose CDF would accurately represent a continuous probability distribution based only on an arbitrary number of quantile/probability pairs $\{(x_i, y_i) \mid i \in 1{:}n\}$. The idea for QPDs was driven by this desire and was originally developed from a simple thought: start with something good and make it better. This thought succeeds in many contexts; for example, Ye et al. (2000) describe a method for genetically engineering rice with the goal of producing golden rice, a grain designed to help the nutritional needs of populations whose diets are deficient of vitamin A. The researchers began with a staple food (rice), and engineered its genes to produce a critical nutrient (beta carotene). In the case of probability distributions, consider the normal distribution. A univariate normal distribution is a function described by two parameters, μ and σ, that can be thought of as analogs to the genes of a living organism, in the sense that modifying either of these parameters modifies the function itself. Ordinarily, μ and σ are parametric constants, but one might also choose to vary them systematically. For example, would smoothly increasing σ over the domain of the PDF yield a right-skewed distribution? Would smoothly increasing μ over the domain of the PDF yield a distribution with a fatter midsection and thinner tails? It turns out that the answer to both of these questions is yes. By starting with something good (a normal distribution), one can make it better for the representation of a broad range of uncertainties by allowing μ and σ to vary. More specifically, one can impart right-skew to a normal distribution by varying the σ parameter as an increasing function of its cumulative probability. Likewise, one can decrease the kurtosis of a normal distribution by varying the μ parameter as an increasing function of its cumulative probability. Varying μ and σ as a function of the cumulative probability of a distribution has the feature of increasing the number of parameters of the distribution while remaining scale invariant.

4. A Simple Quantile-Parameterized Distribution

We illustrate QPDs with an example that we henceforth call the *simple Q-normal*. To construct such a distribution, we take an approach similar to Kirkwood (1976) by changing the parameters of a familiar function. We begin with a normal distribution with random variable $X \sim \mathrm{N}(x; \mu, \sigma)$ and redefine its parameters μ and σ as linear functions[3] of the normal distribution's cumulative probability, $y = F(x)$.

$$\mu(y) = a_1 + a_4 y, \tag{1}$$

[3] This approach is analogous to the generalization of constant absolute risk aversion into hyperbolic absolute risk aversion, where one recasts the decision maker's risk tolerance parameter $\rho(x) = -(u''(x)/u'(x))^{-1}$ as a linear function of wealth, x.

$$\sigma(y) = a_2 + a_3 y. \tag{2}$$

The resulting random variable $X \sim \mathrm{N}(x; \mu(y), \sigma(y))$ is distributed according to a simple Q-normal distribution. Its CDF is an implicit function, and Φ represents the standard normal CDF:

$$F(x) = \Phi\left(\frac{x - (a_1 + a_4 y)}{a_2 + a_3 y}\right) \quad \text{for } x \in (-\infty, \infty). \tag{3}$$

To derive its PDF, start with the chain rule $dF/dx = (d\Phi/dz)(dz/dx)$, where $z = (x - (a_1 + a_4 y))/(a_2 + a_3 y)$, and substitute the standard normal PDF $\phi(z) = d\Phi/dz$,

$$\begin{aligned} \frac{dF}{dx} &= \phi(z) \cdot \left(\frac{1}{\sigma(y)}\frac{dx}{dx} - \frac{1}{\sigma(y)}\frac{d\mu(y)}{dx} - \frac{x - \mu(y)}{\sigma(y)^2}\frac{d\sigma(y)}{dx}\right) \\ &= \frac{\phi(z)}{\sigma(y)}\left(1 - a_4\frac{dF}{dx} - z a_3 \frac{dF}{dx}\right). \end{aligned}$$

Then gather the differential terms and substitute (2) to yield the PDF

$$f(x) = \frac{\phi(z)}{a_2 + a_3 y + \phi(z) \cdot (a_3 z + a_4)} \tag{4}$$

$$\text{given } a_2 + a_3 y + \phi(z) \cdot (a_3 z + a_4) > 0, \quad \text{for all } z \in (-\infty, \infty).$$

Like the CDF, this simple Q-normal PDF in (4) is an implicit function. However, given only the cumulative probability $y = F(x)$, one can determine the remaining variables z and $\Phi(z)$ and hence determine $f(x)$. Note that one can create a three parameter Q-normal distribution by setting any one of the parameters a_1, a_2, a_3, or a_4 equal to zero. Also note that the distribution of (3) reverts to the normal distribution when $a_3 = a_4 = 0$.

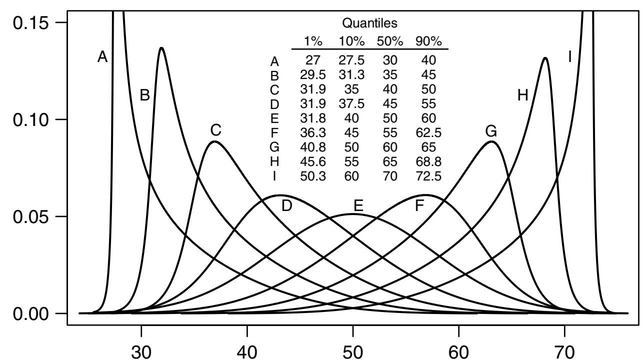
One can readily determine the constants $\{a_i \mid i \in 1{:}4\}$ from a set of four quantile/probability pairs by solving a set of four linear equations. Begin with the equation

$$z = \frac{x - (a_1 + a_4 y)}{a_2 + a_3 y}.$$

We solve for x to yield

$$x = a_1 + a_2 z + a_3 y z + a_4 y.$$

Figure 1 Some Skewed Simple Q-Normal Distributions



Choosing a cumulative probability y determines the standardized variable $z = \Phi^{-1}(y)$. Inputting its associated quantile x then leaves only the four scaling constants $\{a_i \mid i \in 1{:}4\}$ as unknown. One can express these relationships with the system of linear equations

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 & \Phi^{-1}(y_1) & y_1\Phi^{-1}(y_1) & y_1 \\ 1 & \Phi^{-1}(y_2) & y_2\Phi^{-1}(y_2) & y_2 \\ 1 & \Phi^{-1}(y_3) & y_3\Phi^{-1}(y_3) & y_3 \\ 1 & \Phi^{-1}(y_4) & y_4\Phi^{-1}(y_4) & y_4 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix}. \tag{5}$$

We denote this matrix[4] as Y, which represents a linear map $\mathbf{R}^4 \to \mathbf{R}^4$ of the quantiles x to the constants a. In effect, the simple Q-normal is fully parameterized by a set of four quantiles. To determine the constants a, rewrite (5) as $a = Y^{-1}x$. As long as Y is invertible, this method delivers a unique function for any given set of four quantile/probability pairs.

