British Journal of Mathematical and Statistical Psychology (2017)

The British Psychological Society

www.wileyonlinelibrary.com

CDF-quantile distributions for modelling random variables on the unit interval

Michael Smithson* and Yiyun Shou

The Australian National University, Canberra, Australian Capital Territory, Australia

This paper introduces a two-parameter family of distributions for modelling random variables on the (0,1) interval by applying the cumulative distribution function of one 'parent' distribution to the quantile function of another. Family members have explicit probability density functions, cumulative distribution functions and quantiles in a location parameter and a dispersion parameter. They capture a wide variety of shapes that the beta and Kumaraswamy distributions cannot. They are amenable to likelihood inference, and enable a wide variety of quantile regression models, with predictors for both the location and dispersion parameters. We demonstrate their applicability to psychological research problems and their utility in modelling real data.

1. Introduction

To date, the most commonly employed two-parameter distributions for modelling random variables on the (0,1) interval are the beta (e.g., Ferrari & Cribari-Neto, 2004; Smithson & Verkuilen, 2006) and, less commonly, the Kumaraswamy (1980), lambda, and logit-logistic distributions. Other two-parameter distributions have been proposed, including the simplex and triangular distributions (e.g., Barndorff-Nielsen & Jørgensen, 1991; Kotz & Van Dorp, 2004), although these have yet to be widely adopted in practical modelling. Doubly bounded random variables occur throughout psychology and cognate areas such as economics and biology. The most commonplace examples in psychology include proportions and percentages, such as probability judgements, the proportion of the brain's volume occupied by a specific part of the brain, and the proportion of a period of time spent on an activity. Examples from economics include rates, such as fractional repayments on debts, market shares, and capital structure. Many psychological scales are doubly bounded, and in some applications it is sensible to treat the bounds as true scores (rather than as censored scores). For instance, Zou, Carlsson and Quinn (2010) apply beta regression to Likert scale data, and Moberg, Alderling and Meding (2009) apply beta regression to a quality of life index. Likewise, Noël and Dauvier (2007) provide item-response models for doubly bounded continuous scale items using the beta distribution. Verkuilen and Smithson (2012) extended beta regression to mixed (multi-level) models for dealing with dependencies in the data. The handbook of the beta distribution (Gupta & Nadarajah, 2004) includes examples of applications in other disciplines. It would seem that general linear models for proportions and rates have reached a mature stage of development.

*Correspondence should be addressed to Michael Smithson, Research School of Psychology, The Australian National University, Canberra, ACT 2601, Australia (email: Michael.Smithson@anu.edu.au).

DOI:10.1111/bmsp.12091

Nevertheless, in comparison to the rich variety of distributions for modelling continuous variables with a single lower bound, there are relatively few alternatives for modelling doubly bounded random variables. In particular, the Kumaraswamy and logit-logistic distributions are typically the only alternatives available for explicit models of quantiles for such variables. In this paper we introduce a family of two-parameter distributions with support (0,1) that are especially useful for modelling quantiles, and that also sometimes outperform their better-known counterparts. We have simultaneously launched a package, cdfquantreg, in R (Shou & Smithson, 2016) providing the software tools for modelling both parameters as general linear models.

The distribution family presented here has links with earlier families of distributions. Tadikamalla and Johnson (1982) replace the standard normal distribution in Johnson's (1949) SB distribution with the standard logistic distribution, thus producing the logit-logistic distribution. As discussed by Johnson, Kotz and Balakrishnan (1995), the Johnson SB system includes many well-known distributions, but until recently relatively little attention has been paid to distributions whose support is (0,1). A natural extension of this approach is to employ other transformations from (0,1) to either the real line or non-negative half of the real line, and expand the variety of distributions as well.

A related approach is to generate the cumulative distribution function (cdf) of a distribution via another cdf as the upper limit of an integral over the probability density function (pdf) of a random variable whose support is (0,1). Eugene, Lee and Famoye (2002) produced such a family using the beta distribution as a generator. The cdf of a random variable X is defined as

$$G(x) = \int_0^{S(x)} r(t)\mathrm{d}t, \tag{1.1}$$

where $r(t)$ is the pdf of the beta distribution and $S(x)$ is the cdf of any random variable. Cordeiro and de Castro (2011) and Jones (2009) expanded this family by substituting the Kumaraswamy (1980) distribution as $r(t)$. Alzaatreh, Lee and Famoye (2013) made two further extensions that considerably broadened the scope of this family. Their so-called T-X family is defined as follows:

$$G(x) = \int_a^{W(S(x))} r(t)\mathrm{d}t, \tag{1.2}$$

where $r(t)$ is now the pdf of a random variable, $T \in [a,b]$, for $-\infty \leq a < b \leq \infty$. $W(S(x))$ in equation (1.2) satisfies three properties:

1. $W(S(x)) \in [a,b]$.
2. $W(S(x))$ is differentiable and monotonically non-decreasing.
3. $W(S(x)) \to a$ as $x \to -\infty$ and $W(S(x)) \to b$ as $x \to \infty$.

Finally, the cdf in equation (1.2) can be written in terms of the cdf of T:

$$G(x) = R[W(S(x))]. \tag{1.3}$$

Aljarrah, Lee and Famoye (2014) mention that Alzaatreh *et al.* (2013) derived the parameters of their family entirely through the parameters of R and S, without also exploring the properties of W, although they presented several examples of W. Aljarrah et al. extended the T-X family by proposing that W be the quantile function of a third random variable, Y, say, whose support is the same as T.

Given these recent developments, now is a propitious time to exploit them to develop tractable two-parameter candidates that can augment the toolbox of distributions whose support is (0,1). As it happens, a recent development enables us to do so directly. Independently of Aljarrah *et al.* (2014), Smithson and Merkle (2014, p. 158) describe a distribution family that is a special case of Aljarrah et al.'s family but also related to the Johnson SB family.

The rest of this paper investigates a general version of the family described by Smithson and Merkle. The simplest characterization of this family of distributions is in terms of its cdf. Let $G(x,\mu,\sigma)$ denote a cdf for random variable X with support (0,1), a real-valued location parameter μ and positive scale parameter σ. We define G as follows:

$$G(x, \mu, \sigma) = F[U(H^{-1}(x), \mu, \sigma)], \tag{1.4}$$

where F is a cdf with support which we shall denote by D_1, H is an invertible cdf with support denoted by D_2, and $U : D_2 \to D_1$ is an appropriate transform for imposing the location and scale parameters, μ and σ. We limit the domains D_1 and D_2 to pairs taken from $(-\infty,\infty)$ and/or $(0,\infty)$, and consider special cases of U.
For $D_1 = (-\infty, \infty)$ and $D_2 = (-\infty, \infty)$, we put

$$U(y, \mu, \sigma) = (y - \mu)/\sigma. \tag{1.5}$$

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

$$U(y, \mu, \sigma) = (\log(y) - \mu)/\sigma. \tag{1.6}$$

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

$$U(y, \mu, \sigma) = \exp(-\mu/\sigma)\exp(y/\sigma). \tag{1.7}$$

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

$$U(y, \mu, \sigma) = \exp(-\mu/\sigma)y^{1/\sigma}. \tag{1.8}$$

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 (1.5)–(1.8) are as follows. For $D_1 = (-\infty, \infty)$ and $D_2 = (-\infty, \infty)$ we put

$$G^{-1}(\gamma, \mu, \sigma) = H\left[\sigma F^{-1}(\gamma) + \mu\right]. \tag{1.9}$$

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

$$G^{-1}(\gamma, \mu, \sigma) = H\left[\exp\left(\sigma F^{-1}(\gamma) + \mu\right)\right]. \tag{1.10}$$

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

$$G^{-1}(\gamma, \mu, \sigma) = H\left[\mu + \sigma \log\left(F^{-1}(\gamma)\right)\right]. \tag{1.11}$$

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

$$G^{-1}(\gamma, \mu, \sigma) = H\left[\exp(\mu)\left(F^{-1}(\gamma)\right)^{\sigma}\right]. \tag{1.12}$$

The distributions in equation (1.4) are related to the T-X family in equation (1.3) by setting $F = R$ and $U[H^{-1}(S(x), \mu, \sigma)] = W(S(x))$, with H differentiable and $x \in (0,1)$. This reduces to equation (1.4) by restricting S to be the uniform cdf, so that $S(x) = x$. In a third, also independent line of work, Lemonte and Bazán (2016) present a family of distributions with support (0,1) as an extension of the Johnson SB family. It is easy to show that their family is a special case of ours with H restricted to the logistic cdf (details available from the first author). Lemonte and Bazán do not cite Alzaatreh *et al.* (2013) or any other related papers, so they appear to have developed their family independently of that group of researchers and of Smithson and Merkle (2014).

If we further restrict F and H to be cdfs of location–scale distributions whose own parameters have been fixed (e.g., F is the standard logistic cdf and H is a standard Cauchy cdf), then we have a two-parameter subfamily whose location is determined by μ and scale by σ. This paper focuses on this two-parameter subfamily, which we will call the cdf-quantile family. We will argue that it has several useful properties:

1. Many of its members are tractable, with explicit pdfs, cdfs, and quantiles.
2. They are amenable to likelihood inference, and thereby both maximum likelihood and Bayesian estimation techniques.
3. They enable 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.
4. The family can model a wide variety of distribution shapes, with different skew and kurtosis coverage from the beta or the Kumaraswamy.
5. The relationships between skew or kurtosis and the distribution parameters are straightforward and well behaved.
6. Explicit quantiles render random generation of variates straightforward.
7. Members of this family fall into four categories regarding behaviour at the boundaries of the (0,1) interval.

