QUANTILE FUNCTION METHODS FOR DECISION ANALYSIS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF MANAGEMENT
SCIENCE AND ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Bradford W. Powley
August 2013

This dissertation is online at: http://purl.stanford.edu/yn842pf8910

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Ronald Howard, Primary Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Ross Shachter

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Tom Keelin

Approved for the Stanford University Committee on Graduate Studies.

Patricia J. Gumport, Vice Provost Graduate Education

This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.

Abstract

Encoding prior probability distributions is a fundamental step in any decision analysis. A decision analyst often elicits an expert's knowledge about a continuous uncertain quantity as a set of quantile-probability pairs (points on a cumulative distribution function) and seeks a probability distribution consistent with them.

Quantile-parameterized distributions (QPDs) are continuous probability distributions characterized by quantile-probability data. This dissertation demonstrates the flexibility of QPDs to represent a wide range of distributional shapes, examines a QPD's range of parametric feasibility, and introduces various means of using QPDs when encoding relevance between uncertainties.

A decision maker may or may not believe a continuous uncertain quantity has well-defined bounds. For the former case, I offer a toolkit for engineering the support of a QPD. For the latter, I develop a theory for comparing tail heaviness between probability distributions and offer methods for engineering the tail behavior of a QPD.

I conclude with an example decision analysis: a pharmaceutical company CEO's decision whether to market or license a drug. This analysis uses QPDs to encode prior probability distributions with bounded and unbounded supports. It introduces three tools that use QPDs: data compression of multivariate probabilistic simulation, sensitivity analysis of the tail heaviness of a prior probability distribution, and valuation of probability assessment.

Acknowledgment

My time as a doctoral student was both enjoyable and rewarding. I am grateful to the people who helped make it so.

Professor Ron Howard guided my academic journey and emphasized that research should be clear and useful to academics as well as practitioners. I hope this dissertation reflects these values. Professor Ross Shachter improved the quality of my research with his eye for precision in mathematics and language. Both Ron and Ross taught me how to teach students with a balance of compassion and firmness. Tom Keelin gave me the internship where I first learned about quantile parameterized distributions—although we didn't know it at the time. He served as a mentor and research collaborator.

Community is important. I received support and encouragement from various members of the Decision Analysis Working Group and the department of Management Science and Engineering, especially Thomas Seyller, Ibrahim alMojel, Somik Raha, Jeremy Stommes, Muhammad alDawood, Wititchai Sachchamarga, Ahren Lacy, Xi Wang, Noah Burbank, Wenhao Liu, and Supakorn Mudchanatongsuk. My journey was not a lonely one.

My parents Randall and Andrea Powley cultivated my love of learning at an early age by encouraging my interests without pressuring me toward high achievement. My wife Lisa Cram graciously agreed to quit her job and suffer an interstate move so that I could pursue graduate studies. She supported me financially and emotionally, and for this I am deeply thankful.

Contents

List of Tables

List of Figures

Notation

Symbol	Description
X	a continuous uncertain quantity
x	a degree of X
$f(x)$	probability density function
$F(x)$	cumulative distribution function
p	cumulative probability
$Q(p)$	quantile function
$q(p)$	quantile density function
$g(p)$	QPD basis function
$\Phi(\cdot)$	CDF of standard normal distribution
$\phi(\cdot)$	PDF of standard normal distribution
$\sim_L$	left tail equivalence relation
$\sim_R$	right tail equivalence relation
$\prec_L$	left tail strict ordering relation
$\prec_R$	right tail strict ordering relation
$\mathrm{dom}(f)$	the domain of the function f
$\mathrm{supp}(F)$	the support of the probability distribution F
$\mathbf{R}$	the real numbers
$\mathbf{R}_{++}$	the positive real numbers
$\mathbf{R}^n$	the set of real n-vectors
$\mathbf{R}^{m\times n}$	the set of real $m \times n$ matrices

Chapter 1

Introduction

All decisions consist of three distinct elements: alternatives (what the decision maker can do); information (what the decision maker knows); and preferences (what the decision maker wants). Howard calls this formal description the *decision basis* [25, 26]. This dissertation is about the information component. More specifically, it introduces, characterizes, and develops tools for representing a decision maker's knowledge on a continuous uncertain quantity—one that can take an uncountable number of possible values. An example is the uncertain amount of time until a competitor introduces a particular product. The goal is to quantify a decision maker's knowledge in order to combine it with her preferences to evaluate her best alternative.

1.1 Probability Encoding

From the advent of the field, decision analysts have taken the Bayesian view that probability is a measure of an individual's knowledge about a particular uncertain distinction [21, 47]. In his earliest paper on the subject of decision analysis, Howard observes "...the most significant part of the [Bayesian] revolution is not Bayes's theorem or conjugate distributions but rather the concept of probability as a state of mind, a 200 year old concept." In his 1814 essay, Laplace took a Bayesian viewpoint describing probabilities as "degrees of credence" [35, page 8]. A century later, Ramsey, another Bayesian, asserted that since "it is not easy to ascribe numbers to the intensities of feelings," one should quantify degree of belief by expressing it "as the extent to which we are prepared to act on it" [49].

1.1.1 Probability and Quantile Elicitation

As the discipline of decision analysis developed in the 1960s, the encoding of prior probability distributions was noted as a fundamental step [21, page 107]. Naturally, the question of how to best elicit an expert's prior probability distribution arose. Soon, Tversky and Kahneman began documenting a classification of the various cognitive biases that they observed in decision-making experimentation, including poor performance in judging probability [65]. Concurrently, Spetzler and Staël von Holstein [61] detailed a process for translating a decision maker's knowledge about a continuous uncertain quantity into a cumulative distribution function (CDF) by way of a sequence of bets. A probability encoder asks these betting questions in a manner designed to guide the decision maker away from Tversky and Kahneman's cognitive biases. Techniques for addressing motivational biases are different from the techniques for addressing cognitive ones. Winkler and Matheson [39] introduced scoring rules that incent a decision maker to truthfully report her probabilities on a continuous uncertain quantity, and José and Winkler [31] introduced a family of scoring rules for assessing quantiles rather than probabilities. In all of these approaches, the output of the elicitation process on a continuous uncertain quantity is a finite sequence of quantiles and associated cumulative probabilities (the points shown in Figure 1.1). The nature of these data is a feature that distinguishes decision analysis from other fields of study. For example, a Bayesian statistician might begin with diffuse prior probability distributions and update them with a set of prior observations from an uncertain process [30], and a physicist might construct prior probability distributions by maximizing entropy under observed constraints such as a set of moments [29].

1.1.2 Choosing a Continuous Probability Distribution

Once the decision analyst asserts that he has sufficient quantile-probability data, he strives to find a continuous probability distribution that is consistent with the

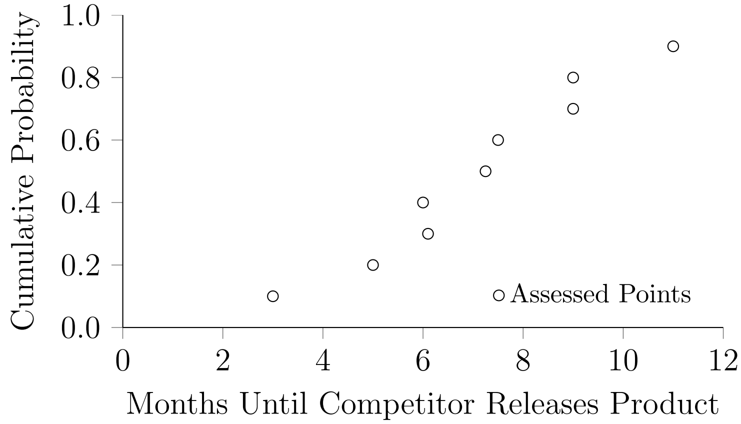

Figure 1.1: A set of quantile-probability pairs

quantile-probability data (Figure 1.2). Spetzler and Staël von Holstein note that the process of elicitation can result in quantile-probability pairs that are inconsistent (e.g., different quantiles for the same probability or pairs of points that are decreasing), in which case the probability encoder can ask further betting questions of the decision maker until her answers show consistency. Alternatively, given a set of inconsistent quantile-probability pairs, the decision analyst can construct a continuous distribution that is a reasonable representation of the given information.

In the early history of decision analysis, this probability distribution was a hand-drawn CDF [21, page 109][69, page 781][47, page 166], or perhaps a member of the canon of commonly used probability distributions like the exponential, normal, logistic, and so on. Both approaches suffice for describing continuous prior probability distributions in a decision analysis, but both have deficiencies. Hand-drawn curves lack the convenience of a concise functional description, and canonical probability distributions tend to have a small number of parameters and therefore lack flexibility in representing a wide diversity of distributional shapes.

Subsequent decision analytic research addressed these shortcomings by using quantile-probability data to parameterize alternate continuous probability distributions. Poland examines mixtures of Gaussian distributions [46] with an eye toward more flexible prior distributions for Gaussian influence diagram models [54]. In a study estimating the moments of a continuous distribution, Runde parameterizes a CDF with

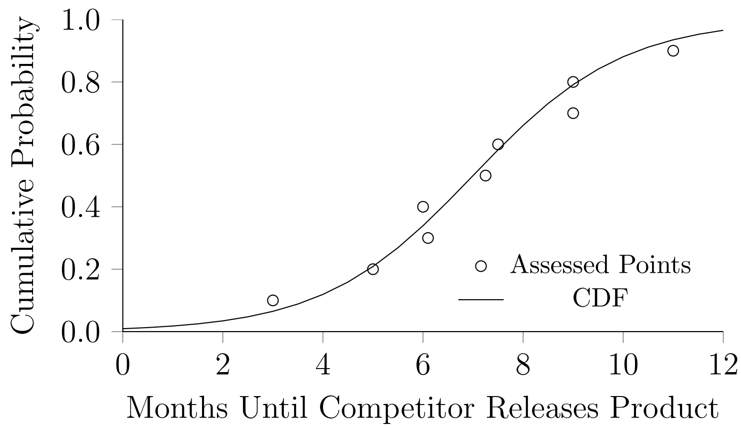

Figure 1.2: A set of quantile-probability pairs and a CDF representing them

quantile-probability data by using Hermite tension splines [51]. The tension parameters of these splines allow control over the monotonicity of the spline—a necessity for the function to be a CDF. Abbas constructs continuous probability distributions by maximizing entropy. His method strives to add no information to the distribution beyond that which the decision analyst knows [29]. When limiting the constraint set to quantile-probability pairs (and the Kolmogorov axioms), the maximum entropy distribution has a piecewise-linear CDF and a stairstep probability density function (PDF)—a function that Abbas [2, 1] calls a *fractile maximum entropy distribution* (FMED). He introduces the *midpoint maximum entropy distribution* (MMED) by beginning with the FMED and adding the restriction that the PDF must cross each interval of the FMED at its midpoint. Adding this midpoint heuristic makes the PDF of the MMED continuous.

1.1.3 Approximating the Certain Equivalent

Once a decision analyst specifies a continuous prior probability distribution within a decision model, he will likely approximate the decision maker's certain equivalent by approximating her distribution over some measure of value. As Smith writes, "Decision models involving continuous probability distributions almost always require some form of approximation" [57]. The decision analyst can approximate a decision maker's certain equivalent by discretizing one or more continuous prior probability distributions and exactly computing an approximate certain equivalent using the resulting discrete decision tree or influence diagram. Alternatively, he may sample from the continuous probability distribution via probabilistic (Monte Carlo) simulation and approximate the certain equivalent by using the simulates as probabilistic input to the value model. He can combine these two approaches by discretizing input continuous probability distributions and using Monte Carlo sampling from discrete uncertainties to approximate the certain equivalent.

The half-century existence of the formal discipline of decision analysis has seen computational power increase dramatically. Early decision analysis software computed certain equivalents by solving discrete decision trees—an approach that enabled computational tractability. Perhaps this is why there are many techniques for discretizing continuous probability distributions. One discretization technique is to first hand-draw a curve through the assessed points, and then apply an algorithm that uses the curve to choose discrete points on the value axis. Abt et al. [8] propose the *bracket-mean method*, a common practice of the decision analysis group at SRI as far back as the early 1970s.[1] The bracket-mean method is a graphical approach that preserves the mean of the original distribution. First, one divides the domain of the CDF into an arbitrary number of brackets. Next, one chooses the single value within each bracket that makes the area to the left of the value and below the CDF equal the area to the right of the value and above the CDF. This value is the conditional mean of the uncertain quantity, given the bracket. The probability assigned to the conditional mean is the cumulative probability within the bracket. Smith [58] mentions the bracket median method—similar in every way to the bracket mean method, except that one chooses the conditional median for each bracket rather than the conditional mean. Pearson and Tukey [45] introduce a method to estimate the first and second moments of a continuous distribution using a discrete approximation with three points of support. Keefer and Bodily build on this approach with their *extended Pearson-Tukey* method—one shown by Reilly [50] to be empirically robust.

[1] Email correspondence with Jim Matheson, former director of the Decision Analysis group at SRI (June 2010).

Miller and Rice [40] introduce an algorithm for an n-point discrete probability distribution that matches the first $2n-1$ moments of the continuous distribution. In the case where the decision analyst does not want to specify a continuous probability distribution (leaving moments unknown), they detail a discretization method (page 356) designed to approximate the moments of a continuous probability distribution using only the assessed quantile-probability pairs by twice applying the Gaussian quadrature prodedure. Bypassing the specification of a continuous probability distribution comes at the cost of a second approximation, and possibly a third (page 362), if the decision maker wants to improve accuracy while reducing the number of points of support. The result of both algorithms is a discrete probability distribution with an arbitrary number of points of support. It is unclear whether this algorithm will tolerate incosistencies in quantile-probability data.

1.2 Research Objectives

This dissertation focuses on the choice of continuous probability distribution detailed in §1.1.2. For decision analysis, it is desirable to have continuous probability distributions that are

- readily parameterized by quantile-probability data;
- capable of taking a variety of useful shapes;
- convenient for evaluating decision models both through direct probabilistic sampling and discretization.

While the three classes of distributions in §1.1.2 each have their own strengths, all fall short of at least one of the three desiderata. Mixtures of Gaussians display a certain lack of flexibility in distributional form—they tend to achieve skewness at the cost of multimodality and/or many basis functions, and they will always have Gaussian tails. Runde introduces Hermite tension spline CDFs as a means toward discretization, not probabilistic simulation. Therefore, his work does not discuss how convenient it is to sample a probability distribution whose CDF is a Hermite tension spline. Common probabilistic simulation methods use either a continuous probability distribution's quantile function (inverse transform sampling) or its PDF (acceptance-rejection sampling) [36]. In general, CDF splines do not have closed-form quantile functions. Also, Runde chooses to compute PDFs via numerical integration, an indication that these piecewise functions might not be expedient to represent in closed form. Abbas's FMED and MMED both lack flexibility in representing a diversity of distributional shapes. Recalling §1.1.2, the PDF of an FMED is stairstep, and that of an MMED is piecewise linear. Moreover, if the continuous uncertain quantity has one or more infinite tails, the quantile-probability pairs are insufficient data for representing its probability distribution. In order to continue with a maximum entropy formalism, the decision analyst must elicit a mean that is conditioned on the uncertain quantity exceeding the most extreme quantile datapoint for each infinite tail. In other words, the quantile-probability data itself is not sufficient to parameterize it.

In the case of a continuous uncertain quantity, the decision analyst desires a probability distribution that is both consistent with the available quantile-probability data and efficient to model. This desire motivates the three objectives of my research:

1. identify a class of probability distributions that meets the desiderata;
2. characterize this class of probability distributions;
3. demonstrate how decision analysts might use such distributions in practice.

In 2011, Tom Keelin and I addressed the first research objective by introducing *quantile-parameterized distributions* (QPDs) [34], the distributions of this dissertation. QPDs are probability distributions whose quantile function is a linear combination of basis functions.

1.3 Research Summary

The structure of this dissertation is linear—to a degree, the information in each chapter is a requirement for understanding the material in the following chapter. Chapter

2 serves as a reference chapter, giving a brief overview of quantile functions and creating distinctions used by the theory introduced in subsequent chapters. Although the results in this chapter are not new, the reader may see unfamiliar material—even if that person has a strong background in probability theory. Chapter 3 reviews quantile-parameterized distributions summarizing our original work [34]. New material in this section includes a refined interpretation of QPDs, further characterization of their parametric limits, and a section on encoding relevance (probabilistic dependence) using QPDs. The content of this chapter is a theoretical contribution that satisfies the first research objective and begins to address the second.

Chapter 4 further addresses the second research objective by building a theory and tools to engineer the support of a QPD, including constrained optimization, truncation, and transformation. Chapter 5 completes my contribution to the second research objective by building a general theory of tail behavior applicable to all probability distributions with strictly increasing, twice-differentiable quantile functions. In it, I develop tail characterization tools based on a probability distribution's quantile function. This theory culminates in some QPD-specific results that give methods for engineering the tail behavior of a QPD.

Chapter 6 addresses the third research objective, by way of an example decision analysis—the CEO of a pharmaceutical company choosing whether to license or market a drug. It introduces three QPD-based decision analytic techniques: data compression of the output of a probabilistic simulation, sensitivity analysis on tail heaviness of a prior probability distribution, and valuing probability assessment.

Chapter 2

An Overview of Quantile Functions

It is appropriate to begin a dissertation titled *Quantile Function Methods for Decision Analysis* with a brief introduction to the quantile function. This is a reference chapter, filled with definitions, propositions, and corollaries relevant to the remainder of this dissertation. I include all proofs within the text to develop the reader's intuition for describing probability distributions with quantile functions.

2.1 Some Basic Definitions

For an uncertain quantity X, a quantile function $Q(p)$ is a generalized inverse of the cumulative distribution function $F(x) \equiv \mathrm{P}\{X \leq x\}$. Like a CDF, it characterizes a probability distribution.

$$Q(p) = \inf\{x \in \mathbf{R} \mid p \leq F(x)\} \tag{2.1}$$

Many commonly used probability distributions have quantile functions that are inverse CDFs. Section 2.2 discusses the conditions under which $Q = F^{-1}$.

In order to satisfy the axioms of probability, a quantile function must meet a few criteria. First, it must be defined over the domain of $p \in (0, 1)$. For any given cumulative probability p, a quantile function $Q(p)$ must return a value, called a *quantile*. Second, a quantile function must be nondecreasing over its domain. This

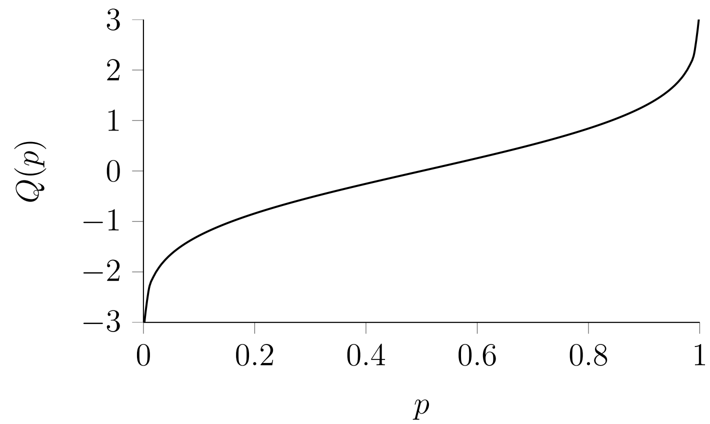

Figure 2.1: The quantile function of the standard normal distribution

requirement corresponds to a nondecreasing CDF. In other words, it guarantees that the probability on any interval $P\{a < X \leq b\} \geq 0$ for all $a, b \in \mathbf{R}$. Figure 2.1 shows the graph of a quantile function.

These two functional requirements lead to methods for constructing quantile functions using other functions as building blocks. For a clear and concise introduction to quantile functions, see Gilchrist [17, section 3.2]. I adapt some of his language and notation in this section and add some elementary proofs to condition the reader to think about probability distributions as represented by quantile functions rather than the more traditional representations like CDFs and PDFs.

Definition 1. *Given a differentiable quantile function $Q(p)$, a quantile density function (QDF) $q(p)$ is the derivative $Q'(p)$.*

Definition 2. *A p-probability density function, or pPDF, is the composition of a probability density function and its associated quantile function, $f(Q(p))$*

Tukey [64] gave the name *sparsity* to what Parzen [44] would later call the *quantile density function*. And Parzen gave the name *density quantile function* to what Gilchrist [17] would later call the *p-probability density function*. Figures 2.2 and 2.3 show the graphs of a QDF and pPDF.

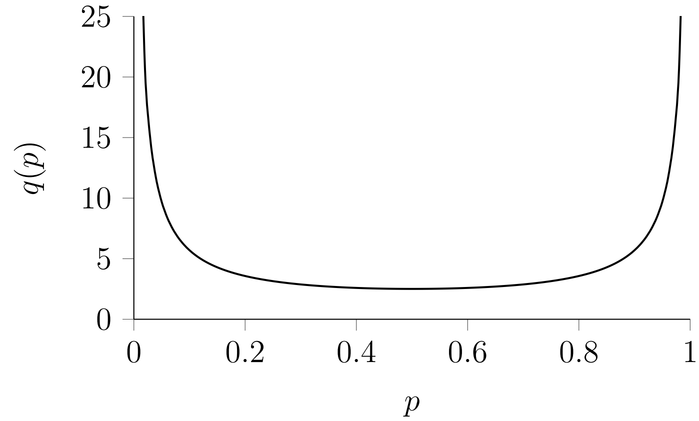

Figure 2.2: The quantile density function of the standard normal distribution

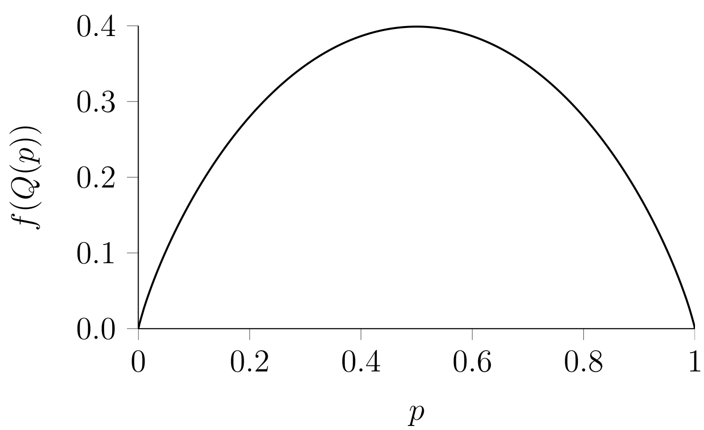

Figure 2.3: The p-probability density function of the standard normal distribution

Corollary 1. *The pPDF is the reciprocal of the QDF.*

Proof. A PDF $f(x) = \frac{dF(x)}{dx}$. Change variables $x = Q(p)$ and $p = F(x)$, and write $f(Q(p)) = \frac{dp}{dQ(p)}$. By Definition 1,

$$f(Q(p)) = \frac{1}{q(p)}. \tag{2.2}$$

□

Like CDFs, PDFs (when they are defined), and quantile functions, the QDF and its reciprocal the pPDF (when they are defined), completely characterize a probability distribution. Both the QDF and pPDF are particularly useful when modeling with QPDs.

2.2 The Continuity of Probability Distributions and Quantile Functions

A continuous probability distribution is one whose CDF is a continuous function. However, a continuous probability distribution does not imply a continuous quantile function, nor does a continuous quantile function imply a continuous probability distribution. When both CDF and quantile function are continuous, $Q = F^{-1}$, so that $F(Q(p)) = p$ and $Q(F(x)) = x$.

Figure 2.4 shows a continuous quantile function whose probability distribution is mixed discrete and continuous. The quantile associated with the flat region of the quantile function is the quantile with probability mass.

Proposition 1. *A probability distribution F is continuous if and only if its quantile function $Q(p)$ is strictly increasing.*

Proof. First, assume that Q is not strictly increasing so that $Q(p_1) = Q(p_2)$ for some $p_1, p_2 \in (0,1),\ p_1 < p_2$. Then there exists a point mass of probability $\mathrm{P}\{X = x_0\} \geq p_2 - p_1 > 0$, implying that F is not a continuous function, a contradiction.

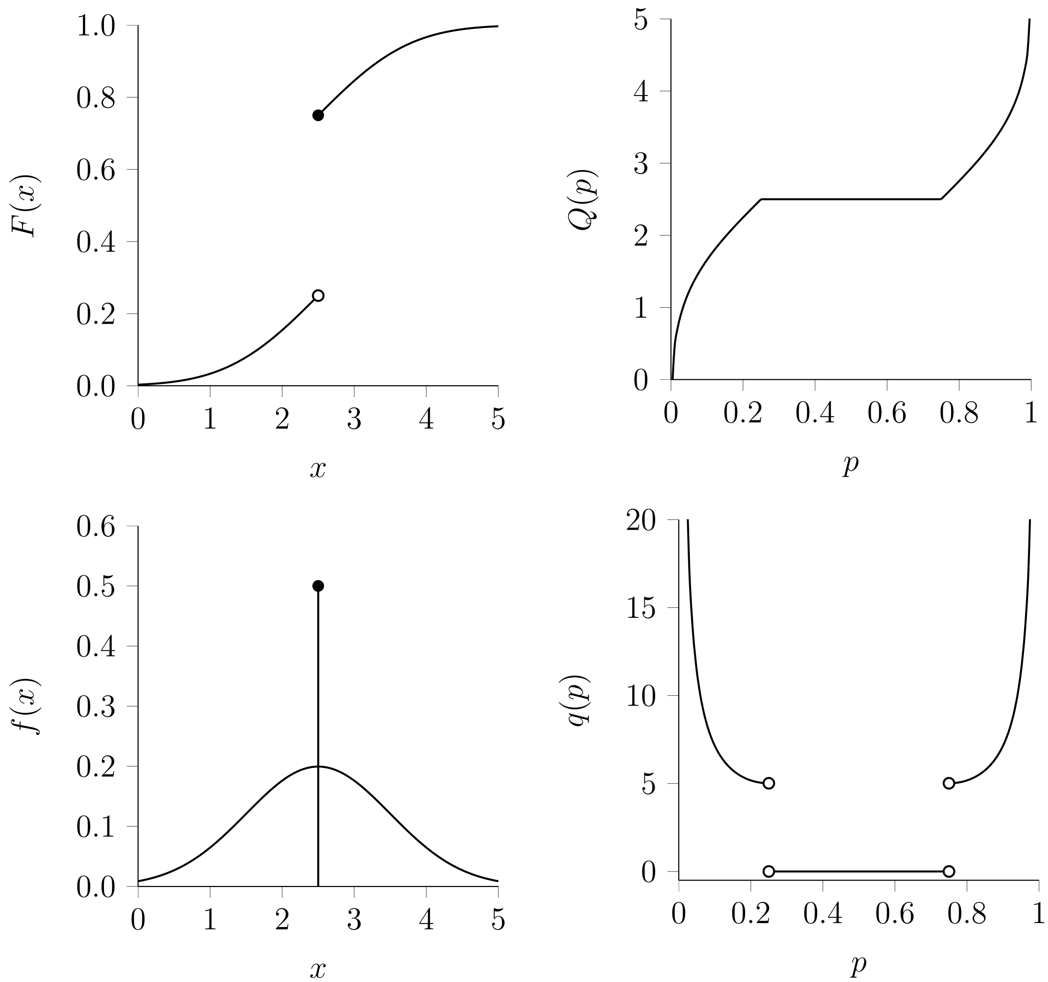

Figure 2.4: A discontinuous probability distribution

Now assume that F is not continuous and choose $x_0 \in \mathrm{dom}(F)$ at a point of discontinuity so that $p_L < p_R$, where $p_L = \lim_{x \to x_0^-} F(x)$ and $p_R = \lim_{x \to x_0^+} F(x)$. Choose p_0 so that $p_L < p_0 < p_R$. Applying (2.1), $Q(p_0) = \inf\{x \mid p_0 \leq F(x)\} = \inf\{x \mid p_R \leq F(x)\} = Q(p_R)$ or $Q(p_0) = Q(p_R)$, which contradicts the original statement that $Q(p)$ is strictly increasing. □

Figure 2.5 depicts a continuous probability distribution whose quantile function is not continuous. The discontinuous region of the quantile function corresponds to an interval of zero probability density. For a quantile function that is *both* continuous and strictly increasing, a condition beyond that of Proposition 1 must hold.

Proposition 2. *The quantile function Q of a probability distribution F is continuous if and only if F is strictly increasing on the interval $\{x \mid 0 < F(x) < 1\}$.*

Proof. First assume $F(x)$ is not strictly increasing and choose three points from its domain, $x_1 < x_2 < x_3$ so that $F(x_1) = F(x_2) < F(x_3)$. Now $Q(F(x_1)) = Q(F(x_2)) < Q(F(x_3))$. So Q is not continuous, because it does not map to the point x_2, a contradiction.

Now assume Q is not continuous and choose $p_0 \in (0, 1)$ at a point of discontinuity so that $x_L < x_R$, where $x_L = \lim_{p \to p_0^-} Q(p)$ and $x_R = \lim_{p \to p_0^+} Q(p)$. Choose x_0 so that $x_L < x_0 < x_R$. By (2.1) $Q(F(x_L)) = Q(F(x_0))$, which implies $F(x_L) = F(x_0)$, contradicting the statement that $F(x)$ is strictly increasing. □

The focus of this dissertation is probability distributions whose CDFs are both continuous *and* strictly increasing, making them and their quantile functions bijective so that $Q = F^{-1}$. Unless otherwise stated, references to probability distributions in the subsequent chapters make this assumption.

2.3 Transforming Quantiles and Cumulative Probabilities

One can modify probability distributions through various transformations of probability and quantile. The resulting distribution, when represented as quantile function,

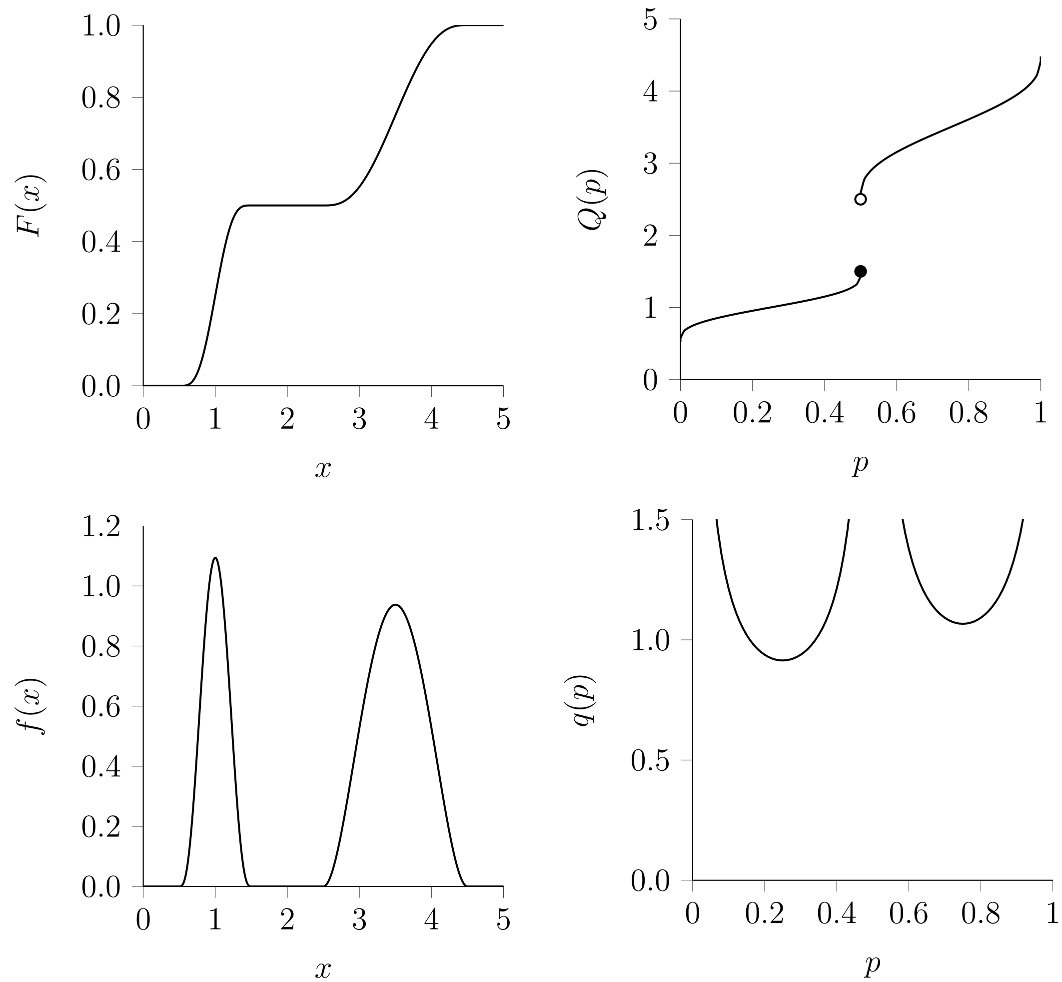

Figure 2.5: A probability distribution whose CDF is not strictly increasing

has an analogous form when represented as a PDF or CDF. For example, applying a positive linear transformation to a quantile function $a+bQ(p)$ is analogous to shifting and scaling the argument of its CDF $F(\frac{x-a}{b})$.

