Partial, Conditional and Multiplicative Correlation Coefficients

by

Kunihiro Baba

E-mail: baba@sfc.keio.ac.jp

School of Fundamental Science and Technology
Graduate School of Science and Technology
Keio University
3-14-1, Hiyoshi, Kohoku-ku, Yokohama, 223-8522, JAPAN

A dissertation
submitted to Keio University
in partial fulfillment of the requirements
for Ph. D

2004

Contents

Chapter 1

Introduction

In this thesis, we develop several theorems around the second moments of multivariate distribution. Equivalence between partial and conditional correlations and properties of multiplicative correlations are investigated.

Dependence

In multivariate analysis, pairwise dependence of variates is a main issue and various measures have been proposed. Such dependence measures are classified into two groups, one is a rank based measure like Kendall's or Spearman's rank correlation coefficient and the other is a second moment based measure like Pearson's correlation coefficient,

$$\rho(X, Y) = \mathrm{cov}(X, Y)/\sqrt{\mathrm{var}(X)\mathrm{var}(Y)}.$$

Advantage of the rank based measure is in its robustness and it has a close link to general idea of dependence called *copula* (see, for example, Nelsen, 1999). The copula is a function $C : [0, 1]^2 \to [0, 1]$ such that

$$F(x, y) = C(F_1(x), F_2(y))$$

where $F_1(x)$ and $F_2(y)$ are the marginal distribution functions of X and Y. Since the independence of two variates is equivalent to $C(u, v) = uv$, we may derive various measures of dependence from discrepancy of the copula from uv. In fact, Spearman's rank correlation coefficient is written as

$$\rho_s = 12 \int_{[0,1]^2} \{C(u, v) - uv\} dudv.$$

However, such a rank based correlation coefficient captures a specific aspect of dependence and it is rather difficult to understand the exact meaning.

One of the advantages of Pearson's correlation coefficient is mathematically simple, considered to be a normalized inner product of two variates, so that it is easier to understand the meaning. Another advantage of Pearson's correlation coefficient is considered when it is extended to a measure of conditional dependence. Conditional correlation coefficient or partial correlation coefficient can be easily derived from Pearson's correlation coefficient, but similar measures are hard to be derived from rank based measures. In this thesis, we will concentrate our attention on such a Pearson-type correlation measure because of such advantages.

Partial correlation and conditional correlation

As we will give the exact definition in Section 2.1, the partial correlation of (X_1, X_2) given $\boldsymbol{Y}$ is the residual correlation of the linear least squares predictor of (X_1, X_2) to $\boldsymbol{Y}$, and represent correlations without the effects of the other variables. The partial correlation coefficient can be easily calculated from the inverse of the variance-covariance matrices.

On the other hand, the conditional correlation of (X_1, X_2) given $\boldsymbol{Y}$ is the ordinary correlation operator applied to (X_1, X_2) but evaluated with reference to the conditional distribution of (X_1, X_2) given $\boldsymbol{Y}$ rather than the marginal distribution of (X_1, X_2).

A purpose of this thesis is to seek the conditions for coincidence of the above two correlations which seem similar at first sight but are entirely different, and to search the multivariate distributions of which both correlations are equal. The motivation is derived from the graphical modeling to search causalities. In the graphical modeling, two vertexes are connected if and only if the corresponding variables are *not* conditionally independent. In order to confirm the conditional independence, particularly when the variables are continuous, it is a common practice to check whether or not the partial correlation coefficient is close enough to zero. See, for example, Whittaker (1990, Section 3.2) and Edwards (1995, Section 1.3). In the background of this practice, it is assumed that zero partial correlation coefficient suggests that the variables are conditionally independent, or nearly so.

One of the questions we ask is whether this assumption is true when we depart from normal distributions. Our answer is negative. Other than normal distribution there are distributions in which zero conditional correlation coefficient implies conditional independence. At least, if the conditional distribution of (X_1, X_2) given $\boldsymbol{Y}$ is bivariate normal distribution, then for any monotone increasing (decreasing) transforms ψ_1 and ψ_2, zero conditional correlation coefficient of $(Z_1, Z_2) =$

$(\psi_1(X_1), \psi_2(X_2))$ given $\boldsymbol{Y}$ is equivalent to the conditional independence of (Z_1, Z_2) given $\boldsymbol{Y}$ (see, Corollary 3 in Baba, Shibata and Sibuya, 2004). Thus, for log-normal distributions, zero conditional correlation coefficient implies conditional independence. However, whether zero partial correlation coefficient implies conditional independence for non-normal is doubtful. Actually, in the class of elliptical distributions which is a natural extension of normal distribution, the normal is only one which zero partial correlation coefficient implies conditional independence (see, Theorem 3 in Baba, Shibata and Sibuya, 2004), although in elliptical distributions partial correlation is identical to conditional correlation as Example 2.2.1 in Section 2.2.

Conditional independence is rather restrictive as a relation between two random variables, though it appeals to common sense and is based on probability theory. It is more reasonable in practice to replace conditional independence with zero partial correlation coefficient or zero conditional correlation coefficient. If it is replaced by zero partial correlation coefficient, disconnected vertices in a graphical model can be considered to be orthogonal to each other after the effects of other variables are removed by projection. Partial correlation coefficient is calculated more easily than conditional correlation coefficient which depends on the shape of the distribution, but it may further depart from conditional independence. Therefore, zero conditional correlation coefficient is preferable to zero partial correlation coefficient, considering that conditional independence is more meaningful than the conditional orthogonality.

Our next interest is equivalence of partial and conditional correlations apart from conditional independence, and we investigate this problem in Chapter 2.

Multiplicative correlation

A simple example of multivariate distribution which has a multiplicative correlation is the one led by so called *reduction method* (Mardia, 1970, p.74). For example, multivariate Poisson or gamma distribution is the case. Such distributions are derived as a joint distribution of $X_i = Z_0 + Z_i,\ i = 1, \ldots, n$. Here $Z_0, Z_1, \ldots, Z_n$ are independent Poisson or gamma distributed random variables (see, Johnson *et al.*, 1997, p.139; Kotz *et al.*, 2000, p.454). If we denote $\mathrm{E}(Z_i) = \mathrm{var}(Z_i) = \theta_i$ $(i = 0, 1, \ldots, n)$, then $\boldsymbol{X} = (X_1, \ldots, X_n)$ has a multiplicative correlation,

$$\rho(X_i, X_j) = \sqrt{\theta_0/(\theta_0 + \theta_i)}\sqrt{\theta_0/(\theta_0 + \theta_j)} \quad (i \neq j = 1, \ldots, n).$$

This is a specific example for multiplicative correlations, since it is not only multiplicative, but also *equi-covariance*, where any pairs of variables share the same covariance. A correlation structure similar to the equi-covariance is *equi-correlation* employed, for example in neuro science. Abbott and Dayan (1999) proposed, as a model of covariance of firing activities of neurons i and j,

$$Q_{ij} = \sigma^2\{\delta_{ij} + c(1-\delta_{ij})\}f_i(x)f_j(x),$$

where $f_i(x)$ is the mean firing rate of the neuron i for the input x, δ_{ij} is Kronecker delta, and c is a parameter. In fact, in this model the same correlation c is shared by any pairs or variables and it is clearly multiplicative. However, multiplicative correlation is not limited to such an equi-covariance or equi-correlation. Consider multinomial distribution, then the correlation matrix is multiplicative,

$$\rho(X_i, X_j) = -\sqrt{p_i p_j/(1-p_i)(1-p_j)},$$

but which is not equi-covariance nor equi-correlation. In Chapter 3, we will investigate the reason why such a multiplicative correlation or covariance appears so frequently.

For the discussion of multiplicative correlations, it is better to distinguish two types of matrices which are parameterized by $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)$, positive multiplicative $R^+(\boldsymbol{\delta}) = \mathrm{diag}(1-\boldsymbol{\delta}^2) + \boldsymbol{\delta}\boldsymbol{\delta}^\top$ or negative multiplicative $R^-(\boldsymbol{\delta}) = \mathrm{diag}(1+\boldsymbol{\delta}^2) - \boldsymbol{\delta}\boldsymbol{\delta}^\top$. Here $\mathrm{diag}(1-\boldsymbol{\delta}^2)$ is a diagonal matrix with its diagonal elements $1-\delta_1^2, \dots, 1-\delta_n^2$. It is easily seen that those two types of multiplicative matrices are exclusive, as far as $\boldsymbol{\delta}$ has more than three non zero elements. Note that the parameterization is unique except the sign of $\boldsymbol{\delta}$, in each class of positive or negative multiplicative matrices, provided that $\boldsymbol{\delta}$ has more than three non zero elements again. Royen (1991) derived a multivariate gamma distribution so as to have multiplicative correlations, where such a correlation structure is called *one-factorial*. The positive multiplicative correlation is also called *structure l* in Khatri (1967). Apparently, the covariance matrix corresponding to a multiplicative correlation is of the form $\mathrm{diag}(\boldsymbol{b}) \pm \boldsymbol{a}\boldsymbol{a}^\top$. We may discuss the multiplicative structure either through correlation or covariance, but we first discuss it through correlation, which is simpler and less redundant.

Plan of this thesis

In Chapter 2, we will discuss equivalence of partial and conditional correlations. We first prove a theorem which provides a necessary and sufficient condition for the coincidence of the partial covariance with the expectation of the conditional covariance which is another measure of conditional independence in Section 2.1. The condition is, essentially, linearity of the conditional expectation. A corollary of this theorem shows a necessary and sufficient condition for the partial variance-covariance being identical to the conditional variance-covariance. We show that this necessary and sufficient condition holds true for the order statistics of a random sample from exponential or geometric distribution which are *memoryless* distributions.

In Section 2.2, another corollary gives us that the partial correlation is identical to the conditional correlation if the conditional correlation coefficient is independent of the value of the condition, and also if the conditional expectation is linear. We call these two conditions Condition C, and investigate multivariate distributions satisfying Condition C. Simple examples satisfying Condition C are a family of elliptical distributions, and the order statistics of a random sample from the generalized Pareto distributions. A wider class satisfying Condition C is a family of the random vectors of which components are independent and reproductive with sum constraint. We also give a slight extension of the condition of this class, but its practical merit may be minor.

In Chapter 3, we will discuss multiplicative correlations. In Section 3.1, we first give a necessary and sufficient condition for the multiplicative matrix parameterized by a $\boldsymbol{\delta}$ to be a real correlation matrix, that is, non-negative (or positive) definite and all elements are less or equal to one in absolute value. The condition is extended for multiplicative covariance matrices, too.

In Section 3.2, we will investigate implications of multiplicative correlation to the structure in the elements of random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$. One is a factorization theorem. Each X_i is factorized into the sum of a common variable and an individual variable. The individual variables are uncorrelated but not always uncorrelated with the common variable. The individual variables are always correlated with the common variable for the case when the correlation is negative multiplicative. Therefore, the theorem is not strong enough to explain negative multiplicative correlations. We will give another theorem which shows that a specific type of negative multiplicative correlation implies an existence of a constraint such that $\sum_{i=1}^{n} X_i$ is almost surely constant, and vice versa.

In Section 3.3, we will show that several interesting relations hold true among various families of multivariate distributions with multiplicative correlation, through the invariance property of multiplicative correlation structure by unconditioning. Such families of multivariate distributions include homogeneous distribution or Liouville distribution. We also show that partial correlations or covariances are multiplicative for multiplicative correlation or covariances, and a simple relation holds true among the parameters.

In Section 3.4, we show that whether the multivariate distributions which are studied in Johnson it et al. (1997) and Kotz *et al.* (2000) have, or can have multiplicative correlations with the reasons as the final settlement of accounts of this chapter.

In Chapter 4, we will discuss two classes of multivariate distribution which is generated from independent samples from the natural exponential family (NEF). We will separately treat equivalence of partial and conditional correlations (in Chapter 2) and multiplicative correlation (in Chapter 3). But, two classes discussed in this chapter have same partial and conditional correlations, and also have multiplicative correlations. One is a class of the conditional distributions of independent NEF samples given the sum. We will show in Section 4.1 that this class has six distributions which include multinomial, hypergeometric, negative hypergeometric and Dirichlet distributions. Another is a class of distributions of independent NEF with quadratic variance function (NEF-QVF) samples when the parameter is randomized by the conjugate prior distribution. We will explain in Section 4.2 that the six distributions which include negative multinomial and multivariate beta type 2 distributions are members of this class.

Chapter 2

Partial Correlation and Conditional Correlation

2.1 Partial and conditional covariance

Let $\boldsymbol{X} = (X_1, \ldots, X_p)$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_q)$ be random vectors in $\mathbb{R}^p$ $(p \geq 2)$ and in $\mathbb{R}^q$ $(q \geq 1)$. In this section, we are interested in the partial and conditional correlations of X_i and X_j $(i \neq j = 1, \ldots, p)$ given $\boldsymbol{Y}$. The notation $i \neq j = 1, \ldots, p$ denotes that $i = 1, \ldots, p,\ j = 1, \ldots, p$ and $i \neq j$. We hereafter assume, for simplicity, that the variance-covariance matrix of $\boldsymbol{Y}$ is positive definite. The partial variance-covariance matrix for $\boldsymbol{X}$ is defined as

$$\boldsymbol{\Sigma}_{\boldsymbol{XX}\cdot\boldsymbol{Y}} = (\sigma_{ij\cdot\boldsymbol{Y}})_{i,j=1,\ldots,p}$$

which can be calculated as $\boldsymbol{\Sigma}_{\boldsymbol{XX}\cdot\boldsymbol{Y}} = \boldsymbol{\Sigma}_{\boldsymbol{XX}} - \boldsymbol{\Sigma}_{\boldsymbol{XY}}\boldsymbol{\Sigma}_{\boldsymbol{YY}}{}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{YX}}$ by partitioning the variance-covariance matrix of $(\boldsymbol{X},\ \boldsymbol{Y})$ into

$$\mathrm{V}\left(\begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{pmatrix}\right) = \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{XX}} & \boldsymbol{\Sigma}_{\boldsymbol{XY}} \\ \boldsymbol{\Sigma}_{\boldsymbol{YX}} & \boldsymbol{\Sigma}_{\boldsymbol{YY}} \end{pmatrix}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{XX}}$ is $p \times p$, $\boldsymbol{\Sigma}_{\boldsymbol{XY}}$ is $p \times q$, and $\boldsymbol{\Sigma}_{\boldsymbol{YY}}$ is $q \times q$. The partial correlation coefficient of X_i and X_j given $\boldsymbol{Y}$ is then

$$\rho_{ij\cdot\boldsymbol{Y}} = \frac{\sigma_{ij\cdot\boldsymbol{Y}}}{\sqrt{\sigma_{ii\cdot\boldsymbol{Y}}\sigma_{jj\cdot\boldsymbol{Y}}}}.$$

The partial variance or covariance given $\boldsymbol{Y}$ can be considered as the variance or covariance between residuals of projections of $\boldsymbol{X}$ on the linear space spanned by elements of $\boldsymbol{Y}$, that is

$$\sigma_{ij\cdot\boldsymbol{Y}} = \mathrm{cov}\left(X_i - \hat{X}_i(\boldsymbol{Y}),\ X_j - \hat{X}_j(\boldsymbol{Y})\right) \qquad \text{for} \quad i, j = 1, \ldots, p$$

where $\hat{\boldsymbol{X}}(\boldsymbol{Y}) = \mathrm{E}(\boldsymbol{X}) + \boldsymbol{\Sigma}_{\boldsymbol{X}\boldsymbol{Y}}\boldsymbol{\Sigma}_{\boldsymbol{Y}\boldsymbol{Y}}^{-1}(\boldsymbol{Y} - \mathrm{E}(\boldsymbol{Y}))$.

