Published in final edited form as:

Commun Stat Theory Methods. 2019 ; 48(12): 3005–3024. doi:10.1080/03610926.2018.1473600.

An Alternative Skew Exponential Power Distribution Formulation

Alan D. Hutson* [Professor of Oncology]

*Roswell Park Cancer Institute, Department of Biostatistics and Bioinformatics, Elm and Carlton Streets, Buffalo, NY 14623.

Abstract

In this note we propose a newly formulated skew exponential power distribution that behaves substantially better than previously defined versions. This new model performs very well in terms of the large sample behavior of the maximum likelihood estimation procedure when compared to the classically defined four parameter model defined by Azzalini (1986). More recently, approaches to defining a skew exponential power distribution have used five or more parameters. Our approach improves upon previous attempts to extend the symmetric power exponential family to include skew alternatives by maintaining a minimum set of four parameters corresponding directly to location, scale, skewness and kurtosis. We illustrate the utility of our proposed model using translational and clinical data sets.

Keywords

skew normal distribution; skew Laplace distribution; beta normal distribution; quantiles; expectiles

1 Introduction

Over the last three decades there has been substantial efforts in developing skew counterparts to classic symmetric distributions. Many of these approaches center around overcoming the assumptions of normality found in the inferential components of many parametric models. However, this technique of creating skew counterparts actually has an even longer history. As observed by Runnenberg (1978), an alternative skew extension obtained by splicing two half-normal distributions differing in scale appears in Fechner (1897) in the context of a mean-median-mode ordering result. Historically, there have always been doubts about the unqualified use of normality, and resistance to methods resulting from the normality assumptions. This is reflected in the quote attributed by Poincaré to Lippmann; see Stigler (1977): “Everyone believes in the normal law, the experimenters because they imagine it a mathematical theorem, and the mathematicians because they think it an experimental fact.” Note also the provocative assertion by R. C. Geary (1947): “Normality is a myth; there never was, and never will be, a normal distribution.”

The most highly cited article pertaining to a skew generalization of a symmetric distribution is authored by Azzalini (1985), who developed the form a skew normal distribution whose density and distribution functions have the form

$$f(x) = \frac{2}{\sigma} \Phi(\lambda z) \phi(z), \quad (1.1)$$

$$F(x) = \Phi(z) - 2T(z, \lambda), \quad (1.2)$$

where $-\infty < \theta, \lambda < \infty$, $\sigma > 0$, $z = (x - \theta)/\sigma$, $\phi(z)$ is a standard normal density, $\Phi(z)$ is a standard normal distribution function and $T(z, \lambda)$ denotes Owen's T function (Owen 1956). A few limitations of this form of a skew normal distribution and the subsequent extensions is that the distribution function is not invertible and it requires Owen's T function to be available in a given software package in order for it to be readily calculated. Issues pertaining to maximum likelihood singularities in the information matrix are touched on below in the broader context of the skew exponential power family of distributions.

Mudholkar and Hutson (1999) developed an alternative parameterization of the skew normal distribution termed the epsilon-skew-normal distribution $ESN(\theta, \sigma, \epsilon)$, which is a unimodal distribution with the mode at θ and probability mass $(1 + \epsilon)/2$ below the mode. The probability density function, the distribution function, and quantile function are given as follows:

$$f_0(z) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{z^2}{2(1+\epsilon)^2}\right), & \text{if } z < 0, \\ \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{z^2}{2(1-\epsilon)^2}\right), & \text{if } z \geq 0, \end{cases} \quad (1.3)$$

$$F_0(z) = \begin{cases} (1+\epsilon)\Phi\left(\frac{z}{1+\epsilon}\right), & \text{if } z < 0, \\ \epsilon + (1-\epsilon)\Phi\left(\frac{z}{1-\epsilon}\right), & \text{if } z \geq 0, \end{cases} \quad (1.4)$$

and

$$Q_0(u) = F_0^{-1}(u) = \theta + \sigma \begin{cases} (1+\epsilon)\Phi^{-1}\left(\frac{u}{1+\epsilon}\right), & \text{if } 0 < u < (1+\epsilon)/2, \\ (1-\epsilon)\Phi^{-1}\left(\frac{u-\epsilon}{1-\epsilon}\right), & \text{if } (1+\epsilon)/2 \leq u < 1, \end{cases} \quad (1.5)$$

where $-1 < \epsilon < 1$, $-\infty < \theta < \infty$, $\sigma > 0$, $z = (x - \theta)/\sigma$, $\phi(x)$ is a standard normal density and $\Phi(x)$ is a standard normal distribution function. The limiting cases of (1.3) as $\epsilon \rightarrow \pm 1$ are the well known half-normal distributions. The standard epsilon-skew-normal distribution,

$ESN(0, 1, \epsilon)$, is a mixture of two half-normal distributions, and reduces to the standard normal distribution, when $\epsilon = 0$. In other words the $ESN(0, 1, \epsilon)$ may be viewed as generated by $(1 - U)(1 - \epsilon)|N_1| - U(1 + \epsilon)|N_2|$, where U, N_1, N_2 are independent, $P(U = 1) = (1 + \epsilon)/2 = 1 - P(U = 0)$, and N_1, N_2 are independent standard normal variates. There are many variations of this type of parameterization scheme in the literature as noted by Jones (2006). Also see, Arellano-Valle *et al.* (2005).

Work has been done on extending the exponential power distribution in a similar fashion to a skewed family extension. The symmetric exponential power distribution dates back to Subbotin (1923). We will use the parameterization given in Box and Tiao (1973), where additional early references to the distributional family's history can be found. The Box and Tiao parametrization of the exponential power density function is given as

$$f(z) = \frac{k}{\sigma} \exp\left(-\frac{1}{2}|z|^{2/(1+\beta)}\right), \quad (1.6)$$

where $k^{-1} = \Gamma\left(1 + \frac{1+\beta}{2}\right)2^{1+\frac{1}{2}(1+\beta)}$, $-1 < \beta \leq 1$, $-\infty < \theta < \infty$, $\sigma > 0$, $z = (x - \theta)/\sigma$ and $\Gamma(x)$ is the standard gamma function. When $\beta = 0$ the density at (1.6) reduces to the standard normal density and when $\beta = 1$ the density at (1.6) reduces to the Laplace or double exponential distribution. In addition, when $\beta \rightarrow -1$ the density tends to a uniform distribution. The exponential power distribution has found a wide range of applications, particularly in the area of business applications, e.g. see Kobayashi (2016) and the references there within. Multivariate generalizations are given by Fernández *et al.* (1995). A Bayesian treatment of the subject may be found in Naranjo *et al.* (2012). Other Bayesian models include a two-piece form as given in Rubio and Steel (2015). A mixture model approach towards creating skew distributions may be found in Wichitaksorn *et al.* (2014).

