Author(s): M Kolossiatis, JE Griffin and MFJ Steel

Article Title: Modelling overdispersion with the Normalized Tempered Stable distribution

Year of publication: 2011

Link to published article: http://dx.doi.org/10.1016/j.csda.2011.01.016

Modelling overdispersion with the Normalized Tempered Stable distribution

M. Kolossiatis,* J. E. Griffin† and M. F. J. Steel*

Abstract

This paper discusses a multivariate distribution which generalizes the Dirichlet distribution and demonstrates its usefulness for modelling overdispersion in count data. The distribution is constructed by normalizing a vector of independent Tempered Stable random variables. General formulae for all moments and cross-moments of the distribution are derived and they are found to have similar forms to those for the Dirichlet distribution. The univariate version of the distribution can be used as a mixing distribution for the success probability of a Binomial distribution to define an alternative to the well-studied Beta-Binomial distribution. Examples of fitting this model to simulated and real data are presented.

Keywords: Mice Fetal Mortality; Normalized Random Measures; Overdispersion

1 Introduction

In many experiments we observe data as the number of observations in a sample with some property. The Binomial distribution is a natural model for this type of data. However, the data are often found to be overdispersed relative to that model. This is often explained and modelled through differences in the binomial success probability p for different units. It is assumed that x_i successes (*i.e.* observations possessing a certain property) are observed from n_i observations and that $x_i \sim \text{Bi}(n_i, p_i)$ where p_i are drawn independent from some

*Department of Statistics, University of Warwick, Coventry, CV4 7AL, U.K. and † School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, CT2 7NF, U.K. The authors would like to thank Professor Martin Ridout for helpful suggestions

mixing distribution on $(0, 1)$. The Beta distribution is a natural choice since the likelihood has an analytic form. However, although analytically convenient, estimates will be biased if the data does not not support this assumption. Many other choices of mixing distribution have been studied including the Logistic-Normal-Binomial model (Williams, 1982) and the Probit-Normal-Binomial model (Ochi and Prentice, 1984). Altham (1978) and Kupper and Haseman (1978) propose a two-parameter distribution, the Correlated-Binomial, which allows for direct interpretation and assignment of the correlation between any two of the underlying Bernoulli observations of a Binomial random variable through one of its two parameters. Paul (1985) proposes a three-parameter generalization of the Beta-Binomial distribution, the Beta-Correlated Binomial distribution. There is also a modified version of the latter in Paul (1987). Brooks *et al* (1997) use finite mixture distributions to provide a flexible specification. However, the introduction of a mixture distribution leads to more complicated inference and harder interpretation of parameters. Kuk (2004) suggests the q-power distribution which models the joint success probabilities of all orders by a power family of completely monotone functions which extends the folded logistic class of George and Dowman (1995). Pang and Kuk (2005) define a shared response model by allowing each response to be independent of all others with probability π or taking a value Z with probability $1 - \pi$. Therefore more than one observation can take the common value Z. They argue that this is more interpretable than the q-power distribution of Kuk (2004). Rodriguez-Avi *et al* (2007) use a Generalized Beta distribution as the mixing distribution.

In this paper, we consider an alternative specification for the distribution of p_i for which all cross-moments are available analytically. A random variable X on $(0, 1)$ can be defined by considering two independent positive random variables V_1 and V_2 and taking

$$X = \frac{V_1}{V_1 + V_2}.$$

A popular choice is $V_1 \sim \text{Ga}(a_1, 1)$ and $V_2 \sim \text{Ga}(a_2, 1)$ where $\text{Ga}(a, b)$ represents a Gamma distribution with shape a and mean a/b. This choice implies that X follows a Beta distribution with parameters a_1 and a_2. The Tempered Stable distribution is a three-parameter generalization of the Gamma distribution which has the Inverse-Gaussian distribution as a special case. The extra parameter influences the heaviness of the tails of the distribution. The distribution was proposed by Hougaard (1986) as a generalization of the stable distribution to model frailties in survival analysis. Recently, it has been found that the additional flexibility can be useful in the modelling of cell generation times (Palmer *et al*, 2008). The Normalized Tempered Stable class is defined by choosing V_1 and V_2 to be Tempered Stable random variables and is a rich family of distributions which generalizes the Beta distribution and, more generally, the Dirichlet distribution, for random variables defined on the unit simplex. Normalized Tempered Stable distributions are indexed by a single additional parameter and can accommodate heavier tails and more skewness than would be possible for the Dirichlet distribution.

The paper is organized as follows: Section 2 discusses some background ideas: the Tempered Stable distribution and some distributions on the unit simplex. Section 3 describes the Normalized Tempered Stable distribution and the form of its moments and cross-moments, Section 4 considers the use of this distribution as a mixing distribution in a Binomial mixture, Section 5 illustrates the use of the model for simulated and real data and compares its fit to other specifications and, finally, Section 6 concludes.

2 Background

This section describes the Tempered Stable distribution and some properties of the Dirichlet and Normalized Inverse-Gaussian distributions.

2.1 Tempered Stable distribution

The Tempered Stable (TS) distribution was introduced by Tweedie (1984).

Definition 1. *A random variable X defined on $\mathbb{R}_+$ follows a Tempered Stable distribution with parameters κ, δ and γ ($0 < \kappa < 1$, $\delta > 0$ and $\gamma > 0$) if its Lévy density is*

$$u(x) = \delta 2^{\kappa} \frac{\kappa}{\Gamma(1-\kappa)} x^{-1-\kappa} \exp\left\{-\frac{1}{2}\gamma^{1/\kappa} x\right\}.$$

We will write $X \sim TS(\kappa, \delta, \gamma)$.

In general, the probability density function is not available analytically but can be expressed through a series representation due to its relationship to the positive stable distribution (see Feller, 1971)

$$f(x) = c \sum_{k=1}^{\infty} (-1)^{(k-1)} \sin(k\pi\kappa) \frac{\Gamma(k\kappa+1)}{k!} 2^{k\kappa+1} (x\delta^{-1/\kappa})^{(-k\kappa-1)} \exp\left\{-\frac{1}{2}\gamma^{1/\kappa} x\right\}$$

where $c = \frac{1}{2\pi}\delta^{-1/\kappa}\exp\{\delta\gamma\}$. The expectation of X is $2\kappa\delta\gamma^{(\kappa-1)/\kappa}$ and its variance is $4\kappa(1-\kappa)\delta\gamma^{(\kappa-2)/\kappa}$. The moment generating function will be important for our derivations and is given by

$$\mathrm{E}(\exp\{tx\}) = \exp\{\delta\gamma - \delta(\gamma^{1/\kappa} - 2t)^{\kappa}\}. \tag{2.1}$$

There are two important subclasses. A TS $\left(\kappa, \frac{\nu}{\kappa\psi^{2\kappa}}, \psi^{2\kappa}\right)$ will limit in probability as $\kappa \to 0$ to a Gamma distribution which has probability density function

$$f(x) = \frac{(\psi^2/2)^\nu}{\Gamma(\nu)} x^{\nu-1} \exp\left\{-\frac{1}{2}\psi^2 x\right\}$$

and the Inverse-Gaussian distribution arises when $\kappa = \frac{1}{2}$ which has probability density function

$$f(x) = \frac{\delta}{\sqrt{2\pi}} \exp\{\delta\gamma\} x^{-3/2} \exp\left\{-\frac{1}{2}(\delta^2 x^{-1} + \gamma^2 x)\right\}$$

The Tempered Stable distribution is infinitely divisible and self-decomposable.

2.2 Dirichlet and Normalized Inverse-Gaussian distribution

The Dirichlet distribution is a commonly used distribution on the unit simplex.

Definition 2. *An n-dimensional random variable $\boldsymbol{W} = (W_1, W_2, \ldots, W_n)$ is said to follow a Dirichlet distribution with parameters $a_1, a_2, \ldots, a_{n+1} > 0$, denoted Dir($a_1, a_2, \ldots, a_{n+1}$), if its density is:*

$$f(\boldsymbol{w}) = \frac{\Gamma(a_1 + a_2 + \cdots + a_{n+1})}{\Gamma(a_1)\Gamma(a_2)\cdots\Gamma(a_{n+1})} \prod_{i=1}^{n} w_i^{a_i - 1} \left(1 - \sum_{j=1}^{n} w_j\right)^{a_{n+1}-1},$$

where $\boldsymbol{w} = (w_1, \ldots, w_n)$ is such that $w_1, \ldots, w_n \geq 0$ and $0 < \sum_{k=1}^{n} w_k < 1$.

The Dirichlet distribution can be generated through an obvious generalization of the normalization idea in Section 1 to a vector of $(n+1)$ independent random variables for which the i-th entry has a Ga$(a_i, 1)$ distribution. Its moments and cross-moments are easy to calculate and are as follows:

Proposition 1. *Let $\boldsymbol{W} = (W_1, \ldots, W_n)$ ~Dir($a_1, a_2, \ldots, a_{n+1}$) and let $N_1, N_2, N \in \mathbb{N}$. Defining $S = \sum_{i=1}^{n+1} a_i$ and $\mu_j = a_j/S$, we can state*

1. $E(W_i^N) = \frac{\Gamma(S)\Gamma(a_i+N)}{\Gamma(S+N)\Gamma(a_i)} = \prod_{k=0}^{N-1}\left(\frac{a_i+k}{S+k}\right),\ i = 1, 2, \ldots, n.$
2. $E(W_i^{N_1} W_j^{N_2}) = \frac{\Gamma(S)\Gamma(a_i+N_1)\Gamma(a_j+N_2)}{\Gamma(S+N_1+N_2)\Gamma(a_i)\Gamma(a_j)} = \prod_{k=0}^{N_1-1}\left(\frac{a_i+k}{S+k}\right)\prod_{l=0}^{N_2-1}\left(\frac{a_j+l}{S+N_1+l}\right),\ i, j = 1, 2, \ldots, n,\ i \neq j.$
3. $E(W_i) = \mu_i$
4. $Var(W_i) = \frac{\mu_i(1-\mu_i)}{S+1}$
5. $Cov(W_i, W_j) = -\frac{\mu_i\mu_j}{S+1}.$

The Inverse-Gaussian distribution is an interesting sub-class of Tempered Stable distributions since its probability density function is available analytically. This is also true for the Normalized Inverse-Gaussian distribution (Lijoi *et al*, 2005) which is constructed by normalizing a vector of $(n+1)$ independent random variables for which all entries have Inverse-Gaussian distributions. It is defined in the following way.