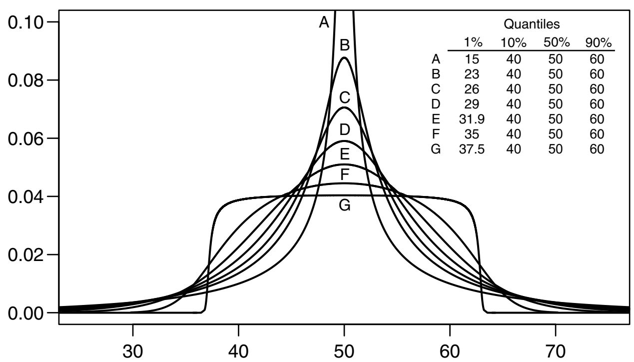
The preceding formulation reveals some key features of this simple Q-normal distribution. To begin with, it results in a wide range of probability distributions that are consistent with a diversity of quantiles, shown briefly in Figures 1 and 2.

The simple Q-normal is supported over the real number line and allows for adjustment of its shape and moments through the modulation of its quantile input parameters. This simple Q-normal can be also be fully described by its inverse CDF, which is an explicit function of y:

$$F^{-1}(y) = a_1 + a_2\Phi^{-1}(y) + a_3 y \Phi^{-1}(y) + a_4 y, \quad \text{for } y \in (0, 1). \tag{6}$$

[4] We use Y to denote this matrix because it depends only on $y_1, \ldots, y_n$ and not on $x_1, \ldots, x_n$.

Figure 2 Some Symmetric Simple Q-Normal Distributions



This makes it well suited to probabilistic simulation—a feature that we soon shall see is true of QPDs in general. Generating a simple Q-normal random variate via the inverse transformation method is as straightforward as computing a uniform(0, 1) random variate and substituting it for the variable y in (6).

We now offer a general definition of a QPD. Note that the parameter substitution method we used to derive the simple Q-normal is neither an attribute nor a requirement of this general definition.

5. Definition: Quantile-Parameterized Distribution

Let $\{g_i(y) \mid i \in 1{:}n,\ y \in (0,1)\}$ be a set of continuously differentiable and linearly independent functions of the cumulative probability y. We henceforth call these *basis functions*. Further, let $\{a_i \mid i \in 1{:}n\}$ be a set of real constants.

Definition 1. A continuous probability distribution is a QPD if and only if its inverse CDF can be written as follows:

$$F^{-1}(y) = \begin{cases} L_0 & y = 0, \\ \sum_{i=1}^{n} a_i g_i(y) & 0 < y < 1, \\ L_1 & y = 1, \end{cases} \tag{7}$$

where the constants L_0 and L_1 are the right-handed limit

$$L_0 = \lim_{y \to 0^+} F^{-1}(y) \tag{8}$$

and the left-handed limit

$$L_1 = \lim_{y \to 1^-} F^{-1}(y). \tag{9}$$

By Definition 1 the simple Q-normal is a QPD because its inverse CDF is of the form of (7), where

$$g_1(y) = 1,$$
$$g_2(y) = \Phi^{-1}(y),$$
$$g_3(y) = y\Phi^{-1}(y),$$
$$g_4(y) = y.$$

Some familiar probability distributions are also QPDs, including the normal, the exponential, the logistic, and the uniform.[5]

We include L_0 and L_1 in the construction of (7) so as not to restrict the set of allowable basis functions to those with ranges over finite intervals. For example, the basis function $g_2(y) = \Phi^{-1}(y)$ of the simple Q-normal has limits $L_0 = -\infty$ and $L_1 = +\infty$. This construction removes any restriction on the support of a QPD. That is, a QPD can be supported over any connected subset of the real numbers depending only on its basis functions and constants $a \in \mathbf{R}^n$. We now explore several properties of a QPD that are implied by its definition.

Proposition 1. *The probability density function of a QPD is given by*

$$f(x) = \left(\sum_{i=1}^{n} a_i \frac{dg_i(y)}{dy} \right)^{-1}, \tag{10}$$

where $x = F^{-1}(y)$.

Proof. Differentiate (7) with respect to y:

$$\frac{dF^{-1}(y)}{dy} = \frac{d}{dy} \sum_{i=1}^{n} a_i g_i(y).$$

Because $x = F^{-1}(y)$ by definition, and the differential operator is linear,

$$\frac{dx}{dy} = \sum_{i=1}^{n} a_i \frac{dg_i(y)}{dy}. \tag{11}$$

Taking the reciprocal of (11) yields PDF (10). □

[5] The QPD definition introduced in this paper is similar to the parametric family of distributions that Karvanen (2006) introduces for the purpose of estimating a probability distribution using L-moment statistics. In contrast, our QPD definition removes the restriction that the basis functions be quantile functions (i.e., nondecreasing) and adds the restriction that they be linearly independent and continuously differentiable.

Proposition 1 is a general equation for deriving the PDF of a specific QPD. As an example, one can derive the PDF of the simple Q-normal given in (4) by applying the Q-normal's basis functions $\{g_1(y)=1;\ g_2(y)=\Phi^{-1}(y);\ g_3(y)=y\Phi^{-1}(y);\ g_4(y)=y\}$ to (10).

Proposition 2. *The mth moment of a QPD is*

$$E[x^m]=\int_{y=0}^{1}\left(\sum_{i=1}^{n}a_i g_i(y)\right)^m dy. \quad (12)$$

Proof. The definition of the mth moment of a probability distribution $f(x)$ is

$$E[x^m]=\int_{x=-\infty}^{+\infty}x^m f(x)\,dx. \quad (13)$$

Because $y=F(x)$ and $f(x)=dF/dx$, $dy=f(x)dx$. By substituting, $dy=f(x)dx$, and $x=F^{-1}(y)$, (13) becomes

$$E[x^m]=\int_{y=0}^{1}(F^{-1}(y))^m\,dy. \quad (14)$$

Substituting (7) into (14) gives (12). □

Proposition 2 is particularly useful when computing moments of a QPD whose PDF is not an explicit function of x, a circumstance that is often the case with QPDs. In such an instance, the integral (13) is not of an explicit form. In contrast, the integral (14) is an explicit function of y.