1.1. Examples

We present three examples from the cdf-quantile family. Our first example is the logit-Cauchy distribution. This distribution employs the logistic cdf $F(z) = \exp(z)/(1 + \exp(z))$ and the Cauchy cdf $H(z) = \tan^{-1}(z)/\pi + 1/2$. It therefore has domains $D_1 = (-\infty, \infty)$ and $D_2 = (-\infty, \infty)$. Inverting H and applying it and F to equation (1.4) gives

$$G(x,\mu,\sigma) = \frac{1}{1 + \exp\left(\frac{\mu+\cot(\pi x)}{\sigma}\right)}, \tag{1.13}$$

and differentiating it gives the pdf

$$g(x,\mu,\sigma) = \frac{\pi\csc^2(\pi x)\exp\left(\frac{\mu+\cot(\pi x)}{\sigma}\right)}{\sigma\left(\exp\left(\frac{\mu+\cot(\pi x)}{\sigma}\right) + 1\right)^2}. \tag{1.14}$$

Clearly $x \in (0,1)$, $\mu \in (-\infty,\infty)$ and $\sigma \in (0,\infty)$.

Inverting F and the appropriate substitutions into equation (1.13) give us the quantile

$$G^{-1}(\gamma,\mu,\sigma) = \frac{\tan^{-1}\left(\sigma\left(\frac{\mu}{\sigma} - \log\left(\frac{1}{\gamma} - 1\right)\right)\right)}{\pi} + \frac{1}{2}. \tag{1.15}$$

When $\gamma = 1/2$ equation (1.15) reduces to

$$G^{-1}(1/2,\mu,\sigma) = \frac{\tan^{-1}(\mu)}{\pi} + \frac{1}{2},$$

so we see that the median is solely a function of μ. For example, if $\mu = 2$ the median is $\tan^{-1}(2)/\pi + 0.5 = 0.852$, while if $\mu = -2$ the median is $\tan^{-1}(-2)/\pi + 0.5 = 0.148$.

We also can see that σ is a simple function of μ and the $e/(1 + e)$ quantile. First, we observe that

$$G^{-1}(e/(e+1),\mu,\sigma) = \frac{\tan^{-1}(\mu + \sigma)}{\pi} + \frac{1}{2}.$$

Solving for σ yields

$$\sigma = \tan\left[\pi(G^{-1}(e/(1+e),\mu,\sigma) - 1/2)\right] - \mu.$$

Our second example is the Cauchit-arcsinh distribution, which employs the Cauchy cdf $F(z) = (\tan^{-1}(z))/\pi + 1/2$ and the arcsinh cdf $H(z) = 1/\left(\exp\left(-\sinh^{-1}(z)\right) + 1\right)$. It therefore also has domains $D_1 = (-\infty,\infty)$ and $D_2 = (-\infty,\infty)$. Inverting H and applying it and F to equation (1.4) gives

$$G(x,\mu,\sigma) = \frac{\tan^{-1}\left(\left(\frac{1-2x}{2(x-1)x} - \mu\right)/\sigma\right)}{\pi} + \frac{1}{2} \tag{1.16}$$

and the pdf

$$g(x,\mu,\sigma) = \frac{2\sigma(2(x-1)x + 1)}{\pi\left(4\sigma^2(x-1)^2x^2 + (2\mu(x-1)x + 2x - 1)^2\right)}. \tag{1.17}$$

It can be shown that, unlike the beta distribution, the density at the boundaries of the unit interval is finite, with limits $g(0,\mu,\sigma) = g(1,\mu,\sigma) = 2\sigma/\pi$.

Inverting F and the appropriate substitutions into equation (1.16) give us the quantile

$$G^{-1}(\gamma, \mu, \sigma) = \frac{\sqrt{-2\mu\sigma\cot(\pi\gamma) + \sigma^2\cot^2(\pi\gamma) + \mu^2 + 1} - \sigma\cot(\pi\gamma) + \mu - 1}{2\mu - 2\sigma\cot(\pi\gamma)}. \tag{1.18}$$

When $\gamma = 1/2$ equation (1.18) reduces to

$$G^{-1}(1/2, \mu, \sigma) = \frac{\sqrt{\mu^2 + 1} + \mu - 1}{2\mu},$$

so again we see that the median is solely a function of μ. We also can see that σ is a simple function of μ and the first quartile. First, we observe that

$$G^{-1}(1/4, \mu, \sigma) = \frac{\sqrt{(\mu - \sigma)^2 + 1} + (\mu - \sigma) - 1}{2(\mu - \sigma)},$$

and solving for σ yields

$$\sigma = \frac{2G^{-1}(1/4, \mu, \sigma) - 1}{2(G^{-1}(1/4, \mu, \sigma) - 1)G^{-1}(1/4, \mu, \sigma)} - \mu.$$

Our third example is the logit-$F(1,1)$ distribution, which employs the logistic cdf $F(z) = \exp(z)/(1 + \exp(z))$ and the $F(1,1)$ cdf $H(z) = 2\tan^{-1}(\sqrt{x})/\pi$. It therefore has domains $D_1 = (-\infty, \infty)$ and $D_2 = (0, \infty)$. Inverting H and applying it and F to equation (1.4) gives

$$G(x, \mu, \sigma) = \frac{1}{\exp\left(-\frac{\log\left(\tan^2\left(\frac{\pi x}{2}\right)\right) - \mu}{\sigma}\right) + 1} \tag{1.19}$$

and the pdf

$$g(x, \mu, \sigma) = \frac{2\pi\exp(\mu/\sigma)\csc(\pi x)\tan^2\left(\frac{\pi x}{2}\right)^{1/\sigma}}{\sigma\left(\exp(\mu/\sigma) + \tan^2\left(\frac{\pi x}{2}\right)^{1/\sigma}\right)^2}. \tag{1.20}$$

Inverting F and the appropriate substitutions into equation (1.19) give us the quantile

$$G^{-1}(\gamma, \mu, \sigma) = \frac{2\tan^{-1}\left(\sqrt{\exp\left(\mu - \sigma\log\left(\frac{1-\gamma}{\gamma}\right)\right)}\right)}{\pi}. \tag{1.21}$$

When $\gamma = 1/2$ equation (1.21) reduces to

$$G^{-1}(1/2, \mu, \sigma) = \frac{2\tan^{-1}\left(\sqrt{\exp(\mu)}\right)}{\pi}.$$

By a similar procedure to the preceding two examples, it can be shown that

$$\sigma = \log\left[\tan(\pi G^{-1}(e/(e+1), \mu, \sigma)/2)^2\right] - \mu.$$

These examples are illustrations that in this family the distribution location may be described by μ and the scale (or dispersion) by σ. These characterizations will be proved in the next section and exploited in Section 4 when we present the general linear model based on this family and its associated estimation methods.

The remainder of this paper consists of six sections. The next section describes the characteristics of the cdf-quantile family, elaborating its interpretation and the nature of its parameters. The section thereafter forcuses on a subfamily that is particularly useful for general linear models of proportions and rates, and that captures four kinds of distribution shape. Section 4 then moves to maximum likelihood inference, including the construction and evaluation of general linear models that employ the subfamily developed in Section 3. Section 5 presents an example from real data sets, demonstrating that cdf-quantile distributions can fit data better than the beta distribution (a second example is presented in the Supporting Information). The final section concludes with prospects for further developments and applications.

2. Basic properties

We begin by offering some interpretations of the cdf-quantile distribution family. Perhaps the most fundamental question raised by introducing a new distribution family is what processes or phenomena are analogous to these distributions. The beta distribution has a derivation as the ratio of one gamma-distributed random variable to its sum with another gamma random variable, both of which share one parameter. An analogous example is the proportion of a task's duration devoted to a particular subtask.

A useful interpretation of the cdf-quantile family is that H transforms X to a random variable, Y, say, in D_1, and then Y is fitted by a location–scale distribution whose cdf is $F(U(y,\mu,\sigma))$. In this sense, H and the appropriate U function are analogous to the link function in a general linear model (GLM). Because X and G share the same domain, an obvious analogy is to consider G as a redistribution of X.

There are many analogous phenomena in psychology. For example, if we ask a random sample of automobile drivers to estimate the percentage of drivers who are worse at driving than they are, we would expect a well-calibrated sample to exhibit a uniform distribution, that is, $G(x,\mu,\sigma) = x$ so that $\mu = 0$ and $\sigma = 1$. An overconfident sample, on the other hand, should return a positive estimate of μ, with the median driver claiming to have driving ability above the median, and $\sigma < 1$. Another example (elaborated later in this paper) is a comparison of positive and negative probability expressions, such as 'likely' and 'unlikely'. If people's personal numerical translations of these terms have the probability associated with 'likely' being equivalent to the probability associated with 'unlikely' subtracted from 1, then we should find that $G(x, \mu, \sigma)$ for the former is identical to $1-G(x, -\mu, \sigma)$ for the latter.

The redistribution of X by G potentially has two components, one if $F \neq H$ and the other if $\mu \neq 0$ and/or $\sigma \neq 1$. It also immediately follows that if $F = H$, then when $\mu = 0$ and $\sigma = 1$, $G(x,\mu,\sigma) = x$ and therefore G includes the uniform pdf as a special case.

We now elaborate four important properties of members of the cdf-quantile family, namely that (1) under certain conditions they are self-dual in an analogous fashion to the beta distribution, (2) μ is a location parameter in the sense that the median is solely a function of μ, (3) σ is a dispersion parameter in the sense that it controls the spread of quantiles, and (4) for members of this distribution family there is a relationship between the tail behaviour of the pdf and the first derivative of the quantile function.

Proposition 1. *Under conditions specified in the cases below,* g(x, μ, σ) *is self-dual in this respect:*

$$g(x, \mu, \sigma) = g(1 - x, -\mu, \sigma). \tag{2.1}$$

Proof. There are four cases, corresponding to the domain pairs for equations (1.5)–(1.8).