The two forms of the skew exponential power distribution that follow similar to either the Azzalini (1985) or Mudholkar and Hutson (2000) formulation given above for skew normal families are due to Azzalini (1986) and Kobayashi (2016), respectively, see also Salinas *et al.* (2007). The Azzalini form has a skew exponential power density function given as

$$f(x) = \frac{2}{\sigma c} \Phi(w) \exp(-|z|^\alpha/\alpha) \quad (1.7)$$

where $c = 2\alpha^{1/\alpha-1}\Gamma(1/\alpha)$, $\lambda > 0$, $w = \text{sign}(z)|z|^{\alpha/2} \times \lambda\sqrt{2/\alpha}$, $-\infty < \theta < \infty$, $\sigma > 0$, $z = (x - \theta)/\sigma$, $\Phi(x)$ is a standard normal distribution function and $\Gamma(x)$ is the standard gamma function. DiCiccio and Monti (2005) provide a detailed maximum likelihood framework for parameter estimation for the density at (1.7). Specific issues about maximum likelihood estimation in the context of this form pertain to near-normal distributions causing singularity issues in the information matrix. In addition, the simulation study performed by DiCiccio and Monti (2005) illustrate the general instability of the Azzalini maximum likelihood estimators up until sample sizes surpass $n \gg 1000$, which in turn makes inference about

specific parameters infeasible. More stable estimation strategies for this model are provided by Monti (2003), which may overcome certain difficulties. An alternative form of this model may be found in Zhu and Zinde-Walsh (2009) with the goal of increasing the range of skewness values that can be accommodated.

The Kobayashi (2016) form of the skew exponential power distribution consists of a five parameter model having a similar form to what was originally proposed by Fernandez *et al.* (1995) and given as

$$f(x) = c^{-1} \times \begin{cases} \exp\left(-(1-p)|z|^{\theta_1}\right), & \text{if } z < 0, \\ \exp\left(-p|z|^{\theta_2}\right), & \text{if } z \geq 0 \end{cases}, \quad (1.8)$$

where

$$c = \frac{\sigma}{(1-p)^{1/\theta_1}} \Gamma(1 + 1/\theta_1) + \frac{\sigma}{p^{1/\theta_2}} \Gamma(1 + 1/\theta_2),$$

$\theta_1, \theta_2, \sigma > 0, -\infty < \theta < \infty, 0 < p < 1, z = (x - \theta)/\sigma$ and $\Gamma(x)$ is the standard gamma function. In this framework the mode of the distribution is θ such that

$$P(X < \theta) = \frac{(1-p)^{-1/\theta_1} \Gamma(1 + 1/\theta_1)}{(1-p)^{-1/\theta_1} \Gamma(1 + 1/\theta_1) + p^{-1/\theta_2} \Gamma(1 + 1/\theta_2)}. \quad (1.9)$$

Kobayashi (2016) does not provide a frequentist maximum likelihood treatment pertaining to this parameterization. In the Kobayashi paper the estimation approach centers on Markov chain Monte Carlo methods.

While each of the above parameterizations have their respective merits there are some shortcomings in both approaches that can be readily overcome with the parameterization that we present in the next section. The form of the density defined by Azzalini (1986) at (1.7) only includes the skew normal distribution as a special case of the exponential power distribution, e.g. the Laplace density is a special case of the symmetric exponential power family yet there is no skew Laplace density available given this form of the model. In fact for any symmetric exponential power set of parameters in the Azzalini (1986) model the only skew counterpart is the skew normal density. One straightforward benefit of having skew counterparts to symmetric power exponential densities is that goodness-of-fit tests for submodels can be readily accommodated via the single skewing parameter.

The Azzalini (1986) formulation suffers from a singular information matrix under the normal submodel, e.g. see DiCiccio and Monti (2005), which in turn requires a less than desirable additional layer of theory and conditional checks within the maximum likelihood

fitting routine. This is not the case under our formulation. The Kobayashi (2016) formulation may not suffer from this type of singularity issue, however maximum likelihood methods were not investigated as part of their work. Our approach allows that for all symmetric power exponential parameterizations there will be a direct skew counterpart. Another disadvantage of the Azzalini formulation is that the cumulative distribution and the quantile function are not readily available without numerical integration and inversion procedures.

As compared to the Azzalini (1986) formulation the model developed by Kobayashi (2016) does have skew counterparts to all symmetric power exponential densities given in the submodels. More simply, the shapes covered by five parameters in the model by Kobayashi (2016) at (1.8) can be accommodated by a simpler four parameter model that we propose. Our model will also readily be able to utilize the classic maximum likelihood framework. The reduction in the number of parameters from five to four without losing any real shape information will reduce the complexity of the estimation procedures, whether one wishes to utilize a Bayesian or frequentist approach. Our approach will also have much cleaner representations relative to expressions such as those found at (1.9) in terms of the probability above or below the mode θ and other similar expressions, e.g. our density formulation does not need to be broken out into two distinct parameterizations relative to the probability support being above or below the mode or some other cut-point.

An alternative approach, which contains the Azzalini form as a special case, is what is commonly referred to as 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 $z = (x - \mu)/\sigma$,

$$F(x) = F_0(z), \quad (1.10)$$

$$f(x) = \frac{1}{\sigma} f_0(z), \quad (1.11)$$

$$Q(u) = \mu + \sigma Q_0(u), \quad (1.12)$$

where μ and σ are the location and scale parameters. 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

$$G(x) = B_{p,q}[F_0(z)], \quad (1.13)$$

$$g(x) = \frac{1}{\sigma} b_{p,q}[F_0(z)] f_0(z), \quad (1.14)$$

$$P(u) = \mu + \sigma Q_0[B_{p,q}^{-1}(u)], \quad (1.15)$$

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

$b_{p,q}(t) = \frac{1}{\beta(p,q)} t^{p-1} (1-t)^{q-1}$ and $\beta(p,q)$ is the normalizing constant. In this case we can substitute the exponential power density at (1.6) in for f_0 at (1.11).

Relative to this approach Hutson and Vexler (2017) highlight theoretically and numerically the concept of limiting alias distributions within the beta family of distributions, which in turn relate to numerical instabilities in the numerical maximum likelihood fitting routines for these families of distributions. For example, 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 $p = q > 0$. Then Hutson and Vexler (2017) show that the

$$\lim_{p \rightarrow \infty} \log\left(\frac{f(x)}{g(x)}\right) = 0, \quad (1.16)$$

where f denotes the standard normal density function with $\mu = 0$, $\sigma = 1$ and $p = q = 1$. Hence the Kulback-Liebler distance $D = \int_{-\infty}^{\infty} \log\left(\frac{f(x)}{g(x)}\right) f(x) dx$ goes to 0 for extremely divergent values of $p = q$ different from 1. Hence, the alternative beta family approach to generalizing the power exponential family would be deemed infeasible.

In Section 2 we define a new parameterization for the skew exponential power distribution and define the density, distribution and quantile functions. In the same section we outline some properties for the new distribution, provide graphical representation of various density shapes and derive its moments. In Section 3 we derive the score and expected information expressions needed to obtain the maximum likelihood estimators via numerical methods. In Section 4 we perform a Monte Carlo simulation study to examine the large sample behavior of the maximum likelihood estimators. In Section 5 we apply our model to two real world clinical and translational datasets.

