Aust. N. Z. J. Stat. **46**(4), 2004, 657–664

PARTIAL CORRELATION AND CONDITIONAL CORRELATION AS MEASURES OF CONDITIONAL INDEPENDENCE

KUNIHIRO BABA[1*], RITEI SHIBATA[1] AND MASAAKI SIBUYA[2]

Keio University and Takachiho University

Summary

This paper investigates the roles of partial correlation and conditional correlation as measures of the conditional independence of two random variables. It first establishes a sufficient condition for the coincidence of the partial correlation with the conditional correlation. The condition is satisfied not only for multivariate normal but also for elliptical, multivariate hypergeometric, multivariate negative hypergeometric, multinomial and Dirichlet distributions. Such families of distributions are characterized by a semigroup property as a parametric family of distributions. A necessary and sufficient condition for the coincidence of the partial covariance with the conditional covariance is also derived. However, a known family of multivariate distributions which satisfies this condition cannot be found, except for the multivariate normal. The paper also shows that conditional independence has no close ties with zero partial correlation except in the case of the multivariate normal distribution; it has rather close ties to the zero conditional correlation. It shows that the equivalence between zero conditional covariance and conditional independence for normal variables is retained by any monotone transformation of each variable. The results suggest that care must be taken when using such correlations as measures of conditional independence unless the joint distribution is known to be normal. Otherwise a new concept of conditional independence may need to be introduced in place of conditional independence through zero conditional correlation or other statistics.

Key words: elliptical distribution; exchangeability; graphical modelling; monotone transformation.

1. Introduction

Conditional independence is a key notion in multivariate analyses such as graphical modelling, where two vertices are connected if and only if the corresponding variables are *not* conditionally independent; see, for example, Whittaker (1990 Section 3.2) and Edwards (1995 Section 1.3). To confirm the conditional independence, particularly when the variables are continuous, it is a common practice to check whether or not the partial correlation is close enough to zero. This is done because it is assumed that zero partial correlation suggests that the variables are conditionally independent, or nearly so. One of the questions we ask in this paper is whether this assumption is true when we depart from normal distributions. As is shown in the next section, the answer is negative. Even the importance of conditional independence itself is somewhat doubtful.

First we prove a theorem which provides a necessary and sufficient condition for the coincidence of the partial covariance with the expectation of the conditional covariance — another

Received October 2002; revised November 2003; accepted February 2004.

*Author to whom correspondence should be addressed.

[1] Dept of Mathematics, Keio University, Yokohama, 223-8522, Japan. e-mail: baba@math.keio.ac.jp

[2] School of Business Administration, Takachiho University, Suginami Tokyo, 168-8508, Japan.

measure of conditional independence. The condition is, essentially, linearity of the conditional expectation. A corollary of this theorem shows that the partial correlation is identical to the conditional correlation if the conditional correlation is independent of the value of the condition, and also if the conditional expectation is linear. A family of elliptical distributions provides a simple example, but is not the only case. We characterize families of distributions which include multinomial, hypergeometric, negative hypergeometric and Dirichlet distributions. Negative multinomial or multivariate Pareto distributions are not included in this characterization, but they satisfy the conditions of the corollary. Another corollary gives us a necessary and sufficient condition for the partial variance–covariance being identical to the conditional variance–covariance.

We then proceed to the equivalence between the zero partial correlation or the zero conditional correlation and the conditional independence. Since it is hard to establish a general theory, we investigate two cases instead. In the first case where the family of distributions is elliptical, the variables are conditionally independent given the rest, if and only if the distribution is multivariate normal. We do not see any conditional independence if multivariate normal is excluded from this family. Second, we take the case of transformed normal distributions. We show that the equivalence between the zero conditional covariance and the conditional independence is retained by any monotone transformation of each variable. However, this is not true for the zero partial correlation.

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 or zero conditional correlation. If it is replaced by zero partial correlation, 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 is calculated more easily than conditional covariance, which depends on the shape of the distribution, but it may further depart from conditional independence. Therefore, zero conditional correlation is preferable to zero partial correlation, considering that conditional independence is more meaningful than conditional orthogonality.