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),

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

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),*

$$\begin{aligned} 1 - G(1 - x, -\mu, \sigma) &= 1 - F\left[\left(-\log\left(H^{-1}(x)\right) + \mu\right)/\sigma\right] = F\left[\left(\log\left(H^{-1}(x)\right) - \mu\right)/\sigma\right] \\ &= G(x, \mu, \sigma). \end{aligned}$$

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),*

$$\begin{aligned} 1 - G(1 - x, -\mu, \sigma) &= 1 - F\left[\left(H^{-1}(1 - x)\exp(\mu)\right)^{1/\sigma}\right] = F\left[\left(H^{-1}(x)\exp(-\mu)\right)^{1/\sigma}\right] \\ &= G(x, \mu, \sigma). \end{aligned}$$

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),*

$$\begin{aligned} 1 - G(1 - x, -\mu, \sigma) &= 1 - F\left[\exp\left(\left(-H^{-1}(x) + \mu\right)/\sigma\right)\right] = F\left[\exp\left(\left(H^{-1}(x) - \mu\right)/\sigma\right)\right] \\ &= G(x, \mu, \sigma). \end{aligned}$$

For each of these cases, we have shown that $1-G(1-x, -\mu, \sigma) = G(x, \mu, \sigma)$, which gives our result.

There are conditions under which the quantile may be expressed purely in terms of the location parameter μ. For modelling purposes, it is useful to know when the median has

this property because that identifies μ with the median (i.e., location) submodel, as will be elaborated later in this paper. For cdf-quantile distributions whose median is solely a function of μ, μ is a location parameter. We already have presented examples in Section 1.1 where this is the case. The following proposition presents the conditions under which this occurs.

Proposition 2. *For support* $D_1 = (-\infty, \infty)$ *the requirement is that for some real number c, which does not depend on* σ*,*

$$F^{-1}(1/2) = c/\sigma, \tag{2.2}$$

with the restriction that the constant $c \neq -\mu$*. For support* $D_1 = (0, \infty)$ *the requirement is*

$$F^{-1}(1/2) = b^{c/\sigma}, \tag{2.3}$$

where b is a non-negative real number, with the restriction that $b^{c/\sigma} \neq \exp(-\mu/\sigma)$*.*

Proof. Both parts of the proposition follow immediately from the quantile functions in equations (1.9) to (1.12).

We now turn to considering the conditions under which σ is a dispersion parameter. Following Mitnik and Baek (2013, pp. 181–182), we show that σ is a dispersion parameter via an argument based on the 'quantile spread order' from Townsend and Colonius (2005). Define quantile spread as

$$Q(\gamma, \mu, \sigma) = G^{-1}(1 - \gamma, \mu, \sigma) - G^{-1}(\gamma, \mu, \sigma), \tag{2.4}$$

for $0 < \gamma < 1/2$. If it can be shown that $Q(\gamma, \mu, \sigma_1) < Q(\gamma, \mu, \sigma_2)$ if and only if $\sigma_1 < \sigma_2$, then σ has been shown to be a dispersion parameter. We prove that this holds under conditions that are a special case of the conditions for Proposition 2.

Proposition 3. *For* $D_1 = (-\infty, \infty)$ *and* $F^{-1}(1/2) = 0$*, and for* $D_1 = (0, \infty)$ *and* $F^{-1}(1/2) = 1$*,* $Q(\gamma, \mu, \sigma_1) < Q(\gamma, \mu, \sigma_2)$ *if and only if* $\sigma_1 < \sigma_2$*.*

Proof. As in Proposition 1, we have four cases.
Cases 1 and 2. For $D_1 = (-\infty, \infty)$, $D_2 = (-\infty, \infty)$, and $F^{-1}(1/2) = 0$,

$$Q(\gamma, \mu, \sigma_1) = H\left[\sigma_1 F^{-1}(1 - \gamma) + \mu\right] - H\left[\sigma_1 F^{-1}(\gamma) + \mu\right]$$

and

$$Q(\gamma, \mu, \sigma_2) = H\left[\sigma_2 F^{-1}(1 - \gamma) + \mu\right] - H\left[\sigma_2 F^{-1}(\gamma) + \mu\right].$$

Because $F^{-1}(1/2) = 0$, we have $F^{-1}(\gamma) < 0$ and $F^{-1}(1 - \gamma) > 0$. Thus,

$$H\left[\sigma_2 F^{-1}(1-\gamma) + \mu\right] > H\left[\sigma_1 F^{-1}(1-\gamma) + \mu\right],$$

and

$$H\left[\sigma_2 F^{-1}(\gamma) + \mu\right] < H\left[\sigma_1 F^{-1}(\gamma) + \mu\right].$$

Therefore, $Q(\gamma, \mu, \sigma_1) < Q(\gamma, \mu, \sigma_2)$ if and only if $\sigma_1 < \sigma_2$. A similar argument goes through for Case 2, where $D_1 = (-\infty, \infty)$ and $D_2 = (0, \infty)$.

Cases 3 and 4. For $D_1 = (0, \infty)$, $D_2 = (-\infty, \infty)$, and $F^{-1}(1/2) = 1$,

$$Q(\gamma, \mu, \sigma_1) = H\left[\mu + \sigma_1 \log\left(F^{-1}(1-\gamma)\right)\right] - H\left[\mu + \sigma_1 \log\left(F^{-1}(\gamma)\right)\right]$$

and

$$Q(\gamma, \mu, \sigma_2) = H\left[\mu + \sigma_2 \log\left(F^{-1}(1-\gamma)\right)\right] - H\left[\mu + \sigma_2 \log\left(F^{-1}(\gamma)\right)\right].$$

Because $F^{-1}(1/2) = 1$, we have $\log\left(F^{-1}(\gamma)\right) < 0$ and $\log\left(F^{-1}(1-\gamma)\right) > 0$. Thus,

$$H\left[\mu + \sigma_2 \log\left(F^{-1}(1-\gamma)\right)\right] > H\left[\mu + \sigma_1 \log\left(F^{-1}(1-\gamma)\right)\right]$$

and

$$H\left[\mu + \sigma_2 \log\left(F^{-1}(\gamma)\right)\right] < H\left[\mu + \sigma_1 \log\left(F^{-1}(\gamma)\right)\right].$$

Therefore, $Q(\gamma, \mu, \sigma_1) < Q(\gamma, \mu, \sigma_2)$ if and only if $\sigma_1 < \sigma_2$. A similar argument goes through for Case 4, where $D_1 = (0, \infty)$ and $D_2 = (0, \infty)$, but only for $1 \leq \sigma_1$ and $1 \leq \sigma_2$.

There is a relation that is apparent from examining pairs of these distributions in which F and H exchange roles. These pairs 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. We will denote these distributions with the nomenclature F–H (e.g., Cauchit-logistic and logit-Cauchy). Jones (2002) refers to these as 'complementary distributions', and applied this exchange to the beta distribution, but observed later (Jones, 2009) that doing so with the beta yields little practical gain. We shall see that in the cdf-quantile family, this exchange produces distributions that differ in important respects from both their quantile-dual counterparts. The chief difference between them stems from the following proposition.

Proposition 4. *For any cdf G whose support is* (0,1)*,*

$$\lim_{x \to 0} \frac{\partial G(x, \mu, \sigma)}{\partial x} = \lim_{\gamma \to 0} \frac{1}{\partial G^{-1}(\gamma, \mu, \sigma)/\partial \gamma}$$

and

$$\lim_{x\to 1}\frac{\partial G(x,\mu,\sigma)}{\partial x}=\lim_{\gamma\to 1}\frac{1}{\partial G^{-1}(\gamma,\mu,\sigma)/\partial\gamma}.$$

Proof. Setting $G(x,\mu,\sigma)=\gamma$, by the inverse rule of differentiation, for every x in (0,1) there exists γ in (0,1) such that

$$\frac{\partial G(x,\mu,\sigma)}{\partial x}=\frac{1}{\partial G^{-1}(\gamma,\mu,\sigma)/\partial\gamma}.$$

The result immediately follows by observing that $G(0,\mu,\sigma)=0$ and $G(1,\mu,\sigma)=1$.

3. Distributions with identical parent distribution supports

For the remainder of the paper, we focus on those $F-H$ distributions for which the parent distribution supports are $D_1=D_2=(-\infty,\infty)$. Our investigations thus far have indicated that these are the most promising for applications. We note in passing that neither Aljarrah *et al.* (2014) nor Lemonte and Bazán (2016) systematically investigate the kinds of shapes that these distributions can fit. Lemonte and Bazán, for instance, give examples of unimodal and trimodal distributions without remarking on this. We have found that these distributions fall into four kinds of characteristic shapes, which can be described by their density's tail behaviour 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
4. $\lim_{x\to 0} g(x,\mu,\sigma)=\lim_{x\to 1} g(x,\mu,\sigma)=\infty$.

The first subgroup, which we denote here by LL, is typified by the logit-logistic distribution. The LL distributions share some properties with the beta and Kumaraswamy distributions, but have a different coverage of skewness and kurtosis (see the Supporting Information). Some LL distributions include the uniform as a special case, but not all do. When $\sigma>s$ the distribution is unimodal, and when $\sigma<s$ the distribution has a bathtub shape with modes at 0 and 1. The logit-logistic distribution's $v(\mu)=\exp(\mu)$ when $\sigma=s=1$, and other members of this subgroup have similar functions of μ at the threshold value for σ. It follows from Proposition 4 that if the $F-H$ distribution is an LL distribution, then so is its quantile-dual $H-F$ distribution. Figure 1 displays several examples of the logit-logistic distribution for selected values of μ and σ.

The second subgroup, referred to here as the bimodal (BM), is capable of being bimodal on the (0,1) interval, thanks to the limit of 0 for the density at 0 and 1, which the beta, Kumaraswamy, and LL distributions are not. Also because of this limit, the second subgroup does not include the uniform distribution as a special case, but has symmetric special cases at $x=1/2$. Figure 2 displays several examples from the T2-Cauchy distribution, a typical BM distribution (T2 refers to a t distribution with 2 degrees of freedom; see Table 1 for the T2 and Cauchy functions that define this distribution).

