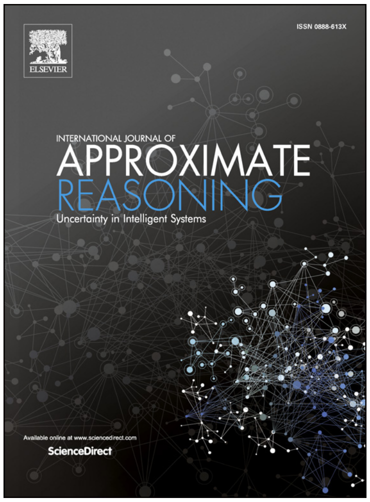
Accepted Manuscript



New distributions for modeling subjective lower and upper probabilities

Michael Smithson, Parker Blakey

PII: S0888-613X(17)30585-6
DOI: https://doi.org/10.1016/j.ijar.2018.05.007
Reference: IJA 8219

To appear in: *International Journal of Approximate Reasoning*

Received date: 20 September 2017
Revised date: 30 April 2018
Accepted date: 29 May 2018

Please cite this article in press as: M. Smithson, P. Blakey, New distributions for modeling subjective lower and upper probabilities, *Int. J. Approx. Reason.* (2018), https://doi.org/10.1016/j.ijar.2018.05.007

Highlights

- A new approach for modeling lower-upper probabilities is presented.
- The model tests the conjugacy relation between lower and upper probabilities.
- The resulting new distributions are tractable and have desirable properties.
- Two applications show that the new distributions are a good fit to data.

New Distributions for Modeling Subjective Lower and Upper Probabilities

Michael Smithson
Parker Blakey

The Australian National University, Canberra, Australia
Michael.Smithson@anu.edu.au

Abstract

This paper presents an investigation of a relatively unstudied approach to modeling lower and upper subjective probabilities. It is based on the fact that every cumulative distribution function (CDF) with support (0,1) has a "dual" CDF that obeys the conjugacy relation between coherent lower and upper probabilities. A new 2-parameter family of "CDF-Quantile" distributions with support (0,1) is extended via a third parameter for the purpose of modeling lower-upper probabilities via this approach. The extension exploits certain properties of the CDF-Quantile family, and the fact that continuous CDFs on (0,1) random variables form an algebraic group that is closed under composition. This extension also yields methods for testing specific models of lower-upper probability assignments. Finally, the new models are applied to real data-sets, and compared with alternative approaches for their relative advantages and drawbacks.

Keywords: Probability judgment, distribution, quantile regression, generalized linear model.

1. Introduction

Techniques for modeling lower and upper pairs of subjective probabilities comprise a relatively neglected topic. Motivations for investigating this topic stem from two sources. First, it is motivated by the many applications in which interval-valued probability assignments play a role in human probability judgments, whether as input into decision making and forecasting or as risk communication (e.g., [1]). Note that this kind of data differs from the Bernoulli random variate data that is modeled by imprecise probability distributions such as the well-known imprecise Dirichlet model (IDM) [2] or Coolen and Augustin's nonparametric predictive model [3]. Here, we wish to model a process that generates pairs of samples from two random variables on (0,1), one of whose cumulative probability distributions dominates the other.

A second motivation arises from the lack of a systematic framework for testing relevant hypotheses about the relationships between lower and upper probability assignments, as generated by human judges or other sources. Perhaps the most clearly important hypothesis is the so-called "conjugacy" relationship, whereby $p_L(A) = 1 - p_U(A)$, where $p_L(A)$ and $p_U(A)$ denote the lower and upper probabilities of event A, respectively. Although [4] presents a statistical test for this hypothesis, it is not embedded in a suitable framework that would enable a model of the distributions of $p_L(A)$ and $p_U(A)$ assuming that conjugacy holds to be compared against a model of the distributions that differs from the preceding model only in relaxing this assumption. However, prospects for such a framework have been raised by recent developments for modeling random variables on the (0,1) interval, which have resulted in a new family of probability distributions with (0,1) support, described by [5] and elaborated in [6].

Accordingly, this paper's primary purposes are as follows.

1. Introduce the concepts required to construct pairs of distributions of random variables on the (0,1) interval that are connected via the conjugacy relationship;
2. Develop distribution pairs based on the distribution family from [6], elaborate their properties, and investigate their viability for statistical modeling;
3. Show that when such distribution pairs are based on the distribution family from [6] they can successfully model real data; and
4. Briefly indicate points of general theoretical and practical interest that are raised by these new distributions.

The paper also extends the conference paper on which it is based [7], in the following respects.

- The properties of the extended distributions are elaborated more extensively, including new results regarding their tail behavior.

- The statistical models utilizing these distributions are evaluated via simulations for their Type I error accuracy, estimation bias, and parameter-estimate collinearity.

- A second data-fitting example has been added, to illustrate modeling decisions specific to conjugate-pair distribution models.

We begin with a brief description of conventional methods for modeling lower-upper probabilities, followed by the introduction of a new modeling approach that can potentially test the conjugacy relationship as well as modeling the distribution of the data. Then the new family of distributions is introduced, and extended for the purpose of modeling lower-upper probabilities and testing whether they obey the conjugacy relation. Models using the new distributions are evaluated for desirable statistical properties, and applied to real data-sets. The paper concludes with a brief enumeration of theoretical and practical prospects raised by the development of the new distributions and their corresponding statistical models.

Conventional statistical approaches to modeling lower-upper probability assignments treat them as a pair of dependent random variables, although lower and upper probabilities need not be statistically related to each other. One type of method ignores which is lower and which is upper, and simply models the pairs of probabilities as deviations from their respective means or as samples from a bivariate distribution with dependency handled by estimating a correlation. A somewhat more sophisticated regression-style approach uses a binary dummy predictor that takes a value of 0 for the lower probabilities and 1 for the upper probabilities and respects the ordering by restricting the regression coefficient for the dummy variable to being non-negative by exponentiating it (e.g., [8]).

This paper introduces another approach to modeling lower-upper probabilities, in which the probability distributions modeling the lower and upper probability assignments share parameters but take two different forms satisfying the conjugacy relation between coherent lower and upper probabilities. To begin, let $p_L(A) = W(p(A), \theta)$ be a lower probability with respect to probability $p(A)$ so that $0 \leq W(p(A), \theta) \leq p(A)$, for real-valued θ. The conjugate upper probability is $p_U(A) = 1 - p_L(\sim A)$, so that $p_U(A) = 1 - W(1 - p(A), \theta)$. Probability intervals on singletons A of a finite possibility space are special cases of 2-monotone capacities [9], but of course not all 2-monotone capacities can be expressed as probability intervals of the kind dealt with here.

A specific instance of the conjugacy relationship may be identified with pairs cumulative distribution functions (CDFs) for random variables on the (0,1) interval, where the CDF behaves as W. To fix ideas, we begin with two simple examples. First, consider a CDF, $G(x, \theta)$, for $0 \leq x \leq 1$, with a location parameter, θ, so that $G(0, \theta) = 0$, $G(1, \theta) = 1$, and G is monotonically increasing in x. We let x take the role of $p(A)$, and G take the role of W so that $G(x, \theta)$ behaves as $p_L(A)$. We require that $G(x, \theta) \leq x$ for all x. Define $G_D(x, \theta) = 1 - G(1 - x, \theta)$, which also is a CDF if G is continuous from below. Then G_D is the *conjugate dual* of G, and takes the role of p_U with the corresponding restriction that $G_D(x, \theta) \geq x$ for all x.

As a specific example, consider $G(x, \theta) = x^\theta$, for $\theta > 0$. Then $G_D(x, \theta) = 1 - (1 - x)^\theta$. When $\theta < 1$ G is the upper CDF, when $\theta = 1$ we have the uniform distribution so that $G = G_D$, and when $\theta > 1$ G is the lower CDF. The "middle" CDF straddled by G and G_D is simply x. For instance, setting $\theta = 2$, for $x = .2$ $G(x, \theta) = .2^2 = .04$. The conjugate upper probability is $1 - G(1 - x, \theta)$, i.e., $G_D(x, \theta) = 1 - (1 - .2)^2 = .36$. Thus, .04 and .36 are a conjugate pair of probabilities with respect to .2.