Proposition 3. *A positive linear transformation of a quantile function is a quantile function.*

This proof and the others that immediately follow, take two parts: show that the function is 1) defined and 2) nondecreasing over the interval $(0,1)$.

Proof. Given a quantile function $Q(p)$ and real numbers a and $b > 0$, let $\tilde{Q}(p) = a + bQ(p)$. Since $Q(p)$ is a quantile function, it is defined over $(0,1)$, therefore $\tilde{Q}(p)$ must be also defined over $(0,1)$. Also, its derivative $Q'(p) \geq 0$, making the derivative $\tilde{Q}'(p) = bQ'(p) \geq 0$ since $b > 0$. Thus, $\tilde{Q}(p)$ is nondecreasing. □

This result naturally leads to a discussion of *standard* probability distributions. Gilchrist [17, page 64] defines them using quantiles, and I adapt his definition here.

Definition 3. *A probability distribution is a standard probability distribution if some measure of position (e.g., its median) is zero, and some linear measure of its variablility (e.g., its inter-quartile range) is one.*

A standard normal distribution fits the above definition because its median (and its mean) is zero and its standard deviation, a constant multiple of its inter-quartile range, is one. This definition allows for two others.

Definition 4. *A probability distribution with quantile function $Q(p)$ has location parameter a and scale parameter b if $Q(p) = a + bQ_s(p)$, where $Q_s(p)$ is the quantile function of a standard probability distribution.*

A location parameter is a scalar that translates the origin of the graph of a standard probability distribution's PDF and CDF. A familiar location parameter is the mean μ of a normal distribution. One can use a scale parameter to change the units of an uncertain quantity. As an example, suppose X is an uncertain quantity that represents your belief over the minimum width of your car's brake pads in units of millimeters. Let $Z = X/10$ represent that same width in centimeters. Given the quantile function representing X is $Q(p)$, then the quantile function that represents Z is $\frac{1}{10}Q(p)$. A familiar scale parameter is the standard deviation σ of a normal distribution.

Proposition 3 shows that a positive linear transform of a quantile function is a quantile function. This is not the only operation guaranteed to yield a quantile function—the same is true for certain nonlinear transforms of quantile functions and even sums of quantile functions.

Proposition 4. *A positive linear combination of a finite number of quantile functions is a quantile function.*

Proof. By Proposition 3, this proof can proceed without loss of generality by replacing linear combinations of quantile functions with the summation of quantile functions. Let $p_1 \leq p_2$, where $p_1, p_2 \in (0, 1)$. Since quantile functions are non-decreasing, $Q_1(p_1) \leq Q_1(p_2)$ and $Q_2(p_1) \leq Q_2(p_2)$, making the sum $Q_1(p_1) + Q_2(p_1) \leq Q_1(p_2) + Q_2(p_2)$. Therefore, the sum of two quantile functions is a quantile function. By an inductive argument, a positive sum of any finite number of quantile functions is a quantile function. $\square$

Proposition 4 allows the creation of new quantile functions that are positive linear combinations of other quantile functions. Chapter 3 builds on this result.

Proposition 5. *Given a nondecreasing function h and a quantile function $Q(p)$, the composition $h(Q(p))$ is a quantile function.*

Proof. Let $p_1 \leq p_2$, where $p_1, p_2 \in (0, 1)$. Because Q is a quantile function, this implies $Q(p_1) \leq Q(p_2)$. Since h is a nondecreasing function, it is true that $h(Q(p_1)) \leq h(Q(p_2))$. $\square$

Proposition 5 serves as an alternate proof of Proposition 3 since a positive linear transformation is a nondecreasing function. Probability distributions described by transforms of quantile functions have a few features of note. The graph of both the quantile function (and CDF) of a quantile function $\tilde{Q}(p) = h(Q(p))$ is the graph of the original quantile function $Q(p)$ with the function h applied to its quantiles. However, the graph of its pPDF (and PDF) requires an additional transformation of the probability density function f.

Corollary 2. *A probability distribution whose quantile function $\tilde{Q}(p) = h(Q(p))$, where $Q(p)$ is a quantile function and h is a nondecreasing function has a pPDF $\tilde{f}(\tilde{Q}(p)) = f(Q(p))/h'(Q(p))$ wherever f and h' are defined.*

Proof.

$$\begin{aligned}
\tilde{f}(\tilde{Q}(p)) &= \frac{1}{\frac{d}{dp}h(Q(p))} && \text{by Proposition 12;} \\
&= \frac{1}{h'(Q(p))Q'(p)} && \text{by the derivative chain rule;} \\
&= \frac{f(Q(p))}{h'(Q(p))} && \text{by Proposition 12.}
\end{aligned}$$

□

Proposition 3 shows that any function of the form $a + bQ(p)$, where $Q(p)$ is a quantile function and $b > 0$, must also be a quantile function. Applying a positive affine transformation to the cumulative probability $p = F(x)$ rather than quantile function $Q(p)$ provides another feature of quantile functions.

Proposition 6. *Let F be a probability distribution with PDF $f(x)$, and quantile function $Q = F^{-1}$. The quantile function $\hat{Q}(p) = Q(\frac{p-a}{b-a})$, $0 \leq a < b \leq 1$ if and only if its associated probability distribution $\hat{F}$ is the probability distribution F conditioned on the interval $(Q(a), Q(b))$.*

Proof. Let $\alpha = Q(a)$, $\beta = Q(b)$, and $\hat{p} = \frac{p-a}{b-a}$. Let $\hat{F}(x) \equiv F(\hat{x}) \equiv \hat{p}$ be the CDF corresponding to $\hat{Q}(p) \equiv Q(\hat{p}) \equiv \hat{x}$.

$$
\begin{aligned}
&\hat{Q}(p) = Q\left(\frac{p-a}{b-a}\right) \\
\Leftrightarrow\ & F(\hat{Q}(p)) = \frac{p-a}{b-a} && \text{by applying } F; \\
\Leftrightarrow\ & \hat{F}(x) = \frac{F(x) - F(\alpha)}{F(\beta) - F(\alpha)} && \hat{Q}(p) = Q(\hat{p}) \Leftrightarrow F(\hat{Q}(p)) = \hat{F}(x); \\
\Leftrightarrow\ & \hat{F}(x) = \frac{\int_\alpha^x f(t)dt}{F(\beta) - F(\alpha)} && \text{by the fundamental theorem of calculus.}
\end{aligned}
$$

This is the definition of conditional probability for the PDF of a probability distribution F over the interval $(Q(a), Q(b))$. □

Proposition 6 shows that one need not compute a new quantile function when conditioning a probability distribution F over an interval $(Q(a), Q(b))$. It suffices to use the composition of the quantile function Q with the affine transformation $\frac{p-a}{b-a}$.

The next proposition shows how to reflect a distribution about the origin. Let $Q(p)$ and $\tilde{Q}(p)$ be the quantile functions corresponding to PDFs $f(x)$ and $\tilde{f}(x)$, respectively.

Proposition 7. *Given $f(x) > 0$, the quantile function $\tilde{Q}(p) = -Q(1-p)$ if and only if the PDF $\tilde{f}(x)$ is the PDF $f(x)$ reflected about the point $x = 0$.*

Proof. It suffices to show that $\tilde{Q}(p) = -Q(1-p)$ if and only if $\tilde{f}(x) = f(-x)$.

$$
\begin{aligned}
&\tilde{Q}(p) = -Q(1-p) \\
\Leftrightarrow\ & \tilde{Q}'(p) = Q'(1-p) \\
\Leftrightarrow\ & \tilde{f}(\tilde{Q}(p)) = f(Q(1-p)) && \text{by Proposition 12;} \\
\Leftrightarrow\ & \tilde{f}(x) = f(-x) && \text{letting } x = \tilde{Q}(p).
\end{aligned}
$$

□

Proposition 7 is helpful when discussing the left and right tail behavior of probability distributions—the subject of Chapter 5.

2.4 Integrals of Quantile Functions

The mathematics of probability theory involves the integration of various functions. For example, for a continuous uncertain quantity X, the m^{th} moment $\mathrm{E}[X^m]$ is defined by the integral $\int x^m f(x)dx$. One can also compute these integrals involving PDFs by integrating various functions of quantile functions.

Proposition 8. *The m^{th} moment of a probability distribution F with quantile function Q is $\int (Q(p))^m dp$.*

Proof. Let $f(x) = \frac{dF}{dx}$. The definition of the m^{th} moment of F is

$$\mathrm{E}[X^m] = \int_{-\infty}^{+\infty} x^m f(x)dx. \tag{2.3}$$

Since $p \equiv F(x)$ and $f(x) \equiv \frac{dF}{dx}$, $dp = f(x)dx$. By substituting, $dp = f(x)dx$, and $x = Q(p)$, (2.3) becomes

$$\mathrm{E}[(Q(p))^m] = \int_0^1 (Q(p))^m \, dp. \tag{2.4}$$

. □

Another important integral in probability theory is that of entropy. One can compute the entropy of a probability distribution using quantile functions. The focus here is on relative entropy—that of continuous probability distributions.

Proposition 9. *The differential entropy of a continuous probability distribution F with quantile function Q is $\int \log(Q'(p))dp$.*

Proof. Let $f(x) = \frac{dF}{dx}$. The integral $-\int f(x)\log(f(x))dx$ is the definition of the differential entropy of F. Substituting $dp = f(x)dx$, $x = Q(p)$, and $f(Q(p)) = 1/Q'(p)$ gives $\int \log(Q'(p))dp$. □

Kullback-Leibler (KL) divergence quantifies the information lost when approximating a probability distribution F with another distribution G. The KL divergence $D_{KL}(F,G) \equiv \int \log\left(\frac{f(x)}{g(x)}\right) f(x)dx$. As with differential entropy, one can express KL divergence in terms of quantile functions. The first integral of (2.5) is the cross-entropy from G to F, and the second is the entropy of distribution F.

Proposition 10. *The KL divergence of a probability distribution G from a reference probability distribution F is*

$$\int \log(Q'_G(G(Q_F(p))))dp - \int \log(Q'_F(p))dp. \tag{2.5}$$

Proof. The definition of KL divergence is $D_{KL}(F,G) \equiv \int \log\left(\frac{f(x)}{g(x)}\right) f(x)dx$, which equals $-\int \log(g(x))f(x)dx + \int \log(f(x))f(x)dx$. Substituting $dp = f(x)dx$, and $x = Q_F(p)$ and by (2.2),

$$D_{KL}(F,G) = \int \log(Q'_G(G(Q_F(p))))dp - \int \log(Q'_F(p))dp.$$

□

This completes the introduction to quantile functions, and gives the mathematical foundation for the tools and theory that follow. The first research objective beckons—identify a useful class of probability distributions that is readily parameterized by a decision analyst's quantile-probability data.

Chapter 3

Quantile-Parameterized Distributions

Although their use is not common in decision analysis, quantile functions have some advantages over traditional functional representations of continuous probability distributions like CDFs and PDFs. First, transforming a quantile function transforms the uncertain variable that it represents. Second, a convenient method for probabilistic simulation (inverse transform sampling) uses a probability distribution's quantile function. Third, linear combinations of quantile functions often enjoy modeling flexibility over mixtures of CDFs or PDFs for representing quantile-probability data, because there is no constraint that requires coefficients to be positive and sum to unity.

This chapter gives a summary of quantile-parameterized distributions, the probability distributions we introduced in 2011 [34]. A QPD is defined by its quantile function, thus affording a decision analyst all of the advantages of modeling with quantile functions. In addition, QPDs are parameterized by quantile-probability data by way of a linear map—a desirable feature highlighted in the next section.

3.1 Definition

In order to formally define a QPD, it is convenient to first introduce three distinctions.

Definition 5. *A set of functions $\{g_i(p) \mid i \in 1:n\}$ is linearly independent over an interval $\mathcal{I} \subseteq \bigcap_{i=1}^n \mathrm{dom}(g_i)$ if the linear combination $\sum_{i=1}^n \beta_i g_i(p) = 0$ for all $p \in \mathcal{I}$ implies that all n components of β equal zero.*

Definition 6. *The members of a set of linearly independent functions $\{g_i(p) \mid i \in 1:n, p \in (0,1)\}$ are called basis functions.*

Definition 7. *A set of basis functions $\{g_i(p) \mid i \in 1:n, p \in (0,1)\}$ is regular if each basis function is continuously differentiable, and the set of basis function derivatives $\{g_i'(p) \mid i \in 1:n, p \in (0,1)\}$ is linearly independent over the interval (a,b), for all $0 < a < b < 1$.*

By definition of linear independence, given at least one nonzero coefficient, a linear combination of the derivatives of a regular set of basis functions will not map to zero over an interval. Because of this, Definition 7 implies that a linear combination of a set of regular basis functions will have no "flat spots." Unless otherwise specified, the notation $g_i(p)$ will refer to a member of a regular set of basis functions.

A particular family of QPDs corresponds to a particular regular set of basis functions. The set defines the QPD's flexibility to represent various distributional shapes, its support, and its tail behavior—features that chapters 3, 4, and 5, respectively, examine. The notion of a regular set of basis function sets the stage for the key definition of this chapter.

Definition 8. *A QPD is a probability distribution whose quantile function can be written:*

$$Q(p) = \sum_{i=1}^n \beta_i g_i(p) \quad 0 < p < 1 \tag{3.1}$$

where $\beta \in \mathbf{R}^n$, and $\{g_i(p) \mid i \in 1:n, p \in (0,1)\}$ is a regular set of basis functions.

By Definition 8, the normal, exponential, logistic, and uniform probability distributions are all QPDs. QPDs are similar to the parametric family of distributions that Karvanen introduces for estimating the parameters of a probability distribution by way of L-moment statistics [33]. The QPD definition removes Karvanen's restriction that the basis functions be quantile functions (i.e., nondecreasing), and it adds the regularity condition of Definition 7.

Proposition 11. *F is a QPD if and only if its quantile function Q is strictly increasing and continuously differentiable on the interval* $(0, 1)$.

Proof. First show that if F is a QPD, then Q is continuous and strictly increasing. Given F is a QPD, Definitions 7 and 8 state that Q is a linear combination of continuously differentiable basis functions $\{g_i(p) \mid i \in 1 : n, p \in (0, 1)\}$. Therefore, Q is continuous. It also implies that Q' is a linear combination of the derivatives of those functions. By Definition 7, the set of the derivatives of the basis functions $\{g_i'(p) \mid i \in 1 : n, p \in (0, 1)\}$ is linearly independent over the interval $p \in (a, b)$, for all $0 < a < b < 1$. This implies that there exists no interval $(a, b),\ 0 < a < b < 1$ over which the derivative $Q'(p)$ is everywhere zero. Therefore, Q is strictly increasing.

Next, show that if Q is strictly increasing and continuously differentiable on the interval $(0, 1)$, then F is a QPD. Given that Q is strictly increasing on the interval $(0, 1)$, there exists no interval $p \in (a, b),\ 0 < a < b < 1$ over which its derivative $Q'(p)$ is everywhere zero. This, along with the fact that Q is continuously differentiable, implies that the singleton $\{Q\}$ meets the criteria of a regular set of basis functions. To complete the proof, use Definition 8 to construct a QPD with lone basis function Q and coefficient $\beta_1 = 1$. □

This equivalence relationship gives a different perspective on the definition of a QPD. It also gives rise to some corollaries that aid in the understanding of QPDs.

Corollary 3. *A QPD is a continuous probability distribution.*

Corollary 4. *A QPD has a CDF $F(x)$ that is strictly increasing for all* $\{x \mid 0 < F(x) < 1\}$.

Corollary 5. *A QPD's quantile function Q is the inverse function of its CDF,* $Q = F^{-1}$.

These three corollaries are true as a consequence of Proposition 11. Corollary 3, because Q is strictly increasing; Corollary 4, because Q is continuous. Corollary 5 combines these two facts. An example of a continuous probability distribution that is not a QPD is a probability distribution whose CDF is piecewise linear and strictly increasing (making its quantile function not continuously differentiable). An example of a strictly increasing CDF that does not represent a QPD is $F(x) = x^3 + 0.5$. Its quantile function $Q(p) = \max\{0, p - 0.5\}^{1/3} + \min\{0, p - 0.5\}^{1/3}$ is not differentiable at $p = 0.5$.

The quantile function equation (3.1) of a QPD also allows for the concise representation of various functions of quantile functions. Two such useful examples are the pPDF and moments of a QPD.

Proposition 12. *Where defined, the equation for the p-probability density function of a QPD is*

$$f(Q(p)) = \left(\sum_{i=1}^{n} \beta_i \frac{dg_i(p)}{dp} \right)^{-1}. \tag{3.2}$$

Proof. Substitute (3.1), the definition of a QPD into (2.2), the definition of a pPDF. □

Proposition 12 is also a general equation for computing the PDF of a specific QPD. As an example, take the QPD with quantile function $Q(p) = \beta_1 p + \beta_2(-\log(1-p))$. Neither its PDF nor CDF have a closed form, a circumstance that is often the case for QPDs. However, using its pPDF $f(Q(p)) = \frac{1-p}{\beta_1(1-p)+\beta_2}$, one can graph its PDF by plotting $f(Q(p))$ versus $Q(p)$. Figure 3.1 graphs the PDF of this QPD with $\beta_1 = 2$ and $\beta_2 = 1$.

Proposition 13. *The m^{th} moment of a QPD is*

$$\int_0^1 \left(\sum_{i=1}^{n} \beta_i g_i(p) \right)^m dp. \tag{3.3}$$

Proof. Substitute (3.1), the definition of a QPD into (2.4), the definition of the m^{th} moment of a probability distribution in terms of its quantile function. □

Proposition 13 is useful for computing moments of a QPD whose PDF is not an explicit function of x, making (2.3) an integral of an implicit function of x. In contrast,

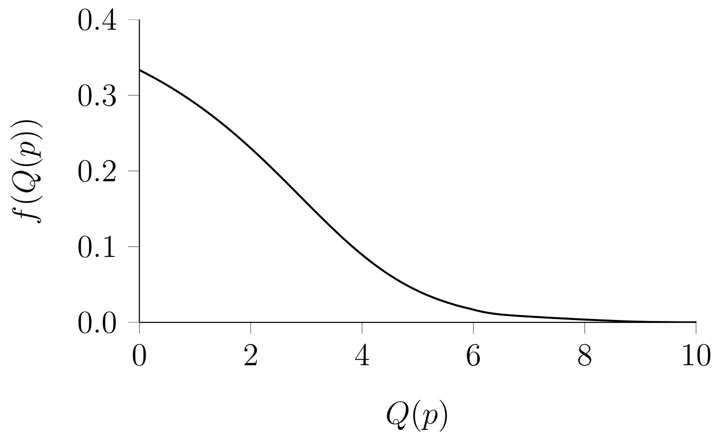

Figure 3.1: A graph of the PDF of the probability distribution with quantile function $Q(p) = 2p - \log(1-p)$

the integral (3.3) is an explicit function of p. This equation allows a more straightforward means of computing moments, for example, when using moment matching to discretize a QPD.

3.2 Parameterization and Feasibility

While every QPD has a quantile function of the form (3.1), not all functions of the form (3.1) characterize a QPD. Just as parametric probability distributions have infeasible parametric regions (e.g., a normal distribution's standard deviation must be positive), various vectors of coefficients β yield functions that aren't quantile functions because they are decreasing over a particular interval. This section explores the feasibility of QPDs and how to parameterize a QPD using quantile-probability data—a requirement for meeting the first desideratum of Chapter 1.

Proposition 14. *A function of the form* (3.1) *characterizes a QPD if and only if*

$$\sum_{i=1}^{n} \beta_i \frac{dg_i(p)}{dp} \geq 0 \quad \textit{all } p \in (0,1) \tag{3.4}$$

Proof. According to Proposition 11, the quantile function $Q(p)$ of a QPD is a continuously differentiable function that is strictly increasing in p. Q meets these conditions if and only if (3.4) holds ($Q'(p)$ is nonnegative). $\square$

Given the regularity condition of Definition 7, the set $\{p \in (0,1) \mid \sum_{i=1}^{n} \beta_i \frac{dg_i(p)}{dp} = 0\}$ has measure zero ($\sum_{i=1}^{n} \beta_i \frac{dg_i(p)}{dp} > 0$ almost everywhere). An example of a QPD where the number of elements in the set $\{p \in (0,1) \mid \sum_{i=1}^{n} \beta_i \frac{dg_i(p)}{dp} = 0\}$ is nonzero is the QPD with quantile function $Q(p) = (p - 0.5)^3$. Its derivative is zero at $p = 0.5$ and positive otherwise. Proposition 14 is important because it gives a method to verify whether a function of the form of (3.1) characterizes a continuous probability distribution. The condition in (3.4) 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 vectors

$$S_\beta = \left\{ \beta \in \mathbf{R}^n \;\middle|\; \sum_{i=1}^{n} \beta_i \frac{dg_i(p)}{dp} \geq 0, p \in (0,1) \right\}$$

that make (3.1) a QPD.

Proposition 15. *A QPD's set of feasible vectors S_β is a proper subset of $\mathbf{R}^n$.*

Proof. Finding a single infeasible vector proves this proposition. Take a QPD with quantile function $Q(p) = \sum_{i=1}^{n} \beta_i g_i(p)$. Since $Q(p)$ is a quantile function, at least one feasible vector $\hat{\beta} \in S_\beta$ must exist, so that $\sum_{i=1}^{n} \hat{\beta}_i \frac{dg_i(p)}{dp} > 0$ for some $p \in (0,1)$. Let $\breve{\beta} = -\hat{\beta}$. This makes $\sum_{i=1}^{n} \breve{\beta}_i \frac{dg_i(p)}{dp} < 0$ for some $p \in (0,1)$; so $\breve{\beta} \in \mathbf{R}^n$, while $\breve{\beta} \notin S_\beta$. Therefore, $S_\beta \subset \mathbf{R}^n$. $\square$

There will always exist coefficient vectors β that result in a function of the form (3.1) that is not a quantile function. This result is important when updating a QPD's coefficients using Bayes's theorem, a fact that §6.5 highlights.

Proposition 16. *A QPD's set of feasible vectors S_β is convex.*

Proof. One can express S_β as an infinite intersection of sets $\bigcap_{p \in (0,1)} S_p$, where S_p is the halfspace $\{\beta \in \mathbf{R}^n \mid b^T \beta \geq 0\}$ and the vector $b = \left(\frac{dg_1(p)}{dp}, \cdots, \frac{dg_n(p)}{dp} \right)^T$. All halfspaces are convex sets, and any intersection of convex sets yields a convex set, therefore S_β is convex. $\square$

Corollary 6. *A QPD's set of feasible vectors S_β is a convex cone.*

Proof. Let β_2 and β_3 be QPD-feasible vectors, and let $b = \left(\frac{dg_1(p)}{dp}, \cdots, \frac{dg_n(p)}{dp}\right)^T$ so that $b^T\beta_2 \geq 0$ and $b^T\beta_3 \geq 0$, all $p \in (0,1)$. Choose two nonnegative constants, $c, d \in \mathbf{R}$, and create the linear combination $c \cdot (b^T\beta_2) + d \cdot (b^T\beta_3)$, which is also nonnegative because it is a positive linear combination of positive real numbers. Rearranging terms, $b^T(c \cdot \beta_2 + d \cdot \beta_3) \geq 0$, completing the proof. $\square$

Since convex optimization requires convex feasible sets, Proposition 16 is directly relevant to optimization problems involving QPDs. Perhaps more importantly, Proposition 16 is useful for determining whether a function of the form (3.1) is a quantile function. The feasibility of input quantiles will return with the discussion of the parametric limits of an example QPD.

Hearkening back to Proposition 5, transforming any quantile function with a nondecreasing function h yields a quantile function. With a further restriction to the transforming function h, an analogous statement holds for the quantile functions of QPDs.

Proposition 17. *Given a function h with inverse h^{-1}, both strictly increasing and continuously differentiable, the quantile function $\tilde{Q}(p) = h^{-1}(Q(p))$ describes a QPD if and only if $Q(p)$ describes a QPD.*

Proof. First, show that a QPD with quantile function $Q(p)$ implies a QPD with quantile function $\tilde{Q}(p)$. By Proposition 11 it suffices to show that $\tilde{Q}(p) = h^{-1}(Q(p))$ is strictly increasing and continuously differentiable (SICD) on the interval $(0,1)$. Because $Q(p)$ represents a QPD, Proposition 11 states that Q is SICD on the interval $(0,1)$. Transforming an SICD function Q by applying another SICD function h^{-1} yields an SICD function.

Next, show that a QPD with quantile function $\tilde{Q}(p)$ implies a QPD with quantile function $Q(p)$. Apply h to both sides of $\tilde{Q}(p) = h^{-1}(Q(p))$ to yield $h(\tilde{Q}(p)) = Q(p)$. Invoke the argument from the preceding paragraph to this equation. $\square$

The convention of setting $\tilde{Q}(p) = h^{-1}(Q(p))$ rather than $h(Q(p))$ is because h transforms the variable $x = \tilde{Q}(p)$. As an example, the transform for the log-normal distribution is $h(x) = \log(x)$. Applying this transform to the standard normal distribution and makes a lognormal distribution with quantile function $\tilde{Q}(p) = \exp(\Phi^{-1}(p))$ and pPDF $f(\tilde{Q}(p)) = \phi(\Phi^{-1}(p))/\exp(\Phi^{-1}(p))$, where ϕ is the standard normal PDF, and the parameters of the lognormal are $\mu = 0$, and $\sigma = 1$.

Corollary 7. *Given a function h with inverse h^{-1}, both strictly increasing and continuously differentiable, a QPD with quantile function $Q(p) = \sum_{i=1}^{n} \beta_i g_i(p)$ and a QPD with quantile function $\tilde{Q}(p) = h^{-1}(Q(p))$ have the same set of feasible vectors β.*

Corollary 7 is true by Proposition 17. It proves useful when checking the feasibility of a QPD whose quantile function is an h transform of another QPD. Take a QPD with quantile function $Q(p) = \mu + \sigma\Phi^{-1}(p)$. This quantile function describes a normal distribution with mean μ and standard deviation σ for all $\mu \in \mathbf{R}$ and all $\sigma \in \mathbf{R}_{++}$. By Corollary 7, the quantile function $Q(p) = \exp(\mu + \sigma\Phi^{-1}(p))$ describes a log-normal distribution for all $\mu \in \mathbf{R}$ and all $\sigma \in \mathbf{R}_{++}$. The parametric limits are identical.

The following theorem shows that quantile-probability data can uniquely determine a QPD's β vector. In such cases, one can think of these quantile-probability data as QPD parameters themselves.

Theorem 1 (Quantile Parameters Theorem)**.** *A set of n distinct points $\{(x_i, p_i) \mid i \in 1:n\}$ uniquely determine $\beta \in \mathbf{R}^n$ of a QPD by the matrix equation*

$$\beta = Y^{-1}x, \tag{3.5}$$

Where $\beta, x \in \mathbf{R}^n$, the set of basis functions $\{g_i(p) \mid i \in 1:n, p \in (0,1)\}$ is regular, and

$$Y = \begin{bmatrix} g_1(p_1) & \cdots & g_n(p_1) \\ \vdots & \ddots & \vdots \\ g_1(p_n) & \cdots & g_n(p_n) \end{bmatrix}, \tag{3.6}$$

if and only if:

I. *the matrix Y is invertible, and*

II. $\sum_{i=1}^{n} \beta_i \frac{dg_i(p)}{dp} \geq 0$, *all* $p \in (0, 1)$.

Proof. Condition I is true if and only if equation (3.5) holds, and the resulting function (3.1) is unique to the quantile inputs $x \in \mathbf{R}^n$. To show this, set up a system of n equations according to (3.1). This yields the matrix equation $x = Y\beta$, using the definition of Y from (3.6). Equation (3.5) holds if and only if Y is invertible. Since Y is square, it defines a one-to-one mapping of the quantiles $x \in \mathbf{R}^n$ to the coefficients $\beta \in \mathbf{R}^n$.

It is possible that the function (3.1) resulting from a set of input quantiles $x \in \mathbf{R}^n$ does not represent a probability distribution. A set of coefficients $\beta \in \mathbf{R}^n$ characterize a QPD if and only if condition II holds. This is true by Proposition 14. □

The power of the Quantile Parameters Theorem is that a decision analyst does not need to directly assess the β vector. Instead, (3.5) uniquely maps n pairs of quantile-probability data to $\beta \in \mathbf{R}^n$. Moreover, one can use other sources of quantile-probability data, such as the results of a probabilistic simulation, to parameterize a QPD.

A note on condition I: because the set of basis functions is linearly independent, the matrix Y is invertible in all but pathological cases. If such a case should occur, a small perturbation will solve the problem. A note on condition II: Proposition 15 shows that it is possible to choose a set of points $\{(x_i, p_i) \mid i \in 1 : n\}$ such that this condition is not satisfied for a given set of basis functions—even when those points are coherent with the axioms of probability (nondecreasing in x). This chapter will return to the discussion of feasibility with an example QPD.

3.3 An Example QPD: The Simple Q-normal

To this point, the discussion of QPDs has been rather abstract. An example will demonstrate some of the many features and limitations of QPDs. This prototypical QPD is the *simple Q-normal*.

In accordance with (3.1), all QPDs are completely characterized by a set of basis functions. The simple Q-normal is defined as the QPD whose basis functions form the set $\{1, p, \Phi^{-1}(p), p\Phi^{-1}(p)\}$. This makes its quantile function

$$Q(p) = \beta_1 + \beta_2 p + \beta_3 \Phi^{-1}(p) + \beta_4 p \Phi^{-1}(p), \quad p \in (0,1) \tag{3.7}$$

As noted in Keelin and Powley [34], this form is equivalent to an uncertain variable x described by the standard normal CDF $\Phi(\mu(p), \sigma(p))$ whose parameters are linear functions of $p \equiv \Phi(x)$.

$$\mu(p) = \beta_1 + \beta_2 p \tag{3.8}$$

$$\sigma(p) = \beta_3 + \beta_4 p \tag{3.9}$$

This makes its CDF and PDF implicit functions of x. To compute its pPDF, differentiate (3.7) and take its reciprocal.

$$f(Q(p)) = \frac{\phi(\Phi^{-1}(p))}{\beta_3 + \beta_4 p + \phi(\Phi^{-1}(p))\left(\beta_2 + \beta_4 \Phi^{-1}(p)\right)} \tag{3.10}$$

To create a three parameter Q-normal distribution, set any one of β_1, β_2, β_3, or β_4 equal to zero. The simple Q-normal reverts to the normal distribution when $\beta_2 = \beta_4 = 0$.

To compute the coefficients $\beta \in \mathbf{R}^4$ from a set of four quantile-probability pairs, solve the set of four linear equations (3.11), denoted by the matrix equation $x = Y\beta$.

