J. Japan Statist. Soc.
Vol. 35 No. 1 2005 1–19

EQUIVALENCE OF PARTIAL AND CONDITIONAL CORRELATION COEFFICIENTS

Kunihiro Baba* and Masaaki Sibuya**

Some classes of multivariate distributions, which have the same partial and conditional correlation coefficients, are obtained. In one class, the sum of components is fixed and all the components have negative correlation coefficients. Another class is constructed as mixtures of independent samples from the NEF-QVF mixed by conjugate prior, and all the components are positively correlated. Implications of the equivalence of partial and conditional covariances, and simple covariance structures are briefly discussed.

Key words and phrases: Conjugate prior, generalized Pareto distributions, Morris class, negative dependence, negative multinomial distributions, Neyman type A distributions, positive dependence by mixture.

1. Introduction

1.1. Equivalence of partial and conditional correlation coefficients

In the graphical modeling to search causalities, two random variables are regarded as unrelated if they are independent when all the other components are fixed (see, for example, Whittaker (1990)). If the set of random variables in a graphical model follows the multivariate normal distribution, the conditional correlation coefficient of a pair of component variables, given all the other variables, is equal to the partial correlation coefficient of the pair. A popular procedure, proposed by Dempster (1972), for constructing a graphical model is, therefore, to select sequentially pairs with the least sample partial correlation coefficient.

The question is what happens if the normality cannot be assumed. First, the zero partial correlation coefficient does not always imply conditional independence. Second, partial correlation coefficients are not always equal to conditional correlation coefficients. Baba, Shibata and Sibuya (2004) discussed the relationship among the notions of conditional and partial covariances, conditional and partial correlation coefficients, and also the relationship among conditional independence, zero conditional and partial correlation coefficients. However, the picture is not clear, and remains rather opaque.

The main purpose of this paper is to show that two classes of multivariate distributions have the same partial and conditional correlation coefficients. In one class, all the components have negative correlations, and the class is very wide. In the other class, all the components have positive correlations, and the class is rather restricted, but it includes very popular distributions like normal

Received October 4, 2004. Revised December 13, 2004. Accepted February 2, 2005.

*Department of Mathematics, Keio University, Yokohama 223-8522, Japan. Email: baba@stat.math.keio.ac.jp

**School of Business Administration, Takachiho University, Suginami, Tokyo 168-8508, Japan.

and negative multinomial. There is a common feature in these two classes; they are constructed as partial sums of the components of a multivariate random variable.

In the graphical modeling, some pairs may be thought to be conditionally dependent before the observation, and some pairs may be judged as not conditionally independent from data. Hence, conditional dependence, rather than conditional independence, can be more important for some sets of variables. For these subsets, equivalence of conditional and partial correlation coefficients is important and practical. Moreover, obtaining conditional correlation coefficients is tedious, but that of partial ones is relatively easier. The equivalence is proved based on a sufficient condition, which will be named Condition C.

In the remainder of this section, Condition C introduced in a previous paper is recalled. In Section 2, the first model with the sum of components fixed, is shown to satisfy Condition C. The variance-covariance matrix of a sample with a given sum from the natural exponential family is shown, and the results are summarized in Table 1.

In Section 3, the second model is introduced as independent observations from the natural exponential family with a quadratic variance function and mixed by the conjugate prior distribution (NEF-QVF-CP), the mixture is shown to satisfy Condition C, and the results are summarized in Table 2. It is shown by counter-examples, that mixture does not always imply Condition C.

In Section 4, concluding discussions, a simple covariance structure appearing in this paper is noted. Upper order statistics of generalized Pareto distributions are shown to satisfy Condition C, when lower order statistics are given. Further, conditional covariances are equal to partial covariances in memoryless distributions, under the same restriction on conditioning.

In Appendix A, six types of NEF-QVF-CP discussed in Section 3, are listed. In Appendix B, the details of counter-examples in Section 3 are explained.

1.2. Condition C

Let M denote a subset of the index set $\{1,\dots,m\}$ of a random variable (rv) $\boldsymbol{Y} = (Y_1,\dots,Y_m)$, $m \geq 3$, $|M| \geq 2$, and let M^c denote its non-empty complement. Divide the variance-covariance matrix of $\boldsymbol{Y}$ into 2×2 block matrix

$$\mathrm{Var}(\boldsymbol{Y}) = \begin{pmatrix} \Lambda_M & \Lambda_{MM^c} \\ \Lambda_{M^cM} & \Lambda_{M^c} \end{pmatrix}.$$

The partial covariance of a set of components $\boldsymbol{Y}_M$ given $\boldsymbol{Y}_{M^c}$, the rest of the components such as Λ_{M^c} is positive definite, is defined by

$$\mathrm{pV}(\boldsymbol{Y}_M;\boldsymbol{Y}_{M^c}) = \Lambda_M - \Lambda_{MM^c}\Lambda_{M^c}^{-1}\Lambda_{M^cM},$$

which is the variance-covariance matrix of $\boldsymbol{Y}_M - \hat{\boldsymbol{Y}}_M$ where $\hat{\boldsymbol{Y}}_M$ is the least-squares linear estimate of $\boldsymbol{Y}_M$ by $\boldsymbol{Y}_{M^c}$. The correlation matrix $\mathrm{pCor}(\boldsymbol{Y}_M;\boldsymbol{Y}_{M^c})$ of $\mathrm{pV}(\boldsymbol{Y}_M;\boldsymbol{Y}_{M^c})$ is the partial correlation matrix of $\boldsymbol{Y}_M$ given $\boldsymbol{Y}_{M^c}$.

Studying partial covariances of multivariate rv in general, Baba *et al.* (2004, Theorem 1) showed that the linearity of conditional expectation $\mathrm{E}(\boldsymbol{Y}_M \mid \boldsymbol{Y}_{M^c})$ in regressor $\boldsymbol{Y}_{M^c}$ is equivalent to that $\mathrm{pV}(\boldsymbol{Y}_M; \boldsymbol{Y}_{M^c}) = \mathrm{E}(\mathrm{Var}(\boldsymbol{Y}_M \mid \boldsymbol{Y}_{M^c}))$. Hence, if $\mathrm{Var}(\boldsymbol{Y}_M \mid \boldsymbol{Y}_{M^c})$ is independent of $\boldsymbol{Y}_{M^c}$, partial and conditional variance-covariance matrices are equal. They failed to find rv, other than normal, satisfying this condition. However, see Section 4 of the present paper for a special case.

If partial and conditional correlation coefficients, rather than covariances, are taken into consideration, the following condition is sufficient for $\mathrm{pCor}(\boldsymbol{Y}_M; \boldsymbol{Y}_{M^c}) = \mathrm{Cor}(\boldsymbol{Y}_M \mid \boldsymbol{Y}_{M^c})$. In Corollary 1 in Baba *et al.* (2004), $\boldsymbol{Y}_M$ is restricted to 2×2 matrices, i.e. $|M| = 2$, for simplicity. The extension to the general case, $|M| \geq 2$, is straightforward, and the extended version is used in this paper. For convenience, the sufficient condition is called Condition C in this paper.

Definition. Condition C.

(i) $\mathrm{E}(\boldsymbol{Y}_M \mid \boldsymbol{Y}_{M^c}) = \boldsymbol{a} + B\boldsymbol{Y}_{M^c}$, *for a constant vector* $\boldsymbol{a}$ *and a constant matrix* B, *and*

(ii) *the conditional correlation matrix* $\mathrm{Cor}(\boldsymbol{Y}_M \mid \boldsymbol{Y}_{M^c})$ *is independent of* $\boldsymbol{Y}_{M^c}$.

The condition is applicable to a wider class of distributions. A typical example of distributions satisfying Condition C is elliptical distributions (Baba *et al.* (2004)), which is not discussed in this paper.

2. Negatively dependent case

2.1. A general result

To show a class of multivariate rv's satisfying Condition C, let us introduce "partial sums $\boldsymbol{Y}$ of $\boldsymbol{X}$" as follows. $\boldsymbol{X} = (X_1, \ldots, X_n)$ is a rv on $\mathcal{R}^n$ with $n \geq 3$ fixed. Partition the index set $\{1, \ldots, n\}$ into m parts $L_1, L_2, \ldots, L_m$ where $|L_j| = l_j > 0$ ($\sum_{j=1}^m l_j = n$), and define $\boldsymbol{Y} = (Y_1, \ldots, Y_m)$ where $Y_j = \sum_{i=1}^n \mathrm{I}[i \in L_j] X_i$.