2. Partial correlation and conditional correlation

We are interested in the conditional independence of X_1 and X_2 given $\boldsymbol{Y} = (Y_1, \ldots, Y_p)$. We hereafter assume, for simplicity, that the variance–covariance matrix of $\boldsymbol{Y}$ is positive definite. The partial variance–covariance matrix for $\boldsymbol{X} = (X_1, X_2)$ is defined as

$$\boldsymbol{\Sigma}_{\boldsymbol{XX}.\boldsymbol{Y}} = \begin{bmatrix} \sigma_{11.\boldsymbol{Y}} & \sigma_{12.\boldsymbol{Y}} \\ \sigma_{21.\boldsymbol{Y}} & \sigma_{22.\boldsymbol{Y}} \end{bmatrix},$$

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

$$\mathrm{V}\left(\begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix}\right) = \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{XX}} & \boldsymbol{\Sigma}_{\boldsymbol{XY}} \\ \boldsymbol{\Sigma}_{\boldsymbol{YX}} & \boldsymbol{\Sigma}_{\boldsymbol{YY}} \end{bmatrix}, \text{ where } \boldsymbol{\Sigma}_{\boldsymbol{XX}} \text{ is } 2 \times 2,\ \boldsymbol{\Sigma}_{\boldsymbol{XY}} \text{ is } 2 \times p, \text{ and } \boldsymbol{\Sigma}_{\boldsymbol{YY}} \text{ is } p \times p\,.$$

The partial correlation is then

$$\rho_{12.\boldsymbol{Y}} = \frac{\sigma_{12.\boldsymbol{Y}}}{\sqrt{\sigma_{11.\boldsymbol{Y}}\, \sigma_{22.\boldsymbol{Y}}}}\,.$$

The partial variance or covariance given $\boldsymbol{Y}$ can be considered as the variance or covariance between residuals of projections of X_1 and X_2 on the linear space spanned by $\boldsymbol{Y}$,

$$\sigma_{ij.\boldsymbol{Y}} = \operatorname{cov}\left(X_i - \hat{X}_i(\boldsymbol{Y}),\ X_j - \hat{X}_j(\boldsymbol{Y})\right) \quad \text{for } i, j = 1, 2\,.$$

Here $\hat{\boldsymbol{X}}(\boldsymbol{Y}) = \mathrm{E}(\boldsymbol{X}) + \boldsymbol{\Sigma}_{\boldsymbol{XY}}\boldsymbol{\Sigma}_{\boldsymbol{YY}}^{-1}(\boldsymbol{Y} - \mathrm{E}(\boldsymbol{Y}))$ is the projection of $\boldsymbol{X}$; that is, the conditional expectation of $\boldsymbol{X}$ given $\boldsymbol{Y}$.

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

$$\operatorname{cov}(X_i, X_j \mid \boldsymbol{Y}) = \mathrm{E}\big(\big(X_i - \mathrm{E}(X_i \mid \boldsymbol{Y})\big)\big(X_j - \mathrm{E}(X_j \mid \boldsymbol{Y})\big) \mid \boldsymbol{Y}\big)\,.$$

We use the following notation for the conditional covariance matrix

$$\boldsymbol{\Sigma}_{\boldsymbol{XX}\mid\boldsymbol{Y}} = \begin{bmatrix} \sigma_{11\mid\boldsymbol{Y}} & \sigma_{12\mid\boldsymbol{Y}} \\ \sigma_{21\mid\boldsymbol{Y}} & \sigma_{22\mid\boldsymbol{Y}} \end{bmatrix},$$

and for the conditional correlation

$$\rho_{12\mid\boldsymbol{Y}} = \frac{\sigma_{12\mid\boldsymbol{Y}}}{\sqrt{\sigma_{11\mid\boldsymbol{Y}}\sigma_{22\mid\boldsymbol{Y}}}}\,.$$

The expectation of the conditional covariance is not necessarily equal to the covariance; that is, $\mathrm{E}(\boldsymbol{\Sigma}_{\boldsymbol{XX}\mid\boldsymbol{Y}}) \neq \mathrm{V}(\boldsymbol{X})$.