Definition 3. *We say that an n-dimensional random variable $\boldsymbol{W} = (W_1, W_2, \dots, W_n)$ follows a Normalised Inverse-Gaussian (N-IG) distribution with parameters $\nu_1, \nu_2, \dots, \nu_{n+1} > 0$, or $\boldsymbol{W} \sim$ N-IG$(\nu_1, \nu_2, \dots, \nu_{n+1})$, if*

$$f(\boldsymbol{w}) = \frac{\exp\left\{\sum_{i=1}^{n+1} \nu_i\right\} \prod_{i=1}^{n+1} \nu_i}{2^{(n+1)/2-1}\pi^{(n+1)/2} A_{n+1}(\boldsymbol{w})^{(n+1)/4}} K_{-(n+1)/2}\left(\sqrt{A_{n+1}(\boldsymbol{w})}\right) \prod_{i=1}^{n} w_i^{-3/2} \left(1 - \sum_{j=1}^{n} w_j\right)^{-3/2} \tag{2.2}$$

where $\boldsymbol{w} = (w_1, \dots, w_n)$ is such that $w_1, \dots, w_n \geq 0$ and $0 < \sum_{k=1}^{n-1} w_k < 1$, $A_{n+1}(\boldsymbol{w}) = \sum_{i=1}^{n} \frac{\nu_i^2}{w_i} + \frac{\nu_{n+1}^2}{1 - \sum_{j=1}^{m-1} w_j}$ and K is the modified Bessel function of the third type.

The expectation and variance of any component W_i, as well as the covariance structure of any two components, W_i and W_j of a N-IG-distributed random vector $\boldsymbol{W}$ are given in Lijoi *et al* (2005):

Proposition 2. *Let $\boldsymbol{W} = (W_1, \dots, W_n) \sim$ N-IG$(\nu_1, \nu_2, \dots, \nu_{n+1})$. Then,*

1. *$E(W_i) = \frac{\nu_i}{S} =: \mu_i,\ i = 1, 2, \dots, n.$*
2. *$Var(W_i) = \mu_i(1-\mu_i)S^2 \exp\{S\}\Gamma(-2, S),\ i = 1, 2, \dots, n.$*
3. *$Cov(W_i, W_j) = -\mu_i\mu_j S^2 \exp\{S\}\Gamma(-2, S),\ i, j = 1, 2, \dots, n,\ i \neq j.$*

where $S = \sum_{i=1}^{n+1} \nu_i$ and $\Gamma(a, x) = \int_x^\infty t^{a-1} \exp\{-t\} dt$ is the incomplete Gamma function.

3 Multivariate Normalized Tempered Stable distribution

We define the Multivariate Normalized Tempered Stable (MNTS) distribution in the following way.

Definition 4. *Let $0 < \kappa < 1$ and $\boldsymbol{\nu} = (\nu_1, \nu_2, \dots, \nu_{n+1})$ be a vector of positive real numbers. If $V_1, V_2, \dots, V_{n+1}$ are independent Tempered Stable random variables with $V_i \sim$*

TS$(\kappa, \frac{\nu_i}{\kappa}, 1)$ *and*

$$W_i = \frac{V_i}{V_1 + V_2 + \dots V_{n+1}}$$

then $\boldsymbol{W} = (W_1, W_2, \dots, W_n)$ *follows a Multivariate Normalized Tempered Stable distribution with parameters* $\boldsymbol{\nu}$ *and* κ *which we denote as MNTS*$(\nu_1, \nu_2, \dots, \nu_{n+1}; \kappa)$.

There are two special cases of this distribution. The Dirichlet distribution arises as $\kappa \to 0$ and the Normalized Inverse-Gaussian distribution arises if $\kappa = 1/2$:

$$\text{MNTS}(\nu_1, \nu_2, \dots, \nu_{n+1}; \kappa) \stackrel{\kappa \to 0}{\longrightarrow} \text{Dir}(\nu_1, \nu_2, \dots, \nu_{n+1})$$

in probability, and

$$\text{MNTS}(\nu_1, \nu_2, \dots, \nu_{n+1}; 1/2) \equiv \text{N-IG}(2\nu_1, 2\nu_2, \dots, 2\nu_{n+1}).$$

All the moments and cross-moments of the n-dimensional MNTS distribution exist (since the distribution is defined on the n dimensional unit simplex) and the following theorem gives their analytic form.

Theorem 1. *Suppose that* $\boldsymbol{W} = (W_1, W_2, \dots, W_n) \sim$ *MNTS*$(\nu_1, \nu_2, \dots, \nu_{n+1}; \kappa)$ *and let* $N_1, N_2, N \in \mathbb{N}$. *Then, defining* $S = \sum_{i=1}^{n+1} \nu_i$ *and* $\mu_i = \nu_i / S$, *we have that*

1. $E\left(W_i^N\right) = \sum_{l=1}^{N} \sum_{j=0}^{N-1} b_N(l, j)\, \Gamma\left(l - j/\kappa, \frac{S}{\kappa}\right)$

2. $E\left(W_i^{N_1} W_j^{N_2}\right) = \sum_{l=1}^{N_1} \sum_{m=1}^{N_2} \sum_{t=0}^{N_1+N_2-1} c_{N_1,N_2}(l, m, t)\, \Gamma\left(l + m - t/\kappa, \frac{S}{\kappa}\right)$

where

$$b_N(l, j) = \left(\begin{array}{c} N-1 \\ j \end{array}\right) \frac{(-1)^{N+j}\ (S/\kappa)^{j/\kappa}\ \exp\left\{\frac{S}{\kappa}\right\}\ d_N(\kappa, l)}{\Gamma(N)\ l!\kappa} \mu_i^l,$$

$$c_{N_1,N_2}(l, m, t) = \left(\begin{array}{c} N_1 + N_2 - 1 \\ t \end{array}\right) \frac{(-1)^{N_1+N_2+t}\ (S/\kappa)^{t/\kappa}\ \exp\left\{\frac{S}{\kappa}\right\}\ d_{N_1}(\kappa, l)\ d_{N_2}(\kappa, m)}{\Gamma(N_1 + N_2)\ l!\ m!\ \kappa} \mu_i^l \mu_j^m,$$

and $d_N(\kappa, l) = \sum_{i=1}^{l} \left(\begin{array}{c} l \\ i \end{array}\right) (-1)^i \prod_{c=0}^{N-1} (\kappa i - c).$

The proof is given in the appendix and extends the method of Lijoi *et al* (2005) and James *et al* (2006). The function $d_N(\kappa, l)$ defined above is related to the generalized Stirling numbers, or generalized factorial coefficients, $G(n,k,\sigma)$ (see, for example, Charalambides and Singh (1988) and Charalambides (2005)), through the simple formula

$$d_N(\kappa,l) = (-1)^N l! G(N,l,k).$$

The expressions for the moments are weighted sums of incomplete Gamma functions which will have negative arguments. It was necessary to use a symbolic language, such as Mathematica, to accurately evaluate the cross-moments when N_1 and N_2 becomes large (over, say, 10). The corresponding code is freely available from:
http://www.warwick.ac.uk/go/msteel/steel_homepage/software.

Corollary 1. *If* $\mathbf{W} = (W_1, W_2, \ldots, W_n) \sim$ *MNTS*$(\nu_1, \nu_2, \ldots, \nu_{n+1}; \kappa)$ *then*

$$E(W_i) = \mu_i,$$

$$\mathit{Var}(W_i) = (1-\kappa)\mu_i(1-\mu_i)\left[1 - \left(\frac{S}{\kappa}\right)^{1/\kappa} \exp\left\{\frac{S}{\kappa}\right\} \Gamma\left(1 - 1/\kappa, \frac{S}{\kappa}\right)\right],$$

$$\begin{aligned} \mathit{Cov}(W_i, W_j) &= \mu_i\mu_j\left[\kappa + \kappa\frac{S}{\kappa} - \kappa\exp\left\{\frac{S}{\kappa}\right\}\left(\frac{S}{\kappa}\right)^{1/\kappa}\Gamma\left(2 - 1/\kappa, \frac{S}{\kappa}\right) - 1\right] \\ &= \mu_i\mu_j(1-\kappa)\left[\exp\left\{\frac{S}{\kappa}\right\}\left(\frac{S}{\kappa}\right)^{1/\kappa}\Gamma\left(1 - 1/\kappa, \frac{S}{\kappa}\right) - 1\right], \end{aligned}$$

and

$$\mathit{Corr}(W_i, W_j) = -\sqrt{\frac{\mu_i}{1-\mu_i}\frac{\mu_j}{1-\mu_j}}.$$

The expectation of W_i and the correlation between W_i and W_j do not depend on κ and have the same form associated with the Dirichlet and the Normalized Inverse-Gaussian distributions. The form of the variance generalizes the form for the Dirichlet and Normalized Inverse-Gaussian distributions since the variance of W_i only depends on $\boldsymbol{\nu}$ through $S = \sum_{i=1}^n \nu_i$ and the mean $\mu_i = \nu_i / S$ and we can write $\text{Var}(W_i) = \alpha(\kappa, S)\mu_i(1-\mu_i)$ for some function α.

