New Distributions for Modeling Subjective Lower and Upper Probabilities

Michael Smithson

The Australian National University

9-14 July 2017 /SIPTA 2017

Modeling lower and upper subjective probability judgment data

This is an investigation of a new approach to modeling lower and upper probability judgments. We will cover the following topics:

- Conjugate lower and upper probability distributions on the unit interval
- The “CDF-Quantile” distribution family and its extension for providing lower-upper distributions
- An example application to modeling data
- Advantages and drawbacks of this approach

Traditional methods

The most commonly employed methods for modeling lower-upper probability assignments are:

- A "repeated-measures" subject-effect parameter or a covariance parameter
- A regression-style model with a dummy variable to distinguish between lower and upper probabilities
- A regression-style model with a dummy variable and exponentiated coefficient to respect the order

None of these test specific models– e.g., none of them test a model incorporating “coherence” in any sense of the term.

Basics

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 $p_U(A) = 1 - W(1 - p(A), \theta)$.

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. Define $G_D(x, \theta) = 1 - G(1 - x, \theta)$, which clearly also is a CDF. G_D is the *conjugate dual* of G, which follows by observing that

$$1 - G_D(1 - x, \theta) = 1 - [1 - G(1 - (1 - x), \theta)] = G(x, \theta) \qquad (1)$$

Basics

One- and two-parameter distributions of the kinds illustrated here have very limited flexibility regarding the location of G and G_D; typically the 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 have such applications when applied to the family of distributions introduced in the next section.

CDF-Quantile family

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:

$$G(x, \mu, \sigma) = F[U(H^{-1}(x), \mu, \sigma)] \tag{2}$$

where F is a CDF with support denoted 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 incorporating parameters μ and σ. We limit the domains to $D_1 = (-\infty, \infty)$ and $D_2 = (-\infty, \infty)$ and put

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

Quantile function

If F is invertible, then for every γ such that $G(x, \mu, \sigma) = \gamma$, the quantile function is: For $D_1 = (-\infty, \infty)$ and $D_2 = (-\infty, \infty)$ we put

$$G^{-1}(\gamma, \mu, \sigma) = H\left[\sigma F^{-1}(\gamma) + \mu\right]. \tag{4}$$

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.

Properties

Smithson and Shou (2017) 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. Members are self-dual in the sense that $g(x, \mu, \sigma) = g(1 - x, -\mu, \sigma)$. Moreover, $G = G_D$, so the conjugate-CDF duals in this family consists of identical distributions.
3. The median is solely a function of μ, so that μ is genuinely a location parameter.
4. The parameter σ is a dispersion parameter.
5. 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.

Introducing a third parameter to the CDF-Quantile Family

Marshall and Olkin (2007) state that the class **G** of CDFs G whose support is (0,1) form an algebraic group, which is closed under composition. This is true of continuous CDFs. Applying an invertible $(0,1) \rightarrow (0,1)$ transformation W to the innermost level of the CDF, for instance, we have

$$G(x, \mu, \sigma, \theta) = F\left[U\left(H^{-1}\left(W\left(x, \theta\right)\right), \mu, \sigma\right)\right] \tag{5}$$

and

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

Introducing a third parameter to the CDF-Quantile Family

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

$$G_D(x,\mu,\sigma,\theta) = F\left[U\left(H^{-1}\left(1 - W(1-x,\theta)\right),\mu,\sigma\right)\right]. \tag{7}$$

Conjugate lower-upper distributions

Proposition 5.1: Let $W(x, \theta)$ be defined as earlier, so that it behaves as a CDF. Let

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

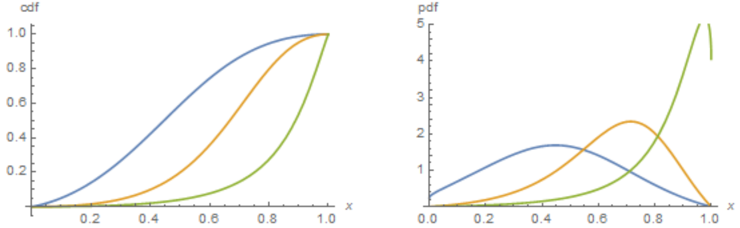
Then if the CDFs F and H satisfy certain symmetry conditions,

$$1 - G(1 - x, -\mu, \sigma, \theta) = G_D(x, \mu, \sigma, \theta), \tag{8}$$

and the quantiles $G^{-1}(\gamma, \mu, \sigma, \theta)$ and $G_D^{-1}(\gamma, \mu, \sigma, \theta)$ behave as conjugate lower-upper probabilities.