Our second example is the beta distribution. It is easy to show that if X is distributed beta(ω, τ) then G_D is the CDF of a random variable, X_D, say, that is distributed beta(τ, ω), i.e., the probability density function (PDF) of X flipped around 1/2. The absolute difference between their means, $|(\omega - \tau)/(\omega + \tau)|$, gives a convenient index of the distance between the lower and upper distributions. Reparameterizing the beta distribution so that the parameters are the mean, $\mu = \omega/(\omega + \tau)$, and precision, $\phi = \omega + \tau$, it is clear that the mean and precision of X jointly determine the magnitude of the difference between its distribution and that of its conjugate dual X_D.

One- and two-parameter distributions of the kinds illustrated here have very limited flexibility regarding the location of G and G_D; these CDFs simply straddle x, the CDF of the uniform distribution, and so their corresponding PDFs are mirror-images of one another reflected around 1/2. Nevertheless, while these pairs of distributions may not be very useful for modeling real data, the concepts involved turn out to be fruitful when applied to the family of distributions introduced in the next section.

2. CDF-Quantile Distributions

The family of distributions presented here is elaborated in [6] and [10] implements them in the R package `cdfquantile` for generalized linear modeling. This family is a special case of the T-X family presented by Aljarrah, et al. [11], although it was independently described in [5]. The following definitions reprise the introduction to the family in [6]. Let $G(x, \mu, \sigma)$ denote a CDF for random variable X with support (0, 1), and parameters μ and σ. We define G as follows:

$$G(x, \mu, \sigma) = F[U(H^{-1}(x), \mu, \sigma)] \tag{1}$$

where F is a standard CDF with support denoted by D_1, H is an invertible standard CDF with support denoted by D_2, and $U : D_2 \to D_1$ is an appropriate transform for incorporating parameters μ and σ. By a "standard" CDF we mean a distribution function whose parameters are fixed at the values conventionally used to describe a standard distribution (e.g., a mean of 0 and standard deviation of 1 for the logit-logistic, normal, and t distributions). We limit the domains D_1 and D_2 to pairs taken from $(-\infty, \infty)$ and/or $(0, \infty)$, and the following cases of U.

For $D_1 = (-\infty, \infty)$ and $D_2 = (-\infty, \infty)$ we put

$$U(y, \mu, \sigma) = (y - \mu)/\sigma. \tag{2}$$

For $D_1 = (-\infty, \infty)$ and $D_2 = (0, \infty)$ we put

$$U(y, \mu, \sigma) = (\log(y) - \mu)/\sigma. \tag{3}$$

For $D_1 = (0, \infty)$ and $D_2 = (-\infty, \infty)$ we put

$$U(y, \mu, \sigma) = \exp(-\mu/\sigma)\exp(y/\sigma). \tag{4}$$

Finally, for $D_1 = (0, \infty)$ and $D_2 = (0, \infty)$ we put

$$U(y, \mu, \sigma) = \exp(-\mu/\sigma)\, y^{1/\sigma}. \tag{5}$$

For all four pairs of domains, μ can take any value on the real line and σ must be a positive number because of the roles these parameters play in $U : D_2 \to D_1$. If all the functions are differentiable then the PDF $g(x, \mu, \sigma)$ has an explicit expression. If F is invertible, then for every γ such that $G(x, \mu, \sigma) = \gamma$, the quantile functions corresponding to the cases described in equations (2) to (5) are as follows. For $D_1 = (-\infty, \infty)$ and $D_2 = (-\infty, \infty)$ we put

$$G^{-1}(\gamma, \mu, \sigma) = H\left[\sigma F^{-1}(\gamma) + \mu\right]. \tag{6}$$

For $D_1 = (-\infty, \infty)$ and $D_2 = (0, \infty)$ we put

$$G^{-1}(\gamma, \mu, \sigma) = H\left[\exp\left(\sigma F^{-1}(\gamma) + \mu\right)\right]. \tag{7}$$

For $D_1 = (0, \infty)$ and $D_2 = (-\infty, \infty)$ we put

$$G^{-1}(\gamma, \mu, \sigma) = H\left[\mu + \sigma \log\left(F^{-1}(\gamma)\right)\right]. \tag{8}$$

Finally, for $D_1 = (0, \infty)$ and $D_2 = (0, \infty)$ we put

$$G^{-1}(\gamma, \mu, \sigma) = H\left[\exp(\mu)\left(F^{-1}(\gamma)\right)^{\sigma}\right]. \tag{9}$$

Smithson and Shou [6] present 36 members of the *CDF-Quantile* family where $D_1 = (-\infty, \infty)$ and $D_2 = (-\infty, \infty)$ by employing six standard distributions for F and H: The logistic, Cauchy, t with df = 2, arc-sinh, Burr VII, and Burr VIII distributions. All of these have explicit PDF, CDF, and quantile functions. Smithson and Shou observe that F and H may exchange roles. The resulting pairs of distributions are "quantile-duals" of one another in the sense that one's CDF is the other's quantile, with the appropriate parameterization. This duality is due to the fact that (0, 1) is both the domain and range of these functions. Smithson and Shou denote these distributions with the nomenclature *F*-*H* (e.g., Cauchit-Logistic and Logit-Cauchy).

Smithson and Shou [6] show that the CDF-Quantile family members share the following properties:

1. The family can model a wide variety of distribution shapes, with different skew and kurtosis coverage from the beta or the Kumaraswamy.
2. (Proposition 1, from [6]) Members are self-dual in the sense that $g(x, \mu, \sigma) = g(1 - x, -\mu, \sigma)$. Note that $-\mu$ has the effect of mirror-reversing the PDF. Moreover, $G = G_D$, so the conjugate-CDF duals in this family consist of identical distributions. Readers may consult [6] for details and proof.
3. (Proposition 2) The median is solely a function of μ, so that μ is genuinely a location parameter. Readers may consult [6] for details and proof.
4. (Proposition 3) The parameter σ is a dispersion parameter, in the sense that it controls how far other quantiles are from the median. Details and proof are in [6].

5. (Proposition 4) Members of this family fall into four subfamilies distinguished by behavior at the boundaries of the [0, 1] interval, including a subfamily whose density is finite in the limits at 0 and at 1.

For illustration, consider the Cauchit-Cauchy distribution, for which F and H both are standard Cauchy CDFs. The CDF G is

$$G\left(x, \mu, \sigma\right) = \frac{1}{2} + \frac{\arctan\left(\left(\tan\left(\left(2\pi x - \pi\right)/2\right) - \mu\right)/\sigma\right)}{\pi} \tag{10}$$

and the quantile function is

$$G^{-1}\left(\gamma, \mu, \sigma\right) = \frac{1}{2} + \frac{\arctan\left(\mu - \sigma\cot(\pi\gamma)\right)}{\pi}. \tag{11}$$

In (10) $1 - x$ will yield the negative of the tangent term for x and therefore $1 - x$ combined with $-\mu$ will yield the negative of the combination of x and μ. Therefore, $G(1 - x, -\mu, \sigma) = 1 - G(x, \mu, \sigma)$ and Proposition 1 is satisfied. Likewise, from (11) we can see that when $\gamma = 1/2$ the cotangent term is 0, so the median is solely a function of μ as per Proposition 2.

Thus, the CDF-Quantile family enables a wide variety of quantile regression models for random variables on the (0, 1) interval with predictors for both location and dispersion parameters, and simple interpretations of those parameters. Smithson and Shou [6] demonstrate that members of the family can out-perform the beta and other two-parameter distributions in fitting real data. Because they have explicit CDFs and quantile functions, the CDF-Quantile family is well-suited for multivariate models using copulas, and an example of this application will be presented later in this paper. Shou and Smithson [12] fit a trivariate copula model to real data as a demonstration of how this may be done using their `cdfquantreg` package in conjunction with the R package `copula`.