Proposition 3. *A function of the form* (7) *characterizes a continuous probability distribution* (*and therefore a QPD*) *if and only if*

$$\sum_{i=1}^{n}a_i\frac{dg_i(y)}{dy}>0,\quad \textit{for all } y\in(0,1). \quad (15)$$

Proof. The CDF of a continuous probability distribution is increasing over its support if and only if its inverse CDF is strictly increasing in y over the interval $(0,1)$. Equation (15) is the latter condition. □

Proposition 3 is important because it gives a method to verify whether a function of the form of (7) characterizes a probability distribution. For example, one can derive the parametric constraint of the simple Q-normal given in (4) by applying the Q-normal's basis functions $\{g_1(y)=1;\ g_2(y)=\Phi^{-1}(y);\ g_3(y)=y\Phi^{-1}(y);\ g_4(y)=y\}$ to (15). The condition in (15) also serves as a feasibility constraint for any optimization formulation relating to a QPD. Henceforth, any reference to *feasibility* in relation to a QPD indicates the set of constants and/or input quantiles that make (7) an inverse CDF—one that is consistent with the axioms of probability.

Proposition 4. *A QPD's set of feasible constants* $S_a=\{a\in\mathbf{R}^n \mid \sum_{i=1}^{n}a_i(dg_i(y)/dy)>0,\ \textit{all } y\in(0,1)\}$ *is convex.*

Proof. The set S_a can be equivalently expressed as an infinite intersection of sets $\bigcap_{y\in(0,1)}S_y$, where S_y is the halfspace $\{a\in\mathbf{R}^n \mid b^T a>0\}$ and the vector $b=(dg_1(y)/dy,\ldots,dg_n(y)/dy)$. Because all halfspaces are convex sets and because any intersection of convex sets yields a convex set, S_a is a convex set. □

Proposition 4 is useful when one wishes to quickly determine whether a function of the form (7) yields a QPD. We will explore the feasibility of input quantiles in more detail when we later test the parametric limits of the simple Q-normal. Finally, because convex optimization problems require convex feasible sets, Proposition 4 directly applies to optimization problems involving QPDs.

The following theorem shows that points on the CDF can uniquely determine the constants a_i. In such cases, points on the CDF are the parameters of a QPD.

Theorem 1 (Quantile Parameters Theorem). *A set of n distinct points* $\{(x_i,y_i)\mid i\in 1{:}n\}$ *uniquely determines the constants* $\{a_i\mid i\in 1{:}n\}$ *of a QPD by the matrix equation*

$$a=Y^{-1}x, \quad (16)$$

where $a,x\in\mathbf{R}^n$, *and*

$$Y=\begin{bmatrix} g_1(y_1) & \cdots & g_n(y_1)\\ \vdots & \ddots & \vdots\\ g_1(y_n) & \cdots & g_n(y_n)\end{bmatrix} \quad (17)$$

if and only if

I. *the matrix* Y *is invertible*, *and*

II. $\sum_{i=1}^{n}a_i(dg_i(y)/dy)>0$, *for all* $y\in(0,1)$.

Proof. We begin by showing that condition I is true if and only if Equation (16) holds and that the resulting function (7) is unique to the quantile inputs $x\in\mathbf{R}^n$. Set up a system of n equations according to (7). This yields the matrix equation $x=Ya$, using the definition of Y from (17). Equation (16) holds if and only if Y is invertible. Because Y is square, it defines a one-to-one mapping of the quantiles $x \in \mathbf{R}^n$ to the constants $a \in \mathbf{R}^n$.

However, the function (7) resulting from a set of input quantiles $x \in \mathbf{R}^n$ may not characterize a probability distribution. We need to show a set of constants $a \in \mathbf{R}^n$ characterize a QPD if and only if condition II holds. This is true by Proposition 3. □

The power of the Quantile Parameters Theorem is that constants $a \in \mathbf{R}^n$ need not be assessed. Instead, points on the CDF uniquely determine these constants according to (16). One can either assess these points directly using probability elicitation methods or take them from other sources of data like scientific measurements, stock movements, or the results of a probabilistic simulation. In the latter case, one can replace the histogram display of a probabilistic simulation with a smooth QPD representation as appropriate.

Regarding condition I, because the basis functions are linearly independent, the invertibility of Y is guaranteed except in pathological cases. If such a case were to occur, a small perturbation would solve the problem. In practical applications, we have never encountered a case where Y is singular.

In contrast, it is very possible to choose a set of basis functions and points $\{(x_i, y_i) \mid i \in 1{:}n\}$ such that condition II is not satisfied. The art of constructing a QPD lies in (a) choosing a set of basis functions that is capable of representing a decision maker's uncertainty and (b) specifying the points to determine that representation. For our simple Q-normal, we will specifically derive a wide range of feasibility conditions that satisfy condition II. For QPDs with other basis functions, one can derive similar conditions.

We offer no axiomatic basis for choosing the basis functions. Their suitability is solely determined by a decision maker's declaration that his uncertainty is appropriately represented. In practice, using QPDs such as the Q-normal, we have found that this is generally the case. Nonetheless, we can offer several practical guidelines for choosing the basis functions, based on our professional experience. These guidelines derive from the observation that a QPD's inverse CDF is a linear combination of its basis functions. Using the simple Q-normal as an example, the constant a_1 is a location parameter. The constant a_2 multiplies an inverse CDF of the standard normal distribution and thus allows the QPD to be supported over the real numbers. The constant a_3, which multiplies the product of uniform and normal inverse CDFs, adds skewness. Positive and negative values for a_3 result, respectively, in right-skewed and left-skewed PDFs, as shown in Figure 1. For symmetric distributions, $a_3 = 0$. The constant a_4 multiplies a uniform distribution. Adding this function to the first two terms reduces or increases kurtosis, as shown in Figure 2, depending on whether a_4 is positive or negative.

