Published in final edited form as:

Am Stat. 2018 ; 72(2): 121–129. doi:10.1080/00031305.2016.1213661.

A Cautionary Note on Beta Families of Distributions and the Aliases Within

Alan D. Hutson, Albert Vexler

Alan D. Hutson is Professor of Oncology, Roswell Park Cancer Institute, Department of Biostatistics & Bioinformatics, Elm & Carlton Streets, Buffalo, NY 14623. Albert Vexler is Associate Professor, University at Buffalo, Department of Biostatistics, 706 Kimball Tower, 3435 Main St., Buffalo, NY 14214-3000.

Abstract

In this note we examine the four parameter beta family of distributions in the context of the beta-normal and beta-logistic distributions. In the process we highlight the concept of numerical and limiting alias distributions, which in turn relate to numerical instabilities in the numerical maximum likelihood fitting routines for these families of distributions. We conjecture that the numerical issues pertaining to fitting these multiparameter distributions may be more widespread than has originally been reported across several families of distributions.

Keywords

numerical methods; maximum likelihood estimation; parametric models

1 Introduction

There has been a long standing tradition in mathematical statistics of extending classic univariate two parameter families of distributions by adding a 3rd, 4th or even 5th parameter, e.g. the exponentiated Weibull and the epsilon-skew-normal distributions of Mudholkar and Hutson (1996,2000) are oft-cited distributions defined by adding a 3rd parameter to the well-known Weibull and normal distributions, respectively. There are literally hundreds of examples of such papers that are too numerous to cite in one article. The general goal to this approach of extending classical models is to add additional parameters, similar to a nonparametric kernel smoothing type model, while retaining some of the advantages of the parametric approach, e.g. direct inference about the data through the parameters or the ability to link a given parameter to a set of covariates via a regression type framework, e.g. see Hutson (2004). Oftentimes, parametric model fit is tested via additional parameters given the model of interest is a subclass of the family, e.g. testing a two parameter Weibull fit as a submodel of the three parameter exponentiated Weibull family, e.g. see Mudholkar and Hutson (1996). The proliferation of these 4 and 5 parameter families of distributions is most likely due to the rapid expansion of easy-to-use statistical maximization software routines. The downside to these types of models are the general subtleties of the underlying

numerical methods, such as picking appropriate starting values in the fitting routine relative to non-identifiability issues (in the numerical sense), as well as the sometimes high degree of interdependence of the maximum likelihood estimates relative to large sample approximations for inferential procedures given a finite sample setting.

We turn our focus on one very oft-cited approach for extending classical models through the incorporation of two additional parameters, which is due in part to the work of M.C. Jones (2004), and consists of embedding the cumulative distribution function (c.d.f.) of interest, denoted F , within the beta c.d.f., say B , described in more detail below. There are strong parallels to this approach to the single order statistic distribution functional form. Within the *Test* volume that contains the Jones (2004) article there is a discussion and rebuttal section relative to the merits of this approach for generating new classes of distributions. The paper also includes several important historical references to prior extensions to classical distributions and serves as good history of this type of approach. Even though we focus on this particular family of distributions this note should serve as a cautionary alarm relative to the general methodology of adding additional parameters to parametric families.

The beta family of distributions, e.g. see Jones (2004), arose from the work of Eugene *et al.* (2002) who first introduced the specific case of the beta-normal distribution. As background relative to defining specific distributions within this family first let $z = (x - \mu)/\sigma$,

$$F(x) = F_0(z), \quad (1.1)$$

$$f(x) = \frac{1}{\sigma} f_0(z), \quad (1.2)$$

$$Q(u) = \mu + \sigma Q_0(u), \quad (1.3)$$

where μ and σ are the location and scale parameters. The beta-normal and beta-logistic symmetric location-scale base families of distributions, are defined as per Table 1. If we let $B_{p,q}(x)$ denote the beta c.d.f., $b_{p,q}(x)$ denote the beta p.d.f. and $B_{p,q}^{-1}(u)$ denote the beta quantile function then the corresponding c.d.f., p.d.f. and quantile function for the beta family of distributions take the form

$$G(x) = B_{p,q}[F_0(z)], \quad (1.4)$$

$$g(x) = \frac{1}{\sigma} b_{p,q}[F_0(z)] f_0(z), \quad (1.5)$$

$$P(u) = \mu + \sigma Q_0[B_{p,q}^{-1}(u)], \quad (1.6)$$

respectively, where the form of the beta p.d.f. used in this note is given as

$$b_{p,q}(t) = \frac{1}{\beta(p,q)} t^{p-1} (1-t)^{q-1} \text{ and } \beta(p,q) \text{ is the normalizing constant. The c.d.f. and}$$

quantile function follow suit. The results in this note extend to several other symmetric base distributions, but for the purpose of illustration we focus on the beta-normal and beta-logistic distributions.