Applying the results of Theorem 1 to the special case of the Normalized Inverse-Gaussian distribution allows us to extend the results of Lijoi *et al* (2005) (as given in Proposition 2) to more general cross-moments:

Corollary 2. *Let* $\mathbf{W} = (W_1, W_2, \ldots, W_n) \sim$ *N-IG* $(\nu_1, \nu_2, \ldots, \nu_{n+1})$ *then, for* $N, N_1, N_2 \in \mathbb{N}$

1. $E\left[W_i^N\right] = \sum_{l=1}^{N} \sum_{j=0}^{N-1} b_N(l,j)\, \Gamma\left(l-2j, 2S\right)$

2. $E\left[W_i^{N_1} W_j^{N_2}\right] = \sum_{l=1}^{N_1} \sum_{m=1}^{N_2} \sum_{t=0}^{N_1+N_2-1} c_{N_1,N_2}(l,m,t)\, \Gamma\left(l+m-2t, 2S\right)$

where

$$b_N(l,j) = \left(\begin{array}{c} N-1 \\ j \end{array} \right) \frac{(-1)^{N+j}\ S^{2j}\ 2^{2j+1}\ \exp\{2S\}\ \ d_N(l)}{\Gamma(N)\ l!} \mu_i^l,$$

$$c_{N_1,N_2}(l,m,t) = \left(\begin{array}{c} N_1+N_2-1 \\ t \end{array} \right) \frac{(-1)^{N_1+N_2+t}\ S^{2t}\ 2^{2t+1}\ \exp\{2S\}\, d_{N_1}(l)\, d_{N_2}(m)}{\Gamma(N_1+N_2)\ l!\ m!}\ \mu_i^l\, \mu_j^m,$$

$$d_N(l) = \sum_{i=1}^{l} \left(\begin{array}{c} l \\ i \end{array} \right) (-1)^i \prod_{c=0}^{N-1} \left(\frac{i}{2} - c\right),\ S = \sum_{i=1}^{n+1} \nu_i \text{ and } \mu_i = \frac{\nu_i}{S}.$$

In the calculation of $d_N(l)$ above, we only need to calculate the additive terms for odd values of i, as the other terms will be 0.

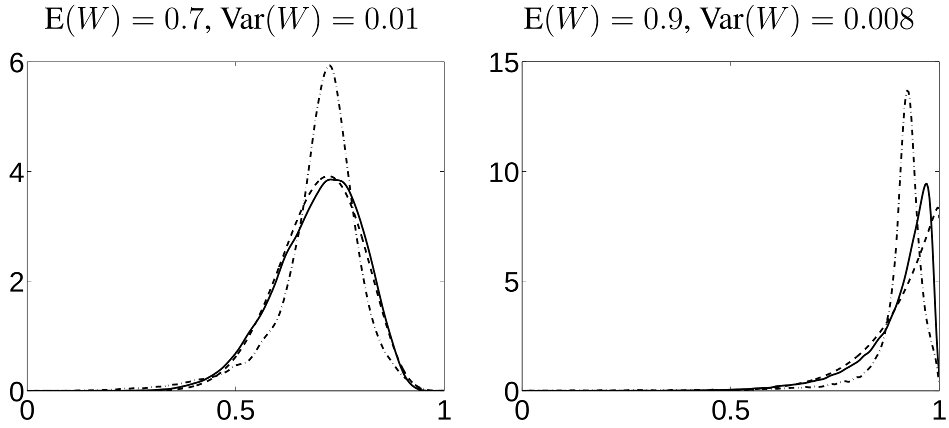

Figure 1: The densities of NTS distributions with the same mean and variance but different values of κ. In each graph the values are: $\kappa = 0$ (dashed line), $\kappa = 0.5$ (solid line), $\kappa = 0.9$ (dot-dashed line)

The univariate MNTS, which we term the Normalized Tempered Stable (with parameters ν_1, ν_2 and κ, written $\text{NTS}(\nu_1, \nu_2; \kappa)$), is an important case and we study its properties in detail. As $\kappa \to 0$, the distribution tends to a Beta distribution with parameters ν_1 and ν_2. As κ increases, the Tempered Stable distribution becomes heavier tailed and this carries over to the Normalized Tempered Stable distribution. This effect is illustrated in Figure 1. In both cases, as κ increases the distribution becomes more peaked around the modal value, with heavier tails. The effect is more pronounced when the mean of W is close to 1 (or 0). Figure 2 shows how the variance changes with κ (left), how the kurtosis changes with κ (middle) and the relationship between the two, for a NTS$(\nu, \nu; \kappa)$ distribution. Kurtosis is defined as the standardized fourth central moment:

$$\text{Kurt}(X) = \frac{\text{E}(X - \text{E}(X))^4}{(\text{Var}(X))^2}.$$

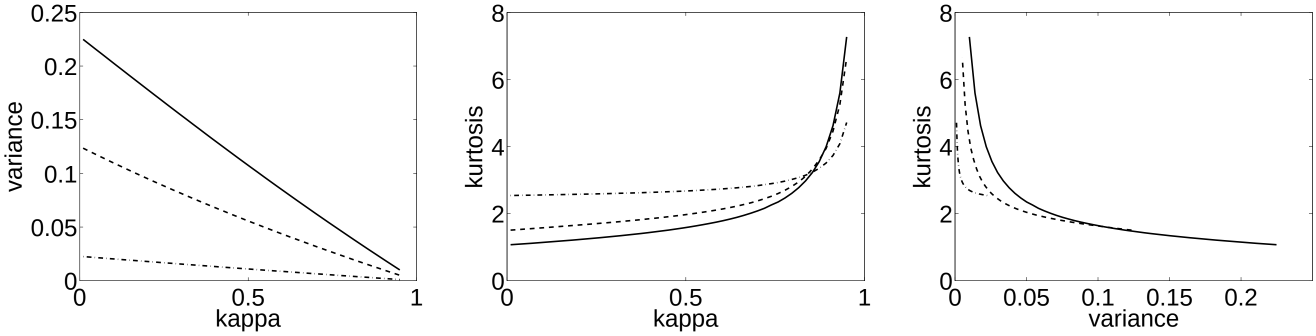

Figure 2: The Variance and kurtosis of NTS distribution with mean 0.5: (a) shows κ versus the variance, (b) shows κ versus the kurtosis and (c) shows variance versus kurtosis. In each graph: $S = 0.1$ (solid line), $S = 1$ (dashed line) and $S = 10$ (dot-dashed line).

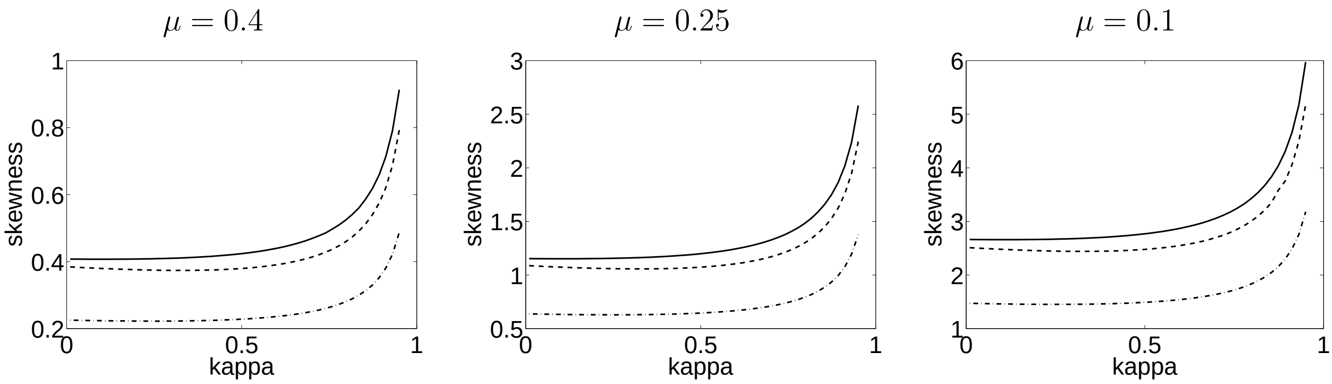

Figure 3: Skewness against κ for various values of the mean for some NTS distributions. In each graph: $S = 0.1$ (solid line), $S = 1$ (dashed line) and $S = 10$ (dot-dashed line).

This graph shows that the variance decreases as κ increases. For other values of the first moment $\mu = \frac{\nu_1}{\nu_1+\nu_2}$ the relationship between the variance and κ is the same but the variance becomes smaller as μ moves further from 1/2. Also note that, for any given κ, the variance decreases with $S = \nu_1 + \nu_2$ (in line with the Beta and the N-IG behaviour).