3. Introducing a Third Parameter to the CDF-Quantile Family

The fact that $G = G_D$ for the entire CDF-Quantile family implies that they may be well-suited to testing the conjugate-CDF model of lower and upper probabilities via the introduction of a third parameter. Unlike two-parameter distributions such as the beta distribution, for a three-parameter distribution the third parameter can determine the difference between a CDF and its conjugate dual CDF.

There are several ways to introduce a third parameter, but we will focus on doing so through a composition operator. Marshall and Olkin [13, pp. 494-495] state that the class $\mathbf{G}$ of CDFs G whose support is (0,1) form an algebraic group. This is true of absolutely continuous CDFs. The class of absolutely continuous CDFs is closed under the composition operation $G_1 \bullet G_2 = G_1(G_2)$, and this operation also is associative. The uniform distribution is the identity. Likewise, for any G in $\mathbf{G}$, the quantile function G^{-1} also is in $\mathbf{G}$. The quantile-dual relation described in the preceding section is a special case of this type of closure.

A straightforward way to introduce a third parameter is via an invertible monotonic function applied either at the outermost or innermost level of the CDF or the quantile function. Applying an invertible $(0,1) \to (0,1)$ transformation W to the innermost level of the CDF, for instance, we have

$$G(x, \mu, \sigma, \theta) = F\left[U\left(H^{-1}(W(x, \theta)), \mu, \sigma\right)\right] \tag{12}$$

and

$$G^{-1}(\gamma, \mu, \sigma, \theta) = W^{-1}\left[H\left(U^{-1}\left(F^{-1}(\gamma), \mu, \sigma\right)\right), \theta\right] \tag{13}$$

If we additionally require that $W(0, \theta) = 0$, $W(1, \theta) = 1$ and W is monotonically increasing in x and continuous from above everywhere, then W behaves as a CDF. The conjugate dual CDF therefore is

$$G_D(x, \mu, \sigma, \theta) = F\left[U\left(H^{-1}(1 - W(1 - x, \theta)), \mu, \sigma\right)\right]. \tag{14}$$

Several kinds of CDFs for W and applications of the CDF-composition operator are available from the literature on lifetime distributions. A power (resilience) parameter or a frailty parameter can be introduced in this way, by applying the CDF-composition operator. The relevant CDF is x^{θ}, for some $\theta > 0$. Slightly less obviously, introducing a tilt parameter also involves a CDF-composition, because, for $\theta > 0$, it is a composition of the CDF $x/(x + \theta(1 - x))$ with $G(x, \mu, \sigma)$. Likewise, a hazard parameter can be introduced via composition using the CDF
$1 - \exp\left[-(-\log(1 - x))^{\theta}\right]$, for $\theta > 0$; and a Laplace transform parameter with the CDF
$\left(1 - e^{-\theta x}\right)/\left(1 - e^{-\theta}\right)$, for real θ.

In the cases where the composition is $G \bullet W$, the introduction of the third parameter yields a three-parameter CDF-Quantile family with distinct CDFs and conjugate dual CDFs (i.e., $G \neq G_D$) and possessing certain properties paralleling those derived by [6] for the two-parameter family. The following Proposition is an extension of Proposition 1 (the self-dual property) from [6].

Let $W(x,\theta)$ be described as earlier, so that it behaves as a CDF. Let

$$G\left(W\left(x,\theta\right),\mu,\sigma\right) = F\left[U\left(H^{-1}\left(W\left(x,\theta\right)\right),\mu,\sigma\right)\right].$$

Then if the CDFs F and H satisfy certain symmetry conditions (in the 4 cases detailed below),

$$1 - G\left(W\left(1-x,\theta\right),-\mu,\sigma\right) = G\left(1-W\left(1-x,\theta\right),\mu,\sigma\right). \tag{15}$$

Now define

$$G^{-1}\left(Z_1\left(\gamma,\mu,\sigma\right),\theta\right) = W^{-1}\left[H\left(U^{-1}\left(F^{-1}\left(\gamma\right),\mu,\sigma\right)\right),\theta\right],$$

and

$$G^{-1}\left(Z_2\left(\gamma,\mu,\sigma\right),\theta\right) = 1 - W^{-1}\left[1-H\left(U^{-1}\left(F^{-1}\left(\gamma\right),\mu,\sigma\right)\right),\theta\right].$$

These are the quantile functions corresponding to the conjugate dual CDFs $G\left(W\left(x,\theta\right),\mu,\sigma\right)$ and $G\left(1-W\left(1-x,\theta\right),\mu,\sigma\right)$, respectively.

Proposition 1: $G^{-1}\left(Z_1\left(\gamma,\mu,\sigma\right),\theta\right)$ and $G^{-1}\left(Z_2\left(\gamma,\mu,\sigma\right),\theta\right)$ behave as conjugate lower-upper probabilities.

Proof: The identity in equation (15) has four cases, corresponding to the four combinations of domains in the CDF-Quantile family.

Case 1: For $D_1 = (-\infty,\infty)$ and $D_2 = (-\infty,\infty)$ when $-H^{-1}(x) = H^{-1}(1-x)$ and $f(x) = f(-x)$, $1 - G\left(W\left(1-x,\theta\right),-\mu,\sigma,\theta\right) = 1 - F\left[\left(H^{-1}\left(W\left(1-x,\theta\right)\right)+\mu\right)/\sigma\right]$
$= 1 - F\left[\left(-H^{-1}\left(1-W\left(1-x,\theta\right)\right)+\mu\right)/\sigma\right] = F\left[\left(H^{-1}\left(1-W\left(1-x,\theta\right)\right)-\mu\right)/\sigma\right]$
$= G\left(1-W\left(1-x,\theta\right),\mu,\sigma,\theta\right).$

Case 2: For $D_1 = (-\infty,\infty)$ and $D_2 = (0,\infty)$ when $H^{-1}(x) = 1/H^{-1}(1-x)$ and $f(x) = f(-x)$, $1 - G\left(W^{-1}\left(1-x,\theta\right),-\mu,\sigma,\theta\right) = 1 - F\left[\left(\log\left(H^{-1}\left(W\left(1-x,\theta\right)\right)\right)+\mu\right)/\sigma\right]$
$= 1 - F\left[\left(-\log\left(H^{-1}\left(1-W\left(1-x,\theta\right)\right)\right)+\mu\right)/\sigma\right] = F\left[\left(\log\left(H^{-1}\left(1-W\left(1-x,\theta\right)\right)\right)-\mu\right)/\sigma\right]$
$= G\left(1-W\left(1-x,\theta\right),\mu,\sigma,\theta\right).$

Case 3: For $D_1 = (0,\infty)$ and $D_2 = (-\infty,\infty)$ when $H^{-1}(x) = 1/H^{-1}(1-x)$ and $F(x) = 1 - F(1/x)$, $1 - G\left(1-x,-\mu,\sigma\right) = 1 - F\left[\left(H^{-1}\left(W\left(1-x,\theta\right)\right)\exp\left(\mu\right)\right)^{1/\sigma}\right]$
$= 1 - F\left[\left(H^{-1}\left(1-W\left(1-x,\theta\right)\right)\right)^{\sigma}\left(\exp\left(\mu\right)\right)^{1/\sigma}\right] = F\left[\left(H^{-1}\left(1-W\left(1-x,\theta\right)\right)\exp\left(-\mu\right)\right)^{1/\sigma}\right]$
$= G\left(1-W\left(1-x,\theta\right),\mu,\sigma,\theta\right).$

Case 4. For $D_1 = (0, \infty)$ and $D_2 = (0, \infty)$ when $-H^{-1}(x) = H^{-1}(1-x)$ and $F(x) = 1 - F(1/x)$, $1 - G(1-x, -\mu, \sigma) = 1 - F\left[\exp\left(\left(-H^{-1}(W(1-x,\theta)) + \mu\right)/\sigma\right)\right]$
$= 1 - F\left[\exp\left(\left(-H^{-1}(1 - W(1-x,\theta)) + \mu\right)/\sigma\right)\right] = F\left[\exp\left(\left(H^{-1}(1 - W(1-x,\theta)) - \mu\right)/\sigma\right)\right]$
$= G(1 - W(1-x,\theta), \mu, \sigma, \theta)$.