2 Skew Exponential Power Distribution

The motivation for our parameterization of the skew exponential power distribution follows from a version of the definition of the α -quantile for the random variable X given by the value of θ satisfying

$$\inf_{\theta \in \mathcal{R}} E \left\{ \frac{|X - \theta| + (2\alpha - 1)(X - \theta)}{2} - \frac{|X| + (2\alpha - 1)X}{2} \right\}. \quad (2.1)$$

The properties of the expression at (2.1) are outlined in Abdous and Theodorescu (1992) who generalize the definition of the α -quantile to R^k space, $k \geq 1$. Most notably, the assumption that $E(X) < \infty$, $\alpha \in (0, 1)$, in (2.1) is not a necessary condition for the existence of the α -quantile. By noting heuristically that α serves the role as a skewing parameter relative to the norm within the expectation at (2.1) allows one to consider using it to construct a density function similar to how other norms have been used in the past, e.g. similar to how the L_1 and L_2 norms are linked to the Laplace and normal distributions, respectively. Towards this end we propose modifying the density at (1.6) using α as our skewing parameter. This leads to our new definition for a skew exponential power density function parameterized as follows:

$$f(x) = \frac{k}{\sigma} \exp \left(-\frac{1}{2} (|z| + (2\alpha - 1)z)^{2/(1+\beta)} \right), \quad (2.2)$$

where $k^{-1} = \Gamma \left(1 + \frac{1+\beta}{2} \right) 2^{1 + \frac{1}{2}(1+\beta)} / (4\alpha(1-\alpha))$, $0 < \alpha < 1$, $-1 < \beta \leq 1$, $-\infty < \theta < \infty$, $\sigma > 0$, $z = (x - \theta)/\sigma$ and $\Gamma(\cdot)$ is the Euler gamma function. In terms of this parameterization θ serves as the location parameter and σ serves as the scale parameter. As we will see below α and β are shape parameters.

The distribution function is then given as

$$F(x) = \begin{cases} \frac{\alpha \Gamma \left(\frac{\beta+1}{2}, 2^{\frac{2}{\beta+1}-1} ((\alpha-1)z)^{\frac{2}{\beta+1}} \right)}{\Gamma \left(\frac{\beta+1}{2} \right)}, & \text{if } z \leq 0, \\ 1 - \frac{(1-\alpha) \Gamma \left(\frac{\beta+1}{2}, 2^{\frac{2}{\beta+1}-1} \left(\frac{1}{\alpha z} \right)^{-\frac{2}{\beta+1}} \right)}{\Gamma \left(\frac{\beta+1}{2} \right)}, & \text{if } z > 0, \end{cases}, \quad (2.3)$$

where $-1 < \beta \leq 1$, $-\infty < \theta < \infty$, $\sigma > 0$, $z = (x - \theta)/\sigma$, $\Gamma(\cdot, \cdot)$ is the incomplete gamma function and $\Gamma(\cdot)$ is the complete gamma function. We can write the cumulative distribution function (c.d.f.) more compactly as

$$F(x) = \begin{cases} \alpha P\left(\frac{\beta+1}{2}, 2^{\frac{2}{\beta+1}-1}((\alpha-1)z)^{\frac{2}{\beta+1}}\right), & \text{if } z \leq 0, \\ 1 - (1-\alpha)P\left(\frac{\beta+1}{2}, 2^{\frac{2}{\beta+1}-1}\left(\frac{1}{\alpha z}\right)^{-\frac{2}{\beta+1}}\right), & \text{if } z > 0, \end{cases}, \quad (2.4)$$

where $P(a, z) = \Gamma(a, z)/\Gamma(a)$ is the regularized gamma function or the survival function of a standard gamma distribution. Again we have $-1 < \beta < 1$, $-\infty < \theta < \infty$, $\sigma > 0$ and $z = (x - \theta)/\sigma$.

Inverting the c.d.f. at (2.4) yields the corresponding quantile function given as

$$Q(u) = \begin{cases} \theta - \sigma \frac{\left(2^{\left(\frac{1}{2} - \frac{1}{\beta+1}\right)} \sqrt{P^{-1}\left(\frac{\beta+1}{2}, \frac{u}{\alpha}\right)}\right)^{\beta+1}}{1-\alpha}, & \text{if } u \leq \alpha, \\ \theta + \sigma \frac{\left(\frac{2^{\left(\frac{1}{\beta+1} - \frac{1}{2}\right)}}{\sqrt{P^{-1}\left(\frac{\beta+1}{2}, \frac{1-u}{1-\alpha}\right)}}\right)^{-(\beta+1)}}{\alpha}, & \text{if } u > \alpha, \end{cases}, \quad (2.5)$$

where $0 < u < 1$. From a practical point of view the quantile function representation (2.5) can be rewritten in terms of the built in gamma quantile functions found in most statistical software as $P^{-1}(a, u) = Q_{ga}(1-u)$. This allows one to generate simulated data quite easily using most standard statistical software packages.

Mode.

The mode of the skew exponential power distribution is θ . This can be noted by the fact that derivative of the p.d.f. at (2.2)

$$f'(x) = -\frac{\sigma^{-\frac{2}{\beta+1}}(2\alpha-1+\text{sgn}(x-\theta))(|x-\theta|+(2\alpha-1)(x-\theta))^{\frac{2}{\beta+1}-1}}{\beta+1}, \quad (2.6)$$

which is 0 at $x = \theta$. Then we have

$$P(X < \theta) = \alpha, \quad (2.7)$$

which in turn leads to the expression for the odds of being below the mode given as

$$P(X < \theta)/P(X > \theta) = \alpha/(1 - \alpha). \quad (2.8)$$

Special Cases.

Below are some well-known distributions, which are sub models of the skew exponential power distribution for specific parameter values.

1. Exponential power distribution ($\alpha = 1/2$).
2. Skew Laplace distribution ($\beta = 1$).
3. Laplace distribution ($\beta = 1, \alpha = 1/2$).
4. Skew Normal distribution ($\beta = 0$).
5. Normal distribution ($\beta = 0, \alpha = 1/2$).
6. Uniform distribution ($\beta \rightarrow -1$).

Moments.