Lemma 1. *Assume that* $\boldsymbol{X}$ *has the following conditional moments given* $T = \sum_{i=1}^n X_i = t$:

$$\begin{aligned} &\mathrm{E}(X_j \mid T = t) = t/n, \qquad \mathrm{Var}(X_j \mid T = t) = \sigma_t^2, \qquad \textit{and} \\ &\mathrm{Cov}(X_i, X_j \mid T = t) = \kappa_t, \qquad (i \neq j = 1, \ldots, n). \end{aligned}$$

The conditional expectation, variance-covariance matrix and correlations of partial sums $\boldsymbol{Y}$ *of* $\boldsymbol{X}$ *given* $T = t$ *are* $\mathrm{E}(\boldsymbol{Y} \mid T = t) = t\boldsymbol{\xi}$,

$$\mathrm{Var}(\boldsymbol{Y} \mid T = t) = -n^2 \kappa_t (\mathrm{diag}(\boldsymbol{\xi}) - \boldsymbol{\xi}\boldsymbol{\xi}^{\mathsf{T}}), \qquad \boldsymbol{\xi} = (\xi_1, \ldots, \xi_m)^{\mathsf{T}}, \tag{2.1}$$

and

$$\mathrm{Cor}(Y_i, Y_j \mid T = t) = -\sqrt{\xi_i \xi_j / (1 - \xi_i)(1 - \xi_j)}, \qquad (i \neq j = 1, \ldots, m), \tag{2.2}$$

where $\xi_j = l_j / n$ $(j = 1, \ldots, m)$.

PROOF. The conditional moments of $\boldsymbol{Y}$ given $T=t$ are $\mathrm{E}(Y_j \mid t) = l_j t/n$,

$$\mathrm{Var}(Y_j \mid t) = \mathrm{E}(Y_j^2 \mid t) - \mathrm{E}(Y_j \mid t)^2 = l_j\sigma_t^2 + l_j(l_j-1)\kappa_t,$$

and

$$\mathrm{Cov}(Y_i, Y_j \mid t) = \mathrm{E}(Y_iY_j \mid t) - \mathrm{E}(Y_i \mid t)\mathrm{E}(Y_j \mid t) = l_il_j\kappa_t, \qquad (i \neq j = 1, \dots, m).$$

They are equal to (2.1) since $\sigma_t^2 = -(n-1)\kappa_t$ from

$$\begin{aligned}\mathrm{E}(X_j^2 \mid T=t) &= \mathrm{E}\left\{ X_j \left(t - \sum_{i\neq j} X_i \right) \middle| T=t \right\} \\ &= t\mathrm{E}(X_j \mid T=t) - \sum_{i\neq j} \mathrm{E}(X_iX_j \mid T=t).\end{aligned}$$ □

Remark 1. If $\boldsymbol{X}$ has a degenerate distribution concentrated on a hyperplane $\sum_{j=1}^n X_j = t$, and if its components have constant moments of the first and second orders, Lemma 1 holds with σ_t^2 and κ_t replaced by constants σ^2 and κ, respectively. The following simple example illustrates this remark, but partial sums are not treated.

Example 1. Let $\boldsymbol{Z} = (Z_1, \dots, Z_m)$ be a set of uncorrelated rv's with a common variance τ^2, and let $\overline{Z}$ denote its sample mean. Put $\boldsymbol{Y} = (Y_1, \dots, Y_m)$, $Y_j = Z_j - \overline{Z}$ $(j = 1, \dots, m)$, then $\boldsymbol{Y}$ is degenerated on a $(m-1)$-dimensional hyperplane and satisfies the assumptions of the above remark with $\mu = 0$, $\sigma^2 = (1-1/m)\tau^2$, and $\kappa = -\tau^2/m$. Hence, $\mathrm{Var}(\boldsymbol{Y}) = m\tau^2(\mathrm{diag}(\boldsymbol{\xi}) - \boldsymbol{\xi}\boldsymbol{\xi}^\mathsf{T})$, where $\boldsymbol{\xi} = \mathbf{1}/m$.

The following is an example of Lemma 1, but partial sums are not treated.

Example 2. Let (Y_1, Y_2, Y_3) be a random sample from the uniform distribution on $(0,1)$. (Y_1, Y_2, Y_3) given $T = Y_1 + Y_2 + Y_3 = t$ has a uniform density on triangles $(0 < t \leq 1$ or $2 \leq t < 3)$ or hexagons $(1 < t < 2)$, and has moments $\mathrm{E}(Y_i \mid T=t) = t/3$, $(0 < t < 3)$,

$$\mathrm{Var}(\boldsymbol{Y} \mid T=t) = -9\kappa_t(\mathrm{diag}(\boldsymbol{\xi}) - \boldsymbol{\xi}\boldsymbol{\xi}^\mathsf{T}); \quad \boldsymbol{\xi} = \mathbf{1}/3,$$

$$-36\kappa_t = \begin{cases} t^2, & 0 < t \leq 1, \\ \frac{-2t^2(3-t)^2 + 12t(3-t) - 15}{2t(3-t)-3}, & 1 < t < 2, \\ (3-t)^2, & 2 \leq t < 3, \end{cases}$$

and $\mathrm{Cor}(Y_j, Y_k \mid T=t) = -1/2$, $(0 < t < 3)$ where $i, j \neq k = 1, 2, 3$. If the sample size, m, is larger or if partial sums are treated, the calculation of variance and covariance is tedious. However, the conditional correlation is easily calculated by (2.2), and is equal to $-1/(m-1)$ if partial sums are not treated.

The following theorem is a slight extension of Theorem 2 in Baba *et al.* (2004), where $\boldsymbol{X}$ is assumed to be independent and reproductive.

THEOREM 1. *Let $\boldsymbol{Y} = (\boldsymbol{Y}_M, \boldsymbol{Y}_{M^c})$ be a partition of partial sums $\boldsymbol{Y}$ of $\boldsymbol{X}$. Under the condition that $T = \sum_{j=1}^m Y_j$ and $\boldsymbol{Y}_{M^c}$ are given, if the first and second order conditional moments of $\{X_i; i \in \cup_{j\in M} L_j\}$, the original components of $\boldsymbol{Y}_M$, are all the same for i, then $(\boldsymbol{Y}_M, \boldsymbol{Y}_{M^c} \mid T)$ satisfies Condition C.*

PROOF. Given $\boldsymbol{Y}_{M^c} = \boldsymbol{y}_{M^c} = \{y_i; i \in M^c\}$, let y_* and l_* denote $y_* = \sum_{i\in M^c} y_i$ and $l_* = \sum_{i\in M^c} l_i$, respectively. Under the condition that $T = t$ and $\boldsymbol{Y}_{M^c} = \boldsymbol{y}_{M^c}$, $\boldsymbol{Y}_M$ has the same stochastic structure as Lemma 1 with t and n replaced by $t - y_*$ and $n - l_*$, respectively. Thus, it holds true that

$$\mathrm{E}(\boldsymbol{Y}_M \mid T = t, \boldsymbol{Y}_{M^c} = \boldsymbol{y}_{M^c}) = (t - y_*)\tilde{\boldsymbol{\xi}}_M \qquad \text{and}$$
$$\mathrm{Var}(\boldsymbol{Y}_M \mid T = t, \boldsymbol{Y}_{M^c} = \boldsymbol{y}_{M^c}) = -(n - l_*)^2 \kappa_{t,\boldsymbol{y}_{M^c}} \{\mathrm{diag}(\tilde{\boldsymbol{\xi}}_M) - \tilde{\boldsymbol{\xi}}_M \tilde{\boldsymbol{\xi}}_M^\top\}$$

where $\tilde{\boldsymbol{\xi}}_M = \boldsymbol{l}_M/(n - l_*)$, $\boldsymbol{l}_M = (l_j; j \in M)$ and $\kappa_{t,\boldsymbol{y}_{M^c}} = \mathrm{Cov}(Y_i, Y_j \mid T = t, \boldsymbol{Y}_{M^c} = \boldsymbol{y}_{M^c})$. Hence Condition C is satisfied. □

If the original $(X_1, \ldots, X_n)$ are identical and independently distributed (iid) rv's, the assumptions of Lemma 1 and Theorem 1 are satisfied. Further if X_i is infinitely divisible, X_i is expressed as a sum of any number of iid rv's. The parameter $l = |L|$, $L \subset \{1, \ldots, n\}$ to define $Y_l = \sum_{i\in L} X_i$ can be extended to $\mathcal{R}_+$, and $E(Y_k)$ can be any real number. See, for example, Lauritzen (1975) and Jørgensen (1997). This fact is illustrated in the next subsection, where size parameter of Y_j is denoted by ν_j rather than l_j. When X_i is not infinitely divisible, ν_j is restricted to be an integer.

2.2. Natural exponential family

A typical application of Theorem 1 is the distribution of a random sample under the condition that the sum of the observations is given. Well-known degenerate distributions are listed on Table 1, except for Morris' distribution which is rarely used. The distributions of Table 1 have the variance-covariance matrix of the following form (2.5), and satisfy Condition C. The concrete expression of the multiplier $c(t, \nu)$ in (2.5) is shown in the last column of Table 1, and it is derived in a unified way as follows:

Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be a random sample from

$$p(x, \theta) = a(x) \exp(\theta x - \psi(\theta)), \qquad \theta \in \Theta \subset \mathcal{R}, \tag{2.3}$$

which is a pdf or a probability mass function (pmf) of the univariate natural exponential family with the cumulant function $\psi(\theta)$. It is specified by the mean $\mu(\theta) = \psi'(\theta)$ and denoted by $\mathsf{NEF}(\mu(\theta))$. See, for example, Letac and Mora (1990), Jørgensen (1997), and Casalis (2000).