If the support of a decision maker's probability distribution is a bounded interval, then one can substitute a bounded distribution's inverse CDF (such as the beta) for the inverse CDF of the standard normal. If one desires the support of a distribution to be bounded below and unbounded above, then one can use the inverse CDF of a lognormal distribution. One can tune the tails of a QPD both by the choice of the basis functions and by the quantile probability pairs. For example, if a distribution is supported over the real numbers, and one desires that $F^{-1}(0.999)$ take a particular value x_0, then a simple Q-normal parameterized by $(x_0, 0.999)$ may suffice. The space of probability distributions governed by QPDs may hold great possibilities for future research.

Indeed, the use of QPDs is not even limited to n quantile/probability pairs, where n is the number of basis functions. Later in this paper we will briefly explore a QPD constructed with n basis functions and m quantile/probability pairs when $m > n$. This makes the matrix $Y \in \mathbf{R}^{m\times n}$.

We now return to exploring the properties of the simple Q-normal, as one example of a useful QPD. We begin by computing its first two central moments.

6. Moments of the Simple Q-Normal

Because the PDF for the simple Q-normal is implicit, we use (14) to determine its central moments. Substituting (6) we write the formula for the mean of the simple Q-normal

$$\mathrm{E}[x^m] = \int_{y=0}^{1} (a_1 + a_2\Phi^{-1}(y) + a_3 y\Phi^{-1}(y) + a_4 y)\, dy.$$

Some further simplification yields the equation

$$\mathrm{E}[x^m] = a_1 + \frac{a_4}{2} + a_2 \int_{y=0}^{1} \Phi^{-1}(y) dy + a_3 \int_{y=0}^{1} y\Phi^{-1}(y)\, dy. \qquad (18)$$

According to (14), the first of the two remaining integrals in (18) is the mean of the standard normal distribution, which equals zero. For the second integral, we change the variable of integration from y to z and integrate by parts to yield

$$\left[-\Phi(z)\phi(z)+\sqrt{\frac{1}{16\pi}}\,\mathrm{erf}(z)\right]_{z=-\infty}^{\infty},$$

where $\mathrm{erf}(z)$ represents the error function. This quantity equals $\sqrt{1/4\pi}$, so the mean of our simple Q-normal equals

$$a_1+\frac{a_3}{\sqrt{4\pi}}+\frac{a_4}{2}. \tag{19}$$

Using the same method, the variance of the simple Q-normal is approximately

$$a_2{}^2+a_2a_3+a_3{}^2\left(\frac{1}{3}+\frac{1}{2\pi\sqrt{3}}-\frac{1}{4\pi}\right)+\frac{a_2a_4}{\sqrt{\pi}}+0.282a_3a_4+\frac{a_4{}^2}{12}, \tag{20}$$

where the constant 0.282 approximates the integral $\int_{y=0}^{1} y^2\Phi^{-1}(y)\,dy$. These first two central moments reveal some items of note. First, we see that the mean of the simple Q-normal is not a function of a_2, the constant term of (2), just as the mean of a normal distribution is not a function of its variance. Similarly, the variance of the simple Q-normal is not a function of a_1, the constant term of (1), just as the variance of a normal distribution is not a function of its mean. Recall that the simple Q-normal reduces to the normal distribution when $a_3=a_4=0$. In this case, the mean must equal a_1 and the variance must equal $a_2{}^2$, a further demonstration that (19) and (20) are consistent with (1) and (2).

7. Parameterizing the Q-Normal Using Quantiles from Familiar Probability Distributions

Suppose an expert asserts 1st, 10th, 50th, and 90th quantiles[6] consistent with an underlying familiar, henceforth *named* distribution (beta, logistic, student's t, etc.). One could use these quantiles to parameterize the simple Q-normal distribution. But how close an approximation might it be? Would the Q-normal provide a representation sufficiently accurate for practical use? To explore these questions, we take these four quantiles from some named probability distributions and use them to parameterize the simple Q-normal. We use a selection of named distributions with a diversity of distributional shapes. Next we compute the 1st, 10th, 50th, and 90th quantiles of each of these distributions and use these quantiles as parameters for a simple Q-normal distribution. We use the Kolmogorov–Smirnoff (K-S) distance (maximum y-deviation) as a measure of accuracy. We show these data in Table 1. Figure 3 gives plots of the PDF and CDF of each of the named distributions overlaid by the simple Q-normal parameterized by the quantiles. Note that for the CDF plots, it is impossible to discriminate between the named distribution and the simple Q-normal parameterized by its 1st, 10th, 50th, and 90th quantiles.

Table 1 **Deviation Between of the Simple Q-Normal and Various Named Distributions**

Named distribution	1%	10%	50%	90%	K-S distance
Beta(2, 4)	0.33	0.11	0.31	0.58	0.010
Logistic(30, 1)	25	28	30	32	0.009
Student's t(8)	−2.9	−1.4	0	1.4	0.010
Lognormal(0, 0.5)	0.31	0.53	1	1.9	0.017
Weibull(10, 5)	3.2	4.0	4.8	5.4	0.014
Normal(30, 7.8)	12	20	30	40	0

8. Range of Flexibility of the Simple Q-Normal Distribution

One can interpret the parametric limits associated with (4) in terms of two ratios: r_1 and r_2. The first gives an indication of distributional symmetry