If X has the p.d.f. of the form at (2.2) then the random variable $Z = (X - \theta)/\sigma$ has the canonical form of the skew exponential power density given as

$$f(z) = k \exp\left(-\frac{1}{2}(|z| + (2\alpha - 1)z)\right)^{2/(1+\beta)}, \quad (2.9)$$

where $k^{-1} = \Gamma\left(1 + \frac{1+\beta}{2}\right) 2^{1 + \frac{1}{2}(1+\beta)} / (4\alpha(1-\alpha))$, $0 < \alpha < 1 - 1 < \beta \leq 1$, $-\infty < \theta < \infty$, and $\Gamma(\cdot)$ is the Euler gamma function. It is readily seen that l th moment for Z is

$$E(z^l) = \frac{(\beta + 1)(-2)^l \alpha^{l+1} + (1 - \alpha)(2(1 - \alpha))^l 2^{\frac{1}{2}(\beta - 3)l - 1} \Gamma\left(\frac{1}{2}(\beta + 1)(l + 1)\right)}{(1 - \alpha)^l \alpha^l \Gamma\left(\frac{\beta + 3}{2}\right)} \quad (2.10)$$

In particular the first four moments of Z are given as

$$E(z) = \frac{(\beta + 1)(1 - 2\alpha) 2^{\frac{\beta - 3}{2}} \Gamma(\beta + 1)}{(1 - \alpha)\alpha \Gamma\left(\frac{\beta + 3}{2}\right)}, \quad (2.11)$$

$$E(z^2) = \frac{(\beta + 1)(3\alpha^2 - 3\alpha + 1) 2^{\beta - 2} \Gamma\left(\frac{3(\beta + 1)}{2}\right)}{(1 - \alpha)^2 \alpha^2 \Gamma\left(\frac{\beta + 3}{2}\right)}, \quad (2.12)$$

$$E(z^3) = \frac{(\beta + 1)(-4\alpha^3 + 6\alpha^2 - 4\alpha + 1)2^{\frac{3\beta}{2} - \frac{5}{2}}\Gamma(2(\beta + 1))}{(1 - \alpha)^3\alpha^3\Gamma(\frac{\beta + 3}{2})}, \quad (2.13)$$

and

$$E(z^4) = \frac{(\beta + 1)(5\alpha^4 - 10\alpha^3 + 10\alpha^2 - 5\alpha + 1)2^{2\beta - 3}\Gamma(\frac{5(\beta + 1)}{2})}{(1 - \alpha)^4\alpha^4\Gamma(\frac{\beta + 3}{2})}. \quad (2.14)$$

The n th moment for X then can be expressed as

$$\begin{aligned} E(X^n) &= E[(\theta + \sigma Z)^n], \\ &= \sum_{l=0}^n \binom{n}{l} \theta^{n-l} \sigma^l E(Z^l), \end{aligned} \quad (2.15)$$

where $E(Z^l)$ is given above at (2.10). In particular the first four moments of X are given as

$$E(X) = \theta + \sigma E(Z), \quad (2.16)$$

$$E(X^2) = \theta^2 + 2\theta\sigma E(Z) + \sigma^2 E(Z^2), \quad (2.17)$$

$$E(X^3) = \theta^3 + 3\theta^2\sigma E(Z) + 3\theta\sigma^2 E(Z^2) + \sigma^3 E(Z^3) \quad (2.18)$$

and

$$E(X^4) = \theta^4 + 4\theta^3\sigma E(Z) + 6\theta^2\sigma^2 E(Z^2) + 4\theta\sigma^3 E(Z^3) + \sigma^4 E(Z^4). \quad (2.19)$$

The skewness and kurtosis of the skew exponential power distribution are functions of α and β , respectively. This can be seen by noting that the skew exponential power distribution has the form of a location-scale family of distribution with two additional parameters. However, the actual equations in terms of α and β are quite expansive. In compact form we can write the variance, skewness and kurtosis of X using the standardized moments of Z at (2.10) and given as:

$$\text{Var}(X) = \sigma^2(E(Z^2) - E(Z)^2) \quad (2.20)$$

$$\text{Skewness}(X) = \frac{E(Z^3) - 3E(Z)E(Z^2) + 2E(Z)^3}{(E(Z^2) - E(Z)^2)^{\frac{3}{2}}} \quad (2.21)$$

$$\text{Kurtosis}(X) = \frac{E(Z^4) - 4E(Z^3)E(Z) + 6E(Z^2)E(Z)^2 - 3E(Z)^4}{(E(Z^2) - E(Z)^2)^2} - 3. \quad (2.22)$$

The bounds for skewness are $-2 < \text{Skewness}(X) < 2$ at $\alpha \rightarrow 0, \beta = 1$ and $\alpha \rightarrow 1, \beta = 1$, respectively. The bounds for kurtosis are $-6/5 < \text{Kurtosis}(X) < 6$ at $\beta \rightarrow -1$ (uniform) and $\alpha \rightarrow (0, 1), \beta = 1$, respectively. In Figures 1–3 we provide some examples of density shapes for the skew exponential power distribution over various sets of skewness and kurtosis parameters.

3 Maximum Likelihood Estimation

Let $X_1, X_2, \dots, X_n$ be an independent and identically distributed (i.i.d.) sample drawn from the skew exponential power distribution (2.2) with corresponding vector of parameters $\lambda = (\theta, \sigma, \alpha, \beta)$. Denote the log-likelihood function for λ as $\sum_{i=1}^n l(\lambda, X_i)$ where the likelihood per observation is given as

$$\begin{aligned} l(\lambda, X) = & -\frac{(|x - \theta| + (2\alpha - 1)(x - \theta))^{\frac{2}{\beta+1}}}{2\sigma} + \log(\alpha(1 - \alpha)) - \left(1 + \frac{1}{2}(\beta + 1)\right)\log(2) \\ & - \log\left(\Gamma\left(\frac{\beta + 1}{2} + 1\right)\right) - \log(\sigma). \end{aligned}$$

(3.1)

The score function is $\sum_{i=1}^n S(\lambda, X_i)$, where the scores per observation, $S_{\lambda_j}(X_i) = \partial l(\lambda, X_i) / \partial \lambda_j$, are given as

$$s_{\theta}(x) = \frac{\sigma^{-\frac{2}{\beta+1}}(|\theta - x| + (2\alpha - 1)(x - \theta))^{\frac{2}{\beta+1}}}{(\beta + 1)(x - \theta)}, \quad (3.2)$$

$$s_{\sigma}(x) = \frac{\sigma^{-\frac{2}{\beta+1}-1}(|x-\theta| + (2\alpha-1)(x-\theta))^{\frac{2}{\beta+1}}}{\beta+1} - \frac{1}{\sigma}, \quad (3.3)$$

$$s_{\alpha}(x) = -\frac{2\sigma^{-\frac{2}{\beta+1}}(x-\theta)(|\theta-x| + (2\alpha-1)(x-\theta))^{\frac{2}{\beta+1}-1}}{\beta+1} - \frac{(2\alpha-1)}{\alpha(1-\alpha)}, \quad (3.4)$$

and

$$\begin{aligned} s_{\beta}(x) &= -\frac{\sigma^{-\frac{2}{\beta+1}}\log(\sigma)(|\theta-x| + (2\alpha-1)(x-\theta))^{\frac{2}{\beta+1}}}{(\beta+1)^2} \\ &+ \frac{\sigma^{-\frac{2}{\beta+1}}(|\theta-x| + (2\alpha-1)(x-\theta))^{\frac{2}{\beta+1}}\log(|\theta-x| + (2\alpha-1)(x-\theta))}{(\beta+1)^2} \\ &- \frac{1}{2}\psi\left(\frac{\beta+3}{2}\right) - \frac{\log(2)}{2}, \end{aligned} \quad (3.5)$$

where λ_j denotes the j th element of the vector λ , $j = 1, 2, 3, 4$ and $\psi(\cdot)$ denotes the digamma function. The derivative of $|\theta - x|$ with respect to θ used to simplify the expressions above was $|\theta - x| \wedge (x - \theta)$, which corresponds to the sign of $(x - \theta)$.