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 & p_1 & \Phi^{-1}(p_1) & p_1\Phi^{-1}(p_1) \\ 1 & p_2 & \Phi^{-1}(p_2) & p_1\Phi^{-1}(p_2) \\ 1 & p_3 & \Phi^{-1}(p_3) & p_3\Phi^{-1}(p_3) \\ 1 & p_4 & \Phi^{-1}(p_4) & p_4\Phi^{-1}(p_4) \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}. \tag{3.11}$$

This matrix Y represents a linear map $\mathbf{R}^4 \to \mathbf{R}^4$ of the quantiles x to the coefficients β. In accordance with the Quantile Parameters Theorem, the simple Q-normal is fully parameterized by any set of four quantiles that result in a nondecreasing quantile function. To compute β, rewrite (3.11) as $\beta = Y^{-1}x$. As long as Y is invertible, this

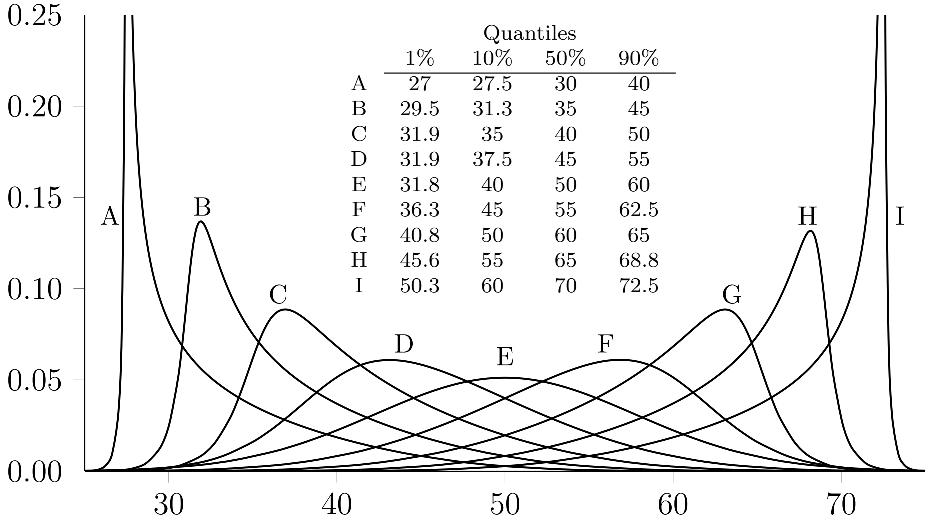

Figure 3.2: Some skewed simple Q-normal distributions

method delivers a unique function for any given set of four quantile-probability pairs.

Figures 3.2 and 3.3 show that it is possible to parameterize the simple Q-normal into a rich set of distributional forms, each consistent with a different set of quantiles. The simple Q-normal is supported over the real numbers, and it allows for the adjustment of its first four moments through the adjustment of its quantiles or coefficients. And because a QPD is constructed explicitly as a quantile function, it is well suited to probabilistic simulation—a feature that is true of QPDs in general. To generate a simple Q-normal random variate via the inverse transformation method, sample a uniform$(0, 1)$ random variate and substitute it for the variable p in (3.7).

3.4 Moments of the Simple Q-normal

Since the probability density function for the simple Q-normal is implicit, one can use (3.3) to determine its central moments. Substituting (3.7), the mean of the simple Q-normal is

$$\mathrm{E}[x^m] = \int_0^1 \left(\beta_1 + \beta_3\Phi^{-1}(p) + \beta_4 p\Phi^{-1}(p) + \beta_2 p\right) dp$$

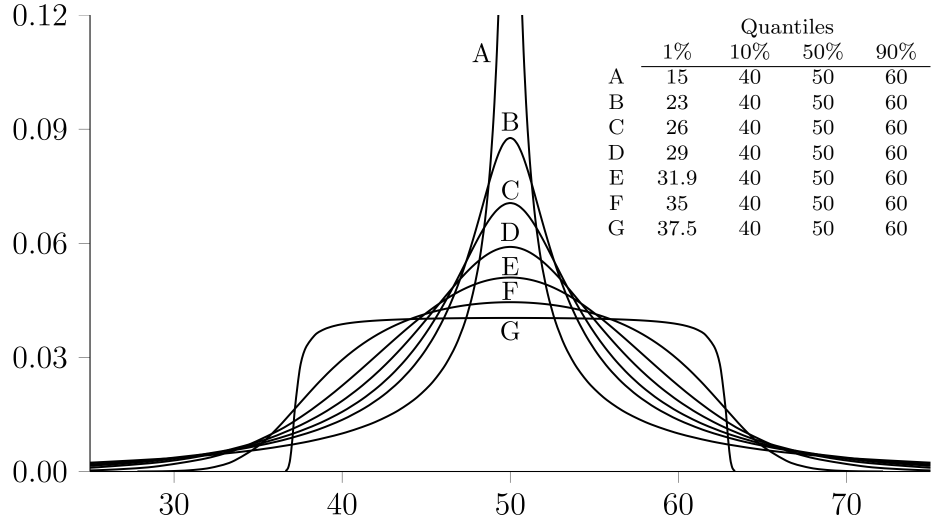

Figure 3.3: Some symmetric simple Q-normal distributions

Some further simplification yields the equation

$$\mathrm{E}[x^m] = \beta_1 + \frac{\beta_2}{2} + \beta_3 \int_0^1 \Phi^{-1}(p)dp + \beta_4 \int_{p=0}^1 p\Phi^{-1}(p)dp \tag{3.12}$$

According to (3.3), the first of the two remaining integrals in (3.12) is the mean of the standard normal distribution, which equals zero. For the second integral, change the variable of integration from p to z and integrate by parts

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

where $\mathrm{erf}(z)$ represents the error function. This quantity equals $\sqrt{\frac{1}{4\pi}}$. So the simple Q-normal's mean equals

$$\beta_1 + \frac{\beta_4}{\sqrt{4\pi}} + \frac{\beta_2}{2}. \tag{3.13}$$

By the same method, the simple Q-normal's variance is approximately

$${\beta_3}^2 + \beta_3\beta_4 + {\beta_4}^2\left(\frac{1}{3} + \frac{1}{2\pi\sqrt{3}} - \frac{1}{4\pi}\right) + \frac{\beta_3\beta_2}{\sqrt{\pi}} + 0.282\beta_4\beta_2 + \frac{{\beta_2}^2}{12}, \tag{3.14}$$

where 0.282 approximates the integral $\int_0^1 p^2\Phi^{-1}(p)dp$. The simple Q-normal's mean is not a function of β_3, the constant term of (3.9), just as the normal distribution's mean is not a function of its variance. Similarly, the simple Q-normal's variance is not a function of β_1, the constant term of (3.8), just as the normal distribution's variance is not a function of its mean. Recall that the simple Q-normal reduces to the normal distribution when $\beta_2 = \beta_4 = 0$. In this case, the mean must equal β_1, and the variance must equal ${\beta_3}^2$, a final demonstration that (3.13) and (3.14) are consistent with (3.8) and (3.9).

3.5 Parameterizing the Simple Q-normal with Quantiles from Familiar Probability Distributions

Suppose an expert assigns 1st, 10th, 50th, and 90th quantiles consistent with an underlying familiar, named distribution like the beta, logistic, student's t, and so on. By the Quantile Parameters Theorem, one can use these quantiles to parameterize the simple Q-normal distribution. But how good is the approximation? Taking these four quantiles from a diverse list of named probability distributions and using them to parameterize the simple Q-normal explores the fidelity of the approximation. Figure 3.4 depicts the CDF and PDF of each named distribution along with the simple Q-normal that approximates it. Note that the simple Q-normal's CDF is barely possible to discriminate from the CDF of the named distribution used to parameterize it. Table 3.1 shows the 1st, 10th, 50th, and 90th quantile for each named distribution along with two accuracy metrics:[1] the Kolmogorov-Smirnoff distance (maximum p-deviation) and Kullback-Leibler divergence of the simple Q-normal from each of the named distributions using (2.5).

[1]Each of these two metrics were computed by probabilistic simulation.

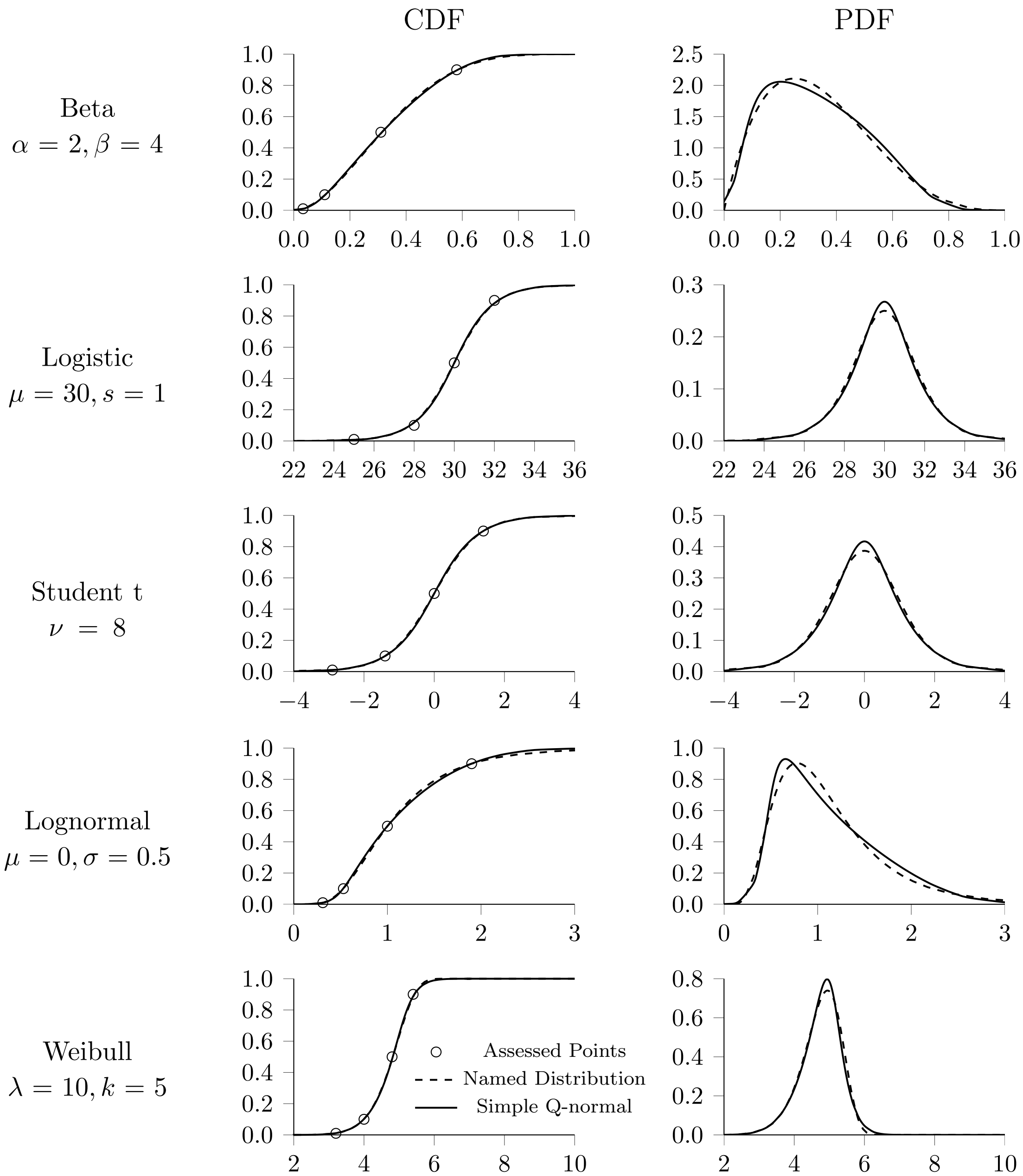

Figure 3.4: The simple Q-normal as parameterized by the quantiles of some named probability distributions

Table 3.1: Deviation between the simple Q-normal and various named distributions

Named Distribution	1%	10%	50%	90%	K-S Distance	KL Divergence
beta$(2, 4)$	0.033	0.11	0.31	0.58	0.010	0.008
logistic$(30, 1)$	25	28	30	32	0.009	0.035
student's t(8)	-2.9	-1.4	0	1.4	0.010	0.003
lognormal$(0, 0.5)$	0.31	0.53	1	1.9	0.017	0.043
Weibull$(10, 5)$	3.2	4.0	4.8	5.4	0.014	0.010
normal$(30, 7.8)$	12	20	30	40	0	0

3.6 Flexibility of the Simple Q-normal

One can transform the parametric limits associated with (3.10) in terms of two ratios: r_1 and r_2. The first indicates distributional symmetry

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

where x_i is the i^{th} quantile. For all symmetric distributions, $r_1 = 0.5$, whereas the right-skewed exponential distribution has an r_1 equal to 0.365 regardless of its rate parameter. The ratio r_2 indicates tail width

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

Projecting the quantile vector $x \in \mathbf{R}^4$ onto the $r_1 - r_2$ plane gives a better visualization of the parametric limits of the simple Q-normal. Plotting these limits demonstrates the flexibility that even a basic QPD like this one offers. The ovoid shape of Figure 3.5 represents the limits of feasible $r_1 - r_2$ pairs; feasible quantile ratios for the simple Q-normal lie within the ovoid; infeasible quantiles lie without.

Figure 3.6 shows how the limits of some named distributions, such as the normal and exponential, are points in the $r_1 - r_2$ plane. Other named distributions, 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. Its feasible region maps to an area in the $r_1 - r_2$

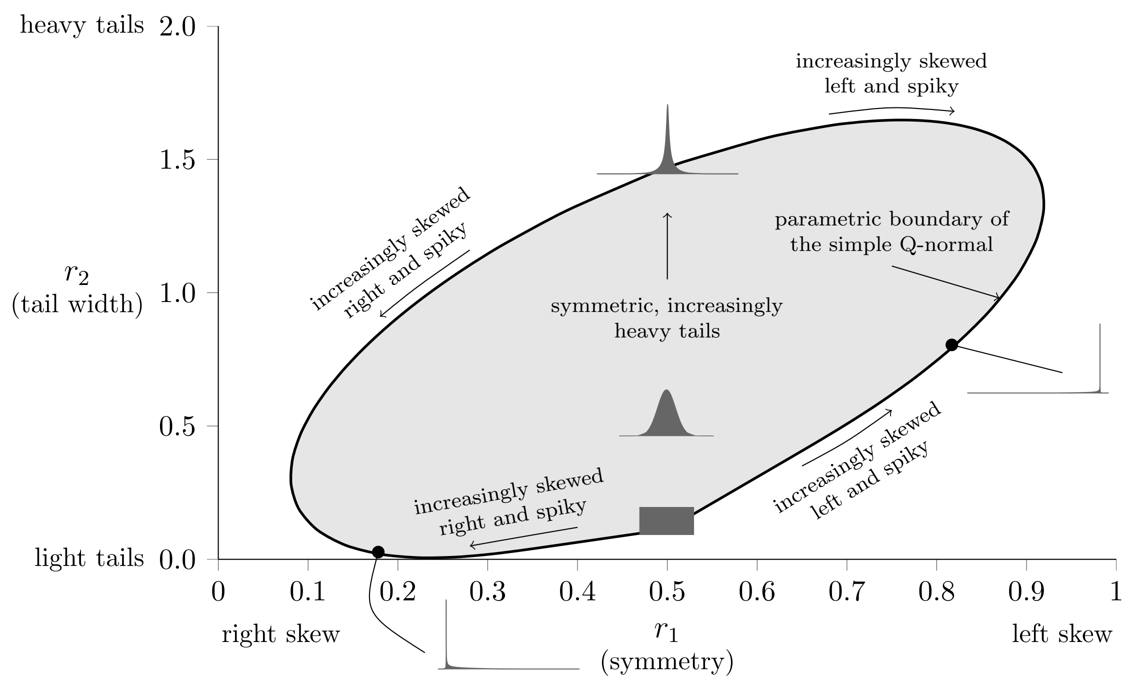

Figure 3.5: Simple Q-normal parametric limits in the $r_1 - r_2$ plane

plane. Figure 3.6 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 substitute for a wide range of named probability distributions in representing uncertainty.

Proposition 18. *The set of feasible quantile ratios $r = (r_1, r_2)$ for a QPD is convex.*

Proof. Let $\psi : (x_1,\ x_{10},\ x_{50},\ x_{90}) \rightarrow \left(\frac{x_{50}-x_{10}}{x_{90}-x_{10}}, \frac{x_{10}-x_1}{x_{90}-x_{10}}\right)$ 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 $S_x = \{x \in \mathbf{R}^4 \mid x = Y\beta, \beta \in S_\beta\}$ is the set of quantile vectors that yield a Q-normal probability distribution, and $S_\beta = \{\beta \in \mathbf{R}^4 \mid \sum_{i=1}^{n} \beta_i \frac{dg_i(p)}{dp} > 0$, all $p \in (0,1)\}$ is the set of feasible β coefficients. From Proposition 16, S_β is convex. Any linear transformation of a convex set is convex, so S_x is also convex. Since ψ is a linear fractional function, and linear fractional functions preserve convexity, S_r is

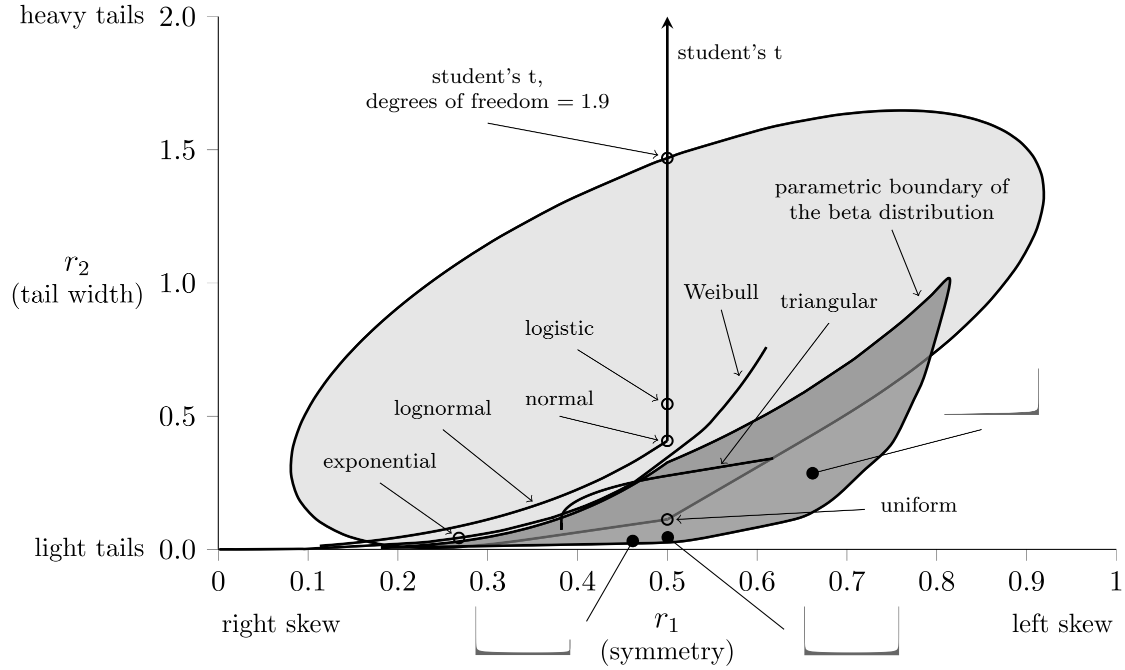

Figure 3.6: Simple Q-normal parametric limits and the limits of some named distributions

also convex. □

The convexity of the simple Q-normal's ovoid facilitates quality control. Imagine a computer program that asks a user for the 1st, 10th, 50th, and 90th quantiles $x \in \mathbf{R}^4$ in order to parameterize the simple Q-normal. Is the β vector in the QPD-feasible set? One might answer this question by exhaustively computing (3.4) using the input quantiles x over a grid of $p \in (0, 1)$ to a desired accuracy. Alternatively, one might 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 pre-computed feasible boundary points must yield a Q-normal probability distribution. The convexity of the ovoid also allows the use of a bisection algorithm for solving the quasi-convex optimization problems of computing these upper and lower limits. See Boyd and Vandenberghe [11] for a discussion on using bisection to solve quasi-convex optimization problems.

3.7 Parameterizing QPDs Using Overdetermined Systems of Equations

Experiments in decision making reveal evidence that a set of probability assessment data can be incoherent [67, 37]—meaning the dataset is inconsistent with the axioms of probability. Spetzler and Staël von Holstein acknowledge that probability encoding procedures may lead to what they term as inconsistencies in data [61]. If a decision analyst makes enough assessments such that the number of quantile-probability pairs exceeds the number of coefficients β, a toolkit is available for finding a QPD that reasonably represents the dataset, whether or not it is incoherent.

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 dimension of β. In such cases, one can use a QPD to provide a smooth representation of the data as an alternative to a histogram. Chapter 6 highlights the use of QPDs in compressing the multivariate data ouput of a probabilistic simulation.

This section illustrates various methods for dealing with such overdetermined systems using the set of quantile-probability data in Table 3.2. Despite an incoherent dataset, a simple Q-normal distribution might exist that decision maker finds satisfactory. Figure 3.7 depicts four examples. Each approach computes the β vector

Table 3.2: A set of inconsistent quantile-probability data

Probability	0.05	0.15	0.20	0.50	0.65	0.80	0.85	0.90
Quantile	0.0	2.5	1.5	4.0	5.0	7.0	6.0	8.0

that minimizes the sum of squares between the QPD's quantile function and the quantile-probability data. A set of m quantile-probability pairs and QPD with n linearly independent basis functions gives a matrix $Y \in \mathbf{R}^{m\times n}$. The matrix for the

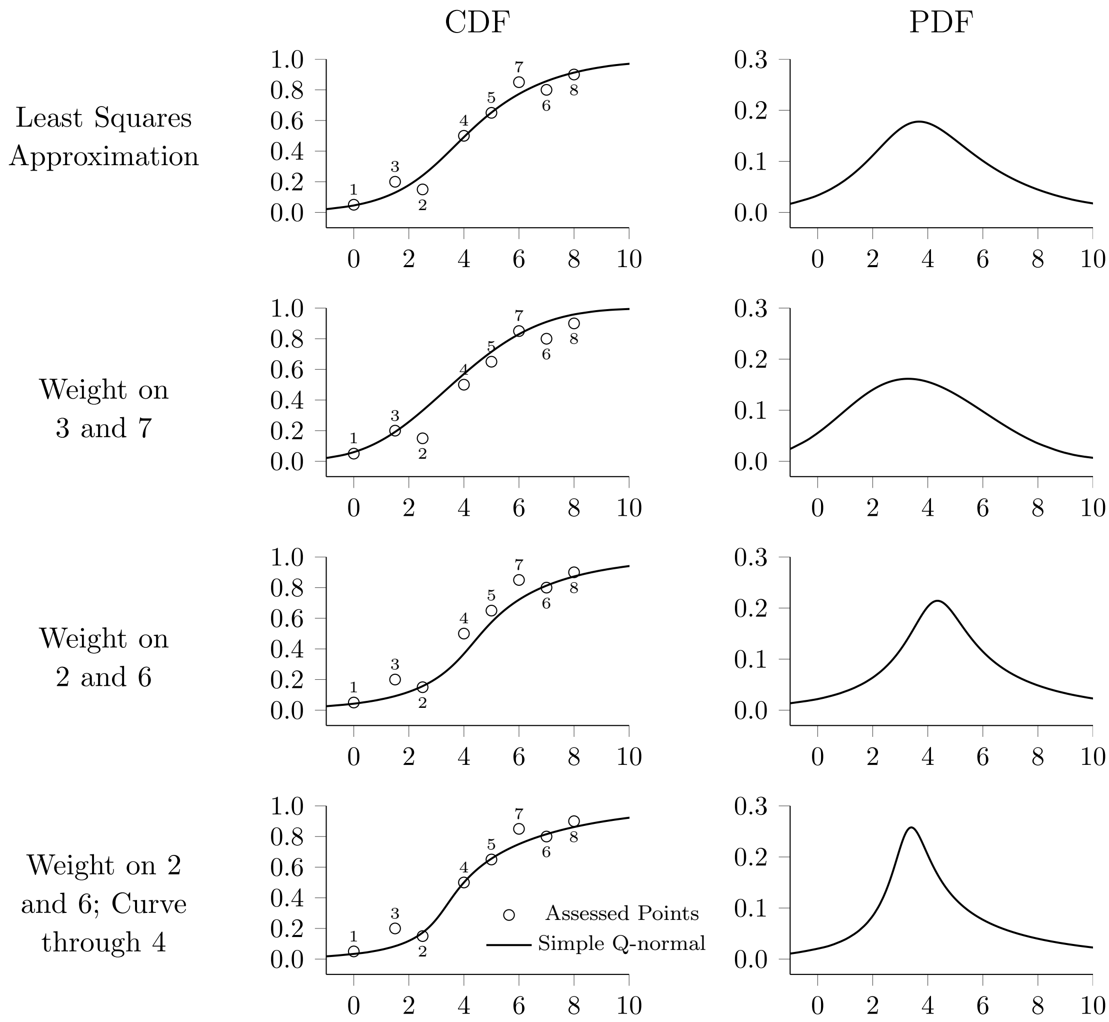

Figure 3.7: Various Q-normal approximations derived from incoherent data

simple Q-normal, built with the probabilities from Table 3.2, is:

$$Y = \begin{bmatrix} 1 & \Phi^{-1}(0.05) & 0.05\Phi^{-1}(0.05) & 0.05 \\ 1 & \Phi^{-1}(0.15) & 0.15\Phi^{-1}(0.15) & 0.15 \\ 1 & \Phi^{-1}(0.20) & 0.20\Phi^{-1}(0.20) & 0.20 \\ 1 & \Phi^{-1}(0.50) & 0.50\Phi^{-1}(0.50) & 0.50 \\ 1 & \Phi^{-1}(0.65) & 0.65\Phi^{-1}(0.65) & 0.65 \\ 1 & \Phi^{-1}(0.80) & 0.80\Phi^{-1}(0.80) & 0.80 \\ 1 & \Phi^{-1}(0.85) & 0.85\Phi^{-1}(0.85) & 0.85 \\ 1 & \Phi^{-1}(0.90) & 0.90\Phi^{-1}(0.90) & 0.90 \end{bmatrix}. \tag{3.15}$$

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

$$\beta = \left(Y^T Y\right)^{-1} Y^T x, \tag{3.16}$$

where x is the vector of quantiles from Table 3.2, and β is the vector of coefficients that specify the quantile function of the simple Q-normal distribution. The simple Q-normal generated by least-squares approximation gives the result shown in the plots on the first row of Figure 3.7.

The second and third rows of plots show how to adjust the simple Q-normal from one extreme of the quantile-probability pairs to the other by applying a weighted least-squares approximation. The second row of plots uses a weighting vector in order to shift the curve toward points 3 and 7. The third row uses weights shifted toward points 2 and 6. The β vector computed from the weighted least-squares approximation is

$$\beta = \left(Y^T W Y\right)^{-1} (WY)^T x,$$

where $W \in \mathbf{R}^m$ is a diagonal matrix whose diagonal elements are the weighting vector applied to each residual. Table 3.3 shows the weights for rows two and three of Figure 3.7.

The final row of plots in Figure 3.7 is a constrained, weighted least-squares approximation. It uses the weighting vector from the third row constrained so that the

Table 3.3: 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

Q-normal passes through the assessed median $(4, 0.5)$, which is point 4 of Figure 3.7. To solve for the vector $\beta \in \mathbf{R}^n$, use the matrix equation

$$\begin{bmatrix} \beta \\ \nu \end{bmatrix} = \begin{bmatrix} 2Y^TWY & c \\ c^T & 0 \end{bmatrix}^{-1} \begin{bmatrix} 2Y^TWx \\ 4 \end{bmatrix},$$

where ν is the Lagrange multiplier for the constraint on the median, and the vector $c = [1, 0, 0, 0.5]^T$ is the vector of basis functions $[1, Q(p), pQ(p), p]$ of equation (3.7) evaluated at $p = 0.5$. These four methods are part of a toolkit for parameterizing the simple Q-normal to blend a set of quantile-probability data. Figure 3.7 shows how the methods lead to diverse CDFs and PDFs. The ability to make such adjustments allows the probability encoder to give feedback in the probability encoding process, and creates a new set of possibilities for probabilistic sensitivity analysis in a decision analysis. For example, a decision analyst can check whether the best alternative changes as the quantile-probability pairs that parameterize a QPD change from one extreme to the other.

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

3.8 Encoding Relevance Between QPDs

The notion of probabilistic dependence, or *relevance* [27] between uncertainties, is an important concept in probability theory. In simplest terms, if observing one uncertain variable changes a person's probability distribution over another, then those two uncertainties are relevant. Relevance is a mutual property so if, for example, an expert's annual precipitation *Forecast* for the coming year in your city is relevant to the sum total inches of *Precipitation* measured in your rain gage over that year, then *Precipitation* is, in turn, relevant to *Forecast*.

One reason for encoding relevance is for model fidelity—accurate representation of the decision maker's beliefs. A second reason is that encoding relevance allows one to update the probability distribution over one uncertainty after observing a relevant uncertain variable. This is a necessary condition for valuing information. If the decision maker were a farmer deciding what crop to plant, how much should he be willing to pay in order to observe *Forecast*, given that such an observation would change his probability distribution over *Precipiatation*? Without encoding relevance, he cannot answer this question. Bayes's theorem governs the updating of information between two uncertainties A and B.

$$P\{B \mid A\} = \frac{P\{B\}P\{A \mid B\}}{P\{A\}}$$

It is a simple equation, but can be difficult to implement for continuous uncertain quantities, except for those whose prior distributions and likelihood functions are conjugate. For a thorough discussion on conjugate distributions, see Raiffa and Schlaifer [48]. Unfortunately, most pairs of QPDs lack the property of conjugacy. When a prior and likelihood function are not conjugate, their posterior distribution $P\{B \mid A\}$ lacks a closed form. The difficulty lies in computing the normalizing constant $P\{A\}$. For arbitrary prior-likelihood combinations, a decision analyst can turn to simulation methods that sample from the posterior distribution whether or not it has a closed form. Methods of simulation include, but are not limited to, Markov-Chain Monte Carlo (MCMC) methods. Because most pairs of QPDs lack the property of conjugacy, Bayesian updating will require a simulation method. See Gelman [16] for an overview of MCMC methods and algorithms for implementing them. See Shachter and Peot [55] for a discussion and comparison of various simulation approaches for probabilistic inference using relevance diagrams.

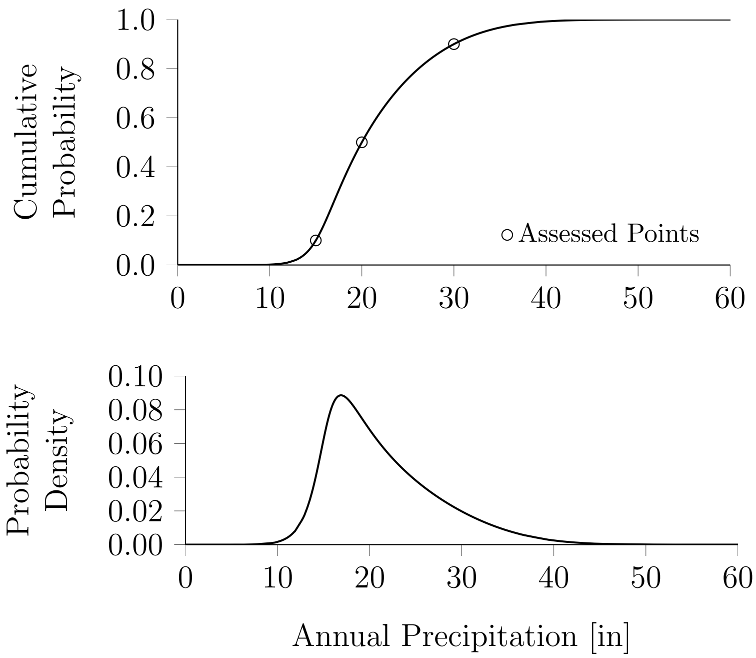

Figure 3.8: A probability distribution over *Precipitation*

Encoding relevance between two continuous uncertain quantities poses a difficulty: complete characterization of their joint probability distribution requires an uncountably infinite number of assessments. As with encoding marginal continuous distributions with a finite number of quantile assessments, encoding joint continuous probability distributions requires modeling approximations.