In a similar way, the conditional covariance of X_i and X_j given $\boldsymbol{Y}$ is defined through

$$\mathrm{cov}(X_i, X_j|\boldsymbol{Y}) = \mathrm{E}\left((X_i - \mathrm{E}(X_i|\boldsymbol{Y}))\,(X_j - \mathrm{E}(X_j|\boldsymbol{Y}))\,|\boldsymbol{Y}\right).$$

We use the following notation for the conditional covariance matrix

$$\boldsymbol{\Sigma}_{\boldsymbol{X}\boldsymbol{X}|\boldsymbol{Y}} = (\sigma_{ij|\boldsymbol{Y}})_{i,j=1,\ldots,p}$$

and for the conditional correlation coefficient

$$\rho_{ij|\boldsymbol{Y}} = \frac{\sigma_{ij|\boldsymbol{Y}}}{\sqrt{\sigma_{ii|\boldsymbol{Y}}\sigma_{jj|\boldsymbol{Y}}}}.$$

The expectation of the conditional covariance is not necessarily equal to the covariance, that is, $\mathrm{E}(\boldsymbol{\Sigma}_{\boldsymbol{X}\boldsymbol{X}|\boldsymbol{Y}}) \not\equiv \mathrm{V}(\boldsymbol{X})$. Notice that partial correlation coefficient, $\rho_{ij\cdot\boldsymbol{Y}}$, is deterministic, but conditional correlation coefficient, $\rho_{ij|\boldsymbol{Y}}$, is random.

The following example illustrates a relationship between the partial covariance and the conditional covariance.

Example 2.1.1. Consider a random 3×1 vector (X_1, X_2, Y). To investigate conditional correlation and partial correlation, it is sufficient for us to specify the conditional distribution of (X_1, X_2) given $Y = y$ as

$$H\left(\frac{x_1 - \mu_1(y)}{\sigma_1(y)}, \frac{x_2 - \mu_2(y)}{\sigma_2(y)}\right),\ \mu_1(y) \in \mathbb{R},\ \mu_2(y) \in \mathbb{R},\ \sigma_1(y),\ \sigma_2(y) > 0$$

with a two-dimensional distribution function H. This type of specification is similar to a copula introduced in Introduction. If we further assume that H has zero means, unit variances and correlation coefficient θ, then the conditional expectations of X_1 and X_2 are $\mu_1(y)$ and $\mu_2(y)$, respectively, and the conditional covariance matrix is written as

$$\begin{pmatrix} \sigma_1(y)^2 & \theta\sigma_1(y)\sigma_2(y) \\ \theta\sigma_1(y)\sigma_2(y) & \sigma_2(y)^2 \end{pmatrix}.$$

Hence, the conditional correlation coefficient is the constant θ but the conditional covariance depends on y. To see how conditional covariance or correlation coefficient varies with y and when it coincides with the partial covariance or correlation coefficient, consider the following three cases.

(i) If $\mu_i(y) = a_i + b_i y$ and $\sigma_i(y) = \sigma_i$ for $i = 1, 2$, then $\sigma_{12|Y} = \theta\sigma_1\sigma_2 = \sigma_{12\cdot Y}$.

(ii) If $\mu_i(y) = a_i + b_i y$ for $i = 1, 2$, but $\sigma_1(y) = \sigma_2(y) = \sigma(y)$ depends on y, then $\sigma_{12|Y} = \theta\sigma(Y)^2$ and $\sigma_{12\cdot Y} = \theta\mathrm{E}\left(\sigma(Y)^2\right)$ so that these two expressions are only equal in the mean. Hence the conditional variance is not necessarily equal to the partial covariance. However, the conditional and partial correlation coefficients are equal; $\rho_{12|Y} = \theta = \rho_{12\cdot Y}$.

(iii) If $\mu_i(y) = y^2$ but $\sigma_i(y) = \sigma_i$ for $i = 1, 2$, then $\sigma_{12|Y} = \theta\sigma_1\sigma_2$ and $\sigma_{12\cdot Y} = (\theta + \mathrm{var}(Y^2)(1 - \rho(Y, Y^2)^2))\,\sigma_1\sigma_2$. Therefore, such variances or correlations are equal to each other only if $\rho(Y, Y^2) = \pm 1$.

These examples suggest that the linearity of conditional expectation is a key to the equivalence of the conditional covariance and the partial covariance.

The following theorem gives us a necessary and sufficient condition for the equivalence of the partial variance-covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{XX}\cdot\boldsymbol{Y}}$ to the expectation of the conditional variance-covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{XX}|\boldsymbol{Y}}$. In fact, Corollary 2.1.1 is a generalization of Example 2.1.1 (i), and Corollary 2.2.1 is a generalization of Example 2.1.1 (ii).

Theorem 2.1.1. *For any random vectors $\boldsymbol{X} = (X_1, \ldots, X_p)$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_q)$, the following two conditions are equivalent.*

(i) $\mathrm{E}(\boldsymbol{X}|\boldsymbol{Y}) = \boldsymbol{\alpha} + \boldsymbol{BY}$ *for a vector $\boldsymbol{\alpha}$ and a matrix $\boldsymbol{B}$,*

(ii) $\boldsymbol{\Sigma}_{\boldsymbol{XX}\cdot\boldsymbol{Y}} = \mathrm{E}\left(\boldsymbol{\Sigma}_{\boldsymbol{XX}|\boldsymbol{Y}}\right)$.

Proof. Since $\boldsymbol{\Sigma}_{\boldsymbol{XX}\cdot\boldsymbol{Y}} = \mathrm{E}\left(\boldsymbol{\Sigma}_{\boldsymbol{XX}|\boldsymbol{Y}}\right) + \mathrm{V}\left(\mathrm{E}\left(\boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Y})\middle|\,\boldsymbol{Y}\right)\right)$, it follows that $\boldsymbol{\Sigma}_{\boldsymbol{XX}\cdot\boldsymbol{Y}} = \mathrm{E}\left(\boldsymbol{\Sigma}_{\boldsymbol{XX}|\boldsymbol{Y}}\right)$ is equivalent to $\mathrm{V}\left(\mathrm{E}\left(\boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Y})\middle|\,\boldsymbol{Y}\right)\right) = O$. This is further equivalent to $\mathrm{E}\left(\boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Y})\middle|\,\boldsymbol{Y}\right) = \boldsymbol{\beta}$ a.s. for a constant vector $\boldsymbol{\beta}$. We get the result by letting $\boldsymbol{B} = \boldsymbol{\Sigma}_{\boldsymbol{XY}}\boldsymbol{\Sigma}_{\boldsymbol{YY}}{}^{-1}$ and $\boldsymbol{\alpha} = \boldsymbol{\beta} + \mathrm{E}(\boldsymbol{X}) - \boldsymbol{B}\mathrm{E}(\boldsymbol{Y})$. □

Lawrance (1976, Results II) showed that (i) implies (ii) for the case when $\boldsymbol{Y}$ is a scalar variable. Now, we have the following corollary as a direct consequence of Theorem 2.1.1.

Corollary 2.1.1. *For any random vectors $\boldsymbol{X} = (X_1, \ldots, X_p)$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_q)$, the following two conditions are equivalent.*

(i) $\mathrm{E}(\boldsymbol{X}|\boldsymbol{Y}) = \boldsymbol{\alpha} + \boldsymbol{BY}$ *for a vector $\boldsymbol{\alpha}$ and a matrix $\boldsymbol{B}$ and $\boldsymbol{\Sigma}_{\boldsymbol{XX}|\boldsymbol{Y}}$ is independent of $\boldsymbol{Y}$,*

(ii) $\boldsymbol{\Sigma}_{\boldsymbol{XX}\cdot\boldsymbol{Y}} = \boldsymbol{\Sigma}_{\boldsymbol{XX}|\boldsymbol{Y}}$ *a.s.*

Kelker (1970, Theorem 6 and Theorem 7) showed that the multivariate normal distribution is the only distribution that satisfies the condition (i) of Corollary 2.1.1 in the class of elliptical distributions.

Other than the normally distributed random vector we found a random vector for which Corollary 2.1.1 holds true. It is the order statistics of a random sample from exponential or geometric distribution, and we show that in the following example. However, since we do not know any other vectors for which Corollary 2.1.1 hold except these two, we consider Corollary 2.1.1 holds only in the restrictive cases.

Example 2.1.2. Assume that a random vector $\boldsymbol{Z} = (Z_1, \ldots, Z_{p+q})$ is the order statistics of a random sample from exponential or geometric distribution such that $Z_1 \geq \cdots \geq Z_{p+q}$, and define p-dimensional random vector $\boldsymbol{X}$ and q-dimensional random vector $\boldsymbol{Y}$ as $\boldsymbol{X} = (Z_1, \ldots, Z_p)$ and $\boldsymbol{Y} = (Z_{p+1}, \ldots, Z_{p+q})$.

Since an exponential and a geometric distributions are memoryless, it holds true

$$\mathrm{P}(X_1 - y_1 \geq x_1, \cdots, X_p - y_1 \geq x_p | Y_1 = y_1) = \mathrm{P}(X_1 \geq x_1, \cdots, X_p \geq x_p)$$

for $Y_1 = y_1$. Since $\boldsymbol{Y}$ is order statistics, conditioning by $\boldsymbol{Y} = \boldsymbol{y} = (y_1, \ldots, y_q)$ is equivalence to by $Y_1 = y_1$. Thus, we have

$$\mathrm{P}(X_1 - y_1 \geq x_1, \cdots, X_p - y_1 \geq x_p | \boldsymbol{Y} = \boldsymbol{y}) = \mathrm{P}(X_1 \geq x_1, \cdots, X_p \geq x_p),$$

and

$$\mathrm{E}(\boldsymbol{X}|\boldsymbol{Y} = \boldsymbol{y}) = y_1 + \mathrm{E}(\boldsymbol{X}) \quad \text{and} \quad \mathrm{V}(\boldsymbol{X}|\boldsymbol{Y} = \boldsymbol{y}) = \mathrm{V}(\boldsymbol{X}).$$

From Corollary 2.1.1, it holds true $\boldsymbol{\Sigma}_{\boldsymbol{XX}\cdot\boldsymbol{Y}} = \boldsymbol{\Sigma}_{\boldsymbol{XX}|\boldsymbol{Y}}$ a.s.

2.2 Equivalence of partial and conditional correlations

As a corollary of Theorem 2.1.1, we have the following.

Corollary 2.2.1. *For any random vectors $\boldsymbol{X} = (X_1, \ldots, X_p)$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_q)$, if there exists a vector $\boldsymbol{\alpha}$ and a matrix $\boldsymbol{B}$ such that*

$$\mathrm{E}(\boldsymbol{X}|\boldsymbol{Y}) = \boldsymbol{\alpha} + \boldsymbol{BY} \quad \textit{and} \quad \rho_{ij|\boldsymbol{Y}} \textit{ does not depend on } \boldsymbol{Y} \quad \textit{for} \quad i \neq j = 1, \ldots, p,$$

then $\rho_{ij\cdot\boldsymbol{Y}} = \rho_{ij|\boldsymbol{Y}}$ *a.s.*

Proof. From Theorem 2.1.1, if $\mathrm{E}(\boldsymbol{X}|\boldsymbol{Y})$ is a linear function of $\boldsymbol{Y}$, then $\rho_{ij\cdot\boldsymbol{Y}} = \mathrm{E}\left(\rho_{ij|\boldsymbol{Y}}\right)$. The assertion of the corollary holds true, since $\rho_{ij|\boldsymbol{Y}}$ is independent of $\boldsymbol{Y}$.

□

We investigate the classes of multivariate distributions for which Corollary 2.2.1 holds true. For convenience, we say the assumption of Corollary 2.2.1 *Condition C* as follows:

Definition (Condition C). *For random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$,*

(i) there exists a vector $\boldsymbol{\alpha}$ and a matrix $\boldsymbol{B}$ such that $\mathrm{E}(\boldsymbol{X}|\boldsymbol{Y}) = \boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{Y}$ *and*

(ii) $\rho_{ij|\boldsymbol{Y}}$ does not depend on $\boldsymbol{Y}$.

The following two examples satisfy the Condition C.

Example 2.2.1 (Elliptical distribution). Elliptical distribution is a family of distributions whose characteristic function takes the form

$$\Psi(\boldsymbol{t}) = \exp\left(i\boldsymbol{t}^{\mathsf{T}}\boldsymbol{\mu}\right)\phi\left(\boldsymbol{t}^{\mathsf{T}}\boldsymbol{\Sigma}\boldsymbol{t}\right)$$

for some scalar function ϕ (see, for example, Fang *et al.*, 1990 p.31). This family is denoted by $\mathrm{EC}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$. From Cambanis *et al.* (1981 Corollary 5), if $(\boldsymbol{X}, \boldsymbol{Y}) \sim \mathrm{EC}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$, then

$$\mathrm{E}(\boldsymbol{X}|\boldsymbol{Y}) = \mathrm{E}(\boldsymbol{X}) + \boldsymbol{\Sigma}_{\boldsymbol{XY}}\boldsymbol{\Sigma}_{\boldsymbol{YY}}{}^{-1}(\boldsymbol{Y} - \mathrm{E}(\boldsymbol{Y})) \quad \text{and} \quad \boldsymbol{\Sigma}_{\boldsymbol{XX}|\boldsymbol{Y}} = s(\boldsymbol{Y})\boldsymbol{\Sigma}^*,$$

where s is a function and the matrix $\boldsymbol{\Sigma}^*$ is independent of the value of $\boldsymbol{Y}$. Note that the conditional distribution is also elliptical. This shows that the Condition C is satisfied for the elliptical distributions.

The elliptical distribution is a natural generalization of the multivariate normal distribution. However, other than normal distribution partial covariance is not equal to conditional covariance as mentioned in Section 2.1, although partial correlation coincides with conditional correlation. Moreover, Theorem 3 in Baba, Shibata and Sibuya (2004) showed that zero partial correlation coefficient or zero conditional correlation coefficient does not imply conditional independence except for normal distribution.

Example 2.2.2 (Distribution generated from generalized Pareto distribution). Generalized Pareto distribution is defined as

$$\bar{F}(x;\gamma,a) = \begin{cases} (1+\gamma x/a)^{-1/\gamma} & \text{if } \gamma \neq 0 \\ \exp(-x/a) & \text{if } \gamma = 0 \end{cases} \qquad (a > 0),$$

where $\bar{F} = 1 - F$ is the survival function (sf). The distribution is concentrated on $x > 0$ if $\gamma \geq 0$ and on $0 < x < -a/\gamma$ otherwise. The details of the distribution can be found in, for example, Johnson *et al.* (1994), Embrechts *et al.* (1997) and Coles (2001). The reason why this family of distribution is called generalized Pareto distribution is that it covers wider range of distributions like exponential ($\gamma = 0$) or uniform ($\gamma = -1$) distribution by allowing $\gamma \leq 0$. It is known that the rth moment is finite if and only if $r < 1/\gamma$.

If this distribution, denoted by $\mathsf{GPrt}(\gamma, a)$, is left truncated at u, the truncated sf is

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

which is $\mathsf{GPrt}(\gamma, a+\gamma u)$.