$T = \sum_{j=1}^n X_j$ is a sufficient statistic, and is an $\mathsf{NEF}(n\mu(\theta))$ rv with the density

$$p(t; \theta) = b(t; n) \exp(\theta t - n\psi(\theta)),$$

where $b(t; n)$ is the n-convolution of $a(t)$. If X_i is infinitely divisible as noted in the end of Subsection 2.1, the sample size n is extended to real numbers, and the family $\mathsf{NEF}(\nu\mu(\theta))$, $\nu \in \mathcal{R}$, can be defined.

Now, let $\boldsymbol{Y} = (Y_1, \dots, Y_m)$ be independent and $Y_j \sim \mathsf{NEF}(\nu_j\mu(\theta))$. $T = \sum_{j=1}^m Y_j$ is a sufficient statistic, and $T \sim \mathsf{NEF}(\nu\mu(\theta))$, $\nu = \sum_{j=1}^m \nu_j$. The conditional density of $\boldsymbol{Y}$ given $T = t$ is

$$\prod_{j=1}^m b(y_j; \nu_j)/b(t; \nu). \tag{2.4}$$

Lemma 1 shows that $\mathrm{E}(Y_j \mid t) = t\nu_j/\nu$, and the correlations of $(Y_1, \dots, Y_m) \mid t$ are given by (2.2), with $\boldsymbol{\xi} = (\nu_1/\nu, \dots, \nu_m/\nu)$. Let $M \cup M^c$ be a partition of $\{1, \dots, m\}$. The conditional density of $\boldsymbol{Y}_M = (Y_j, j \in M)$ given $T = t$ and $\boldsymbol{Y}_{M^c} = \boldsymbol{y}_{M^c}$ is

$$\prod_{j \in M} b(y_j; \nu_j)/b(t - y_*; \nu - \nu_*), \qquad y_* = \sum_{j \in M^c} y_j, \qquad \nu_* = \sum_{j \in M^c} \nu_j$$

which satisfies the assumption of Theorem 1.

So far the variance function is not restricted, and κ_t can not be expressed explicitly. Now assume that the variance function, $V(\mu) = \psi''(\theta)$, is quadratic (NEF-QVF, or Morris class), then variance-covariance matrices are explicitly obtained. The following is an extension of Morris (1983, Section 4).

Proposition 1. *Assume that the variance function of NEF($\mu(\theta)$) is $V(\mu) = v_0 + v_1\mu + v_2\mu^2$. The variance-covariance matrix of $\boldsymbol{Y} = (Y_1, \dots, Y_m)$, $Y_j \sim$ NEF-QVF($\nu_j\mu(\theta)$), given $\sum_{j=1}^m Y_j = t$ is*

$$\begin{aligned} \mathrm{Var}(\boldsymbol{Y} \mid t) &= c(t, \nu)(\mathrm{diag}(\boldsymbol{\xi}) - \boldsymbol{\xi}\boldsymbol{\xi}^\mathsf{T}), \\ \nu &= \sum_{j=1}^m \nu_j; \ \xi_j = \nu_j/\nu, \ j = 1, \dots, m \\ c(t, \nu) &= \frac{\nu^2}{\nu + v_2} V\left(\frac{t}{\nu}\right). \end{aligned} \tag{2.5}$$

Proof. The following equalities hold, since the expectation with respect to the complete sufficient statistic t of both sides are equal (see, e.g. Morris (1983), Lemma 4.1),

$$\begin{aligned} \mathrm{E}(Y_i^2 \mid t) &= \left(\frac{\nu_i t}{\nu}\right)^2 + \frac{\nu_i(\nu - \nu_i)}{\nu + v_2} V\left(\frac{t}{\nu}\right), \\ \mathrm{E}(Y_i Y_j \mid t) &= \nu_i\nu_j\left(\frac{t}{\nu}\right)^2 - \frac{\nu_i\nu_j}{\nu + v_2} V\left(\frac{t}{\nu}\right), \end{aligned}$$

and (2.5) is obtained. □

The following example shows that if the sufficient statistic of an exponential family is not the sum of the sample, then the conditional distribution given the sufficient statistic does not satisfy the condition of Lemma 1.

Table 1. iid variables and conditional multivariate distributions satisfying Condition C.

$\boldsymbol{X}$ iid	$(\boldsymbol{Y} \mid T=t)$ conditional	$\mathrm{V}(\mu)$	$b(t;\nu)$	$c(t,\nu)$
normal	normal	1	$\frac{1}{\sqrt{2\pi\nu}}e^{-\frac{t^2}{2\nu}}$	ν
Poisson	multinomial	μ	$\frac{\nu^t}{t!}$	t
binomial, Bernoulli	mv. hypergeometric	$\mu(1-\mu)$	$\binom{\nu}{t,\nu\in\mathcal{N}}$	$\frac{t(\nu-t)}{\nu-1}$
neg. binomial, geometric	mv. neg. hypergeometric	$\mu(1+\mu)$	$\binom{t+\nu-1}{t}$	$\frac{t(\nu+t)}{\nu+1}$
gamma, exponential	Dirichlet	μ^2	$\frac{t^{\nu-1}}{\Gamma(\nu)}$	$\frac{t^2}{\nu+1}$
hyperbolic secant	Morris (1983)	$1+\mu^2$	$*$	$\frac{\nu(1+t^2)}{\nu+1}$

$* \quad = \frac{2^{\nu-2}}{\pi\Gamma(\nu)}|\Gamma(\frac{\nu}{2}+i\frac{t}{2})|^2 = \frac{2^{\nu-2}(\Gamma(\nu/2))^2}{\pi\Gamma(\nu)}\prod_{k=0}^{\infty}(1+(\frac{t}{\nu+2k})^2)^{-1}$

Example 3. (counter-example) Let $\boldsymbol{Y} = (Y_1, \ldots, Y_m)$ be a random sample from the uniform distribution on $(0, \theta)$, and $T = \max_{j=1}^m Y_j$ is a sufficient statistic. The conditional marginal distributions of Y_j under $T = t$ are

$$P\{Y_j = t \mid t\} = \frac{1}{m}, \qquad P\{Y_j \le y \mid t\} = \frac{(m-1)y}{mt}, \qquad y < t,$$
$$P\{Y_i = t \cup Y_j = t \mid t\} = \frac{2}{m},$$
$$P\{Y_i \le x \cap Y_j \le y \mid t\} = \frac{(m-1)^2 xy}{(mt)^2}, \qquad x, y < t;\ i \ne j.$$

The conditional moments of Y_j are

$$\mathrm{E}(Y_j \mid t) = \frac{(m+1)t}{2m}, \qquad \mathrm{Var}(Y_j \mid t) = \frac{(m+2)t^2}{3m}, \qquad \mathrm{Cov}(Y_i, Y_j \mid t) = \frac{-t^2}{4m^2}.$$

3. Positively dependent case

In this section, a class of six multivariate distributions of Table 2, satisfying Condition C, is introduced. The class includes negative multinomial distributions and is characterized as "NEF-QVF samples mixed by conjugate prior (NEF-QVF-CP)."

3.1. NEF-QVF samples mixed by conjugate prior

Let $\boldsymbol{X}_\theta = (X_1, \ldots, X_n)$ be a random sample from the distribution function $F_0(x;\theta)$. Suppose that the real parameter θ is a random variable with the distribution function $G(\theta)$, and $\boldsymbol{W} = \boldsymbol{X}_\theta \mid \theta \sim G$ has the joint distribution function

$$F(x_1, \ldots, x_n) = \int F_0(x_1;\theta)\cdots F_0(x_n;\theta)dG(\theta). \tag{3.1}$$

The distribution F or the rv $\boldsymbol{W}$ is called mixture by the mixing distribution G, denoted by Gurland's notation (Gurland (1957))

$$\prod_{i=1}^n F_0(x_i;\theta) \bigwedge_\theta G \quad \text{or} \quad \boldsymbol{X}_\theta \bigwedge_\theta G$$

which is known to be positively dependent (see, Shaked (1971); Marshall and Olkin (1979)). The conditional distribution of $(W_1, \ldots, W_{n-1})$ given W_n is

$$F(w_1, \ldots, w_{n-1} \mid w_n) = \int \left(\prod_{j=1}^{n-1} F_0(w_i; \theta) \right) dG(\theta \mid w_n),$$
$$dG(\theta \mid w_n) = \frac{F_0(w_n; \theta) dG(\theta)}{\int F_0(w_n; \theta) dG(\theta)},$$

that is the conditional distribution is obtained by changing the prior distribution $G(\theta)$ to its posterior distribution $G(\theta \mid w_n)$.