A motivation for this note pertains to our own empirical observations and to two passages, one from Jones (2004) and one from Eugene *et al.* (2002), respectively. Jones states that “Incorporation of four-parameter families of distributions in modelling exercises automatically allows for the effects of skewness and (heavy) tail weights. Of course, for small to moderate sample sizes, there will typically be insufficient information in the data to estimate the parameters a and b [in our notation p and q] well (very large sample sizes will normally be necessary for that). But also this will not usually be very important. Interest would normally be in location and perhaps scale parameters” and Eugene *et al.* (2002) who state the following with respect to beta-normal fitting routines: “A simple procedure for the choice of initial values $(\alpha, \beta, \mu, \sigma)$ [in our notation p, q, μ, σ] is as follows: Evaluate the likelihood at the points $(\alpha, \beta, \mu, \sigma) = (1, 1, \bar{x}, s)$, $(\alpha, \beta, \mu, \sigma) = (0.1, 0.1, \bar{x}, s)$ and start the iteration from the point that has higher likelihood. Our experience indicates that using this technique is sufficiently fast. For most of the data sets, the Newton Raphson algorithm converged in less than 20 iterations.”

We originally noticed numerical fitting issues for values of p and q near 1 in terms of the numerical consistency of the parameter estimates given a symmetric base distribution such as normal or logistic for substantially large sample sizes. This is of particular importance if one is looking to reduce the model via the test $H_0 : p = 1, q = 1$ or via confidence intervals about p and q . Heuristically, if one looks at the form of the p.d.f. $g(x)$ at (1.5) there are two components of scale at play: 1) the standard scale parameter σ and 2) the beta normalizing constant $\beta(p, q)$. For near symmetric distributions this lends itself to somewhat of an identifiability issue in terms of the numerical fitting routines relative to scale, which we will illustrate. A general fix to this problem is to constrain the model such that σ is a function of p and/or q with the same relative goodness-of-fit. This appears to be a non-issue for asymmetric base distributions such as the Weibull distribution, which extends to the beta-Weibull distribution, e.g. see Famoye *et al.* (2005) who introduced the distribution and Cordeiro *et al.* (2013) who provide additional mathematical results pertaining to the beta-Weibull distribution.

The other interesting issue, which arose while investigating the beta family of distributions is the concept of distributional aliases from a distance perspective, which we outline in the next section. For a given base distribution parameterization, say standard normal, we found that there were sets of vastly different p and q beta values for which the two distributions were virtually identical from a numerical standpoint, i.e. no matter what the sample size we could find pairs of distributions that were so close in terms of having virtually the same probability measure that a typical fitting routine would not be able to distinguish between the two sets of parameters, not even for massively large sample sizes. In fact, in the next section we prove the counterintuitive result that if we let $p = q$ and σ go to infinity jointly at a certain rate the limiting distribution is the normal distribution within the beta-normal family of distributions.

The issues outlined above are quite relevant in the context of testing for subclasses of distributions within the beta family of distributions, e.g. testing for normality as a submodel of the beta-normal distribution. In addition, as we will illustrate theoretically, large sample variance approximations of maximum likelihood estimators are vastly different between a given distribution and its respective alias, even when the actual density functions are virtually identical. This in turn impacts inferential procedures around such quantities as $E(X)$ ($= \mu$ in the symmetric case). More recently the beta-normal has been incorporated into other research fields such as psychometric testing, crop yield and risk assessment, e.g. see Ranger and Kuhn (2015), Hennessy (2009) and Razzaghi (2009), respectively, without regards to the issues outlined above.

In Section 2 we provide an approach to determining aliases in the beta family of distributions and provide some relevant limiting results. In Section 3 we provide the general form for the likelihood function and its related quantities in terms of a general beta family of distributions. A small simulation study is carried forth to demonstrate the numerical instabilities relative to maximum likelihood estimation between a given distribution and its alias at both moderate to large samples. We then provide some concluding remarks.

2 Aliases

In this section we illustrate how for a fixed set of parameter values for a given beta-normal or beta-logistic distribution there are alias distributions that are virtually identical relative to minimizing the Kulback-Liebler distance

$$D = \int_{-\infty}^{\infty} \log\left(\frac{f(x)}{g(x)}\right) f(x) dx. \quad (2.1)$$

The underlying reason why we can find two symmetrical distributions with very distinct parameter values was pointed out in the introduction in that the standard scale parameter σ and the beta-normalizing constant $\beta(p, q)$ are in a sense both scale factors. Interestingly, the major problem here appears to occur with respect to the most important case of choosing between a normal model and a beta-normal model.

As an example, let us set $\mu = 0$, $\sigma = 1$ and $p = q = 1$ for the beta-normal distribution, i.e. the standard normal density. We can arbitrarily set $\mu = 0$ and $p = q = 4$ and minimize the Kulback-Liebler distance at (2.1) with respect to σ . In this scenario we arrive at the alias setting $\sigma = 2.180$ with $D = 0.000015$. If we set $\mu = 0$ and $p = q = 10$ we arrive at the alias to the standard normal distribution by setting $\sigma = 3.518$ with $D = 0.0000034$, even smaller than before, i.e. as $p = q$ move away from 1 the distance D between the standard normal and the beta-normal distributions decreases. The three density functions are plotted together in Figure 1. This observation led to a counterintuitive mathematical result. In fact we can show that in a limiting sense we can prove the convergence of the beta-normal to standard normal if we let $p = q$ go to ∞ in a specific fashion.