Properties of the conjugate pairs

The conjugate dual CDFs straddle the CDF $G(x, \mu, \sigma)$ and the resultant lower and upper quantile functions straddle the quantile function $G^{-1}(\gamma, \mu, \sigma)$.



Data-set

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. (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.

Example: “The Greenland ice sheet and other Arctic ice fields likely contributed no more than 4 m of the observed sea level rise.” What probability did the authors mean by “likely”?

Model comparisons

I fitted 11 models 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.

1. Model 1 is just the two-parameter distribution, as defined in equation (3): $\hat{\mu} = \beta_0$ and $\hat{\sigma} = \exp(\delta_0)$.
2. Model 2: $\hat{\mu} = \beta_0 + \beta_1 x$ and $\hat{\sigma} = \exp(\delta_0 + \delta_1 x)$, where $x = 0$ for lower probabilities and $x = 1$ for upper probabilities.
3. Model 3 estimates the dependency between the lower and upper estimates via a t-copula with Model 2 margins.

Model comparisons

The remaining models are based on the three-parameter CDF-Quantile distribution.

4. Model 4 has intercept-only submodels $\hat{\mu} = \beta_0$, $\hat{\sigma} = \exp(\delta_0)$, and $\hat{\theta} = \exp(\gamma_0)$.
5. Model 5 is the conjugate-dual model, as defined in equations (5) and (7), with Model 4 parameters plus a $0, 1$ mixture parameter applying G to the upper and G_D to the lower probabilities.
6. Model 6 has $\hat{\mu} = \beta_0 + \beta_1 x$ and $\hat{\sigma} = \exp(\delta_0 + \delta_1 x)$ with $x = 0$ and 1 for lower and upper probabilities, but $\hat{\theta} = \exp(\gamma_0)$.
7. Model 7 has Model 6 μ and σ plus $\hat{\theta} = \exp(\gamma_0 + \gamma_1 x)$.
8. Models 8-11 are based on a “tilt-parameter” version of the three-parameter CDF-Quantile distribution (omitted due to space limitations).

Results

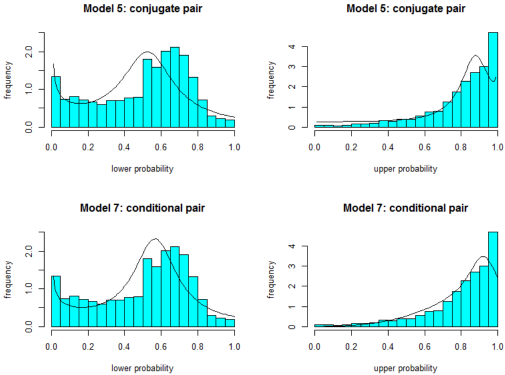
The best fitting distribution is the Cauchit-Cauchy, so all 11 models are based on this distribution.

Table : 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

Results

The figure below shows histograms of the lower and upper probabilities with the fitted distributions from Model 5 (the conjugate-dual model) and Model 7 (the 6-parameter conditional exponentiated model).



Prospects and problems

1. The 2-parameter CDF-Quantile family is available for generalized linear modeling via the `cdfquantreg` package in R and a SAS macro, and Smithson and Shgou (2017) also have shown that these distributions can model data better than other two-parameter distributions such as the beta.
2. The 3-parameter conjugate-dual CDF-Quantile family may be of theoretical interest, has potential for applications, and can test specific models of lower-upper probability judgments.
3. High correlations between parameter estimates may be a pervasive problem for three-parameter distributions on the unit interval, but the conjugate-dual model does not seem to suffer from this.
4. Much remains to be developed and explored regarding parameter estimation methods and model diagnostics, even for the two-parameter CDF-Quantile family.

The End

Thanks!