The conjugacy relationship immediately follows by observing that, in the definition of the quantile functions, $H\left(U^{-1}\left(F^{-1}(\gamma), \mu, \sigma\right)\right)$ fulfills the role of x in the function W^{-1}. *End of proof.*

The conjugate dual CDFs straddle the CDF $G(x, \mu, \sigma)$ and the resultant lower and upper quantile functions straddle the quantile function $G^{-1}(\gamma, \mu, \sigma)$. That is, the location of the conjugate-dual pair is determined by μ, which makes them flexible enough to be worthy candidates for modeling real data. Propositions 2-4 in [6] also hold for these three-parameter CDF-Quantile distributions because W is monotonically increasing in x and we can write the quantile function as $W^{-1}\left[H\left(U^{-1}\left(F^{-1}(\gamma), \mu, \sigma\right)\right), \theta\right]$. Thus, the median is solely a function of μ and θ, and σ still is a dispersion parameter. Moreover, the θ parameter has an interpretation as a risk-attitude parameter, because it determines the difference between the lower and upper CDFs (and likewise the difference between the corresponding quantile functions). This parameter plays a rather different role from the imprecision parameter in the IDM. The IDM parameter is set by the modeler, whereas θ is determined by the data. A larger θ indicates a more cautious or risk-averse set of probability assignments. This three-parameter family therefore is suited to ascertaining whether samples of lower and upper probability assignments behave as though they come from populations with conjugate dual distributions.

Because these are conjugate dual CDFs, the lower and upper probability pairs modeled by them are comonotone, i.e., $G(x_1, \mu, \sigma, \theta) < G(x_2, \mu, \sigma, \theta) \Rightarrow G_D(x_1, \mu, \sigma, \theta) \leq G_D(x_2, \mu, \sigma, \theta)$, and $G_D(x_1, \mu, \sigma, \theta) < G_D(x_2, \mu, \sigma, \theta) \Rightarrow G(x_1, \mu, \sigma, \theta) \leq G(x_2, \mu, \sigma, \theta)$. Pairs of conjugate lower and upper probabilities are not generally comonotone, so this constitutes a restriction on these distributions. That said, the same restriction is shared by some of the most popular models of lower and upper probabilities, such as certain neighbourhood models (pari-mutuel, constant-odds-ratio, and neighbourhood-contaminated) and p-boxes. Moreover, in regression models where the θ parameter's value is conditioned by predictors, comonotonicity is relaxed and these distributions are then capable to at least some extent of modeling conjugate lower and upper probabilities whose pairs are not comonotone. It is beyond the scope of this paper to pursue this issue further, but we note that this is an active topic of research.

3.1. $G \bullet W$ Conjugate Duals Examples

In this subsection we briefly survey two examples of three-parameter CDF-Quantile distributions of the $G \bullet W$ type, each one corresponding to a well-known kind of parameterization borrowed from the life distributions literature. These include the power parameter (which in this case corresponds to a frailty parameter) and the tilt parameter. The Cauchit-Cauchy distribution will be used throughout this subsection for illustrative purposes (it also is employed in a data-fitting example in the next subsection).

Starting with the power parameter, $W(x,\theta) = x^\theta$ and so $1 - W(1-x,\theta) = 1-(1-x)^\theta$. Applied to the Cauchit-Cauchy distribution, we have the conjugate CDF duals. As its name suggests, both F and H are Cauchy CDFs, the power parameter (exponentiated) model simply replaces x with x^θ, and the conjugate-dual CDF pair is

$$G\left(x,\mu,\sigma,\theta\right) = \frac{1}{2} + \frac{\arctan\left(\left(\tan\left((2\pi x^\theta - \pi)/2\right) - \mu\right)/\sigma\right)}{\pi} \tag{16}$$

and

$$G_D\left(x,\mu,\sigma,\theta\right) = \frac{1}{2} + \frac{\arctan\left(\left(\tan\left((2\pi\left(1-(1-x)^\theta\right) - \pi)/2\right) - \mu\right)/\sigma\right)}{\pi} \tag{17}$$

When $\theta < 1$ then $G > G_D$, and when $\theta > 1$ then $G < G_D$.

The tilt parameter, as mentioned earlier, uses the CDF $W(x,\theta) = x/(x+\theta\,(1-x))$. Applying it to the Cauchit-Cauchy distribution yields the conjugate CDF duals

$$G\left(x,\mu,\sigma,\theta\right) = \frac{1}{2} + \frac{\arctan\left(\left(\tan\left((2\pi x/(x+\theta\,(1-x)) - \pi)/2\right) - \mu\right)/\sigma\right)}{\pi} \tag{18}$$

and

$$G_D\left(x,\mu,\sigma,\theta\right) = \frac{1}{2} + \frac{\arctan\left(\left(\tan\left((2\pi\theta x/(1+x\,(\theta-1)) - \pi)/2\right) - \mu\right)/\sigma\right)}{\pi} \tag{19}$$

This model behaves as a rescaled version of the constant-odds-ratio imprecise probability model described in [14] and elsewhere. When $\theta < 1$ then $G > G_D$, and when $\theta > 1$ then $G < G_D$. Figure 1 displays the pairs of CDFs and PDFs for the exponentiated and tilt parameter models when $\mu = 0.1, \sigma = 0.5$, and $\theta = 1.5$.

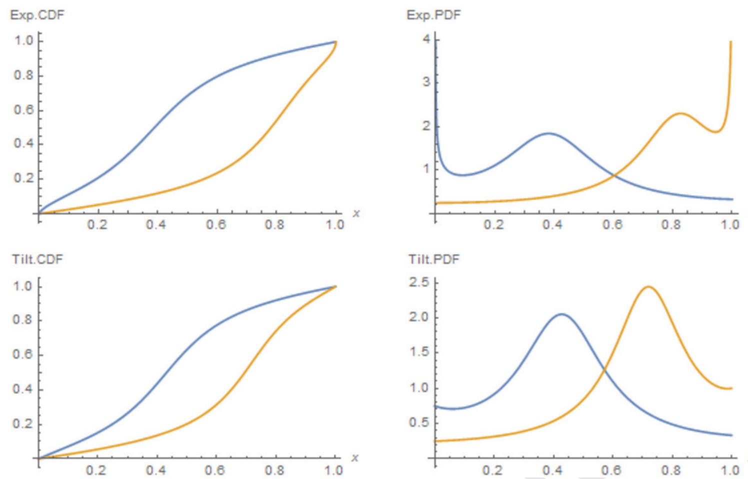


Figure 1: Power- and Tilt-Parameter Conjugate Dual Distributions

3.2. W-Functions and Distribution Subfamilies

Because any CDF whose support is (0,1) can play the role of W, a one-parameter version of any member of the CDF-Quantile family may be used in that capacity, with θ as the location parameter. These alternatives would seem to present a forbiddingly large variety of models for analysts to consider. However, it turns out that under some conditions all of them can be very similar to one another with appropriate choices of θ. For many practical modeling purposes we may restrict attention to a subset of such models, such as the power and tilt parameter models, but at this stage of research on these models the criteria and procedure for selecting among them remain an open topic of research. Two kinds of criteria can be considered here: Distribution properties and sample estimation behavior. This subsection deals with distribution properties, particularly the tail behavior of distributions in the different subfamilies of the CDF-Quantile family when extended via the conjugate-dual parameter.