Now, assume that a random vector $\boldsymbol{Z} = (Z_1, \ldots, Z_{p+q})$ is the order statistics of a random sample from $\mathsf{GPrt}(\gamma, a)$ such that $Z_1 \geq \cdots \geq Z_{p+q}$, and define p-dimensional random vector $\boldsymbol{X}$ and q-dimensional random vector $\boldsymbol{Y}$ as $\boldsymbol{X} = (Z_1, \ldots, Z_p)$ and $\boldsymbol{Y} = (Z_{p+1}, \ldots, Z_{p+q})$. We have

$$(X_i|\boldsymbol{Y} = \boldsymbol{y}) \ \stackrel{d}{=}\ (X_i|Y_1 = y_1) \ \stackrel{d}{=}\ y_1 + (1+\gamma y_1/a)X_i \quad (i = 1, \ldots, p)$$

where $\stackrel{d}{=}$ is the symbol for equality in distribution. Thus, the conditional moments of $\boldsymbol{X}$ given $\boldsymbol{Y} = \boldsymbol{y}$ are

$$\begin{aligned}\mathrm{E}(X_i|\boldsymbol{Y}) &= y_1 + (1+\gamma y_1/a)\mathrm{E}(X_i), \quad \mathrm{var}(X_i|\boldsymbol{Y}) = (1+\gamma y_1/a)^2\mathrm{var}(X_i),\\ \mathrm{cov}(X_i, X_j|\boldsymbol{Y}) &= (1+\gamma y_1/a)^2\mathrm{cov}(X_i, X_j), \quad \text{and} \quad \rho(X_i, X_j|\boldsymbol{Y}) = \rho(X_i, X_j),\end{aligned}$$

where $i \neq j = 1, \ldots, p$. Hence, the Condition C is satisfied for the order statistics of the generalized Pareto distribution, $\mathsf{GPrt}(\gamma, a)$, with $\gamma < 1/2$.

Next, we show a wider class of multivariate distributions satisfying the Condition C rather than Example 2.2.1 and Example 2.2.2. The class is a family of random vectors of which components are independent and reproductive with sum constraint, and includes multinomial and multivariate hypergeometric distributions. At first, we define the reproductive family of distributions, $\mathcal{F}$, as follows:

Definition (A Family of Distributions $\mathcal{F}$). $\mathcal{F} = \{F_\theta : \ \theta \in \Theta\}$ *is defined as the family of distribution functions which have a semigroup property such that* $F_{\theta_1} * F_{\theta_2} = F_{\theta_1+\theta_2} \in \mathcal{F}$ *for the convolution of any* $F_{\theta_1}, F_{\theta_2} \in \mathcal{F}$. *Parameter space* Θ *is assumed to be* $(0, \infty)$ *or the set of all natural numbers. (If* $\Theta = (0, \infty)$, $\mathcal{F}$ *is a class of infinitely divisible distributions.)*

A distribution of the natural exponential family (see, Chapter 4) with density (4.2) is a member of $\mathcal{F}$ when we consider the convolution parameter ν as θ. Other than the natural exponential family, Cauchy distribution is a member of $\mathcal{F}$. If a random variable X has density

$$f(x) = (\pi\lambda)^{-1}\left[1 + \left\{\frac{x-\theta}{\lambda}\right\}^2\right]^{-1}, \quad \lambda > 0,$$

X has Cauchy distribution and we denote $X \sim \mathrm{Ca}(\theta, \lambda)$. If $X_1, \ldots, X_n$ are independent and are distributed $X_i \sim \mathrm{Ca}(\theta_i, \lambda_i)$, $i = 1, \ldots, n$, then $\sum_{i=1} X_i \sim \mathrm{Ca}(\sum_{i=1}^n \theta_i, \sum_{i=1}^n \lambda_i)$ (see, Johnson *et al.*, 1994, p. 301). Thus, Cauchy distribution $\mathrm{Ca}(\theta, \lambda)$ is a member of $\mathcal{F}$ in terms of θ and λ.

We consider the class of random vectors of which components are independent, distributed as $F_\theta \in \mathcal{F}$, and have sum constraint. The following lemma shows conditional moments of the random vector given its total sum.

Lemma 2.2.1. *If random variables $\boldsymbol{Z} = (Z_1, \ldots, Z_n)$ are mutually independent and $Z_i \overset{d}{=} F_{\theta_i} \in \mathcal{F}$ $(i = 1, \ldots, n)$, then conditional expectations and correlation coefficients of $\boldsymbol{Z}$ given $T = \sum_{k=1}^n Z_k = t$ are*

$$\mathrm{E}(Z_i|T=t) = t\xi_i \quad \textit{and} \quad \rho(Z_i, Z_j|T=t) = -\sqrt{\frac{\xi_i\xi_j}{(1-\xi_i)(1-\xi_j)}} \quad (i \neq j = 1, \ldots, n)$$

where $\xi_i = \theta_i \,/\sum_{k=1}^n \theta_k$.

Proof. From the linearity of the expectation, we have

$$\mathrm{E}(Z_i|T=t) = \mathrm{E}\left(t - \sum_{k\neq i}^{n} Z_k \,\middle|\, T=t\right) = t - \sum_{k\neq i}^{n} \mathrm{E}(Z_k|T=t) \quad (i = 1, \ldots, n).$$

It is sufficient to consider the case when θ_i $(i = 1, 2, \ldots, n)$ are rational numbers since $\mathcal{F}$ is a continuous parametric family of distributions when $\Theta = (0, \infty)$. Then we can write $\theta_i = b_i/a$ where a and the b_i are positive integers. From the property of $\mathcal{F}$, there exist independent random variables $(Z_{i1}, Z_{i2}, \ldots, Z_{ib_i})$ for each Z_i such that

$$Z_i = \sum_{j=1}^{b_i} Z_{ij} \overset{d}{=} F_{\theta_i} \in \mathcal{F} \quad \text{and} \quad Z_{ij} \overset{d}{=} F_{\frac{1}{a}} \in \mathcal{F} \quad (j = 1, 2, \ldots, b_i).$$

We have then

$$\mathrm{E}(Z_i|T=t) = \sum_{j=1}^{b_i} \mathrm{E}(Z_{1j}|T=t) = t - \sum_{k\neq i}^{n}\sum_{j=1}^{b_k} \mathrm{E}(Z_{kj}|T=t).$$

Since the variables $\{Z_{ij} :\ j = 1, 2, \ldots, b_i\}$ are exchangeable given $T = t$, $i = 1, \ldots, n$, we have

$$\mathrm{E}(Z_i|T = t) = b_i \mathrm{E}(Z_{ij}|T = t) = tb_i \Bigg/ \left(\sum_{k=1}^{n} b_k\right) = t\theta_i \Bigg/ \left(\sum_{k=1}^{n} \theta_k\right).$$

For the conditional second moments we can write

$$\mathrm{E}\left({Z_i}^2 \middle| T = t\right) = \mathrm{E}\left(Z_i \left(t - \sum_{k \neq i}^{n} Z_k \right) \middle| T = t \right) = b_1 t^2 \Bigg/ \left(\sum_{k=1}^{n} b_k\right) - b_1 \left(\sum_{k \neq i}^{n} b_k\right) \eta$$

and

$$\mathrm{E}(Z_i Z_j|T = t) = \mathrm{E}\left(\left(\sum_{k=1}^{b_i} Z_{ik}\right) \left(\sum_{k=1}^{b_j} Z_{jk}\right) \middle| T = t \right) = b_i b_j \eta,$$

with $\eta = \mathrm{E}(Z_{ik_i} Z_{jk_j}|T = t)$ $(k_i = 1, \ldots, b_i,\ k_j = 1, \ldots, b_j)$. From these results, the conditional variance-covariance matrix of $\boldsymbol{Z}$ given T is

$$\mathrm{V}(\boldsymbol{Z}|T = t) = (t^2 - \eta') \left\{ \mathrm{diag}(\boldsymbol{\xi}) - \boldsymbol{\xi}\boldsymbol{\xi}^{\top} \right\} \quad \boldsymbol{\xi} = (\xi_1, \ldots, \xi_n) \tag{2.1}$$

where $\eta' = a^2\theta^2\eta$, and we have the latter part of the result. □

The following theorem shows that if $\boldsymbol{Z}$ in Lemma 2.2.1 is partitioned into $(\boldsymbol{X}, \boldsymbol{Y})$, the Condition C is satisfied for $(\boldsymbol{X}, \boldsymbol{Y})$ provided that T is given.

Theorem 2.2.1. *Assume that random variables $Z_1, \ldots, Z_{p+q}$ are mutually independent and distributed as $Z_i \overset{d}{=} F_{\theta_i} \in \mathcal{F}$ $(i = 1, \ldots, p + q)$, and define a p-dimensional random vector $\boldsymbol{X}$ and a q-dimensional random vector $\boldsymbol{Y}$ as $\boldsymbol{X} = (Z_1, \ldots, Z_p)$ and $\boldsymbol{Y} = (Z_{p+1}, \ldots, Z_{p+q})$. The conditional expectations and correlation coefficients of $\boldsymbol{X}$ given $\boldsymbol{Y}$ and $T = \sum_{i=1}^{p} X_i + \sum_{i=1}^{q} Y_i$ are*

$$\mathrm{E}(X_i|\boldsymbol{Y} = \boldsymbol{y}, T = t) = \left(t - \sum_{k=1}^{q} y_k \right) \tau_i, \qquad \tau_i = \theta_i \Bigg/ \sum_{k=1}^{p} \theta_k\ ,$$

and

$$\rho(X_i, X_j|\boldsymbol{Y} = \boldsymbol{y}, T = t) = -\sqrt{\frac{\tau_i \tau_j}{(1 - \tau_i)(1 - \tau_j)}} \quad (i \neq j = 1, \ldots, n)$$

for $\boldsymbol{y} = (y_1, \ldots, y_q)$, and the Condition C is satisfied for $(\boldsymbol{X}, \boldsymbol{Y})$ provided that T is given.

Proof. Under the condition that T and $\boldsymbol{Y}$ are given, $\boldsymbol{X}$ has the same stochastic structure as Lemma 2.2.1 with t and $\sum_{k=1}^{p+q}\theta_k$ replaced by $t-\sum_{i=1}^{q}y_i$ and $\sum_{k=1}^{p}\theta_k$, respectively. Thus, we directly have the conditional expectations and correlation coefficients of $\boldsymbol{X}$ given $\boldsymbol{Y}$ and T from Lemma 2.2.1. Hence, the Condition C is satisfied provided that T is given. □

We will show the examples which are distributions satisfying this theorem in Section 4.1. They are conditional distributions of independent samples distributed the natural exponential family given the sum.

In the proofs of Lemma 2.2.1 and Theorem 2.2.1, we use the fact that to assume that the independent variables have a distribution of $\mathcal{F}$ and that total sum of them is given implies that they can be factorized into variables of minimal unit ($\{Z_{ij}\}_{i=1,\ldots,n,j=1,\ldots,b_i}$ in Lemma 2.2.1) which are conditionally exchangeable when $\boldsymbol{Y}$ and T are given. However, for satisfying the Condition C, the assumption of conditional exchangeability is too strong, and it is sufficient that conditional first and second order moments of variables of minimal unit are assumed to be all same.

Hence to extend Theorem 2.2.1, let us introduce "partial sums of $\boldsymbol{Z}$" by which the property of the distribution family $\mathcal{F}$ is replaced, as follows:

Definition (Partial sums of $\boldsymbol{Z}$). *For a random vector $\boldsymbol{Z}=(Z_1,\ldots,Z_n)$, partition the index set $\{1,\ldots,n\}$ into $(p+q)$ parts $L_1,L_2,\ldots,L_{p+q}$ where $|L_j|=l_j>0$ ($\sum_{j=1}^{p+q}l_j=n$). Define a partial sums $(\boldsymbol{X},\boldsymbol{Y})=(X_1,\ldots,X_p,Y_1,\ldots,Y_q)$ of $\boldsymbol{Z}$ as*

$$X_j=\sum_{i=1}^{n}\mathrm{I}[i\in L_j]\,Z_i\ (j=1,\ldots,p)\quad\text{and}\quad Y_j=\sum_{i=1}^{n}\mathrm{I}[i\in L_{p+j}]\,Z_i\ (j=1,\ldots,q)$$

where I is the indicator function.

The following theorem is a slight extension of Theorem 2.2.1.

Theorem 2.2.2. *Let a random vector $(\boldsymbol{X},\boldsymbol{Y})=(X_1,\ldots,X_p,Y_1,\ldots,Y_q)$ be a partial sums of a random vector $\boldsymbol{Z}=(Z_1,\ldots,Z_n)$. If $\mathrm{E}(Z_i|\boldsymbol{Y},T=\sum_{k=1}^{n}Z_k)$, $\mathrm{var}(Z_i|\boldsymbol{Y},T)$ and $\mathrm{cov}(Z_i,Z_j|\boldsymbol{Y},T)$ are constants for any $i\neq j=1,\ldots,n$, then the Condition C is satisfied for $(\boldsymbol{X},\boldsymbol{Y})$ provided that T is given.*

Proof. Give $\boldsymbol{Y}=\boldsymbol{y}$ and $T=t$, and let the index subset $\cup_{j\in\{i,\ldots,p\}}L_j$ denote $L_{\boldsymbol{X}}$. Since $\sum_{i\in L_{\boldsymbol{X}}}\mathrm{E}(Z_i|\boldsymbol{Y},T)=t-y$ where $y=\sum_{j=1}^{q}y_j$, it holds true $\mathrm{E}(Z_i|\boldsymbol{Y},T)=(t-y)/(n-l_y)$ where $l_y=\sum_{j=p+1}^{p+q}$ for any $i\in L_{\boldsymbol{X}}$. Thus, we have $\mathrm{E}(X_j|\boldsymbol{Y},T)=l_j(t-y)/(n-l_y)$ $(j=1,\ldots,p)$, and they are linear combination of $\boldsymbol{y}$.

Let $\mathrm{var}(Z_i|\boldsymbol{Y},T)$ and $\mathrm{cov}(Z_i,Z_j|\boldsymbol{Y},T)$ for $i \neq j \in L_{\boldsymbol{X}}$ denote $\sigma^2_{\boldsymbol{y},t}$ and $\kappa_{\boldsymbol{y},t}$, respectively. Then we have

$$\mathrm{var}(X_j|\boldsymbol{Y},T) = l_j\sigma^2_{\boldsymbol{y},t} + l_j(l_j-1)\kappa_{\boldsymbol{y},t} \quad \text{and} \quad \mathrm{cov}(X_i,X_j|\boldsymbol{Y},T) = l_il_j\kappa_{\boldsymbol{y},t}$$

for $i \neq j = 1,\ldots,p$. Since

$$\begin{aligned}\mathrm{E}(Z_j^2|\boldsymbol{Y},T) &= \mathrm{E}\left\{Z_j\left(t-y-\sum_{i\neq j}Z_i\right)|\boldsymbol{Y},T\right\} \\ &= (t-y)\mathrm{E}(Z_j|\boldsymbol{Y},T)-(n-l_y-1)\mathrm{E}(Z_iZ_j|\boldsymbol{Y},T),\end{aligned}$$

$\sigma^2_{\boldsymbol{y},t}$ and $\kappa_{\boldsymbol{y},t}$ have the restriction $\sigma^2_{\boldsymbol{y},t} = -(n-l_y-1)\kappa_{\boldsymbol{y},t}$. Therefore, we have

$$\mathrm{V}(\boldsymbol{X}|\boldsymbol{Y},T) = -(n-l_y)^2\kappa_{\boldsymbol{y},t}\left(\,\mathrm{diag}(\boldsymbol{\upsilon})-\boldsymbol{\upsilon}\boldsymbol{\upsilon}^\top\right) \tag{2.2}$$

where $\upsilon_j = l_j/(n-l_y)$ $(j=1,\ldots,p)$, and $\rho(X_i,X_j|\boldsymbol{Y},T) = -\sqrt{\upsilon_i\upsilon_j/(1-\upsilon_i)(1-\upsilon_j)}$ $(i\neq j=1,\ldots,p)$ is not dependent of $\boldsymbol{y}$. Hence, Condition C is satisfied. □

There is one more class satisfying the Condition C, it is NEF-QVF-CP. However, we will explain the class in Section 4.2.