From the shape of the graph of kurtosis plotted against κ, one can see that as κ increases, the tails of the underlying TS distributions become heavier. There is a dramatic increase in kurtosis for values of κ greater than 0.8. The shape of this graph is preserved for all values of the other parameters, ν_1 and ν_2, and we note that the minimum kurtosis is not always achieved for the limiting case as $\kappa \to 0$, although the value at this limit is very close to the overall minimum. For $\mu \to 0$, the value of κ that gives the minimum value of kurtosis tends to 0.2, whereas for not very small values of μ, the case $\kappa \simeq 0$ seems to provide the smallest kurtosis. There is also symmetry around $\mu = 1/2$. The values of kurtosis increase as $|\mu - 1/2|$ increases, whereas for large values of κ, kurtosis decreases as $S = \nu_1 + \nu_2$ increases, and for small values of κ kurtosis increases as S increases. In other words, the range of possible kurtosis values decreases with S.

In the right graph in Figure 2 we see the relationship between the variance and the kurtosis for $\mu = 0.5$. The shape again is the same for other values of the parameters ν_1 and ν_2 and the graph is exactly the same for μ and $1 - \mu$. The Beta distribution corresponds to the point at the left end of the graph (i.e. for smallest variance).

Let skewness of a distribution denote its standardized third central moment, i.e.

$$\text{Skew}(X) = \frac{\text{E}(X - \text{E}(X))^3}{(\text{Var}(X))^{3/2}}.$$

If $\mu = 1/2$ the skewness is zero for all values of κ. Figure 3 shows the skewness against κ for various values of μ and S. We only plot the skewness $\text{Skew}(\mu, S, \kappa)$ for $\mu < 0.5$ since $\text{Skew}(\mu, S, \kappa) = -\text{Skew}(1 - \mu, S, \kappa)$ (which follows from the construction of the distribution). As the value of μ moves away from 1/2, the value of skewness also increases in absolute terms. On the other hand, when the value of $S = \nu_1 + \nu_2$ increases, skewness decreases in absolute value. Finally, note that, as for kurtosis, the minimum skewness (maximum, for $\mu > 1/2$) is not achieved for $\kappa \simeq 0$, but (usually) for some value between 0 and 0.6.

In Figure 4 we plot the relationship between skewness and variance (left) and kurtosis and skewness (right), for distributions with $\mu < 0.5$. The shape of the graph of skewness versus the variance does not change as ν_1 and ν_2 change and the Beta distribution corresponds to the point at the right end of the graph (i.e. with largest variance), whereas the minimum skewness is not necessarily found at the same point. The graph of kurtosis versus skewness is the most intriguing one. The reason for this is the little curl of the curve at its endpoint where we have the smallest values of kurtosis. This is caused by the fact that the minimum skewness is not achieved at the limiting case $\kappa = 0$, whereas it is for the kurtosis (except for very small values of μ). The endpoint of the graph for which we have minimum kurtosis corresponds to the Beta distribution. For cases with very small values for the mean,

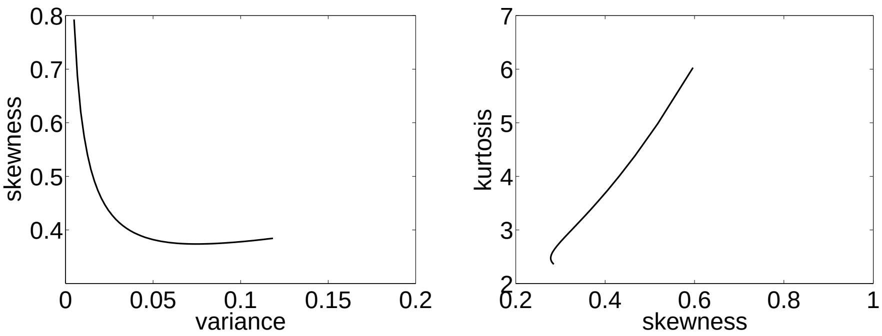

Figure 4: Skewness against variance and kurtosis against skewness for some NTS distributions

the same endpoint corresponds to the Beta distribution, but to neither minimum skewness, nor to minimum kurtosis and the graph curls for both those quantities (rather than only for the skewness, as in Figure 4).

4 The NTS-Binomial distributions

In this section, we consider using the Normalized Tempered Stable distribution as a mixing distribution in a Binomial model. This can be written as

$$X_i \sim \text{Bi}(n_i, P_i), \qquad P_i \sim \text{NTS}(\nu_1, \nu_2; \kappa).$$

The $\text{NTS}(\nu_1, \nu_2; \kappa)$ mixing distribution represents the heterogeneity in the probability of success across the different observed groups. The response can be written as the sum of n Bernoulli random variables, $X_i = \sum_{j=1}^{n_i} Z_{i,j}$ where $Z_{i,1}, Z_{i,2}, \ldots, Z_{i,n_i}$ are i.i.d. Bernoulli with success probability P_i. The intra-group correlation is defined as $\text{Corr}(Z_{i,k}, Z_{i,j})$ which in our model has the form

$$\rho = (1-\kappa)\left[1 - \left(\frac{S}{\kappa}\right)^{1/\kappa} \exp\left\{\frac{S}{\kappa}\right\} \Gamma\left(1 - 1/\kappa, \frac{S}{\kappa}\right)\right]. \tag{4.3}$$

where $S = \nu_1 + \nu_2$. The variance of X_i can be written as

$$\text{Var}(X_i) = n_i^2 \text{Var}(P_i) + n_i \text{E}[P_i(1-P_i)]$$

where $P_i \sim \text{NTS}(\nu_1, \nu_2; \kappa)$. The first term of this sum can be interpreted as the variance due to differences between individuals in the sample (between-subject variance) whereas the second term represents the intra-subject variability. In our model these have a simple form: $\text{Var}(P_i) = \mu(1-\mu)\rho$ and $\text{E}[P_i(1-P_i)] = \mu(1-\mu)(1-\rho)$, where $\mu = \text{E}(P_i) = \frac{\nu_1}{\nu_1+\nu_2}$.

The formulae for the moments derived in Theorem 1 can now be used to derive the likelihood for an observed sample $\mathbf{x} = (x_1, x_2, \ldots, x_n)$:

$$\begin{aligned}
f(\mathbf{x}) &= \int_0^1 \cdots \int_0^1 f(\mathbf{x}|p_1, \ldots, p_N) f(p_1, \ldots, p_N) dp_1 \ldots dp_N \\
&= \prod_{i=1}^{N} \int_0^1 f(x_i|p_i) f(p_i) dp_i \\
&= \prod_{i=1}^{N} \left(\begin{array}{c} n_i \\ x_i \end{array} \right) \mathrm{E}_{p_i} \left(p_i^{x_i} (1-p_i)^{n_i - x_i} \right)
\end{aligned}$$

where the expectation is given by the result in Theorem 1.

5 Examples

5.1 Simulated data

This simulation study considers the effectiveness of Maximum Likelihood estimation for the NTS-Binomial and compares its results with the Beta-Binomial model. We created 18 data sets by simulating from the NTS-Binomial model with different parameter values. The model is parameterized by (λ, μ, ρ) where $\lambda = \log \kappa - \log(1-\kappa)$, $\mu = \nu_1/(\nu_1 + \nu_2)$ and ρ is the intra-group correlation defined in (4.3). The number of observations for each unit n_i is set equal to a common value n. The NTS-Binomial and the Beta-Binomial models were fitted to each data set. The Beta-Binomial is parameterized as $\mu = \nu_1/(\nu_1 + \nu_2)$ and $\rho = 1/(\nu_1 + \nu_2 + 1)$ for which $\mathrm{E}[P_i] = \mu$ and $\mathrm{Var}[P_i] = \mu(1-\mu)\rho$ in a similar way to the NTS-Binomial. Table 1 shows the true parameter value and the maximum likelihood estimates for all data sets with (asymptotic) standard errors shown in parentheses using the NTS-Binomial model. The standard errors calculated using the (λ, μ, ρ)-parametrization were found to be more reliable than those using the original parameterisation for these data. The parameters μ and ρ are well-estimated in all cases. The parameter λ is harder to estimate but becomes increasingly better estimated as n or N are increased. The estimates of μ and ρ with the Beta-Binomial model are very similar to those with the NTS-Binomial model, illustrating robustness of these parameters to the model misspecification. However, the skewness and kurtosis are sensitive to the choice of models for some parameter values as shown in Table 2. The skewness (in absolute terms) and the kurtosis tend to be underestimated by the Beta-Binomial model and the problem becomes more pronounced as μ moves away from $1/2$. The standard errors for μ and ρ are only slightly larger for

data set		n	N	$\hat{\lambda}$	$\hat{\mu}$	$\hat{\rho}$
1	$\lambda = -0.69$,	6	500	-0.71 (*5.43*)	0.67 (*0.022*)	0.26 (*0.017*)
2	$\mu = 0.67$,		1000	-1.07 (*5.26*)	0.67 (*0.014*)	0.26 (*0.012*)
3	$\rho = 0.25$		1500	-0.71 (*3.34*)	0.67 (*0.013*)	0.24 (*0.0098*)
4		12	500	-0.75 (*5.17*)	0.67 (*0.041*)	0.26 (*0.023*)
5			1000	-0.29 (*3.11*)	0.68 (*0.040*)	0.24 (*0.018*)
6			1500	-0.47 (*2.82*)	0.67 (*0.029*)	0.26 (*0.012*)
7	$\lambda = 0.41$,	6	500	1.24 (*4.00*)	0.19 (*0.015*)	0.051 (*0.037*)
8	$\mu = 0.2$,		1000	0.16 (*0.51*)	0.20 (*0.016*)	0.052 (*0.030*)
9	$\rho = 0.062$		2000	-0.23 (*0.000002*)	0.20 (*0.0067*)	0.063 (*0.023*)
10		12	500	0.41 (*0.15*)	0.21 (*0.080*)	0.064 (*0.079*)
11			1000	0.40 (*0.00008*)	0.19 (*0.030*)	0.065 (*0.050*)
12			1500	0.41 (*0.0019*)	0.20 (*0.025*)	0.058 (*0.039*)
13	$\lambda = 1.39$,	6	500	0.50 (*16.67*)	0.31 (*0.016*)	0.044 (*0.021*)
14	$\mu = 0.33$,		1000	2.29 (*3.70*)	0.33 (*0.014*)	0.031 (*0.018*)
15	$\rho = 0.026$		2000	2.40 (*14.58*)	0.33 (*0.010*)	0.011 (*0.011*)
16		12	500	2.56 (*15.62*)	0.33 (*0.078*)	0.012 (*0.055*)
17			1000	1.50 (*15.62*)	0.33 (*0.053*)	0.027 (*0.028*)
18			1500	1.73 (*6.13*)	0.33 (*0.047*)	0.026 (*0.033*)