Towards this end, and without loss of generality with respect to the parameter μ , let p.d.f. $f(x)$ correspond to the beta-normal density function with $\mu = 0$, $\sigma = 1$ and $p = q = 1$ and let

the p.d.f. $g(x)$ correspond to the symmetric beta-normal density function with $\mu = 0$, $\sigma > 0$ and $p = q > 0$. An interesting limiting result occurs if for p.d.f. g we let $p = q$ (symmetric beta-normal) and link σ to p as follows:

$$\sigma = \frac{4^{1-p}}{\beta(p, p)} = \frac{2\Gamma\left(p + \frac{1}{2}\right)}{\sqrt{\pi}\Gamma(p)}. \quad (2.2)$$

This is the point where the value for σ at (2.2) is given by solving $f(0) = g(0)$.

If we then substitute σ at (2.2) into g then the p.d.f. g is a symmetric beta-normal density function with $\mu = 0$, $\sigma = \frac{4^{1-p}}{\beta(p, p)} = \frac{2\Gamma\left(p + \frac{1}{2}\right)}{\sqrt{\pi}\Gamma(p)}$ and $q = p$ such that g has the single parameter p . Then the

$$\lim_{p \rightarrow \infty} \log\left(\frac{f(x)}{g(x)}\right) = 0, \quad (2.3)$$

where f denotes the standard normal density function with $\mu = 0$, $\sigma = 1$ and $p = q = 1$. Hence the Kulback-Liebler distance go to 0 for extremely divergent values of $p = q$ different from 1. This results holds for general and equivalent $\mu = 0$ values for the classic normal densities versus symmetric beta normal densities.

The result at (2.3) follows by noting that

$$\log\left(\frac{f(x)}{g(x)}\right) = -\frac{x^2}{2} + \frac{x^2}{8}\pi\gamma(p)^2 + (1-p)\log\left[1 - \text{Erf}\left(x\frac{\sqrt{2\pi}}{4}\gamma(p)\right)^2\right], \quad (2.4)$$

where $\gamma(p) = \Gamma(p)/\Gamma(p + 1/2)$ and $\text{Erf}(x) = 2\int_0^x \exp(-t^2)dt/\sqrt{\pi}$ is the error function. The rest of the steps in the proof are given in Appendix A. In this case we present in Appendix A a detailed proof of that $D = O(p^{-1}) \rightarrow 0$, as $p \rightarrow \infty$, where D is defined in (2.1).

For the more general symmetric beta-normal case let $\mu = 0$, $\sigma = 1$ and $p_0 = q_0$ be the beta parameters for f for a beta-normal comparison and let g have parameterization $\mu = 0$, $\sigma = 1$ and $p = q$ then

$$\sigma = \frac{4^{p_0-p}\beta(p_0, p_0)}{\beta(p, p)} = \frac{\Gamma\left(p + \frac{1}{2}\right)\Gamma(p_0)}{\Gamma\left(p_0 + \frac{1}{2}\right)\Gamma(p)}. \quad (2.5)$$

Using similar arguments to that of the standard normal case for fixed x we arrive at

$$\log\left(\frac{f(x)}{g(x)}\right) = \frac{4x^2\Gamma\left(\frac{1}{2} + p_0\right)^2 - \pi\Gamma(p_0)^2\left(x^2 - 2(p_0 - 1)\log\left(1 - \operatorname{Erf}\left(\frac{x}{\sqrt{2}}\right)\right)\right)}{2\pi\Gamma(p_0)^2} \quad (2.6)$$

$$+ O(p^{-1}),$$

as $p \rightarrow \infty$. See Appendix A for remarks pertaining to this result. This results holds for general and equivalent $\mu = 0$ values for symmetric beta-normal density functions versus asymmetric beta normal density functions. We can see from (2.6) this limit does not converge to 0. However, for any given fixed value of x the limit gets quite small for large values of p different from $p_0 = q_0$, which has implications in the numerical fitting process, which we describe in the next section.

For the more general symmetric beta-logistic case let $\mu = 0$, $\sigma = 1$ and $p_0 = q_0$ be the beta parameters for f for a beta-logistic comparison and let g have parameterization $\mu = 0$, $\sigma = 1$ and $p = q$. Using similar arguments to that of the normal case for fixed x we arrive at the limiting result between two symmetric beta-logistic density functions at fixed value for x in the form

$$\log\left(\frac{f(x)}{g(x)}\right) = \frac{x^2\Gamma\left(\frac{1}{2} + p_0\right)^2}{4\Gamma(p_0)^2} + p_0(x + \log(4) - 2\log[1 + \exp(x)]) + O(p^{-1}) \quad (2.7)$$

See Appendix A for details. This results holds for general and equivalent $\mu = 0$ values for symmetric beta-logistic distributions versus symmetric beta normal distributions. The standard logistic comparison to the symmetric beta-logistic distribution does not produce the same interesting limiting argument as was presented in the beta-normal case at $p_0 = 1$.

Numerically we can show similar behavior in the asymmetrical cases. For example if we take a beta-normal density function f with $\mu = 0$, $\sigma = 1$ and $p = 2$ and $q = 4$ and minimize the Kulback-Liebler distance with respect to σ and q and setting $\mu = 0$ and $p = 10$ we get for values of $\sigma = 2.147$ and $q = 14.288$ that the distance from (2.3) is 0.00005. The plots of the two density functions are overlaid in Figure 2. In the next section we illustrate the numerical estimation issues caused by the alias phenomena.