This section is a brief discussion of practical methods for encoding relevance between continuous uncertain quantities using one or more QPDs. All three approaches begin with a marginal distribution over one continuous uncertainty, *Precipitation*. Figure 3.8 encodes this marginal distribution by parameterizing the QPD whose basis functions are $\{1, \Phi^{-1}(p), p\Phi^{-1}(p)\}$ with the quantile-probability pairs $\{(15, 0.1), (20, 0.5), (30, 0.9)\}$ making $\beta = [20.0\ 3.41\ 4.88]^T$.

3.8.1 Conditioning on a Discretized Marginal Distribution

One approach for encoding relevance is to first discretize the decision maker's continuous distribution over *Precipitation*. Simulating the first five moments of the probability distribution over *Precipitation* and discretizing it using Miller and Rice's algorithm [40] yields the discrete probability distribution whose points of support are 15.8, 24.6, and 37.9 inches with probabilities 0.06, 0.48, and 0.46, respectively. The second step is to assess three distributions for *Forecast*, conditioned on each of the three points of support. Each set of conditional quantiles can parameterize a QPD, for example, with basis functions $\{1, \Phi^{-1}(p), p\Phi^{-1}(p)\}$. Table 3.4 shows these data and Figure 3.9 depicts each continuous distribution over *Forecast*, conditioned on the three values of *Precipitation*. The top image of Figure 3.11 depicts $1,000$ samples from the joint *Precipitation-Forecast* distribution. The key tradeoff is between the accuracy of the discretization approximation and the number of conditional assessments. Another downside is that the number of assessments grows exponentially in the number of conditioning uncertainties.

Table 3.4: Quantile assessments for *Forecast* given *Precipitation*

Forecast | Precip. [in.]→	Quantiles		
Precipitation [in.] ↓	0.10	0.50	0.90
15.8	13	16	21
24.6	20	26	33
37.9	31	39	51

3.8.2 Conditioning on a Continuous Marginal Distribution

A second approach for encoding relevance between variables is to assess continuous probability distributions over one or more of the parameters of a conditional distribution. In this case, the decision analyst may use QPDs in multiple ways. He may encode the conditional distribution over *Forecast* as a QPD, describing one or more of its coefficients as functions of the *Precipitation* uncertainty. Or he may directly encode the decision maker's uncertainty over the quantile assessments of an

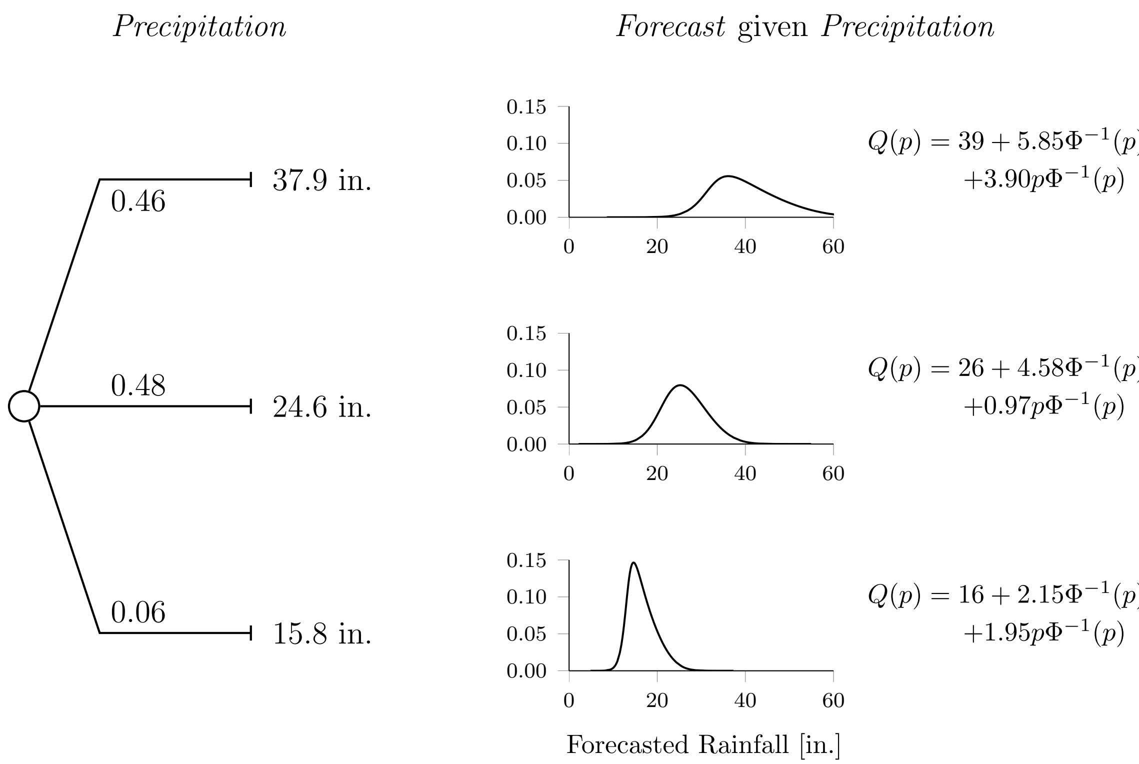

Figure 3.9: *Forecast* conditioned on a discretized distribution over *Precipitation*

expert. The following example demonstrates the former approach; §6.5 demonstrates the latter approach in the application of valuing probability assessment.

The decision analyst can begin with a set of conditional assessments like those in Table 3.4 and use that data to choose a continuous function that maps a particular observation of *Precipitation* to one or more of its parameters. Take as an example the conditional QPD whose quantile function is

$$Q(p \mid R) = R + R/20\left(1 + 4\Phi^{-1}(p) + 2p\Phi^{-1}(p)\right),$$

where the variable R represents a given observation of the *Precipitation* uncertainty. This conditional probability distribution is equivalent to *Forecast* having an additive, heteroskedastic error term described by the QPD with quantile function $Q(p) = (1/20)(1 + 4\Phi^{-1}(p) + 2p\Phi^{-1}(p))$, scaled by *Precipitation*. The middle image of Figure

3.11 shows $1,000$ points sampled from the joint probability distributions generated by the conditional probability distribution described in this subsection.

The advantage to the method of encoding a probability distribution conditioned on a continuous marginal distribution is the same as the advantage of encoding a continuous marginal distribution over a naturally continuous variable—fidelity to the decision maker's beliefs. However, there may be difficulty in choosing a conditional probability distribution that is flexible enough to adequately reflect the decision maker's knowledge while being reasonably parameterizable.

3.8.3 Relating Uncertain Quantities with Copulas

Clemen and Reilly [13] write, "A disadvantage with the typical marginal-and-conditional approach is that the required number of probability assessments can grow exponentially with the number of variables." A third approach is to encode the decision maker's knowledge of the second uncertainty as a marginal QPD and separately encode relevance. The decision analyst now has two marginal continuous probability distributions, but no function for relating them. Copulas are a class of such functions. Sklar's theorem [43, page 21] shows that one can express any joint cumulative distribution function H as a copula C, which is a function that maps two or more marginal cumulative probabilities to a joint probability.

$$H(x_1, \ldots, x_n) = C\left(F(x_1), \ldots, F(x_n)\right)$$

As a functional form, copulas work particularly well with QPDs because a) copulas accept arbitrary marginal distributions and b) both quantile functions and copulas take cumulative probabilities as arguments.

To use the copula method, the decision analyst must encode the decision maker's marginal distribution over *Forecast*. Figure 3.10 encodes this marginal distribution by parameterizing the QPD whose basis functions are $\{1, \Phi^{-1}(p), p\Phi^{-1}(p)\}$ with the quantile-probability pairs $\{(14, 0.1), (21, 0.5), (35, 0.9)\}$ making $\beta = [21.0\ 4.78\ 6.83]^T$.

The next step is to choose a copula and parameterize it in a manner consistent with the decision maker's knowledge relating *Precipitation* and *Forecast*. The decision

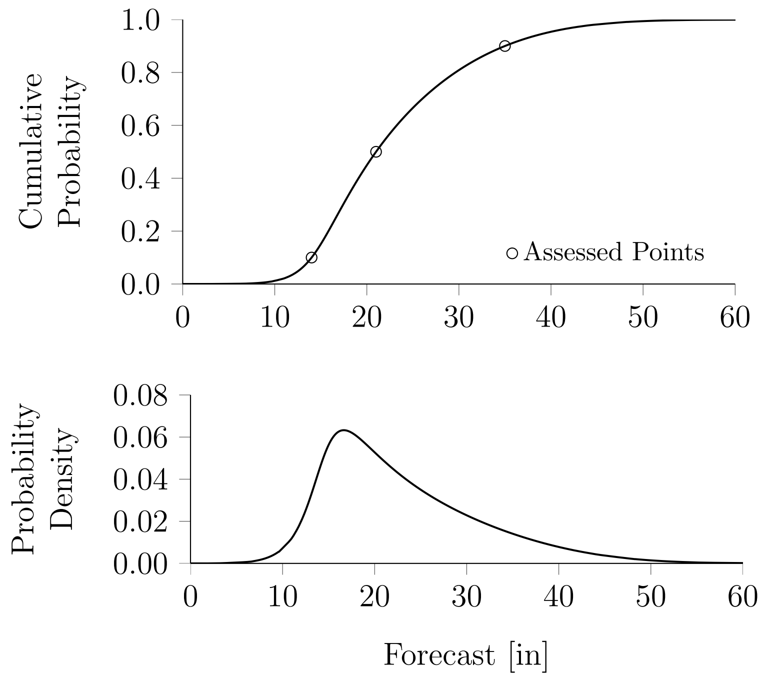

Figure 3.10: A probability distribution over *Forecast*

analyst chooses the multivariate Gaussian copula—perhaps the most commonly used copula. In order to parameterize it, he chooses the approach detailed by Clemen, Fischer, and Winkler [12]. They present an empirical study testing six assessment methods, including *conditional fractile* and *concordance probability*, which one can use to estimate two pairwise rank-correlation coefficients, Spearman's ρ and Kendall's τ, respectively. One can use either of these rank-correlation coefficents to compute the corresponding Pearson's correlation coefficients r_{ij} necessary for creating the correlation matrix for a Gaussian copula. These assessment methods require conditional expectations to be monotone in the conditioning variable (e.g., no $\cup$ or $\cap$-shaped relationships).

$$r_{ij} = 2\sin\left(\frac{\pi\rho_{ij}}{6}\right) \tag{3.17}$$

$$r_{ij} = \sin\left(\frac{\pi\tau_{ij}}{2}\right)$$

Through a series of assessment questions like those used in the study of Clemen et al., the decision analyst encodes Spearman's ρ as 0.67 and computes the copula's covariance matrix shown in (3.18) using (3.17).

$$\Sigma = \begin{bmatrix} 1 & 0.65 \\ 0.65 & 1 \end{bmatrix} \tag{3.18}$$

The bottom image of Figure 3.11 depicts 1,000 samples from the joint *Precipitation–Forecast* distribution. The advantage to a copula-based approach is that it requires relatively few parameters to describe complex relationships between uncertainties. Clemen, Fischer, and Winkler note that their subjects performed well at assessing correlation directly, and that much work could be done to frame assessment questions and train decision makers in a way that might improve assessments. There remains another disadvantage of the copula method: the problem of assessing a marginal distribution that may not be easy to think about—the reading of a detector, for example. A more natural assessment might be one conditioned on *Precipitation*, lending itself to one of the first two methods of this section.

3.9 Engineering QPDs

A decision maker's declaration that a QPD represents her uncertainty is a simultaneous declaration that its set of basis functions is satisfactory. Beyond that, this chapter offers no axiomatic foundation for choosing a QPD's basis functions. The choice of basis functions does, however, affect the characteristics of a QPD. Return to the simple Q-normal as an example. The first basis function $g_0(p) = 1$ makes β_1 a location parameter according to Definition 4; $g_1(p) = \Phi^{-1}(p)$ makes the support of the simple Q-normal the real number line; $g_2(p) = p\Phi^{-1}(p)$ imparts the element of skewness through β_3; and $g_3(p) = p$ exerts an effect tail heaviness through β_4.

Since a QPD's basis functions determine many of its characteristics, choosing a set of basis functions is one way to engineer a QPD. The next two chapters focus on two key elements of a probability distribution that one may want to engineer. Chapter 4 introduces tools for engineering the support of a QPD, and Chapter 5 introduces a theory for engineering its tail behavior.

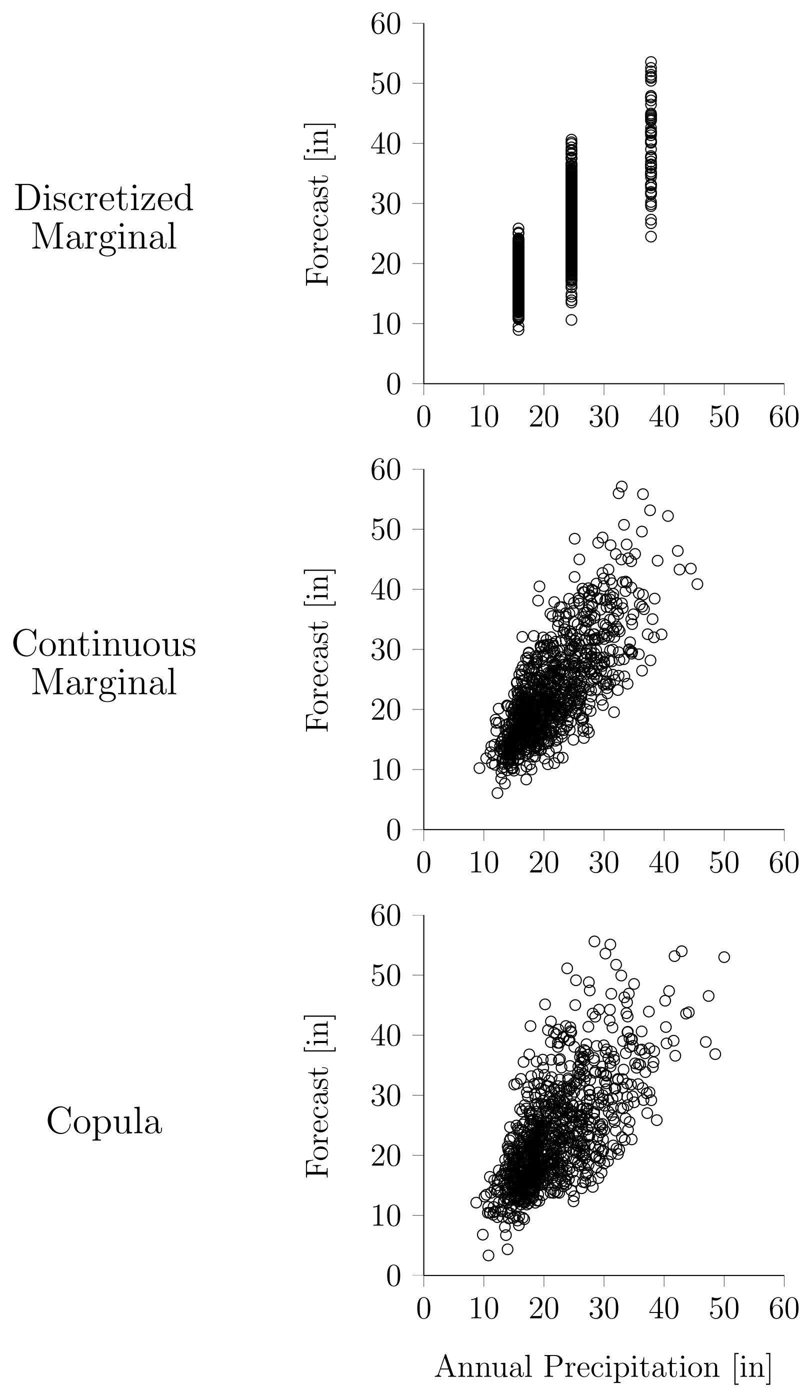

Figure 3.11: One thousand samples from the joint probability distributions over *Forecast* and *Rain* compare the methods of discretized and continuous marginal distributions and copulas

Chapter 4

Engineering the Support of a Quantile-Parameterized Distribution

Many continuous uncertain quantities have one or more well-defined limits. For example, the volume of recoverable oil in a reservoir can't be less than zero barrels, and the market share of a product must be between zero and one hundred percent inclusive. When encoding a decision maker's knowledge about a continuous uncertain quantity, it is important to consider the range of possible values that it can take. This range of possible values is called the *support*.

Definition 9. *The support of a continuous probability distribution F, denoted* $\mathrm{supp}(F)$*, is the subset of the domain of its PDF that has nonzero density,* $\{x \in \mathbf{R} \mid f(x) > 0\}$.

Proposition 19. *The support of a continuous probability distribution F with quantile function $Q(p)$ is its image $\{Q(p) \mid p \in (0,1)\}$.*

Proof. By the definition of support,

$$\begin{aligned} & \{x \in \mathbf{R} \mid f(x) > 0\} \\ \Leftrightarrow\ & \{Q(p) \mid f(Q(p)) > 0, p \in (0,1)\} \\ \Leftrightarrow\ & \left\{Q(p) \;\middle|\; \frac{1}{Q'(p)} > 0, p \in (0,1)\right\} \\ \Leftrightarrow\ & \{Q(p) \mid p \in (0,1)\} \end{aligned}$$

The first equivalence is the substitution $x = Q(p)$; the second equivalence is the direct application of (2.2). The third equivalence holds because $Q(p)$ is the quantile function of a continuous probability distribution F; therefore, $Q'(p) > 0$ for $p \in (0,1)$, and $\frac{1}{Q'(p)} > 0$ for all $p \in (0,1)$. □

4.1 How Basis Functions Affect a QPD's Support

Because of Proposition 19, one need look no further than the image of a continuous probability distribution's quantile function in order to determine that distribution's support. This result naturally extends the discussion of support to QPDs. The choice of basis functions has implications on the support of a QPD before the decision analyst makes a single probability assessment.

Theorem 2 (QPD support theorem)**.** *Given the set $X_i = \{g_i(p) \mid p \in (0,1)\}$ is the image of the ith basis function $g_i(p)$ of a QPD F, the following statements are true.*

1. *X_i is a finite interval for **all** $i \in \{1 : n\}$ if and only if* supp(F) *is a finite interval.*

2. *X_i is a semi-infinite interval for **exactly** one basis function i, and X_j is a finite interval for all $j \neq i$ implies supp(F) is semi-infinite.*

3. *$X_i = (-\infty, \infty)$ for **any** $i \in \{1 : n\}$ implies* $\text{supp}(F) = (-\infty, \infty)$.

See the appendix for the proof of the QPD support theorem. Note that its statements hold, regardless of QPD-feasible vectors $\beta \in \mathbf{R}^n$. For a QPD whose support is a finite interval, Theorem 2 mandates that the image of each of its basis functions be finite, for example $\{1, p, p^2\}$. Likewise, for a QPD with semi-infinite support begin with a set of $n-1$ basis functions whose images are finite intervals and add one basis function whose image is a semi-infinite interval, for example $\{1, p, p^2, \log(p)\}$. Finally, for a QPD whose support is $(-\infty, \infty)$, Theorem 2 suggests choosing at least one basis function whose image is $(-\infty, \infty)$, for example $\{1, p, p^2, \log(p), \Phi^{-1}(p)\}$. The choice of basis functions grants a useful but limited control over a QPD's support. Precise support control is available through a different set of tools that I will introduce later in this chapter.

The theory that follows makes no effort to lengthen an uncertainty's support from a semi-infinite or finite interval into a semi-infinite or infinite interval. Well-behaved basis functions that make for QPDs whose support is a semi-infinite or infinite interval are readily available. The more useful way to engineer the support of a probability distribution is to shorten it from a semi-infinite or infinite interval into a specific finite or semi-infinite interval. The focus of this section is to show methods for turning a QPD with a given support into a new QPD whose support is a precise subset of the original's. For example, a decision maker who is uncertain about the fraction of current customers that will cancel a given subscription in the coming year knows for certain that the value must lie between zero and one, inclusive. All fractions outside of the unit interval should be assigned zero probability density. A QPD such as the shifted version of Gilchrist's skew-logistic, whose basis functions are $\{1, \log(p), -\log(1-p)\}$, will not meet this criterion because its support is the real line. Figure 4.1 depicts this QPD parameterized by the set of quantile-probability pairs $\{(0.1, 0.1), (0.4, 0.5), (0.8, 0.9)\}$, making its quantile function $Q(p) = 0.33 + 0.11\log(p) + 0.21\,(-\log(1-p))$.

4.2 Defining Support with Extreme Quantiles

It is tempting to think that choosing a QPD whose basis functions map to the interval $(0, 1)$ would lead to a QPD whose support is also $(0, 1)$. Unfortunately, this is not true. Theorem 2 is the most general statement I have regarding such a QPD: its support

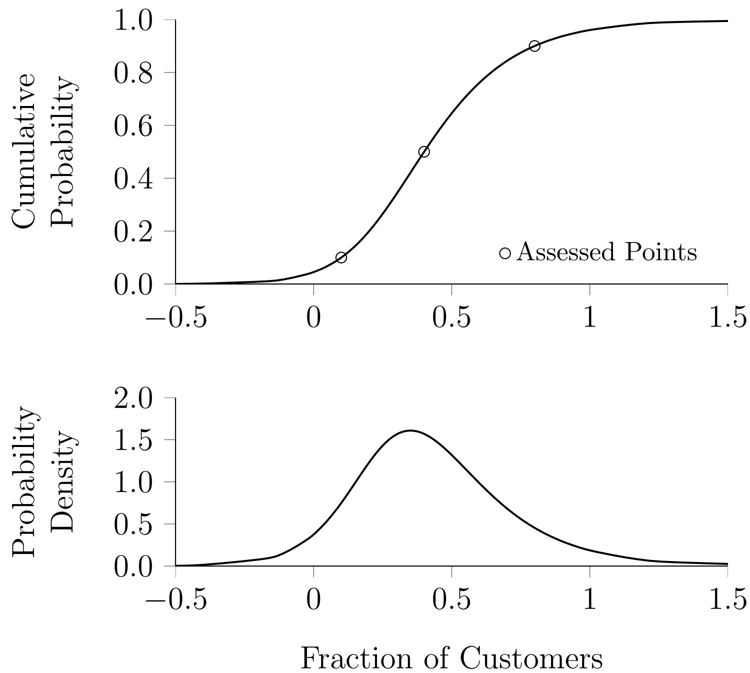

Figure 4.1: A QPD with infinite support

is a finite interval. As an example, take the QPD with quantile function $Q(p) = 5Q_{B1}(p) + 2Q_{B2}(p) + Q_{B3}(p)$, where $Q_{B1}(p), Q_{B2}(p)$ and $Q_{B3}(p)$ are the quantile function of the beta distribution with (α, β) parameters $(1,5), (3,3)$, and $(5,1)$. The three basis functions $\{Q_{B1}(p), Q_{B2}(p), Q_{B3}(p)\}$ are each supported on $(0,1)$, while $Q(p)$ has support $(0,8)$.

An alternative is to assess the extreme quantiles. Doing so makes the assessed quantiles of the *Fraction of Customers* uncertainty become $\{(0,0), (0.1,0.1), (0.4,0.5), (0.8,0.9), (1,1)\}$. Returning to the basis functions $\{Q_{B1}(p),\ Q_{B2}(p),\ Q_{B3}(p)\}$, the system is now overdetermined—five sets of quantile-probability pairs and three basis functions. From here, a decision analyst can choose two new basis functions, each of whose images is a finite interval, making sure that the resulting set of basis functions is regular. Or one can use a constrained least-squares approximation similar to §3.7 with the constraint that the QPD must pass through both extreme quantiles. This is equivalent to a quadratic program with β constrained to the three-dimensional simplex,

$$\begin{aligned}
\underset{\beta}{\text{minimize}} \quad & (Y\beta - x)^T (Y\beta - x) \\
\text{subject to} \quad & \sum_{i=1}^{3} \beta_i = 1 \\
& \beta_i \geq 0, i = 1, \ldots, 3,
\end{aligned}$$

where $x = (0.1, 0.4, 0.8)$ and

$$Y = \begin{bmatrix} Q_{B1}(0.1) & Q_{B2}(0.1) & Q_{B3}(0.1) \\ Q_{B1}(0.5) & Q_{B2}(0.5) & Q_{B3}(0.5) \\ Q_{B1}(0.9) & Q_{B2}(0.9) & Q_{B3}(0.9) \end{bmatrix}.$$

This constrained least-squares approximation yields the QPD $Q(p) = 0.16Q_{B1}(p) + 0.84Q_{B2}(p)$ shown in Figure 4.2. There is negligible weight on the third basis function $Q_{B3}(p)$. The downside of this method is that it is not guaranteed to pass through the original three quantile-probability pairs.

4.3 Defining Support by Truncation

A more straightforward approach for defining the support of a QPD, or a quantile function in general, is to begin with a probability distribution whose support extends too far and then truncate its support by "chopping off" one or both of its tails. To truncate the support of a distribution F with quantile function $Q(p)$, apply the affine transform

$$\hat{p} = p \cdot (b - a) + a; \quad a, b \in (0, 1) \tag{4.1}$$

to the cumulative probability p. The resulting quantile function $Q(\hat{p}) \equiv \hat{Q}(p)$ represents the probability distribution $\hat{F}$, whose support is $(Q(a), Q(b))$. This is the inverse of the conditional probability transform of Proposition 6. When using this truncation method on a QPD, apply the p-transform after first computing the coefficients β. This preserves the quantile function's linearity in its basis functions, maintaining the

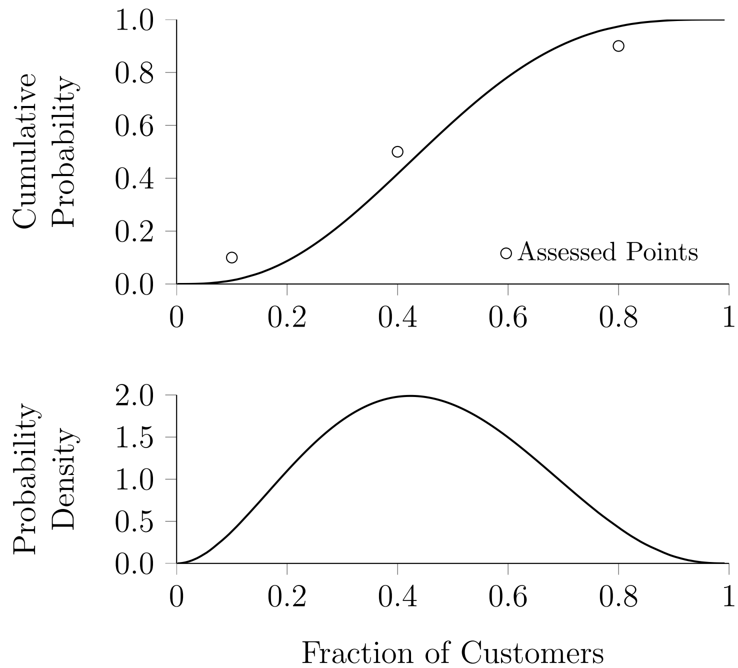

Figure 4.2: A QPD with $(0, 1)$ support computed via constrained least-squares approximation

ease of computing the coefficients β from a set of assessed quantiles.

Take as an example the QPD of Figure 4.1. To truncate this distribution, begin by transforming the cumulative probability p of its quantile function $Q(p) = 0.33 + 0.11\log(p) + 0.21(-\log(1-p))$. In order to apply the transform given by (4.1), compute the cumulative probabilities associated with the endpoints of the desired support, which is $(0, 1)$. This makes the constant $a = F(0)$ and the constant $b = F(1)$. By iteration, a and b are approximately 0.05 and 0.96, respectively. The resulting quantile function of the truncated distribution $\hat{F}$ shown in Figure 4.3 is $\hat{Q}(p) = 0.33 + 0.11\log(0.91p + 0.05) + 0.21\left(-\log\left(1 - (0.91p + 0.05)\right)\right)$.

Two nice features of this approach are: 1) the coefficients β need not be recalculated, and 2) the p-transformation renormalizes the PDF so that it sums to unity. Figure 4.3 higlights the two negative features of the truncation method 1) its distribution is no longer guaranteed to pass through the assessed quantiles, and 2) its chopped-off tails may be a poor representation of the decision maker's knowledge.

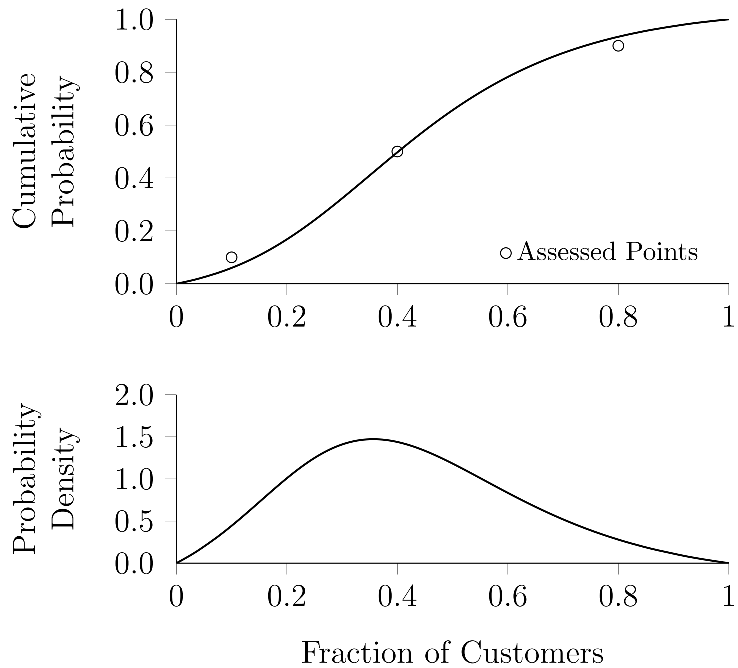

Figure 4.3: A truncated QPD with finite support

Chopping off more tail probability can worsen these negative factors. The modeler must decide whether the convenience of the truncation method outweighs the loss in fidelity of the resulting probability distribution.

When modeling, one can simulate from the p-transformed distribution $\hat{F}$ using the inverse-transform method. Equivalently, one can simulate from F and reject any sampled value that lies outside the interval $(Q(a), Q(b))$. Although this rejection sampling method is straightforward, rejecting simulates comes with a computational cost. The number of rejections, and therefore the cost, is proportional to the amount of tail probability that is chopped off. This fact, along with the aforementioned two negative features, suggests that the truncation method is best when the total probability truncated from the original distribution is small.

4.4 Defining Support by Quantile Function Transformation

Instead of transforming p, one can control an uncertainty's support by applying a transform to $Q(p)$, the quantile function itself. According to Proposition 5, if the transforming function is nondecreasing, it maps one quantile function to another quantile function. And Proposition 17 states that if the transforming function h is strictly increasing and continuously differentiable (SICD) with an SICD inverse h^{-1}, it maps the quantile function of one QPD to the quantile function of another QPD. This section makes use of these results to develop tools for transforming quantile functions, with a focus on QPDs. It uses the convention $h : \mathbf{R} \to \mathbf{R}$ is an SICD function with SICD inverse h^{-1}.

One can solve for the β coefficients of a QPD created by transforming the quantile function of another QPD using the method introduced in the Quantile Parameters Theorem (theorem 1), namely:

$$\tilde{\beta} = Y^{-1}\tilde{x}, \tag{4.2}$$

Where $a, \tilde{x} \in \mathbf{R}^n$, and

$$\tilde{x} = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_n) \end{bmatrix}, \tag{4.3}$$

The matrix Y is defined by (3.11).