A closed form expression for the maximum likelihood estimator for σ is available and given as

$$\hat{\sigma} = \left(\frac{\sum_{i=1}^n (|x_i - \theta| + (2\alpha-1)(x_i - \theta))^{\frac{2}{\beta+1}}}{n(\beta+1)} \right)^{\frac{\beta+1}{2}}. \quad (3.6)$$

The other maximum likelihood estimators need to be derived numerically. However, finding even this one closed form expression for $\hat{\sigma}$ greatly reduces the overall numerical complexity and stability of the numerical maximum likelihood fitting routine. As we can see for the normal case at $\beta = 0$ and $\alpha = 1/2$ the expression at (3.6) reduces to $\left(\sum_{i=1}^n (x_i - \theta)^2 / n \right)^{1/2}$.

Likewise we can see for the Laplace case at $\beta = 1$ and $\alpha = 1/2$ the expression at (3.6) reduces to $\sum_{i=1}^n |x_i - \theta| / (2n)$ as per classic theory.

The elements of the sample information and co-information per observation for each combination of parameter values, $i_{\lambda_j \lambda_k}(X_i) = -\partial s_{\lambda_j}(X_i) / \partial \lambda_k$, $j = 1, 2, 3, 4$, $k = 1, 2, 3, 4$, $j \leq k$ are given in the appendix at equations (A.1)–(A.10). The elements involved in the upper triangle

of the 4×4 expected information matrix are given by n times the expected information per observation $I_{\lambda_j \lambda_k} = E \left(E i_{\lambda_j \lambda_k} (X_i) \right)$ for $j = 1, 2, 3, 4, k = 1, 2, 3, 4, j \neq k$ where:

$$I_{\theta\theta}(x) = \frac{(1-\alpha)\alpha 2^{3-\beta} \Gamma\left(\frac{3-\beta}{2}\right)}{(\beta+1)^2 \sigma^2 \Gamma\left(\frac{\beta+1}{2}\right)}, \quad (3.7)$$

$$I_{\theta\sigma}(x) = 0, \quad (3.8)$$

$$I_{\theta\alpha}(x) = -\frac{2^{\frac{3}{2}-\frac{\beta}{2}}}{(\beta+1)\sigma \Gamma\left(\frac{\beta+3}{2}\right)}, \quad (3.9)$$

$$I_{\theta\beta}(x) = \frac{(1-\alpha)\alpha 2^{\frac{1}{2}-\frac{\beta}{2}} \left(2\beta \log\left(-\frac{\alpha-1}{\sigma}\right) + (\beta+1) \log\left(\frac{1}{2} \left(\frac{2-2\alpha}{\sigma}\right)^{\frac{2}{\beta+1}}\right) \right)}{(\beta+1)^4 \sigma \Gamma\left(\frac{\beta+3}{2}\right)} \quad (3.10)$$

$$- \frac{2(\beta+1) \log(2(1-\alpha)) + 2\beta \log(\sigma) + \beta \log(8) + 2 \log(\sigma) + \log(2)}{(\beta+1)^4 \sigma \Gamma\left(\frac{\beta+3}{2}\right)},$$

$$I_{\sigma\sigma}(x) = \frac{2}{(\beta+1)\sigma^2}, \quad (3.11)$$

$$I_{\sigma\alpha}(x) = \frac{2(2\alpha-1)}{(1-\alpha)\alpha(\beta+1)\sigma}, \quad (3.12)$$

$$I_{\sigma\beta}(x) = \frac{\psi\left(\frac{\beta}{2} + \frac{3}{2}\right) + 1 + \log(2)}{(\beta+1)\sigma}, \quad (3.13)$$

$$I_{\alpha\alpha}(x) = \frac{(1-\alpha)\alpha(\beta-5)+2}{(1-\alpha)^2 \alpha^2 (\beta+1)}, \quad (3.14)$$

$$I_{\alpha\beta}(x) = \frac{(2\alpha - 1)\left(\psi\left(\frac{\beta+3}{2}\right) + 1 + \log(2)\right)}{(1-\alpha)\alpha(\beta+1)} \quad (3.15)$$

$$I_{\beta\beta}(x) = \frac{2\psi\left(\frac{\beta+3}{2}\right)^2 + (\beta+3)\psi_1\left(\frac{\beta+3}{2}\right) + 4(1+\log(2))\psi\left(\frac{\beta+3}{2}\right) + 2\log^2(2) + \log(16)}{4(\beta+1)}, \quad (3.16)$$

where $\psi_1(\cdot)$ is the trigamma function.

Unlike the Azzalini (1986) skew exponential power distribution maximum likelihood routine there is no singularity issue for the normal submodel ($\beta = 0$, $\alpha = 1/2$) under our formulation, where the value for the determinant of the information matrix under the normal parametrization is

$$\frac{(3\pi - 8)(3\pi^2 - 28)}{\pi\sigma^4}. \quad (3.17)$$

In general, we will illustrate the relative stability of the numerically derived maximum likelihood estimates for reasonably small sample sizes in a Monte Carlo study in the next section.

Specific Tests of Interest.

Tests of symmetry $H_0 : \alpha = 1/2$ should perform well even for alternative distributions that are not contained in the skew exponential power family of distributions. Similarly, tests about kurtosis, jointly about kurtosis and symmetry are possible. Similarly, tests of normality or skew normality as subclass are possible, e.g. $H_0 : \alpha = 1/2$, $\beta = 0$ or $H_0 : \beta = 0$, respectively. These tests are not the focus of this work, but might be considered for future study including a more formal examination of the specific skew-normal submodel formulation as compared to other skew-normal models currently found in the literature.

4 Monte Carlo Studies

DiCiccio and Monti (2004) performed a simulation study regarding the behavior of the maximum likelihood estimators for the Azzalini skew exponential power density formulation at (1.7) for the special situation of near normal data and in general. They examined the bias and root mean squared error over a series of parameter combinations and sample sizes. In general, the bias of their formulation does not start to become negligible until the sample size is $n = 1000$. In fact, for samples of size $n = 250$ there is still substantial bias in the parameter estimates obtained via maximum likelihood. As noted by DiCiccio and

Monti (2004) “if, instead, the whole parameter set is of interest, large sample size is needed.” Again, as noted above we do not need to be concerned about special cases for our formulation, e.g. the normal parameterization is no more different than any other scenario and we will illustrate that the maximum likelihood estimates behave well for what one might typically expect in terms of sample sizes and convergence rates.