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

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

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

with a two-dimensional distribution function H. This type of specification is similar to a copula (see e.g. Nelsen, 1999 Chapter 2). If we further assume that H has zero means, unit variances and correlation 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

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

Hence, the conditional correlation is the constant θ but the conditional covariance depends on y. To see how conditional covariance or correlation varies with y and when it coincides with the partial covariance or correlation, 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 $\sigma_{12\mid Y} = \theta\sigma_1\sigma_2 = \sigma_{12.Y}$.

(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 $\sigma_{12\mid Y} = \theta\sigma(Y)^2$ and $\sigma_{12.Y} = \theta\,\mathrm{E}(\sigma(Y)^2)$ 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 correlations are equal; $\rho_{12\mid Y} = \theta = \rho_{12.Y}$.

(iii) If $\mu_i(y) = y^2$, $\sigma_i(y) = \sigma_i$ for $i = 1, 2,$ then $\sigma_{12.Y} = (\theta + \operatorname{var}(Y^2)(1 - \rho(Y, Y^2)^2))\sigma_1\sigma_2$ and $\sigma_{12\mid Y} = \theta\sigma_1\sigma_2$. 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 between the partial variance–covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{XX}.\boldsymbol{Y}}$ and the expectation of the conditional variance–covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{XX}\mid\boldsymbol{Y}}$. In fact, Corollary 1 is a generalization of Example 1 (ii), and Corollary 2 is a generalization of Example 1 (i).

Theorem 1. *For any random vectors* $\boldsymbol{X} = (X_1, X_2)$ *and* $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_p)$ *the following two conditions are equivalent.*

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

(ii) $\boldsymbol{\Sigma}_{\boldsymbol{XX}.\boldsymbol{Y}} = \mathrm{E}(\boldsymbol{\Sigma}_{\boldsymbol{XX}\mid\boldsymbol{Y}})$.

Proof. Since $\boldsymbol{\Sigma}_{\boldsymbol{XX}.\boldsymbol{Y}} = \mathrm{E}(\boldsymbol{\Sigma}_{\boldsymbol{XX}\mid\boldsymbol{Y}}) + \mathrm{V}(\mathrm{E}(\boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Y})\mid\boldsymbol{Y}))$, we have $\boldsymbol{\Sigma}_{\boldsymbol{XX}.\boldsymbol{Y}} = \mathrm{E}(\boldsymbol{\Sigma}_{\boldsymbol{XX}\mid\boldsymbol{Y}})$ equivalent to $\mathrm{V}(\mathrm{E}(\boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Y})\mid\boldsymbol{Y})) = 0$, and also to $\mathrm{E}(\boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Y})\mid\boldsymbol{Y}) = \boldsymbol{\beta}$ 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. As a corollary of Theorem 1, we have Corollary 1.

Corollary 1. *For any random vectors* $\boldsymbol{X} = (X_1, X_2)$ *and* $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_p)$, *if there exists a vector* $\boldsymbol{\alpha}$ *and a matrix* $\boldsymbol{B}$ *such that*

$$\mathrm{E}(\boldsymbol{X}\mid\boldsymbol{Y}) = \boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{Y} \quad \textit{and } \rho_{12\mid\boldsymbol{Y}} \textit{ does not depend on } \boldsymbol{Y}\,,$$

then $\rho_{12.\boldsymbol{Y}} = \rho_{12\mid\boldsymbol{Y}}$ *a.s.*

Proof. From Theorem 1, if $\mathrm{E}(\boldsymbol{X}\mid\boldsymbol{Y})$ is a linear function of $\boldsymbol{Y}$, then $\rho_{12.\boldsymbol{Y}} = \mathrm{E}(\rho_{12\mid\boldsymbol{Y}})$. The assertion of the corollary holds true, since $\rho_{12\mid\boldsymbol{Y}}$ is independent of $\boldsymbol{Y}$.