Now, (2.1) in Lemma 2.2.1 and (2.2) in Theorem 2.2.2 shows that the variance-covariance matrices are a special form, $\mathrm{diag}(\boldsymbol{b})-\boldsymbol{a}\boldsymbol{a}^\top$. We say this form *multiplicative covariance matrix*, and will discuss the structure in the following chapter.

Chapter 3

Multiplicative Correlation

3.1 Feasible value of parameters

We give a necessary and sufficient condition for a multiplicative matrix to be a correlation matrix, separately for the case of positive multiplicative matrix $R^{+}(\boldsymbol{\delta})$ and for negative multiplicative matrix $R^{-}(\boldsymbol{\delta})$.

3.1.1 Positive multiplicative correlation

We need the following lemma for the proof of Theorem 3.1.1.

Lemma 3.1.1. *The following inequality holds true for the eigenvalues* $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ *of positive multiplicative matrix* $R^{+}(\boldsymbol{\delta})$*.*

$$1-\delta_{k_1}^2 = \lambda_1 = \cdots = \lambda_{n_1-1} < \lambda_{n_1} < 1-\delta_{k_2}^2 = \lambda_{n_1+1} = \cdots = \lambda_{n_1+n_2-1} < \lambda_{n_1+n_2} < \cdots$$

$$\cdots < \lambda_{n-n_m} < 1-\delta_{k_m}^2 = \lambda_{n-n_m+1} = \cdots = \lambda_{n-1} < \lambda_n, \tag{3.1}$$

where $\delta_{k_1}^2 > \delta_{k_2}^2 > \cdots > \delta_{k_m}^2$ *are* m *distinct values in* $\delta_1^2, \ldots, \delta_n^2$ *and* $n_1, \ldots, n_m$ $(\sum_{i=1}^m n_i = n)$ *are the multiplicities of* $\delta_{k_1}^2, \ldots, \delta_{k_m}^2$*, respectively.*

In Lemma 3.1.1, we have used a convention that the equation in (3.1) becomes empty if the last suffix of λ is less than the first suffix in the equation. For example, $1-\delta_{k_1}^2 = \lambda_1 = \cdots = \lambda_{n_1-1} < \lambda_{n_1}$ reduces to a simple inequality $1-\delta_{k_1}^2 < \lambda_1$ if $n_1 = 1$.

Proof. The characteristic equation of $R^{+}(\boldsymbol{\delta})$ is

$$\prod_{i=1}^{m}(1-\delta_{k_i}^2-\lambda)^{n_i-1}\det\left(\mathrm{diag}(1-\boldsymbol{\delta}_k^2-\lambda)+\mathrm{diag}(\boldsymbol{n})\boldsymbol{\delta}_k\boldsymbol{\delta}_k^{\mathsf{T}}\right)=0, \tag{3.2}$$

where $\boldsymbol{\delta}_k = (\delta_{k_1}, \ldots, \delta_{k_m})$ and $\boldsymbol{n} = (n_1, \ldots, n_m)$. As is easily seen, the trivial solutions are $\lambda = 1 - \delta_{k_i}^2$ for which $n_i > 1$. Then it is enough to find other m solutions, since each of the eigenvalues we found has the multiplicity $n_i - 1$. We seek for other solutions, provided that they are different from any of $1 - \delta_{k_i}^2,\ i = 1, \ldots, m$. Then the equation (3.2) is equivalent to

$$\sum_{i=1}^{m} (n_i \delta_{k_i}^2)/(1 - \delta_{k_i}^2 - \lambda) = -1.$$

We have used here the formula $\det(A \pm \boldsymbol{b}\boldsymbol{b}^\mathsf{T}) = \det(A)(1 \pm \boldsymbol{b}^\mathsf{T} A^{-1} \boldsymbol{b})$ for the non-singular matrix $A = \mathrm{diag}(1 - \boldsymbol{\delta}_k^2 - \lambda)$. Note that the function $f(\lambda) = \sum_{i=1}^{m} (n_i \delta_{k_i}^2)/(1 - \delta_{k_i}^2 - \lambda)$ is a strictly monotone increasing function of λ on each interval $(1 - \delta_{k_i}^2, 1 - \delta_{k_{i+1}}^2)$, and diverges to negative or positive infinity on the boundaries of each intervals for $i = 1, \ldots, m-1$. Then we found a solution on each interval. Furthermore, we can find one more solution λ_n on the interval $(1 - \delta_{k_m}^2, \infty)$, since $f(\lambda)$ is also a monotone increasing function of λ on this interval and diverges to negative infinity on the left boundary and zero for large enough λ. We have now found the remaining m solutions such that $1 - \delta_{k_1}^2 < \lambda_{n_1} < 1 - \delta_{k_2}^2 < \lambda_{n_1+n_2} < \cdots < \lambda_{n-n_m} < 1 - \delta_{k_m}^2 < \lambda_n$. $\square$

We note that (3.1) in Lemma 3.1.1 is not a direct consequence of the well known inequality for eigenvalues of A and B such that $A \le B$ nor of a more sophisticated inequality in the framework of majorization (for example, Theorem 16.F.1 and Theorem 9.G.1.c in Marshall and Olkin, 1979). In fact, (3.1) is much stronger because Lemma 3.1.1 is specialized for the matrix like $R^+(\boldsymbol{\delta}) = \mathrm{diag}(1 - \boldsymbol{\delta}^2) + \boldsymbol{\delta}\boldsymbol{\delta}^\mathsf{T}$.

We have now the following theorem.

Theorem 3.1.1. *Assume that $|\delta_n| \le \cdots \le |\delta_2| \le |\delta_1|$. $R^+(\boldsymbol{\delta})$ is a correlation matrix if and only if*

$$|\delta_1| \le 1 \qquad \textit{or} \qquad |\delta_2| < 1 < |\delta_1| \quad \textit{and} \quad \sum_{i=1}^{n} \delta_i^2/(1 - \delta_i^2) \le -1.$$

Furthermore, it is proper if and only if

$$1 \ne |\delta_2| \le |\delta_1| \le 1 \qquad \textit{or} \qquad |\delta_2| < 1 < |\delta_1| \quad \textit{and} \quad \sum_{i=1}^{n} \delta_i^2/(1 - \delta_i^2) < -1.$$

Proof. From Lemma 3.1.1, non-negativeness of the smallest eigenvalue λ_1 is equivalent to $1 - \delta_1^2 \ge 0$ if $\delta_1 = \delta_2$ and $\lambda_1 \ge 0$ otherwise. For the latter case, since

$|\delta_2| < 1 < |\delta_1|$ and $1 - \delta_{k_1}^2 < \lambda_1 < 1 - \delta_{k_2}^2$ from Lemma 3.1.1, we see that $\lambda_1 \geq 0$ is equivalent to

$$f(0) = \sum_{j=1}^{m} (n_j \delta_{k_j}^2)/(1 - \delta_{k_j}^2) = \sum_{i=1}^{n} \delta_i^2/(1 - \delta_i^2) \leq -1.$$

By similar discussion, we have the necessary and sufficient condition for the positiveness of λ_1. The condition $|\delta_i \delta_j| \leq 1$ $(i \neq j = 1, \dots, n)$ is always satisfied when either $|\delta_1| \leq 1$, or $|\delta_2| < 1 < |\delta_1|$ and $\sum_{i=1}^{n} \delta_i^2/(1 - \delta_i^2) \leq -1$ holds true. For the latter case, it is enough to note that the following inequality holds true:

$$0 \geq 1 + \sum_{i=1}^{n} \delta_i^2/(1 - \delta_i^2) = (1 - \delta_1^2 \delta_2^2)/\{(1 - \delta_1^2)(1 - \delta_2^2)\} + \sum_{i=3}^{n} \delta_i^2/(1 - \delta_i^2). \quad \square$$

It is interesting to note that all $|\delta_i|$'s are not necessarily less or equal to 1. The largest δ_1 can be greater than 1 in absolute value. However, $|\delta_1|$ can not be far away from 1 to satisfy the side condition $\sum_{i=1}^{n} \delta_i^2/(1 - \delta_i^2) < -1$ unless all other parameters are very small in absolute value.

3.1.2 Negative multiplicative correlation

For negative multiplicative matrix, we need the following lemma for the proof of Theorem 3.1.2. Lemma 3.1.2 looks similar to Lemma 3.1.1 but not the same.

Lemma 3.1.2. *The following inequality holds true for the eigenvalues* $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ *of negative multiplicative matrix* $R^-(\boldsymbol{\delta})$.

$$\lambda_1 < 1 + \delta_{k_1}^2 = \lambda_2 = \cdots = \lambda_{n_1} < \lambda_{n_1+1} < 1 + \delta_{k_2}^2 = \lambda_{n_1+2} = \cdots = \lambda_{n_1+n_2} < \lambda_{n_1+n_2+1} < \cdots$$

$$\cdots < \lambda_{n-n_m+1} < 1 + \delta_{k_m}^2 = \lambda_{n-n_m+2} = \cdots = \lambda_n,$$

where $\delta_{k_1}^2 < \delta_{k_2}^2 < \cdots < \delta_{k_m}^2$ *are* m *distinct values in* $\delta_1^2, \dots, \delta_n^2$ *and* $n_1, \dots, n_m$ $(\sum_{i=1}^{m} n_i = n)$ *are multiplicities of* $\delta_{k_1}^2, \dots, \delta_{k_m}^2$, *respectively.*

Proof. Since the characteristic equation of $R^-(\boldsymbol{\delta})$ is

$$\prod_{i=1}^{m} (1 + \delta_{k_i}^2 - \lambda)^{n_i - 1} \det\left(\mathrm{diag}(1 + \boldsymbol{\delta}_k^2 - \lambda) + \mathrm{diag}(\boldsymbol{n}) \boldsymbol{\delta}_k \boldsymbol{\delta}_k^{\mathsf{T}}\right) = 0,$$

a similar discussion follows as in the proof of Lemma 3.1.1. To find m non-trivial solutions, it is enough to note that the equation above is equivalent to $\sum_{i=1}^{m} (n_i \delta_{k_i}^2)/(1 + \delta_{k_i}^2 - \lambda) = 1$ provided that λ is equal to none of $1 + \delta_{k_i}^2$'s. We then find a solution on each interval $(1 + \delta_{k_i}^2, 1 + \delta_{k_{i+1}}^2)$ for $i = 1, \dots, m-1$ by noting that $g(\lambda) = \sum_{i=1}^{m} (n_i \delta_{k_i}^2)/(1 + \delta_{k_i}^2 - \lambda)$ has the same properties as those of $f(\lambda)$ in Lemma

3.1.1. A remaining solution can be found on $(-\infty, 1+\delta_{k_1}^2)$ since $g(\lambda)$ is strictly monotone increasing, converges to zero as λ tends to negative infinity and $g(1+\delta_{k_1}^2) = 0$. □

Theorem 3.1.2. *$R^-(\boldsymbol{\delta})$ is a correlation matrix if and only if $\sum_{i=1}^n \delta_i^2/(1+\delta_i^2) \le 1$. It is proper if and only if the strict inequality holds true.*

Proof. From the proof of Lemma 3.1.2, non-negativeness of the minimum eigenvalue λ_1 is equivalent to

$$g(0) = \sum_{j=1}^m (n_j\delta_{k_j}^2)/(1+\delta_{k_j}^2) = \sum_{i=1}^n \delta_i^2/(1+\delta_i^2) \le 1.$$

And the condition $|\delta_1\delta_2| \le 1$ follows from

$$1 \ge \sum_{i=1}^n \delta_i^2/(1+\delta_i^2) \ge \sum_{i=1}^2 \delta_i^2/(1+\delta_i^2).$$ □

Compared with the condition in Theorem 3.1.1 for positive multiplicative correlation matrix, the condition in Theorem 3.1.2 for negative multiplicative correlation matrix looks simpler. However, it seems more restrictive, because even though there is no explicit restriction to any values of $|\delta_i|$'s like most of them should be less than 1, the total contribution through the formula $\delta_i^2/(1+\delta_i^2)$ should not exceed 1.

Example 3.1.1 (Equi-correlation). From Theorem 3.1.1 and Theorem 3.1.2, we can easily see that the choice of parameter c for the equi-correlation model Q_{ij} in the Introduction is limited to the range $-1/(n-1) \le c \le 1$.

3.1.3 Multiplicative covariance

Although multiplicative correlation structure can be identified through the correlation, there are cases where it would be simpler to discuss it through the corresponding covariance. A multiplicative covariance can be derived from multiplicative correlation matrix by giving the variances, $\mathrm{var}(X_i) = \sigma_i^2,\ i = 1,\dots,n$. We hereafter write such a covariance matrix as $\Sigma^+(\boldsymbol{a},\boldsymbol{b}) = \mathrm{diag}(\boldsymbol{b}) + \boldsymbol{a}\boldsymbol{a}^\top$ in case of $R^+(\boldsymbol{\delta})$, and as $\Sigma^-(\boldsymbol{a},\boldsymbol{b}) = \mathrm{diag}(\boldsymbol{b}) - \boldsymbol{a}\boldsymbol{a}^\top$ in case of $R^-(\boldsymbol{\delta})$. Here $\boldsymbol{a} = (a_1,\dots,a_n)$ is a vector of $a_i = \sigma_i\delta_i,\ i = 1,\dots,n$ and $\boldsymbol{b} = (b_1 \dots, b_n)$ is a vector of $b_i = \sigma_i^2(1-\delta_i^2)\ i = 1,\dots,n$ for $R^+(\boldsymbol{\delta})$ and $b_i = \sigma_i^2(1+\delta_i^2)$ for $R^-(\boldsymbol{\delta})$. Of course, the converse does not always hold true since we allow any of σ_i^2 to be 0.

Theorem 3.1.3. *Assume that $b_1 \le b_2 \le \cdots \le b_n$. The matrix $\Sigma^+(\boldsymbol{a}, \boldsymbol{b})$ is a covariance matrix if and only if*

$$0 \le b_1 \qquad \text{or} \qquad b_1 < 0 < b_2 \;\; \text{and} \;\; 1 + \sum_{i=1}^{n} a_i^2/b_i \le 0 \;\; .$$

It is proper covariance if and only if

$$0 \le b_1 \le b_2 \ne 0 \qquad \text{or} \qquad b_1 < 0 < b_2 \;\; \text{and} \;\; 1 + \sum_{i=1}^{n} a_i^2/b_i < 0.$$

The matrix $\Sigma^-(\boldsymbol{a}, \boldsymbol{b})$ is a covariance matrix if and only if $0 < b_1$ and $\sum_{i=1}^n a_i^2/b_i \le 1$. It is proper if the strict inequality holds true.

Theorem 3.1.3 is a direct consequence of Theorem 3.1.1 and Theorem 3.1.2. However, it is worthy of giving here the distribution of the eigenvalues, too. Provided that $\boldsymbol{a}$ has no zero elements, the following inequalities are rather trivial in view of Lemma 3.1.1 and Lemma 3.1.2.