Now, to discuss NEF-QVF-CP, let us return to Y_j of the NEF$(\nu_j \mu(\theta))$ $(j = 1, \ldots, m)$ in Section 2.2, and assume that $\psi''(\theta) = V(\mu(\theta))$, the variance function, is quadratic in μ. This assumption guarantees appropriate properties of the standard conjugate prior of the NEF$(\nu\mu(\theta))$. See, Morris (1983) and Consonni and Veronese (1992).

The NEF$(\nu\mu(\theta))$ with the density

$$p(y; \theta, \nu) = b(y; \nu) \exp(\theta y - \nu\psi(\theta)) \tag{3.2}$$

has the conjugate prior on θ which mimics (3.2), namely,

$$\pi^*(\theta; \eta, \zeta) = K(\eta, \zeta) \exp(\eta\theta - \zeta\psi(\theta)),$$

where η and ζ are the natural and the convolution parameters of $\pi^*(\theta)$, respectively, and K is the normalization constant. The prior π^* is regarded here as a distribution on $\mu(\theta) = \psi'(\theta)$. The density of CP$(\eta, \zeta)$, the conjugate prior on μ, is

$$\begin{aligned} &\pi(\mu; \eta, \zeta) = K(\eta, \zeta) \exp(\eta\theta(\mu) - \zeta\psi(\theta(\mu)))V^{-1}(\mu), \\ &V(\mu) = \psi''(\theta(\mu)). \end{aligned} \tag{3.3}$$

The product $p\pi$ is

$$b(y; \nu)K(\eta, \zeta) \exp((y + \eta)\theta(\mu) - (\nu + \zeta)\psi(\theta(\mu)))V^{-1}(\mu),$$

which, being integrated, turns to

$$p(y; \nu, \eta, \zeta) := b(y; \nu)K(\eta, \zeta)/K(\eta + y, \zeta + \nu), \tag{3.4}$$

and the posterior is

$$\pi(\mu; \eta + y, \zeta + \nu).$$

This simple form is a merit of conjugate prior.

THEOREM 2. *Let $\boldsymbol{Y} = (Y_1, \ldots, Y_m)$ be independent and $Y_j \sim$ NEF-QVF$(\nu_j \mu(\theta))$. If $\pi(\mu; \eta, \zeta)$ is its conjugate prior CP(η, ζ) of NEF-QVF$(\mu(\theta))$, the mixture $\boldsymbol{W}$ of $\boldsymbol{Y}$ mixed by $\pi(\mu; \eta, \zeta)$ has*

$$\begin{aligned} &\mathrm{Var}(\boldsymbol{W}) = \frac{V(\eta/\zeta)}{\zeta - v_2}(\zeta \operatorname{diag}(\boldsymbol{\nu}) + \boldsymbol{\nu}\boldsymbol{\nu}^{\mathsf{T}}), \\ &\mathrm{Cor}(W_i, W_j) = \sqrt{\frac{\nu_i \nu_j}{(\zeta + \nu_i)(\zeta + \nu_j)}}, \end{aligned} \tag{3.5}$$

and satisfies Condition C.

PROOF. Let $V(\mu) = v_0 + v_1\mu + v_2\mu^2$ denote the variance function of NEF-QVF$(\mu(\theta))$. Its CP(η, ζ) has the moments

$$\mathrm{E}(\mu) = \mu_0 \quad \text{and} \quad \mathrm{Var}(\mu) = V(\mu_0)/(\zeta - v_2), \quad \mu_0 = \eta/\zeta,$$

from Morris (1983) Theorem 5.3. Since

$$\mathrm{E}(W_j \mid \mu) = \nu_j\mu, \qquad \mathrm{E}(W_j^2 \mid \mu) = \nu_j V(\mu) + (\nu_j\mu)^2, \quad \text{and}$$
$$\mathrm{E}(W_iW_j \mid \mu) = \nu_i\nu_j\mu^2,$$

the moments of $\boldsymbol{W}$ are $\mathrm{E}(W_j) = \nu_j\mu_0$,

$$\mathrm{E}(W_j^2) = \nu_j V(\mu_0)\left(1 + \frac{\nu_j + v_2}{\zeta - v_2}\right) + (\nu_j\mu_0)^2, \quad \text{and}$$
$$\mathrm{E}(W_iW_j) = \nu_i\nu_j\left(\frac{V(\mu_0)}{\zeta - v_2} + \mu_0^2\right).$$

Hence $\mathrm{Var}(\boldsymbol{W})$ is obtained.

The posterior of CP(η, ζ) given $y \sim$ NEF-QVF$(\nu\mu(\theta))$ is CP$(\eta+y, \zeta+\nu)$, and the conditional variance-covariance of $\boldsymbol{W}_M$ given $\sum_{j\in M^c} W_j = w_*$, has a common factor $V((\eta + w_*)/(\zeta + \nu_*))$ where $\nu_* = \sum_{j\in M^c} \nu_j$, depending on w_*. However, the conditional correlation coefficients, $\sqrt{\nu_i\nu_j/(\zeta + \nu_* + \nu_i)(\zeta + \nu_* + \nu_j)}$, do not depend on w_*, and the conditional expectation, $\mathrm{E}(W_j \mid \boldsymbol{W}_{M^c}) = \nu_j(\eta+w_*)/(\zeta+\nu_*)$, is linear in $\boldsymbol{w}_{M^c}$, hence Condition C is satisfied. □

Remark 2. The parameter ζ is called *prior size*, and correlation coefficients are determined only by the sample and prior sizes.

3.2. Members of NEF-QVF-CP

For better understanding of Theorem 2, related sampling distributions are obtained. Let us start from the mixture $\boldsymbol{W} = (W_1, \dots, W_m)$ by conjugate prior,

$$\prod_{j=1}^{m} \mathrm{NEF}(\nu_j\mu(\theta)) \bigwedge_{\mu} \pi(\mu; \eta, \zeta).$$

This mixture has density, an extension of (3.4),

$$p(\boldsymbol{w}; \boldsymbol{\nu}; \eta, \zeta) = \left(\prod_{j=1}^{m} b(w_j; \nu_j)\right) K(\eta, \zeta) \Big/ K(\eta + t, \zeta + \nu) \tag{3.6}$$

$$= p(t; \nu, \eta, \zeta) \prod_{j=1}^{m} b(w_j; \nu_j)/b(t; \nu), \tag{3.7}$$

$$t = \sum_{j=1}^{m} w_j, \quad \nu = \sum_{j=1}^{m} \nu_j.$$

Compare the last expression with the conditional density (2.4) to find that (3.7) is obtained if t in (2.4) follows $p(t;\nu,\eta,\xi)$. This is another genesis of (3.7).

The expression (3.7) suggests an alternative computation of $\mathrm{Var}(\boldsymbol{W})$ starting from the conditional variance-covariance matrix (2.5), or

$$\mathrm{E}(W_j^2 \mid t) = c(t,\nu)(\xi_j - \xi_j^2) + t^2\xi_j^2, \qquad \mathrm{E}(W_iW_j \mid t) = -c(t,\nu)\xi_i\xi_j + t^2\xi_i\xi_j,$$
$$c(t,\nu) = \frac{\nu_2}{\nu + v_2}\mathrm{V}\left(\frac{t}{\nu}\right).$$

Put $\mathrm{E}(T) := \mu$, $\mathrm{Var}(T) := \sigma^2$, and $\mathrm{E}(c(T,\nu)) := \kappa$, and

$$\mathrm{Var}(\boldsymbol{W}) = \kappa\,\mathrm{diag}(\boldsymbol{\xi}) + (\sigma^2 - \kappa)\boldsymbol{\xi}\boldsymbol{\xi}^{\mathsf{T}}.$$

Now, $\mathrm{V}(t/\nu) = v_0 + v_1 t/\nu + v_2 t^2/\nu^2$, and $\mathrm{E}(W_j)$, $\mathrm{E}(W_j^2)$, $\mathrm{E}(W_iW_j)$ are computed in the above proof. Hence

$$\text{(3.8)} \qquad \kappa = \frac{\nu\zeta}{\zeta - v_2}\mathrm{V}(\mu_0), \qquad \sigma^2 - \kappa = \frac{\nu^2}{\zeta - v_2}\mathrm{V}(\mu_0), \qquad \mu_0 = \eta/\zeta,$$

and $\mathrm{Var}(\boldsymbol{W})$ is as (3.5).

Note that the expression (3.7) shows also that partial sums of the mixture are mixtures of partial sums. Note also that the conditional density of $\boldsymbol{W}_M$ given $\boldsymbol{W}_{M^c}$ is expressed in terms of $b(\cdot;\cdot)$ and $K(\cdot,\cdot)$.

Six classes of mixtures discussed in this subsection are shown in Appendix, and summarized in Table 2, which should be compared with Table 1. The first two columns of Table 2 show multisample mixtures with density, (3.7). The third column is the non-exponential constant part K of conjugate prior $\mathrm{CP}(\eta,\zeta)$, (3.3), which also appears in (3.6). The fourth column is κ in (3.8), and $c(t,\nu)$ is the last column of Table 1. The last column is the distribution of the total size, following the univariate mixture distribution, appearing in (3.7).