The elliptical distribution is a natural generalization of the multivariate normal distribution. Corollary 1 holds true for this distribution. However, zero partial correlation or zero conditional correlation do not imply conditional independence except for normal distributions. This is shown in the next section.

Example 2 (Elliptical distribution). The elliptical distribution is a family of distributions whose characteristic function takes the form

$$\Psi(\boldsymbol{t}) = \exp(i\boldsymbol{t}^{\top}\boldsymbol{\mu})\phi(\boldsymbol{t}^{\top}\boldsymbol{\Sigma}\boldsymbol{t}) \quad \text{for some scalar function } \phi$$

(see e.g. Fang, Kotz & Ng, 1990 p.31). This family is denoted by $\mathrm{EC}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$. From Cambanis, Huang & Simons (1981 Corollary 5), if $(\boldsymbol{X}, \boldsymbol{Y}) \overset{\mathrm{d}}{=} \mathrm{EC}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$, then

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

where s is a function and the matrix $\boldsymbol{\Sigma}^{*}$ is independent of the value of $\boldsymbol{Y}$. The conditional distribution is also elliptical. This shows that Corollary 1 holds true for the elliptical distribution and the partial correlation is identical to the conditional correlation.

To investigate a family of distributions which satisfies the condition given in Corollary 1, we introduce Lemma 1.

Lemma 1. *Let* $\mathcal{F} = \{F_\theta\colon \theta \in \Theta\}$ *denote a family of distribution functions with a parameter space* Θ *which is* $(0, \infty)$ *or the set of all natural numbers. If this family* $\mathcal{F}$ *has a semigroup property such that* $F_{\theta_1} * F_{\theta_2} = F_{\theta_1+\theta_2} \in \mathcal{F}$ *for the convolution of any* $F_{\theta_1}, F_{\theta_2} \in \mathcal{F}$, *then for independent variables* Z_i, *where* $Z_i \overset{\mathrm{d}}{=} F_{\theta_i} \in \mathcal{F}$ $(i = 1, 2, \ldots, n+1)$, *we have*

$$\mathrm{E}(Z_j \mid \mathcal{C}) = \frac{\theta_j(N - \sum_{i=3}^n z_i)}{\theta_1 + \theta_2 + \theta_{n+1}} \quad \textit{for} \quad j = 1, 2$$

and

$$\rho(Z_1, Z_2 \mid \mathcal{C}) = -\sqrt{\frac{\theta_1\theta_2}{(\theta_1 + \theta_{n+1})(\theta_2 + \theta_{n+1})}}\,,$$

where $\mathcal{C}$ *denotes the condition* $\{Z_3 = z_3, \ldots, Z_n = z_n, \sum_{i=1}^{n+1} Z_i = N\}$ *and* N *is a fixed number. (If* $\Theta = (0, \infty)$, $F_\theta \in \mathcal{F}$ *is infinitely divisible.)*

Proof. From the definition of the condition $\mathcal{C}$, we have

$$\mathrm{E}(Z_1 \mid \mathcal{C}) = \mathrm{E}\left(N - Z_2 - \sum_{i=3}^n z_i - Z_{n+1} \,\middle|\, \mathcal{C}\right) = N - \sum_{i=3}^n z_i - \mathrm{E}(Z_2 \mid \mathcal{C}) - \mathrm{E}(Z_{n+1} \mid \mathcal{C})\,.$$

It is sufficient to consider the case when θ_i $(i = 1, 2, \ldots, n+1)$ 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 b_i are positive integers. From the property of $\mathcal{C}$, there exist independent random variables $(Z_{i1}, Z_{i2}, \ldots, Z_{ib_i})$ for each Z_i such that

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

We have then

$$\mathrm{E}(Z_1 \mid \mathcal{C}) = \sum_{j=1}^{b_1} \mathrm{E}(Z_{1j} \mid \mathcal{C}) = N - \sum_{i=3}^n z_i - \sum_{j=1}^{b_2} \mathrm{E}(Z_{2j} \mid \mathcal{C}) - \sum_{j=1}^{b_{n+1}} \mathrm{E}(Z_{n+1,j} \mid \mathcal{C})\,.$$