Table 1: Maximum likelihood estimates for λ, μ, and ρ for the NTS-Binomial model with the simulated data sets.

the NTS-Biomial model compared with the Beta-Binomial model (results not shown). The

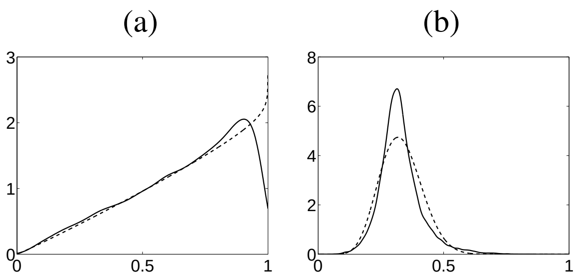

Figure 5: Density of the mixing distribution for the NTS-Binomial (solid line) and Beta-Binomial (dashed line) models evaluated at the maximum likelihood estimates for: a) $\lambda = -0.69$, $\mu = 0.67$, $\rho = 0.25$, $n = 6$, $N = 1500$ and b) $\lambda = 1.39$, $\mu = 0.33$, $\rho = 0.026$, $n = 12$, $N = 1000$

	n	N	True value Skew	True value Kurt	NTS-Binomial Skew	NTS-Binomial Kurt	Beta-Binomial Skew	Beta-Binomial Kurt
$\lambda = -0.69$,	6	500	-0.617	2.39	-0.634	2.38	-0.572	2.38
$\mu = 0.67$,		1000			-0.620	2.43	-0.577	2.51
$\rho = 0.25$		1500			-0.622	2.41	-0.608	2.46
	12	500			-0.643	2.39	-0.591	2.41
		1000			-0.692	2.52	-0.587	2.45
		1500			-0.643	2.41	-0.567	2.41
$\lambda = 0.41$,	6	500	1.07	4.26	1.470	6.04	0.521	3.15
$\mu = 0.2$,		1000			0.922	3.89	0.721	3.38
$\rho = 0.062$		2000			0.916	3.79	0.707	3.38
	12	500			1.040	4.14	0.681	3.32
		1000			1.070	4.31	0.676	3.38
		1500			1.050	4.21	0.678	3.35
$\lambda = 1.39$,	6	500	0.571	3.52	0.516	3.08	0.180	2.96
$\mu = 0.33$,		1000			1.070	6.02	0.249	2.92
$\rho = 0.026$		2000			0.803	5.19	0.143	2.97
	12	500			0.959	6.10	0.152	2.97
		1000			0.625	3.71	0.229	2.93
		1500			0.698	4.04	0.224	2.92

Table 2: The skewness and kurtosis evaluated at the maximum likelihood estimate for the NTS-Binomial and Beta-Binomial models with the simulated data sets.

difference between the two models can be illustrated by plotting the estimated mixing distribution. The density function of the NTS distribution is not available analytically and so a kernel density estimate is calculated from 50 000 random variates simulated from the Tempered Stable distribution using the R code of Palmer *et al* (2008b). Figure 5 displays these densities for data sets 3 and 17, respectively. In both cases, the NTS-Binomial distribution is able to capture the shape of the simulated data. The Beta-Binomial distribution is able to capture the general shape of the distribution but not some features. For example, the densities are very different for values close to 1 for data set 3 and the mass around the mode is very different for data set 17.

5.2 Mice fetal mortality data

Brooks *et al* (1997) analyze six data sets of fetal mortality in mouse litters. The data sets are: E1, E2, HS1, HS2, HS3 and AVSS. E1 and E2 were created by pooling smaller data sets used by James and Smith (1982), HS1, HS2 and HS3 were introduced by Haseman and Soares (1976) and the AVSS data set was first described by Aeschbacher *et al* (1977). In each data set, the data were more dispersed than under a Binomial distribution. Brooks *et al* (1997) show that finite mixture models fit the data better than the standard Beta-Binomial model in all data sets except AVSS. Here, we fit the NTS-Binomial and its special case when $\kappa = 1/2$, the N-IG-Binomial model, as alternatives. Several other models have been applied to these data including the shared response model of Pang and Kuk (2005) and the q-power distribution of Kuk (2004). We use the data with the correction described in Garren *et al* (2001).

Data	N	$\hat{\lambda}$	$\hat{\mu}$	$\hat{\rho}$
E1	205	1.036 (*4.152*)	0.089 (*0.033*)	0.079 (*0.130*)
E2	211	1.062 (*2.862*)	0.110 (*0.044*)	0.116 (*0.122*)
HS1	524	0.979 (*2.767*)	0.090 (*0.023*)	0.080 (*0.091*)
HS2	1328	1.600 (*2.328*)	0.108 (*0.0090*)	0.040 (*0.032*)
HS3	554	1.411 (*1.793*)	0.074 (*0.018*)	0.105 (*0.083*)
AVSS	127	$-\infty$	0.069 (*0.038*)	0.0059 (*0.137*)

Table 3: Maximum likelihood estimates with standard errors for the parameters in NTS-Binomial model for the six mice fetal mortality data sets.

Table 3 shows the maximum likelihood estimates of the parameters of the NTS-Binomial models with their related asymptotic standard errors (shown in parentheses). In all data sets except AVSS the estimate of λ is substantially different from $\lambda = -\infty$ (which corresponds to the Beta-Binomial case). In the other data sets λ is estimated to be between 0.979 and 1.6 which correspond to κ values between 0.74 and 0.83. This suggests that the estimated mixing distributions are substantially different from a Beta distribution. In fact the tails of the distribution are much heavier than those defined by a Beta distribution with the same mean and variance. The estimated mixing distributions are shown in Figure 6 with the mixing distribution for the Beta-Binomial distribution (the AVSS data set is not included since the estimates for NTS-Binomial and Beta-Binomial models imply the same mixing distribution).

Estimated skewness and kurtosis for the two distributions are shown in Table 4. The

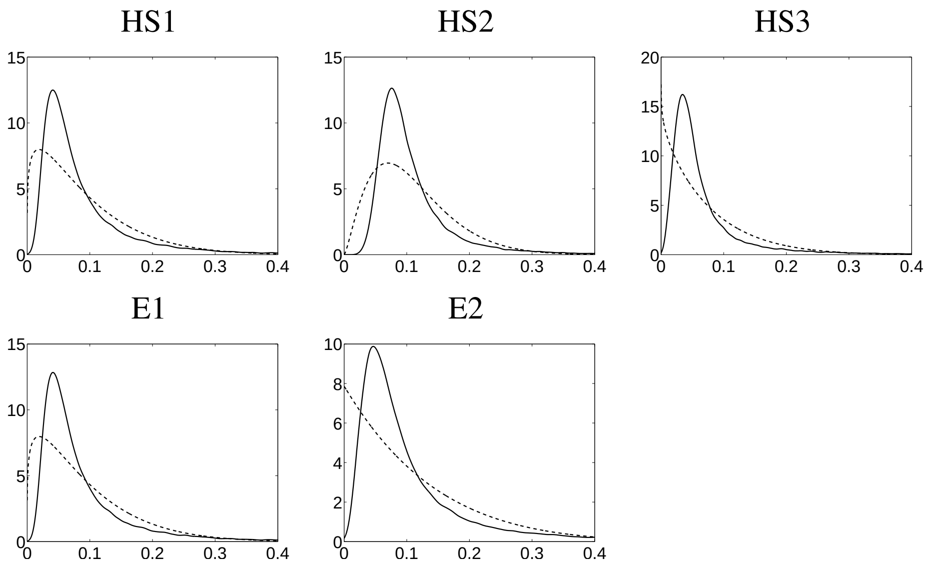

Figure 6: Density of the mixing distribution for the NTS-Binomial (solid line) and Beta-Binomial (dashed line) models evaluated at the maximum likelihood estimates for the mice fetal mortality data set

	NTS-Binomial		Beta-Binomial	
	Skew	Kurt	Skew	Kurt
E1	2.75	13.26	1.41	2.42
E2	2.66	11.99	1.42	2.25
HS1	2.67	12.62	1.41	2.42
HS2	2.66	13.57	1.00	1.21
HS3	3.89	22.93	1.65	3.46

Table 4: Estimates of the skewness and kurtosis of the mixing distributions for the Beta-Binomial and NTS-Binomial distributions

third and especially fourth moments are much larger for the NTS-Binomial model which confirms the interpretation of κ given in Section 3.