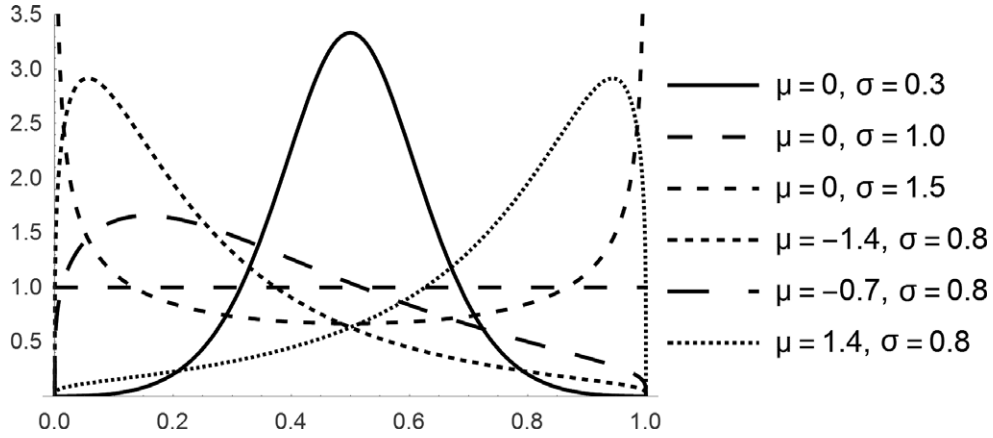

Figure 1. Logit-logistic distribution probability density functions.

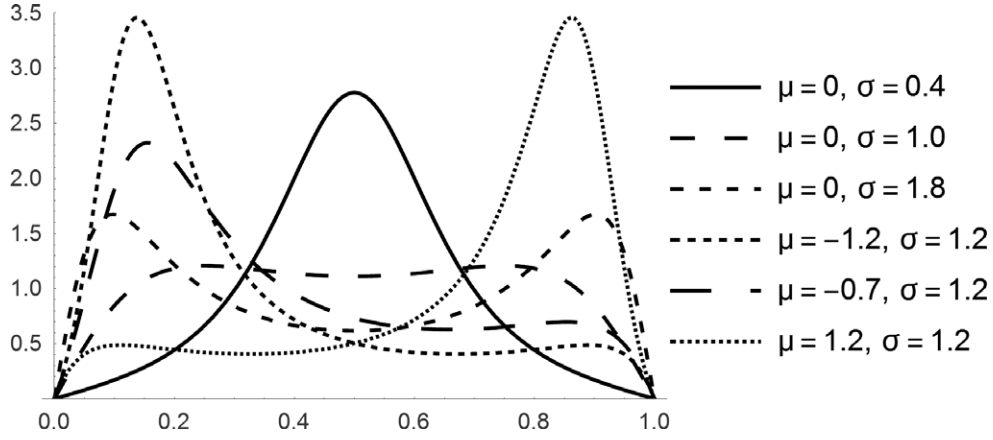

Figure 2. T2-Cauchy distribution probability density functions.

The third subgroup has non-zero but finite density in the limit at 0 and 1 as a function of σ, and we shall refer to it as the finite-tailed (FT) group. Some FT distributions include the uniform pdf, but not all do. The arcsinh-arcsinh distribution's limiting density at 0 and 1 is $u(\sigma) = \sigma$, and for other members of this subgroup $u(\sigma)$ is a simple monotonically increasing function of σ. Although generally unimodal, for sufficiently large σ, these distributions have minor modes at 0 and 1. It follows from Proposition 4 that if the $F{-}H$ distribution is an FT distribution, then so is its quantile-dual $H{-}F$ distribution. Figure 3 displays several examples from the arcsinh-arcsinh distribution (see Table 1 for the definition of the ArcSinh function).

The fourth subgroup is trimodal, with two modes at 0 and 1 due to the infinite densities at the limit. We refer to it as the trimodal (TM) group, and it does not include the uniform pdf but does have symmetric special cases at $x = 1/2$. Figure 4 displays several examples from the Cauchit-T2 distribution. This distribution is the quantile-dual of the T2-Cauchy, whose distribution is graphed in Figure 2. It follows from Proposition 4 that the $F{-}H$ distribution is BM if and only if its quantile-dual $H{-}F$ distribution is TM.

Table 1. Distributions with support $D_1 = D_2 = (-\infty, \infty)$ by shape

			H					
Subfamily	F	f	Logistic	Cauchy	Arcsinh	T2	Burr VII	Burr VIII
Logit	$\frac{\exp(z)}{1+\exp(z)}$	$\frac{\exp(z)}{(\exp(z)+1)^2}$	LL*	BM	BM	BM	LL*	LL
Cauchit	$\frac{\tan^{-1}(z)}{\pi}+\frac{1}{2}$	$\frac{1}{\pi z^2+\pi}$	TM	FT*	FT	TM	TM	TM
Arcsinh	$\frac{1}{\exp(-\sinh^{-1}(z))+1}$	$\frac{\exp(\sinh^{-1}(z))}{\sqrt{z^2+1}(\exp(\sinh^{-1}(z))+1)^2}$	TM	FT	FT*	TM	TM	TM
T2	$\frac{z}{2\sqrt{z^2+2}}+\frac{1}{2}$	$\frac{1}{(z^2+2)^{3/2}}$	TM	BM	BM	FT*	TM	TM
Burr VII	$\frac{\tanh(z)+1}{2}$	$\frac{\operatorname{sech}^2(z)}{2}$	LL*	BM	BM	BM	LL*	LL
Burr VIII	$\frac{2\tan^{-1}(\exp(z))}{\pi}$	$\frac{2\exp(z)}{\pi+\exp(2z)\pi}$	LL	BM	BM	BM	LL	LL*

Notes. LL = logit-logistic shape; BM = can be bimodal on the (0,1) interval; FT = finite-tailed shape; TM = trimodal, with modes at 0 and 1.
*Includes the uniform distribution as a special case.

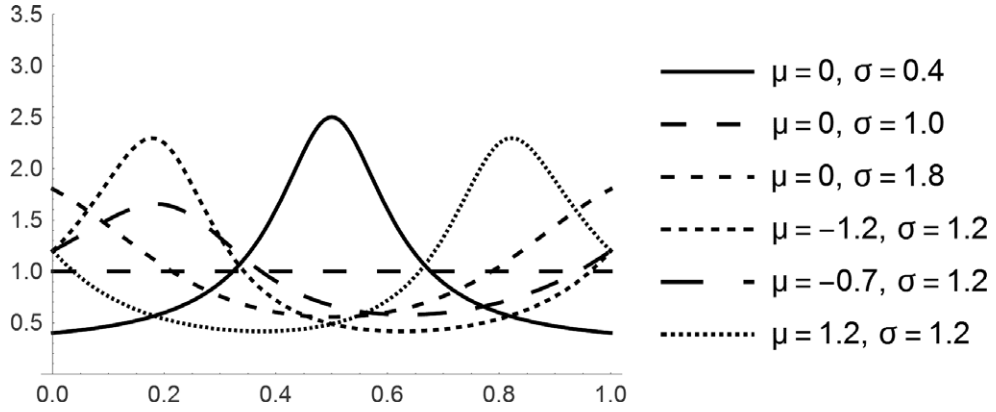

Figure 3. Arcsinh-arcsinh distribution probability density functions.

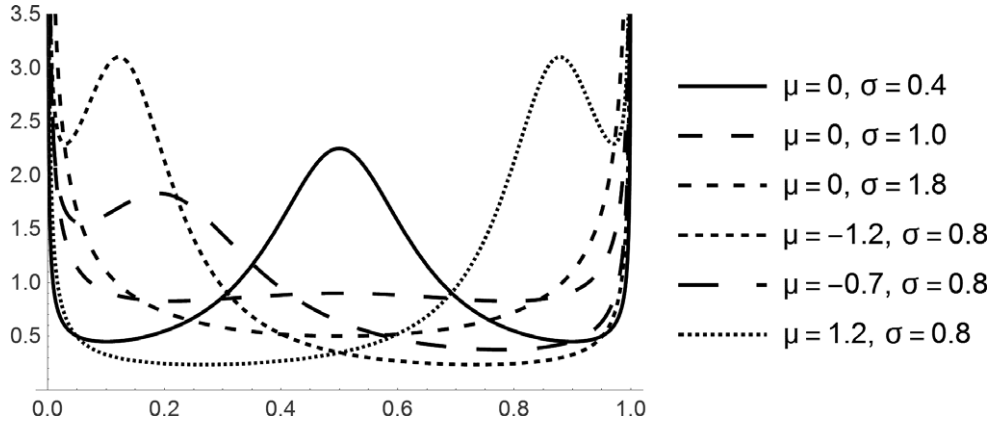

Figure 4. Cauchit-T2 distribution probability density functions.

Table 1 classifies various F–H distributions for which the parent distribution supports are $D_1 = D_2 = (-\infty, \infty)$, by subfamily and shape. All of them are self-dual and therefore have symmetric special cases, but only the LL and some of the FT distributions include the uniform as a special case.

The finding that the subfamily of distributions for which the parent distribution supports are $D_1 = D_2 = [-\infty, \infty]$ divides into four categories of shapes suggests that their skewness–kurtosis maps might fall into corresponding categories. We investigated two kinds of skewness and kurtosis measures, both of which take advantage of an explicit expression for the quantile. First, we employed the classical quantile-based measures of skewness (Bowley, 1917) and kurtosis (Moors, 1988). These are the skewness and kurtosis measures employed by Alzaatreh *et al.* (2013) in exploring the properties of their distribution family. Second, we also used the L-moment measures, which may be computed via linear combinations of probability-weighted moments (Greenwood, Landwehr, Matalas & Wallis, 1979) to provide a check on the findings from the classical measures.

These investigations are detailed in the Supporting Information. As demonstrated therein, the cdf-quantile family is well behaved with regard to skew and kurtosis as functions of its location and dispersion parameters. The distributions there cover a variety of skew and kurtosis combinations that differ from coverage by the Kumaraswamy or the beta. The characteristics of their skew–kurtosis maps correspond to the four shape categories described earlier.