The following two examples are counter-examples of positive dependence by mixture, which do not satisfy Condition C. The first example is a simple mixture of independent discrete uniform rv's by summation of another rv. The second example is a mixture of the NEF-QVF, by a prior which is not conjugate. The details of these examples are explained in Appendix B.

Example 4. In the expression (3.1), if $F_0(x;\theta) = F_0(x-\theta)$, the components of the mixture $\boldsymbol{W}$ is written as $W_i = X_i + X_0$, $X_i \sim F_0(\cdot)$, $X_0 \sim G(\cdot)$, $i = 1,\dots,m$, where $(X_i)_{i=0}^m$ are independent. If X_1 and X_0 are discrete uniform distributions on $\{0,1,\dots,a\}$ and $\{0,1,\dots,b\}$, $a > b$, respectively, correlation coefficient of (W_1,W_2) given $(W_3,\dots,W_m) = (w_3,\dots,w_m)$ depends on $l^{**} = \min(w_3,\dots,w_m,b)$, $l^* = \max(w_3 - a,\dots,w_m - a, 0)$ and is not constant. Hence Condition C is not satisfied.

Example 5. A definition of multivariate Neyman type A distributions is the mixture

$$\boldsymbol{W} = (W_1,\dots,W_m) \sim \prod_{j=1}^m \mathsf{Po}(\nu_j k) \bigwedge_k \mathsf{Po}(\lambda).$$

Table 2. Members of NEF-QVF-CP satisfying Condition C.

Mixed and mixing distribution	Mixture distribution $p(\boldsymbol{x})$; $x=\sum_{j=1}^m x_j = t$, $\nu=\sum_{j=1}^m \nu_j$	$K(\eta,\zeta)$	$\mathrm{E}(c(t,\nu))$	distribution of t
$\prod_{j=1}^m \mathrm{N}(\nu_j\mu,\nu_j)\bigwedge_\mu \mathrm{N}(\eta/\zeta,1/\zeta)$	$\mathsf{N}\left(\frac{\eta}{\zeta}\boldsymbol{\nu},\Lambda\right)$ $\Lambda=\mathrm{diag}(\boldsymbol{\nu})+\frac{1}{\zeta}\boldsymbol{\nu}\boldsymbol{\nu}^\mathsf{T}$ $\Lambda^{-1}=\mathrm{diag}\left(\frac{1}{\boldsymbol{\nu}}\right)-\frac{\zeta^2}{\zeta+\nu}\mathbf{1}\mathbf{1}^\mathsf{T}$	$\sqrt{\frac{\zeta}{2\pi}}\exp\left(-\frac{\eta^2}{2\zeta}\right)$	$\frac{1}{\zeta}$	$\mathsf{N}\left(\frac{\nu\eta}{\zeta},\frac{\nu(\nu+\zeta)}{\zeta}\right)$
$\prod_{j=1}^m \mathrm{Po}(\nu_j\lambda)\bigwedge_\lambda \mathrm{Ga}(\eta,1/\zeta)$	$\mathsf{NgMn}\left(m,\eta,\left(\frac{\zeta}{\zeta+\nu},\frac{\nu_j}{\zeta+\nu}\right)\right)$ $\frac{\Gamma(\eta+x)}{\Gamma(\eta)\prod_{j=1}^m x_j!}\left(\frac{\zeta}{\zeta+\nu}\right)^\eta\prod_j\left(\frac{\nu_j}{\zeta+\nu}\right)^{x_j}$	$\frac{\zeta^\eta}{\Gamma(\eta)}$	$\frac{\eta}{\zeta^2}$	$\mathsf{NgBn}\left(\eta,\frac{\zeta}{\zeta+\nu}\right)$
$\prod_{j=1}^m \mathsf{Bn}(\nu_j,p)\bigwedge_p \mathsf{Be}(\eta,\zeta-\eta)$	$\mathsf{MsNgHg}(m,\boldsymbol{\nu},\eta,\zeta-\eta)$ $\frac{B(\eta+x,\zeta-\eta+\nu-x)}{B(\eta,\zeta-\eta)}\prod_{j=1}^m\binom{\nu_j}{x_j}$	$\frac{1}{B(\eta,\zeta-\eta)}$	$\frac{\eta(\zeta-\eta)}{\zeta^2(\zeta+1)}$	$\mathsf{NgHg}(\nu;\eta,\zeta-\eta)$
$\prod_{j=1}^m \mathsf{NgBn}(\nu_j,p)\bigwedge_p \mathsf{Be}(\zeta+1,\eta)$	$\mathsf{MsGHgB3}(m,\boldsymbol{\nu},\eta,\zeta+1)$ $\frac{B(\zeta+\nu+1,\eta+x)}{B(\zeta+1,\eta)}\prod_{j=1}^m\binom{\nu_j+x_j-1}{x_j}$	$\frac{1}{B(\zeta+1,\eta)}$	$\frac{\eta(\zeta+\eta)}{\zeta^2(\zeta-1)}$	$\mathsf{GHgB3}(\eta,\nu;\zeta+1)$
$\prod_{j=1}^m \mathsf{Ga}(\nu_j,a)\bigwedge_a \mathsf{Ga}(\zeta+1,1/\eta)$	$\mathsf{MsBe2}(\zeta+1,\boldsymbol{\nu})$ $\frac{\Gamma(\zeta+\nu+1)\eta^{\zeta+1}}{\Gamma(\zeta+1)\prod_{j=1}^m\Gamma(\nu_j)}\frac{\prod_{j=1}^m x_j^{\nu_j-1}}{(\eta+x)^{\zeta+\nu+1}}$	$\frac{\eta^{\zeta+1}}{\Gamma(\zeta+1)}$	$\frac{\eta^2}{\zeta^2(\zeta-1)}$	$\mathsf{Be2}(\zeta+1,\nu;\eta)$
mixed: NEF-GHS mixing: Morris' t	$\mathsf{MsMorrisMixture}(\boldsymbol{\nu};\eta,\zeta)$ $\frac{H(\eta+x,\zeta+\nu)}{H(\eta,\zeta)}\prod_{j=1}^m b(x_j;\nu_j)$	$\frac{1}{H(\eta,\zeta)}$	$\frac{\eta^2+\zeta^2}{\zeta^2(\zeta-1)}$	$\mathsf{MorrisMixture}(\nu;\eta,\zeta)$

Since $\mathrm{E}(W_1 \mid W_2 = 0) = \nu_1\lambda e^{-\nu_2}$, $\mathrm{E}(W_1 \mid W_2 = 1) = \nu_1(1 + \lambda e^{-\nu_2})$, and $\mathrm{E}(W_1 \mid W_2 = 2) = \nu_1(1 + 3\lambda e^{-\nu_2} + (\lambda e^{-\nu_2})^2)/(1 + \lambda e^{-\nu_2})$, $\mathrm{E}(W_1 \mid W_2 = w_2)$ is not linear in w_2, and hence $\boldsymbol{W}$ does not satisfy Condition C.

4. Concluding discussions

In this paper, two sufficient conditions for the equivalence of partial and conditional correlation coefficients are given (Theorems 1 and 2). A common feature in these theorems is that the variance-covariance matrix of the concerned multivariate distributions have the form

$$\mathrm{Var}(\boldsymbol{Y}) = a(\mathrm{diag}(\boldsymbol{w}) + b\boldsymbol{w}\boldsymbol{w}^{\mathsf{T}}), \tag{4.1}$$

where $a > 0$ and $\boldsymbol{w}$ is a vector of nonzero components. The matrix $\mathrm{Var}(\boldsymbol{Y})$ of the form (4.1) is connected with Condition C in the negatively dependent case, but not in the positively dependent case. The covariances or the correlation coefficients expressed by outer-product of a vector, like (4.1), will be discussed further in a subsequent paper.

As (4.1) shows, a parameter a reflecting the total size of components is separated from parameters $\boldsymbol{w}$ for the ratios of components. When the total size changes, the variances and covariances will change, but the correlation coefficients remain unchanged.

The separation of the total size from the component ratios appears in a disguised form in the following order statistics: For the generalized Pareto distributions with the survival function

$$\bar{F}(x; \gamma, a) = (1 + \gamma x/a)^{-1/\gamma} I[x > 0 \ \& \ a + \gamma x > 0], \qquad a > 0,\ \gamma \in \mathcal{R},$$

the following assertions hold.

First, the order statistics of a random sample from the generalized Pareto distribution with $\gamma < 1/2$, $Y_{1:n} \geq \cdots \geq Y_{n:n}$, satisfies Condition C, in a limited sense that conditioning of lower order statistics is considered, since $Y_{i:n} \mid (Y_{k+1:n} = u) \stackrel{D}{=} u + (1 + \gamma u/a)X_{i:k}$, $1 \leq i \leq k$, where $X_{1:k} \geq \cdots \geq X_{k:k}$ are the order statistics of a random sample of smaller size from the same distribution.