3 Maximum Likelihood Estimation

For issues related to maximum likelihood related to the beta family of distributions let us define the vector of the parameter of interest as $\theta = (\mu, \sigma, p, q)$ and let $z_i = (x_i - \mu)/\sigma$. For an i.i.d. observed sample of size n , $x_1, x_2, \dots, x_n$ the log-likelihood takes the form

$$l_n(\boldsymbol{\theta}) = \sum_{i=1}^n l(x_i, \boldsymbol{\theta}) \quad (3.1)$$

$$= \sum_{i=1}^n \left\{ \log b_{p,q}[F_0(z_i)] + \log f_0(z_i)/\sigma \right\}, \quad (3.2)$$

where

$$\log b_{p,q}[F_0(z_i)] = \log \Gamma(p+q) - \log \Gamma(p) - \log \Gamma(q) + (p-1) \log F_0(z_i) + (q-1) \log [1 - F_0(z_i)]. \quad (3.3)$$

The scores per observation for each component of $\boldsymbol{\theta}$, given as $s_j(x_i) = \partial l(x_i, \boldsymbol{\theta}) / \partial \theta_j$ are provided in Appendix B in equations (B.1–B.4). Similarly the sample information per observation for each component of $\boldsymbol{\theta}$, given as $i_{kj}(x_i) = -\partial^2 l(x_i, \boldsymbol{\theta}) / \partial \theta_k \partial \theta_j$ are provided in the Appendix B in equations (B.5–B.14).

In Section 2 we illustrated that for the beta-normal and beta-logistic distribution we can find two alias distributions with very divergent values for p and q such that the densities are virtually identical. However, in terms of maximum likelihood estimation it is clear that the information matrix will be quite different for two alias distributions. Note that in the general form the sample information per observation for p and q , i_{33} and i_{44} at (B.12) and (B.14), respectively is equivalent to the expected information. As p and q go to infinity these elements tend towards 0 in the limit. The implication being that for alias families the corresponding numerical estimation procedures will become quite unstable in terms of the numerical inversion of the sample information matrix. For example, for true values of $\mu = 0$, $\sigma = 1$ and $p = q = 1$ we have $i_{33} = i_{44} = 1$. For the alias distribution with $\mu = 0$, $\sigma = 3.518$ and $p = q = 10$ we have $i_{33} = i_{44} = 0.053$.

In terms of inference about the parameters, two alias models will be indistinguishable in terms of numerical model fitting procedures and the actual sample data values, yet inferential quantities about the parameters, e.g. large sample confidence intervals will be vastly different. These numerical issues are compounded for beta family based regression models around the standard parameterization $\mu = X\beta$. Similarly, the beta-normal family is used to test for subclasses of models, e.g. testing normality via testing the hypothesis $H_0: p = 1, q = 1$. These tests will also be unstable given the numerical issues outlined above.

As an illustration of the numerical instability of these models, in terms of maximum likelihood estimation, we could provide countless examples. We will take a few basic examples to illustrate our point. Let us take our three alias cases from the previous section for the beta-normal distribution with case 1 setting $\mu = 0$, $\sigma = 1$ and $p = q = 1$, case 2 setting $\mu = 0$, $\sigma = 2.18$ and $p = q = 4$ and case 3 setting $\mu = 0$, $\sigma = 3.518$ and $p = q = 10$. The plots of the three densities are given in Figure 1. We use a variation of the starting points suggested by Eugene *et al.* (2002) and use the true values for μ and σ for the choice of initial values (p

$= \alpha, q = \beta, \mu, \sigma$) as follows: Evaluate the likelihood at the points $(p, q, \mu, \sigma) = (1, 1, 0, 1)$, $(p, q, \mu, \sigma) = (0.1, 0.1, 0, 1)$ and start the iteration with the highest likelihood.

In Table 2 we utilize Eugene's starting values and provide the quartiles of the maximum likelihood values from 1000 model fits for sample sizes $n = 50$ and $n = 500$. As can be seen the fitted maximum likelihood estimates are vastly different than the true parameter value with the exception of μ in a few cases. Having a large sample size did not improve the instabilities. We repeated this experiment using the actual true parameter values as the starting values. The results are presented in Table 3. As expected, the results are similarly poor. The quartiles themselves were used over the mean since oftentimes there were extreme values for $\hat{p}$ and $\hat{q}$, which make sense given consideration our limiting result given in the previous section.

We repeated our simulation study for an asymmetric case given as case 4 with $\mu = 0, \sigma = 1, p = 2, q = 4$ with numerical alias given as case 5 $\mu = 0, \sigma = 3.518, p = 10, q = 14.28$. Figure 2 provides plots of the overlay of the two distributions. As can be seen in Tables 4 we have similar and dramatic instabilities in the model fitting, which are comparable to the symmetric case.