4. Inference and modelling

4.1. Maximum likelihood inference and regression model

Maximum likelihood inference for this distribution family hinges entirely on the nature of F and H. For $D_1 = D_2 = (-\infty, \infty)$ the pdf may be written as

$$g(x, \mu, \sigma) = \frac{q(x) f\left(\frac{H^{-1}(x)-\mu}{\sigma}\right)}{\sigma}, \tag{4.1}$$

where f is the pdf corresponding to F, and q is the quantile distribution function corresponding to H^{-1}. The log-likelihood function therefore is

$$l(x, \mu, \sigma) = \log\left(f\left(\frac{H^{-1}(x) - \mu}{\sigma}\right)\right) + \log(q(x)) - \log(\sigma). \tag{4.2}$$

Differentiating the log of g with respect to μ and σ drops q and yields

$$\partial \log(g(x, \mu, \sigma))/\partial\mu = -\frac{\partial f\left(\frac{H^{-1}(x)-\mu}{\sigma}\right)/\partial\mu}{\sigma f\left(\frac{H^{-1}(x)-\mu}{\sigma}\right)}, \tag{4.3}$$

$$\partial \log(g(x, \mu, \sigma))/\partial\sigma = \frac{(\mu - H^{-1}(x))\left(\partial f\left(\frac{H^{-1}(x)-\mu}{\sigma}\right)/\partial\sigma\right)}{\sigma^2 f\left(\frac{H^{-1}(x)-\mu}{\sigma}\right)} - \frac{1}{\sigma}. \tag{4.4}$$

Equations (4.3) and (4.4) show that the sole requirement for an explicit gradient is that f is differentiable. A similar argument leads to the requirement that f is twice differentiable to obtain an explicit Hessian, from which standard errors of the parameter estimates may be obtained. All of the examples in Table 1 satisfy these requirements, and so maximum likelihood estimation (MLE) with them may be achieved using standard estimation methods.

The cdf-quantile family and its appropriate estimation methods also are in the tradition of log-location–scale modelling (e.g., Lawless, 2011), with G inheriting important properties of F. For instance, log-concavity of G in $H^{-1}(x)$ is guaranteed if the parent pdf, $f(x, \mu, \sigma) = \partial F(x, \mu, \sigma)/\partial x$, is log-concave (thereby ensuring that the cdf, F, also is log-concave).

We are now in a position to describe and illustrate the GLM of quantiles (i.e., the quantile regression model) for random variables on the (0,1) interval. Consider a random sample of N independent observations from $(Y, \mathbf{V})$, where Y is a random variable with support (0,1) and $\mathbf{V}$ is a vector of predictors. A model for the distribution of Y, conditional on $\mathbf{V}$, has two submodels, the 'location submodel' for μ and the 'dispersion submodel' for

σ. The location submodel has a vector **W** of predictors selected from **V**, and the dispersion model has another vector of predictors **Z** selected from **V**. The sets of predictors in **W** and **Z** may or may not overlap. For the *i*th observation, these submodels may be written as

$$L_{\mu}(\hat{\mu}_i) = \mathbf{w}_i^{\mathrm{T}}\boldsymbol{\beta}, \quad i = 1, \ldots, N, \tag{4.5}$$

$$L_{\sigma}(\hat{\sigma}_i) = \mathbf{z}_i^{\mathrm{T}}\boldsymbol{\delta}, \quad i = 1, \ldots, N, \tag{4.6}$$

where L_{μ} and L_{σ} are link functions in the usual GLM sense of this term, $\hat{\mu}_i$ and $\hat{\sigma}_i$ are maximum likelihood estimates of μ and σ respectively, conditional on $\mathbf{w}_i$ and $\mathbf{z}_i$ which are the observed values of the vectors of the predictors selected from **V**, and $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ are vectors of coefficients yielding the maximum likelihood estimates. For $D_1 = D_2 = (-\infty, \infty)$ a practical location submodel link function is the identity, and a practical link function for the dispersion submodel is the log.

For large samples, under the usual regularity conditions (Pawitan, 2001) for MLE, the joint distribution of the maximum likelihood estimates is approximately multivariate normal, so that

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} \end{pmatrix} \sim N\left(\begin{pmatrix} \beta \\ \gamma \end{pmatrix}, M^{-1}\right), \tag{4.7}$$

where M^{-1} is the inverse Fisher information matrix. The Wald confidence interval for each parameter θ_j in the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta},\boldsymbol{\gamma})$ is

$$\hat{\theta}_j \pm \Phi^{-1}(1 - \alpha/2)se\left(\hat{\theta}_j\right), \tag{4.8}$$

where $\hat{\theta}_j$ is the maximum likelihood estimate of θ_j, Φ^{-1} is the cdf of the normal distribution, $1-\alpha$ is the confidence level, and

$$se\left(\hat{\theta}_j\right) = \left[\mathrm{diag}\left(\mathrm{cov}\left(\hat{\boldsymbol{\theta}}\right)\right)\right]_j^{1/2} \tag{4.9}$$

is the standard error of $\hat{\theta}_j$ in which $\mathrm{cov}\left(\hat{\boldsymbol{\theta}}\right) = M^{-1}\left(\hat{\boldsymbol{\theta}}\right)$ is the asymptotic variance–covariance matrix of $\hat{\boldsymbol{\theta}}$. In the cdf-quantile family the median is a monotonic function of the location parameter μ, so under the conditions specified for Proposition 3 the confidence interval for $\hat{\mu}$ can be converted into a confidence interval for the median by substituting the lower and upper limits into the formula for the median. For our first example, the logit-Cauchy distribution, from equation (1.15) we can see that the confidence interval around the median is $\tan^{-1}(\hat{\mu} \pm \Phi^{-1}(1 - \alpha/2)se(\hat{\mu}))/\pi + 1/2$ For instance, if $\hat{\mu} = 2.5$ and $se(\hat{\mu}) = 0.6$, then the 95% confidence interval for μ is [1.324,3.676], the estimate for the median is 0.879, and the 95% confidence interval around it is [0.794,0.915].

We have conducted simulations investigating the distributions of maximum likelihood estimates for μ and σ from samples from each of the 16 distributions included in the original cdfquantreg R package which comprise a representative selection from the family, for sample sizes of 20, 100, and 200. These distributions are listed in Figures S3 and S4 in the Supporting Information. We assessed each distribution for estimation bias and

the normality of the maximum likelihood estimate distributions. As detailed in the Supporting Information, the maximum likelihood estimates for both parameters exhibit close approximations to normal distributions even for sample sizes of 20. There is little indication of estimator bias for either parameter.

We also have conducted simulations investigating the Type I error-rate coverage of the likelihood-ratio test for four distributions from the family (arcsinh-arcsinh, logit-logistic, T2-Cauchy, and Cauchit-T2) along with the beta distribution, for sample sizes of 20 and 100. Details are available in the Supporting Information. Briefly, Type I error rates were accurate for both the location and dispersion submodels when distributions were used to model random samples from the same distributions, and location submodel error rates also were accurate even when models used distributions differing from the sampled distribution. Dispersion submodel error rates tended to become somewhat inflated when the modelling distribution differed from the sampled distribution, most strongly for the TM and least so for the BM distributions, and to a lesser extent with larger samples. This was more strongly the case for the beta distribution than for most of the cdf-quantile distributions, so it appears to be a generic hazard in modelling doubly bounded random variables and cautions researchers to ensure that their models are not misspecified.

4.2. Fitting distributions

The prospect of model evaluation immediately raises two important issues: the choice of a distribution from the family; and what constitutes appropriate model diagnostics. Regarding the first issue, it should be noted that the current state of the art offers only scanty guidelines for selecting among alternative distributions. The conventional approaches favour one information-criterion measure or another; for example, Rust, Simester, Brodie and Nilikant (1995) advocate the Bayesian information criterion (BIC). Vrieze (2012) presents a thoughtful exposition on the choice between the BIC and the Akaike information criterion (AIC). However, these considerations become irrelevant when comparing alternative distributions on the same data set with the same predictors, because differences between any such information criteria simply reduce to differences between the log-likelihoods (see our example in Section 5). If we turn now to the T-X and related families, none of the relevant papers offer guidelines for choosing one distribution in such a family over another. Lemonte and Bazán (2016) utilize the AIC in their examples when they compare alternative distributions for goodness of fit.

In fact, the empirical approach is the state of the art when it comes to deciding whether to use the beta, simplex, logit-logistic, Kumaraswamy, or any other appropriate distributions available in the current literature. One of the main reasons for this is, as Kieschnick and McCullough (2003, p. 95) point out, that when it comes to proportions 'there is no clear agreement on the data generating process for such data. Thus, we fit the different regression models to a common data set using a common specification of the regressors to determine which regression model best describes the data.'

An in-depth examination of the prospects for an agreed-upon data generating process (or processes) for proportional data is beyond the scope of this paper, but we add one suggestion here, referring to the tail behaviour of the distributions in the cdf-quantile family. There may be theoretical or practical reasons to prefer one type of distribution over another on the basis of its behaviour near the boundaries of the unit interval. For

example, if one expects never to observe values close to 0 or 1 then the BM and LL types would seem preferable to the TM and FT types. On the other hand, as our example demonstrates, a 'long' tail in the observed distribution may favour the FT or TM types for modelling it. Given the current state of the art, we recommend that researchers fit several alternative distributions to their data unless they have compelling reasons to narrow their selection *a priori*.

A systematic investigation of the model-diagnostic issues of model residuals and influence measures for this distribution family is also beyond our scope, and these are active areas of research. However, traditional residuals such as the Pearson, Anscombe, and deviance residuals are applicable here, and so are influence measures such as the dfbetas. We make one suggestion regarding these, namely that residuals allow for terms in both the location and dispersion submodels. For instance, we follow Smithson and Merkle (2014, p. 171) in defining the Pearson residual as $r_i = (y_i - \hat{\mu}_i)/\hat{\sigma}_i$, so that non-constant variance is permitted.