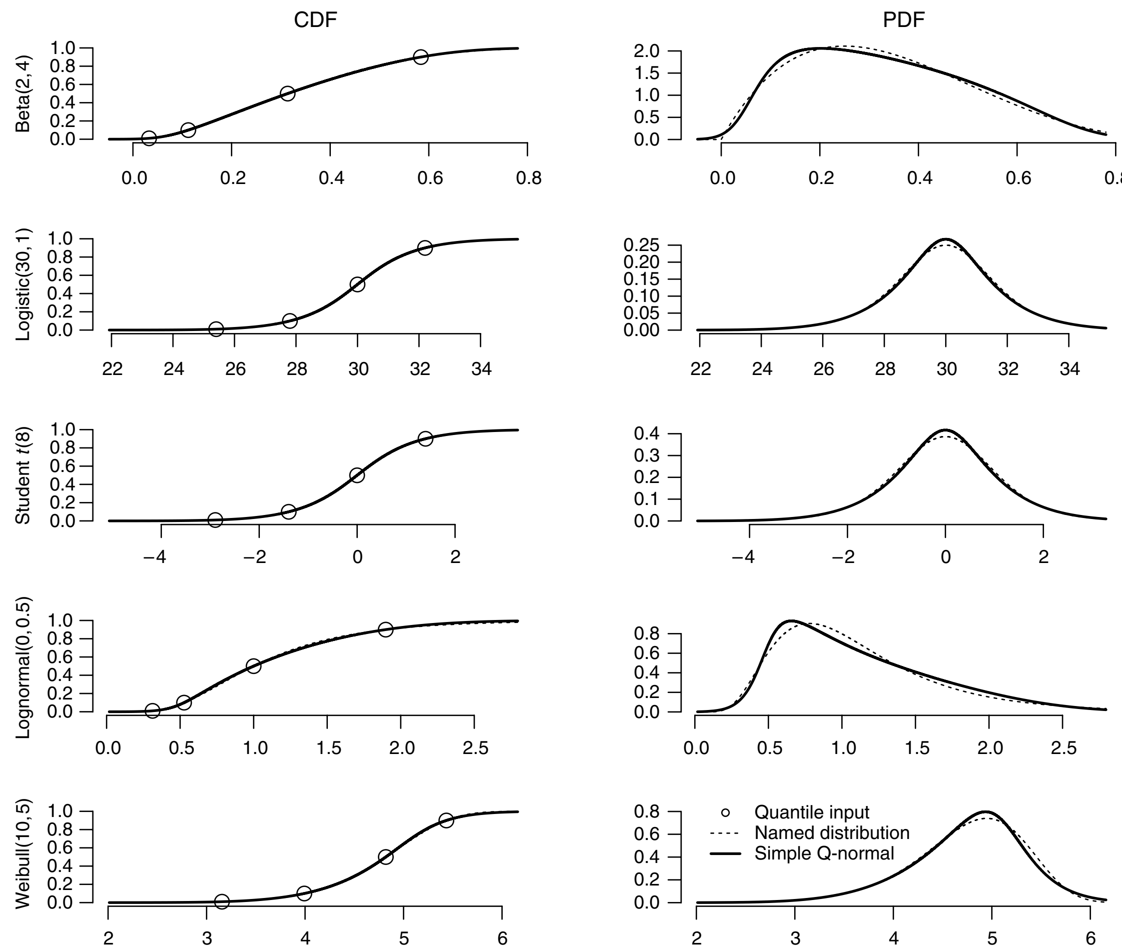
$$r_1=\frac{x_{50}-x_{10}}{x_{90}-x_{10}},$$

where x_i is the ith quantile. To give intuition, all symmetric distributions yield a value $r_1=0.5$, whereas the right-skewed exponential distribution has an r_1 equal to 0.365 regardless of the value of its rate parameter. The second ratio r_2 gives a sense of tail width

$$r_2=\frac{x_{10}-x_1}{x_{90}-x_{10}}.$$

[6] We choose these quantiles to remain consistent with the preceding demonstrations. However, the simple Q-normal is not limited to the 1st, 10th, 50th, and 90th quantiles—one could choose any four unique quantiles to use in this example.

Figure 3 **The Simple Q-Normal as Parameterized by Some Named Probability Distributions**



Note. Both the named distribution and its associated QPD pass through each of the four quantile input points.

We project the quantile vector $x \in \mathbf{R}^4$ onto the $r_1 - r_2$ plane to better visualize the limits of the simple Q-normal. Paradoxically, graphing these limits (Figure 4) clearly demonstrates the flexibility that this simple parameterization of a QPD offers. The limits are an ovoid shape. Inside the ovoid are coordinates of r_1 and r_2 that this simple Q-normal distribution can express; outside are coordinates it cannot.

Figure 5 shows how the limits of some named distributions such as the normal and exponential reveal themselves as points in the $r_1 - r_2$ plane, where the limits of other distributional forms such as the Weibull, lognormal, triangular, and student's t are curves.

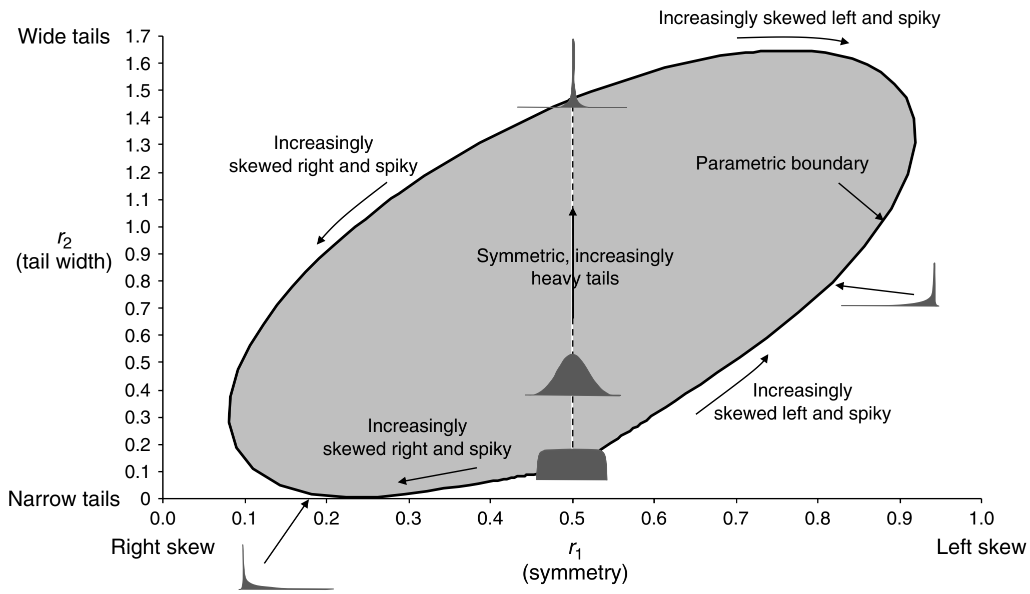
A beta distribution is a very flexible functional form able to represent a wide range of distributional shapes. Indeed its feasible region maps to an area in the $r_1 - r_2$ plane. Yet Figure 5 indicates that despite its flexibility, the beta distribution adds little to the Q-normal's territory beyond some bimodal forms. A QPD of modest functional form like the simple Q-normal demonstrates a flexibility to match quantiles that is not approached by a battery of named probability distributions. From a different perspective, the simple Q-normal has the flexibility to replace a wide range of named probability distributions in representing uncertainty.