Smithson and Shou [6] report that the CDF-Quantile distributions fall into four subfamilies, distinguished by their density's tail behavior at the boundaries of the unit interval:

1. $\forall\sigma < s$, $\lim_{x\to 0} g(x,\mu,\sigma) = \lim_{x\to 1} g(x,\mu,\sigma) = 0$, $\forall\sigma = s$, $\lim_{x\to 0} g(x,\mu,\sigma) = v(-\mu)$ and $\lim_{x\to 1} g(x,\mu,\sigma) = v(\mu)$, and $\forall\sigma > s$, $\lim_{x\to 0} g(x,\mu,\sigma) = \lim_{x\to 1} g(x,\mu,\sigma) = \infty$, where s is a constant and $v(z) \geq 0$;

2. $\lim_{x\to 0} g(x,\mu,\sigma) = \lim_{x\to 1} g(x,\mu,\sigma) = 0$;
3. $\lim_{x\to 0} g(x,\mu,\sigma) = \lim_{x\to 1} g(x,\mu,\sigma) = u(\sigma)$, where $u(\sigma) \geq 0$ and is finite; and
4. $\lim_{x\to 0} g(x,\mu,\sigma) = \lim_{x\to 1} g(x,\mu,\sigma) = \infty$.

For the first subfamily (the LL group, typified by the logit-logistic distribution), when $\sigma > s$ the distribution is uni-modal and when $\sigma < s$ the distribution has a bathtub shape with modes at 0 and 1. The second subfamily (the bimodal or BM group) has a limit of 0 for the density at 0 and 1, and is capable of bimodality on the interior of the unit interval. The third subfamily has nonzero finite density in the limit at 0 and 1 as a function of σ, and is known as the finite-tailed (FT) group. The fourth subfamily has infinite densities at at 0 and 1 and is known as the trimodal (TM) group.

We now examine the impact of the W functions on the tail behavior of these subfamilies. From equation (12), and applying the chain-rule, we have

$$g(x,\mu,\sigma,\theta) = \frac{\partial G[W(x,\theta),\mu,\sigma]}{\partial W(x,\theta)} \cdot \frac{\partial W(x,\theta)}{\partial x}. \tag{20}$$

Because the first term on the right-hand side of this equation is just the PDF of a two-parameter CDF-Quantile distribution with $W(x,\theta)$ substituted for x, we may write

$$\frac{\partial G[W(x,\theta),\mu,\sigma]}{\partial W(x,\theta)} = g^*\left(W(x,\theta),\mu,\sigma\right). \tag{21}$$

Now, $W(0,\theta) = 0$ and $W(1,\theta) = 1$, so for $x = 0$ and 1 respectively, we have

$$\begin{aligned} g(0,\mu,\sigma,\theta) &= g^*\left(0,\mu,\sigma\right)\left(\tfrac{\partial W(x,\theta)}{\partial x}\right)_{x=0} \\ g(1,\mu,\sigma,\theta) &= g^*\left(1,\mu,\sigma\right)\left(\tfrac{\partial W(x,\theta)}{\partial x}\right)_{x=1} \end{aligned} \tag{22}$$

so the partial derivatives of the W functions scale the values of the CDF-Quantile distributions at 0 and 1. We consider four popular candidates for W functions described earlier.

Starting with the power parameter, we have

$$\frac{\partial W(x,\theta)}{\partial x} = \theta x^{\theta-1}. \tag{23}$$

At $x = 1$ we have $g(1, \mu, \sigma, \theta) = \theta g^* (1, \mu, \sigma)$, so the right tails of the distributions are simply scaled by θ. However, as x goes to 0, for $0 \leq \theta < 1$ this expression will go to ∞ and for $\theta \geq 1$ it will go to 0. Thus, the left tails of the distributions and the right tails of their conjugate duals will not generally retain the properties entailed by membership in the four subfamilies identified by [6].

The tilt parameter, on the other hand, is better behaved:

$$\frac{\partial W(x,\theta)}{\partial x} = \frac{\partial \left(x/x\left(\theta(1-x)+x\right)\right)}{\partial x} = \frac{\theta}{\left(\theta(1-x)+x\right)^2}. \tag{24}$$

The right-hand expression goes to $1/\theta$ as x goes to 0 and to θ as x goes to 1. So, the tail behavior properties of the CDF-Quantile subfamilies are preserved in a rescaled form.

The hazard parameter yields

$$\frac{\partial W(x,\theta)}{\partial x} = \frac{\partial \left(1-\exp\left(-(-\log(1-x))^{\theta}\right)\right)}{\partial x} = \frac{\theta e^{-(-\log(1-x))^{\theta}}(-\log(1-x))^{\theta-1}}{1-x}. \tag{25}$$

The right-hand expression goes to 0 as x goes to 0 or to 1 when $\theta > 1$ and to ∞ when $\theta < 1$. As with the power parameter, the tail-behavior properties of the CDF-Quantile subfamilies therefore are not preserved.

Finally, for the Laplace parameter we have

$$\frac{\partial W(x,\theta)}{\partial x} = \frac{\partial \left(\left(1-\exp(-\theta x)\right)/\left(1-\exp(-\theta)\right)\right)}{\partial x} = \frac{\theta e^{\theta-\theta x}}{e^{\theta}-1}. \tag{26}$$

The right hand expression goes to $\theta e^{\theta}\big/\left(e^{\theta}-1\right)$ as as x goes to 0 and to $\theta\big/\left(e^{\theta}-1\right)$ as x goes to 1, and thus the tail-behavior properties of the CDF-Quantile subfamilies are simply rescaled.

If we wish the tail-behavior properties of the CDF-Quantile subfamilies to be preserved when adding a third parameter, then the tilt and Laplace parameters are desirable because they do so. The power parameter does this only for the right tail (as x goes to 1) of the distributions and the left tail of their conjugate dual counterparts, and the hazard parameter does not preserve the properties for either tail. That said, what impact does the non-preservation of these properties have? The θ and σ parameters both influence dispersion, so each can offset the other's effect. Thus, the three-parameter conjugate-dual LL and TM subfamilies' tail behaviors are qualitatively the same as in their respective two-parameter versions, but now are influenced by both θ and σ. The conjugate-dual BM subfamily exhibits conditional bimodality which is violated for sufficiently high θ relative to σ, and likewise the conjugate-dual FT subfamily has finite densities at 0 and 1 conditional on θ sufficiently low relative to σ.

4. Sample Parameter Estimate Collinearity, Bias, and Type I Error Rate Accuracy

Among the most important criteria for selecting specific distributions for modeling data are parameter estimation bias, Type I error-rate accuracy, and parameter estimate collinearity. This section presents the results of simulations in which every member of the conjugate-dual CDF-Quantile family was fitted to samples drawn from its own distribution population. Each simulation had 2500 runs. To assess the effect of sample size, samples of size 10, 25, 50, 100, and 250 were investigated for every distribution. Two types of W functions were investigated: The tilt parameter and the Laplace parameter. The null-model set consisted of $\mu = 0$, $\sigma = 1$, and $\theta = 1$ for the tilt parameter and $\theta = 10^{-7}$ for the Laplace parameter, i.e., values that were found to render the lower and upper conjugate dual distributions identical. The non-null-model set comprised $\mu = 0.4$, $\sigma = 0.5$, and $\theta = 1.5$ for the tilt parameter, and $\theta = -0.85$ for the Laplace parameter, in order to give similar distribution shapes for the two parameters. Thus, there were a total of four models (two W functions by two parameter sets) and five sample sizes applied to 36 distributions.

4.1. Bias and Type I Error Rate Accuracy

We begin with a brief summary of the findings regarding estimation bias for the three parameters (details are in the Supplementary Materials: `https://drive.google.com/open?id=0B2mZom-c8j2_THRJQTZYbnBCVUE`). No systematic bias in μ or θ was observed across all four models, distributions, or sample sizes. However, σ estimates showed a consistent (negative) bias. Bias was more strongly negative for null models than non-null models. The bias decreased in magnitude with increasing sample size, but remained below -.06 for all sample sizes.