Lemonte and Bazán (2016) discuss methods of identifying local influence via Cook's (1986) distance under five perturbation schemes. They also utilize deviance residuals and a modified form of these (see Rocha & Simas, 2011) as leverage measures. We do not investigate these in this paper, but refer readers to these papers.

Lemonte and Bazán mention but do not examine the sensitivity of different distributions to observations near 0 or 1, whereas we have conducted simulations to investigate this issue. As detailed in the Supporting Information, our simulations show that the cdf-quantile distributions' estimators for quantiles such as the 25th, 50th, and 75th percentiles are more stable under the influence of outliers than the corresponding estimators yielded by the beta distribution. The same therefore holds for the corresponding location and dispersion parameter estimates. We conclude on this basis that the cdf-quantile distributions are, overall, better suited than the beta distribution for modelling quantiles of rates and proportions in the presence of outliers near 0 or 1.

5. Example: Lay interpretations of IPCC report probability phrases

We now present an example of model fitting in which distributions from the cdf-quantile family outperform the beta and Kumaraswamy distributions. 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, Broomell and Por (2009) 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 (e.g., estimates of the probability meant by 'likely' in the sentence 'The Greenland ice sheet and other Arctic ice fields likely contributed no more than 4 m of the observed sea level rise'). Budescu et al. found that participants' 'best' estimates were more regressive (towards the middle of the [0, 1] interval) than the IPCC stipulations. In a reanalysis of their data using beta regression, Smithson, Budescu, Broomell and Por (2012) reported that this tendency was stronger for negatively worded phrases (e.g., 'unlikely') than for positively worded phrases. Moreover, they found greater dispersion of responses (i.e., less consensus) for negative than for positive phrases.

We reprise part of this modelling exercise here, comparing the performance of the beta regression model with its logit-logistic and T2–T2 counterparts, as examples from the

Table 2. Model fits and parameter estimates for 'best' estimates

Item	Distribution	$\hat{\mu}$	*SE*	$\log(\hat{\sigma})$	*SE*	–LL	$Q_{.25}$	$Q_{.5}$	$Q_{.75}$
Q4 ('likely')	Empirical	0.658					0.570	0.699	0.804
	Beta	0.637	0.062	3.940	0.340	−44.164	0.483	0.662	0.813
	Logit-log.	0.825	0.067	−0.519	0.057	−62.180	0.543	0.695	0.814
	T2–T2	0.643	0.045	−0.659	0.075	−76.160	0.577	0.707	0.801
Q5 ('likely')	Empirical	0.646					0.585	0.679	0.769
	Beta	0.643	0.052	6.150	0.546	−76.842	0.519	0.659	0.782
	Logit-log.	0.720	0.054	−0.748	0.057	−87.560	0.550	0.673	0.776
	T2–T2	0.553	0.035	−0.915	0.074	−98.630	0.579	0.682	0.764
Q6 ('likely')	Empirical	0.697					0.610	0.719	0.799
	Beta	0.700	0.057	5.345	0.476	−83.389	0.579	0.726	0.845
	Logit-log.	0.940	0.062	−0.594	0.058	−85.260	0.583	0.719	0.824
	T2–T2	0.653	0.039	−0.797	0.075	−99.030	0.599	0.710	0.793
Q8 ('unlikely')	Empirical	0.652					0.460	0.749	0.838
	Beta	0.614	0.072	2.438	0.201	−19.763	0.413	0.649	0.842
	Logit-log.	0.889	0.093	−0.201	0.057	−35.630	0.497	0.709	0.857
	T2–T2	0.807	0.062	−0.319	0.078	−53.640	0.575	0.748	0.852
Q9 ('unlikely')	Empirical	0.648					0.455	0.759	0.829
	Beta	0.616	0.070	2.666	0.222	−21.880	0.423	0.648	0.833
	Logit-log.	0.870	0.090	−0.249	0.057	−36.810	0.503	0.705	0.849
	T2–T2	0.820	0.058	−0.398	0.081	−56.080	0.594	0.751	0.848
Q10 ('unlikely')	Empirical	0.667					0.520	0.749	0.833
	Beta	0.646	0.069	2.943	0.249	−34.526	0.472	0.683	0.850
	Logit-log.	0.930	0.085	−0.286	0.057	−46.040	0.526	0.717	0.853
	T2–T2	0.769	0.059	−0.393	0.076	−57.370	0.576	0.739	0.841

Notes. The 'unlikely' data (Q8–Q10) have been subtracted from 1 to render them directly comparable with the 'likely' data. This is also the case for the histograms in Figure 5. The $\hat{\mu}$ and $\hat{\sigma}$ parameters in this table do not have the conventional meanings of a mean and standard deviation. They are the mean and precision for the beta distribution, and otherwise they are the parameters for members of the cdf-quantile family as defined in equation (2.1). Logit-log. = logit-logistic

cdf-quantile family (we will bring in other examples shortly). The cdf and pdf of the logit-logistic distribution are

$$G(x, \mu, \sigma) = 1/(\exp[(\mu + \log((1-x)/x))/\sigma] + 1),$$
$$g(x, \mu, \sigma) = \frac{\exp(\mu/\sigma)((1-x)/x)^{1/\sigma-1}}{\left(x + \exp(\mu/\sigma)((1-x)/x)^{1/\sigma}x\right)^2 \sigma} \quad . \tag{5.1}$$

The T2–T2 cdf is

$$G(x, \mu, \sigma) = \frac{w(x) - \mu}{\sigma\left(2\sqrt{((w(x)-\mu)/\sigma)^2 + 2}\right)} + \frac{1}{2}, \tag{5.2}$$

where

Table 3. Parameter estimates, quantiles, and negative log-likelihoods for various distributions

Distribution	$-LL$	param.	β	*SE*	95% lower	CI upper	δ	*SE*	95% lower	CI upper
Beta	−265	Intercept	0.512	0.041			0.977	0.048		
		Valence	0.150	0.053	0.046	0.254	0.603	0.070	0.466	0.740
Kumaraswamy	−269.157	Intercept	0.524	0.052			0.019	0.053		
		Valence	0.573	0.069	0.438	0.708	0.560	0.078	0.408	0.713
Logit-logistic	−345.185	Intercept	0.897	0.052			−0.245	0.033		
		Valence	−0.071	0.062	−0.193	0.051	−0.367	0.047	−0.459	−0.275
T2–T2	−435.294	Intercept	0.798	0.034			−0.368	0.045		
		Valence	−0.186	0.041	−0.267	−0.105	−0.421	0.062	−0.543	−0.299
Arcsinh-Cauchy	−424.615	Intercept	1.023	0.039			−0.597	0.057		
		Valence	−0.339	0.047	−0.431	−0.247	−0.450	0.079	−0.605	−0.295
T2–Burr VII	−412.531	Intercept	0.542	0.021			−0.887	0.046		
		Valence	−0.118	0.025	−0.168	−0.069	−0.362	0.062	−0.485	−0.240
T2–Burr VIII	−425.799	Intercept	0.868	0.034			−0.374	0.045		
		Valence	−0.195	0.042	−0.277	−0.113	−0.379	0.062	−0.501	−0.258
Burr VIII-Burr VII	−375.702	Intercept	0.486	0.024			−0.786	0.035		
		Valence	−0.069	0.029	−0.126	−0.011	−0.365	0.050	−0.463	−0.268

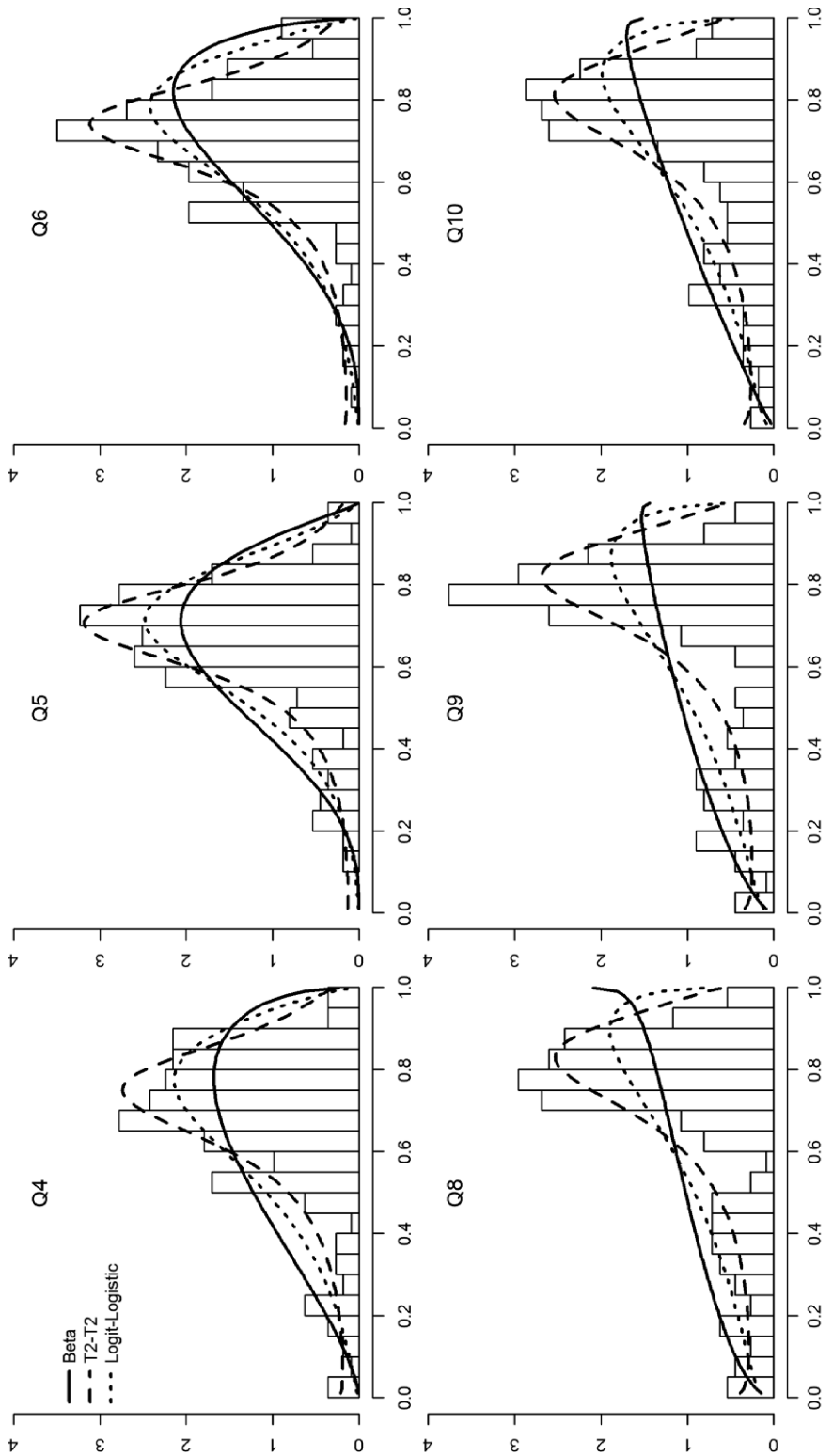

Figure 5. Histograms of the Budescu *et al.* (2009) data, with fitted beta, logit-logistic, and T2–T2 distributions.