4 Conclusions

We have given some theoretical and simulation based results to illustrate the beta family of distributions may in fact be an overparameterized way of modeling data relative to the numerical closeness of given distributions and their respective aliases within these families. Mathematically these models make logical sense, however, when combined with numerical fitting issues there is somewhat of an identifiability issue relative to scale and shape parameters. The numerical maximum likelihood fitting issues are non-trivial in nature. Hence, we advise caution against fitting these models to real data relative to interpretation of the parameter estimates and/or any functions derived from these estimates. We conjecture that other 4 and 5 parameter parametric models may have similar underlying issues, e.g. the Kumaraswamy skew-normal distribution by Mameli (2015) or the beta skew-normal distribution by Mameli and Musio (2013). If one wishes to use the beta family of distributions we suggest linking the parameters p or q with the scale parameter σ in a meaningful way that still captures scale, skewness and tail-weight appropriately, but reduces the beta based model down to 3 parameters.

Acknowledgments

We wish to thank the AE and two reviewers who made important comments, which led to a vastly improved version of this paper. This work was supported by Roswell Park Cancer Institute and National Cancer Institute (NCI) grant P30CA016056

Appendix A: Asymptotic Evaluations

The asymptotic forms involve the ratio of gamma functions. The following general results evaluate the ratio of gamma functions.

Define the ratio $\gamma = \Gamma(p)/\Gamma(p + 1/2)$

Lemma 1. For $1/2 < p \in \mathfrak{R}$, we have

$$\gamma = p^{-1/2} \left[1 + \frac{1}{8p} + O(p^{-2}) \right] \text{ as } p \rightarrow \infty \quad (\text{A.1})$$

and

$$(p + 1/4)^{1/2} < \gamma^{-1} < \left[p - 1/2 + (3/4)^{1/2} \right]^{1/2}. \quad (\text{A.2})$$

Proof. The result (A.1) is presented in Tricomi and Erdélyi (1951). To obtain the inequality (A.2) one can use the Kershaw's Double Inequalities shown, e.g., in Qi (2010) (see, p. 32, where in the context of Qi (2010)'s notations we denote $x = p - 1/2 > 0$ and $s = 1/2$), see also Mortici (2010, p. 426).

Consider equation (2.4). Taking into account (A.1), we use the Taylor theorem with $\text{Erf}(x\sqrt{2\pi}\gamma/4)^2$ around zero to represent

$$\begin{aligned} \log \left[1 - \text{Erf} \left(x \frac{\sqrt{2\pi}}{4} \gamma \right)^2 \right] &= - \text{Erf} \left(x \frac{\sqrt{2\pi}}{4} \gamma \right)^2 - \frac{1}{2} \text{Erf} \left(x \frac{\sqrt{2\pi}}{4} \gamma \right)^4 - \frac{1}{3} \\ &\quad \text{Erf} \left(x \frac{\sqrt{2\pi}}{4} \gamma \right)^6 \left(1 - \varpi \text{Erf} \left(x \frac{\sqrt{2\pi}}{4} \gamma \right)^2 \right)^{-3}, \end{aligned} \quad (\text{A.3})$$

with $\varpi \in (0, 1)$, where the reminder term

$$\begin{aligned} 0 &\leq \frac{1}{3} \text{Erf} \left(x \frac{\sqrt{2\pi}}{4} \gamma \right)^6 \left(1 - \varpi \text{Erf} \left(x \frac{\sqrt{2\pi}}{4} \gamma \right)^2 \right)^{-3} \\ &\leq \frac{1}{3} \text{Erf} \left(x \frac{\sqrt{2\pi}}{4} \gamma \right)^6 \left(1 - \text{Erf} \left(x \frac{\sqrt{2\pi}}{4} \gamma \right)^2 \right)^{-3}. \end{aligned}$$

Here we use the inequality (A.2) that insures that, for arbitrary large p 's, we have

$\text{Erf} \left(x \frac{\sqrt{2\pi}}{4} \gamma \right) < 1/2 + c$, where $0 < c < 1/4$ is a fixed constant. Thus, it is clear that (A.3), the

Taylor expansion of $\text{Erf} \left(x \frac{\sqrt{2\pi}}{4} \gamma \right)$ with γ around zero and (A.1) provide that

$$\begin{aligned} \log \left[1 - \text{Erf} \left(x \frac{\sqrt{2\pi}}{4} \gamma \right)^2 \right] &= - \left[x \frac{1}{\sqrt{2}} \gamma - x^3 \frac{\pi\sqrt{2}}{48} \gamma^3 + O(\gamma^4) \right]^2 - \frac{1}{2} \left[x \frac{1}{\sqrt{2}} \gamma - x^3 \frac{\pi\sqrt{2}}{48} \gamma^3 + O(\gamma^4) \right]^4 = \\ &= - \left[x \frac{1}{\sqrt{2}} p^{-1/2} + x \frac{1}{8\sqrt{2}} p^{-3/2} - x^3 \frac{\pi\sqrt{2}}{48} p^{-3/2} + O(p^{-2}) \right]^2 - \frac{1}{2} \left[x \frac{1}{\sqrt{2}} p^{-1/2} \right]^4 + O(p^{-2}). \end{aligned}$$

This and (A.1) applied to equation (2.4) imply

$$\begin{aligned} \log \left(\frac{f(x)}{g(x)} \right) &= - \frac{x^2}{2} + \frac{x^2}{8} p^{-1/2} + \frac{1}{8} p^{-3/2} \left(1 - (1 - p) \right. \\ &\quad \left. \left[x \frac{1}{\sqrt{2}} p^{-1/2} + x \frac{1}{8\sqrt{2}} p^{-3/2} - x^3 \frac{\pi\sqrt{2}}{48} p^{-3/2} + O(p^{-2}) \right]^2 - \frac{(1 - p)}{2} \left[x \frac{1}{\sqrt{2}} p^{-1/2} \right]^4 + O(p^{-2}) \right). \end{aligned}$$

Then without too much algebra one finds

$$\log\left(\frac{f(x)}{g(x)}\right) = \frac{x^2(x^2 - 3)(3 - \pi)}{24p} + o(p^{-3/2})$$

, where as in the body of the text $f(x)$ denotes the standard normal distribution.