Proposition 5. *The set of feasible quantile ratios $r = (r_1, r_2)$ for the simple Q-normal is convex.*

Proof. Let

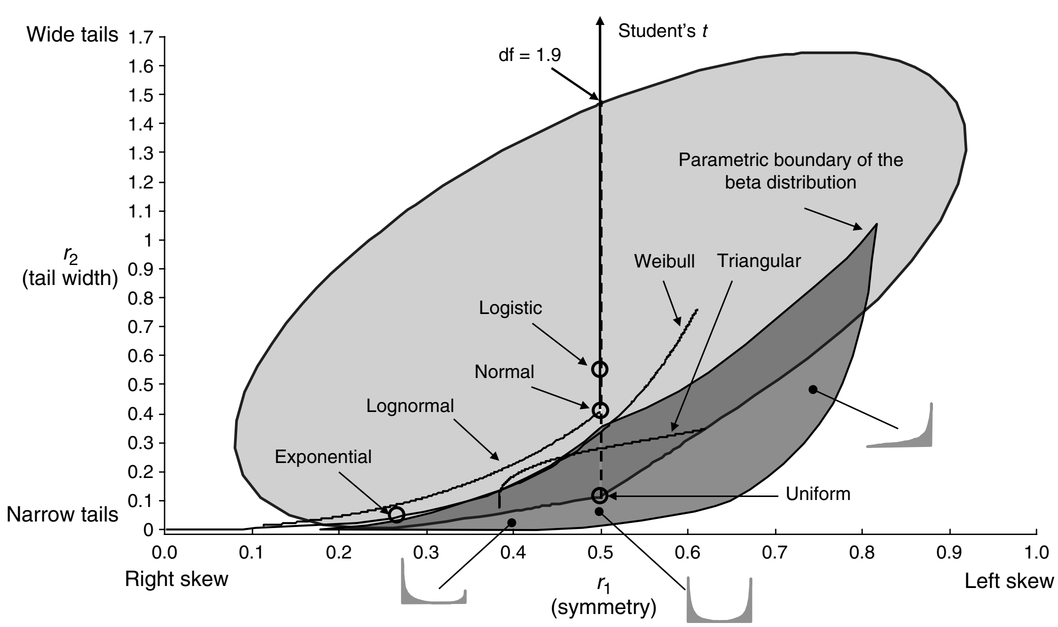
$$\psi\colon (x_1, x_{10}, x_{50}, x_{90}) \to \left(\frac{x_{50} - x_{10}}{x_{90} - x_{10}}, \frac{x_{10} - x_1}{x_{90} - x_{10}}\right)$$

Figure 4 Simple Q-Normal Parametric Limits in the $r_1 - r_2$ Plane



be the function whose image is the vector $r = (r_1, r_2)$. Let S_r be the set of feasible ratio vectors $S_r = \{r \in \mathbf{R}^2 \mid r = \psi(x),\ x \in S_x\}$, where $S_x = \{x \in \mathbf{R}^4 \mid x = Ya,\ a \in S_a\}$ is the set of quantile vectors that yields a Q-normal probability distribution and $S_a = \{a \in \mathbf{R}^4 \mid \sum_{i=1}^{n} a_i(dg_i(y)/dy) > 0,\ y \in (0,1)\}$ is the set of feasible constants. From Proposition 4, we know that S_a is convex. And because any linear transformation of a convex set is convex, it follows that S_x is also convex. Because ψ is a linear fractional function, and linear fractional functions preserve convexity, S_r is convex. □

Figure 5 Simple Q-Normal Parametric Limits and the Limits of Some Named Distributions



The convexity of the simple Q-normal's ovoid serves the practical function of facilitating quality control. Imagine a computer program that asks a user for the 1st, 10th, 50th, and 90th quantiles $x \in \mathbf{R}^4$ for a given uncertain variable. The program contains a subroutine that parameterizes the simple Q-normal with these quantiles. Will the subroutine output a vector of constants $a \in \mathbf{R}^4$ that result in a probability distribution? One might answer this quality control question by exhaustively computing the condition given in (15) using the input quantiles x over a grid of $y \in (0, 1)$ to a desired accuracy. Alternatively, one could compute and store a table of upper and lower limits of the ratio r_2 over a grid of r_1 to a desired accuracy. By the convexity of the ovoid, any input quantile vector x whose ratio vector $r = \psi(x)$ lies within a polygon formed by connecting any subset of these precomputed feasible boundary points must yield a Q-normal probability distribution. Conveniently, the convexity of the ovoid also allows the use of a bisection algorithm for solving the quasiconvex optimization problems of computing these upper and lower limits. See Boyd and Vandenberghe (2009) for a discussion on using bisection to solve quasiconvex optimization problems.

9. Parameterizing QPDs Using Overdetermined Systems of Equations

Many authors, including Wallsten and Budescu (1983) and Lindley et al. (1979), cite evidence that probability assessment data can be incoherent—a term that they use to mean that the data are inconsistent with the axioms of probability. Spetzler and Staël von Holstein (1975) acknowledge that probability encoding procedures can lead to what they term as inconsistencies in data. If one makes enough assessments such that the number of quantile/probability pairs exceeds the number of constants a_i, then a wealth of tools is available for finding a QPD that reasonably represents the incoherent data.

In other cases of overdetermined systems, as in the discrete CDF that results from probabilistic simulation, the number of data points may be far greater than the number of constants a_i. In such cases, one may use a QPD to provide a smooth representation of the data as an alternative to a histogram.

Table 2 **A Set of Inconsistent Quantile/Probability Data**

Probability	0.05	0.15	0.20	0.50	0.65	0.80	0.85	0.85
Quantile	0.0	2.5	1.5	4.0	5.0	7.0	6.0	8.0

We illustrate various methods for dealing with such overdetermined systems using the set of quantile/probability data in Table 2. It is clear that the quantile data are not monotone in probability and therefore are incoherent. Nonetheless, we can use the simple Q-normal distribution as a reasonable representation. See Figure 6 for four such examples.