For our simulation studies we used 1000 Monte Carlo resamples using the quantile function at (2.5) and applying it to random uniform variates for samples of size $n = 35, 50, 100$ across parameter values $\mu = 0$, $\sigma = 1$, $\alpha = 0.5, 0.75, 0.9$ and $\beta = 0, -0.5, 0.5$, where the beta values correspond to mesokurtic, platykurtic and leptokurtic tails, respectively. Similar to DiCiccio and Monti (2004) we provide both the bias and mean-squared errors for each parameter setting. As one can see from the Monte Carlo simulation results provided in Tables 1–3 the bias reduces to near zero for sample sizes of size $n = 50$ and the estimators definitely behave well at sample sizes $n = 100$, which is dramatically better than what was shown by DiCiccio and Monti (2004) for the Azzalini (1986) formulation. The MSE reduces accordingly with the increase in sample sizes. Again, as noted above the Kobayashi (2016) formulation is over parameterized at five parameters as compared to four parameters for our formulation, which covers the same density shapes.

5 Example

For our examples we fit the Azzalini, Kobayashi and Hutson skew exponential power densities at (1.7), (1.8) and (2.2), respectively, with a starting value of $\bar{x}$ for θ and utilizing the expression for $\hat{\sigma}$ at (3.6) to simplify the numerical fitting process. We did a grid search for the additional parameters for each model. Various goodness-of-fit measures for each model are given in Table 10 consisting of -2 times the log-likelihood function, the Akaike(AIC), the finite sample corrected Akaike (AICC) and the Schwartz(BIC) measures. The smaller values for the AIC, AICC and BIC imply a better fit.

Example 1.

Example 1 data comes from screening data from a clinical trial in childhood malaria as described in Planche *et al.* (2003). We fit the systolic and diastolic blood pressure to the each skew exponential power density for a sample size of $n = 1416$ children. The histograms and overlaid densities for the Hutson skew exponential power density are given in Figures 4 and 5, respectively. The corresponding maximum likelihood estimates and 95% confidence intervals for each model are provided in Tables 4–8. It is interesting to note that when looking at the fitted figures at Figures 4 and 5 it appears that the data look normal-like. However, the parameter estimates for α and β along with the corresponding 95% confidence intervals indicate both distributions are slight skewed away from normal. The odds of being below θ for systolic blood pressure is $0.42/(1-.42)=0.72$. Similarly, the odds of being below θ for diastolic blood pressure is $0.36/(1-.36)=0.56$. More striking is that the tail estimated leptokurtic weights are significantly different from normal even though the data visually looks normally distributed, where the normal submodel has tail weights corresponding to $\beta = 1$.

Example 2.

Example 2 data comes from a study of percent weight change in $n = 219$ mice following treatment with cancer drugs, see Demant *et al.* (2015). The histogram and overlaid density is given in Figure 6. The corresponding maximum likelihood estimates and 95% confidence intervals for each model are provided in Tables 4–8.

In the one sample percent change from baseline setting it may be of interest to test using our model $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$, particularly when percent change is not necessarily symmetrically distributed even for large sample sizes. For the percent change in weight the estimated value of α and the corresponding confidence interval indicates asymmetry in the data. In the one sample setting symmetry about a location parameter is a key assumption for many statistical tests, e.g. the one sample t-test and the Wilcoxon sign-rank test. Hence, the ability to test about a location parameter given asymmetry is meaningful. For our percent weight change the corresponding large sample p-value for testing $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$ is < 0.0001 and does not violate the assumptions of normality and or symmetry of the classic tests.

In terms of goodness-of-fit measures across the examples the four parameter Azzalini and Hutson models were quite similar with no negligible difference in the four measures in Table 10. Surprisingly, the five parameter Kobayashi model, though not much different, always had poorer fits than the competing four parameter models. Formal Anderson-Darling goodness-of-fit tests were performed for each model across all examples. For all tests and all models the p-values were > 0.250 for testing the respective model fits.

Acknowledgments

This work was supported by Roswell Park Cancer Institute and National Cancer Institute (NCI) grant P30CA016056, National Science Foundation IIS-1514204 and NRG Oncology Statistical and Data Management Center grant U10CA180822. We wish to thank the three referees whose thoughtful reviews vastly improved this paper.

Appendix A:: Sample Information Per Observation

Expressions for the sample information per observation for the skew exponentiated power distribution (2.2) are given below:

$$i_{\theta\theta}(x) = - \frac{(\beta - 1)\sigma^{-\frac{2}{\beta+1}}(|x - \theta| + (2\alpha - 1)(x - \theta))^{\frac{2}{\beta+1}}}{(\beta + 1)^2(x - \theta)^2} \quad (\text{A.1})$$

$$i_{\theta\sigma}(x) = \frac{2\sigma^{-\frac{\beta+3}{\beta+1}}(|x - \theta| + (2\alpha - 1)(x - \theta))^{\frac{2}{\beta+1}}}{(\beta + 1)^2(x - \theta)} \quad (\text{A.2})$$

$$i_{\theta\alpha}(x) = -\frac{4\sigma^{-\frac{2}{\beta+1}}(|x-\theta|+(2\alpha-1)(x-\theta))^{\frac{2}{\beta+1}}}{(\beta+1)^2((2\alpha-1)(x-\theta)+|x-\theta|)} \quad (\text{A.3})$$

$$\begin{aligned} I_{\theta\beta}(x) &= \frac{\sigma^{-\frac{2}{\beta+1}}(|x-\theta|+(2\alpha-1)(x-\theta))^{\frac{2}{\beta+1}}(2\log(|x-\theta|))}{(\beta+1)^3(x-\theta)} \\ &+ \frac{(2\alpha-1)(x-\theta)+\beta-2\log(\sigma)+1}{(\beta+1)^3(x-\theta)} \end{aligned} \quad (\text{A.4})$$

$$i_{\sigma\sigma}(x) = \frac{(\beta+3)\sigma^{-\frac{2}{\beta+1}}(|x-\theta|+(2\alpha-1)(x-\theta))^{\frac{2}{\beta+1}}-(\beta+1)^2}{(\beta+1)^2\sigma^2} \quad (\text{A.5})$$

$$i_{\sigma\alpha}(x) = -\frac{4\sigma^{-\frac{\beta+3}{\beta+1}}(x-\theta)(|x-\theta|+(2\alpha-1)(x-\theta))^{\frac{2}{\beta+1}-1}}{(\beta+1)^2} \quad (\text{A.6})$$

$$\begin{aligned} I_{\sigma\beta}(x) &= \frac{\sigma^{-\frac{\beta+3}{\beta+1}}(|x-\theta|+(2\alpha-1)(x-\theta))^{\frac{2}{\beta+1}}(2\log(|\theta-x|))}{(\beta+1)^3} \\ &+ \frac{(2\alpha-1)(x-\theta)+\beta-2\log(\sigma)+1}{(\beta+1)^3} \end{aligned} \quad (\text{A.7})$$