Second, a random sample from the exponential and geometric distributions, memolyless distributions, also satisfies Condition C, and furthermore, using the same notations of the above assertion, it holds $\mathrm{Var}(\boldsymbol{Y}_c \mid Y_{k+1:n} = u) = \mathrm{Var}(\boldsymbol{X}), \boldsymbol{Y}_c = (Y_{1:n}, \ldots, Y_{k:n})$. This fact that conditional covariances do not depend on the condition means that conditional covariances are equal to partial covariances. See, Corollary 2 in Baba *et al.* (2004), where they wrote that other than normal they did not know of such distributions.

Appendix A: Properties of NEF-QVF-CP members

Using the notations

$$\mathcal{R}_+ := \{x : 0 < x < \infty\}, \qquad \mathcal{N} := \{1, 2, \ldots\}, \qquad \mathcal{N}_0 := \{0, 1, 2, \ldots\},$$

the m dimensional *unit simplex and lattice simplex* are defined as

$$\Delta(m) := \left\{ \boldsymbol{x} = (x_1, \ldots, x_m) \in \mathcal{R}_+^m : \sum_{j=1}^m x_j = 1 \right\}, \qquad m \in \mathcal{N},$$

$$\Delta(m,n) := \left\{ \boldsymbol{x} = (x_1, \ldots, x_m) \in \mathcal{N}_0^m : \sum_{j=1}^m x_j = n \right\}, \quad m, n \in \mathcal{N},$$

respectively. The latter is the set of all ordered partitions of a positive integer n into a sum of m nonnegative integers.

In the following six subsections, all the possible NEF-QVF($\nu\mu(\theta)$) (3.2), its conjugate prior PC(η, ζ) (3.3), and mixture of samples by PC (3.6) or (3.7) are listed.

A.1 Normal distribution

NEF: $\mathsf{N}(\nu\mu, \nu)$, $\nu > 0$, $\mu \in \mathcal{R}$.
CP: $\mathsf{N}(\eta/\zeta, 1/\zeta)$, $\zeta > 0$, $\eta \in \mathcal{R}$.
Sample mixture:

$$\prod_{j=1}^m \mathsf{N}(\nu_j\mu, \nu_j) \bigwedge_{\mu} \mathsf{N}(\eta/\zeta, 1/\zeta): \qquad \mathsf{N}((\eta/\zeta)\boldsymbol{\nu}, \Lambda), \quad \boldsymbol{\nu} = (\nu_1, \ldots, \nu_m)^T \in \mathcal{R}_+^n.$$

$$f(\boldsymbol{x}) = (2\pi)^{-m/2} |\Lambda|^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{z}^T \Lambda^{-1} \boldsymbol{z}\right)$$
$$= \left(\prod_{j=1}^m \frac{1}{\sqrt{2\pi\nu_j}}\right) \sqrt{\frac{\zeta}{\zeta+\nu}} \exp\left(-\frac{1}{2}\left(\sum_{j=1}^m \frac{x_j^2}{\nu_j} - \frac{(t+\eta)^2}{\zeta+\nu} - \frac{\eta^2}{\zeta}\right)\right),$$
$$\Lambda = \mathrm{diag}(\boldsymbol{\nu}) + \frac{1}{\zeta}\boldsymbol{\nu}\boldsymbol{\nu}^T, \qquad \Lambda^{-1} = \mathrm{diag}(1/\boldsymbol{\nu}) - \frac{1}{\zeta+\nu}\mathbf{1}\mathbf{1}^T,$$
$$\boldsymbol{z} = \boldsymbol{x} - (\eta/\zeta)\boldsymbol{\nu}, \qquad t = \sum_{j=1}^m x_j, \qquad \nu = \sum_{j=1}^m \nu_j.$$

Conditional (sum) with random total size: $\mathsf{N}(t\boldsymbol{\xi}, \nu(\mathrm{diag}(\boldsymbol{\xi}) - \boldsymbol{\xi}\boldsymbol{\xi}^T))$, $\boldsymbol{\xi} = \boldsymbol{\nu}/\nu$

$$(2\pi)^{-(m-1)/2} \sqrt{\frac{\nu}{\prod_j \nu_j}} \exp\left(-\frac{1}{2}\sum_{j=1}^m \frac{1}{\nu_j}\left(x_j - \frac{\nu_j}{\nu}t\right)^2\right),$$

$\mathrm{diag}(1/\boldsymbol{\nu})$ is a generalized inverse of $\nu(\mathrm{diag}(\boldsymbol{\xi}) - \boldsymbol{\xi}\boldsymbol{\xi}^T)$. $t \sim \mathsf{N}(\frac{\nu\eta}{\zeta}, \frac{\nu(\nu+\zeta)}{\zeta})$.

A.2 Negative multinomial distributions

NEF: $\mathsf{Po}(\nu\lambda)$, ν, $\lambda > 0$,

$$\frac{\nu^x}{x!}\lambda^x e^{-\nu\lambda} = \frac{\nu^x}{x!}\exp(\theta x - \nu e^\theta), \qquad \lambda = e^\theta,\ \theta = \log\lambda;\ \theta \in \mathcal{R}.$$

CP: $\mathsf{Ga}(\eta, 1/\zeta)$, ζ, $\eta > 0$, $\mu = \lambda$.

$$\pi(\mu; \eta, \zeta) = \frac{\zeta^\eta}{\Gamma(\eta)} \mu^{\eta-1} e^{-\zeta\mu} = \frac{\zeta^\eta}{\Gamma(\eta)} \exp(\eta \log \mu - \zeta\mu) \frac{1}{\mu}.$$

Sample mixture:

$$\mathsf{NgMn}\left(m, \eta, \left(\frac{\zeta}{\zeta+\nu}, \frac{\boldsymbol{\nu}}{\zeta+\nu}\right)\right) \quad \text{or} \quad \mathsf{NgMn}(m, \eta; \boldsymbol{\xi}),$$

$$\boldsymbol{\nu} = (\nu_1, \dots, \nu_m)^T \in \Delta(m), \qquad \boldsymbol{\xi} = (\xi_0, \dots, \xi_m)^T \in \Delta(m+1),$$
$$\xi_0 = \zeta/(\zeta+\nu), \qquad \xi_j = (1-\xi_0)\nu_j/\nu, \qquad j = 1, \dots, m.$$

$$\prod_{j=1}^m \mathsf{Po}(\nu_j \lambda) \bigwedge_\lambda \mathsf{Ga}(\eta, 1/\zeta),$$

$$p(\boldsymbol{x}) = \frac{\Gamma(\eta+x)}{\Gamma(\eta)\prod_j x_j!} \left(\frac{\zeta}{\zeta+\nu}\right)^\eta \prod_j \left(\frac{\nu_j}{\zeta+\nu}\right)^{x_j} = \frac{\Gamma(\eta+x)}{\Gamma(\eta)\prod_{j=1}^m x_j!} \xi_0^\eta \prod_{j=1}^m \xi_j^{x_j},$$

$$\boldsymbol{x} \in \mathcal{N}_0^m, \qquad x = \sum_{j=1}^m x_j, \qquad \nu = \sum_{j=1}^m \nu_j.$$

$$\mathrm{Var}(\boldsymbol{X}) = \frac{\eta}{\zeta^2}(\zeta\, \mathrm{diag}(\boldsymbol{\nu}) + \boldsymbol{\nu}\boldsymbol{\nu}^T).$$

Conditional (sum) with random total size:

$$\mathsf{Mn}(m, t; \boldsymbol{\nu}/\nu), \qquad \boldsymbol{x} \in \Delta(m, t).$$

$$t \sim \mathsf{NgBn}(\eta, \zeta/(\zeta+\nu)), \qquad p(t; \nu, \eta, \zeta) = \binom{\eta+t-1}{t} \left(\frac{\zeta}{\zeta+\nu}\right)^\eta \left(\frac{\nu}{\zeta+\nu}\right)^t.$$

A.3 Multisample negative hypergeometric distributions

NEF: $\mathsf{Bn}(\nu, p)$, $0 < p < 1$, $\nu > 0$.

$$\binom{\nu}{x} p^x (1-p)^{\nu-x} = \binom{\nu}{x} \exp(\theta x - \nu \log(1+e^\theta)),$$

$$p = \frac{e^\theta}{1+e^\theta}, \quad \theta = \log \frac{p}{1-p} \in \mathcal{R}.$$

CP: $\mathsf{Be}(\eta, \zeta-\eta)$, $\eta > 0$, $\zeta - \eta > 0$, $\mu = p$.

$$\begin{aligned} \pi(\mu; \eta, \zeta) &= \frac{1}{B(\eta, \zeta-\eta)} \mu^{\eta-1}(1-\mu)^{\zeta-\eta-1} \\ &= \frac{1}{B(\eta, \zeta-\eta)} \exp\left(\eta \log \frac{\mu}{1-\mu} + \zeta \log(1-\mu)\right) \frac{1}{\mu(1-\mu)}. \end{aligned}$$