$$w(x) = -\frac{\sqrt{(1-2x)^2}}{\sqrt{2}\sqrt{(1-x)x}}, 0 \le x < \frac{1}{2},$$
$$w(x) = \frac{\sqrt{(1-2x)^2}}{\sqrt{2}\sqrt{(1-x)x}}, \frac{1}{2} \le x \le 1; \tag{5.3}$$

and the pdf is

$$g(x, \mu, \sigma) = \sigma^2 / \left(1 + 2\mu\left(1-2x^2\right)\sqrt{2(1-x)x} + 2x(1-x)(\mu^2 + 2\sigma^2 - 2)\right)^{3/2}. \tag{5.4}$$

We focus on six of the 13 sentences, namely those containing the probability expressions 'likely' and 'unlikely' (Q4–Q6 and Q8–Q10). We also limit our analysis to the 'best' estimates given by the 223 subjects in their interpretations of 'likely' and 'unlikely' (thereby omitting their lower and upper estimates). To begin, we obtain maximum likelihood estimates for the three distributions fitted to this subset of the data. The parameter estimates, standard errors, 25th, 50th, and 75th estimated quantiles, along with the negative log-likelihoods for the fits, are displayed in Table 2. In all cases, the T2–T2 distributions fit the data better than the beta or logit-logistic distributions, while the logit-logistic fits are better than those for the beta.

Figure 5 displays the fitted beta, logit-logistic, and T2–T2 distributions superimposed on histograms of the data whose total areas equal 1. Table 2 shows a consistent tendency for the T2–T2 models to place the $Q_{.25}$ and $Q_{.5}$ estimates higher than the logit-logistic and the beta models. That tendency and these graphs clarify the main reason for the superior fit of the T2–T2 to the data. The distributions not only are skewed but also have a heavy left tail combined with an inflection close to 0.5, so a member of the finite-tailed (FT) group of distributions such as T2–T2 captures them fairly well.

We recall that a key finding by Smithson *et al.* (2012) was that the means of the 'unlikely' distributions were more regressive than the means of the 'likely' distributions. The T2–T2 estimates in Table 3 appear to be suggesting that the opposite is true for the medians. We now model the entire data set (treating the observations as independent for the time being), with 'likely' versus 'unlikely' as a predictor. The location and dispersion submodels are

$$\begin{aligned} \hat{\mu}_i &= \hat{\beta}_0 + \hat{\beta}_1 w_i, \\ \log(\hat{\sigma}_i) &= \hat{\delta}_0 + \delta_1 w_i \end{aligned}, \tag{5.5}$$

where $w_i = 1$ if the phrase is 'likely' and 0 if it is 'unlikely'.

Table 3 displays the parameter estimates, standard errors, and 95% confidence intervals for the beta, Kumaraswamy, and several members of the cdf-quantile family, including the logit-logistic and T2–T2 as before, along with their respective negative log-likelihoods. Table 4 displays the 25th, 50th, and 75th estimated quantiles. The sample quantiles are shown in the first pair of rows of the table, and the quantile estimates from the beta and Kumaraswamy distributions in the second and third pairs of rows. The Supporting Information subsection on Example 1 shows how to obtain

Table 4. Parameter estimates, quantiles, and negative log-likelihoods for various distributions

Distribution	Item	$Q.25$	$Q.5$	$Q.75$
Empirical quantiles	Likely	0.590	0.699	0.789
	Unlikely	0.470	0.749	0.829
Beta	Likely	0.432	0.736	0.928
	Unlikely	0.313	0.726	0.956
Kumaraswamy	Likely	0.743	0.811	0.865
	Unlikely	0.436	0.658	0.839
Logit-logistic	Likely	0.557	0.696	0.806
	Unlikely	0.509	0.710	0.853
T2–T2	Likely	0.584	0.699	0.785
	Unlikely	0.581	0.746	0.847
Arcsinh-Cauchy	Likely	0.568	0.691	0.773
	Unlikely	0.590	0.754	0.835
T2–Burr VII	Likely	0.593	0.700	0.788
	Unlikely	0.601	0.747	0.853
T2–Burr VIII	Likely	0.591	0.700	0.787
	Unlikely	0.596	0.747	0.85
Burr VIII-Burr VII	Likely	0.569	0.697	0.801
	Unlikely	0.542	0.726	0.855

conditional estimates of these quantiles from the regression coefficients, using the T2–T2 model for illustration. Briefly, the coefficients in the location submodel have a qualitatively similar interpretation to traditional regression coefficients; for example, a positive coefficient indicates that higher values of the covariate predict higher values for the median and, *ceteris paribus*, for other quantiles. A positive coefficient in the dispersion submodel implies that higher values of the covariate predict greater spread among the quantiles.

Turning now to the models, the beta and Kumaraswamy models both indicate that the median is higher for the 'likely' than for the 'unlikely' sentences, while the logit-logistic model suggests that they may not differ (the location submodel coefficient is not significant). However, as suggested by the T2–T2 results in Table 3, several other cdf-quantile distribution models with good fits in Table 4 have significant negative coefficients in the location submodel, indicating that the median for the 'likely' sentences is lower, which agrees with the empirical quantiles themselves. Moreover, the T2–T2 and similar models' estimates agree closely with the empirical medians. We note, however, that the interquartile range is greater in the 'unlikely' data and in the distribution models (as indicated by the negative coefficients in the dispersion submodel), which agrees with the Smithson et al. finding that the variability of responses is greater for negative probability phrases.

When we include other examples of distributions that fit the data better than the beta, Kumaraswamy, and logit-logistic models, and compare them with the empirical quantiles, their strong log-likelihood values in Table 3 suggest that the T2–T2 result is not a fluke. Moreover, referring back to the question of how to select distributions to fit data, it should be clear from the histograms in Figure 5 and the tail behaviours distinguishing the four subfamilies that the BM distributions would not be good candidates for these data, because their densities have limits of 0 at both boundaries of the unit interval. This is borne out when we fit BM distributions. For example, the model log-likelihood for the logit-T2 distribution is only 7.5 and for the Burr VIII-Cauchy it is much worse, at −1347.4,

compared with log-likelihoods of 265 to 435.3 for the eight distributions in Table 3. For the same reason, from the published distributions the simplex also would be a poor choice for these data. Indeed, according to the simplexreg R package (Zhang, Qiu & Shi, 2016), its model log-likelihood is −1,755.

6. Conclusion and prospects

We have introduced a family of two-parameter distributions that have several useful properties, as outlined in the Introduction, and provided demonstrations of their applicability to psychology and their utility in modelling real data. The variety of shapes that they can model makes them a viable alternative to other parametric alternative such as mixtures of betas, both because they use fewer parameters than mixture models and because they have a probability-reweighting interpretation that mixture models do not. The advantages that this family offers over other two-parameter alternatives stem primarily from two of its characteristics. First, many of its members are tractable, with both explicit cdfs and quantile functions. Second, unlike other distributions with explicit cdfs and quantile functions such as the Kumaraswamy or those featured in Kotz and Van Dorp (2004), but in common with the beta, its members' parameters explicitly model location and dispersion separately. It thus combines several of the best properties of distributions for modelling random variates on the (0,1) interval, while substantially augmenting the set of such distributions.

Several relevant topics have not been fully addressed in this paper, but remain to be explored and developed in future research on these distributions. This concluding section briefly surveys these topics and their prospects.

As we have mentioned earlier, further investigations into model diagnostics and residuals are under way for GLMs using these distributions. Residuals and diagnostics have been problematic in beta regression, and several papers attempting to deal with these problems have appeared since 2004 (e.g., Espinheira, Ferrari & Cribari-Neto, 2008; Ferrari, Espinheira & Cribari-Neto, 2011). A source of difficulties is dealing with cases near the boundaries of the support, and therefore we anticipate at least some of these issues may arise for this distribution family.

In the same vein, we have omitted to present any conclusive material on estimating random-effects models, but anticipate producing a treatment of the topic to parallel the Verkuilen and Smithson (2012) paper. Likewise, we have not dealt with alternative estimation approaches to MLE. Our package does enable users to fit models via Bayesian Markov chain Monte Carlo, but we have not implemented other alternatives such as weighted least squares. We refer interested readers to Lemonte and Bazán (2016)'s comparisons of MLE and weighted least squares.

Several generalizations and extensions of the cdf-quantile family remain to be explored. Extensions within the two-parameter family include using non-identical domains for the parent functions, replacing the argument x in $G(x,\mu,\sigma)$ with another cdf or appropriate $S(x)$, and adding parameters to the parent distribution cdfs.

We limit discussion here to the question of adding a third parameter. The rationale behind adding a parameter is to expand the variety of shapes that can be modelled (e.g., a distribution whose median is 1/2 but whose pdf is asymmetric). There are at least four straightforward ways of generalizing the cdf-quantile family by adding a parameter.