For the eigenvalues of $\Sigma^+(\boldsymbol{a}, \boldsymbol{b})$

$$b_{k_1} = \lambda_1 = \cdots = \lambda_{n_1 - 1} < \lambda_{n_1} < b_{k_2} = \lambda_{n_1+1} = \cdots = \lambda_{n_1+n_2-1} < \lambda_{n_1+n_2} < \cdots$$

$$\cdots < \lambda_{n-n_m} < b_{k_m} = \lambda_{n-n_m+1} = \cdots = \lambda_{n-1} < \lambda_n,$$

and for the eigenvalues of $\Sigma^-(\boldsymbol{a}, \boldsymbol{b})$

$$\lambda_1 < b_{k_1} = \lambda_2 = \cdots = \lambda_{n_1} < \lambda_{n_1+1} < b_{k_2} = \lambda_{n_1+2} = \cdots = \lambda_{n_1+n_2} < \lambda_{n_1+n_2+1} < \cdots$$

$$\cdots < \lambda_{n-n_m+1} < b_{k_m} = \lambda_{n-n_m+2} = \cdots = \lambda_n.$$

Here $b_{k_1} < b_{k_2} < \cdots < b_{k_m}$ are m distinct elements of $\boldsymbol{b}$ and $n_1, \ldots, n_m$ $(\sum_{i=1}^m n_i = n)$ are the multiplicities of $b_{k_1}, \ldots, b_{k_m}$, as same as in Lemma 3.1.1 or Lemma 3.1.2. It is interesting to note that either of the inequalities depends only on the values of $\boldsymbol{b}$. This can be easily understood, if we note that the characteristic equation here is $\prod_{i=1}^m (b_{k_i} - \lambda)^{n_i - 1} \det\left\{\mathrm{diag}(\boldsymbol{b}_k - \lambda) \pm \boldsymbol{\gamma}\boldsymbol{\gamma}^\mathsf{T}\right\} = 0$, where $\boldsymbol{b}_k = (b_{k_1}, \ldots, b_{k_m})$ and $\boldsymbol{\gamma} = (\gamma_{k_1}, \ldots, \gamma_{k_m})$ with $\gamma_{k_i}^2 = \sum_{j:\ b_j = b_{k_i}} a_j^2$.

If there is a zero element a_i in $\boldsymbol{a}$, the inequalities above should be modified by noting that the corresponding b_i becomes an eigenvalue. Ronning (1982) proved a similar inequality in case of multinomial, Dirichlet or multivariate hypergeometric distributions. In case of multinomial distribution, Watson (1996) independently noted that such an inequality holds true for multinomial distribution. However, our result above is not restricted to such multiplicative covariance matrices but also for any multiplicative matrices in general.

Example 3.1.2 (Equi-covariance). From Theorem 3.1.3, we can easily see that the covariance r for the equi-covariance model, which any pairs of variables share, should satisfy

$$0 \le r \le \sigma_1^2,\ \ \sigma_1^2 < r < \sigma_2^2 \text{ and } \sum_{i=1}^{n} r/(\sigma_i^2 - r) \le -1,\ \text{ or } r < 0 \text{ and } \sum_{i=1}^{n} r/(\sigma_i^2 - r) \ge -1,$$

where σ_1^2 and σ_2^2 are the minimum and the second minimum of the variances of $X_i's$, respectively.

3.2 Implications of multiplicative correlations or covariances

In this section, we investigate implications of multiplicative correlations or covariances. In Section 3.2.1, we will give a factorization theorem. The meaning of the multiplicative correlation becomes clearer through the factorization of variables, at least for the case of positive multiplicative correlations. However, this factorization is not powerful enough for understanding negative multiplicative correlations. Another theorem given in Section 3.2.2 will explain the reason why such negative multiplicative correlation matrices arise so frequently, although they are not exhaustive.

3.2.1 Factorization

In view of the reduction method mentioned in the Introduction, we may expect that X_i's can be represented as a common variable plus individual variables if the correlation is multiplicative. For multivariate normal distribution, such a factorization is almost trivial, and by making use of the factorization, Curnow and Dunnett (1962) or Gupta (1963) showed that a simple calculation of the distribution is possible when the correlation is positive multiplicative. Six (1981) extended their results for the case of negative multiplicative correlation. The factorization as in the following Corollary 3.2.1 is known as *fundamental theorem of factor analysis*, but it is only for the case of positive multiplicative correlations (see, for example, Steiger, 1979, p.158). The following theorem gives us a general factorization theorem for positive and negative multiplicative correlations.

Theorem 3.2.1. *A random vector* $\boldsymbol{X} = (X_1, \ldots, X_n)$ *with zero means has a non-singular multiplicative covariance* $\mathrm{V}(\boldsymbol{X}) = \mathrm{diag}(\boldsymbol{b}) \pm \boldsymbol{a}\boldsymbol{a}^{\top}$ *with* $\boldsymbol{b} > \boldsymbol{0}$ *if and* *only if each element of* $\boldsymbol{X}$ *is written as*

$$X_i = \gamma a_i Z_0 + \sqrt{b_i} Z_i \qquad i = 1, \dots, n, \tag{3.3}$$

where $Z_0, Z_1, \dots, Z_n$ *are random variables with zero means and unit variances, in which* $Z_1, \dots, Z_n$ *are uncorrelated each other but correlated with* Z_0 *as*

$$\rho(Z_0, Z_i) = c a_i / \sqrt{b_i} \quad i = 1, \dots, n,$$

where $|c| \leq 1/\kappa$ *for* $\Sigma^{+}(\boldsymbol{a}, \boldsymbol{b})$ *and* $1 \leq |c| \leq 1/\kappa$ *for* $\Sigma^{-}(\boldsymbol{a}, \boldsymbol{b})$ *with* $\kappa^2 = \sum_{i=1}^{n}(a_i^2/b_i)$*. The constants* γ *and* c *satisfy the equation,*

$$\gamma^2 + 2\gamma c = \begin{cases} 1 & \textit{for} \quad \Sigma^{+}(\boldsymbol{a}, \boldsymbol{b}) \\ -1 & \textit{for} \quad \Sigma^{-}(\boldsymbol{a}, \boldsymbol{b}) \end{cases}.$$

Proof. If $\boldsymbol{X}$ is represented as in (3.3), then a direct calculation yields the multiplicative covariance $\mathrm{V}(\boldsymbol{X}) = \mathrm{diag}(\boldsymbol{b}) \pm \boldsymbol{a}\boldsymbol{a}^{\mathsf{T}}$. On the other hand, if the $\boldsymbol{X}$ has the desired covariance, define a random variable

$$Z_0 = \left\{(1 - c^2\kappa^2)/(1 + \sigma\kappa^2)\right\}^{1/2} X_0 \pm (c^2 + \sigma)^{1/2} \boldsymbol{a}^{\mathsf{T}} \Sigma^{-1} \boldsymbol{X}$$

by introducing a new random variable X_0 with mean zero and unit variance but independent of any X_i's. Here the σ is 1 for $\Sigma^{+}(\boldsymbol{a}, \boldsymbol{b})$ or is -1 for $\Sigma^{-}(\boldsymbol{a}, \boldsymbol{b})$, and the sign $\pm$ shows alternative definitions. It can be easily seen that the Z_0 and $Z_i = [X_i - \{\pm(c^2 + \sigma)^{1/2} - c\} a_i Z_0]/\sqrt{b_i}$ satisfy the desired properties from the non-negative definiteness of the covariance of $\boldsymbol{Z}$ and Theorem 3.1.3. □

If $\gamma = 1$, then $c = 0$ for $\Sigma^{+}(\boldsymbol{a}, \boldsymbol{b})$, and $c = -1$ for $\Sigma^{-}(\boldsymbol{a}, \boldsymbol{b})$, so that we have the following corollary.

Corollary 3.2.1. *A random vector* $\boldsymbol{X} = (X_1, \dots, X_n)$ *with zero means has a non-singular multiplicative covariance* $\mathrm{V}(\boldsymbol{X}) = \mathrm{diag}(\boldsymbol{b}) \pm \boldsymbol{a}\boldsymbol{a}^{\mathsf{T}}$ *with* $\boldsymbol{b} > \boldsymbol{0}$ *if and only if each element of* $\boldsymbol{X}$ *is written as*

$$X_i = a_i Z_0 + \sqrt{b_i} Z_i, \quad i = 1, \dots, n,$$

where $Z_0, Z_1, \dots, Z_n$ *are random variables with zero means and unit variances, in which* $Z_1, \dots, Z_n$ *are uncorrelated each other but it can be correlated with* Z_0 *as*

$$\rho(Z_0, Z_i) = \begin{cases} 0 & \textit{for} \quad \Sigma^{+}(\boldsymbol{a}, \boldsymbol{b}) \\ -a_i/\sqrt{b_i} & \textit{for} \quad \Sigma^{-}(\boldsymbol{a}, \boldsymbol{b}) \end{cases} \qquad i = 1, \dots, n.$$

The factorization of $\Sigma^+(\boldsymbol{a}, \boldsymbol{b})$ in Corollary 3.2.1 is well-known (Steiger, 1979) but Theorem 3.2.1 is much more general. It is interesting to note that no uncorrelated factorization is possible in case of negative multiplicative covariances, but the common variable is orthogonal to the original variable X_is. The following corollary is a direct consequence of Corollary 3.2.1 for multiplicative correlations.

Corollary 3.2.2. *A random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$ with zero means and variances $\sigma_i^2 = \mathrm{var}(X_i), i = 1, \ldots, n$ has a proper multiplicative correlation matrix $R^{\pm}(\boldsymbol{\delta})$ if and only if each element of $\boldsymbol{X}$ is written as*

$$X_i/\sigma_i = \begin{cases} \delta_i Z_0 + (1-\delta_i^2)^{1/2} Z_i & \text{for} \quad R^+(\boldsymbol{\delta}) \\ \delta_i Z_0 + (1+\delta_i^2)^{1/2} Z_i & \text{for} \quad R^-(\boldsymbol{\delta}) \end{cases}, \qquad i = 1, \ldots, n,$$

where $Z_0, Z_1, \ldots, Z_n$ are random variables with zero means and unit variances, in which $Z_1, \ldots, Z_n$ are uncorrelated each other but it can be correlated with Z_0 as

$$\rho(Z_0, Z_i) = \begin{cases} 0 & \text{for} \ R^+(\boldsymbol{\delta}) \\ -\delta_i/(1+\delta_i^2)^{1/2} & \text{for} \ R^-(\boldsymbol{\delta}) \end{cases}, \qquad i = 1, \ldots, n.$$

Recently Kelderman (2004) shows that positive multiplicative covariance $\mathrm{V}(\boldsymbol{X}) = \mathrm{diag}(\boldsymbol{b}) + \boldsymbol{a}\boldsymbol{a}^\top$ with $\boldsymbol{a} > \boldsymbol{0}$ and $\boldsymbol{b} > \boldsymbol{0}$ is equivalent to the fact that the conditional density $f(\boldsymbol{y}_M | \boldsymbol{Y}_{M^c} = \boldsymbol{y}_{M^c})$ for any partition $\boldsymbol{Y} = (\boldsymbol{Y}_M, \boldsymbol{Y}_{M^c})$ is invariant with respect to permutation of the values $\boldsymbol{y}_{M^c}$ after a scale and location transform $\boldsymbol{Y} = \boldsymbol{\alpha} + \mathrm{diag}(\boldsymbol{\beta})\boldsymbol{X}$ has been applied. However, this equivalence is proved only for the case of multivariate normal distribution. This is considered to be a characterization of positive multiplicative correlation, but its applicativity to other distributions is unknown.

3.2.2 A characterization of negative multiplicative covariance

The following theorem characterizes an interesting class of negative covariance matrices.

Theorem 3.2.2. *Assume that an n-dimensional random vector $\boldsymbol{X}$ has a negative multiplicative covariance $\Sigma^-(\boldsymbol{a}, \boldsymbol{b})$. Then, $\boldsymbol{b} = (\sum_{i=1}^n a_i)\boldsymbol{a}$ if and only if $\sum_{i=1}^n X_i$ is almost surely constant.*

Proof. The fact that $\sum_{i=1}^n X_i$ is almost surely constant is equivalent to

$$\mathrm{var}\left(\sum_i X_i\right) = \boldsymbol{1}^\top \left\{\mathrm{diag}(\boldsymbol{b}) - \boldsymbol{a}\boldsymbol{a}^\top\right\} \boldsymbol{1} = \sum_i b_i - \left(\sum_i a_i\right)^2 = 0.$$

Noting the Schwarz's inequality, we have

$$\sum_i b_i = \left(\sum_i a_i\right)^2 = \left(\sum_i \frac{a_i}{\sqrt{b_i}}\sqrt{b_i}\right)^2 \le \left(\sum_i b_i\right)\left(\sum_i \frac{a_i^2}{b_i}\right).$$

Then, it is clear that b_i is proportional to a_i since $\sum_{i=1}^n a_i^2/b_i \le 1$. □

This theorem says that the negative multiplicative covariance matrix takes the form of

$$\mathrm{V}(\boldsymbol{X}) = \left(\sum_{i=1}^n a_i\right)\mathrm{diag}(\boldsymbol{a}) - \boldsymbol{a}\boldsymbol{a}^\top,$$

as far as there is a sum constraint $\sum_{i=1}^n X_i = \text{const a.s.}$ An example of family of distributions which have such a negative covariance matrix is the following Multivariate Pólya-Eggenberger distribution.

Example 3.2.1. The joint probability function of *multivariate Pólya-Eggenberger distribution* is given by

$$p(x_1,\dots,x_n) = \binom{t}{x_1,\dots,x_n}\left\{\prod_{i=1}^n \alpha_i^{[x_i,c]}\right\}\Big/\,\alpha^{[t,c]},$$

where x_i and $\alpha_i, i = 1,\dots,n$ are non-negative integers, c is an integer, $t = \sum_{i=1}^n x_i$, $\alpha = \sum_{i=1}^n \alpha_i$, and $\alpha^{[x,c]} = \alpha(\alpha+c)\cdots\{\alpha+(x-1)c\}$ with $\alpha^{[0,c]} = 1$ (see Johnson *el al.*, 1997, p. 201). Multivariate Pólya-Eggenberger distribution is a wide class of multivariate discrete distributions, which becomes a family of multinomial distributions, multivariate hypergeometric distributions, or multivariate negative hypergeometric distributions by respectively taking $c = 0,\ -1$, or 1. It has a negative multiplicative covariance since

$$\mathrm{E}(\boldsymbol{X}) = t\frac{\boldsymbol{\alpha}}{\alpha} \quad \text{and} \quad \mathrm{V}(\boldsymbol{X}) = \frac{t(\alpha+tc)}{\alpha^2(\alpha+c)}\left\{\alpha\mathrm{diag}(\boldsymbol{\alpha}) - \boldsymbol{\alpha}\boldsymbol{\alpha}^\top\right\},$$

with $\boldsymbol{\alpha} = (\alpha_1,\dots,\alpha_n)$. Since Dirichlet distribution is a limit of multivariate Pólya-Eggenberger distribution when t tends to infinity in such a way that $y_i = \lim_{t\to\infty} x_i/t$, $i = 1,\dots,n$ for the fixed $\nu_i = \alpha_i/c$, $i = 1,\dots,n$, it has the negative multiplicative covariance as in Theorem 3.2.2. In fact, the density function is

$$f(\boldsymbol{y}) = \left(\Gamma(\nu)/\prod_{i=1}^n\Gamma(\nu_i)\right)\prod_{i=1}^n y_i^{\nu_i-1} \text{ where } \sum_{i=1}^n y_i = 1 \text{ and } \sum_{i=1}^n \nu_i = \nu,$$

and the covariance matrix is

$$\mathrm{V}(\boldsymbol{Y}) = \left(\nu\mathrm{diag}(\boldsymbol{\nu}) - \boldsymbol{\nu}\boldsymbol{\nu}^\top\right)/\left\{\nu^2(\nu+1)\right\}.$$

3.3 Invariance of multiplicative covariance

We can understand from Theorem 3.2.2 that the conditional and unconditional covariances are both negative multiplicative as far as the given condition takes a form of $\sum_{i=1}^{n} X_i = const$. We will investigate, in this section, if such a multiplicative property is preserved or not by unconditioning. The following theorem gives us a condition for that.