Sample mixture: $\mathsf{MsNgHg}(m, \boldsymbol{\nu}, \eta, \zeta-\eta)$

$$\prod_{j=1}^m \mathsf{Bn}(\nu_j, p) \bigwedge_p \mathsf{Be}(\eta, \zeta-\eta),$$

$$p(\boldsymbol{x}) = \frac{B(\eta+x,\zeta-\eta+\nu-x)}{B(\eta,\zeta-\eta)} \prod_{j=1}^{m} \binom{\nu_j}{x_j}$$

$$\boldsymbol{x} \in (0,1,\dots,\nu_1)\times\cdots\times(0,1,\dots,\nu_m), \qquad \boldsymbol{\nu}\in\mathcal{N}^m,\ \nu=\sum_{j=1}^{m}\nu_j,\ x=\sum_{j=1}^{m}x_j.$$

$$\operatorname{Var}(\boldsymbol{X}) = \frac{\eta(\zeta-\eta)}{\zeta^2(\zeta+1)}(\zeta\operatorname{diag}(\boldsymbol{\nu})+\boldsymbol{\nu}\boldsymbol{\nu}^T).$$

Conditional (sum) with random total size:

$$\mathsf{MvHg}(m,t;\boldsymbol{\nu}) \qquad \boldsymbol{x}\in\Delta(m,t), \qquad \boldsymbol{\nu}\in\Delta(m,\nu), \qquad \nu\ge\max(m,t).$$

$$t\sim\mathsf{NgHg}(\nu;\eta,\zeta-\eta) \qquad p(t;\nu,\eta,\zeta)=\binom{\nu}{t}\frac{B(\eta+t,\zeta+\nu-\eta-t)}{B(\eta,\zeta-\eta)}.$$

A.4 Multisample GHgB3 distributions

NEF: negative binomial(ν,p), $0<p<1$, $\nu>0$.

$$\binom{\nu+x-1}{x}p^{\nu}(1-p)^x = \binom{\nu+x-1}{x}\exp(\theta x+\nu\log(1-e^{\theta})),$$
$$p=1-e^{\theta}, \qquad \theta=\log(1-p)<0,$$

CP: $\mathsf{Be2}(\zeta+1,\eta)$, $\zeta>1$, $\eta>0$, $\mu=\frac{1-p}{p}>0$; $p=\frac{1}{\mu+1}\sim\mathsf{Be}(\zeta+1,\eta)$.

$$\begin{aligned}\pi(\mu;\eta,\zeta) &= \frac{1}{B(\zeta+1,\eta)}p^{\zeta}(1-p)^{\eta-1}\frac{dp}{d\mu}\\ &= \frac{1}{B(\zeta+1,\eta)}\exp\left(\eta\log\frac{\mu}{\mu+1}-\zeta\log(\mu+1)\right)\frac{\mu+1}{\mu}\end{aligned}$$

Sample mixture: $\mathsf{MsGHgB3}(m,\boldsymbol{\nu},\eta,\zeta+1)$, $\boldsymbol{x}\in\mathcal{N}_0^m$; $\boldsymbol{\nu}\in\mathcal{R}_+^m$; $\eta,\zeta\in\mathcal{R}_+$.

$$\prod_{j=1}^{m}\mathsf{NgBn}(\nu_j,p)\bigwedge_p\mathsf{Be}(\zeta+1,\eta),$$

$$p(\boldsymbol{x})=\frac{B(\zeta+\nu+1,\eta+x)}{B(\zeta+1,\eta)}\prod_{j=1}^{m}\binom{\nu_j+x_j-1}{x_j} \qquad x=\sum_{j=1}^{m}x_j,\ \nu=\sum_{j=1}^{m}\nu_j.$$

$$\operatorname{Var}(\boldsymbol{X})=\frac{\eta(\eta+\zeta)}{\zeta^2(\zeta-1)}\left(\zeta\operatorname{diag}(\boldsymbol{\nu})+\boldsymbol{\nu}\boldsymbol{\nu}^{\mathsf{T}}\right).$$

Conditional (sum) with random total size:

$$\mathsf{MvNgHg}(m,t;\boldsymbol{\nu}) \qquad \boldsymbol{x}\in\Delta(m,t).$$

$$t\sim\mathsf{GHgB3}(\eta,\nu;\zeta+1) \qquad p(t;\nu,\eta,\zeta)=\frac{\Gamma(\nu+t)}{\Gamma(\nu)t!}\frac{B(\zeta+\nu+1,\eta+t)}{B(\zeta+1,\eta)}.$$

A.5 Multivariate beta type 2

NEF: $\mathsf{Ga}(\nu, \text{scale} = 1/a)$, ν, $a > 0$.

$$\frac{x^{\nu-1}}{\Gamma(\nu)} a^{\nu} e^{-ax} = \frac{x^{\nu-1}}{\Gamma(\nu)} \exp(\theta x + \nu \log(-\theta)), \qquad \theta = -a < 0.$$

CP: $\mathsf{RcGa}(\zeta + 1, 1/\eta)$, ζ, $\eta > 0$, $\mu = 1/a > 0$; $a \sim \mathsf{Ga}(\zeta + 1, 1/\eta)$

$$\pi(\mu; \eta, \zeta) = \frac{\eta^{\zeta+1}}{\Gamma(\zeta+1)} \mu^{-\zeta} e^{-\eta/\mu} = \frac{\eta^{\zeta+1}}{\Gamma(\zeta+1)} \exp\left(-\frac{\eta}{\mu} - \zeta \log \mu\right).$$

Sample mixture: $\mathsf{MvBe2}(\zeta + 1, \boldsymbol{\nu})$, $\zeta > 1$, $\boldsymbol{\nu} \in \mathcal{R}_+^m$.

$$\prod_{j=1}^{m} \mathsf{Ga}(\nu_j, a) \bigwedge_a \mathsf{Ga}(\zeta + 1, 1/\eta),$$

$$f(\boldsymbol{x}) = \frac{\Gamma(\zeta + \nu + 1)\eta^{\zeta+1}}{\Gamma(\zeta+1)\prod_{j=1}^m \Gamma(\nu_j)} \frac{\prod_{j=1}^m x_j^{\nu_j - 1}}{(\eta + x)^{\zeta+\nu+1}}, \qquad \boldsymbol{x} \in \mathcal{R}_+^m,\ x = \sum_{j=1}^m x_j,\ \nu = \sum_{j=1}^m \nu_j.$$

$$\mathrm{Var}(\boldsymbol{X}) = \frac{\eta^2}{\zeta^2(\zeta - 1)}(\zeta \mathrm{diag}(\boldsymbol{\nu}) + \boldsymbol{\nu}\boldsymbol{\nu}^T), \qquad \zeta > 1.$$

Conditional (sum) with random total size:

$$\boldsymbol{x}/t \mid t \sim \mathsf{Dir}(\boldsymbol{\nu}) \qquad \boldsymbol{\nu} \in \mathcal{R}_+^m, \qquad t \sim \frac{\Gamma(\nu + \zeta + 1)}{\Gamma(\nu)\Gamma(\zeta+1)} \frac{\eta^{\zeta+1} t^{\nu-1}}{(\eta + x)^{\nu+\zeta+1}}.$$

A.6 Hyperbolic secant distribution

NEF: $\mathsf{NEF\text{-}GHS}(\nu, \mu) : \nu > 0$, $\mu \in \mathcal{R}$

$$b(x; \nu) \exp\left(x \tan^{-1} \mu - \frac{\nu}{2} \log(1 + \mu^2)\right) = b(x; \nu) \exp(\theta x + \nu \log \cos \theta),$$

$$b(x; \nu) = \frac{2^{\nu-2}}{\pi \Gamma(\nu)} \left| \Gamma\left(\frac{\nu}{2} + i\frac{x}{2}\right)\right|^2 = \frac{2^{\nu-2}(\Gamma(\nu/2))^2}{\pi\Gamma(\nu)} \prod_{k=0}^{\infty} \left(1 + \left(\frac{x}{\nu + 2k}\right)^2\right)^{-1},$$

$$x \in \mathcal{R}, \qquad -\frac{\pi}{2} < \theta < \frac{\pi}{2}, \qquad \mu = \tan\theta.$$

CP: Morris' t

$$\pi(\mu; \eta, \zeta) = H(\eta, \zeta)^{-1} \exp\left(\eta \tan^{-1} \mu - \frac{\zeta}{2} \log(1 + \mu^2)\right),$$

$$H(\eta; \zeta) = \int_{-\pi/2}^{\pi/2} e^{\eta\theta} \cos^{\zeta-2} \theta d\theta, \qquad \mu \in \mathcal{R},\ \eta \in \mathcal{R},\ \zeta > 0.$$