Consider the Kulback-Liebler measure (2.1) based on the form (2.4). In general we note that

$$D = \int \log\left(\frac{f(x)}{g(x)}\right) f(x) dx \geq 0, \quad (\text{A.4})$$

since

$$D = \int \log\left(\frac{f(x)}{g(x)}\right) f(x) dx = - \int \log\left(\frac{g(x)}{f(x)}\right) f(x) dx \geq - \log\left(\int \frac{g(x)}{f(x)} f(x) dx\right) = 0.$$

Taking into account (2.4), we have, for large values of $p > 1$,

$$(1 - p) \int \log\left[1 - \text{Erf}\left(x \frac{\sqrt{2\pi}}{4} \gamma\right)^2\right] f(x) dx \leq (p - 1) \int \text{Erf}\left(x \frac{\sqrt{2\pi}}{4} \gamma\right)^2 f(x) dx,$$

since, for all $0 < s < 1$, $\log(1 - s) \geq -s$. The Taylor theorem gives

$$\text{Erf}\left(x \frac{\sqrt{2\pi}}{4} \gamma\right) = x \frac{1}{\sqrt{2}} \gamma - x^3 \frac{\sqrt{2\pi}}{16} \gamma^3 \exp\left[-\left(x \frac{\sqrt{2\pi}}{4} \gamma\right)^2\right], \quad \gamma \in (0, 1).$$

That is

$$\begin{aligned} (1 - p) \int \log\left[1 - \text{Erf}\left(x \frac{\sqrt{2\pi}}{4} \gamma\right)^2\right] f(x) dx &\leq (p - 1) \int \left[x^2 \frac{1}{2} \gamma^2 + x^6 \frac{\pi^2}{128} \gamma^6 \exp\left[-2\left(x \frac{\sqrt{2\pi}}{4} \gamma\right)^2\right]\right] f(x) dx \leq \\ (p - 1) \int \left[x^2 \frac{1}{2} \gamma^2 + x^6 \frac{\pi^2}{128} \gamma^6\right] f(x) dx &= (p - 1) \frac{1}{2} \gamma^2 + (p - 1) O(\gamma^6). \end{aligned}$$

Thus, by virtue of (2.4), we have

$$\begin{aligned} D &\leq \int \left(-\frac{x^2}{2} + \frac{x^2}{8} \pi \gamma^2\right) f(x) dx + (p - 1) \frac{1}{2} \gamma^2 + (p - 1) O(\gamma^6) \\ &= -\frac{1}{2} + (p - 1) \frac{1}{2} \gamma^2 + O(\gamma^2) + (p - 1) O(\gamma^6) = O(p^{-1}) \rightarrow 0, \end{aligned}$$

as $p \rightarrow \infty$, since (A.1). This result and (A.4) lead to $D = O(p^{-1}) \rightarrow 0$, as $p \rightarrow \infty$.

Remark. The asymptotic results (2.6) and (2.7) can be derived in a similar manner to that mentioned above. For example, to evaluate

$$\log \left[1 - \operatorname{Erf} \left(x \frac{\sqrt{2\pi} \Gamma(p_0 + 1/2)}{4 \Gamma(p_0)} \gamma \right) \right]^2$$

a simple transformation $y = x\Gamma(p_0 + 1/2)/\Gamma(p_0)$ can be used for providing the proof scheme presented above.

Appendix B: Scores and Information

Scores per observation for the general beta family of distributions corresponding to the parameter vector $\theta = (\mu, \sigma, p, q)$ corresponding to the likelihood at (3.2) are given as:

$$s_1(x_i) = \frac{-(p-1)f_0(z_i)}{\sigma F_0(z_i)} + \frac{(q-1)f_0(z_i)}{\sigma(1-F_0(z_i))} - \frac{f'_0(z_i)}{\sigma f_0(z_i)}, \quad (\text{B.1})$$

$$s_2(x_i) = \frac{-1}{\sigma} - z_i s_1(x_i), \quad (\text{B.2})$$

$$s_3(x_i) = \psi(p+q) - \psi(p) + \log F_0(z_i), \quad (\text{B.3})$$

$$s_4(x_i) = \psi(p+q) - \psi(q) + \log[1 - F_0(z_i)], \quad (\text{B.4})$$

where ψ denotes the di-gamma function and the indices $j = 1, 2, 3, 4$ pertain to the elements of θ .