Continuing with the earlier example from this chapter, one can represent the decision maker's uncertainty about the fraction of customers canceling a subscription in the coming year by transforming a QPD. Beginning with the same basis functions $\{1, \log(p), -\log(1-p)\}$ and applying a logit transform yields the equation for the coefficients $\tilde{\beta}$.

$$\tilde{\beta} = \begin{bmatrix} 1 & \log(0.1) & -\log(1-0.1) \\ 1 & \log(0.5) & -\log(1-0.5) \\ 1 & \log(0.9) & -\log(1-0.9) \end{bmatrix}^{-1} \begin{bmatrix} \log\left(\frac{0.1}{1-0.1}\right) \\ \log\left(\frac{0.4}{1-0.4}\right) \\ \log\left(\frac{0.8}{1-0.8}\right) \end{bmatrix}$$

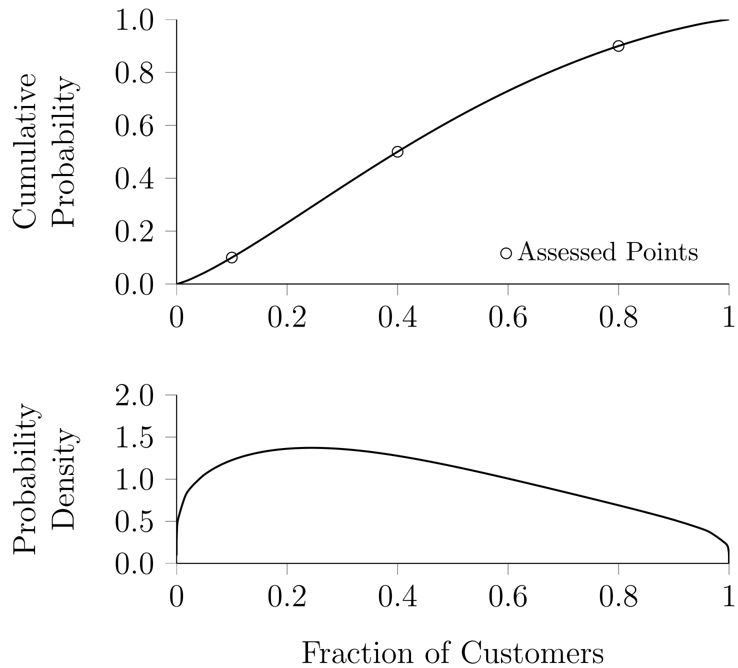

Figure 4.4: A logit-transformed QPD with finite support

Figure 4.4 shows this transformed QPD and its $(0, 1)$ support. Its quantile function is $Q(p) = L(-0.41 + 0.82 \log(p) + 0.82(-\log(1 - p)))$, where the function $L(x) = \exp(x)/(1 + \exp(x))$. It shows a distribution whose PDF has a very different shape from those in Figures 4.1 and 4.3. The appropriate method for engineering the support of a probability distribution is the one that the decision maker declares as an appropriate representation of her knowledge. Transforming a QPD allows one to achieve the desired support while passing through the assessed quantiles. See Table 4.1 for some more useful transforms that control support. Recall from Corollary 2 that the pPDF of a transformed quantile function requires an additional transform $f(\tilde{Q}(p)) = f(Q(p))/(h^{-1})'(Q(p))$. Table 4.2 details the effects of various transforms on a probability distribution's pPDF.

Applying a transform to the quantile function of a QPD is the last tool for encoding continuous uncertain quantities with well-defined limits. Occasionally, a decision maker might not want to assign precise limits. The next chapter introduces a theory

Table 4.1: Some useful transforms for controlling support

name	$h(x)$	old support	new support
positive affine	$cx + d, c > 0$	$[a, b]$	$[ac + d, bc + d]$
logarithmic	$\log(x)$	$(-\infty, \infty)$	$(0, \infty)$
reflected logarithmic	$\log(-x)$	$(-\infty, \infty)$	$(-\infty, 0)$
logit	$\log\left(\frac{x}{1-x}\right)$	$(-\infty, \infty)$	$(0, 1)$
probit	$\Phi^{-1}(x)$	$(-\infty, \infty)$	$(0, 1)$

Table 4.2: Transform effects on the pPDF

transform $h(\cdot)$	$\tilde{Q}(p)$	pPDF $f(\tilde{Q}'(p))$
positive affine	$\frac{Q(p)-a}{b}$	$b \cdot f(Q(p))$
logarithmic	$\exp(Q(p))$	$\frac{1}{\exp(Q(p))} f(Q(p))$
reflected logarithmic	$-\exp(Q(p))$	$\frac{1}{\exp(Q(p))} f(Q(p))$
logit	$\frac{\exp(Q(p))}{1+\exp(Q(p))}$	$\frac{(1+\exp(Q(p)))^2}{\exp(Q(p))} f(Q(p))$
probit	$\Phi(Q(p))$	$\frac{1}{\phi(Q(p))} f(Q(p))$

tailored to help a decision analyst think clearly when encoding a probability distribution in the absence of bounds.

Chapter 5

A Theory of Tail Behavior

One can argue that precise bounds for a continuous uncertain quantity might not naturally exist, or they might not be available because a decision maker simply does not want to expend the energy to assign them. Regardless, it can be useful for the decision analyst to encode the decision maker's probability distribution with one or more infinite tails. However, an infinite tail of one probability distribution might be heavier than that of another, adding a wrinkle to probability encoding. This chapter defines *heavier tails* and introduces a theory of tail behavior tailored to help a decision analyst encode a probability distribution with one or more infinite tails.

5.1 Motivating a Theory of Tail Behavior

A decision maker might not want to assign precise limits to a continuous uncertain quantity, but she may be willing to assign extreme quantiles, thereby giving evidence as to the tail heaviness of its probability distribution. There are two approaches to assessing quantiles that lie in the tails of a decision maker's distribution—those less than the 10th or greater than the 90th quantiles. The first approach is to ask quantile-probability assessment questions directly. Spetzler and Staël von Holstein [61, page 349] note that their probability wheel encoding method has a disadvantage in direct assessment of the tails of a distribution:

> . . . because it is difficult for a subject to discriminate between the sizes of

> very small sectors, the wheel is most useful for evaluating probabilities in the range from 0.1 to 0.9.

Recent work in the field of risk communication for medical decision making addresses this problem with some graphical techniques for communicating low-frequency events [14, 71, 19]. Although graphics such as these show promise for probability elicitation, they remain untested. Since Spetzler and Staël von Holstein's standard techniques don't perform well for rare events, such as those described by the tail of a continuous probability distribution, they recommend a second approach: create a more detailed probabilistic model with uncertainties that do not describe low-probability events.

> ...our experience with probability encoding for rare events indicates that probabilistic modeling is generally more effective than direct encoding. For example, in order for an event to occur it may be necessary that a sequence of other events occur. These intermediate events may not be low-probability events and standard encoding procedures can then be used.

Their recommendation of adding more distinctions to a model is an expansion of the decision maker's frame. Sensitivity analysis is a phase of the decision analysis cycle that allows a decision analyst to test whether the existing frame merits expansion. A desideratum, which I will demonstrate in §6.4, is to describe a method that enables a sensitivity analysis to tail behavior. Having a notion of how sensitive a decision is to the tail behavior of one of the decision maker's uncertainties can help an analyst recommend whether or not to add complexity to a model and also whether or not to spend resources assessing a rare-event probability. In order to test sensitivity to tail behavior, one must first have a method for classifying tail behavior.

5.2 Characterizing Tail Behavior

Research on the tail behavior of probability distributions comes from the literature of nonparametric statistics, specifically the study of rank tests. This research consists of both asymptotic classifiers and nonasymptotic orderings. Two asymptotic classifiers of note include the Parzen [44] and the refined Parzen [52]. To compute the right-tailed Parzen exponent classifier, rewrite the pPDF of a probability distribution as a function of the form

$$f\left(Q(p)\right) = L(p)(1-p)^{\alpha}, \tag{5.1}$$

where the function $L(p)$ is a slowly-varying function, defined as $\lim_{u \to 1^-} L(\lambda p)/L(p) = 1$ for all $\lambda > 0$. Parzen uses the sign of the exponent $\alpha[<] = [>]0$ to determine the three classifications [short] medium [long] tails. Substituting $1-p$ for p in equation (5.1) and taking the limit as $p \to 0^+$ gives the left tail classifier. One complaint regarding this classification is that it does not discriminate between normal and exponential tails (both are medium-tailed). Schuster addresses this problem with his refined Parzen classifier, which uses the hazard function as a means of breaking up the medium tail category into three components: medium-short, medium-medium, and medium-long.

The nonasymptotic tail characterization methods predate the asymptotic. All of the methods that follow induce a partial ordering[1] over a set of univariate probability distributions with certain regularity conditions—some involve first and second derivatives and therefore apply only to probability distributions in which these derivatives are defined.

Definition 10. *A binary relation $\preceq$ over a set $\mathcal{X}$ is a partial ordering if it meets the following criteria for all $F, G, H \in \mathcal{X}$:*

Reflexivity: $F \preceq F$

Antisymmetry: If $F \preceq G$ and $G \preceq F$ then F is G, written $F \sim G$

Transitivity: If $F \preceq G$ and $G \preceq H$ then $F \preceq H$

Definition 11. *Given a partial ordering $\preceq$ over a set $\mathcal{X}$, the binary relation $F \prec G$ holds when $F \preceq G$ and $G \not\preceq F$, $F, G \in \mathcal{X}$.*

[1] In decision analysis, a familiar partial ordering of probability distributions is the binary relation over a set of deals that defines second-order stochastic dominance.

In contrast to a partial ordering, a *total* ordering makes the additional requirement of completeness: either $F \preceq G$ or $G \preceq F$ (or both) for all $F, G \in \mathcal{X}$. Because a partial ordering lacks this requirement, and because all of the tail classifiers that follow are partial orderings, one can find probability distributions that one or more of the following three methods cannot order.

In a study of convex transformations of random variables, van Zwet [68, Ch.4] introduces a binary relation that partially orders a set of probability distributions $\mathcal{F}$, according to their tail behavior. He defines this relationship by choosing two probability distributions $F, G \in \mathcal{F}$ over a given domain; where $\mathcal{F}$ [68, pg. 24] is the subset of univariate probability distributions whose quantile function $Q(p)$ is twice differentiable, whose second derivative $Q''(p)$ is continuous, whose first derivative $Q'(p) > 0$, and there exist nonnegative integers a and b such that $|Q(p)p^a(1-p)^b|$ is bounded for $p \in (0, 1)$. The last condition is to ensure that probability distributions have defined order statistics. Van Zwet defines c-ordering as $F \prec_c G$ if and only if the function $Q_G(F(x))$ is convex over I, the support of F. In his definition, $F(x)$ and $G(x)$ are the respective CDFs of the probability distributions F and G, and $Q_G(p)$ is the quantile function of G. Van Zwet goes on to focus on symmetric distributions and applies his ordering to the domain $I = (0.5, 1)$ using the notation $\prec_s$. The same index function $Q''(p)/Q'(p)$ applies to van Zwet's s-ordering.

Hájek [18, pg.150] uses a different tail function to generate a partial order over $\mathcal{H}$, the set of symmetric, log-concave probability distributions centered about zero. He defines a binary relation $F \prec_H G$ if and only if the function $a(p) = Q_G(p)/Q_F(p)$ is nondecreasing and nonconstant over the interval $(0.5, 1)$. In his paper, Hájek investigates the relationship between the score function $\varphi(p) = -\frac{f'(Q(p))}{f(Q(p))}$ and tail ordering.[2] For $F, G \in \mathcal{H}$, he shows that if the ratio of score functions $\varphi_G(p)/\varphi_F(p)$ between two probability distributions is nondecreasing, then $F \prec_s G$, which implies that $F \prec_H G$.

Gastwirth [15] uses Hájek's ratio of score functions as a new partial ordering. The probability distribution $F \prec_G G$ if and only if $b(p) = \varphi_G(p)/\varphi_F(p)$ is nondecreasing and nonconstant over the interval $(0.5, 1)$. For probability distributions with

[2] Applying the score function to a cumulative distribution function $\varphi(F(x))$ yields the Arrow-Pratt risk aversion function.

symmetric log-concave densities centered about zero and twice differentiable quantile functions whose second derivative is continuous, Gastwirth ordering implies van Zwet's s-ordering implies Hájek ordering. Gastwirth notes that the converse is not true. The question of which ordering is most useful when considering candidate distributions for decision analysis remains.

5.3 Defining a Binary Relation for Tail Behavior

Before defining a binary relation that distinguishes whether one probability distribution has heavier tails than another, it is important to first clarify what *heavier tails* means.

Definition 12. *The probability distribution G has a heavier right tail than F when there exists a point $\hat{x} \in \mathbf{R}$ such that $G(x) < F(x)$, all $x \in (\hat{x}, \infty)$. Likewise, the probability distribution G has a heavier left tail than F when there exists a point $\hat{x} \in \mathbf{R}$ such that $F(x) < G(x)$, all $x \in (-\infty, \hat{x})$.*

For the purposes of decision analysis, it is not enough to say that G has heavier right [left] tails than F in the limit when there is no quantile data in a large interval like $p \in (0.9, 1)$. Therefore, I remove the asymptotic classifiers from consideration since the Parzen and refined Parzen classifiers consider tail behavior only as $p \to 1$ [$p \to 0$ for left tails]. In contrast, the van Zwet, Hájek, and Gastwirth orderings apply to an interval $p \in (0.5, 1)$. Unfortunately, the set of probability distributions over which the van Zwet ($\mathcal{F}$), Hájek, and Gastwirth ($\mathcal{H}$) orderings apply is too restrictive. Decision analysts do not confine themselves to the use of the symmetric, log-concave distributions of $\mathcal{H}$, and they do not distinguish distributions with defined order statistics as $\mathcal{F}$ stipulates. Since none of the preceding tail classifiers is satisfactory, I introduce two new binary relations $\prec_R$ and $\prec_L$ based on van Zwet ordering. One can apply these relations to the set $\mathcal{D}$ of univariate probability distributions whose quantile function $Q(p)$ is twice differentiable, whose second derivative $Q''(p)$ is continuous, and whose first derivative $Q'(p)$ is positive. See Table 5.1 for a comparison of the various sets of probability distributions over which the various tail orderings apply. The definition

Table 5.1: Sets of probability distributions relevant to tail classification

set	description	tail classification
$\mathcal{F}$	$Q''(p)$ exists and is continuous; $Q'(p) > 0$; $\lvert Q(p)p^a(1-p)^b \rvert$ is bounded for $p \in (0,1)$ for some nonnegative integers a and b	van Zwet
$\mathcal{H}$	$Q''(p)$ exists and is continuous; PDFs are log-concave	Hájek, Gastwirth
$\mathcal{D}$	$Q''(p)$ exists and is continuous; $Q'(p) > 0$	R ordering, L ordering

of these new binary relations uses van Zwet's function

$$\zeta(x) = Q_G(F(x)). \tag{5.2}$$

Definition 13. *The probability distribution $F \prec_R G$, for $F, G \in \mathcal{D}$ when there exist two points $x_0, x_1 \in \mathbf{R}$ such that $\zeta(x)$ is convex over $x \in (x_0, \infty)$, $\zeta'(x)|_{x_1} > 1$, and $x_1 \in (x_0, \infty)$.*

Definition 14. *The probability distribution $F \prec_L G$, for $F, G \in \mathcal{D}$ when there exist two points $x_0, x_1 \in \mathbf{R}$ such that $\zeta(x)$ is concave over $x \in (-\infty, x_0)$, $\zeta'(x)|_{x_1} > 1$, and $x_1 \in (-\infty, x_0)$.*

These definitions differ from van Zwet's, Hájek's, and Gastwirth's binary relations in a few other important ways. First, they separate the ordering of left and right tails.[3] Second, they do not restrict themselves to the domain $p \in (0.5, 1)$. This allows one to focus only on the tail when comparing two distributions. Finally, they carry an important implication.

Proposition 20. *If $F \prec_R [\prec_L] G$, then G has heavier right [left] tails than F.*

[3] A note on notation: the results that follow use the convention of making explicit statements for R-ordering while including the corresponding L-ordering statements in square brackets. The proofs cover only R-ordering—the L-ordering proofs follow by argument of symmetry.

Proof. It suffices to show that there exists a point $\hat{x} \in \mathbf{R}$ such that $F(x) > G(x)$, all $x \in (\hat{x}, \infty)$. By definition of $\prec_R$, $\zeta'(x)|_{x_1} > 1$, $x_1 \in (x_0, \infty)$, and $\zeta(x)$ is convex. Therefore, $\zeta(x) > x$, $x \in (x_1, \infty)$ because $\zeta(x)$ is convex and increasing. Let $\hat{x} = x_1$, and apply G to both sides of $\zeta(x) > x$ to yield $F(x) > G(x)$, $x \in (\hat{x}, \infty)$. □

The binary relation $\prec_R$ [$\prec_L$] does not form a partial ordering because it lacks an equivalence relation. Defining the equivalence relation $\sim_R$ [$\sim_L$] and combining it with the binary relation $\prec_R$ [$\prec_L$] creates a partial ordering over the set $\mathcal{D}$. I refer to this ordering as *R-ordering* [*L-ordering*].

Definition 15. *The probability distribution $F \sim_R G$, [$F \sim_L G$], for $F, G \in \mathcal{D}$ when there exists a point $p_0 \in (0, 1)$ such that $Q_F(p) = Q_G(p)$ $p \in (p_0, 1)$ [$p \in (0, p_0)$].*

In plainer language, any two probability distributions whose CDF has right [left] tails that are identical over a right[left]-unbounded interval are part of the same equivalence class.

5.4 Relating R- and L-Ordering to Quantile Functions

Van Zwet shows that the $F \prec_c G$ over $x \in I$ if and only if the index function $Q''_F(p)/Q'_F(p) \leq Q''_G(p)/Q'_G(p)$ over $p \in I$. Thus, the ratio of the second derivative of a quantile function to its first derivative serves as an index function for ordering probability distributions by their tail behavior. As a further consequence, he shows that $Q''_F(p)/Q'_F(p) \leq Q''_G(p)/Q'_G(p)$ if and only if $Q'_G(p)/Q'_F(p)$ is nondecreasing over $p \in I$. These same conditions apply to $\prec_R$ [$\prec_L$] over $p \in (p_0, 1)$ [$p \in (0, p_0)$], and they can serve as a convenient verification tool for these two binary relations.

Proposition 21. *For distributions $F, G \in \mathcal{D}$, $F \prec_R$ [$\prec_L$] G if and only if there exists points p_0, p_1 such that $Q''_F(p)/Q'_F(p) <$ [$>$]$Q''_G(p)/Q'_G(p), p \in (p_0, 1)[(0, p_0)]$ and $Q'_F(p)|_{p_1} <$ [$>$]$Q'_G(p)|_{p_1}, p_1 \in (p_0, 1)[(0, p_0)]$.*

Proof. Begin with the condition that $\zeta(x)$ is concave over the interval (x_0, ∞), where $x_0 = Q_F(p_0)$. This is true if and only if:

$$\begin{aligned}
&\frac{d^2}{dx^2}\left(Q_G(F(x))\right) > 0 \\
\Leftrightarrow\ &\frac{d}{dx}\left(Q'_G(F(x))f(x)\right) > 0 \\
\Leftrightarrow\ &\frac{d}{dx}\left(\frac{Q'_G(F(x))}{Q'_F(F(x))}\right) > 0 \\
\Leftrightarrow\ &\left(\frac{Q''_G(F(x))}{[Q'_F(F(x))]^2} - \frac{Q'_G(F(x))Q''_F(F(x))}{[Q'_F(F(x))]^3}\right) > 0 \\
\Leftrightarrow\ &\left(\frac{Q''_G(p)}{Q'_G(p)} - \frac{Q''_F(p)}{Q'_F(p)}\right) > 0, p \in (p_0, 1).
\end{aligned}$$

The third step makes use of the fact that by differentiating $Q_F(F(x)) = x$, $f(x) = 1/Q'_F(F(x))$. The fourth step divides by $Q'_G(F(x))$, which must be positive. The last step sets $p = F(x)$. Continue with the condition $\zeta'(x_1) > x_1, x \in (x_0, \infty)$, and let $p_1 = F(x_1)$. Begin by differentiating $\zeta(x) = Q_G(F(x))$:

$$\begin{aligned}
&\frac{d}{dx}\left(Q_G(F(x))\right|_{x_1} > 1 \\
\Leftrightarrow\ &Q'_G(F(x_1))f(x_1) > 1 \\
\Leftrightarrow\ &\frac{Q'_G(F(x_1))}{Q'_F(F(x_1))} > 1 \\
\Leftrightarrow\ &Q'_G(p_1) > Q'_F(p_1).
\end{aligned}$$

The fourth step uses the fact that $Q'_F(p) \geq 0$ because F is a probability distribution. One can prove this proposition for the relation $\prec_L$ in an analogous fashion.

□

Table 5.2 shows some common probability distributions and the functions $Q(p)$, $Q'(p)$, and $Q''(p)/Q'(p)$ that one can use in verifying tail ordering according to Proposition 21. Applying the relation $\prec_R$ to a set of familiar distributions yields the ordering: *uniform* $\prec_R$ *normal* $\prec_R$ *logistic* $\prec_R$ *exponential* $\prec_R$ *Cauchy*. Figure 5.1 shows the van Zwet index function $Q''(p)/Q'(p)$ for the normal distribution $\frac{\Phi^{-1}(p)}{\phi(\Phi^{-1}(p))}$

Table 5.2: Tail functions for various probability distributions

distribution	$Q(p)$	$Q'(p)$	$Q''(p)/Q'(p)$
uniform	p	1	0
normal	$\Phi^{-1}(p)$	$\frac{1}{\phi(\Phi^{-1}(p))}$	$\frac{\Phi^{-1}(p)}{\phi(\Phi^{-1}(p))}$
logistic	$\log(\frac{p}{1-p})$	$\frac{1}{p(1-p)}$	$\frac{2p-1}{p(1-p)}$
exponential	$-\log(1-p)$	$\frac{1}{1-p}$	$\frac{1}{1-p}$
Cauchy	$\tan(\pi(p-0.5))$	$\pi\sec^2(x)$	$2\pi\tan(\pi(p-0.5))$

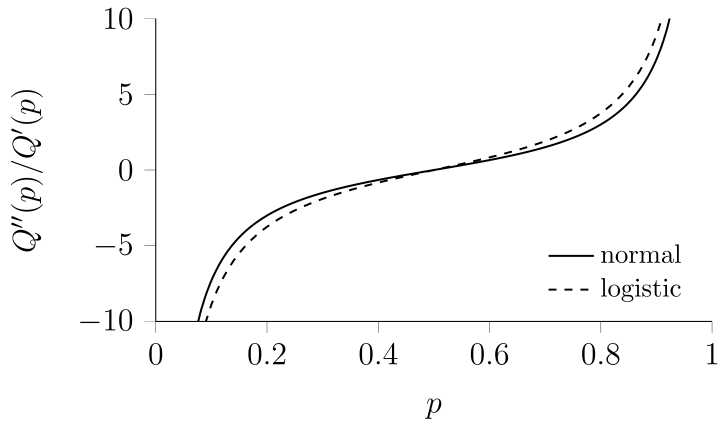

Figure 5.1: The van Zwet index function of the normal and logistic distributions

and the logistic distribution $\frac{2p-1}{p(1-p)}$. The information in this figure is consistent with the fact that *normal* $\prec_L$ *logistic* and *normal* $\prec_R$ *logistic*. There is one additional item of note: the index functions $Q''(p)/Q'(p)$ are themselves quantile functions. Indeed, each of the distributions shown in Table 5.2, are themselves quantile functions because they are both nondecreasing and defined for all $p \in (0,1)$. In general, it is not true that a probability distribution's index function is a quantile function, but it is always true for the important class of *log-concave* probability distributions with twice-differentiable PDFs.[4]

Definition 16. *A continuous probability distribution F whose PDF $f(x)$ is a log-concave function is a log-concave probability distribution.*

[4]All probability distributions in Table 5.2 are log-concave except for the Cauchy distribution.

In mathematical terms, $f(\theta x_1+(1-\theta)x_2) \geq f(x_1)^\theta f(x_2)^{1-\theta}$ for all $x_1, x_2 \in \mathrm{dom}(f)$ and $\theta \in (0,1)$. Equivalently, $\frac{d^2}{dx^2}\log(f(x)) \leq 0$. A function $f(x)$ is log-concave if $\log(f(x))$ is concave.

Proposition 22. *A log-concave probability distribution has a van Zwet index function $Q''(p)/Q'(p)$ that is itself a quantile function.*

The proof of Proposition 22 is in the appendix.

5.5 Tail Behavior of QPDs

The probability distribution tail theory introduced in the previous sections leads to some useful results for QPDs. This discussion begins by studying the tail behavior of QPDs whose basis functions are all quantile functions. The following three propositions refer to a set of continuous, strictly increasing quantile functions $Q_1(p), \cdots, Q_n(p)$ with corresponding probability distributions $F_1, \cdots, F_n$. They apply when $n > 1$ and do not address the trivial case where $n = 1$. See the appendix for proofs of the following three propositions.

Proposition 23. *Let $F_i \prec_R F_n$ $[F_i \prec_L F_n]$ for all $i \in \{1 : n-1\}$ and let $\tilde{F}_n$ be a probability distribution whose quantile function is $\left(\sum_{i=1}^n \beta_i\right) Q_n(p)$. If a QPD with quantile function $\tilde{Q}(p) = \beta_1 Q_1(p) + \cdots + \beta_n Q_n(p)$ has coefficients $\{\beta_i \mid i \in 1, \cdots, n\}$ that are all positive, then $\tilde{F} \prec_R \tilde{F}_n$ $[\tilde{F} \prec_L \tilde{F}_n]$.*

Proposition 24. *Let $F_1 \prec_R F_i$ $[F_1 \prec_L F_i]$ for all $i \in \{2 : n\}$ and let $\tilde{F}_1$ be probability distributions whose quantile function is $\left(\sum_{i=1}^n \beta_i\right) Q_1(p)$. If a QPD with quantile function $\tilde{Q}(p) = \beta_1 Q_1(p) + \cdots + \beta_n Q_n(p)$ has coefficients $\{\beta_i \mid i \in 1, \cdots, n\}$ that are all positive, then $\tilde{F}_1 \prec_R \tilde{F}$ $[\tilde{F}_1 \prec_L \tilde{F}]$.*

Propositions 23 and 24 give upper and lower tail bounds to any QPD whose basis functions are quantile functions whenever the basis functions meet specific ordering conditions and its coefficients are all positive. As an example, the tail behavior of a QPD with quantile function $Q(p) = 3 \cdot (2 + \Phi^{-1}(p)) + 2\log\left(\frac{p}{1-p}\right) + 4(-\log(1-p))$ is R-ordered and L-ordered between a normal distribution with mean 18 and variance

81, (quantile function $(3 + 2 + 4) \cdot (2 + \Phi^{-1}(p))$), and an exponential distribution whose parameter is 9 (quantile function $(3 + 2 + 4)(-\log(1 - p))$). These bounds hold because $F_{normal} \prec_R F_{logistic} \prec_R F_{exponential}$. While it is infeasible for a QPD whose basis functions are quantile functions to have all negative coefficients, it is possible for such QPDs to have one or more negative coefficients, for example, the QPD whose quantile function is $Q(p) = \log(\frac{p}{1-p}) - \Phi^{-1}(p)$. The following proposition applies to cases where one or more coefficients of a QPD are negative numbers, with the added restriction that the heaviest tailed basis function $Q_n(p)$ has an infinite right [left] tail. For this proposition, define the set of indices $\mathcal{I}_+ \equiv \{i \mid \beta_i > 0\}$ and $\mathcal{I}_- \equiv \{i \mid \beta_i < 0\}$ so that $\mathcal{I}_+ \cup \mathcal{I}_- = \{1, \cdots, n\}$. Using this notation, one can write $\tilde{Q}(p) = \sum_{i \in \mathcal{I}_+} \beta_i Q_i(p) + \sum_{i \in \mathcal{I}_-} \beta_i Q_i(p)$, thereby separating $\tilde{Q}(p)$ into a positive weighted sum of quantile functions plus a negative weighted sum of quantile functions.

Proposition 25. *Let $F_i \prec_R F_n$ [$F_i \prec_L F_n$] for all $i \in \{1 : n-1\}$ and $\lim_{p \to 1} Q_n(p) = \infty$ [$\lim_{p \to 0} Q_n(p) = -\infty$] meaning F_n has an infinite right [left] tail. If the probability distribution $\tilde{F}$ has a quantile function of the form $\tilde{Q}(p) = \beta_1 Q_1(p) + \cdots + \beta_n Q_n(p)$, then there exists a constant $\kappa \in \mathbf{R}_{++}$ such that $\tilde{F} \prec_R \tilde{F}_n$ [$\tilde{F} \prec_L \tilde{F}_n$], where $\tilde{F}_n$ is a probability distribution whose quantile function is $\kappa Q_n(p)$.*

The implication of Proposition 25 is that the probability distribution $\tilde{F}$ is right [left] tail-bounded by a positively scaled version of its heaviest tailed basis function $Q_n(p)$. One last generalization remains—removing the restriction that the basis functions are quantile functions. The following three propositions do this, making statements on tail behavior that apply to any QPD whose basis functions each have a right [left] *tail-matching* quantile function that is continuous and strictly increasing.

Definition 17. *The quantile function $Q(p)$ right [left] tail-matches a QPD basis function $g(p)$ when there exists a point $p_t \in (0, 1)$ such that $g(p) = Q(p)$, all $p \in (p_t, 1)$ [$(0, p_t)$].*

A tail-matching quantile function will exist for any basis function whose right [left] tail $p \in (p_t, 1)$ [$(0, p_t)$] is nondecreasing in p. Therefore, a tail-matching quantile function will not exist for all basis functions, for example, $g(p) = \sin(\pi p),\ p \in (0, 1)$

does not have a right tail-matching quantile function. A remedy is to redifine $g(p) = -\sin(\pi p)$, leaving its QPD unchanged, except for a corresponding sign change to the coefficient that multiplies it. Although basis functions exist that negation will not remedy, like $g(p) = \sin\left(\frac{1}{1-p}\right)$, or $g(p) = \sin\left(\frac{1}{p}\right)$, it is difficult to imagine a reason to choose such functions as basis functions.

Propositions 26, 27, and 28 take a QPD F with quantile function $Q(p)$ and basis functions $g_i(p), \cdots, g_n(p)$. Each basis function has continuous, strictly increasing right [left] tail-matching quantile functions $Q_1(p), \cdots, Q_n(p)$ with corresponding probability distributions $F_1, \cdots, F_n$.

Proposition 26. *If $F_i \prec_R F_n$ $[F_i \prec_L F_n]$ for all $i \in \{1 : n-1\}$, and the coefficients $\beta_i > 0,\ i \in \{1 : n\}$, there exists a point $p_0 \in (0, 1)$ such that* $Q(p) < \left(\sum_{i=1}^{n} \beta_i\right) g_n(p),\ p \in (p_0, 1)\ \left[\left(\sum_{i=1}^{n} \beta_i\right) g_n(p) < Q(p),\ p \in (0, p_0)\right]$.

Proposition 27. *If $F_1 \prec_R F_i$ $[F_1 \prec_L F_i]$ for all $i \in \{2 : n\}$, and the coefficients $\beta_i > 0,\ i \in \{1 : n\}$, there exists a point $p_0 \in (0, 1)$ such that* $\left(\sum_{i=1}^{n} \beta_i\right) g_1(p) < Q(p),\ p \in (p_0, 1)\ \left[Q(p) < \left(\sum_{i=1}^{n} \beta_i\right) g_1(p),\ p \in (0, p_0)\right]$.

Proposition 28. *If $F_i \prec_R F_n$ $[F_i \prec_L F_n]$ for all $i \in \{1 : n-1\}$, and $Q_n(p)$ has an infinite right [left] tail, there exists a point $p_0 \in (0, 1)$ and a constant $\kappa \in \mathbf{R}_{++}$ such that $Q(p) < \kappa g_n(p),\ p \in (p_0, 1)\ [\kappa g_n(p) < Q(p),\ p \in (0, p_0)]$.*

Propositions 26, 27, and 28 are restatements of Propositions 23, 24, and 25 and are generalized to apply to basis functions that are not quantile functions, yet behave like quantile functions in their tails. The proofs of these propositions follow from the definition of a tail-matching quantile function and Propositions 23, 24, and 25.