Tendencies in estimation bias had consequences for Type I error-rate accuracy, with an absence of bias generally ensuring that the observed Type I error rates were close to the target .05 rate. Across all four models, μ

Type I error rates tended to stabilize for sample sizes close to 50. These error rates were similar for all distributions, with a mean at $n = 50$ of $.0637 \pm .0065$. Likewise, θ error rates at $n = 50$ averaged $.0627 \pm .0129$ for 34 of the 36 distributions, with exceptions being the Cauchit-BurrVII distribution ($.1055 \pm .0507$) and the ArcSinh-BurrVII distribution ($.1507 \pm .0799$). The latter was due to a slight positive estimation bias for θ. Type I error rates in σ increased with sample size, due to the negative estimation bias and the fact that it did not decrease sufficiently rapidly with increased sample size. For example, the average error rate at $n = 50$ was $.1108 \pm .0327$, and increased to $.1547 \pm .0620$ at $n = 100$.

4.2. Parameter Estimate Collinearity

Parameter estimate collinearity is a relatively unstudied problem for distributions with doubly-bounded support. Its potential importance stems from the fact that location, dispersion, skew, and other aspects of distribution shape are not independent of one another. For instance, as location approaches either boundary of the support interval, skew must increase if dispersion does not decrease, and vice-versa. Our preliminary investigations into adding a third parameter to the CDF-Quantile family and other two-parameter distributions for random variables on the unit interval have frequently encountered high collinearity among parameter estimates.

However, parameter estimate correlations in conjugate-dual distribution models averaged close to 0 for all pairs of parameters, across all models, distributions, and sample sizes. Variability in these correlations decreased with increasing sample size, with variability tending to be somewhat greater for distributions whose F distribution components were Cauchit, arc-sinh, and T2. As mentioned earlier, more details regarding these findings are available in the Supplementary Materials.

5. Examples and Applications

5.1. Interpretations of Verbal Probability Phrases

We now present an example of model-fitting that compares the conjugate lower-upper distributions with appropriate alternatives for modeling lower-upper probability assignments. The fourth Intergovernmental Panel on Climate Change (IPCC) report utilizes verbal phrases such as “likely” and “unlikely” to describe the uncertainties in climate science. Budescu et al. [15] conducted an experimental study of lay interpretations of these phrases, using 13 sentences from the IPCC report, in which they asked 223 participants to provide lower, "best", and upper numerical estimates of the probabilities to which they believed each sentence referred. For example, participants were presented with the sentence "The Greenland ice sheet and other Arctic ice fields likely contributed no more than 4 m of the observed sea level rise.", and asked to consider the probability they thought the report authors may have had in mind for the term "likely" in this sentence. Participants were required to provide their lowest, highest, and their best numerical estimates of this probability. Budescu et al. found that participants' "best" estimates were more regressive (toward the middle of the [0, 1] interval) than the IPCC stipulations, but they did not report systematic analyses of the lower and upper estimates.

We present 11 models fitted to the lower and upper probability estimates in the Budescu et al. data. The first three models are based on the two-parameter CDF-Quantile distribution. Model 1 is just the two-parameter distribution, as defined in equation (2), with intercept-only submodels $\hat{\mu} = \beta_0$ and $\hat{\sigma} = \exp(\delta_0)$. Model 2 has conditional parameter estimates, with submodels $\hat{\mu} = \beta_0 + \beta_1 x$ and $\hat{\sigma} = \exp(\delta_0 + \delta_1 x)$, where $x = 0$ for lower probabilities and $x = 1$ for upper probabilities. Model 3, in addition to the submodels from Model 2, also estimates the dependency between the lower and upper estimates via a t-copula with CDF-Quantile margins. This model therefore also includes estimates of the t-copula dependency parameter, ρ, and degrees of freedom parameter, ϕ.

Models 4-7 are based on the 3-parameter power (exponentiated) CDF-Quantile distribution, as in the CDF defined in equation (12) with $W(x, \theta) = x^{\theta}$. Model 4 has intercept-only submodels $\hat{\mu} = \beta_0$, $\hat{\sigma} = \exp(\delta_0)$, and $\hat{\theta} = \exp(\gamma_0)$. Model 5 is the conjugate-dual model, as defined in equations (12) and (14). This has the same intercept-only submodels as Model 4 but is a two-component distribution mixture model with a fixed mixture parameter, so that the first CDF, G, is weighted 1 and the second, G_D, is weighted 0 for the upper probabilities and the reverse weighting is applied to the lower probabilities. Technically, it is a four-parameter model although the mixture parameter is not being estimated. Model 6 has conditional parameter estimates, $\hat{\mu} = \beta_0 + \beta_1 x$ and $\hat{\sigma} = \exp(\delta_0 + \delta_1 x)$ with $x = 0$ and 1 for lower and upper probabilities, but an intercept-only submodel $\hat{\theta} = \exp(\gamma_0)$. Model 7 has the conditional μ and σ submodels in Model 6 plus $\hat{\theta} = \exp(\gamma_0 + \gamma_1 x)$. Finally, models 8-11 are based on the tilt-parameter CDF-Quantile distribution, as in the CDF defined in equation

(12) with $W(x, \theta) = x/(x + \theta(1 - x))$. These models have the same variants as Models 4-7.

The best-fitting models from the CDF-Quantile family are from the FT subfamily, whose members have defined, finite densities at 0 and 1 as shown in [6]. The best-fitting distribution from this subfamily is the Cauchit-Cauchy, so the models considered here are mainly limited to that distribution. Table 1 displays goodness-of-fit statistics for the 11 models. The top section of the table presents these results for the three models using the two-parameter Cauchit-Cauchy. The middle section contains the power-parameter (exponentiated) models, and the lower section contains the tilted-parameter models. The "Params" column displays the number of parameters in each model, the "2LL" column shows twice the log-likelihood of the fitted models, and the "AIC" column is the Akaike Information Criterion, $AIC = -2LL + 2p$, where p is the number of parameters in the Params column.

Remarkably, the 4-parameter conjugate-dual models fit the data better than most of the 5- and 6-parameter conditional models and better than the 6-parameter copula model. The conjugate-dual power-parameter model is superior to the conjugate-dual tilted-parameter model, and is out-performed only by the 6-parameter conditional tilted-parameter model. Likewise, the conjugate-dual tilted-parameter model is out-performed only by the 5- and 6-parameter conditional tilted-parameter models and the 6-parameter conditional power-parameter model.

These results are not due to some kind of fluke in the Cauchit-Cauchy distribution. Other members of the FT subfamily have similar fits for their conjugate-dual models. For instance, the T2-T2 and the Cauchit-ArcSinh conjugate-dual power-parameter models have AIC's of -2159 and -2062, respectively, and both of these out-perform their respective 5- and 6-parameter conditional power-parameter counterparts.

Figure 2 shows the fitted distributions from the conjugate-dual model (top half of the figure) and the 6-parameter conditional exponentiated model. The two pairs of fitted distributions are strikingly similar and the conjugate-dual AIC is the better of the two. The facts that the 4-parameter conjugate-dual model fits the data better than a regression model with 6 parameters and that the fitted distribution shapes are reasonably similar to the empirical distributions lend plausibility to the seemingly unlikely conjecture that human lower-upper probability judgments are distributed approximately as conjugate-dual distributions.