Each approach computes a QPD's a vector using a variant of least squares. A total of m quantile/probability pairs and QPD with n parameters gives a matrix $Y \in \mathbf{R}^{m\times n}$. Applying the simple Q-normal to the data in Table 2 gives the following matrix:

$$Y = \begin{bmatrix} 1 & \Phi^{-1}(0.05) & 0.05\Phi^{-1}(0.05) & 0.05 \\ 1 & \Phi^{-1}(0.15) & 0.15\Phi^{-1}(0.15) & 0.15 \\ 1 & \Phi^{-1}(0.20) & 0.20\Phi^{-1}(0.20) & 0.20 \\ 1 & \Phi^{-1}(0.50) & 0.50\Phi^{-1}(0.50) & 0.50 \\ 1 & \Phi^{-1}(0.65) & 0.65\Phi^{-1}(0.65) & 0.65 \\ 1 & \Phi^{-1}(0.80) & 0.80\Phi^{-1}(0.80) & 0.80 \\ 1 & \Phi^{-1}(0.85) & 0.85\Phi^{-1}(0.85) & 0.85 \\ 1 & \Phi^{-1}(0.85) & 0.85\Phi^{-1}(0.85) & 0.85 \end{bmatrix}.$$

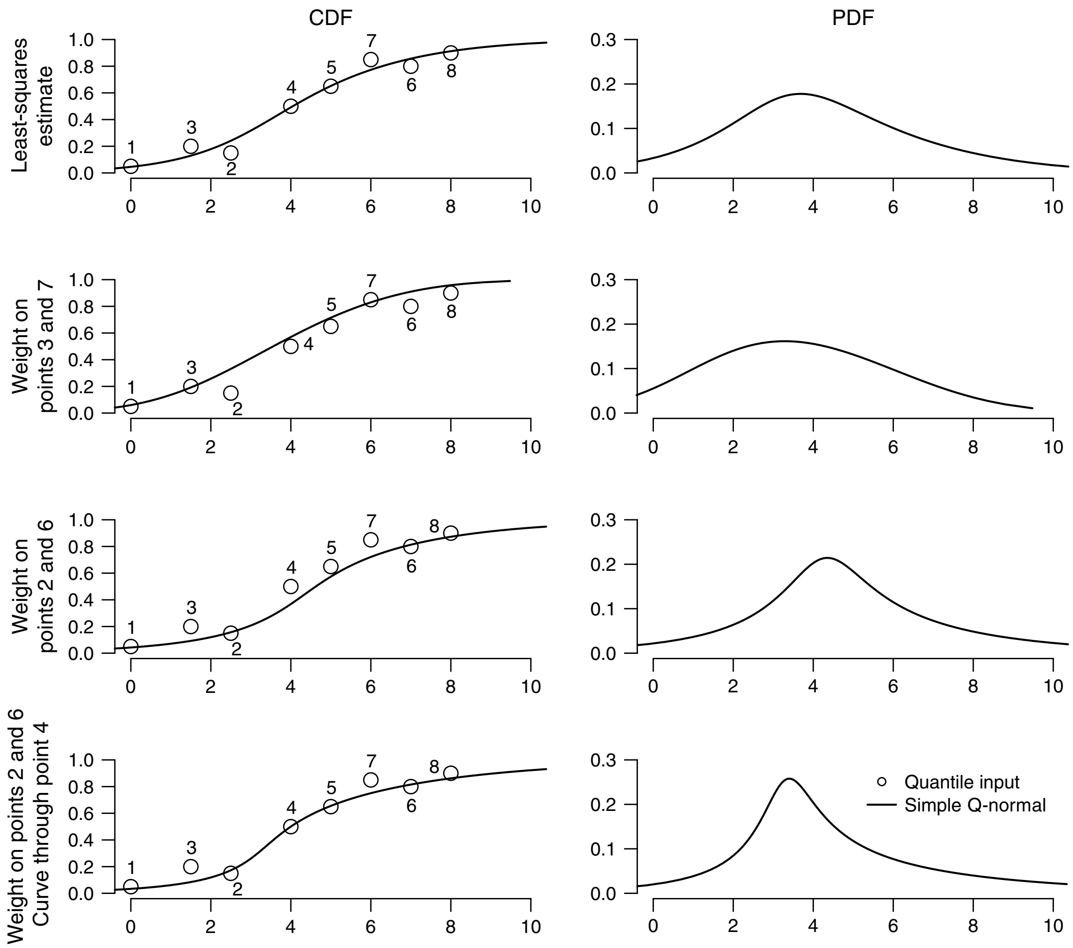
Choosing a vector of constants $a \in \mathbf{R}^n$ that minimizes the Euclidean norm of the vector of residuals $\|x - Ya\|_2$ yields the well-known closed-form equation for the least-squares approximation (providing Y is full rank):

$$a = (Y^T Y)^{-1} Y^T x,$$

where x is the vector of quantiles from Table 2, and $a \in \mathbf{R}^4$ is the vector of constants that specifies the inverse CDF of the simple Q-normal distribution. The simple Q-normal generated by least-squares approximation gives the very reasonable result shown in the plots on the first row of Figure 6.

The second and third rows of plots in Figure 6 show how one can quickly adjust the simple Q-normal from one extreme of the quantile/probability pairs to the next by applying a weighting vector to the squared

Figure 6 Various Q-Normal Approximations Derived from Incoherent Data



residuals. In the second row of plots, we apply a weighting vector to shift the curve toward points 3 and 7. In the third row, we change the weights toward points 2 and 6. The QPD vector $a \in \mathbf{R}^n$ computed from the weighted least-squares approximation is given by

$$a = (Y^T W Y)^{-1} (WY)^T x,$$

where $W \in \mathbf{R}^m$ is a diagonal matrix whose diagonal elements are given by the weighting vector. Table 3 shows the weights we used in the plots of rows two and three of Figure 6.

The fourth and final row of plots in Figure 6 is a weighted least-squares approximation (using the weighting vector from the third row) constrained so that the Q-normal passes through the median $(4, 0.5)$, which is point 4 of Figure 6. We solve for the vector $a \in \mathbf{R}^n$ with the equation

$$\begin{bmatrix} a \\ \nu \end{bmatrix} = \begin{bmatrix} 2Y^T W Y & c \\ c^T & 0 \end{bmatrix}^{-1} \begin{bmatrix} 2Y^T W x \\ 4 \end{bmatrix},$$

where ν is the Lagrange multiplier associated with the constraint on the median and $c = (1, 0, 0, 0.5)$, the vector resulting from evaluating the coefficients 1, $F^{-1}(y)$, $yF^{-1}(y)$, and y of Equation (6) at $y = 0.5$. These four methods show how readily one can parameterize the simple Q-normal to blend incoherent quantile/

Table 3 Two Weighting Vectors

Point	1	2	3	4	5	6	7	8
Weighting vector of row 2	0.05	0	0.4	0.05	0.05	0	0.4	0.05
Weighting vector of row 3	0.05	0.4	0	0.05	0.05	0.4	0	0.05

probability data. Figure 6 shows how the methods lead to very different CDFs and PDFs. The ability to make such adjustments has use in giving feedback in the probability encoding process as well as facilitating probabilistic sensitivity analysis in a decision analysis. For example, one can answer the question of whether the best alternative will change when one changes a QPD from one extreme of quantile/probability pairs to the next.