Theorem 3.3.1. *Let* $(\boldsymbol{X}, T)$ *be an* $(n+1)$*-dimensional random vector and assume that the conditional variance of* $\boldsymbol{X}$ *given* T *is*

$$\mathrm{V}(\boldsymbol{X}|\ T=t) = \sigma(t)\left(\mathrm{diag}(\boldsymbol{b}) \pm \boldsymbol{a}\boldsymbol{a}^{\top}\right)$$

for a $\sigma(t) > 0$. *If the conditional expectation is written as*

$$\mathrm{E}(\boldsymbol{X}|\ T=t) = \mu(t)\boldsymbol{a} + \boldsymbol{c}$$

for an n*-dimensional constant vector* $\boldsymbol{c}$*, then the unconditional covariance is again multiplicative,*

$$\mathrm{V}(\boldsymbol{X}) = \mathrm{E}\left(\sigma(T)\right)\mathrm{diag}(\boldsymbol{b}) + \left\{\mathrm{var}(\mu(T)) \pm \mathrm{E}\left(\sigma(T)\right)\right\}\boldsymbol{a}\boldsymbol{a}^{\top}.$$

Proof. It is almost clear since

$$\mathrm{E}(X_i|\ T=t) = a_i\mu(t) + c_i, \quad \mathrm{E}(X_i^2|\ T=t) = (b_i \pm a_i^2)\sigma(t) + \left(a_i\mu(t) + c_i\right)^2,$$

and

$$\mathrm{E}(X_iX_j|\ T=t) = \pm a_ia_j\sigma(t) + \left(a_i\mu(t) + c_i\right)\left(a_j\mu(t) + c_j\right). \qquad \square$$

It is worthy of noting that the unconditional covariance can be positive and negative multiplicative irrespective of positiveness or negativeness of the conditional covariance.

Example 3.3.1 (Homogeneous Distribution). It is known that $\boldsymbol{X}$ is distributed as a multivariate homogeneous distribution if and only if the conditional distribution of $\boldsymbol{X}$ given the sum $\sum_{i=1}^{n} X_i$, is multinomial (see, Johnson *el al.*, 1997, p. 20). Since the conditional expectation and covariance matrix are written as $\mathrm{E}(\boldsymbol{X}|T=t) = t\boldsymbol{p}$ and $\mathrm{V}(\boldsymbol{X}|T=t) = t\left(\mathrm{diag}(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}^{\top}\right)$, the conditions in Theorem 3.3.1 are satisfied by taking $\sigma(t) = \mu(t) = t$ and $\boldsymbol{c} = \boldsymbol{0}$. Therefore, we see that homogeneous distribution always has a multiplicative covariance such as

$$\mathrm{V}(\boldsymbol{X}) = \mathrm{E}(T)\mathrm{diag}(\boldsymbol{p}) + \left(\mathrm{var}(T) - \mathrm{E}(T)\right)\boldsymbol{p}\boldsymbol{p}^{\top}.$$

The sign of $\text{var}(T)-\text{E}(T)$ depends on the distribution of T. For example, it is always negative multiplicative if the distribution of T is binomial and positive multiplicative if it is negative binomial. Although it can be seen from the fact that the resulting distribution of $\boldsymbol{X}$ is multinomial or negative multinomial respectively, it can be shown by a direct calculation of $\text{var}(T) - \text{E}(T)$ as $-k\xi^2 < 0$ for $\text{Bn}(k, \xi)$ and $k(1 - \xi)^2/\xi^2 > 0$ for $\text{NgBn}(k, \xi)$. It is trivial but interesting to note that $\boldsymbol{X}$ is a vector of orthogonal variable if T is Poisson distributed because $\text{var}(T) = \text{E}(T)$. Let us back to the multivariate Pólya-Eggenberger distributions in Example 3.2.1. The t there is a parameter and can be replaced by a non-negative integer valued random variable T. Then it is clear from Theorem 3.3.1 that the covariance matrix of $\boldsymbol{X}$ is again multiplicative for any randomization T.

Example 3.3.2 (Random Scaling). It is clear from Theorem 3.3.1 that the randomly scaled $\boldsymbol{X} = T\boldsymbol{Y}$ has a multiplicative covariance, as far as

$$\text{E}(\boldsymbol{Y}) = k\boldsymbol{a} \quad \text{for a constant } k, \quad \text{V}(\boldsymbol{Y}) = \text{diag}(\boldsymbol{b}) \pm \boldsymbol{a}\boldsymbol{a}^{\mathsf{T}},$$

and T is independent of $\boldsymbol{Y}$. Several multivariate continuous distributions are derived by such a random scaling. For example, multivariate Liouville distribution or multivariate second kind beta (or inverted Dirichlet) distribution are derived from Dirichlet distribution by taking respectively Liouville distributed T or second kind beta distributed T(see, Kotz *et al.*, 2000, p.491, 530; Gupta and Richards, 2001). In view of Theorem 6, multiplicative covariance structures are preserved through such a derivation of multivariate distribution. It is a good contrast to the reduction method for multivariate discrete distributions.

Unfortunately the converse of Theorem 3.3.1 is not so simple. It heavily depends on the shape of distribution. We leave this problem for future investigation. We end up this section by giving another invariance property. It is about partial correlations or covariances. Although $\boldsymbol{X} = (X_1, \ldots, X_n)$ is partitioned as $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$ with m-dimensional vector $\boldsymbol{X}_1$ and $(n-m)$-dimensional vector $\boldsymbol{X}_2$ and its parameter vectors are also partitioned as $\boldsymbol{a} = (\boldsymbol{a}_1, \boldsymbol{a}_2)$ and $\boldsymbol{b} = (\boldsymbol{b}_1, \boldsymbol{b}_2)$ in the theorem, but it is only for the convenience. The result holds true for any partial covariances.

Theorem 3.3.2. *Assume that $\boldsymbol{X}$ has a multiplicative covariance $\Sigma^{\pm}(\boldsymbol{a}, \boldsymbol{b})$. If $\text{V}(\boldsymbol{X}_2)$ is non-singular and all elements of $\mathbf{b}_2$ are positive, then the partial covariance of $\boldsymbol{X}_1$ given $\boldsymbol{X}_2$ is also multiplicative and*

$$\text{diag}(\boldsymbol{b}_1) \pm \boldsymbol{a}_1\boldsymbol{a}_1^{\mathsf{T}}/\{1 \pm \boldsymbol{a}_2^{\mathsf{T}}\text{diag}(\boldsymbol{b}_2)^{-1}\boldsymbol{a}_2\}.$$

Proof. Partition $\Sigma^{\pm}(\boldsymbol{a}, \boldsymbol{b})$ into

$$\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} \mathrm{diag}(\boldsymbol{b}_1) \pm \boldsymbol{a}_1\boldsymbol{a}_1^\top & \pm\boldsymbol{a}_1\boldsymbol{a}_2^\top \\ \pm\boldsymbol{a}_2\boldsymbol{a}_1^\top & \mathrm{diag}(\boldsymbol{b}_2) \pm \boldsymbol{a}_2\boldsymbol{a}_2^\top \end{pmatrix}.$$

Then the partial covariance matrix of $\boldsymbol{X}_1$ given $\boldsymbol{X}_2$ is written as

$$\begin{aligned} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} &= \mathrm{diag}(\boldsymbol{b}_1) \pm \boldsymbol{a}_1\boldsymbol{a}_1^\top - \boldsymbol{a}_1\boldsymbol{a}_2^\top\left(\mathrm{diag}(\boldsymbol{b}_2) \pm \boldsymbol{a}_2\boldsymbol{a}_2^\top\right)^{-1}\boldsymbol{a}_2\boldsymbol{a}_1^\top \\ &= \mathrm{diag}(\boldsymbol{b}_1) \pm \boldsymbol{a}_1\boldsymbol{a}_1^\top/\{1 \pm \boldsymbol{a}_2^\top\mathrm{diag}(\boldsymbol{b}_2)^{-1}\boldsymbol{a}_2\}, \end{aligned}$$

since

$$(\mathrm{diag}(\boldsymbol{b}_2)+\boldsymbol{a}_2\boldsymbol{a}_2^\top)^{-1} = \mathrm{diag}(\boldsymbol{b}_2)^{-1}-\mathrm{diag}(\boldsymbol{b}_2)^{-1}\boldsymbol{a}_2\boldsymbol{a}_2^\top\mathrm{diag}(\boldsymbol{b}_2)^{-1}/(1+\boldsymbol{a}_2^\top\mathrm{diag}(\boldsymbol{b}_2)^{-1}\boldsymbol{a}_2).\Box$$

For multiplicative correlations, the following corollary holds true, which is a direct consequence of Theorem 3.3.2.

Corollary 3.3.1. *If $\boldsymbol{X}$ has a multiplicative correlation $R^{\pm}(\boldsymbol{\delta})$, then the partial correlation of $\boldsymbol{X}_1$ given $\boldsymbol{X}_2$ is also multiplicative and $R^{\pm}(\tilde{\boldsymbol{\delta}})$ with*

$$\tilde{\delta}_i = \frac{\delta_i}{\{1 \pm c(1 \mp \delta_i^2)\}^{1/2}}, \quad i = 1, \ldots, m,$$

where $c = \sum_{j=m+1}^{n} \delta_j^2/(1 \mp \delta_j^2)$.

We see that the partial covariance is proportional to the original covariance but the partial correlation coefficient is not so, although the multiplicative property is preserved. An important implication of Theorem 3.3.2 or Corollary 3.3.1 is that it is enough to check if the zero correlation coefficient for the check of zero partial correlation coefficient. This is due to the multiplicative parameterization of the correlation or covariances, and it is not always true without such a parameterization. A simplest example is for the case of $n = 3$. The partial covariance between X_1 and X_2 is $\sigma_{12} - \sigma_{13}\sigma_{23}/\sigma_{33}$, which is not necessarily zero even if the original covariance $\sigma_{12} = 0$. However, σ_{13} or σ_{23} becomes zero if $\sigma_{12} = 0$ under the multiplicative parameterization of the covariance, so that the zero covariance implies the zero partial covariance.

3.4 Family of distributions which have multiplicative correlation

In this section, we show that various known multivariate distributions have multiplicative correlation or covariances. It is explained by one of the reasons in the previous sections.

In Johnson *et al.* (1997), eight families of discrete distributions have been introduced as in Table 3.1. This table shows if each family of distributions has multiplicative correlation or not with the reason. As has been discussed before, multinomial, negative multinomial, Poisson, hypergeometric and Pólya-Eggenberger distribution have multiplicative correlations. Although power series distributions or multivariate distributions of order s have no multiplicative correlation as a whole family of distributions, subfamilies like logarithmic series distributions (Johnson *et al.*, 1997, p.157), multivariate negative multinomial of order s (p.255) or multivariate logarithmic series distributions of order s (p.260) have multiplicative correlations. However, at this stage, we do not know the exact reason why such subfamilies have multiplicative correlations. Apparently, Ewens distributions have no multiplicative correlations.

Table 3.1: Discrete Multivariate Distributions in Johnson *et al.* (1997).

Family	Subfamily	Positive or Negative	Reason
35 Multinomial		Negative	Example 3.2.1
36 Negative multinomial		Positive	Example 3.3.1
37 Poisson		Positive	Reduction method
38 Power series	Logarithmic series	Positive	?
39 Hypergeometric		Negative	Example 3.2.1
40 Pólya-Eggenberger		Negative	Example 3.2.1
41 Ewens	—	—	—
42 Distributions of	Negative binomial of order s	Positive	?
order s	Logarithmic distr. of order s	Negative	?

In terms of continuous distributions, eight families of continuous distributions have been introduced in Kotz *et al.* (2000). Table 3.2 shows if each family has multiplicative correlation or not in the same manner as in Table 3.1. For multivariate normal distributions, that is, no explicit restriction to the correlations, we may define a subfamily of the distribution so that the correlation is multiplicative. We call it as multiplicatively correlated normal. Although multivariate exponential, multivariate gamma, multivariate logistic or multivariate Pareto distributions have no multiplicative correlations as a whole family, subfamilies like Moran and Downton's multivariate exponential distributions (Kotz *et al.*, 2000, p.400), Cheriyan and Ramabhadran's multivariate gamma distributions (p.454), Gumbel-Malik-Abraham's(p.552) and Farlie-Gumbel-Morgenstern's (p.561) multivariate lo-

Table 3.2: Continuous Multivariate Distributions in Kotz *et al.* (2000).

Family	Subfamily	Positive or Negative	Reason
45 Normal	Multiplicatively correlated normal	Both	
47 Exponential	Moran and Downton's	Positive	Equi-correlation
48 Gamma	Cheriyan and Ramabhadran's	Positive	Reduction Method
49 Dirichlet		Positive	Example 3.2.1
49 Inverted Dirichlet		Negative	Example 3.3.2
50 Liouvill		Both	Example 3.3.2
51 Logistic	Gumbel-Malik-Abraham	Positive	Equi-correlation
	Farlie-Gumbel-Morgenstern	Negative	Equi-correlation
52 Pareto	The first kind	Positive	Equi-correlation
53 Extreme value	—	—	—

gistic distributions, or multivariate Pareto distributions of the first kind (p.599) have equi-covariance or equi-correlation which is multiplicative. It has already shown in Example 3.2.1 or Example 3.3.2 that Dirichlet, inverted Dirichlet or multivariate Liouville distributions have multiplicative correlations. Apparently, the correlations of multivariate extreme value distributions are not multiplicative.

In this chapter, we could clarify several reasons why such a multiplicative correlation or covariance appears so frequently. One of reasons is that equi-covariance is a special case of multiplicative correlation or covariance. This typically arises in an introduction of new family of multivariate distributions by reduction method. Its generalization is so called "common variable plus individual variables model". The equi-correlation is another source of multiplicative correlation. A different reasons are "sum constraint" or "random scaling or mixing". However, we could discuss such different sources and reasons simultaneously from the view point of the multiplicative property of the correlation or the covariance.

A practical importance of multiplicative correlation or covariances is in its simplicity as a statistical model. It models only the covariance matrix but has several nice properties, so that it would be a quite powerful vehicle for analyzing complicated large scale phenomena. Inference of the parameter $\boldsymbol{\delta}$ or $\boldsymbol{a}$ and $\boldsymbol{b}$ is an open problem and left for future investigation.

Chapter 4

Multivariate Distribution Generated from Natural Exponential Family

In this chapter, we discuss two classes of multivariate distribution which is generated from independent samples from the natural exponential family (NEF). Both classes have two common features: satisfying the Condition C, that is, coincidence with partial and conditional correlations (see Chapter 2), and having multiplicative correlations (see Chapter 3).

One is the class of conditional distribution of independent NEF samples given the sum. Another is the class of distribution of independent NEF with quadratic variance function (NEF-QVF) samples when the parameter is randomized by the conjugate prior distributions. Representative distributions of the former are multinomial and Dirichlet distributions, and ones of the latter are negative multinomial and multivariate beta type two distributions. We treat the former in Section 4.1, and the latter in Section 4.2.

At first, we introduce NEF. When a random variable Z has a pdf or a pmf

$$p(z;\theta) = a(z)\ \exp(\theta z - \psi(\theta)), \tag{4.1}$$

it is said that Z is distributed as a univariate *natural exponential family (NEF)* with the *cumulant function* $\psi(\theta)$. (see, for example, Letac and Mora, 1990; Jørgensen, 1997; and Kotz *et al.*, 2000, Chapter 54). The θ is called the *natural parameter* and the natural parameter space Θ is the largest subset of $\mathbb{R}$ for which $p(z;\theta)$ is well defined. The mean and variance are

$$\mu := \mathrm{E}(Z) = \psi'(\theta) \quad \text{and} \quad V(\mu) := \mathrm{var}(Z) = \psi''(\theta),$$

and μ and $V(\mu)$ are called *mean parameter* and *variance function*, respectively. Since a distribution of NEF is specified by μ and $V(\mu)$, we think of (4.1) as a distribution on μ, and the distribution with the density

$$p(z;\mu) = a(z)\ \exp(\theta(\mu)z - \psi(\theta(\mu)))$$

where $\theta(\mu)$ is the inverse function of $\mu = \psi'(\theta)$.