The exponentiated Cauchit-Cauchy 4-parameter conjugate-dual and 6-

Table 1: Cauchit-Cauchy Models and Fits

Model	Description	Params.	2LL	AIC
1	2-parameter	2	595	-591
2	2-parameter condit. μ, σ	4	1378	-1370
3	2-parameter condit. t-copula	6	1584	-1572
4	exponentiated 3-param.	3	616	-609
5	conjugate-dual exponentiated	4	2378	-2372
6	exponentiated condit. μ, σ	5	1392	-1382
7	exponentiated condit. μ, σ, θ	6	1967	-1955
8	tilted 3-param.	3	880	-874
9	conjugate-dual tilted	4	1736	-1730
10	tilted condit. μ, σ	5	2152	-2142
11	tilted condit. μ, σ, θ	6	3118	-3106

Table 2: Quantiles and Exponentiated Model Quantile Estimates

Model	Estimate	.1	.25	.5	.75	.9
	empirical lower	0.092	0.301	0.570	0.699	0.779
5	conjugate-dual lower	0.059	0.303	0.535	0.688	0.825
7	conditional lower	0.091	0.378	0.584	0.713	0.834
	empirical upper	0.540	0.729	0.858	0.948	0.998
5	conjugate-dual upper	0.298	0.684	0.863	0.935	0.977
7	conditional upper	0.495	0.672	0.846	0.935	0.975

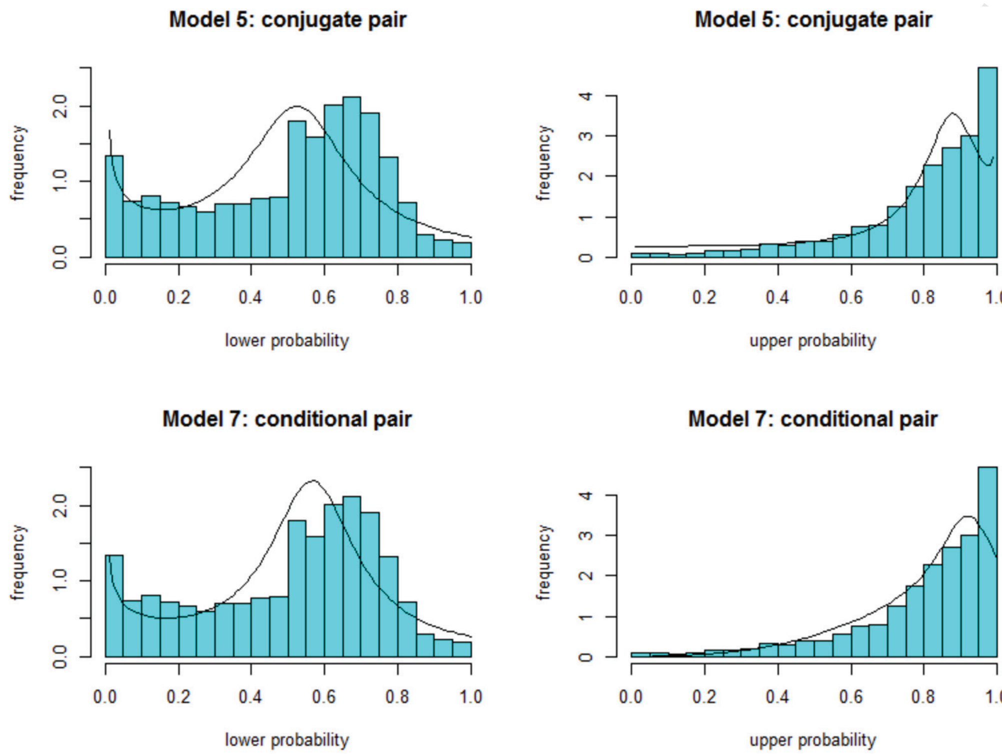


Figure 2: IPCC Data and Fitted Distributions

parameter conditional regression models may be compared further via the 5-number summaries in Table 2. When compared with their empirical counterparts (rows 1 and 4 in the table), the conditional model is more accurate than the conjugate-dual model at the 10^{th} quantile, but the reverse is the case for most of the other quantiles. Both models appear to be fairly accurate in the middle 50% of the distributions. Again, this is an intriguing outcome for the conjugate-dual model, given that only three of its four parameters are being estimated from the data.

5.2. Partiton Priming Effects Study

Our second example probes the problem of how to choose between two possible pairings of the conjugate-dual distributions. In principle, either G or G_D may take the role of the upper distribution; swapping them simply reverses the sign of the θ parameter. These alternative models may be fitted by reversing the mixing dummy-variable that specifies which parts of the data are to be treated as lower and which as upper distribution data (i.e., z vs $1-z$). However, the magnitudes of the alternative θs may not be identical, and the maximum-likelihood estimates for μ and σ alternatives may differ, along with the goodness of fit. These considerations raise the question of when a case can be put that one G-G_D pairing should be preferred to its alternative. This example illustrates this type of decision.

The "Weather task" in the study reported by [16] asked participants to judge how likely Sunday is to be the hottest day of next week. This replicates a task used by [17]. Using the terminology adopted by Fox and Rottenstreich, participants were randomly assigned to a Case Prime condition (see below) which invokes a two-fold partition by focusing on whether Sunday will or will not be hottest, or a Class Prime condition (see below) which invokes a seven-fold partition by focusing on the hottest day of the week. This task has a "corrrect" partition, namely the seven-fold. The main hypothesis was that the tendency to provide probability intervals that included 1/2 would be more pronounced in the Case than in the Class condition. The primes are as follows:

Case Prime

- "[What is the probability that] the temperature at Canberra airport on Sunday will be higher than every other day next week?"
- "[What is the probability that] the temperature at Canberra airport on Sunday will not be higher than every other day next week?"

Class Prime

- "[What is the probability that] the highest temperature of the week at Canberra airport will occur on Sunday?"
- "[What is the probability that] the highest temperature of the week at Canberra airport will not occur on Sunday?"

Participants were instructed to provide lower and upper probability estimates:

> For next few questions, please assign a lower and an upper probability estimate for each event described in the spaces provided. For example, if you think the probability of an event could range from 20 percent to 40 percent then record those figures in the space provided.

Thus, participants were asked to provide lower and upper estimates of the probability that Sunday would be the hottest day (the "yes" judgments), and of the probability that it would not be the hottest (the "no" judgments). The data from 171 participants are included in the analyses presented here. To simplify interpretations of the conjugate-dual models, the "no" probabilities have been subtracted from 1 (i.e., reversed), thereby enabling them to be compared directly with the "yes" probabilities.

We initially fit four conjugate-dual distributions to the Weather data, one from each of the four subfamilies (logit-logistic from the LL subfamily, T2-T2 from the FT subfamily, logit-T2 from the BM subfamily, and T2-logistic from the TM subfamily). The log-likelihoods for these models are shown in Table 3. These clearly suggest that the T2-T2 and T2-logistic distributions have the best, and nearly identical, fits. However, we selected the T2-T2 model as the more plausible of the two, because there is no evidence of anti-modes at 0 or 1 in the data (as predicted by the T2-logistic model).

Log-likelihoods also were compared to determine which of alternative conjugate-dual distribution pair should be used as the upper and which as the lower distribution. The best-fitting alternatives are those whose results are shown in Table 5. For the logit-logistic, T2-T2, and T2-logistic models, the log-likelihoods for these alternatives did not differ greatly (0.79 to 3.78). However, the logit-T2 the model displayed in Table 5 has a log-likelihood of 123.94 whereas its alternative conjugate pair's log-likelihood is only 38.50.

Table 3: Log-Likelihoods for Four Models

Distribution	log-lik.
logit-logistic	168.9132
T2-T2	188.9316
T2-logistic	188.7894
logit-T2	123.9383

The T2-T2 model was then extended to include effects from valence (yes vs no: Whether participants were asked to estimate the probability that Sunday would or would not be the hottest day) and partition (two-fold vs seven-fold). The final model coefficients and log-likelihood are shown in Table 4. In the left-hand column, "partition" refers to the seven-fold partition condition, and "valence" refers to the "no" judgments. The intercept coefficients therefore index the two-fold partition and the "yes" judgments. This model has main effects for valence in all three parameters' submodels, main effects for partition in the μ and σ submodels, and a valence-by-partition interaction term in σ. Adding other terms did not significantly improve model fit, whereas subtracting any of the current terms significantly decreased model fit. The model broadly confirms the findings reported by [16], yielding a significantly lower median and quantile spread for the seven-fold partition condition, with the partition effect on quantile-spread enhanced in the "no" judgments.