Sample information per observation for the general beta family of distributions:

$$\begin{aligned} i_{11}(x_i) = & \frac{(p-1)[f_0(z_i)^2 - F_0(z_i)f'_0(z_i)]}{\sigma^2 F_0(z_i)^2} \\ & - \frac{(q-1)[f_0(z_i)^2 + (1-F_0(z_i))f'_0(z_i)]}{\sigma^2 [1-F_0(z_i)]^2} + \frac{f'_0(z_i)^2 - f_0(z_i)f''_0(z_i)}{\sigma^2 f_0(z_i)^2} \end{aligned} \quad (\text{B.5})$$

$$\begin{aligned} i_{12}(x_i) = & \frac{z_i(p-1)[f_0(z_i)^2 - F_0(z_i)f'_0(z_i)]}{\sigma^2 F_0(z_i)^2} \\ & + \frac{z_i(q-1)[f_0(z_i)^2 + (1-F_0(z_i))f'_0(z_i)]}{\sigma^2 [1-F_0(z_i)]^2} + \frac{z_i[f'_0(z_i)^2 - f_0(z_i)f''_0(z_i)]}{\sigma^2 f_0(z_i)^2} + \frac{s_{11}}{\sigma} \end{aligned} \quad (\text{B.6})$$

$$i_{13}(x_i) = \frac{f_0(z_i)}{\sigma F_0(z_i)} \quad (\text{B.7})$$

$$i_{14}(x_i) = \frac{-f_0(z_i)}{\sigma(1 - F_0(z_i))} \quad (\text{B.8})$$

$$i_{22}(x_i) = \frac{1}{\sigma^2} + z_i^2 i_{11} + 2z_i \left\{ \frac{-(p-1)f_0(z_i)}{\sigma^2 F_0(z_i)} + \frac{(q-1)f_0(z_i)}{\sigma^2 [1 - F_0(z_i)]} + \frac{f'_0(z_i)}{f_0(z_i)} \right\} \quad (\text{B.9})$$

$$i_{23}(x_i) = -z_i i_{13} \quad (\text{B.10})$$

$$i_{24}(x_i) = -z_i i_{14} \quad (\text{B.11})$$

$$i_{33}(x_i) = \psi'(p) - \psi'(p+q) \quad (\text{B.12})$$

$$i_{34}(x_i) = -\psi'(p+q) \quad (\text{B.13})$$

$$i_{44}(x_i) = \psi'(q) - \psi'(p+q) \quad (\text{B.14})$$

References

- Cordeiro G, Nadarajah S and Ortega EMM (2013) General results for the beta Weibull distribution. *Journal of Statistical Computation and Simulation* 83, 1082–1114
- Eugene N, Lee C and Famoye F (2002) Beta-Normal Distribution and Its Applications. *Communications in Statistics-Theory and Methods*, 31 497–512.
- Famoye F, Lee C and Olumolade O (2005) The beta-Weibull distribution. *Journal of Statistical Theory and Applications*, 4 121–136.
- Hennessy DA (2009) Crop Yield Skewness Under Law of the Minimum Technology. *American Journal of Agricultural Economics* 91 197–208.
- Hutson AD (2004) Utilizing the Flexibility of the Epsilon-Skew-Normal Distribution for Common Regression Problems. *Journal of Applied Statistical Science*, 31 673–683.
- Jones MC (2004) Families of distributions arising from distributions of order statistics. *Test*, 13 1–43.
- Mameli V and Musio M (2013) A Generalization of the Skew-Normal Distribution: The Beta Skew-Normal. *Communications in Statistics - Theory and Methods* 42 2229–2244.
- Mameli V (2015) The Kumaraswamy skew-normal distribution. *Statistics and Probability Letters* 104 75–81.
- Mortici C (2010). New approximation formulas for evaluating the ratio of gamma functions. *Mathematical and Computing Modelling*, 52 425–433.
- Mudholkar GS and Hutson AD (1996) A Study of the Exponentiated Weibull Family of Distributions. *Communications in Statistics-Theory and Methods*, 25 3059–3083
- Mudholkar GS and Hutson AD (2000) The Epsilon-Skew-Normal Distribution for Analyzing Near-Normal Data. *Journal of Statistical Planning and Inference*, 83 291–309.
- Qi F (2010). Bounds for the ratio of two gamma functions. *Journal Inequalities and Applications*, 2010 1–84 (Available online at 10.1155/2010/493058)
- Ranger J and Kuhn J-T (2015) Modeling Information Accumulation in Psychological Tests Using Item Response Times. *Journal of Educational and Behavioral Statistics* 40 274–306.

- Razzaghi M (2009) Modeling Information Accumulation in Psychological Tests Using Item Response Times. *Environmental and Ecological Statistics*, 16 25–36.
- Tricomi FG and Erdélyi A (1951). The asymptotic expansion of a ratio of gamma functions. *Pacific Journal of Mathematics*, 1 133–142.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1: Standard normal distribution overlaid with its alias beta-normal distributions $\mu = 0$, $p = q = 4$ and $\sigma = 2.180$ and $\mu = 0$, $p = q = 10$ and $\sigma = 3.518$.

Figure 2:
Beta-normal distribution with $\mu=0$, $p=2$, $q=4$ and $\sigma=1$ overlayed with its alias beta-normal distributions $\mu=0$, $p=10$, $q=14.288$ and $\sigma=2.147$.