First, raising any cdf or quantile function in the family to a positive exponent yields tractable and invertible three-parameter cdfs. However, this sacrifices some of the other

attractive properties of their two-parameter counterparts (e.g., modelling the median in one parameter).

Second, because we have used distributions with fixed parameters for F and H, thereby restricting this family to two parameters (μ,σ), a simple generalization is to add parameters that augment the cdfs whose standardized versions we have used here. This generalization is effective for the Burr (1942) distributions of Types II (logistic), VII, and VIII, whose support is $(-\infty,\infty)$, because all of them are invertible and therefore have explicit expressions for their quantiles. Lemonte and Bazán (2016) treatment does this for F but not H, and they do not systematically investigate how best to add such parameters.

Third, our treatment, as mentioned before, is a special case of the Aljarrah *et al.* (2014) family with $S(x)$ restricted to the uniform cdf. A simple generalization would be to employ a power-function distribution, with $S(x) = x^{\gamma}$, for $\gamma>0$, or any other monotonic $[0,1]\rightarrow[0,1]$ function.

Fourth, in some cases one may apply an appropriate one-parameter transformation such as the Libby and Novick (1982) three-parameter generalized beta distribution. For instance, the arcsinh-logistic distribution is generalizable in this way, and can yield a distribution whose median is 1/2 but whose pdf is asymmetric.

There is a natural extension of the GLM for the cdf-quantile family to handle random effects, following the approach in Verkuilen and Smithson (2012). When the support is $[-\infty,\infty]$, we may add random error terms onto appropriate coefficients in the submodel of μ. When the support is $[0,\infty]$, the submodel of μ is linear in the log-scale, so the error terms would be added onto the log-scale submodel. In principle we may do the same in the log-scale submodel of σ. An elaboration of these models is beyond the scope of this paper, but we provide one illustration in the Supporting Information by applying this extension to the location submodel in equation (5.1). It also has not escaped our notice that the cdf-quantile family is immediately extensible to multivariate distribution models via copulas, because of its explict cdf and quantile functions.

References

Aljarrah, M. A., Lee, C., & Famoye, F. (2014). On generating T-X family of distributions using quantile functions. *Journal of Statistical Distributions and Applications*, *1*(1), 2. doi:10.1186/2195-5832-1-2

Alzaatreh, A., Lee, C., & Famoye, F. (2013). A new method for generating families of continuous distributions. *Metron*, *71*(1), 63–79. doi:10.1007/s40300-013-0007-y

Barndorff-Nielsen, O., & Jørgensen, B. (1991). Some parametric models on the simplex. *Journal of Multivariate Analysis*, *39*, 106–116. doi:10.1016/0047-259X(91)90008-P

Bowley, A. L. (1917). *Elements of statistics*. London, UK: P. S. King & Son.

Budescu, D. V., Broomell, S., & Por, H. H. (2009). Improving communication of uncertainty in the reports of the Intergovernmental Panel on Climate Change. *Psychological Science*, *20*, 299–308. doi:10.1111/J.1467-9280.2009.02284.X

Burr, I. W. (1942). Cumulative frequency functions. *The Annals of Mathematical Statistics*, *13*(2), 215–232. doi:10.1214/aoms/1177731607

Cook, R. D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society, Series B (Methodological)*, *48*(2), 133–169.

Cordeiro, G. M., & de Castro, M. (2011). A new family of generalized distributions. *Journal of Statistical Computation and Simulation*, *81*, 883–898. doi:10.1080/00949650903530745

Espinheira, P. L., Ferrari, S. L., & Cribari-Neto, F. (2008). On beta regression residuals. *Journal of Applied Statistics*, *35*, 407–419. doi:10.1080/02664760701834931

Eugene, N., Lee, C., & Famoye, F. (2002). Beta-normal distribution and its applications. *Communications in Statistics – Theory and Methods*, *31*, 497–512. doi:10.1081/STA-120003130

Ferrari, S. L., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, *31*(7), 799–815. doi:10.1080/0266476042000214501

Ferrari, S. L., Espinheira, P. L., & Cribari-Neto, F. (2011). Diagnostic tools in beta regression with varying dispersion. *Statistica Neerlandica*, *65*, 337–351. doi:10.1111/j.1467-9574.2011.00488

Greenwood, J. A., Landwehr, J. M., Matalas, N. C., & Wallis, J. R. (1979). Probability weighted moments: Definition and relation to parameters of several distributions expressable in inverse form. *Water Resources Research*, *15*, 1049–1054. doi:10.1029/WR015i005p01049

Gupta, A. K., & Nadarajah, S. (2004). *Handbook of beta distribution and its applications*. Boca Raton, FL: CRC Press.

Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, *36*(1-2), 149–176. doi:10.1093/biomet/36.1-2.149

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1995). *Continuous univariate distributions* (Vol. 2). New York, NY: Wiley.

Jones, M. C. (2002). The complementary beta distribution. *Journal of Statistical Planning and Inference*, *104*, 329–337. doi:10.1016/S0378-3758(01)00260-9

Jones, M. C. (2009). Kumaraswamy's distribution: A beta-type distribution with some tractability advantages. *Statistical Methodology*, *6*(1), 70–81. doi:10.1016/j.stamet.2008.04.001

Kieschnick, R., & McCullough, B. D. (2003). Regression analysis of variates observed on (0, 1): Percentages, proportions and fractions. *Statistical Modelling*, *3*, 193–213. doi:10.1191/1471082X03st053oa

Kotz, S., & Van Dorp, J. R. (2004). *Beyond beta: Other continuous families of distributions with bounded support and applications*. Singapore: World Scientific.

Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random processes. *Journal of Hydrology*, *46*, 79–88. doi:10.1016/0022-1694(80)90036-0

Lawless, J. F. (2011). *Statistical models and methods for lifetime data*. Hoboken, NJ: John Wiley.

Lemonte, A. J., & Bazán, J. L. (2016). New class of Johnson SB distributions and its associated regression model for rates and proportions. *Biometrical Journal*, *58*, 727–746. doi:10.1002/bimj.201500030

Libby, D. L., & Novick, M. R. (1982). Multivariate generalized beta distributions with applications to utility assessment. *Journal of Educational Statistics*, *7*(4), 271–294. doi:10.3102/10769986007004271

Mitnik, P. A., & Baek, S. (2013). The Kumaraswamy distribution: Median-dispersion re-parameterizations for regression modelling and simulation-based estimation. *Statistical Papers*, *54*(1), 177–192. doi:10.1007/s00362-011-0417-y

Moberg, C., Alderling, M., & Meding, B. (2009). Hand eczema and quality of life: a population-based study. *British Journal of Dermatology*, *161*, 397–403. doi:10.1111/j.1365-2133.2009.09099.x

Moors, J. J. A. (1988). A quantile alternative for kurtosis. *The Statistician*, *37*(1), 25–32. doi:10.2307/2348376

Noël, Y., & Dauvier, B. (2007). A beta item response model for continuous bounded responses. *Applied Psychological Measurement*, *31*(1), 47–73. doi:10.1177/0146621605287691

Pawitan, Y. (2001). *In all likelihood: Statistical modelling and inference using likelihood*. Oxford, UK: Oxford University Press.

Rocha, A. V., & Simas, A. B. (2011). Influence diagnostics in a general class of beta regression models. *Test*, *20*(1), 95–119. doi:10.1007/s11749-010-0189-z

Rust, R. T., Simester, D., Brodie, R. J., & Nilikant, V. (1995). Model selection criteria: An investigation of relative accuracy, posterior probabilities, and combinations of criteria. *Management Science*, *41*, 322–333. doi:10.1287/mnsc.41.2.322

Shou, Y., & Smithson, M. (2016). *cdfquantreg:Quantileregressionforrandomvariableson theunitinterval*[*Computersoftwaremanual*]. Retrieved from https://CRAN.R-project.org/package=cdfquantreg (R package version 1.0.4).

Smithson, M., & Merkle, E. C. (2014). *Generalized linear models for categorical and continuous limited dependent variables*. Boca Raton, FL: Chapman & Hall/CRC Press.

Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with betadistributed dependent variables. *Psychological Methods*, *11*(1), 54–71. doi:10.1037/1082-989X.11.1.54

Smithson, M., Budescu, D. V., Broomell, S. B., & Por, H. H. (2012). Never say 'not': Impact of negative wording in probability phrases on imprecise probability judgments. *International Journal of Approximate Reasoning*, *53*, 1262–1270. doi:10.1016/j.ijar.2012.06.019

Tadikamalla, P., & Johnson, N. L. (1982). Systems of frequency curves generated by transformations of logistic variables. *Biometrika*, *69*, 461–465. doi:10.1093/biomet/69.2.461

Townsend, J. T., & Colonius, H. (2005). Variability of the MAX and MIN statistic: A theory of the quantile spread as a function of sample size. *Psychometrika*, *70*, 759–772. doi:10.1007/s11336-001-0916-1

Verkuilen, J., & Smithson, M. (2012). Mixed and mixture regression models for continuous bounded responses using the beta distribution. *Journal of Educational and Behavioral Statistics*, *37*, 82–113. doi:10.3102/1076998610396895

Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods*, *17*, 228–243. doi:10.1037/a0027127

Zhang, P., Qiu, Z., & Shi, C. (2016). simplexreg: An R package for regression analysis of proportional data using the simplex distribution. *Journal of Statistical Software*, *71*(11), 1–21. doi:10.18637/jss.v071.i11

Zou, K. H., Carlsson, M. O., & Quinn, S. A. (2010). Beta-mapping and beta-regression for changes of ordinalrating measurements on Likert scales: A comparison of the change scores among multiple treatment groups. *Statistics in Medicine*, *29*, 2486–2500. doi:10.1002/sim.4012

Received 13 April 2016; revised version received 25 October 2016

Supporting Information

The following supporting information may be found in the online edition of the article:

Data S1. CDFquantile supplementary materials.