Now, if Z has the density

$$p(z;\mu,\nu) = b(z;\nu)\exp(\theta(\mu)z - \nu\psi(\theta(\mu))), \tag{4.2}$$

we denote as $Z \sim \mathsf{NEF}(\nu\mu, \nu V(\mu))$. This is only a notational generalization of the above model when $\nu \neq 1$.

Especially, when the variance function is quadratic in μ as $V(\mu) = v_0 + v_1\mu + v_2\mu^2$, we say that Z is distributed *natural exponential family with quadratic variance function* (NEF-QVF, or Morris class). If Z has the density (4.2) and $V(\mu)$ is quadratic, we denote as $Z \sim \mathsf{NEF\text{-}QVF}\ (\nu\mu, \nu V(\mu))$. It is known that only six distributions are members of univariate NEF-QVF (see, Morris, 1982). They are summarized in Table 4.1.

Table 4.1: Members of NEF-QVF.

distribution	$\mathsf{NEF}(\nu\mu, \nu V(\mu))$	θ	$\psi(\theta)$	μ	$V(\mu)$	$b(z;\nu)$
Normal	$\mathrm{N}(\nu\mu, \nu)$	μ	$\theta^2/2$	μ	1	$\frac{1}{\sqrt{2\pi\nu}}e^{-\frac{z^2}{2\nu}}$
Poisson	$\mathrm{Po}(\nu\lambda)$	$\log\lambda$	e^θ	λ	μ	$\frac{\nu^z}{z!}$
Binomial	$\mathrm{Bn}(\nu, p)$	$\log(p/1-p)$	$\log(1+e^\theta)$	p	$\mu(1-\mu)$	$\binom{\nu}{z}, \nu \in \mathcal{N}$
Negative binomial	$\mathrm{NgBn}(\nu, p)$	$\log(1-p)$	$-\log(1-e^\theta)$	$(1-p)/p$	$\mu(1+\mu)$	$\binom{z+\nu-1}{z}$
Gamma	$\mathrm{Ga}(\nu, 1/a)$	$-a$	$-\log(-\theta)$	$1/a$	μ^2	$\frac{z^{\nu-1}}{\Gamma(\nu)}$
Hyperbolic secant	$\mathrm{NEF\text{-}GHS}(\nu, \mu)$	$\tan^{-1}\mu$	$-\log\cos\theta$	μ	$1+\mu^2$	$*$

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

4.1 Conditional distribution of independent NEF samples given the sum

Let $\boldsymbol{Z} = (Z_1, \dots, Z_n)$ be independent and $Z_j \sim \mathsf{NEF}(\nu_j\mu, \nu_j V(\mu))$. $T = \sum_{j=1}^n Z_j$ is a sufficient statistic, and $T \sim \mathsf{NEF}(\nu\mu, \nu V(\mu))$, $\nu = \sum_{j=1}^n \nu_j$. The conditional density of $\boldsymbol{Z}$ given $T = t$ is

$$\prod_{j=1}^n b(z_j; \nu_j)/b(t; \nu). \tag{4.3}$$

The following theorem shows that $(\boldsymbol{Z}|T)$ has a multiplicative correlation.

Theorem 4.1.1 (Multiplicative correlation). *If $\boldsymbol{Z} = (Z_1, \dots, Z_n)$ are independent and $Z_j \sim \mathsf{NEF}(\nu_j\mu, \nu_j V(\mu))$, then the conditional variance-covariance matrix of $\boldsymbol{Z}$ given $T = \sum_{j=1}^n Z_j$ is multiplicative.*

Proof. Since $\mathsf{NEF}(\nu_j\mu, \nu_j V(\mu)) \in \mathcal{F}$ where $\mathcal{F}$ was defined in Section 2.3 and θ_i is replaced with ν_i, then Lemma 2.2.1 holds true. From (2.1) in the proof of Lemma 2.2.1, we yield the conclusion. □

Next, partition $\boldsymbol{Z}$ as $\boldsymbol{Z} = (\boldsymbol{X}, \boldsymbol{Y}) = (X_1, \dots, X_p, Y_1, \dots, Y_q)$ where $p + q = n$. The conditional density of $\boldsymbol{X}$ given $T = t$ and $\boldsymbol{Y} = (y_1, \dots, y_q)$ is

$$\prod_{j=1}^p b(z_j; \nu_j)/b(t - y; \nu - \nu_y) \qquad \left(y = \sum_{j=1}^q y_j, \quad \nu_y = \sum_{j=p+1}^{p+q} \nu_j \right).$$

The following theorem is a direct consequence of Theorem 2.2.1.

Theorem 4.1.2 (Condition C). *Suppose that $\boldsymbol{Z} = (Z_1, \dots, Z_n)$ are random variables in Theorem 4.1.1. If the distribution of $(\boldsymbol{X}, \boldsymbol{Y}) = (X_1, \dots, X_p, Y_1, \dots, Y_q)$ $(p + q = n)$ is equal to the distribution of $\boldsymbol{Z}$, then $(\boldsymbol{X}, \boldsymbol{Y})$ satisfies the Condition C provided that $T = \sum_{i=1}^p X_i + \sum_{i=1}^q Y_i$ is given.*

So far the variance function is not restricted, and η' in (2.1) can not be expressed explicitly. Now, we assume that the variance function is quadratic as $V(\mu) = v_0 + v_1\mu + v_2\mu^2$, the conditional variance-covariance matrix is explicitly obtained. That is, if $\boldsymbol{Z} = (Z_1, \dots, Z_n)$ are independent and $Z_j \sim \mathsf{NEF\text{-}QVF}(\nu_j\mu, \nu_j V(\mu))$, the conditional variance-covariance matrix of $\boldsymbol{Z}$ given $\sum_{j=1}^n Z_j = t$ is

$$\mathrm{V}(\boldsymbol{Z}|t) = c(t, \nu)(\mathrm{diag}(\boldsymbol{\xi}) - \boldsymbol{\xi}\boldsymbol{\xi}^\mathsf{T}), \quad \nu = \sum_{j=1}^n \nu_j; \ \xi_j = \nu_j/\nu, \ j = 1, \dots, n; \tag{4.4}$$

$$c(t, \nu) = \frac{\nu^2}{\nu + v_2} V\left(\frac{t}{\nu}\right).$$

The result is an extension of Morris (1983; Section 4).

Six conditional distributions of independent NEF-QVF samples given the sum are familiar, and listed in Table 4.2. The concrete expression of the multiplier $c(t,\nu)$ in (4.4) is shown in the last column.

Table 4.2: Conditional distribution of independent NEF-QVF samples given sum.

Z_i	$(\boldsymbol{Z}\mid T=t)$ conditional	$c(t,\nu)$
$\mathrm{N}(\nu_i\mu,\nu)$	$\mathrm{N}\big(t\boldsymbol{\xi},\nu(\mathrm{diag}(\boldsymbol{\xi})-\boldsymbol{\xi}\boldsymbol{\xi}^\top)\big)^*$	ν
$\mathrm{Po}(\nu_i\lambda)$	$\mathrm{Mn}(t,\boldsymbol{\xi})$	t
$\mathrm{Bn}(\nu_i,p)$	$\mathrm{MvHg}(t,\boldsymbol{\nu})$	$\frac{t(\nu-t)}{\nu-1}$
$\mathrm{NgBn}(\nu_i,p)$	$\mathrm{MvNgHg}(t,\boldsymbol{\nu})$	$\frac{t(\nu+t)}{\nu+1}$
$\mathrm{Ga}(\nu_i,1/a)$	$\boldsymbol{x}/t\mid t\sim\mathrm{Dir}(\boldsymbol{\nu})$	$\frac{t^2}{\nu+1}$
$\text{NEF-GHS}(\nu_i,\mu)$	Morris(1983)	$\frac{\nu(1+t^2)}{\nu+1}$

* $\mathrm{diag}(1/\boldsymbol{\nu})$ is a generalized inverse of $\nu(\mathrm{diag}(\boldsymbol{\xi})-\boldsymbol{\xi}\boldsymbol{\xi}^\top)$

4.2 Distribution of independent NEF-QVF samples when the parameter is randomized

In this section, we discuss the distributions generated from independent NEF-QVF samples when the mean parameter is randomized. We discuss the cases where the parameter is randomized by the conjugate prior in Section 4.2.1 and by other priors in Section 4.2.2. For the former case, we will show that the distributions generated from independent NEF-QVF samples have multiplicative correlations, and that they satisfy the Condition C, that is, their partial correlations are equal to conditional correlations. On the other hand, we will give an example in which the distribution has multiplicative correlation but do not satisfy the Condition C for the latter case.

At first, we introduce the conjugate prior distributions of NEF and the mixture of distributions.

Conjugate prior of NEF

For the $\mathrm{NEF}(\nu\mu,\nu V(\mu))$ with the density

$$p(z;\mu,\nu)=b(z;\nu)\ \exp(\theta(\mu)z-\nu\psi(\theta(\mu))), \tag{4.5}$$

the conjugate prior distribution on θ mimics (4.5), being

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

where K is the normalization constant. Here $\pi^*(\theta; \eta, \zeta)$ is a two parameter family on densities for θ having NEF with natural parameter η and convolution parameter ζ. We think of π^* as a distribution on $\mu = \psi'(\theta)$, and not on θ. This usually is a non-linear transformation of the NEF for θ and therefore is an exponential family that is not a NEF except in the case of the normal distribution. The density of μ is

$$\pi(\mu; \eta, \zeta) = K(\eta, \zeta)\, \exp(\eta\, \theta(\mu) - \zeta\, \psi(\theta(\mu)))\, V^{-1}(\mu), \quad V(\mu) = \psi''(\theta(\mu)), \qquad (4.6)$$

and is called the *conjugate prior distribution* of $\mathsf{NEF}(\mu, V(\mu))$. We describe it as $\mathsf{CP}(\eta, \zeta)$. For a member of NEF, its conjugate prior has appropriate properties (See, Morris, 1983; Consonni and Veronese, 1992).

Mixture of distributions

Let $\boldsymbol{Z} = (Z_1, \ldots, Z_n)$ be independent variables and Z_i has the distribution function $F_i(z_i; \mu)$ $(i = 1, \ldots, n)$. Suppose that the real parameter μ is a random variable with the distribution function $G(\mu)$ which is independent of $\boldsymbol{Z}$, and $\boldsymbol{Z}$ has the joint distribution function

$$F(z_1, \ldots, z_n) = \int F_1(z_1; \mu) \cdots F_n(z_n; \mu)\, dG(\mu). \qquad (4.7)$$

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

$$\prod_{i=1}^{n} F_i(z_i;\ \mu) \bigwedge_{\mu} G(\mu).$$

In the following subsection, we consider the case where $F_i(z_i; \mu)$ is $\mathsf{NEF\text{-}QVF}(\nu_i\mu, \nu_i V(\mu))$ and $G(\mu)$ is the conjugate prior distribution $\mathsf{CP}(\eta, \zeta)$.

4.2.1 Randomized by the conjugate prior

In this subsection, we show that the distributions generated from independent NEF-QVF samples mixed by the conjugate prior (NEF-QVF-CP) have multiplicative correlations and satisfy the Condition C, and list six concrete distributions of NEF-QVF-CP.

If Z is distributed $\mathsf{NEF\text{-}QVF}(\nu\mu, \nu V(\mu))$ with the density (4.5) and μ is distributed $\mathsf{CP}(\eta, \zeta)$ with the density (4.6), then the product $p\pi$ is

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

being integrated, the density of mixture Z becomes

$$p(z;\nu,\eta,\zeta) := b(z;\nu)\, K(\eta,\zeta) \,/\, K(\eta+z, \zeta+\nu). \tag{4.8}$$

Now, if each component of an n-dimensional random vector is distributed $\mathsf{NEF}(\nu_j\mu, \nu_j V(\mu))$ and μ is distributed $\mathsf{CP}(\eta, \zeta)$, then the mixture $\boldsymbol{Z} = (Z_1, \ldots, Z_n)$,

$$\prod_{j=1}^{n} \mathrm{NEF}(\nu_j\mu,\ \nu_j V(\mu)) \quad \bigwedge_{\mu} \quad \pi(\mu;\eta,\zeta),$$

has density

$$p(\boldsymbol{z};\boldsymbol{\nu};\eta,\zeta) = \left(\prod_{j=1}^{n} b(z_j;\nu_j)\right) K(\eta,\zeta) \Bigg/ K(\eta+t, \zeta+\nu) \tag{4.9}$$

$$= p(t;\nu,\eta,\zeta) \prod_{j=1}^{n} b(z_j;\nu_j) \,/\, b(t;\nu) \quad \text{where} \quad t = \sum_{j=1}^{n} z_j \quad \text{and} \quad \nu = \sum_{j=1}^{n} \nu_j. \tag{4.10}$$

The following theorem gives the first and second order moments of NEF-QVF-CP, and shows that the variance-covariance matrices are multiplicative. The following will be also used in the proof of Theorem 4.2.2.

Theorem 4.2.1 (Multiplicative correlation). *If the components of an n-dimensional random vector are independent and are distributed* $\mathsf{NEF\text{-}QVF}(\nu_j\mu, \nu_j V(\mu))$ *$(j = 1, \ldots, n)$ and μ is distributed the conjugate prior* $\mathrm{CP}(\eta, \zeta)$ *of* $\mathsf{NEF\text{-}QVF}(\mu, V(\mu))$*, then the mixture $\boldsymbol{Z}$ has*

$$\mathrm{E}(\boldsymbol{Z}) = \eta\boldsymbol{\nu}/\zeta, \qquad \mathrm{V}(\boldsymbol{Z}) = \frac{V(\eta/\zeta)}{\zeta - v_2}(\zeta \mathrm{diag}(\boldsymbol{\nu}) + \boldsymbol{\nu}\boldsymbol{\nu}^{\mathsf{T}}),$$

$$\textit{and} \qquad \rho(Z_i, Z_j) = \sqrt{\frac{\nu_i\nu_j}{(\zeta+\nu_i)(\zeta+\nu_j)}} \quad (i \neq j = 1, \ldots, n). \tag{4.11}$$

Proof. Let $V(\mu) = v_0 + v_1\mu + v_2\mu^2$ denote the variance function of $\mathsf{NEF\text{-}QVF}(\mu, V(\mu))$. Its $\mathrm{CP}(\eta, \zeta)$ 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 Theorem 5.3 in Morris (1983). Since

$$\mathrm{E}(Z_j|\mu) = \nu_j\mu, \quad \mathrm{E}(Z_j^2|\mu) = \nu_j V(\mu) + (\nu_j\mu)^2, \quad \text{and} \quad \mathrm{E}(Z_iZ_j|\mu) = \nu_i\nu_j\mu^2,$$

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

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

Hence $\mathrm{V}(\boldsymbol{Z})$ is obtained. □

Remark. Correlation coefficients of $\boldsymbol{Z}$ are always positive. The mixture of distributions, (4.7), is known to have positively dependence (see, Shaked, 1971; Marshall and Olkin, 1979).

Finally, partition $\boldsymbol{Z}$ as $\boldsymbol{Z} = (\boldsymbol{X}, \boldsymbol{Y}) = (X_1, \ldots, X_p, Y_1, \ldots, Y_q)$ $(p + q = n)$, and let us obtain the conditional density of $\boldsymbol{X}$ given $\boldsymbol{Y} = (y_1, \ldots, y_q)$.

In (4.7), the conditional distribution of $(Z_1, \ldots, Z_{n-1})$ given Z_n is

$$F(z_1, \ldots, z_{n-1}|z_n) = \int \left(\prod_{j=1}^{n-1} F_j(z_j;\mu)\right) dG(\mu|z_n), \quad dG(\mu|z_n) = \frac{F_n(z_n;\mu)\, dG(\mu)}{\int F_n(z_n;\mu)\, dG(\mu)},$$

that is, the conditional distribution is obtained by changing the prior distribution $G(\mu)$ to its posterior distribution $G(\mu|z_n)$. From (4.8), when the prior is the conjugate, the posterior is

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

This simple form is a merit of conjugate prior.