$$\begin{aligned} i_{\alpha\alpha}(x) &= -\frac{4(\beta-1)\sigma^{-\frac{2}{\beta+1}}(\theta-x)^2(|x-\theta|+(2\alpha-1)(x-\theta))^{-\frac{2\beta}{\beta+1}}}{(\beta+1)^2} + \frac{1}{\alpha^2} + \frac{1}{(\alpha-1)^2} \\ & \quad (\text{A.8}) \end{aligned}$$

$$I_{\alpha\beta}(x) = \frac{2\sigma^{-\frac{2}{\beta+1}}(\theta-x)(|x-\theta|+(2\alpha-1)(x-\theta))^{\frac{2}{\beta+1}}}{(\beta+1)^3((2\alpha-1)(x-\theta)+|x-\theta|)} \quad (\text{A.9})$$

$$\times \frac{(2\log(|x-\theta|+(2\alpha-1)(x-\theta))+\beta-2\log(\sigma)+1)}{(\beta+1)^3((2\alpha-1)(\theta-x)-|x-\theta|)}$$

$$I_{\beta\beta}(x) = \frac{\sigma^{-\frac{2}{\beta+1}} \left(8(|x-\theta|+(2\alpha-1)(x-\theta))^{\frac{2}{\beta+1}} (\log(\sigma) - \log(|x-\theta|+(2\alpha-1)(x-\theta))) \right)}{4(\beta+1)^4}$$

$$\times \frac{(-\log(|x-\theta|+(2\alpha-1)(x-\theta)) - \beta + \log(\sigma) - 1) + (\beta+1)^4 \sigma^{\frac{2}{\beta+1}} \psi\left(\frac{\beta+3}{2}\right)}{4(\beta+1)^4}$$

(A.10)

6 References

- Abdous B and Theodorescu R (1992). "Note on the spatial quantile of a random vector," *Statistics & Probability Letters*, 13, 333–336.
- Arellano-Valle RB, Gomez HW and Quintana FA (2005) Statistical inference for a general class of asymmetric distributions. *Journal of Statistical Planning and Inference* 128 427–443.
- Azzalini A (1985) A Class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* 12 171–178.
- Azzalini A (1986) Further results on a class of distributions which includes the normal ones. *Statistica* 46 199–208.
- Box GE and Tiao GC (1973) *Bayesian Inference in Statistical Analysis* Addison-Wesley Publishing Company, Reading, MA.
- Demant P, Thudium KE, Sepporta MV, Affronti H, Meijer G, Prendergast L, Mathijssen RHJ, Burhans WC, Ma WW, Hutson A, Fetterly GJ, Dittmar AJ (2015) Comprehensive genetic definition of susceptibility to toxic side effects of irinotecan in mice. *Developmental Therapeutics Clinical Pharmacology and Experimental Therapeutics Abstract* e13566.
- DiCiccio TJ and Monti AC (2004) Inferential aspects of the skew exponential power distribution. *Journal of the American Statistical Association* 99 439–450.
- Eugene N, Lee C and Famoye F (2002) *Beta-Normal Distribution and Its Applications*. *Communications in Statistics-Theory and Methods*, 31 497–512.
- Fechner GT (1897) *Kollektivmasslehre* Leipzig, Engleman.
- Férrandez C, Osiewalski J, Steel MFJ (1995). Modeling and inference with $[nu]$ spherical distributions. *Journal of the American Statistical Association* 90 1331–1340.
- Geary RC (1947) Testing for normality. *Biometrika* 34 209–242. [PubMed: 18918691]

- Hutson AD and Vexler A (2017) A Cautionary Note on Beta Families of Distributions and the Aliases Within. *The American Statistician* (in press, available online).
- Jones MC (2004) Families of distributions arising from distributions of order statistics. *Test*, 13 1–43.
- Jones MC (2006) A Note on rescalings, reparameterizations and classes of distributions. *Journal of Statistical Planning and Inference* 136 2730–3733.
- Kobayashi G (2016) Skew exponential power stochastic volatility model for analysis of skewness, non-normal tails, quantiles and expectiles. *Computational Statistics* 31 48–88.
- Monti AC (2003) A note on the estimation of the skew normal and the skew exponential power distributions. *Metron* 61 205–219.
- 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.
- Naranjo L and Pérez CJ, Martn J (2012) Bayesian analysis of a skewed exponential power distribution. *Proceedings of COMPSTAT 2012, 20th International Conference on Computational Statistics* 641–652.
- Owen D (1956) Tables for computing bivariate normal probabilities. *Annals of Mathematical Statistics* 27 1075–1090.
- Planche T, Agbenyega T, Bedu-Addo G, Ansong D, Owusu-Ofori A, Micah F, Anakwa C, Asfo-Agyei E, Hutson A, Stacpoole P, and Krishna S (2003) A Prospective Comparison of Malaria with Other Severe Diseases in African Children: Prognosis and Optimization of Management. *Clinical Infectious Diseases* 37 890–897. [PubMed: 13130399]
- Rubio FJ and Steel MFJ (2015) Bayesian modelling of skewness and kurtosis with two-piece scale and shape distributions. *Electronic Journal of Statistics* 9 1884–1912.
- Runnenberg JT (1978) Mean, median, mode. *Statistica Neerlandica* 32 73–79.
- Salinas HS, Arellano-Valle RB and Gomez HW (2007) The extended skew-exponential power distribution and its derivation. *Communications in Statistics Theory and Methods* 36 1673–1689.
- Subbotin MT (1923) On the law of frequency of error. *Matematicheskii Sbornik* 31 296–301.
- Stigler SM (1977) Do robust estimators work with real data? *The Annals of Statistics* 5 1055–1098.
- Wichitaksorn N, Choy STB, Gerlach R (2014) A generalized class of skew distributions and associated robust quantile regression models. *The Canadian Journal of Statistics* 42 579–596.
- Zhu D and Zinde-Walsh V (2009) Properties and estimation of asymmetric exponential power distribution. *Journal of Econometrics* 148 86–99.

Figure 1:

Plot of the skew exponential power distribution for $\theta = 0$, $\sigma = 1$, $\alpha = 1/10, 1/4, 0, 3/4, 9/10$ and $\beta = -9/10, -1/2$

Figure 2:

Plot of the skew exponential power distribution for $\theta = 0$, $\sigma = 1$, $\alpha = 1/10, 1/4, 0, 3/4, 9/10$ and $\beta = 0, 1/2$

Figure 3:

Plot of the skew exponential power distribution for $\theta = 0$, $\sigma = 1$, $\alpha = 1/10, 1/4, 0, 3/4, 9/10$ and $\beta = 7/10, 1$

Figure 4:
Systolic blood pressure from childhood malaria screening data.

Figure 5:
Diastolic blood pressure from childhood malaria screening data.

Figure 6:
Percent change in body weight in mice.

Table 1:

Simulation study for various skewness and mesokurtic settings.