Parameterizing QPDs using overdetermined systems of equations is not limited to the quadratic penalty functions of least-squares approaches. For example, one might instead choose to minimize the sum of the absolute values of the residuals. Regardless of approach, the probability distribution resulting from any probability encoding method should pass the ultimate test of whether the decision maker declares that it reflects his beliefs.

10. Conclusion

This paper introduces a new class of probability distributions that take points on the CDF as parameters. Using the example of a simple Q-normal distribution, we demonstrate that QPDs can flexibly represent nonphysical-process based uncertainties as typically arise in business, technology, and science. For such applications, points on the CDF are the natural and intuitive parameters.

Beyond the simple Q-normal, this paper provides a theoretical foundation that enables research on other QPD formulations. In addition, we show that QPDs are well suited to probabilistic simulation because of their inverse CDF formulation, that they provide a new alternative for smooth continuous representations of histograms, and that one can use them to reasonably represent incoherent quantile/probability data.

Over the last two years, QPDs have proven their value in our decision analysis consulting practice, in which we model dozens uncertain variables on each client engagement. We now routinely use QPDs to represent continuous uncertainties instead of using the traditional three-branch, discrete-approximation methods. We have found that use of QPDs facilitates probability assessments, enables modeling of decision makers' probabilistic information (including tails) with greater fidelity, and provides an improved method for communicating and visualizing probabilistic information.

Acknowledgments

The authors' gratitude goes to Jim Matheson, Ron Howard, Ross Shachter, Michael Mischke-Reeds, John Selig, Thomas Seyller, and Ahren Lacy for helpful suggestions and to Cameron MacKenzie for his recommendations for computing the moments of the simple Q-normal distribution. Finally, the authors thank the reviewers for useful feedback.

References

Abbas, A. E. 2003. Entropy methods for univariate distributions in decision analysis. C. Williams, ed. *Bayesian Inference and Maximum Entropy Methods Sci. Engrg.: 22nd Internat. Workshop*, American Institute of Physics, Melville, NY, 339–349.

Abbas, A. E. 2005. Maximum entropy distributions between upper and lower bounds. *AIP Conf. Proc.* **803**(1) 25–42.

Abt, R., M. Borja, M. M. Menke, J. P. Pezier. 1979. The dangerous quest for certainty in market forecasting. *Long Range Planning* **12**(2) 52–62.

Boneva, L. I., D. Kendall, I. Stefanov. 1971. Spline transformations: Three new diagnostic aids for the statistical data-analyst. *J. Royal Statist. Soc., Ser. B* **33**(1) 1–71.

Boyd, S., L. Vandenberghe. 2009. *Convex Optimization*. Cambridge University Press, Cambridge, UK.

Hilger, F., H. Ney. 2001. Quantile based histogram equalization for noise robust speech recognition. *Proc. Eur. Conf. Speech Comm. Tech., Aalborg, Denmark*, 1135–1138.

Howard, R. A. 1966. Decision analysis: Applied decision theory. D. B. Hertz, J. Melese. eds. *Proc. 4th Internat. Conf. Oper. Res.*, Wiley-Interscience, New York, 55–77.

Howard, R. A. 1988. Decision analysis: Practice and promise. *Management Sci.* **34**(6) 679–695.

José, V. R., R. L. Winkler. 2009. Evaluating quantile assessments. *Oper. Res.* **57**(5) 1287–1297.

Karvanen, J. 2006. Estimation of quantile mixtures via L-moments and trimmed L-moments. *Comput. Statist. Data Anal.* **51**(2) 947–959.

Keefer, D. L., S. E. Bodily. 1983. Three-point approximations for continuous random variables. *Management Sci.* **29**(5) 595–609.

Kirkwood, C. 1976. Parametrically dependent preferences for multiattributed consequences. *Oper. Res.* **24**(1) 92–103.

Korn, F., T. Johnson, H. V. Jagadish. 1999. Range selectivity estimation for continuous attributes. *Proc. 11th Internat. Conf. Sci. Statist. Database Management*, IEEE Computer Society, Washington, DC.

Lindley, D. V., A. Tversky, R. V. Brown. 1979. On the reconciliation of probability assessments. *J. Royal Statist. Soc., Ser. A* **142**(2) 146–180.

Matheson, J. E., R. L. Winkler. 1976. Scoring rules for continuous probability distributions. *Management Sci.* **22**(10) 1087–1096.

Miller, A. C., III, T. R. Rice. 1983. Discrete approximations of probability distributions. *Management Sci.* **29**(3) 352–362.

Pearson, E. S., J. W. Tukey. 1965. Approximate means and standard deviations based on distances between percentage points of frequency curves. *Biometrika* **52**(3) 533–546.

Reilly, T. 2002. Estimating moments of subjectively assessed distributions. *Decision Sci.* **33**(1) 133–147.

Runde, A. S. 1997. Estimating distributions, moments, and discrete approximations of a continuous random variable using Hermite tension splines. Ph.D. dissertation, University of Oregon, Eugene.

Smith, J. E. 1993. Moment methods for decision analysis. *Management Sci.* **39**(3) 340–358.

Spetzler, C. S., C.-A. S. Staël von Holstein. 1975. Probability encoding in decision analysis. *Management Sci.* **22**(3) 340–358.

Tversky, A., D. Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* **185**(4157) 1124–1131.

Wallsten, T. S., D. V. Budescu. 1983. Encoding subjective probabilities: A psychological and psychometric review. *Management Sci.* **29**(2) 151–173.

Ye, X., S. Al-Babili, A. Klöti, J. Zhang, P. Lucca, P. Beyer, I. Potrykus. 2000. Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science* **287**(5451) 303–305.