The maximum likelihood values for the NTS-Binomial model are substantially better than those for the Beta-Binomial model. In addition, we compare the NTS-Binomial model to the main competitors in the literature: the finite mixture models of Brooks *et al* (1997), the shared response model of Pang and Kuk (2005), the Correlated-Binomial of Altham (1978) and Kupper and Haseman (1978), the Beta-Correlated Binomial of Paul (1985) and the q-power distribution of Kuk (2004). We also consider the normalized Inverse-Gaussian distribution, which fixes $\kappa = 1/2$ in the NTS-Binomial model. The Akaike Information

Criterion (AIC) and the Bayesian Information Criterion (BIC) are used as measures of fit. If L is the maximum likelihood value, n is the data size and k is the number of parameters, the AIC $= -2\log L + 2k$ and the BIC $= -2\log L + k\log n$.

Model	E1	E2	AVSS	HS1	HS2	HS3
Beta-Binomial	+8.3	+6.0	**341.9**	+9.3	+43.9	+32.8
NTS-Binomial	+4.8	+1.1	+2.0	+1.2	+24.1	+3.4
N-IG-Binomial	+5.1	+0.6	+0.5	+3.1	+38.9	+18.8
B-B/B mixture	+3.5	+0.4	+3.8	**1550.3**	**3274.7**	**1373.9**
2-d binom. mixture	+1.7	+0.6	+1.9	+20.6	+20.1	+4.6
3-d binom. mixture	+5.1	+1.9	+5.7	+1.1	+4.1	+1.8
Best binom. mixture	+5.1	+1.9	+9.7	+4.5	+1.7	+5.5
Shared response	**563.1**	**687.8**	+6.1	+20.3	+42.5	+8.9
q-power	+6.6	+8.4	+8.0	+6.5	+2.1	+0.5
Correlated-Binomial	+26.2	+36.7	+1.2	+57.0	+67.3	+94.5
B-C-B	+7.8	+1.3	+2.0	+8.4	+35.7	+24.0

Table 5: AIC values for the competing models for each data set. The smallest value for each data set is shown in bold and other AIC values are shown as differences from that minimum.

Results for each data set are given in Table 5 (for AIC) and Table 6 (for BIC). The best model has the smallest value of the information criterion. In both tables, the smallest value of AIC/BIC for each of the data sets are given in bold and the values for the other models are given as differences from the best model. In the tables, B-B/B mixture is the Beta-Binomial/Binomial mixture (i.e. a mixture consisting of a Beta-Binomial part and a Binomial part), 2-d/3-d correspond to mixtures of two or three Binomials respectively, B-C-B is the Beta-Correlated-Binomial model, and N-IG is the normalised Inverse-Gaussian distribution. Finally, the best Binomial mixture is a mixture of Binomials with the number of components unknown. The number of components to be fitted is derived using the program C.A.MAN, using directional derivative methods (see Bohning *et al* (1992)). We also found that the log likelihood value given by Pang and Kuk (2005) for E1 was not consistent with their estimates. Their value seems to correspond to the data set given by Brooks *et al* (1997) rather than the corrected version given by Garren *et al* (2001).

The NTS-Binomial model performs consistently well for all six data sets, especially in terms of BIC. The N-IG-Binomial model also performs quite well (although it does worse for HS2 and HS3). For the E1 and E2 data, the shared response model has the lowest AIC and BIC values. However, the N-IG-Binomial model has a very similar value. For the

Model	E1	E2	AVSS	HS1	HS2	HS3
Beta-Binomial	+8.3	+6.0	**347.6**	+6.2	+41.9	+32.3
NTS-Binomial	+8.1	+4.5	+4.9	+2.4	+27.2	+7.2
N-IG-Binomial	+5.1	+0.6	+0.5	**1561.9**	+36.8	+18.3
B-B/B mixture	+10.2	+4.6	+9.5	+5.4	+8.3	+8.2
2-d binom. mixture	+5.0	+4.0	+4.8	+21.7	+23.2	+8.4
3-d binom. mixture	+15.1	+12.0	+14.2	+10.8	+17.5	+14.3
Best binom. mixture	+15.1	+12.0	+23.9	+22.7	+25.6	+26.6
Shared response	**569.7**	**694.5**	+6.1	+17.2	+40.4	+8.4
q-power	+6.6	+7.9	+9.0	+3.4	**3287.2**	**1383.0**
Correlated-Binomial	+26.2	+36.7	+1.2	+53.9	+65.2	+94.0
B-C-B	+11.1	+4.6	+4.9	+9.6	+38.8	+27.8

Table 6: BIC values for the competing models for each data set. The smallest value for each data set are shown in bold and other BIC values are shown as differences from that minimum.

AVSS data, the Beta-Binomial model is the best model with the N-IG-Binomial coming a close second. Considering the HS1-HS3 data sets, the differences between the criterion values of the different models are larger than in the other data sets, due to the much larger data size. For HS1, the N-IG-Binomial model is actually the best in terms of the BIC, whereas the B-B/B mixture performs best in terms of the AIC. The value of AIC for the N-IG-Binomial model and the value of BIC for the B-B/B mixture are, as one would expect, close to the smallest values. The NTS-Binomial model is the second best in terms of BIC and third (but also very close to the second) in terms of AIC. The best model for the HS2 and HS3 data is the B-B/B mixture in terms of AIC and the q-power model in terms of BIC. For HS3, the NTS-Binomial is second best in terms of BIC and fourth in terms of AIC.

6 Discussion

This paper has introduced a new class of distributions for random variables on the unit simplex by normalising Tempered Stable distributions. The resulting Multivariate Normalized Tempered Stable (MNTS) distribution is a natural generalisation of the Dirichlet distribution and many properties carry over. This distribution only involves a single extra parameter which can be clearly linked to the heaviness of the tails. The cross-moments can be calculated analytically which allows easy likelihood inference for an NTS-Binomial model, which is the Binomial model mixed with an univariate MNTS distribution on the success probability. In our examples the NTS-Binomial outperforms the Beta-Binomial and is competitive with other previously proposed models.

References

Abramowitz, M. and Stegun, I. A. (1964): "Handbook of mathematical functions with formulas, graphs, and mathematical tables, Volume 55 of National Bureau of Standards Applied Mathematics Series,"

Aeschbacher, H. U., Vuataz, L., Sotek, J., and Stalder, R. (1977): "The use of the Beta-Binomial distribution in dominant-lethal testing for weak mutagenic activity (Part 1)," *Mutation Research*, 44, 369-390.

Altham, P. M. E. (1978): "Two generalizations of the Binomial distribution," *Applied Statistics*, 27, 162-167.

Billingsley, P. (1995): "Probability and measure," John Wiley & Sons: New York.

Bohning, D., Schlattmann, P. and Lindsay, B. (1992): "Computer-Assisted Analysis of Mixtures (C.A.MAN): Statistical Algorithms," *Biometrics*, 48, 283-303.

Brix, A. (1999): "Generalized Gamma measures and shot-noise Cox processes," *Advances in Applied Probability*, 31, 929-953.

Brooks, S. P. (2001): "On Bayesian analyses and finite mixtures for proportions," *Statistics and Computing*, 11, 179-190.

Brooks, S. P., Morgan, B. J. T., Ridout, M. S. and Pack, S. E. (1997): "Finite Mixture Models for Proportions," *Biometrics*, 53, 1097-1115.

Charalambides, Ch. A. (2005): "Combinatorial methods in discrete distributions" John Wiley & Sons: Hoboken, New Jersey.

Charalambides, Ch. A. and Singh, J. (1988): "A review of the Stirling numbers, their generalizations and statistical applications," *Communications in Statistics-Theory and Methods*, 17, 2533-2595.

Feller, W. (1971): "An introduction to probability theory and it applications," John Wiley & Sons: New York.

Garren, S. T., Smith, R. L. and Piegorsch, W. (2001): "Bootstrap goodness-of-fit test for the Beta-Binomial model," *Journal of Applied Statistics*, 28, 561-571.

George, E. O. and Bowman, D. (1995): “A full likelihood procedure for analysing exchangeable binary data,” *Biometrics*, 51, 512-523.

Gradshteyn, I. S. and Ryzhik, I. M. (1994): “Table of integrals, series and products,” Academic Press: Boston.

Haseman, J.K . and Soares, E. R. (1976): “The distribution of fetal death in control mice and its implications on statistical tests for dominant lethal effects,” *Mutation Research*, 41, 272-288.

Hougaard, P. (1986): “Survival models for heterogeneous populations derived from stable distributions,” *Biometrika*, 73, 387-396.

James, D. A. and Smith, D. M. (1982): “Analysis of results from a collaborative study of the dominant lethal assay,” *Mutation Research*, 97, 303-314

James, L. F., Lijoi, A. and Prünster, I. (2006): “Conjugacy as a distinctive feature of the Dirichlet process,” *Scandinavian Journal of Statistics*, 33, 105-120.

Johnson, W. P. (2002): “The curious history of Faá di Bruno’s formula,” *American Mathmetical Monthly*, 109, 217-234.

Jorgensen, B. (1987): “Exponential dispersion models (with discussion),” *Journal of the Royal Statistical Society B*, 49, 127-162.

Kuk, A. Y. C. (2004): “A litter-based approach to risk assessment in developmental toxicity studies via a power family of completely monotone functions,” *Journal of the Royal Statistical Society. Series C.*, 53, 369-386.

Kupper, L. L. and Haseman, J. K. (1978): “The use of a correlated Binomial model for the analysis of certain toxicological experiments,” *Biometrics*, 35, 281-293.

Lijoi, A., Mena, R. H. and Prünster, I. (2005): “Hierarchical mixture modeling with normalized inverse-Gaussian priors,” *Journal of the American Statistical Association*, 100, 1278-1291.