Now, when $\boldsymbol{Y} = (y_1, \ldots, y_q)$ are given, conditional distribution of $\boldsymbol{X}$ given $\boldsymbol{Y}$ is obtained by changing the prior $\mathrm{CP}(\eta, \zeta)$ to posterior $\mathrm{CP}(\eta + y, \zeta + \nu_y)$ where $y = \sum_{j=1}^{p} y_j$ and $\nu_y = \sum_{j=p+1}^{p+q} \nu_j$. Hence the conditional distribution of $\boldsymbol{X}$ given $\boldsymbol{Y}$ is

$$\prod_{j=1}^{p} \mathrm{NEF}(\nu_j\mu, \nu_j V(\mu)) \quad \bigwedge_{\mu} \quad \pi(\mu; \eta + y, \zeta + \nu_y),$$

with the density

$$\left(\prod_{j=1}^{p} b(x_j; \nu_j)\right) K(\eta + y, \zeta + \nu_y) \Big/ K(\eta + t, \zeta + \nu).$$

The following theorem shows that partial and conditional correlations coincide for NEF-QVF-CP.

Theorem 4.2.2 (Condition C). *Suppose that* $\boldsymbol{Z} = (Z_1, \ldots, Z_n)$ *are random variables in Theorem 4.2.1. If the distribution of* $(\boldsymbol{X}, \boldsymbol{Y}) = (X_1, \ldots, X_p, Y_1, \ldots, Y_q)$ $(p + q = n)$ *is equal to the distribution of* $\boldsymbol{Z}$*, then* $(\boldsymbol{X}, \boldsymbol{Y})$ *satisfies the Condition C.*

Proof. Since the conditional distribution of $\boldsymbol{X}$ given $\boldsymbol{Y}$ is obtained by changing the prior $\mathrm{CP}(\eta, \zeta)$ to posterior $\mathrm{CP}(\eta + y, \zeta + \nu_y)$, we have

$$\mathrm{E}(X_j|\boldsymbol{Y}) = \nu_j(\eta + y)/(\zeta + \nu_y) \quad \text{and} \quad \rho(X_i, X_j|\boldsymbol{Y}) = \sqrt{\nu_i\nu_j/(\zeta + \nu_y + \nu_i)(\zeta + \nu_y + \nu_j)}$$

where $i \neq j = 1, \ldots, p$ from Theorem 4.2.1. Hence, the Condition C is satisfied. □

Six classes of mixtures, NEF-QVF-CP are shown in Table 4.3. The first two columns of Table 4.3 shows multisample mixtures with density, (4.9). The third column shows the non-exponential constant part K of conjugate prior $\mathrm{CP}(\eta, \zeta)$, (4.6), which also appears in (4.9). The fourth column shows the expectation of $c(t, \nu)$ which is the last column of Table 4.2. The last column shows the distribution of the total size, following the univariate mixture distribution, appearing in (4.10).

Table 4.3: Members of NEF-QVF-CP.

Mixed and mixing distrb.	Mixture distribution $p(\boldsymbol{x});\ x = \sum_{j=1}^m x_j = t,\ \nu = \sum_{j=1}^m \nu_j$	$K(\eta, \zeta)$	$\mathrm{E}(c(t, \nu))$	distrb. 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	$\mathsf{MsMorrisMixture}(\boldsymbol{\nu}; \eta, \zeta)$ $\frac{H(\eta + x, \zeta + \nu)}{H(\eta, \zeta)} \prod_{j=1}^m b(x_j; \nu_j)$	$\frac{1}{H(\eta, \zeta)}$	$\frac{\eta^2 + \zeta^2}{\zeta^2(\zeta - 1)}$	$\mathsf{MorrisMixture}(\nu; \eta, \zeta)$

4.2.2 Randomized by other priors

This subsection shows that when the parameter is randomized by a non-conjugate prior a distribution of independent NEF-QVF samples has a multiplicative correlation but do not satisfy the Condition C in Example 4.2.1, and that when the location parameter is randomized a distribution of independent non-NEF-QVF samples do not satisfy the Condition C in Remark.

Example 4.2.1. Neyman type A distribution is the mixture $Z \sim \mathsf{Po}(\nu k) \bigwedge_k \mathsf{Po}(\lambda)$, with the pmf

$$p(z) = \mathrm{E}^k(p(z|k)) = \frac{\nu^z}{z!}\ \mathrm{E}^k(e^{-\nu k}k^z),$$

and the factorial moments

$$\mathrm{E}(Z^{\underline{r}}) = \mathrm{E}^k(Z^{\underline{r}}|k) = \mathrm{E}^k((\nu k)^r) = \nu^r \sum_{l=1}^{r} \left\{ {r \atop l} \right\} \lambda^l,$$

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

A definition of multivariate Neyman type A distributions is the mixture $\boldsymbol{Z} = (Z_1, \ldots, Z_n) \sim \prod_{j=1}^n \mathsf{Po}(\nu_j k) \bigwedge_k \mathsf{Po}(\lambda)$, with the factorial moments

$$\mathrm{E}\left(\prod_{j=1}^{n} Z_j^{\underline{r_j}}\right) = \mathrm{E}^k\left(\prod_{j=1}^{n}(\nu_j k)^{\underline{r_j}}\right) = \left(\prod_{j=1}^{n} \nu_j^{\underline{r_j}}\right)\sum_{l=1}^{r}\left\{ {r \atop l} \right\} \lambda^l, \quad r = \sum_{j=1}^{n} r_j.$$

Hence, its variance-covariance matrix is

$$\mathrm{V}(\boldsymbol{Z}) = \lambda(\mathrm{diag}(\boldsymbol{\nu}) + \boldsymbol{\nu}\boldsymbol{\nu}^{\mathsf{T}}), \quad \boldsymbol{\nu} = (\nu_1, \ldots, \nu_n).$$

To find conditional moments $\mathrm{E}(Z_i^{\underline{r}}|Z_n)$, $\mathrm{E}(Z_1^{\underline{r}}|Z_2 = z_2)$ is calculated without loss of generality.

$$\begin{aligned}
\mathrm{E}(Z_1^{\underline{r}}|z_2) &= \mathrm{E}^k\left(\mathrm{E}(Z_1^{\underline{r}}|k)\ p(z_2|k)\right) / \mathrm{E}^k(p(z_2|k)) \\
&= \mathrm{E}^k\left((\nu_1 k^r)\ e^{-\nu_2 k}k^{z_2}\right) / \mathrm{E}^k(e^{-\nu_2 k}\ k^{z_2}) \\
&= \sum_{l=1}^{z_2+r}\left\{ {z_2 + r \atop l} \right\} (\lambda\ e^{-\nu_2})^l \Bigg/ \sum_{l=1}^{z_2}\left\{ {z_2 \atop l} \right\} (\lambda\ e^{-\nu_2})^l,
\end{aligned}$$

where the numerator is 1 if $z_2 = 0$. Hence

$$\begin{aligned}
\mathrm{E}(Z_1|Z_2 = 0) &= \nu_1\ \lambda\ e^{-\nu_2}, \\
\mathrm{E}(Z_1|Z_2 = 1) &= \nu_1\ (1 + \lambda\ e^{-\nu_2}), \\
\mathrm{E}(Z_1|Z_2 = 2) &= \nu_1\ (1 + 3\lambda\ e^{-\nu_2} + (\lambda\ e^{-\nu_2})^2) / (1 + \lambda\ e^{-\nu_2}),
\end{aligned}$$

and so on. $\mathrm{E}(Z_1|Z_2)$ is not linear in Z_2, and $\boldsymbol{Z}$ does not satisfy the Condition C.

Remark. We consider a distribution of independent discrete uniform, not NEF, samples when the location parameter is randomized as follows.

In the expression (4.7) , if $F_i(z_i;\ \mu) = F_i(z_i - \mu)$, the components of the mixture $\boldsymbol{Z}$ is written as $Z_i = W_i + W_0,\ W_i \sim F_i(\cdot),\ W_0 \sim G(\cdot),\ i = 1, \ldots, n$, where $(W_i)_{i=0}^n$ are independent.

Let $(W_i)_{i=1}^n$ be iid random variables with a pmf p_1, and W_0 be independent of $(W_i)_{i=1}^n$ with a pmf p_2. Define $Z_i = W_i + W_0, \quad i = 1, \ldots, n$, and the conditional pmf of (Z_1, Z_2) given $Z_3 = z_3, \ldots, Z_m = z_n$, is

$$p(z_1, z_2 | z_3, \ldots, z_n) = \sum_w (\prod_{j=1}^n p_1(z_j - w)) p_0(w) \Big/ \sum_w (\prod_{j=3}^n p_1(z_j - w)) p_0(w).$$

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

$$\begin{aligned} p(z_1, \ldots, z_n) &= (a+1)^{-n}(b+1)^{-1} \sum_w I[0 \le z_1 - w \le a, \ldots,\ 0 \le z_n - w \le a, 0 \le w \le b] \\ &= (a+1)^{-n}(b+1)^{-1}(n^{**} - n^* + 1) I[n^* \le n^{**}], \\ n^{**} &= \min(z_1, \ldots, z_n, b),\ n^* = \max(z_1 - a, z_n - a, 0), \end{aligned}$$

and

$$\begin{aligned} p(z_1, z_2 | z_3, \ldots, z_n) &= \frac{(a+1)^{n-2}(b+1)}{(a+1)^n (b+1)} \frac{(n^{**} - n^* + 1) I[n^* \le n^{**}]}{(l^{**} - l^* + 1)} \\ &= \frac{1}{(a+1)^2 (l^{**} - l^* + 1)} (\min(z_1, z_2, l^{**} - \max(z_1 - a, z_2 - a, l^*)), \\ l^{**} &= \min(z_3, \ldots, z_n, b),\ l^* = \max(z_3 - a, \ldots, z_n - a, 0),\ 0 \le l^* \le l^{**} \le b. \end{aligned}$$

For $a = 2, b = 1, n \ge 3$, the conditional distributions given $(l^{**}, l^*),\ 1 \ge l^{**} \ge l^* \ge 0$, are shown in **Table** 4.4 with correlations. The conditional correlations depend on the statistics (l^{**}, l^*) which are not constant. Hence the Condition C is not satisfied.

Table 4.4: Conditional distributions and correlations.

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

probabilities×9

Z_1/Z_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

$\rho(Z_1, Z_2|l^*, l^{**}) = 0$,
$\mathrm{E}(Z_i|l^*, l^{**}) = 1$

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

probabilities×18

Z_1/Z_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

$\rho(Z_1, Z_2|l^*, l^{**}) = 3/11$,
$\mathrm{E}(Z_i|l^*, l^{**}) = 1.5$

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

probabilities×9

Z_1/Z_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

$\rho(Z_1, Z_2|l^*, l^{**}) = 0$,
$\mathrm{E}(Z_i|l^*, l^{**}) = 2$

Acknowledgements

I am deeply indebted to my supervisor, Professor Ritei Shibata, who had also been the supervisor for my bachelor's course. After my teaching experience at Keio Shonan-Fujisawa Junior and Senior High School for five years, he suggested that I should finalize my interest into doctorate. Thus, his excellent advice and encouragement have always been with me for over fifteen years. I am very glad to have studied under his guidance.

I would like to thank Professor Makoto Maejima, Professor Kunio Shimizu, and Professor Nobuo Shinozaki, for their careful reading of the thesis and a number of useful suggestions which led to this final draft.

My sincere gratitude is especially conveyed to Professor Masaaki Sibuya at Takachiho University, whose enormous knowledge aided me in the development of the study.

I was fortunate to have been surrounded by the supportive members of Data Science Division at Keio University, and the staff of Keio Shonan-Fujisawa Junior and Senior High School, who allowed me a fruitful sabbatical year.

Finally, I wish to say special thanks to my wife, daughter, and parents. Without their affectionate support and encouragement, I would not have been able to accomplish this work.

Kunihiro Baba

Bibliography

Abbott, F. and Dayan, P. (1999) The effect of correlated variability on the accuracy of a population code, *Neural Computation*, **11**, 91–101.

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.

Baba, K. and Shibata, R. (2005) Multiplicative correlations, *Ann. Inst. Statist. Math.*, (submitted).

Baba, K. and Sibuya, M. (2005) Equivalence of partial and conditional correlation coefficients, *J. Japan Statist. Soc.*, (to appear).

Cambanis, S., Huang, S., and Simons, G. (1981) On the theory of elliptically contoured distributions, *J. Multivariate Anal.*, **11**, 368–385.

Coles, S. (2001) *An Introduction to Statistical Modeling of Extreme Values*, Springer, London.

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

Curnow, R. N., and Dunnett, C. W. (1962) The numerical evaluation of certain multivariate normal integrals, *Ann. Math. Statist.*, **33**, 571–579.

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

Edwards, D.(1995) *Introduction to Graphical Modelling*, Springer: New York.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin.

Fang, K. T., Kotz, S., and Ng, K. W. (1990) *Symmetric Multivariate and Related Distributions*, Chapman and Hall: New York.

Gupta, S. S. (1963) Probability integrals of multivariate normal and multivariate t, *Ann. Math. Statist.*, **34**, 792-828.

Gupta, R. D., and Richards, D. S. P. (2001) The covariance structure of the multivariate Liouville distributions, *Algebraic Methods in Statistics and Probability*, American Mathematical Society, 125–138.

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

Joe, H. (1997) *Multivariate Models and Dependence Concepts*, Chapman and Hall: New York.

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

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1993) *Continuous Univariate Distributions*, Vol. 1, New York: John Wiley & Sons.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1997) *Discrete Multivariate Distributions*, New York: John Wiley & Sons.

Kelderman, H. (2004) Measurement exchangeability and normal one-factor models, *Biometrika*, **91**, 738–742.

Kelker, D. (1970) Distibution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā*, **A, 32**, 419–430.

Khatri, C. G. (1967) On certain inequalities for normal distributions and their applications to simultaneous confidence bounds, *Ann. Math. Statist.*, **38**, No. 6, 1853–1867.

Kotz, S., Balakrishnan, N., and Johnson, N. L. (2000) *Continuous Multivariate Distributions*, New York: John Wiley & Sons.

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

Lawrance, A. J. (1976) On conditional and partial correlation. *The American Statistician*, **30**, 146–149.

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

Mardia, K. V. (1970) *Families of Bivariate Distributions*, London: Griffin.

Mari, D. D. and Kotz, S. (2001) *Correlation and Dependence*, Imperial College Press: London.

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.

Nelsen, R. B. (1999) *An Introduction to Copulas*, Springer: New York.

Ronning, G. (1982) Characteristic values and triangular factorization of the covariance matrix for multinomial, Dirichlet and multivariate hypergeometric distributions and some related results, *Statist. Hefte*, **23**, No. 3, 152–176.

Royen, T. (1991) Multivariate gamma distributions with one-factorial accompanying correlation matrices and applications to the distribution of the multivariate range, *Metrika*, **38**, 299–315.

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

Sibuya, M. (1960) Bivariate extreme statistics. *Ann. Inst. Statist. Math.*, **11**, 195–210.

Six, F. B. (1981) Representations of multivariate normal distributions with special correlation structures, *Comm. Statist. A—Theory Methods* 10 (13), 1285–1295.

Steiger, J. H. (1979) Factor indeterminacy in the 1930's and the 1970's some interesting parallels, *Psychometrika*, **44**, No. 1, 157–167.

Watson, G. S. (1996) Spectral decomposition of the covariance matrix of a multinomial, *J. Roy. Statist. Soc.* B. **58**, No. 1, 289–291.

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