Table 4: Final T2-T2 Model

Effect	Coeff.	Estimate	S.Error	z	p
μ intercept	b_0	-0.6098	0.0341		
μ valence	b_1	-0.2625	0.0354	-7.4079	$< .0001$
μ partition	b_2	-0.1666	0.0329	-5.0611	$< .0001$
σ intercept	d_0	-0.4747	0.0421		
σ valence	d_1	-0.1707	0.0422	-4.0430	$< .0001$
σ partition	d_2	-0.0818	0.0419	-1.9545	0.0253
σ val*partit	d_3	-0.1326	0.0418	-3.1745	$< .0001$
θ intercept	g_0	-0.3560	0.0266		
θ valence	g_1	0.0640	0.0262	2.4428	0.0073
log-lik.	252.6550				

This model is fairly complex, so by way of assessing it, Table 5 compares the model predictions of the proportion of participants providing lower and/or upper probabilities greater than 0.5 (denoted here by $P(x > 0.5)$) with the corresponding proportions in the data. The 0.5 benchmark corresponds with the main hypothesis that invoking a two-fold partition (either Sunday will or will not be the hottest day of the week), participants would anchor their judgments around a 0.5 probability. The model captures the expected partition effects in both the "yes" and "no" data, with $P(x > 0.5)$ clearly higher for the two-fold than for the seven-fold partition conditions, and this difference being larger in the "no" data. The proportions also are higher for the "no" probabilities (in the lower half of the table), and the model captures this reasonably well, despite over-estimating the lower-yes-two-fold probability. The model also echoes the clear differences in $P(x > 0.5)$ between the lower and upper distributions. Except for the lower-yes-two-fold probability, the magnitudes of the residuals are less than 0.1, suggesting a reasonably good fit by the model overall.

Table 5: Model vs Empirical $P(x > 0.5)$

distribution	valence	partition	model $P(x > 0.5)$	empirical $P(x > 0.5)$	residual
lower	yes	two-fold	0.1349	0.0112	0.1237
upper	yes	two-fold	0.2884	0.3034	-0.0150
lower	yes	seven-fold	0.0449	0.0000	0.0449
upper	yes	seven-fold	0.1067	0.1829	-0.0762
lower	no	two-fold	0.2493	0.1798	0.0695
upper	no	two-fold	0.6051	0.5169	0.0882
lower	no	seven-fold	0.1814	0.1220	0.0594
upper	no	seven-fold	0.4457	0.4269	0.0188

6. Conclusions and Future Directions

A new family of probability distributions, the CDF-Quantile family, shows promise in modeling probability judgments. The two-parameter version of the family has been sufficiently well-explored by [6] to have been made available for generalized linear modeling via the `cdfquantreg` package in R and a SAS macro, as presented by Shou and Smithson ([10], [12]), and those authors also have demonstrated that these distributions can model probabilities better than other two-parameter distributions such as the beta. This paper has presented an investigation of the application of the CDF-Quantile family to modeling pairs of lower and upper probability assignments or estimates, by extending it to incorporate a third parameter.

Because absolutely continuous CDFs whose support is the (0,1) interval are closed under composition, and due to the properties of the CDF-Quantile distributions, three-parameter extensions via the composition of CDF functions yield conjugate dual pairs of CDFs. This result may hold some theoretical interest. A future line of research may elaborate the connections between these conjugate duals and imprecise probability frameworks. There is a natural link with probability boxes (p-boxes, as coined by [18]), given that the conjugate-dual CDFs form a p-box. This can be seen from the fact that one CDF bounds the other one from below for every $x \in (0, 1)$, as implied by the restriction that $W(x, \theta) < (>)x$ and $1 - W(1 - x, \theta) > (<)x$ and inspection of equations (12) and (14). Conjugate duals are noteworthy cases of p-boxes because the "width" of the gap between them is determined in a different way from the data-driven methods to which Ferson et al. [18] refer. To our awareness, p-boxes have not been systematically studied regarding methods of fitting them to lower-upper probability data.

Some conjugate-dual models, in turn, have been found to fit two data-sets reasonably well, raising the possibility that human lower-upper probability assignments may approximate a conjugacy relationship in their CDFs. Further research will determine whether these findings generalize to other such data-sets, if elicitation methods influence the results, and what cognitive mechanisms or heuristics account for the phenomenon. However, perhaps the first priority is to ascertain the connections between the θ parameter, measurement error, and sampling error.

Finally, the three-parameter CDF-Quantile distributions also beg for further investigation. The overview in this paper only skims their characteristics, and little is known about the advantages and drawbacks of alternative parameterization methods for θ (e.g., power versus tilt parameters). Preliminary investigations suggest that the high correlations between parameter estimates may be a pervasive problem for three-parameter distributions on the unit interval (including three-parameter generalizations of the beta distribution). Moreover, as [6] observes, model diagnostics and related aspects of model evaluation for the 2-parameter CDF-Quantile family have yet to be completely thought through, and the same holds for their 3-parameter extensions. Thus, the questions of effective estimation procedures and diagnostics for models using these distributions are active topics of research. The primary goals regarding this paper have been to introduce this extension of the CDF-Quantile family and to make a case that it holds some promise for modeling distributions of lower-upper probability assignments while also testing the conjugacy relationship. However, this paper also may be regarded as a preliminary exploration of three-parameter CDF-Quantile distributions, with the unexpected finding that conjugate-dual distributions may be useful for modeling lower-upper probability assignments made by human judges.

[1] D. V. Budescu, H.-H. Por, S. B. Broomell, M. Smithson, The interpretation of ipcc probabilistic statements around the world, Nature Climate Change 4 (6) (2014) 508–512.

[2] P. Walley, Inferences from multinomial data: learning about a bag of

marbles, Journal of the Royal Statistical Society. Series B (Methodological) ("1996") 3–57.

[3] F. P. Coolen, T. Augustin, Learning from multinomial data: a non-parametric predictive alternative to the imprecise dirichlet model., in: ISIPTA, Vol. 5, "2005", pp. 125–134.

[4] M. Smithson, Elicitation, in: T. Augustin, F. P. Coolen, G. de Cooman, M. C. Troffaes (Eds.), Introduction to Imprecise Probabilities, Wiley Online Library, 2014, pp. 318–328.

[5] M. Smithson, E. C. Merkle, Generalized Linear Models for Categorical and Continuous Limited Dependent Variables, Chapman and Hall/CRC Press, 2014.

[6] M. Smithson, Y. Shou, Cdf-quantile distributions for modeling random variables on the unit interval, British Journal of Mathematical and Statistical Psychology (2017) doi: 10.1111/bmsp.12091.

[7] M. Smithson, New distributions for modeling subjective lower and upper probabilities, in: Proceedings of the Tenth International Symposium on Imprecise Probability: Theories and Applications, 2017, pp. 301–312.

[8] M. Smithson, D. V. Budescu, S. B. Broomell, H.-H. Por, Never say not: Impact of negative wording in probability phrases on imprecise probability judgments, International Journal of Approximate Reasoning 53 (8) (2012) 1262–1270.

[9] S. Destercke, D. Dubois, Special cases, in: T. Augustin, F. P. Coolen, G. de Cooman, M. C. Troffaes (Eds.), Introduction to Imprecise Probabilities, Wiley Online Library, 2014, pp. 318–328.

[10] Y. Shou, M. Smithson, cdfquantreg: Quantile Regression for Random Variables on the Unit Interval, R package version 1.1.0 (2016).

[11] M. A. Aljarrah, C. Lee, F. Famoye, On generating t-x family of distributions using quantile functions, Journal of Statistical Distributions and Applications 1 (1) (2014) 2.

[12] Y. Shou, M. Smithson, cdfquantreg: An R package for cdf-quantile regression, Journal of Statistical Software, (forthcoming).

[13] A. W. Marshall, I. Olkin, Life Distributions, Springer, 2007.

[14] P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman and Hall, London, 1991.

[15] D. V. Budescu, S. Broomell, H.-H. Por, Improving communication of uncertainty in the reports of the intergovernmental panel on climate change, Psychological Science 20 (3) (2009) 299–308.

[16] M. Smithson, C. Segale, Partition priming in judgments of imprecise probabilities, Journal of Statistical Theory and Practice 3 (2009) 169–182.

[17] C. R. Fox, Y. Rottenstreich, Partition priming in judgment under uncertainty, Psychological Science 14 (2003) 195–200.

[18] S. Ferson, V. Kreinovich, L. Ginzburg, D. S. Myers, K. Sentz, Constructing probability boxes and dempster-shafer structures, Sandia National Laboratories (2003) 143–180.