To demonstrate Propositions 26 and 27, take the QPD with quantile function $Q(p) = 1 + 2\Phi^{-1}(p) + 3p\Phi^{-1}(p)$. This QPD is right tail-bounded by $4(1 + 2\Phi^{-1}(p))$ (a normal distribution with mean 4 and variance 64), and the basis function $4p\Phi^{-1}(p)$. Note that $p\Phi^{-1}(p)$ is not a quantile function because it is decreasing over part of its domain, but there exist quantile functions that will right tail-match it. To demonstrate Proposition 25, take the QPD with quantile function $Q(p)$ with basis functions $\{1, (p-1)^2, \Phi^{-1}(p)\}$. This theorem implies that there exists a normal distribution with finite variance κ^2 that bounds the right and left tails of $Q(p)$.

One last result ties this theory of tail behavior to the engineering of QPDs—the QPD Tails Theorem. This theorem explicitly indicates how to change the tail behavior of a QPD by adding a basis function. As with the previous three propositions, it takes $Q(p)$ to be a quantile function of a QPD F with basis functions $g_i(p), \cdots, g_n(p)$. Each basis function has continuous, strictly increasing right [left] tail-matching quantile functions $Q_1(p), \cdots, Q_n(p)$ with corresponding probability distributions $F_1, \cdots, F_n$, where $F_i \prec_R F_n,\ i \in \{1 : n-1\}$.

Theorem 3 (QPD Tails Theorem)**.** *Let $g_H(p),\ p \in (0,1)$ be a continuous function with right [left] tail-matching quantile function $Q_H(p)$. Create a new QPD $\tilde{F}$ with quantile function $\tilde{Q}(p)$ by adding the basis function $g_H(p)$ to the QPD F's set of basis functions. If $\lim_{p\to 1} g_H(p)/g_n(p) = \infty$ $[\lim_{p\to 0} g_H(p)/g_n(p) = -\infty]$, then $F \prec_R \tilde{F}$ $[F \prec_L \tilde{F}]$.*

Proof.

$$
\begin{aligned}
& \lim_{p\to 1} \frac{g_H(p)}{g_n(p)} = \infty \\
\Leftrightarrow\ & \lim_{p\to 1} \frac{g_H(p)}{\kappa Q_n(p)} = \infty,\ \kappa \in \mathbf{R}_{++} && \text{by Definition 17} \\
\Rightarrow\ & \lim_{p\to 1} \frac{g_H(p)}{Q(p)} = \infty && \text{by Proposition 28} \\
\Leftrightarrow\ & \lim_{p\to 1} \left(1 + \frac{\beta_H g_H(p)}{Q(p)}\right) = \infty,\ i \in \{1 : n\} && \text{because } \beta_H \in \mathbf{R}_{++} \\
\Leftrightarrow\ & \lim_{p\to 1} \frac{\tilde{Q}(p)}{Q(p)} = \infty && \text{by definition of } \tilde{Q}(p)
\end{aligned}
$$

Since it is given that $\tilde{Q}(p)$ is a quantile function, it must be true that $\beta_H \in \mathbf{R}_{++}$; otherwise, $\tilde{Q}(p)$ would be decreasing in p across some subinterval of its domain. Since $g_H(p)$ and $Q(p)$ are continuous, so too must $\tilde{Q}(p)$ be continuous. Because $\tilde{Q}(p)$ is continuous and $\lim_{p\to 1} \frac{\tilde{Q}(p)}{Q(p)} = \infty$, there must exist a $p_0 \in (0,1)$ such that $Q(p) < \tilde{Q}(p),\ p \in (p_0, 1)$, therefore $F \prec_R \tilde{F}$. □

The QPD Tails Theorem gives a rule for identifying a basis function that, when added to a QPD's set of basis functions, creates a new QPD with increased right and/or left tail heaviness. To demonstrate this theorem, take a QPD F with basis functions $\{1, (p-1)^2, \Phi^{-1}(p)\}$. Create another QPD $\tilde{F}$ by adding $\log\left(\frac{p}{1-p}\right)$ so that its set of basis functions is $\left\{1, (p-1)^2, \Phi^{-1}(p), \log\left(\frac{p}{1-p}\right)\right\}$. The R-ordering is $F \prec_R \tilde{F}$ because $\lim_{p\to 1}\left(\log\left(\frac{p}{1-p}\right)/\Phi^{-1}(p)\right) = \infty$.

5.6 Implications of Transforms on Tail Behavior

A quantile function that is a transform of another quantile function $\tilde{Q}(p) = h^{-1}(Q(p))$ often has different tail behavior than the original quantile function $Q(p)$. Table 5.3 shows that for logistic, power, and exponential transforms, the van Zwet index function $\tilde{Q}''(p)/\tilde{Q}(p)$ is a simple function of the van Zwet index function of the pre-transformed quantile function $Q(p)$. Happily, this gives a clear interpretation of the implications for each transform. A logarithmic transform of any probability distribution thickens its right tail, whereas an exponential transform thins its right tail. Finally, a power transform thickens [thins] its right tail if $\alpha > 1$, [$\alpha < 1$]. For example, when taking an exponential transform of a QPD, the transformed distribution has shorter tails than did its pre-transformed distribution.

Table 5.3: Transform effects on tail behavior

transform $h(\cdot)$	$\tilde{Q}(p)$	$\tilde{Q}'(p)$	$\tilde{Q}''(p)/\tilde{Q}'(p)$
positive affine	$\frac{Q(p)-a}{b}$	$\frac{Q'(p)}{b}$	$\frac{Q''(p)}{Q'(p)}$
logarithmic	$\exp(Q(p))$	$Q'(p)\exp(Q(p))$	$\frac{Q''(p)}{Q'(p)} + Q'(p)$
power	$[Q(p)]^\alpha$	$\alpha[Q(p)]^{\alpha-1}Q'(p)$	$\frac{\alpha-1}{Q(p)} + \frac{Q''(p)}{Q'(p)}$
exponential	$\log(Q(p))$	$\frac{Q'(p)}{Q(p)}$	$\frac{Q''(p)}{Q'(p)} - \left[\frac{Q'(p)}{Q(p)}\right]^2$

Transforms can be useful when modeling with quantile functions. When choosing a transform, as when choosing a QPD's basis functions, a decision analyst should be mindful of the effect it has on tail behavior. Armed with the tools from this and the previous two chapters, the focus shifts from characterization to application.

Chapter 6

A Decision Analysis Using Quantile-Parameterized Distributions

Perhaps the best way to highlight the usefulness of QPDs is to demonstrate how to use them in the modeling of a decision. This chapter takes a CEO's decision through the decision analysis cycle [23, 26]. It begins with the *formulation* stage in §6.1 and §6.2, continues in §6.3 with the *evaluation* stage, and concludes with the *appraisal* stage in §6.4 and §6.5. These final two sections illustrate three methods of decision model appraisal inspired by the use of QPDs.

6.1 Whether to Market or License a Drug

6.1.1 The Market Alternative

A mid-size pharmaceutical company owns a drug candidate that has recently received FDA approval. The CEO has a plan for her company to market the drug. She values this *Market* alternative for the drug according to the equation

$$v_{market} = \sum_{i=1}^{T} \frac{m \cdot s(i) \cdot (p - c)}{(1 + r)^i}. \tag{6.1}$$

The value measure v_{market} is the net present value of *profit* discounted over T time periods with discount rate r. The variable m is the initial *market size* for the drug candidate.

The market share of the drug $s(i)$ for each period i is the ratio of the number of prescriptions written for the drug to the total number of prescriptions written for it and competing therapies. The CEO believes market share will grow each period to account for a forecasted increase in physician acceptance over time. She models this market share to begin at zero and grow according to

$$s(i) = s_p \cdot (1 - \exp(-g \cdot i)). \tag{6.2}$$

The *peak market share* s_p represents her belief about the highest market share that the drug would achieve were it to have an infinite patent life. The growth parameter g represents her belief about the rate that the drug approaches this *peak market share*. Figure 6.1 depicts a market share profile for $s_p = 0.7$, $g = 0.5$ and $T = 10$.

To encode her prior probability distribution on the growth parameter g, the decision analyst assesses t_h, the time until the market share reaches half of its peak, and then relates it to the growth parameter with the equation $g = \frac{1}{t_h} \log(2)$. This changes (6.2) to

$$s(i) = s_p \cdot (1 - 2^{-i/t_h}). \tag{6.3}$$

The remaining two variables necessary for determining v_{market}, are p, the *unit price* paid to the company for each prescription, and c the *unit cost* per prescription. Substituting (6.3) into (6.1) yields an equation that separates into two geometric

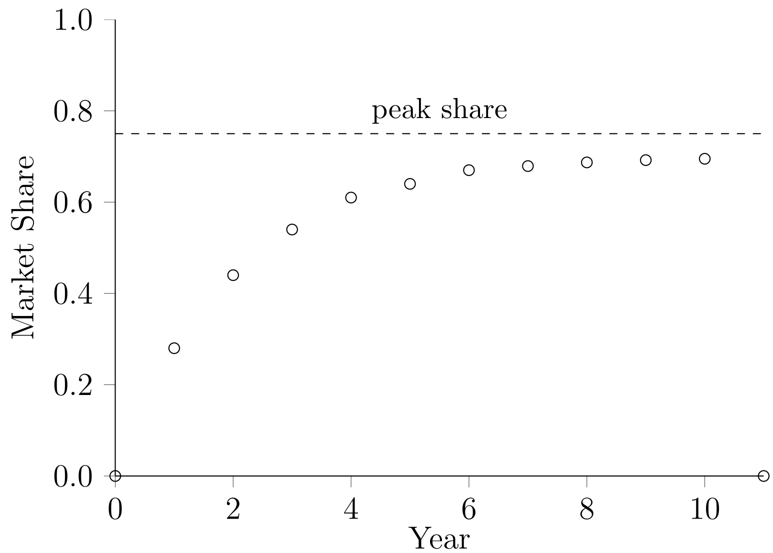

Figure 6.1: A market share profile with $s_p = 0.7$, $g = 1.5$ and $T = 10$

series and simplifies to

$$v_{market} = m \cdot s_p \cdot (p - c) \cdot \left(\frac{1 - \left(\frac{1}{1+r}\right)^{T+1}}{1 - \frac{1}{1+r}} - \frac{1 - \left(\frac{1}{2^{1/t_h}(1+r)}\right)^{T+1}}{1 - \frac{1}{2^{1/t_h}(1+r)}} \right). \tag{6.4}$$

6.1.2 The License Alternative

Before introducing the drug to market, the company's business development team generates a new alternative to license its marketing rights to the drug to a second company. In exchange for these marketing rights, the licensee will pay an *upfront* payment to the CEO's company, as well as an ongoing *royalty* payment that is a fixed percentage of the revenue it garners from the sale of the drug. She values this *License* alternative according to the equation

$$v_{license} = u + \sum_{i=1}^{T} \frac{f \cdot m \cdot s(i) \cdot p}{(1+r)^i}, \tag{6.5}$$

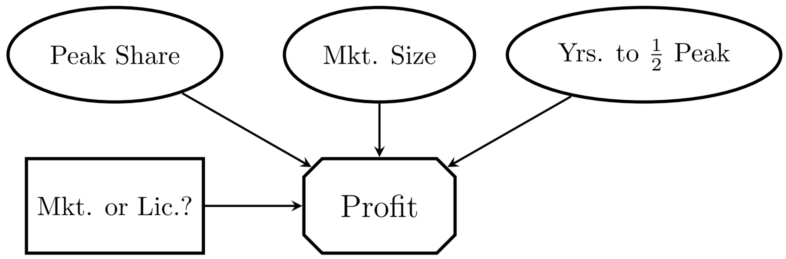

Figure 6.2: The CEO's decision diagram

which simplifies to

$$v_{license} = u + f \cdot m \cdot s_p \cdot p \cdot \left(\frac{1 - \left(\frac{1}{1+r}\right)^{T+1}}{1 - \frac{1}{1+r}} - \frac{1 - \left(\frac{1}{2^{1/t_h}(1+r)}\right)^{T+1}}{1 - \frac{1}{2^{1/t_h}(1+r)}} \right). \tag{6.6}$$

The variables T, r, m, s_p, and t_h in (6.5) are the same as those in (6.1). The *upfront payment* u and the *royalty* on revenue f account for the licensing deal terms. The CEO is uncertain about m, s_p, and t_h. Table 6.1 shows her input values for all other model parameters. The CEO must decide whether to *Market* or *License* the drug; Figure 6.2 depicts her decision.

Table 6.1: The CEO's deterministic inputs

time periods, T	unit price, p	unit cost, c	upfront payment, u	royalty, f
10 years	\$500	\$15	\$75 million	12%

6.2 Encoding Prior Probability Distributions

The decision analyst elicits the CEO's prior probability distribution over *market size*, *peak market share*, and *years to half-peak share* as the sets of quantile/probability pairs in Table 6.2. In addition, the CEO assigns bounds for each uncertainty. Initial *market size* must be a nonnegative integer, and *peak market share* must be limited between 0% and 100%. She also asserts that the number of *years to half-peak share* would never be less than one year. She does not want to expend time thinking about precise upper bounds for *market size* and *years to half-peak share*, so the decision analyst pencils in "very large" for these two values. She agrees that the decision analyst should model the number of prescriptions sold in any period as a nonnegative real number rather than a nonnegative integer. This allows him to use continuous probability distributions to represent her knowledge about these three uncertainties. The analyst chooses the set of basis functions $\{1, \Phi^{-1}(p), p\Phi^{-1}(p)\}$ for *market size*.

Table 6.2: The CEO's elicited quantile-probability pairs

Quantile	0.1	0.5	0.9	Lower Limit	Upper Limit
Market Size [Rx]	$20,000$	$50,000$	$100,000$	0	very large
Market Share	60%	75%	90%	0%	100%
Years to Half-Peak	2	3	5	1	very large

By equation (3.5), the QPD $Q(p) = 50 + 21.5\Phi^{-1}(p) + 19.5p\Phi^{-1}(p)$, in units of $1,000$ annual prescriptions, is consistent with the CEO's quantiles, but it does not conform with her limits because it allows for negative market sizes. After discussion, she chooses the truncated quantile function (6.7) shown in Figure 6.3. According to §4.3, this quantile function is the QPD whose argument is transformed by the function $\hat{p} = p \cdot (b - a) + a$ where $b = 1$ and the constant $a = F(0)$, which is approximately $a = 0.01$.

$$\hat{Q}(p) = 50 + 21.5\Phi^{-1}(0.99p + 0.01) + 19.5(0.99p + 0.01)\Phi^{-1}(0.99p + 0.01) \qquad (6.7)$$

Its support $(0, \infty)$ is consistent with the upper and lower limits she places on *market size*. She feels that this QPD represents her knowledge about *market size*.

The decision analyst again chooses the set of basis functions $\{1, \Phi^{-1}(p), p\Phi^{-1}(p)\}$, this time for the *years to half-peak share* uncertainty. Since this uncertain quantity must lie between one year and a "very large" number of years, he wants a transform that will give a support of $(1, \infty)$. The shifted, log-transformed QPD, with quantile function

$$\tilde{Q}(p) = \exp\left(0.70 + 0.54\Phi^{-1}(p)\right) + 1$$

gives him a support that agrees with this range and is consistent with the CEO's

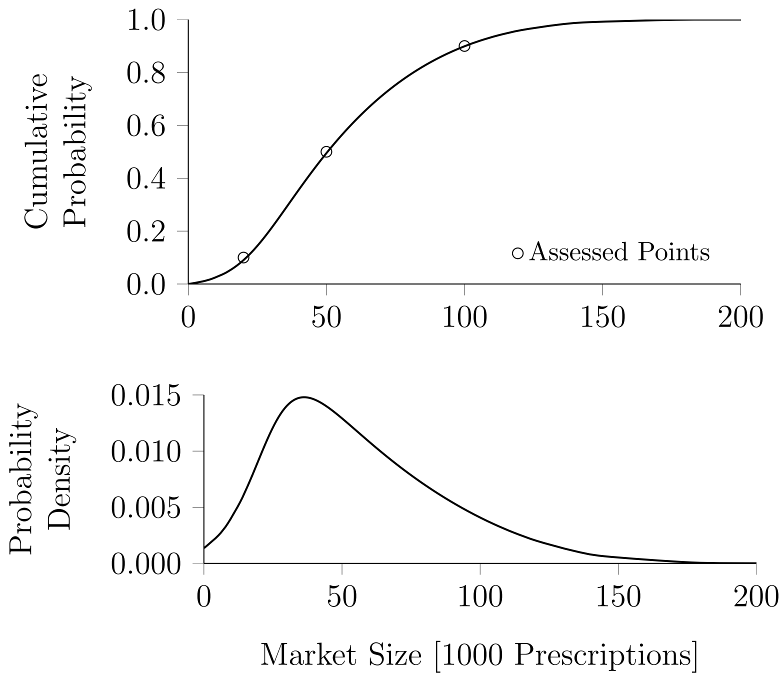

Figure 6.3: The CEO's probability distribution over initial market size

quantiles of Table 6.2. The coefficient β_3 of the basis function $p\Phi^{-1}(p)$ is negligible. Again, after seeing the CDF and PDF in Figure 6.4, the CEO feels this QPD represents her knowledge about *peak market share*.

Continuing with the basis functions $\{1, \Phi^{-1}(p), p\Phi^{-1}(p)\}$, the decision analyst encodes the *peak market share* uncertainty. Since market share must lie between 0% and 100% (0 to 1), he chooses the probit-transformed QPD, with quantile function

$$\tilde{Q}(p) = \Phi\left(0.67 + 0.31\Phi^{-1}(p) + 0.18p\Phi^{-1}(p)\right)$$

to make its support agree with these limits and the CEO's quantiles. Again, after seeing the CDF and PDF in Figure 6.5, she feels this QPD represents her knowledge about *peak market share*. The decision analyst now has the CEO's prior probability distributions encoded and ready for a first-pass analysis.

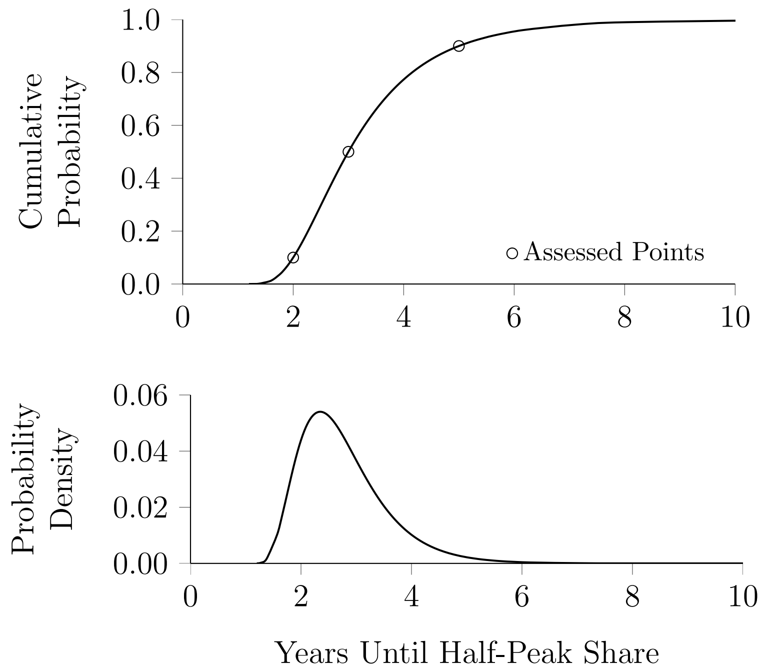

Figure 6.4: The CEO's probability distribution over market growth rate

6.3 Evaluating the Alternatives

The CEO declares that her company has a utility function of the form $u(x) = -\exp(-x/\rho)$. The decision analyst assesses her risk tolerance at $\rho = \$150$ million. For further reading on a corporation's risk tolerance, see Howard [24], Spetzler [60], and Bickel [9]. This information, along with the value functions (6.4) and (6.6) for each alternative and the CEO's prior probability distributions, completes her decision basis. He can now compute her certain equivalent for each alternative. He runs a probabilistic simulation that converges to the values shown in Table 6.3 within one million samples. If deciding immediately, she should choose the *License* alternative.

Table 6.3: The CEO's certain equivalents

Market	License
\$85.0 million	\$86.6 million

She wants to roll up these results into a portfolio analysis to be conducted later

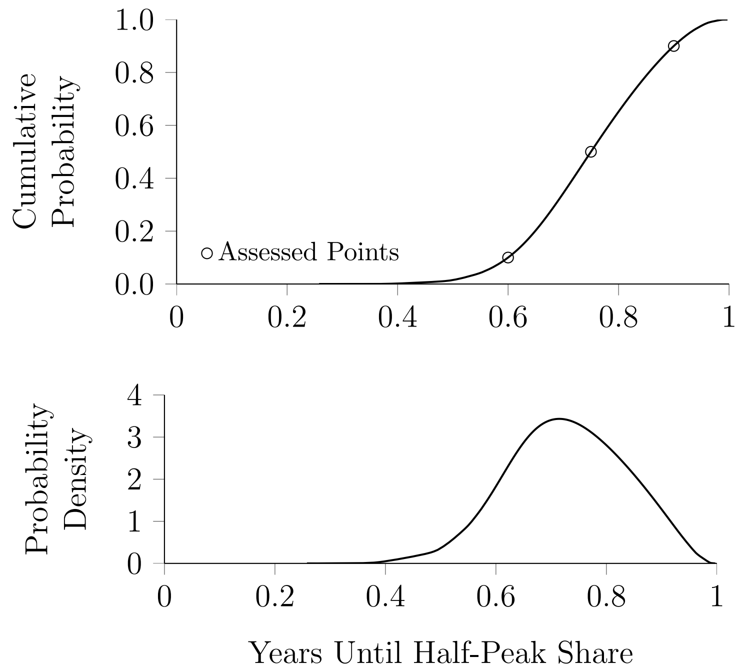

Figure 6.5: The CEO's probability distribution over initial market size

in the year. Rather than store this set of two million datapoints, she asks the decision analyst to compress it for her. He does so using the least-squares approximation method detailed in §3.7. He takes the million simulates for each alternative and builds data vectors x_{Market} and $x_{License}$ with ascending orders, and an equally-spaced probability vector p. Each of these three vectors has one million components. After choosing a set of basis functions, he builds a matrix $Y \in \mathbf{R}^{m \times n}$ similar to that of (3.15), where m is the number of simulates and n the number of basis functions. He then computes the least-squares approximation $\beta_{Market} = \left(Y^T Y\right)^{-1} Y^T x_{Market}$ and $\beta_{License} = \left(Y^T Y\right)^{-1} Y^T x_{License}$. He first tries the simple Q-normal basis functions $\{1, \Phi^{-1}(p), p\Phi^{-1}(p), p\}$. He computes a least-squares approximation for the data with the vectors β_{Market} and $\beta_{License}$, but is unsatisfied with the representation. He adds one more basis function to the set giving a five-coefficientQ-normal $\{1, \Phi^{-1}(p), p\Phi^{-1}(p), p, p(\Phi^{-1}(p))^2\}$, yielding the QPDs with quantile functions

$$Q(p)_{Market} = 50.1 + 20.0\Phi^{-1}(p) + 18.7p\Phi^{-1}(p) + 66.2p + 13.9p(\Phi^{-1}(p))^2 \tag{6.8}$$

and

$$Q(p)_{License} = 81.2 + 2.48\Phi^{-1}(p) + 2.32p\Phi^{-1}(p) + 8.19p + 1.71p(\Phi^{-1}(p))^2 \qquad (6.9)$$

whose units are millions of dollars. This is a QPD representation (Figures 6.6 and 6.7) that satisfies both him and the CEO. For clarity, only 50 of the million value simulates for each alternative are shown as points on the CDF plot. This compression reduces the simulation data representing the probability distribution over the CEO's value of each alternative from two million down to just ten numbers—five coefficients for each alternative. As a check, he generates one million simulates from each of these two QPDs and uses the data to compute her certain equivalent for each alternative. Table 6.4 shows that the approximation is a very good one—accurate to the nearest \$100,000. As an added benefit, this QPD representation allows for a smooth PDF representation for her probability distribution over value for each alternative. The

Table 6.4: The CEO's certain equivalents computed from compressed simulation output

Market	License
\$85.0 million	\$86.6 million
0.055% difference	0.002% difference

CEO expresses a desire to retain more information than just the probability distributions over value. She notes that one strategic decision for her drug portfolio is how much to invest in brand marketing—an intervention that is relevant to *peak market share*, but is irrelevant to *profit* given *peak market share*. The decision diagram of Figure 6.8 depicts her branding decision. The decision analyst already has marginal distributions for both uncertain quantities described by equations (6.8) and (6.9), so he decides to encode the relevance relationship between *peak market share* and *profit* by way of a Gaussian copula. This copula is parameterized by a single number $\rho = 0.234$, which he computes from the $n = 1$ million simulated *peak market share-value* pairs $(s, v) \in \mathbf{R}^n \times \mathbf{R}^n$ using the following steps:

1. $s \xrightarrow{Q_s^{-1}} p_s \in \mathbf{R}^n$

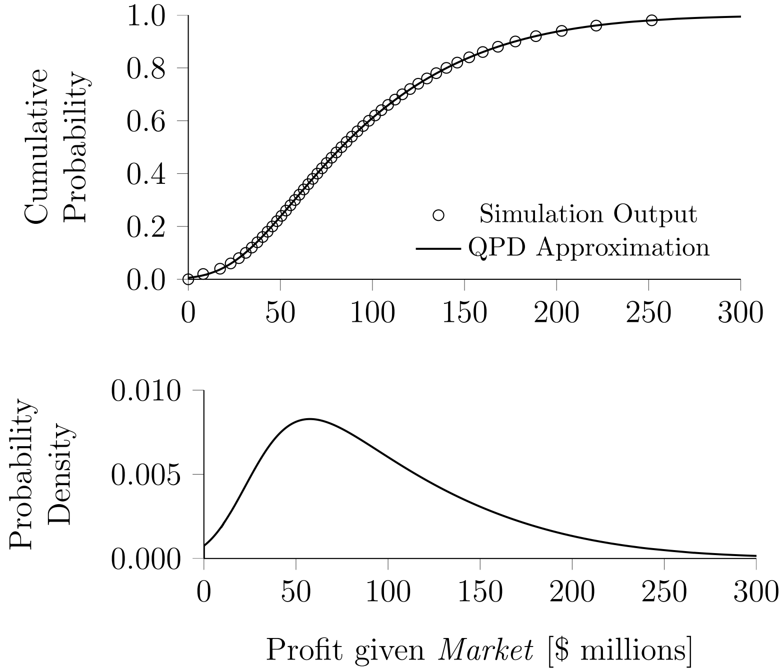

Figure 6.6: The CEO's distribution over the *Market* alternative represented by compressed simulation output

2. $v \to p_v \in \mathbf{R}^n$

3. $p_s \xrightarrow{\Phi^{-1}} x;\ v \xrightarrow{\Phi^{-1}} y,\ x, y \in \mathbf{R}^n$

4. $\rho = (x - m_x)^T (y - m_y)$, where $m_x = \frac{1}{n}\sum_i^n x_i$ and $m_y = \frac{1}{n}\sum_i^n y_i$

The first step iteratively computes the cumulative probability for each *peak market share* simulate by finding the argument p_s that balances the equation $Q(p_s) = s$. The second step estimates the cumulative probability of each component of the simulated value vector p_v through the equation $p_{v_i} = \text{rank}(v_i)/(n+1)$. Both value functions (6.4) and (6.6) are strictly increasing in the same vectors of simulates m,s, and t, therefore $p_{v_{market}} = p_{v_{license}}$, which also implies $\rho_{market} = \rho_{license}$. The third step computes two vectors of standard normal variables $x = \Phi(s)$ and $y = \Phi(v)$. The fourth step computes the maximum likelihood estimator for the Pearson correlation coefficient ρ. The quantities m_x and m_y are the averages of the components of the x and y vectors, respectively.

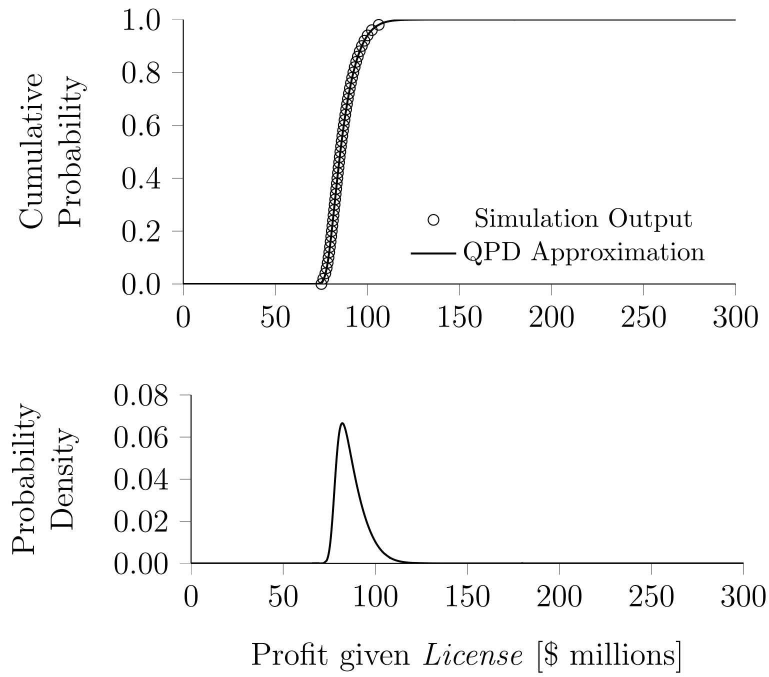

Figure 6.7: The CEO's distribution over the *License* alternative represented by compressed simulation output

The decision analyst returns to the simulation output data and computes the CEO's certain equivalent conditioned on each *peak market share* decile. Then, using equations (6.8), (6.9) and the copula, he samples from each value function conditioned on each decile of *peak market share* and computes the CEO's certain equivalent. Table 6.5 compares the original simulation output and the approximated output. Both the decision analyst and the CEO feel that the approximation is acceptable.

6.4 Sensitivity Analysis

Sensitivity analysis is an important component of the *appraisal* phase of the decision analysis cycle. It highlights how certain equivalents change with changes to the decision model's parameters. Rather than conduct traditional sensitivity analyses to model parameters such as discount rate or risk tolerance (analyses unhindered by the use of QPDs), this section introduces what I believe to be a new sensitivity analysis

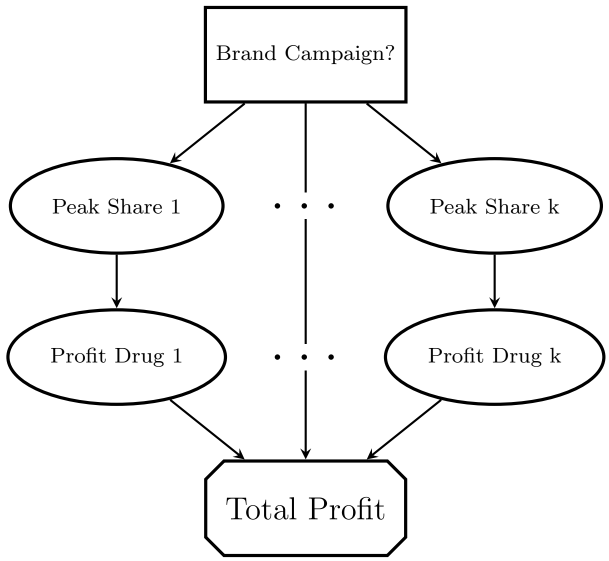

Figure 6.8: The CEO's future brand campaign decision

Peak Mkt. Share Decile	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
CE Mkt. Orig. [$mm]	63.76	72.91	77.27	80.64	83.92	87.38	90.90	94.70	98.89	104.10
CE Mkt. Approx. [$mm]	67.42	74.64	78.27	81.66	84.45	87.65	90.80	94.21	98.97	108.14
Difference [%]	5.7	2.3	1.8	1.6	0.79	0.43	0.073	0.37	0.18	3.9
CE Lic. Orig. [$mm]	83.47	84.76	85.40	85.90	86.38	86.91	87.42	88.01	88.64	89.44
CE Lic. Approx. [$mm]	84.08	85.10	85.59	86.06	86.44	86.89	87.33	87.81	88.48	89.76
Difference [%]	0.74	0.37	0.28	0.24	0.010	0.002	0.11	0.20	0.16	0.36

Table 6.5: The CEO's original value model and the QPD-copula approxmation to that model in terms of certain equivalents conditioned on *peak market share* decile

inspired by the quantile function methods of this research—sensitivity to the tail thickness of an input probability distribution.