Ochi, Y. and Prentice, R. L. (1984): “Likelihood inference in a correlated probit regression models,” *Biometrika*, 71, 531-543.

Palmer, K. J., Ridout, M. S. and Morgan, B. J. T. (2008a): “Modelling cell generation times using the Tempered Stable distribution,” *Journal of the Royal Statistical Society Series C*, 57, 379-397.

Palmer, K. J., Ridout, M. S. and Morgan, B. J. T (2008b): R functions for the Tempered Stable distribution (R code). Available from
http://www.kent.ac.uk/IMS/personal/msr/TemperedStable.html

Pang, Z. and Kuk, A. Y. C. (2005): "A shared response model for clustered binary data in developmental toxicity studies," *Biometrics*, 61, 1076-1084.

Paul, S. R. (1985): "A three-parameter generalisation of the Binomial distribution," *Communications in Statistics Theory and Methods*, 14(6), 1497-1506.

Paul, S. R. (1987): "On the Beta-Correlated Binomial (BCB) distribution in a three parameter generalization of the Binomial distribution," *Communications in Statistics Theory and Methods*, 16(5), 1473-1478.

Tweedie, M. (1984): "An index which distinguishes between some important exponential families," in *Statistics: Applications and New Directions: Proceedings of the Indian statistical Institute Golden Jubilee International Conference*, Eds: J. Ghosh and J. Roy, 579-604.

Williams, D. A. (1982): "Extra-Binomial variation in logistic linear models," *Journal of the Royal Statistical Society C*, 31, 144-148.

Zhang, S. and Jin, J. (1996): "Computation of Special Functions," John Wiley & Sons: New York.

A Proofs

A.1 Proof of Theorem 1, part 1

We exploit the representation of the Normalized Tempered Stable distribution through Tempered Stable random variables. If $\mathbf{W} \sim \text{MNTS}(\nu_1, \nu_2, \ldots, \nu_{n+1}; \kappa)$ then we can write $W_i = \frac{V_i}{V}$ where $V = \sum_{j=1}^{n+1} V_j$, $V_1, V_2, \ldots, V_{n+1}$ are independent and $V_j \sim \text{TS}(\kappa, \frac{\nu_j}{\kappa}, 1)$. Then, we have:

$$
\begin{aligned}
\mathrm{E}\left[W_i^N\right] &= \mathrm{E}\left[\frac{V_i^N}{V^N}\right] \\
&= \mathrm{E}\left[\frac{V_i^N}{\Gamma(N)}\int_0^\infty u^{N-1}\exp\{-uV\}\,du\right] \\
&= \frac{1}{\Gamma(N)}\int_0^\infty u^{N-1}\prod_{j\neq i}\mathrm{E}\left[\exp\{-uV_j\}\right]\,\mathrm{E}\left[V_i^N\exp\{-uV_i\}\right]\,du \qquad \text{(A.4)} \\
&= \frac{(-1)^N}{\Gamma(N)}\int_0^\infty u^{N-1}\prod_{j\neq i}\mathrm{E}\left[\exp\{-uV_j\}\right]\,\mathrm{E}\left[\frac{\partial^N}{\partial u^N}\exp\{-uV_i\}\right]\,du \\
&= \frac{(-1)^N}{\Gamma(N)}\int_0^\infty u^{N-1}\prod_{j\neq i}\mathrm{E}\left[\exp\{-uV_j\}\right]\frac{\partial^N}{\partial u^N}\,\mathrm{E}\left[\exp\{-uV_i\}\right]\,du \qquad \text{(A.5)} \\
&= \frac{(-1)^N}{\Gamma(N)}\exp\left\{\frac{S}{\kappa}\right\}\int_0^\infty u^{N-1}\,\exp\left\{-\sum_{j\neq i}\frac{\nu_j}{\kappa}(1+2u)^\kappa\right\}\underbrace{\frac{\partial^N}{\partial u^N}\exp\left\{-\frac{\nu_i}{\kappa}(1+2u)^\kappa\right\}}_{(1)}du \qquad \text{(A.6)}
\end{aligned}
$$

(A.3) is an application of the Fubini Theorem and (A.4) is an application of Theorem (16.8) in Billingsley (1995). Finally, for (A.5) we used the know form for the moment generating function of the Tempered Stable distribution.

The difficulty is to calculate (1), i.e. the N-th derivative of the function $\exp\left\{-\frac{\nu_i}{\kappa}(1+2u)^\kappa\right\}$. This is possible using Meyer's formula, which is a variation of Faa di Bruno's formula (see, for example, Johnson *et al* (2002)): if f and g are functions with sufficient derivatives then

$$
\frac{\partial^N}{\partial u^N}(g\circ f)(x) = \frac{\partial^N}{\partial u^N}\left[g(f(u))\right] = \sum_{l=0}^N \frac{g^{(l)}(f(u))}{l!}\left\{\frac{\partial^N}{\partial h^N}\left[f(u+h)-f(u)\right]^l\Big|_{h=0}\right\}
$$

In this case, $g(x) = \exp\{x\}$ and $f(x) = -\frac{\nu_i}{\kappa}(1+2x)^\kappa$. We first consider:

$$
\left[f(x+h)-f(x)\right]^l = \frac{\nu_i^l}{\kappa^l}\left[(1+2x)^\kappa - (1+2x+2h)^\kappa\right]^l
$$

and the derivative can be seen to be

$$\frac{\partial^N}{\partial h^N}\left[f(x+h)-f(x)\right]^l = \frac{\nu_i^l}{\kappa^l}\sum_{i=0}^{l}\begin{pmatrix} l \\ i \end{pmatrix}(-1)^i(1+2x)^{\kappa(l-i)}\left[\frac{\partial^N}{\partial h^N}(1+2x+2h)^{\kappa i}\right]$$

The derivative can be expressed as

$$\frac{\partial^N}{\partial h^N}(1+2x+2h)^{\kappa i} = 2^N\kappa i(\kappa i-1)\cdots(\kappa i-N+1)(1+2x+2h)^{\kappa i-N}$$

So, at $h=0$, we get:

$$\begin{aligned}\left.\frac{\partial^N}{\partial h^N}[f(x+h)-f(x)]^l\right|_{h=0} &= 2^N\frac{\nu_i^l}{\kappa^l}\sum_{i=1}^{l}\begin{pmatrix} l \\ i \end{pmatrix}(-1)^i(1+2x)^{\kappa(l-i)}\prod_{c=0}^{N-1}(\kappa i-c)\left.(1+2x+2h)^{\kappa i-N}\right|_{h=0} \\ &= 2^N\frac{\nu_i^l}{\kappa^l}\sum_{i=1}^{l}\begin{pmatrix} l \\ i \end{pmatrix}(-1)^i(1+2x)^{\kappa l-N}\prod_{c=0}^{N-1}(\kappa i-c)\end{aligned}$$

By plugging the above in Meyer's formula, and noting that $g^{(k)}(f(x)) = g(f(x)) = \exp\left\{-\frac{\nu_i}{\kappa}(1+2x)^\kappa\right\}$, we can see that

$$\begin{aligned}\frac{\partial^N}{\partial u^N}\exp\left\{-\frac{\nu_i}{\kappa}(1+2u)^\kappa\right\} &= 2^N\sum_{l=1}^{N}\frac{\exp\left\{-\frac{\nu_i}{\kappa}(1+2u)^\kappa\right\}}{l!}\frac{\nu_i^l}{\kappa^l}\sum_{j=1}^{l}\begin{pmatrix} l \\ j \end{pmatrix}(-1)^j(1+2u)^{\kappa l-N}\prod_{c=0}^{N-1}(\kappa j-c) \\ &= 2^N\sum_{l=1}^{N}\frac{\exp\left\{-\frac{\nu_i}{\kappa}(1+2u)^\kappa\right\}}{l!}\frac{\nu_i^l}{\kappa^l}(1+2u)^{\kappa l-N}d_N(\kappa,l)\end{aligned} \tag{A.7}$$

where

$$d_N(\kappa,l) = \sum_{j=1}^{l}\begin{pmatrix} l \\ j \end{pmatrix}(-1)^j\prod_{c=0}^{N-1}(\kappa j-c) = \sum_{j=1}^{l}\begin{pmatrix} l \\ j \end{pmatrix}(-1)^j\frac{\Gamma(\kappa j+1)}{\Gamma(\kappa j-N+1)}$$

It is now straightforward to verify that

$$\mathrm{E}\left(W_i^N\right) = \frac{(-1)^N2^N\exp\left\{\frac{S}{\kappa}\right\}}{\Gamma(N)}\sum_{l=1}^{N}\frac{\frac{\nu_i^l}{\kappa^l}d_N(\kappa,l)}{l!}\int_0^\infty u^{N-1}\exp\left\{-\frac{S}{\kappa}(1+2u)^\kappa\right\}(1{+}2u)^{\kappa l-N}du.$$