Table 1:

Common Standardized Symmetric Models

Distribution	$F_{\theta}(x)$	$f_{\theta}(x)$	$Q_{\theta}(u)$
Normal	$\Phi(x)$	$\exp(-x^2/2)/\sqrt{2\pi}$	$\Phi^{-1}(u)$
Logistic	$1/(1 + \exp(-x))$	$\exp(x)/(1 + \exp(x))^2$	$\log(u/(1 - u))$

Table 2:

Simulation results for 3 symmetric alias scenarios using Eugene *et al.* (2002) starting points. Quartiles from the simulated maximum likelihood estimates are given in the table.

$n=50$												
Case 1				Case 2				Case 3				
$\mu=0, \sigma=1, p=q=1$				$\mu=0, \sigma=2.18, p=q=4$				$\mu=0, \sigma=3.518, p=q=10$				
	$\hat{\mu}$	$\hat{\sigma}$	$\hat{p}$	$\hat{q}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{p}$	$\hat{q}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{p}$	$\hat{q}$
Q1	-0.99	0.36	0.14	0.14	-0.96	0.37	0.14	0.14	-0.92	0.38	0.14	0.14
Q2	-0.01	0.49	0.47	0.47	-0.03	0.50	0.57	0.50	0.01	0.51	0.49	0.48
Q3	0.90	1.08	1.54	1.45	1.00	1.06	1.52	1.55	0.98	1.14	1.44	1.56
$n=500$												
Case 1				Case 2				Case 3				
$\mu=0, \sigma=1, p=q=1$				$\mu=0, \sigma=2.18, p=q=4$				$\mu=0, \sigma=3.518, p=q=10$				
	$\hat{\mu}$	$\hat{\sigma}$	$\hat{p}$	$\hat{q}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{p}$	$\hat{q}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{p}$	$\hat{q}$
Q1	-0.41	0.57	0.43	0.43	-0.40	0.59	0.50	0.49	-0.49	0.60	0.47	0.47
Q2	0.01	0.84	0.80	0.81	0.01	0.93	1.01	1.01	-0.01	0.92	0.93	0.93
Q3	0.45	1.43	1.78	1.78	0.41	1.42	1.96	1.98	0.41	1.46	1.87	1.87

Table 3:

Simulation results for 3 symmetric alias scenarios using Case 2 true parameter values as starting points. Quartiles from the simulated maximum likelihood estimates are given in the table.

$n=50$												
	Case 1				Case 2				Case 3			
	$\mu=0, \sigma=1, p=q=1$				$\mu=0, \sigma=2.18, p=q=4$				$\mu=0, \sigma=3.518, p=q=10$			
	$\hat{\mu}$	$\hat{\sigma}$	$\hat{p}$	$\hat{q}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{p}$	$\hat{q}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{p}$	$\hat{q}$
Q1	-1.88	0.62	0.96	0.97	-1.75	0.59	0.81	0.96	-1.84	0.58	0.92	0.82
Q2	-0.01	1.41	3.14	2.87	0.42	1.46	2.14	3.15	0.00	1.41	3.06	2.94
Q3	1.84	4.29	16.3	16.51	2.00	4.27	12.38	13.60	1.96	4.37	13.52	13.66
$n=500$												
	Case 1				Case 2				Case 3			
	$\mu=0, \sigma=1, p=q=1$				$\mu=0, \sigma=2.18, p=q=4$				$\mu=0, \sigma=3.518, p=q=10$			
	$\hat{\mu}$	$\hat{\sigma}$	$\hat{p}$	$\hat{q}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{p}$	$\hat{q}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{p}$	$\hat{q}$
Q1	-1.55	1.52	2.16	2.11	-1.18	1.50	2.00	2.02	-1.10	1.49	2.01	2.00
Q2	-0.04	4.23	7.64	7.66	0.03	3.80	6.97	6.70	-0.03	3.88	6.85	7.02
Q3	1.36	4.89	27.28	24.10	1.08	4.89	22.70	23.32	1.26	4.91	23.94	24.32

Table 4:

Simulation results for 2 asymmetric alias scenarios using Eugene *et al.* (2002) starting points. Quartiles from the simulated maximum likelihood estimates are given in the table.

$n = 50$								
Case 4				Case 5				
$\mu = 0, \sigma = 1, p = 2, q = 4$				$\mu = 0, \sigma = 3.518, p = 10, q = 14.28$				
	$\hat{\mu}$	$\hat{\sigma}$	$\hat{p}$	$\hat{q}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{p}$	$\hat{q}$
Q1	-0.83	0.23	0.13	0.19	-1.04	0.22	0.14	0.14
Q2	-0.20	0.29	0.42	0.96	-0.44	0.28	0.49	0.63
Q3	0.28	0.62	1.35	1.79	0.17	0.60	1.49	1.61
$n = 500$								
Case 4				Case 5				
$\mu = 0, \sigma = 1, p = 2, q = 4$				$\mu = 0, \sigma = 3.518, p = 10, q = 14.28$				
	$\hat{\mu}$	$\hat{\sigma}$	$\hat{p}$	$\hat{q}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{p}$	$\hat{q}$
Q1	-0.39	0.35	0.37	0.68	-0.64	0.33	0.41	0.52
Q2	-0.16	0.54	0.75	1.27	-0.39	0.49	0.78	0.94
Q3	0.03	0.84	1.59	2.58	-0.17	0.80	1.73	1.95