In §5.1 I discuss why it is desirable to have a method for testing a decision model's sensitivity to the tail behavior of an input distrubtion. Imagine that QPD describes a decision maker's probability distribution over a specific uncertainty, for example, the QPD with quantile function $Q(p) = 50 + 21.5\Phi^{-1}(0.99p + 0.01) + 19.5p\Phi^{-1}(0.99p + 0.01)$ of (6.7), the CEO's distribution over *market size*. I submit two methods for conducting such a sensitivity analysis on this probability distribution. One method is to add a fourth degree of freedom to the original QPD by adding a fourth basis

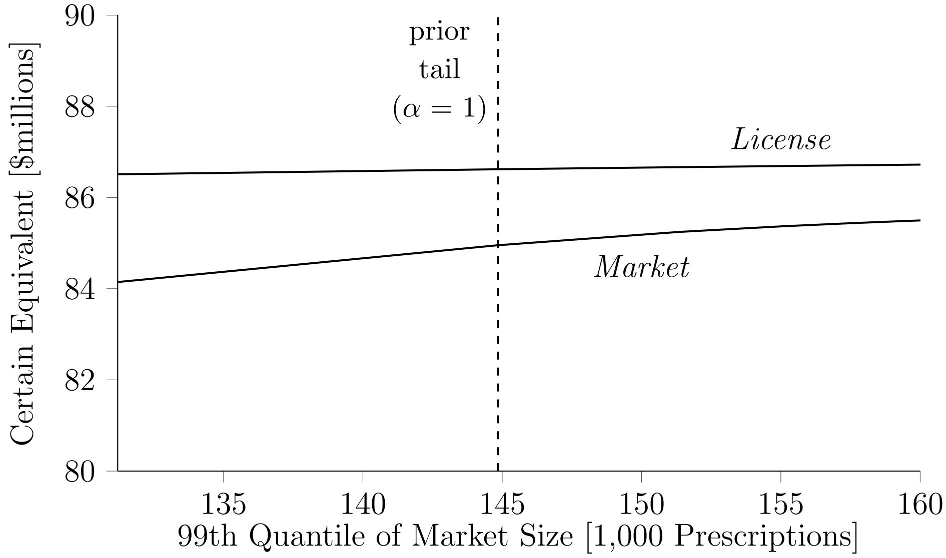

Figure 6.9: Sensitivity to the tail behavior of the *market size* uncertainty

function $g_4(p)$. Then the decision analyst can recompute her certain equivalents after modulating a fourth, extreme quantile. If the addition of this fourth quantile adheres to the requirements of the Quantile Parameters Theorem, the resulting QPD will still pass through the three quantiles from the original analysis. As an alternative approach, he can transform the QPD in such a way as to make the tail lighter or heavier. Table 5.3 shows that the tail index of a distribution power-transformed by the exponent α is $\frac{\alpha-1}{Q(p)} + \frac{Q''(p)}{Q'(p)}$. Thus, the tail-heaviness of a power-transformed distribution is strictly increasing in α, making the power transform a systematic means of modulating the tail-heaviness of any QPD.

The decision analyst decides to use the power-transform method for a sensitivity analysis on the right tail behavior of the CEO's distribution over *market size*, $Q(p) = (50 + 21.5\Phi^{-1}(0.99p + 0.01) + 19.5p\Phi^{-1}(0.99p + 0.01))^{\alpha}$. Figure 6.9 graphs the results of this study for the exponents α from 0.5 to 7 by increments of 0.5. Rather than using the tail exponent α as the dependent variable, he chooses to plot the certain equivalents of each alternative versus the $99th$ quantile of the transformed distribution. This clarifies his communication of this analysis to the CEO.

Figure 6.10 plots the CDFs and PDFs of both the input distribution ($\alpha = 1$) and

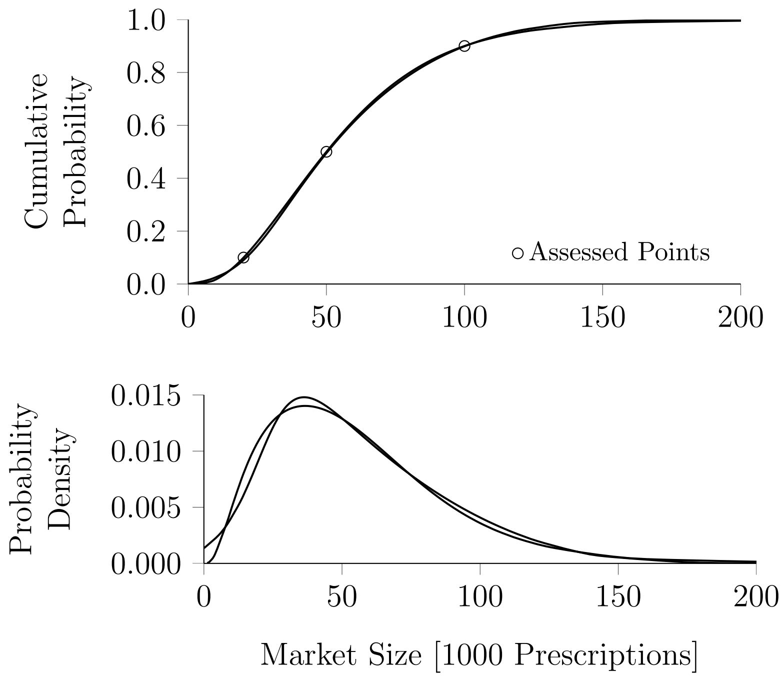

Figure 6.10: CDF and PDF of the initial distribution on the *market size* uncertainty and the distribution with exponent $\alpha = 7$

the heaviest-tailed distribution ($\alpha = 7$). From this study, it is clear to the CEO that spending further time detailing the tail of the *market size* uncertainty is unwarranted.

Figure 6.10 shows that the transformed distribution changes more than just the shape of the tail. The difference is barely perceptible. If the CEO was concerned with this change in the central portion of the distribution, the decision analyst could create a hybrid distribution with the same tail modulation parameter α:

$$\tilde{Q}(p) = \begin{cases} 50 + 21.5\Phi^{-1}(0.99p + 0.01) & \\ +19.5(0.99p + 0.01)\Phi^{-1}(0.99p + 0.01) & 0 \leq p < 0.9 \\ & \\ (\beta_1 + \beta_2\Phi^{-1}(0.99p + 0.01) & \\ +\beta_3(0.99p + 0.01)\Phi^{-1}(0.99p + 0.01))^{\alpha} & 0.9 \leq p < \infty \end{cases} \tag{6.10}$$

The vector $\beta = Y^{-1}x^{\frac{1}{\alpha}}$, where Y is the same basis function matrix used to compute the original coefficients for *market size*. The distribution in (6.10) is identical to the original distribution for all quantiles less than the 90th. It has a CDF that is continuous but not necessarily differentiable.

6.5 Valuing Information

Valuing information is another important component of the *appraisal* phase [22]. Rather than conduct a traditional value of information analysis on one of the three uncertain distinctions of the model (analyses unhindered by the use of QPDs), I instead use QPDs to highlight the valuation of a type of information rarely considered by decision analysts—the valuation of probability assessment.[1] This analysis requires both a decision analyst able to conduct a moderately difficult simulation, and a decision maker who is comfortable assigning probabilities to abstract quantities—perhaps this is one reason for the rarity of the valuation of probability assessment in practice. For a theory on the use of experts in decision analyis, see Morris [41, 42]. For a detailed examination of valuing probability assessment, see Logan [38].

The CEO is aware of a market research firm whose specialty is predicting the market size of drugs about to enter the market. The firm has an expert in the market sizing of drugs similar to the one of this decision. The service comes at a cost; for $\$200,000$, the expert will perform a thorough market analysis and assign his $(0.1, 0.2, \ldots, 0.8, 0.9)$-quantiles on *market size*. This feels expensive to the CEO. Nevertheless, she asks the decision analyst whether he thinks paying for the expert's *market size* quantiles is a good deal. To answer this question, he performs an analysis to determine whether or not the CEO should purchase this every-decile assessment.

When conducting such an analysis, it is important to recognize that different assignments of probabilities and quantiles might be inconsistent, reflecting a lack of information. Two extensive literature reviews on probability judgment and assessment are Slovic, Fischoff, and Lichtenstein [56] and Wallsten and Budescu [67]. In

[1] I use the term *probability assessment* to mean the assessment of a decision maker's uncertainty as points on a CDF, so that the assessed quantity could either be a quantile or a cumulative probability.

their paper, Wallsten and Budescu adopt the linear model

$$x = t + e. \tag{6.11}$$

The model describes the relationship between an elicited probability x as the sum of t, the *true* probability, and an error term e. On page 153 of their aforementioned paper, Wallsten and Budescu choose a definition for *true* probability:

> A convenient way to define the true score, t, is as the expected value of this hypothetical distribution that would be obtained across a series of statistically independent judgments by a given individual. Thus, true score is a hypothetical concept which is determined through the observed value by the expectation operator.

This definition passes the clarity test as defined by Howard [26]. Tani [62, page 1501] defines a more general notion he calls *authentic* probability:

> Let us imagine the existence of a person, called the Probabilist, who is capable of performing upon request any calculation using the rules of probability calculus (e.g., Bayes' Theorem, expansion, or change of variable). Then we can state the following operational definition of autheticity: The authentic probability for an event is the one that we would obtain if we could spend an unlimited amount of time in introspection and if we had the services of the Probabilist.

Logan [38, page 25] builds on Tani's work by introducing a set of axioms "sufficient for the existence and uniqueness of an authentic probability distribution."

Logan bases his work on the idea that one can value probability assessment in the same way one values information. For univariate continuous probability distributions, he proposes valuing assessment over the first four central moments of the distribution [38, page 67]. But Spetzler and Staël von Holstein contend "Subjects are seldom able to express their uncertainty in terms of a density function, a cumulative distribution, or moments of a distribution. Therefore, it is usually not meaningful to try eliciting a distribution or its moments directly" [61, page 351]. One key desideratum of a process for valuing probability assessment is that the assessment data be quantile-probability pairs. For probability distributions defined by QPDs, quantiles are linear functions of coefficients, making them well suited to such a valuation.

The example of §6.5.1 uses the quantile-probability assessment data of Spetzler and Staël von Holstein. This section uses their *V-method* of elicitation—fixing a probability and assessing its associated quantile. Abbas et al. [5] compare what they call *fixed value* versus *fixed probability* (V-method). They slightly favor *fixed value* elicitation. In this section on valuing probability assessment, I adapt the linear model of Wallsten and Budescu and propose a definition for the coefficients β that passes the clarity test.

6.5.1 Modeling the Expert's Responses

In order to value the market-sizing expert's quantiles on *market size*, the decision analyst first creates a model (6.12) that describes the CEO's knowledge about the result of such an assessment. This model uses the set of basis functions $\{1, \Phi^{-1}(p), p\Phi^{-1}(p)\}$, the same as the QPD describing the CEO's knowledge about the *market size* uncertainty from Table 6.2.

$$x_i = \beta_1 + \beta_2\Phi^{-1}(p_i) + \beta_3 p_i\Phi^{-1}(p_i) + \varepsilon_i \tag{6.12}$$

Each p_i-quantile the expert assigns is x_i, and the term ε_i accounts for the expert's inconsistency in making his assignment of the ith quantile. She believes that the expert might give slightly different answers to the same quantile assessment question asked at different times. This is equivalent to Wallsten and Budescu's error term e from (6.11). The CEO believes all such inconsistencies ε_i are irrelevant to one another and distributed normally with mean 0 and uncertain variance σ^2. Further, the CEO believes that the $(p_1, \ldots, p_m)$-quantile data $(x_1, \ldots, x_m)$ elicited from the expert are irrelevant given σ and β.

Similar to Wallsten and Budescu, and Tani, the decision analyst defines the expert's probability distribution over *market size* as the QPD with the aformentioned basis functions whose vector of coefficients β is the answer to the question, "What

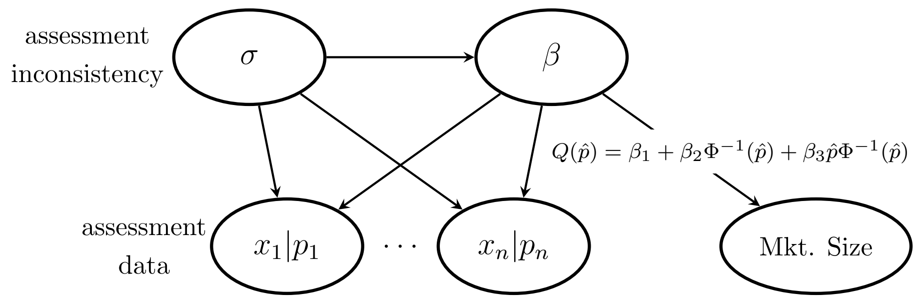

Figure 6.11: The CEO's relevance diagram describing the expert's quantiles on the *market size* uncertainty

vector of coefficients $\beta \in \mathbf{R}^3$ is the least-squares estimator of a set of one thousand quantile assessments the expert would assign to the decision analyst were they to take the time to do so?" Since the clairvoyant can make this computation and answer this question, the CEO and decision analyst agree that it passes the clarity test.

Figure 6.11 depicts these probabilistic relationships. If the CEO were to observe a set of expert-assessed quantiles $\{x_i \mid p_i, i \in 1 : m\}$, she would update the coefficients β of her probability distribution over *market size* whose quantile function is

$$Q(\hat{p}) = \beta_1 + \beta_2\Phi^{-1}(\hat{p}) + \beta_3\hat{p}\Phi^{-1}(\hat{p}), \tag{6.13}$$

and the p-transform

$$\hat{p} = (1 - F(0))\, p + F(0) \tag{6.14}$$

truncates the distribution so that its support is $(0, \infty)$, in accordance with §4.3. As a consequence of the model given by (6.12) and its following paragraph, the likelihood function for the quantile data vector x given p, σ, and β is:

$$P\{x|p, \sigma, \beta\} \propto \sigma^{-m} \exp\left(-\frac{1}{2\sigma^2}(x - Y\beta)^T(x - Y\beta)\right), \tag{6.15}$$

where the matrix

$$Y = \begin{bmatrix} 1 & \Phi^{-1}(p_1) & p_1\Phi^{-1}(p_1) \\ \vdots & \vdots & \vdots \\ 1 & \Phi^{-1}(p_m) & \hat{p}_m\Phi^{-1}(p_m) \end{bmatrix}, \tag{6.16}$$

has the form of (3.6), and m is the total number of quantiles assessed from the expert.

One can think of this model as a Bayesian linear regression, where the regression function (6.12) has the form of the normal linear model $y = \beta^T x + \varepsilon$. Two early texts that cover Bayesian linear regression are Raiffa and Schlaifer [48] and Zellner [70]. Both show that the likelihood function (6.15) has conjugate prior distributions:

$$\frac{1}{\sigma^2} \sim gamma\left(a = \nu_0/2, b = \nu_0 s_0^2/2\right) \tag{6.17}$$

and

$$\beta|\sigma \sim N\left(\mu = \beta_0, \Sigma = \sigma^2\left(Y_0^T Y_0\right)^{-1}\right). \tag{6.18}$$

There are various methods of indirectly assessing the parameters ν_0, s_0^2, β_0, and Y_0 of these conjugate prior distributions. For example, Kadane et al. [32] elicit quantiles over the independent variable (the market size quantile of this example), given fixed independent variable(s) (the cumulative probability p, in this example). The conjugate Bayesian regression model assumes that the domain of the n-dimensional vector β is $\mathbf{R}^n$. However, according to Proposition 15, the feasible set of vectors S_β of a QPD is always a proper subset of $\mathbf{R}^n$. Therefore, this method of valuing probability assessment will never enjoy a Bayesian linear regression model with conjugate priors. Figure 6.12 shows the feasible region for $\beta_2 - \beta_3$ of the family of QPDs chosen for the expert's distribution over *market size*. The feasible region for the location parameter β_1 is always the real numbers.

6.5.2 Assessing Prior Distributions on Coefficients

The model is now well-structured, but a difficulty remains in assessing a prior probability distribution on β from (6.13)—it is not easy to think about the probabilistic relationships between β_1, β_2, and β_3. To address this, the CEO and decision analyst agree to simplify the model from 6.11 through two steps. First, the CEO declares that σ, the standard deviation of the error term ε, is irrelevent to the coefficients β, because observing β will not change her probability distribution on σ. Second, the CEO asserts that while observing a set of the expert's quantiles on *market size*

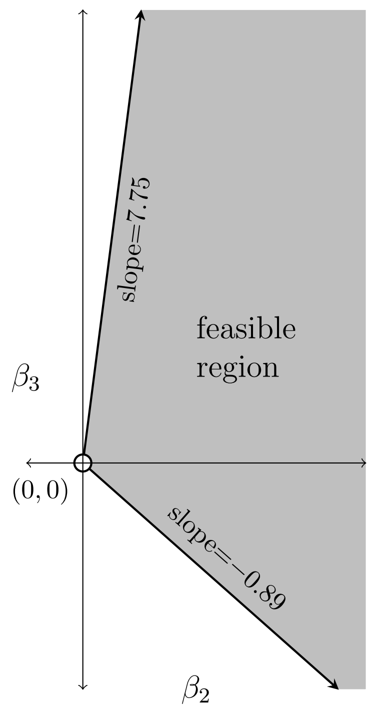

Figure 6.12: The feasible $\beta_2 - \beta_3$ region for the family of QPDs with basis functions $\{1, \Phi^{-1}(p), p\Phi^{-1}(p)\}$

may change the median and interdecile range (0.9-quantile minus 0.1-quantile) of her distribution, it will not change its shape. Equivalently, without the x-axis labels, her posterior distribution will be indistinguishable from Figure 6.5, her prior distribution. Moreover, observing the expert's median will not change her distribution over his interdecile range. In response to these thoughts, the decision analyst changes the normal linear model (6.12) to

$$x_i = \theta_1 + \theta_2 \left(\Phi^{-1}(p_i) + \frac{10}{11} p_i \Phi^{-1}(p_i) \right) + \varepsilon_i. \tag{6.19}$$

He then reformulates the CEO's probability distribution over the *market size* uncertainty of (6.13) into the location-scale family[2]

$$Q(\hat{p}) = \theta_1 + \theta_2 \left(\Phi^{-1}(\hat{p}) + (10/11)\hat{p}\Phi^{-1}(\hat{p}) \right), \quad \hat{p} = (1 - F(0))\, p + F(0). \tag{6.20}$$

[2]Recall the discussion of location and scale parameters in §2.3.

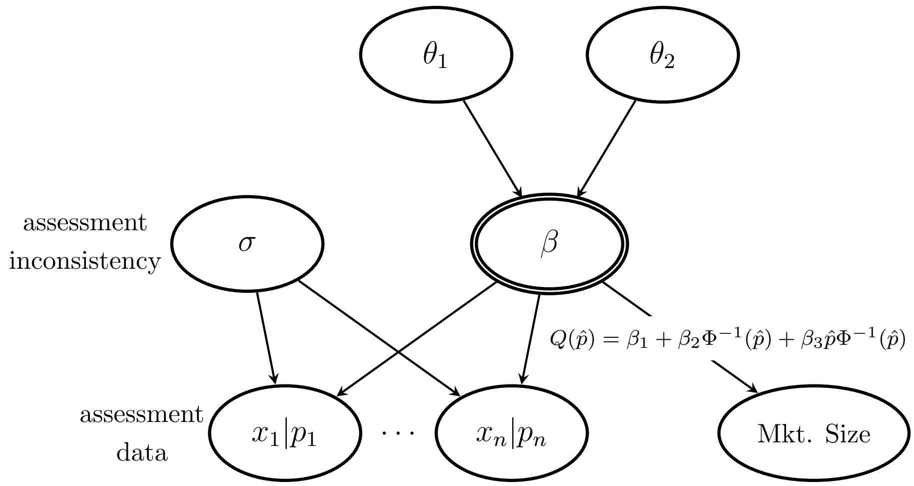

Figure 6.13: The CEO's modified relevance diagram describing the expert's assessed quantiles on the *market size* uncertainty

where the location parameter $\theta_1 = \beta_1$, the scale parameter, $\theta_2 = \beta_2$, and 10/11 is the ratio β_3/β_2 computed from the CEO's quantiles from Table 6.2. This ratio ensures that the CEO's posterior probability distribution on *market share* will have the same shape as her prior. The relevance diagram of Figure 6.13 depicts this model. Note that the arc from σ to β that is present in Figure 6.11 is absent in Figure 6.13. Also, there is no arc between θ_1 and θ_2, indicating that these distinctions are irrelevant given her current state of knowledge. Finally, the β uncertainty node from Figure 6.11 is now a deterministic node, recognizing the new functional relationship $(\beta_1, \beta_2, \beta_3) = (\theta_1, \theta_2, \frac{10}{11}\theta_2)$ that (6.20) implies.

Recall that the CEO is uncertain about the median and interquartile range of the expert's quantiles on the *market size* uncertainty. By (6.20), the expert's median on *market size* equals θ_1 plus the zero mean, normally distributed error term ε. He creates the clarity test definition for θ_1 as the answer to the question, "What median will the expert assign for the *market size* uncertainty?" After assessing the CEO's quantiles on θ_1, the decision analyst encodes the CEO's knowledge as the normal distribution with $\mu_{\theta_1} = 50$, and $\sigma^2_{\theta_1} = 25$ whose PDF Figure 6.14 depicts.

Also by (6.20), the expert's interdecile range on *market size* approximately equals $3.5\theta_2$ plus a normally distributed error term that has mean zero and double the

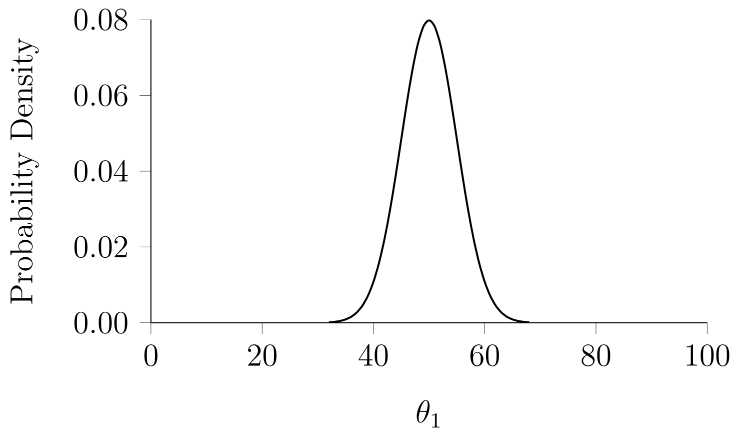

Figure 6.14: The CEO's prior on the location parameter θ_1 of the expert's probability distribution for *market size*

variance of ε. He creates the clarity test definition for θ_2 as the answer to the question, "What interdecile range will the expert assign for the *market size* uncertainty?" The PDF

$$f(x) = \frac{2b^a c^{2a}}{\Gamma(a) x^{2a+1}} \exp\left(-b\left(\frac{c}{x}\right)^2\right), \quad x > 0, \tag{6.21}$$

characterizes the inverse-squareroot-gamma distribution with scale parameter c. See Zellner [70, pages 371–373] for more information on this distribution—he terms it an "inverse gamma" distribution. Zellner examines this distribution because it is the conjugate prior on the standard deviation σ of the error term ε of a Bayesian linear regression. After assessing the CEO's quantiles on her uncertainty about the expert's interdecile range and dividing them by 3.5, the decision analyst encodes the CEO's probability distribution on the scale parameter θ_2 as an inverse-squareroot-gamma distribution with parameters $a_{\theta_2} = 22$, $b_{\theta_2} = 1$, and $c_{\theta_2} = 100$. Figure 6.15 depicts the CEO's PDF on θ_2.

In order to be consistent with the CEO's original probability distribution on *market size*, $\mathrm{E}[\theta_1]$ and $\mathrm{E}[\theta_2]$ should be 50 and 21.5, respectively. These are the values for β_1 and β_2 from the CEO's original quantiles on *market size*. Her distributions' first moments are $\mathrm{E}[\theta_1] = 50$ and $\mathrm{E}[\theta_2] = 21.7$—an acceptably small deviation.

The third and final assessment is the CEO's knowledge about σ, the standard

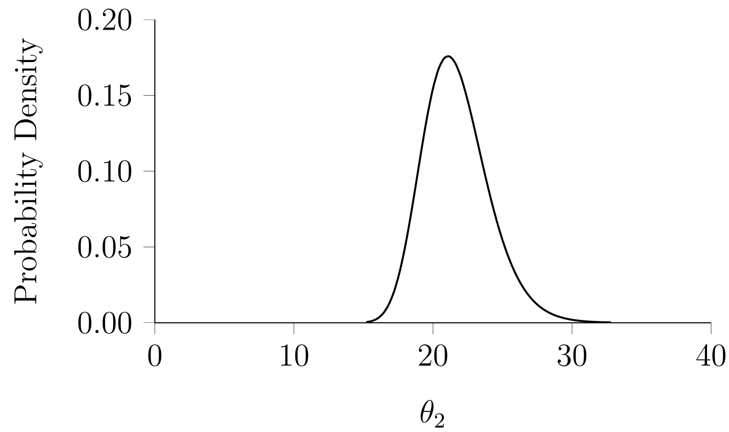

Figure 6.15: The CEO's prior on the scale parameter θ_2 of the expert's probability distribution for *market size*

deviation of the error term ε in (6.19). He creates the clarity test definition for σ as the answer to the question, "What is the sample standard deviation from two or more conditionally irrelevant median assessments on *market size* that the expert would assign?" Once again, he uses the CEO's quantiles on σ to parameterize the inverse-squareroot-gamma distribution with $a_{\theta_2} = 2$, $b_{\theta_2} = 5$, and $c_{\theta_2} = 1.25$. Figure 6.16 depicts the CEO's PDF on σ.

Now the decision analyst is ready to value assessing the expert's *market size* deciles. Such an assessment is material to the CEO's decision whenever a possible assessment exists that might change her distribution on *market size* such that the CEO changes her decision from *License* to *Market*. The CEO will first decide whether or not to pay for the expert's services. After updating her distribution on *market size* with these new quantile assessment data, the CEO will make her decision whether to *Market* or *License* the drug. Figure 6.17 depicts her modified decision situation.

Because this normal linear model lacks conjugate priors, the decision analyst must simulate her valuation of the expert's quantiles on *market size*. He first draws $10,000$ samples from the pre-posterior distribution, which is the CEO's marginal distribution over the quantiles $\tilde{x}|p$ he might elicit from the expert. He then draws $1,000$ samples from her posterior distributions of the location parameter θ_1 and the scale parameter θ_2, given each set of pre-posterior quantile assessment data $\tilde{x}|p$ by way

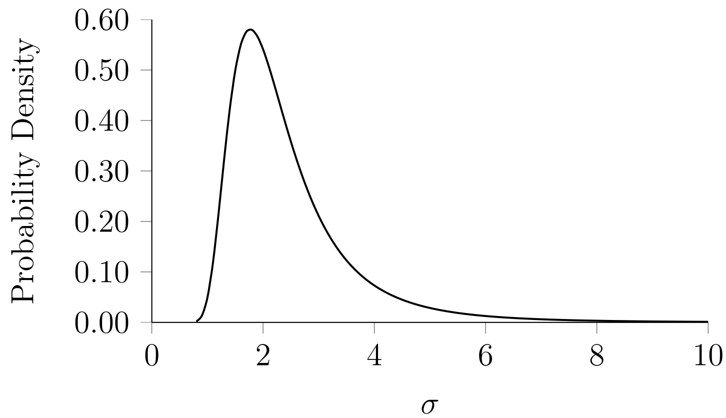

Figure 6.16: The CEO's prior on the standard deviation σ of the expert's assessment inconsistency ε

of the Metropolis-Hastings algorithm.[3] Next, he computes her expected utility for *Market* and *License* by simulating from *years to half-peak share*, *peak market share*, and *market size* and recording the alternative with the highest expected utility. He repeats this process for each posterior $\theta_1|\tilde{x}, p$ and $\theta_2|\tilde{x}, p$ that he simulated previously. He takes the expectation of these utilities over all $\theta_1, \theta_2|\tilde{x}, p$ to compute an expected utility for valuing the expert's deciles.

Since hers is a delta-property utility function [4][23, pages 214–215], her value of probability assessment is the difference between her value of free probability assessment minus her original certain equivalent for the *License* alternative—a difference that is approximately \$2.3 million. The expert's \$200, 000 fee, once regarded as expensive, is clearly a good deal. The decision analyst recommends that the CEO contract the expert's services, update her distribution on *market size* given his deciles, and choose the best alternative given her new information.

[3]The Metropolis-Hastings algorithm is a method that belongs to the Markov Chain Monte Carlo class of probabilistic simulation. Gelman et al. [16] give advice for tuning this algorithm.

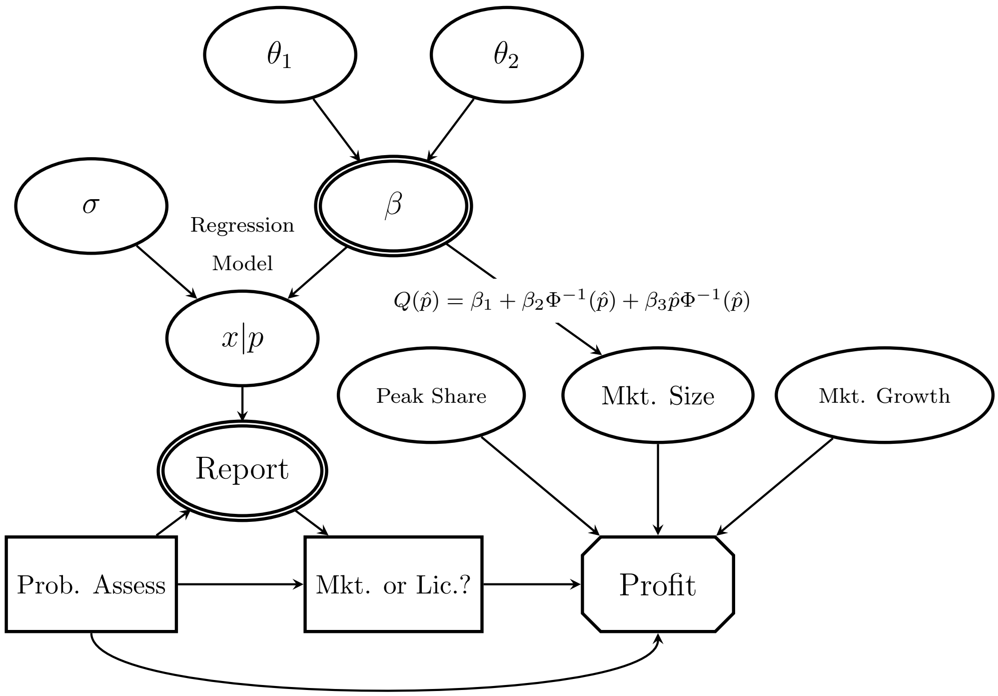

Figure 6.17: The decision diagram for valuing probability assessment

6.6 Summary

This elementary decision analysis shows the power of the QPD toolkit. It demonstrates the encoding of prior probability distributions using quantile-probability data, whether or not the underlying uncertain quantity has well-defined bounds. It highlights the ability of QPDs to represent a large number of quantile-probability points, such as the output of a probabilistic simulation, with very few coefficients. The sensitivity analysis shows a method of using QPDs and the theory of tail behavior to demonstrate the robustness of the *License* alternative to changes in the tail-heaviness of the CEOs distribution on *market size*. It shows how QPDs can ease the valuation of probability assessment by using quantile-probability data.