Sample mixture: multisample Morris' mixture

$$\prod_{j=1}^{m} \mathrm{GHS}(\nu_j, \mu) \bigwedge_{\mu} \pi(\mu; \eta, \zeta),$$

$$f(\boldsymbol{x}) = \frac{H(\eta + x, \zeta + \nu)}{H(\eta, \zeta)} \prod_{j=1}^{m} b(x_j; \nu_j),$$

$$\mathrm{Var}(\boldsymbol{X}) = \frac{\eta^2 + \zeta^2}{\zeta^2(\zeta - 1)}(\zeta\, \mathrm{diag}(\boldsymbol{\nu}) + \boldsymbol{\nu}\boldsymbol{\nu}^T).$$

Conditional (sum) with random total size:

$$\text{conditional multivariate GHS} \qquad f(\boldsymbol{x} \mid t) = \prod_{j=1}^{m} b(x_j;\nu_j)/b(t;\nu),$$

$$t \sim \text{Morris' mixture} \qquad f(t;\eta,\zeta,\nu) = \frac{H(\eta+t,\zeta+\nu)}{H(\eta,\zeta)} b(t;\nu).$$

Appendix B: Details of Examples 4 and 5

Detail of Example 4. Let $(X_i)_{i=1}^m$ be iid random variables with a pmf p_1, and X_0 be independent of $(X_i)_{i=1}^m$ with a pmf p_2. Define $W_i = X_i + X_0$, $i = 1,\dots,m$, and the conditional pmf of (W_1, W_2) given $W_3 = w_3,\dots,W_m = w_m$, is

$$\begin{aligned} &p(w_1, w_2 \mid w_3,\dots,w_m) \\ &\quad = \sum_x \left(\prod_{j=1}^m p_1(w_j - x)\right) p_0(x) \Big/ \sum_x \left(\prod_{j=3}^m p_1(w_j - x)\right) p_0(x). \end{aligned}$$

Now, if p_1 and p_0 are discrete uniform distributions on $\{0,1,\dots,a\}$ and $\{0,1,\dots,b\}$, $a > b$, respectively,

$$\begin{aligned} p(w_1,\dots,w_m) &= (a+1)^{-m}(b+1)^{-1} \\ &\qquad \times \sum_x I[0 \le w_1 - x \le a,\dots,0 \le w_m - x \le a, 0 \le x \le b] \\ &= (a+1)^{-m}(b+1)^{-1}(m^{**} - m^* + 1) I[m^* \le m^{**}], \\ m^{**} &= \min(w_1,\dots,w_m,b), \qquad m^* = \max(w_1 - a, w_m - a, 0), \end{aligned}$$

and

$$\begin{aligned} p(w_1,w_2 \mid w_3,\dots,w_m) &= \frac{(a+1)^{m-2}(b+1)}{(a+1)^m(b+1)} \frac{(m^{**} - m^* + 1)I[m^* \le m^{**}]}{(l^{**} - l^* + 1)} \\ &= \frac{1}{(a+1)^2(l^{**} - l^* + 1)} \\ &\qquad \times (\min(w_1, w_2, l^{**}) - \max(w_1 - a, w_2 - a, l^*)), \\ l^{**} &= \min(w_3,\dots,w_m,b), \\ l^* &= \max(w_3 - a,\dots,w_m - a, 0), \qquad 0 \le l^* \le l^{**} \le b. \end{aligned}$$

For $a = 2$, $b = 1$, $m \ge 3$, the conditional distributions given (l^{**}, l^*), $1 \ge l^{**} \ge l^* \ge 0$, are shown in Table 3 with correlations. The conditional moments depend on the statistics (l^{**}, l^*) and the correlations are not equal.

Detail of Example 5. A definition of multivariate Neyman type A distributions is the mixture $\boldsymbol{W} = (W_1,\dots,W_m) \sim \prod_{j=1}^m \mathsf{Po}(\nu_j k) \bigwedge_k \mathsf{Po}(\lambda)$, with the factorial moments

$$\mathrm{E}\left(\prod_{j=1}^m W_j^{\underline{r_j}}\right) = \mathrm{E}^k\left(\prod_{j=1}^m (\nu_j k)^{\underline{r_j}}\right) = \left(\prod_{j=1}^m \nu_j^{\underline{r_j}}\right) \sum_{l=1}^r \left\{ {r \atop l} \right\} \lambda^l, \qquad r = \sum_{j=1}^m r_j$$

where $\left\{ {r \atop l} \right\}$ is the Stirling number of the second kind.

Table 3. Conditional distributions and correlations.

$(l^{**}, l^*) = (0, 0)$

probabilities$\times 9$

W_1/W_2	0	1	2	3
0	1	1	1	0
1	1	1	1	0
2	1	1	1	0
3	0	0	0	0

$\mathrm{Cor}(W_1, W_2 \mid l^*, l^{**}) = 0$, $\mathrm{E}(W_i \mid l^*, l^{**}) = 1$

$(l^{**}, l^*) = (1, 0)$

probabilities$\times 18$

W_1/W_2	0	1	2	3
0	1	1	1	0
1	1	2	2	1
2	1	2	2	1
3	0	1	1	1

$\mathrm{Cor}(W_1, W_2 \mid l^*, l^{**}) = 3/11$, $\mathrm{E}(W_i \mid l^*, l^{**}) = 1.5$

$(l^{**}, l^*) = (1, 1)$

probabilities$\times 9$

W_1/W_2	0	1	2	3
0	0	0	0	0
1	0	1	1	1
2	0	1	1	1
3	0	1	1	1

$\mathrm{Cor}(W_1, W_2 \mid l^*, l^{**}) = 0$, $\mathrm{E}(W_i \mid l^*, l^{**}) = 2$

To find conditional moments $\mathrm{E}(W_i^{\underline{r}} \mid W_m)$, $\mathrm{E}(W_1^{\underline{r}} \mid W_2 = w_2)$ is calculated without loss of generality.

$$\begin{aligned}\mathrm{E}(W_1^{\underline{r}} \mid w_2) &= \mathrm{E}^k(\mathrm{E}(W_1^{\underline{r}} \mid k)p(w_2 \mid k))/\mathrm{E}^k(p(w_2 \mid k))\\ &= \mathrm{E}^k((\nu_1 k^r)e^{-\nu_2 k}k^{w_2})/\mathrm{E}^k(e^{-\nu_2 k}k^{w_2})\\ &= \sum_{l=1}^{w_2+r}\left\{\begin{matrix} w_2+r\\ l\end{matrix}\right\}(\lambda e^{-\nu_2})^l \Bigg/ \sum_{l=1}^{w_2}\left\{\begin{matrix} w_2\\ l\end{matrix}\right\}(\lambda e^{-\nu_2})^l,\end{aligned}$$

where the numerator is 1 if $w_2 = 0$. Hence $\mathrm{E}(W_1 \mid W_2 = 0) = \nu_1\lambda e^{-\nu_2}$, $\mathrm{E}(W_1 \mid W_2 = 1) = \nu_1(1 + \lambda e^{-\nu_2})$, and $\mathrm{E}(W_1 \mid W_2 = 2) = \nu_1(1 + 3\lambda e^{-\nu_2} + (\lambda e^{-\nu_2})^2)/(1 + \lambda e^{-\nu_2})$, and so on. $\mathrm{E}(W_1 \mid w_2)$ is not linear in w_2, and $\boldsymbol{W}$ does not satisfy Condition C.

Acknowledgements

The authors wish to thank the referees for their constructive suggestions.

References

Baba, K., Shibata, R. and Sibuya, M. (2004). Partial correlation and conditional correlation as measures of conditional independence, *Aust. N. Z. J. Stat.*, **46**, 657–664.

Casalis, M. (2000). Natural exponential families, Chapter 54 of Kotz, S., Balakrishnan, N., and Johnson, N. L., *Continuous Multivariate Distributions*, Volume 1, Wiley, New York.

Consonni, G. and Veronese, P. (1992). Conjugate priors for exponential families having quadratic variance functions, *J. Amer. Statist. Assoc.*, **87**, 1123–1127.

Dempster, A. P. (1972). Covariance selection, *Biometrics*, **28**, 157–175.

Gurland, J. (1957). Some interrelations among compound and generalized distributions, *Biometrika*, **44**, 265–268.

Jørgensen, B. (1997). *The Theory of Dispersion Models*, Chapman and Hall, London.

Lauritzen, S. L. (1975). General exponential models for discrete observations, *Scand. J. Statist.*, **2**, 23–33.

Letac, G. and Mora, M. (1990). Natural real exponential families with cubic variance functions, *Ann. Statist.*, **18**, 1–37.

Marshall, A. W. and Olkin, I. (1979). *Inequalities: Theory of Majorization and Its Applications*, Academic Press, New York.

Morris, C. N. (1983). Natural exponential families with quadratic variance functions: Statistical theory, *Ann. Statist.*, **11**, 515–529.

Shaked, M. (1971). A concept of positive dependence for exchangeable random variables, *Ann. Statist.*, **5**, 505–515.

Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*, Wiley, New York.