		$\mu = 0$		$\sigma = 1$		$\alpha = 0.5$		$\beta = 0$	
n		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
35		0.006	0.259	-0.071	0.133	0.002	0.025	-0.071	0.250
50		0.012	0.197	-0.030	0.095	0.003	0.019	-0.078	0.150
100		0.001	0.087	-0.007	0.044	0.001	0.008	-0.034	0.061
		$\mu = 0$		$\sigma = 1$		$\alpha = 0.75$		$\beta = 0$	
n		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
35		0.025	0.198	-0.055	0.140	0.015	0.011	-0.104	0.198
50		0.020	0.135	-0.041	0.105	0.010	0.008	-0.060	0.124
100		0.001	0.080	-0.015	0.060	0.002	0.004	-0.026	0.056
		$\mu = 0$		$\sigma = 1$		$\alpha = 0.9$		$\beta = 0$	
n		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
35		0.006	0.057	-0.082	0.050	0.011	0.001	0.088	0.077
50		0.010	0.039	-0.060	0.035	0.008	0.001	-0.074	0.057
100		0.005	0.023	-0.028	0.021	0.004	0.001	-0.036	0.027

Table 2:

Simulation study for various skewness and platykurtic settings.

n	$\mu = 0$		$\sigma = 1$		$\alpha = 0.5$		$\beta = -0.5$	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
35	0.027	0.159	-0.082	0.073	0.013	0.029	-0.142	0.133
50	0.001	0.161	-0.093	0.048	-0.001	0.028	-0.086	0.028
100	-0.003	0.099	-0.047	0.020	-0.001	0.016	-0.049	0.029
n	$\mu = 0$		$\sigma = 1$		$\alpha = 0.75$		$\beta = -0.5$	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
35	0.041	0.080	-0.086	0.057	0.023	0.008	-0.135	0.108
50	0.037	0.067	-0.063	0.048	0.017	0.006	-0.088	0.075
100	0.014	0.042	-0.033	0.030	0.007	0.004	-0.032	0.028
n	$\mu = 0$		$\sigma = 1$		$\alpha = 0.9$		$\beta = -0.5$	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
35	-0.046	0.023	-0.053	0.017	0.006	0.001	-0.082	0.046
50	-0.038	0.015	-0.022	0.012	0.002	0.001	-0.058	0.032
100	-0.018	0.008	-0.005	0.006	0.001	0.001	-0.035	0.016

Table 3:

Simulation study for various skewness and leptokurtic settings.

n	$\mu = 0$		$\sigma = 1$		$\alpha = 0.5$		$\beta = -0.5$	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
35	-0.017	0.324	-0.050	0.220	-0.002	0.017	-0.121	0.237
50	-0.004	0.220	0.034	0.170	0.001	0.011	-0.064	0.175
100	-0.024	0.083	0.020	0.094	-0.005	0.004	-0.022	0.090
n	$\mu = 0$		$\sigma = 1$		$\alpha = 0.75$		$\beta = 0.5$	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
35	0.057	0.268	0.008	0.253	0.019	0.009	-0.129	0.197
50	0.031	0.189	0.012	0.191	0.011	0.006	-0.078	0.154
100	0.003	0.101	-0.005	0.092	0.005	0.003	-0.018	0.079
n	$\mu = 0$		$\sigma = 1$		$\alpha = 0.9$		$\beta = 0.5$	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
35	0.006	0.104	-0.068	0.085	0.010	0.001	-0.112	0.081
50	-0.020	0.086	-0.061	0.060	0.007	0.001	-0.054	0.064
100	0.003	0.040	-0.024	0.031	0.004	0.001	-0.035	0.033

Table 4:

Maximum likelihood estimates for childhood malaria systolic and diastolic blood pressure data based on Hutson model (2.2)

Parameter	Systolic Blood Pressure			Diastolic Blood Pressure		
	Estimate	95% confidence interval		Estimate	95% confidence interval	
		Lower	Upper		Lower	Upper
θ	84.85	82.82	86.87	48.30	46.93	49.66
σ	10.44	9.04	11.84	8.87	7.60	10.14
α	0.42	0.38	0.46	0.36	0.32	0.39
β	0.38	0.24	0.51	0.32	0.19	0.46

Table 5:

Maximum likelihood estimates for mouse percent change in weight loss data based on Hutson model (2.2)

Parameter	Estimate	95% confidence interval	
		Lower	Upper
θ	7.95	6.70	9.20
σ	2.59	1.61	3.57
α	0.31	0.22	0.39
β	0.41	0.06	0.76

Table 6:

Maximum likelihood estimates for childhood malaria systolic and diastolic blood pressure data based on Azzalini model (1.7)

Parameter	Systolic Blood Pressure			Diastolic Blood Pressure		
	Estimate	95% confidence interval		Estimate	95% confidence interval	
		Lower	Upper		Lower	Upper
θ	79.57	78.55	80.59	41.91	41.48	42.34
σ	15.56	14.44	16.68	16.32	15.34	17.29
α	1.44	1.30	1.58	1.55	1.40	1.69
λ	0.70	0.56	0.84	1.34	1.15	1.53

Table 7:

Maximum likelihood estimates for mouse percent change in weight loss data based on Azzalini model (1.7)

Parameter	Estimate	95% confidence interval	
		Lower	Upper
θ	6.85	6.33	7.38
σ	5.03	4.14	5.93
α	1.33	1.01	1.65
λ	1.05	0.61	1.49

Table 8:

Maximum likelihood estimates for childhood malaria systolic and diastolic blood pressure data based on Kobayashi model (1.8)

Parameter	Systolic Blood Pressure			Diastolic Blood Pressure		
	Estimate	95% confidence interval		Estimate	95% confidence interval	
		Lower	Upper		Lower	Upper
θ	80.01	79.71	80.30	48.96	45.34	52.56
σ	8.31	6.88	9.73	8.94	7.24	10.65
θ_1	1.13	0.95	1.31	1.56	1.17	1.94
θ_2	1.68	1.50	1.87	1.48	1.17	1.78
p	0.15	0.11	0.20	0.32	0.09	0.57

Table 9:

Maximum likelihood estimates for mouse percent change in weight loss data based on Kobayashi model (1.8)

Parameter	Estimate	95% confidence interval	
		Lower	Upper
θ	9.70	9.32	10.07
σ	3.24	7 2.1	4.32
θ_1	2.15	1.33	2.96
θ_2	1.13	0.82	1.45
p	0.61	0.44	0.78

Table 10:

Goodness-of-fit comparisons between the Hutson, Azzalini and Kobayashi models.

SBP Example				
Model	-2 log likelihood	AIC	AICC	BIC
Hutson	11702	11710	11710	11731
Azzalini	11701	11709	11709	11730
Kobaysi	11709	11719	11719	11745
DBP Example				
Model	-2 log likelihood	AIC	AICC	BIC
Hutson	11182	11190	11190	11211
Azzalini	11181	11189	11189	11210
Kobaysi	111182	11192	11192	11219
Weight Loss Example				
Model	-2 log likelihood	AIC	AICC	BIC
Hutson	1448.9	1456.9	1457.1	1471.0
Azzalini	1450.4	1458.4	1458.6	1472.5
Kobaysi	1447.9	1457.9	1458.2	1475.5