Chapter 7

Conclusion

7.1 Summary

Quantile-parameterized distributions are a new class of probability distributions that are both readily parameterizable by quantiles and amenable to discretization and probabilistic simulation. The Quantile Parameters Theorem is a key result, showing the conditions required for parameterizing a QPD using quantile-probability pairs. Another important characteristic of a QPD is its region of parametric feasibility. While infeasible parameters always exist, the feasible parametric region is convex. The simple Q-normal distribution serves as a demonstration of these results, and despite its simplicity, it carries the flexibility to represent a wide range of distributional shapes. One can parameterize a QPD using overdetermined systems of equations by minimizing an appropriate norm. One can encode relevance with QPDs by way of discretization and conditioning, by creating a QPD prior over the coefficients of a QPD likelihood function, or by relating two QPD marginal distributions by a copula. These methods open the door to probabilistic inference using QPDs.

Continuous uncertain quantities often have well-defined limits. As an example, the volume of recoverable oil in a reservoir cannot be less than zero. In such cases, the decision analyst desires a probability distribution that reflects these limits. The QPD support thepborem shows the effect that a particular set of basis functions has on the support of a QPD. One can engineer the support of a QPD using the following three methods. The first is to set up a constrained optimization problem to solve for a QPD's coefficients. The QPD resulting from this method will pass through the extreme quantiles, but it is not guaranteed to pass through the original quantile-probability points. The second is to control support through truncation. When using this method, the coefficients need not be recalculated, and probabilistic simulation is possible through rejection sampling. However, the resulting QPD is not guaranteed to pass through the quantile-probability points and the chopped-off tails may be a poor representation of the decision maker's beliefs. The third method is to create a transformed QPD. The probability distribution resulting from this method must pass through both the extreme and intermediate quantiles. This does not mean that it is always the best approach—as with any probability encoding process, the encoder should present the candidate probability distribution to the decision maker in order to verify that it is an acceptable representation of her beliefs.

When continuous uncertain quantities do not have well-defined limits, an alternative to extreme quantile assessment is to allow a distribution to have infinite right and/or left tails. The heaviness of the infinite tail of a decision maker's probability distribution may be material to her decision. There are four key contributions of this section. The first is a definition of what it means for one probability distribution to have *heavier tails* than another. The second is the binary relation over the set of univariate probability distributions with twice differentiable quantile functions (R- and L-ordering.) These orderings imply relative tail heaviness—if $F \prec_R [\prec_L] G$, then G has a heavier right [left] tail than F. The third is a set of tail bounds for any QPD with positive coefficients. The fourth contribution is the QPD tails theorem—a statement showing how to choose a new basis function, which, when added to a set of basis functions, is guaranteed to result in a QPD with heavier tails.

The pharmaceutical company CEO's market versus license decision demonstrates the use of QPDs in a decision analysis. All prior probabilities are encoded using QPDs, both truncated and transformed to be consistent with the limits of each variable. The key results from the analysis are three novel QPD-based methods for decision analysis. The first is a data compression method that encodes the information from the probabilistic simulation over the *peak market share–value* joint distribution—one that reduces four million datapoints to eleven parameters. The second is a sensitivity analysis to tail heaviness that shows the decision maker whether the right tail heaviness of the *market share* uncertainty is material to her decision. The third is a valuation of probability assessment, which not only shows the decision analyst what it is worth to assess further quantiles, but also shows how to apply probabilistic inference using QPDs.

7.2 Areas for Future Research

7.2.1 Promising Directions

The areas for future research span a few different disciplines. The first is in the field of artificial intelligence—specifically automated probability encoding systems for continuous uncertain quantities. Recall that this dissertation offers no axioms for the selection of basis functions. Instead, it identifies support and tail behavior as the important characteristics to consider when choosing a probability distribution consistent with quantile-probability data, and it shows how the choice of basis functions and transforms influence these two characteristics. Using this theory, one might also consider an automated system that chooses from a vast library of basis function candidates based on the input of quantile-probability data, support bounds, and/or desired tail behavior. A method of choosing such functions is to minimize a sum squared deviations with a regularization term [11, §6.3.2] that is the sum of absolute values of the components of the coefficient vector. A mathematical formulation of this problem is

$$\hat{\beta} = \text{argmin}_{\beta}\{\|Y\beta - x\|_2 + \gamma\|\beta\|_1\},$$

where $\|v\|_p$ is the p-norm of a vector v. In the statistical literature, this approach has the name "lasso" regression [63]. The regularization term $\gamma\|\beta\|_1$ serves as a heuristic for reducing the number of basis functions by setting various components of β to zero. As a related method, support vector regression [59] also holds promise for the automated selection of basis functions.

A second research area is empirical research characterizing the probability elicitation of rare events using new graphics. The development of methods for debiasing and assessing extreme quantiles would allow for a better selection of basis functions to represent tail behavior. In addition, a characterization of the error of various methods of assessing extreme quantiles might lead to better models for valuing probability assessment. Other empirical research of value is the development and characterization of the verification stage of any quantile-probability based probability encoding method by using QPD-generated PDFs as feedback rather than CDFs.

A third area is further characterization of how sets of basis functions relate to parametric feasibility. Since the simple Q-normal covers such a diversity of distributional shapes, I spent little time characterizing the parametric coverage of different sets of basis functions. One possible measure of such coverage is the volume of the convex set of feasible normalized quantiles like the $r_1 - r_2$ space of Chapter 3. Further study of various sets of basis functions may yield insight about how to choose basis functions for various continuous uncertain quantities. Research comparing the quantile coverage of QPDs to four parameter distributions from the canon of commonly used probability distributions, such as the Pearson family, may stimulate the use of QPDs in research and in the practice of decision analysis.

The last promising research area of note deals with the theory of tail behavior detailed in Chapter 5. One path for future research is finding tighter bounds for the tail behavior of QPDs. Another is in relating parametric feasibility to the tail behavior of QPDs. Yet another is in the study of the function $Q''(p)/Q'(p)$. It is curious that the Cauchy distribution is invariant under the $Q''(p)/Q'(p)$ operation. It is also curious that $Q''(p)/Q'(p)$ is a quantile function for more than just log-concave probability distributions with quantile function $Q(p)$. Exploring these curiosities may lead to useful results.

7.2.2 Culs-de-sac?

Perhaps every Ph.D. student culls promising paths of research in the interest of completing his or her dissertation. This research is no different—there is much litter on the cutting room floor. In the goal of advancing human knowledge, it is important to identify halted research, as well as research success.

The function $Q''(p)/Q'(p)$ has a relationship to the unimodality of QPDs. I spent time investigating why QPDs like the simple Q-normal seem to yield unimodal distributions—especially compared with mixture distributions of CDFs. I have no general statements to make about the unimodality of QPDs at the time of this writing. Research into the source of what makes certain QPDs tend to be unimodal may help in choosing basis functions.

Another investigation that did not make the cut is that of quantile-parameterized utility. Abbas and Matheson introduce the notion of utility-probability duality to show (for continuous, increasing utility functions whose image is the unit interval) that the best of a set of alternatives has both the highest expected utility $\int_x U(x)dF(x)$ and the lowest expected disutility $\int_x F(x)dU(x)$ [7]. Many years earlier, Borch [10] applied a duality approach equivalent to Abbas's and Matheson's to show how a decision maker seeking to minimize his probability of ruin over a sequence of deals induces his utility function. Indeed, parallels between utility and probability appear frequently in economic and decision analytic literature. Abbas and Howard use probability concepts like information entropy and Bayes's theorem to infer a decision maker's utility function from a finite set of utility assessments [3, 6]. The very foundations of decision theory and decision analysis are axiomatized in a way that expresses utility in terms of an indifference probability between two uncertain deals and treats it as a probability (continuity axiom of von Neumann and Morgenstern [66], Howard's equivalence and substitution rules [27, 28]). Given this history of parallels between probability and utility, it follows that one might apply a *quantile-parameterized utility* approach (QPU) to encode a decision maker's risk preference much as one applies a QPD approach to encode his beliefs. This is indeed possible, but the QPU approach tends to perform poorly compared to positive linear combinations of utility functions, at least in one of the three desiderata applied most often to utility functions:

1. the decision maker's utility function is strictly increasing in wealth, $U'(x) > 0$;

2. the decision maker is risk averse over monetary prospects, $U''(x) < 0$;

3. the decision maker's risk aversion function $-U''(x)/U'(x)$ is strictly decreasing in wealth.

Take the inverse utility function $\hat{U}^{-1}(v) = \beta_1 U_1^{-1}(v) + \cdots + \beta_n U_n^{-1}(v)$, where the basis functions are the inverse functions of utility functions that each meet the three desiderata. While $\hat{U}$ meets the first two desiderata, it is not guaranteed to meet the third. However, a positive weighted sum of those basis utility functions $\tilde{U}(x) = \beta_1 U_1(x) + \cdots + \beta_n U_n(x)$, meets all three.

The question mark in the title of this subsection indicates that these research directions may not be dead ends. These, along with the research directions of §7.2.1, may lead to results important to the theory and practice of decision analysis. Indeed, I hope this proves to be so.

Appendix A

Selected Proofs

Proof of the QPD Support Theorem (Theorem 2). This proof relies on the fact that a QPD consists of a regular set of basis functions (definition 7).

Begin with item 1. If X_i is a finite interval for all $i \in \{1 : n\}$, then the quantile function $Q(p)$ of QPD F is a linear combination of finite numbers, therefore $Q(p)$ is finite for all $p \in (0, 1)$ and $\beta \in \mathbf{R}^n$. By Proposition 19, supp(F) is finite.

If supp(F) of a QPD is finite, by Proposition 19, a linear combination of its basis functions is finite for all $p \in (0, 1)$ and $\beta \in \mathbf{R}^n$. Because a QPD's set of basis functions is regular, it is impossible for a linear combination of two basis functions with infinite tails to cancel each other out. This implies each of its basis functions must map to a finite interval for all $p \in (0, 1)$.

Continue with item 2. Take a linear combination of all basis functions X_j, $j \neq i$. The proof for item 1 shows that the image $\tilde{X}_{-i} = \{\sum_{j \neq i} \beta_j g_j(p) \mid p \in (0, 1), \beta \in \mathbf{R}^{n-1}\}$ is a finite interval. The Minkowski sum of a finite interval ($\tilde{X}_{-i}$) and a semi-infinite interval ($\{\beta_i g_i(p) \mid p \in (0, 1), \beta_i \in \mathbf{R}\}$) yields a semi-infinite interval, making supp(F) semi-infinite.

Conclude with item 3. For any QPD, $\lim_{p \to 0} \sum_{i=1}^{n} \beta_i g_i(p) \in \mathbf{R}$ or approaches $-\infty$ and $\lim_{p \to 1} \sum_{i=1}^{n} \beta_i g_i(p) \in \mathbf{R}$ or approaches ∞. It suffices to show that if the image of one or more basis functions is $(-\infty, \infty)$, then the first limit approaches $-\infty$, and the second limit approaches ∞. Proof by contradiction: assume that the first and second limits belong to the set of real numbers. Because a QPD's set of basis functions is regular, the limit of each of its basis functions $\lim_{p\to 1} g_i(p) \in \mathbf{R}$, for all $i \in \{1:n\}$ and $p \in (0,1)$. By the same argument, $\lim_{p\to 0} g_i(p) \in \mathbf{R}$, for all $i \in \{1:n\}$ and $p \in (0,1)$. This contradicts the original statement that the image of at least one basis function $X_i = (-\infty, \infty)$. $\square$

Proof of Proposition 22. It suffices to show that $Q''(p)/Q'(p)$ is nondecreasing. By log-concavity,

$$
\begin{aligned}
& \frac{d^2}{dx^2}\log(f(x)) \le 0 \\
\Leftrightarrow\ & \frac{f''(x)}{f(x)} - \left(\frac{f'(x)}{f(x)}\right)^2 \le 0 \\
\Leftrightarrow\ & \frac{f''(Q(p))}{f(Q(p))} - \left(\frac{f'(Q(p))}{f(Q(p))}\right)^2 \le 0 \text{ , substitute } x = Q(p) \\
\Rightarrow\ & \frac{f''(Q(p))}{f(Q(p))}(Q'(p))^2 - 2\left(\frac{f'(Q(p))}{f(Q(p))}\right)^2 (Q'(p))^2 \le 0 \text{ , since } (Q'(p))^2 \ge 0 \\
\Leftrightarrow\ & \frac{d^2}{dp^2}\log(f(Q(p))) \le 0 \\
\Leftrightarrow\ & \frac{d}{dp}\log(f(Q(p))) \text{ is nonincreasing} \\
\Leftrightarrow\ & \frac{d}{dp}\log\left(\frac{1}{Q'(p)}\right) \text{ is nonincreasing, using (2.2)} \\
\Leftrightarrow\ & \frac{Q''(p)}{Q'(p)} \text{ is increasing}
\end{aligned}
$$

The substitution of $x = Q(p)$ holds because F is a continuous probability distribution with a twice-differentiable quantile function, therefore $Q : (0,1) \to \mathrm{dom}(F)$ is a bijection. $\square$

Proof of Proposition 23. By Proposition 21, it suffices to show that

1. There exists a point p_0 such that $\frac{\tilde{Q}''(p)}{\tilde{Q}'(p)} < \frac{\frac{d^2}{dp^2}\left((\sum_{i=1}^n \beta_i) Q_n(p)\right)}{\frac{d}{dp}\left((\sum_{i=1}^n \beta_i) Q_n(p)\right)}, p \in (p_0, 1)$

2. There exists a point p_1 such that $\tilde{Q}'(p_1) < \frac{d}{dp}\left((\sum_{i=1}^n \beta_i) Q_n(p)\right)\Big|_{p_1}, p_1 \in (p_0, 1)$

Begin with item 1. Because $F_i \prec_R F_n$, all $i \in \{1 : n-1\}$ and by Proposition 21, we know that there exists a point p_0^i such that $\frac{Q_i''(p)}{Q_i'(p)} < \frac{Q_n''(p)}{Q_n'(p)}$, $p \in (p_0^i, 1), i \in \{1 : n-1\}$. Let $p_0 = \max_i\{p_0^i\}$ and restrict the following equations so that $p \in (p_0, 1)$:

$$\begin{aligned}
\frac{\tilde{Q}''(p)}{\tilde{Q}'(p)} &= \frac{\beta_1 Q_1''(p) + \cdots + \beta_n Q_n''(p)}{\beta_1 Q_1'(p) + \cdots + \beta_n Q_n'(p)} \\
&= \left(\frac{\beta_1 Q_1'(p)}{\sum_{i=1}^n \beta_i Q_i'(p)}\right) \frac{Q_1''(p)}{Q_1'(p)} + \cdots + \left(\frac{\beta_n Q_n'(p)}{\sum_{i=1}^n \beta_i Q_i'(p)}\right) \frac{Q_n''(p)}{Q_n'(p)} \\
&< \left(\frac{\beta_1 Q_1'(p)}{\sum_{i=1}^n \beta_i Q_n'(p)}\right) \frac{Q_n''(p)}{Q_n'(p)} + \cdots + \left(\frac{\beta_n Q_n'(p)}{\sum_{i=1}^n \beta_i Q_i'(p)}\right) \frac{Q_n''(p)}{Q_n'(p)} \\
&= \frac{Q_n''(p)}{Q_n'(p)} \\
&= \frac{\frac{d^2}{dp^2}\left((\sum_{i=1}^n \beta_i) Q_n(p)\right)}{\frac{d}{dp}\left((\sum_{i=1}^n \beta_i) Q_n(p)\right)}
\end{aligned}$$

Continue with item 2. By Proposition 21, we know that there exists a point p_1^i such that $Q_i'(p_1) < Q_n'(p_1)$, $p_1 \in (p_0, 1)$. Let $p_1 = \max_i\{p_1^i\}$. Show that $\tilde{Q}'(p_1) < \left(\sum_{i=1}^n \beta_i\right) Q_n'(p_1), p_1 \in (p_0, \infty)$.

$$\begin{aligned}
\tilde{Q}'(p_1) &= \beta_1 Q_1'(p_1) + \cdots + \beta_n Q_n'(p_1) \\
&< \beta_1 Q_n'(p_1) + \cdots + \beta_n Q_n'(p_1) \\
&= \left(\sum_{i=1}^n \beta_i\right) Q_n'(p_1).
\end{aligned}$$

The second step makes use of the fact that $Q_F'(p_1) < Q_G'(p_1)$, $p_1 \in (p_0, 1)$ if and only if $Q_F'(p) < Q_G'(p)$, $p \in (p_1, 1)$ because $\phi(x)$ is convex. $\square$

Proof of Proposition 24. Again, by Proposition 21, it suffices to show that

1. There exists a point p_0 such that $\frac{\frac{d^2}{dp^2}\left((\sum_{i=1}^n \beta_i) Q_1(p)\right)}{\frac{d}{dp}\left((\sum_{i=1}^n \beta_i) Q_1(p)\right)} < \frac{\tilde{Q}''(p)}{\tilde{Q}'(p)}, p \in (p_0, 1)$
2. There exists a point p_1 such that $\frac{d}{dp}\left((\sum_{i=1}^n \beta_i) Q_1(p)\right)\Big|_{p_1} < \tilde{Q}'(p_1), p_1 \in (p_0, 1)$

Begin with item 1. Because $F_1 \prec_R F_i$, all $i \in \{2 : n\}$ and by Proposition 21, we know that there exists a point p_0^i such that $\frac{Q_1''(p)}{Q_1'(p)} < \frac{Q_i''(p)}{Q_i'(p)}$, $p \in (p_0^i, 1), i \in \{2 : n\}$.

Let $p_0 = \max_i\{p_0^i\}$ and restrict the following equations so that $p \in (p_0, 1)$:

$$\begin{aligned}
\frac{\frac{d^2}{dp^2}\left((\sum_{i=1}^n \beta_i)Q_1(p)\right)}{\frac{d}{dp}\left((\sum_{i=1}^n \beta_i)Q_1(p)\right)} &= \frac{Q_1''(p)}{Q_1'(p)} \\
&< \left(\frac{\beta_1 Q_1'(p)}{\sum_{i=1}^n \beta_i Q_i'(p)}\right)\frac{Q_1''(p)}{Q_1'(p)} + \cdots + \left(\frac{\beta_n Q_n'(p)}{\sum_{i=1}^n \beta_i Q_i'(p)}\right)\frac{Q_n''(p)}{Q_n'(p)} \\
&= \frac{\beta_1 Q_1''(p) + \cdots + \beta_n Q_n''(p)}{\beta_1 Q_1'(p) + \cdots + \beta_n Q_n'(p)} \\
&= \frac{\tilde{Q}''(p)}{\tilde{Q}'(p)}
\end{aligned}$$

The proof for item 2 is analogous to the proof for item 2 of Proposition 23:

$$\begin{aligned}
\left(\sum_{i=1}^n \beta_i\right) Q_1'(p_1) &= \beta_1 Q_1'(p_1) + \cdots + \beta_n Q_1'(p_1) \\
&< \beta_1 Q_1'(p_1) + \cdots + \beta_n Q_n'(p_1) \\
&= \tilde{Q}'(p_1)
\end{aligned}$$

□

Proof of Proposition 25. Let $Q_+(p) = \sum_{i\in\mathcal{I}_+} \beta_i Q_i(p)$ and $Q_-(p) = -\sum_{i\in\mathcal{I}_-} \beta_i Q_i(p)$ making $\tilde{Q}(p) = Q_+(p) - Q_-(p)$. By Proposition 4, $Q_+(p)$ is a quantile function because it is a positive weighted sum of quantile functions. Likewise, $Q_-(p)$ is a quantile function because it is the negative of a negative weighted sum of quantile functions.

Let F_+ be the probability distribution whose quantile function is $Q_+(p)$. If $\mathcal{I}_+$ has more than one member, then by Proposition 23, $F_+ \prec_R \tilde{F}_n$; by Proposition 20, $\tilde{F}_n$ has heavier right tails than F_+; and so there exists a point p_0 such that $Q_+(p) < \left(\sum_{i\in\mathcal{I}_+} \beta_i\right) Q_n(p),\ p \in (p_0, 1)$. However, it may be true that $\mathcal{I}_+$ is the singleton $\{n\}$, making $Q_+(p) = \left(\sum_{i\in\mathcal{I}_+} \beta_i\right) Q_n(p)$ for all $p \in (0, 1)$. Regardless of the members of $\mathcal{I}$, it is true that $Q_+(p) \le \left(\sum_{i\in\mathcal{I}_+} \beta_i\right) Q_n(p),\ p \in (p_0, 1)$. Include an arbitrarily small positive quantity ε so that $Q_+(p) < \left(\sum_{i\in\mathcal{I}_+} \beta_i + \varepsilon\right) Q_n(p),\ p \in (p_0, 1)$.

Since $Q_-(p)$ is a quantile function, $\lim_{p\to 1}(-Q_-(p)) \in \{\mathbf{R}, -\infty\}$. In cases where this limit is either a negative real number or $-\infty$, set $\kappa = \sum_{i\in\mathcal{I}_+}\beta_i + \varepsilon$. If, however, the limit is a positive real number, choose an upper bound $B \in (0,\infty)$ so that $B > -Q_-(p),\ p \in (p_0, 1)$. This is possible because $Q_-(p)$ is a quantile function and is therefore nondecreasing in p. Now set $\kappa = \sum_{i\in\mathcal{I}_+}\beta_i + B + \varepsilon$. $\square$

References

[1] Ali E. Abbas. Entropy Methods for Univariate Distributions in Decision Analysis. In C. Williams, editor, *Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 22nd International Workshop*, pages 339–349, Moscow, ID, 2003. American Institute of Physics.

[2] Ali E. Abbas. *Entropy Methods in Decision Analysis*. Ph.D. dissertation, Stanford University, 2003.

[3] Ali E. Abbas. Maximum entropy utility. *Operations Research*, 54(2):277–290, 2006.

[4] Ali E. Abbas. Invariant Utility Functions and Certain Equivalent Transformations. *Decision Analysis*, 4(1):17–31, 2007.

[5] Ali E. Abbas, David V. Budescu, Hsiu-Ting Yu, and Ryan Haggerty. A Comparison of Two Probability Encoding Methods: Fixed Probability fs. Fixed Variable Values. *Decision Analysis*, 5(4):190–202, 2008.

[6] Ali E. Abbas and Ronald A. Howard. Attribute Dominance Utility. *Decision Analysis*, 2(4):185–206, 2005.

[7] Ali E. Abbas and James E. Matheson. Normative Target-Based Decision Making. *Managerial and Decision Economics*, 26:373–385, 2005.

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

[9] J. Eric Bickel. Some Determinants of Corporate Risk Aversion. *Decision Analysis*, 3(4):233–251, 2006.

[10] Karl Borch. Economic Objectives and Decision Problems. *IEEE Transactions on Systems Science and Cybernetics*, 4(3):266–270, 1968.

[11] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, U.K., 2009.

[12] Robert T. Clemen, Gregory W. Fischer, and Robert L. Winkler. Assessing Dependence: Some Experimental Results. *Management Science*, 46(8):1100–1115, 2000.

[13] Robert T. Clemen and Terence Reilly. Correlations and Copulas for Decision and Risk Analysis. *Management Science*, 45(2):208–224, 1999.

[14] Angela Fagerlin, Catharine Wang, and Peter A. Ubel. Reducing the Influence of Anecdotal Reasoning on People's Health Care Decisions: Is a Picture Worth a Thousand Statistics? *Medical Decision Making*, 25:398–405, 2005.

[15] Joseph L. Gastwirth. On Robust Rank Tests. In M.L. Puri, editor, *Nonparametric Techniques in Statistical Inference*, pages 89–109. Cambridge University Press, Cambridge, England, 1970.

[16] Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman & Hall/CRC, Boca Raton, FL, 2nd edition, 2004.

[17] Warren Gilchrist. *Statistical Modelling with Quantile Functions*. CRC Press, 2000.

[18] Jaroslav Hájek. *A course in nonparametric statistics*. Holden-Day, San Francisco, 1969.

[19] Rebecca Hess, Vivianne H.M. Visschers, and Michael Siegrist. Risk communication with pictographs: The role of numeracy and graph procesing. *Judgment and Decision Making*, 6(3):263–274, 2011.

[20] Thomas P. Hettmansperger and Michael A. Keenan. Tailweight, Statistical Inference and Families of Distributions—a Brief Survey. In C. Taillie, Ganapati P. Patil, and Bruno Baldessari, editors, *Statistical Distributions in Scientific Work: Models, Structures, and Characterizations*, pages 161–171. D. Reidel, Dordrecht, Holland, 1981.

[21] Ronald A. Howard. Decision Analysis: Applied Decision Theory. In Hertz D B and Melese J, editors, *Proceedings of the 4th International Conference on Operational Research*, pages 55–77. Wiley-Interscience, 1966.

[22] Ronald A. Howard. Information Value Theory. *IEEE Transactions on Systems Science and Cybernetics*, 2(1):22–26, 1966.

[23] Ronald A. Howard. The Foundations of Decision Analysis. *IEEE Transactions on Systems Science and Cybernetics*, 4(3):211–219, 1968.

[24] Ronald A. Howard. Risk Preference. In James E. Matheson and Ronald A. Howard, editors, *Readings in Decision Analysis*, pages 429–465. Stanford Research Institute, Menlo Park, CA, 2nd edition, 1977.

[25] Ronald A. Howard. The Evolution of Decision Analysis. In Ronald A. Howard and James E. Matheson, editors, *Readings on the Principles and Applications of Decision Analysis*. Stanford Research Institute, 1st edition, 1983.

[26] Ronald A. Howard. Decision Analysis: Practice and Promise. *Management Science*, 34(6):679–695, 1988.

[27] Ronald A. Howard. Speaking of Decisions: Precise Decision Language. *Decision Analysis*, 1(2):71–78, 2004.

[28] Ronald A. Howard. Foundations of Decision Analysis Revisited. In Ward Edwards, Ralph F. Miles, and Detlof Von Winterfeldt, editors, *Advances in Decision Analysis: From Foundations to Applications*, pages 32–56. Cambridge University Press, 2007.

[29] Edwin T. Jaynes. Prior Probabilities. *IEEE Transactions on Systems Science and Cybernetics*, 4(3):227–241, 1968.

[30] Harold Jeffreys. An Invariant Form for the Prior Probability in Estimation Problems. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 186(1007):453–461, 1946.

[31] Victor R. José and Robert L. Winkler. Evaluating Quantile Assessments. *Operations Research*, 57(5):1287–1297, 2009.

[32] Joseph B. Kadane, James M. Dickey, Robert L. Winkler, Wayne S. Smith, and Stephen C. Peters. Interactive Elicitation of Opinion for a Normal Linear Model. *Journal of the American Statistical Association*, 75(372):845–854, 1980.

[33] Juha Karvanen. Estimation of quantile mixtures via L-moments and trimmed L-moments. *Computational Statistics & Data Analysis*, 51(2):947–959, 2006.

[34] Thomas W. Keelin and Bradford W. Powley. Quantile-Parameterized Distributions. *Decision Analysis*, 8(3):206–219, 2011.

[35] Pierre-Simon Laplace. *A Philosophical Essays on Probabilities.* J. Wiley, New York, 1 edition, 1902.

[36] Averill M. Law. *Simulation Modeling and Analysis.* McGraw-Hill, 4th edition, 2007.

[37] Dennis V. Lindley, Amos Tversky, and Rex V. Brown. On the Reconciliation of Probability Assessments. *Journal of the Royal Statistical Society Series A General*, 142(2):146–180, 1979.

[38] Douglas M. Logan. *Value of Probability Assessment.* Ph.D. dissertation, Stanford University, 1985.

[39] James E. Matheson and Robert L. Winkler. Scoring Rules for Continuous Probability Distributions. *Management Science*, 22(10):1087–1096, 1976.

[40] Allen C. Miller III and Thomas R. Rice. Discrete Approximations of Probability Distributions. *Management Science*, 29(3):352–362, 1983.

[41] Peter A. Morris. *Bayesian Expert Resolution*. Ph.D. dissertation, Stanford University, 1971.

[42] Peter A. Morris. Decision Analysis Expert Use. *Management Science*, 20(9):1233–1241, 1974.

[43] Roger B. Nelsen. *An Introduction to Copulas*. Springer, New York, 2nd edition, 2006.

[44] Emanuel Parzen. Nonparametric Statistical Data Modeling. *Journal of the American Statistical Association*, 74(365):105–121, 1979.

[45] Egon S. Pearson and John W. Tukey. Approximate Means and Standard Deviations Based on Distances between Percentage Points of Frequency Curves. *Biometrika*, 52(3):533–546, 1965.

[46] William B Poland. *Decision Analysis with Continuous and Discrete Variables: a Mixture Distribution Approach*. Ph.D. dissertation, Stanford University, 1994.

[47] Howard Raiffa. *Decision Analysis: Introductory Lectures on Choice Under Uncertainty*. Addison-Wesley, Oxford, England, 1968.

[48] Howard Raiffa and Robert Schlaifer. *Applied Statistical Decision Theory*. Division of Research, Graduate School of Business Adminitration, Harvard University, Boston, 1961.

[49] Frank P. Ramsey. Truth and Probability (1926). In *The Foundations of Mathematic and other Logical Essays*, chapter VII, pages 156–198. edited by R. B. Braithwaite, London: Kegan, Paul, Trench, Trubner & Co., New York: Harcourt, Brace & Co., 1931. 1999 electronic edition.

[50] Terence Reilly. Estimating moments of subjectively assessed distributions. *Decision Sciences*, 33(1):133–147, 2002.

[51] Audun S. Runde. *Estimating Distributions, Moments, and Discrete Approximations of a Continuous Random Variable Using Hermite Tension Splines*. Ph.D. dissertation, University of Oregon, 1997.

[52] Eugene F. Schuster. Classification of Probability Laws by Tail Behavior. *Journal of the American Statistical Association*, 79(388):936–939, 1984.

[53] Ross D. Shachter. An ordered examination of influence diagrams. *Networks*, 20(5):535–563, August 1990.

[54] Ross D. Shachter and C. Robert Kenley. Gaussian Influence Diagrams. *Management Science*, 35(5):527–550, 1989.

[55] Ross D. Shachter and Mark A. Peot. Simulation Approaches to General Probabilistic Inference on Belief Networks. *Uncertainty and Artificial Intelligence*, 5:221–231, 1989.

[56] Paul Slovic, Baruch Fischoff, and Sarah Lichtenstein. Behavioral Decision Theory. *Ann. Rev. Psychol.*, 28:1–39, 1977.

[57] James E. Smith. *Moment Methods for Decision Analysis*. Ph.D. dissertation, Stanford University, 1990.

[58] James E. Smith. Moment Methods for Decision Analysis. *Management Science*, 39(3):340–358, 1993.

[59] Alex J. Smola and Bernhard Schölkopf. A tutorial on support vector regression. *Statistics and Computing*, 14:199–222, 2004.

[60] Carl Spetzler. Establishing a Corporate Risk Policy. In James E. Matheson and Ronald A. Howard, editors, *Readings in Decision Analysis*, pages 237–245. Stanford Research Institute, Menlo Park, CA, 2 edition, 1977.

[61] Carl S. Spetzler and Carl-Axel S. Staël von Holstein. Probability encoding in decision analysis. *Management Science*, 22(3):340–358, 1975.

[62] Steven N. Tani. A Perspective on Modeling in Decision Analysis. *Management Science*, 24(14):1500–1506, 1978.

[63] Robert Tibshirani. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society*, 58(1):267–288, 1996.

[64] John W. Tukey. Which Part of the Sample Contains the Information? *Proceedings of the National Academy of Sciences of the United States of America*, 53(1):127–34, January 1965.

[65] Amos Tversky and Daniel Kahneman. Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157):1124–1131, 1974.

[66] John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ, 1944.

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

[68] Willem Rutger van Zwet. *Convex transformations of random variables*, volume 3. Mathematisch Centrum, Amsterdam, 1964.

[69] Robert L. Winkler. The Assessment of Prior Distributions in Bayesian Analysis. *Journal of the American Statistical Association*, 62(319):776–800, 1967.

[70] Arnold Zellner. *An Introduction to Bayesian Inference in Econometrics*. J. Wiley, New York, 1971.

[71] Brian J. Zikmund-Fisher, Angela Fagerlin, and Peter A. Ubel. Improving Understanding of Adjuvant Therapy Options Via Simpler Risk Graphics. *Cancer*, 113(12):3382–3390, 2008.