The integral in the last expression can be simplified using the substitution $y=(1+2u)^\kappa$ and the Binomial theorem:

$$\int_0^\infty u^{N-1}\exp\left\{-\frac{S}{\kappa}(1+2u)^\kappa\right\}(1{+}2u)^{\kappa l-N}du = \frac{1}{2^N\kappa\frac{S^l}{\kappa^l}}\sum_{j=0}^{N-1}\begin{pmatrix} N-1 \\ j \end{pmatrix}(-1)^j\frac{S^{j/\kappa}}{\kappa^{j/\kappa}}\Gamma\left(l-j/\kappa,\frac{S}{\kappa}\right)$$

and therefore

$$\mathrm{E}\left(W_i^N\right) = \sum_{l=1}^{N}\sum_{j=0}^{N-1}\begin{pmatrix} N-1 \\ j \end{pmatrix}\frac{(-1)^{N+j}\exp\left\{\frac{S}{\kappa}\right\}d_N(\kappa,l)}{\Gamma(N)\,l!\kappa}\left(\frac{S}{\kappa}\right)^{j/\kappa}\left(\frac{\nu_i}{S}\right)^l\Gamma\left(l-j/\kappa,\frac{S}{\kappa}\right).\square$$

A.2 Proof of Theorem 1, part 2

For the cross-moments of the MNTS distribution, we have:

$$
\begin{aligned}
\mathrm{E}\left[W_i^{N_1}W_j^{N_2}\right] =& \mathrm{E}\left[\frac{V_i^{N_1}V_j^{N_2}}{V^N}\right], \text{ where } N = N_1 + N_2 \\
=& \mathrm{E}\left[\frac{V_i^{N_1}V_j^{N_2}}{\Gamma(N)}\int_0^\infty u^{N-1}\ \exp\{-uV\}\,du\right] \\
=& \frac{(-1)^N}{\Gamma(N)}\int_0^\infty u^{N-1}\ \mathrm{E}\left[\frac{\partial^{N_1}}{\partial u^{N_1}}\exp\{-uV_i\}\right]\ \mathrm{E}\left[\frac{\partial^{N_2}}{\partial u^{N_2}}\exp\{-uV_j\}\right]\prod_{t\neq i,j}\mathrm{E}\left[\exp\{-uV_t\}\right]du \\
=& \frac{(-1)^N}{\Gamma(N)}\int_0^\infty u^{N-1}\ \frac{\partial^{N_1}}{\partial u^{N_1}}\mathrm{E}\left[\exp\{-uV_i\}\right]\ \frac{\partial^{N_2}}{\partial u^{N_2}}\mathrm{E}\left[\exp\{-uV_j\}\right]\ \prod_{t\neq i,j}\mathrm{E}\left[\exp\{-uV_t\}\right]du \\
=& \frac{(-1)^N}{\Gamma(N)}\exp\left\{\frac{S}{\kappa}\right\}\int_0^\infty u^{N-1}\ \frac{\partial^{N_1}}{\partial u^{N_1}}\left(\exp\left\{-\frac{\nu_i}{\kappa}(1+2u)^\kappa\right\}\right) \\
&\times\frac{\partial^{N_2}}{\partial u^{N_2}}\left(\exp\left\{-\frac{\nu_j}{\kappa}(1+2u)^\kappa\right\}\right)\exp\left\{-\sum_{t\neq i,j}\frac{\nu_t}{\kappa}(1+2u)^\kappa\right\}du
\end{aligned}
$$

Using the result for the $N-$th derivative of the function $\exp\left\{-\frac{\nu_i}{\kappa}(1+2u)^\kappa\right\}$ in (A.6), we find:

$$
\mathrm{E}\left(W_i^{N_1}W_j^{N_2}\right) = \frac{(-1)^N}{\Gamma(N)}\exp\left\{\frac{S}{\kappa}\right\}2^N\sum_{l=1}^{N_1}\sum_{m=1}^{N_2}\frac{d_{N_1}(\kappa,l)d_{N_2}(\kappa,m)}{l!m!}\frac{\nu_i^l}{\kappa^l}\frac{\nu_j^m}{\kappa^m}I^*_{l,m}(\kappa)
$$

where

$$
I^*_{l,m}(\kappa) = \int_0^\infty u^{N-1}\ \exp\left\{-\frac{S}{\kappa}(1+2u)^\kappa\right\}(1+2u)^{\kappa(l+m)-N}\ du.
$$

Using the same substitution as above, $y = (1+2u)^\kappa$, in $I^*_{l,m}(\kappa)$, together with the Binomial theorem, we find:

$$
\mathrm{E}\left(W_i^{N_1}W_j^{N_2}\right) = \sum_{l=1}^{N_1}\sum_{m=1}^{N_2}\sum_{t=0}^{N_1+N_2-1}c_{N_1,N_2}(l,m,t)\Gamma\left(l+m-t/\kappa,\frac{S}{\kappa}\right) \qquad \text{(A.8)}
$$

where

$$
c_{N_1,N_2}(l,m,t) = \begin{pmatrix} N_1+N_2-1 \\ t\end{pmatrix}\frac{(-1)^{N_1+N_2+t}\ (S/\kappa)^{t/\kappa}\ \exp\left\{\frac{S}{\kappa}\right\}\ d_{N_1}(\kappa,l)\ d_{N_2}(\kappa,m)}{\Gamma(N_1+N_2)\ l!\ m!\ \kappa}\mu_i^l\mu_j^m. \quad \square
$$

A.3 Proof of Corollary 1

For the first moment we have:

$$
b_1(1,0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}\frac{(-1)^1\exp\left\{\frac{S}{\kappa}\right\}\nu_i^1}{\Gamma(1)1!\kappa^{1+0/\kappa}\ S^{1-0/\kappa}}(-\kappa),\ \text{and } d_1(\kappa,1) = -\kappa \Rightarrow b_1(1,0) = \exp\left\{\frac{S}{\kappa}\right\}\frac{\nu_i}{S}.
$$

The result follows from noting that $\Gamma\left(1-0/\kappa, \frac{S}{\kappa}\right) = \Gamma\left(1, \frac{S}{\kappa}\right) = \int_{\frac{S}{\kappa}}^{\infty} \exp\{-t\}dt = \exp\left\{-\frac{S}{\kappa}\right\}$.

The second moment is

$$\mathrm{E}\left[W_i^2\right] = -(1-\kappa)\mu_i(1-\mu_i)\frac{S^{1/\kappa}}{\kappa^{1/\kappa}}\exp\left\{\frac{S}{\kappa}\right\}\Gamma\left(1-1/\kappa, \frac{S}{\kappa}\right) + \mu_i(1-\kappa+\mu_i\kappa)$$

since $d_2(\kappa, 1) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} (-1)^1(\kappa-0)(\kappa-1) = \kappa(1-\kappa)$ and $d_2(\kappa, 2) = 2\kappa^2$, which implies that $b_2(1,0) = (1-\kappa)\exp\left\{\frac{S}{\kappa}\right\}\mu_i$, $b_2(1,1) = -(1-\kappa)\exp\left\{\frac{S}{\kappa}\right\}\mu_i\frac{S^{1/\kappa}}{\kappa^{1/\kappa}}$, $b_2(2,0) = \kappa\exp\left\{\frac{S}{\kappa}\right\}\mu_i^2$ and $b_2(2,1) = -\kappa\exp\left\{\frac{S}{\kappa}\right\}\mu_i^2\frac{S^{1/\kappa}}{\kappa^{1/\kappa}}$. The result follows from noting that $\Gamma\left(1-0/\kappa, \frac{S}{\kappa}\right) = \exp\left\{-\frac{S}{\kappa}\right\}$ and $\Gamma\left(2-0/\kappa, \frac{S}{\kappa}\right) = \exp\left\{-\frac{S}{\kappa}\right\}\left(1+\frac{S}{\kappa}\right)$.

In order to calculate the covariance we only have to calculate $c_{1,1}(1,1,0)$ and $c_{1,1}(1,1,1)$. Noting that $d_1(\kappa, 1) = -\kappa$ it follows that

$$c_{1,1}(1,1,0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \frac{(-1)^2\exp\left\{\frac{S}{\kappa}\right\}(-\kappa)^2\frac{\nu_i^1}{\kappa^1}\frac{\nu_j^1}{\kappa}}{\Gamma(2)1!1!\kappa\frac{S^{1+1-0/\kappa}}{\kappa^{1+1-0/\kappa}}} = \exp\left\{\frac{S}{\kappa}\right\}\kappa\frac{\nu_i\nu_j}{S^2} = \exp\left\{\frac{S}{\kappa}\right\}\kappa\mu_i\mu_j$$

$$c_{1,1}(1,1,1) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \frac{(-1)^3\exp\left\{\frac{S}{\kappa}\right\}\nu_i^1\nu_j^1}{\Gamma(2)1!1!\kappa\frac{S^{1+1-1/\kappa}}{\kappa^{1+1-1/\kappa}}} = -\exp\left\{\frac{S}{\kappa}\right\}\kappa\frac{S^{1/\kappa}}{\kappa^{1/\kappa}}\frac{\nu_i\nu_j}{S^2} = -\exp\left\{\frac{S}{\kappa}\right\}\kappa\mu_i\mu_j\frac{S^{1/\kappa}}{\kappa^{1/\kappa}}$$

The fact that $\Gamma\left(1+1-0/\kappa, \frac{S}{\kappa}\right) = \Gamma\left(2, \frac{S}{\kappa}\right) = \exp\left\{-\frac{S}{\kappa}\right\}\left(1+\frac{S}{\kappa}\right)$ implies that

$$\mathrm{E}\left[W_iW_j\right] = \kappa\mu_i\mu_j\left[1+\frac{S}{\kappa} - \exp\left\{\frac{S}{\kappa}\right\}\left(\frac{S}{\kappa}\right)^{1/\kappa}\Gamma\left(2-1/\kappa, \frac{S}{\kappa}\right)\right].$$

Subtracting $\mathrm{E}(W_i)\mathrm{E}(W_j)$ from the above, we derive the formula for the covariance of W_i and W_j, and the result follows. $\square$