Since the variables $\{Z_{ij}\colon j = 1, 2, \ldots, b_i\}$ are exchangeable given $\mathcal{C}$, $i = 1, 2, n+1$, we have

$$(b_1 + b_2 + b_{n+1})\mathrm{E}(Z_{ij} \mid \mathcal{C}) = N - \sum_{i=3}^n z_i\,.$$

Therefore we have

$$\mathrm{E}(Z_1 \mid \mathcal{C}) = b_1\mathrm{E}(Z_{ij} \mid \mathcal{C}) = \frac{b_1}{b_1 + b_2 + b_{n+1}}\left(N - \sum_{i=3}^n z_i\right) = \frac{\theta_1}{\theta_1 + \theta_2 + \theta_{n+1}}\left(N - \sum_{i=3}^n z_i\right).$$

The same discussion can be applied for $\mathrm{E}(Z_2 \mid \mathcal{C})$. For the conditional second moments we can write

$$\mathrm{E}(Z_1^2 \mid \mathcal{C}) = \mathrm{E}\left(Z_1\left(N - \sum_{i=3}^n z_i - Z_2 - Z_{n+1}\right) \,\middle|\, \mathcal{C}\right) = \frac{b_1(N - \sum_{i=3}^n z_i)^2}{b_1 + b_2 + b_{n+1}} - b_1(b_2 + b_{n+1})\eta$$

and
$$\mathrm{E}(Z_1 Z_2 \mid \mathcal{C}) = \mathrm{E}\Big(\Big(\sum_{j=1}^{b_1} Z_{1j}\Big)\Big(\sum_{k=1}^{b_2} Z_{2k}\Big) \,\Big|\, \mathcal{C}\Big) = b_1 b_2 \eta\,,$$

with $\eta = \mathrm{E}(Z_{1j} Z_{2k} \,|\, \mathcal{C})$ $(j = 1, 2, \ldots, b_1,\ k = 1, 2, \ldots, b_2)$. Combining these results, we have the latter part of the result from

$$\mathrm{V}\big((Z_1, Z_2) \mid \mathcal{C}\big) = \Big(\Big(\frac{N - \sum_{i=3}^{n} z_i}{b_1 + b_2 + b_{n+1}}\Big)^2 - \eta\Big)\begin{bmatrix} b_1(b_2 + b_{n+1}) & -b_1 b_2 \\ -b_1 b_2 & b_2(b_1 + b_{n+1}) \end{bmatrix}.$$

The following theorem is a direct consequence of this lemma and Corollary 1.

Theorem 2. *Suppose that $Z_1, Z_2, \ldots, Z_{n+1}$ are random variables in Lemma 1. If the distribution of $(X_1, X_2, Y_1, Y_2, \ldots, Y_{n-2})$ is equal to the conditional distribution of $(Z_1, Z_2, \ldots, Z_n)$ given $\sum_{i=1}^{n+1} Z_i = N$, then $\rho_{X_1 X_2 \boldsymbol{\cdot} \boldsymbol{Y}} = \rho(X_1, X_2 \,|\, \boldsymbol{Y})$ a.s., which is always negative.*

In fact, various types of multivariate distributions satisfy the condition of this theorem.

Example 3. If Z_i $(i = 1, 2, \ldots, n+1)$ are independent binomial variables, then the conditional distribution of $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_n)$ given $\sum_{i=1}^{n+1} Z_i = N$ is the multivariate hypergeometric distribution. Therefore, the partial correlation equals the conditional correlation as described in the theorem. Similarly, we see that it also holds true for the multivariate negative hypergeometric, multinomial and Dirichlet distributions if the Z_i $(i = 1, 2, \ldots, n+1)$ are taken to be negative binomial, Poisson and chi-squared variables, respectively. On the other hand, we can show that negative multinomial or multivariate Pareto distributions are beyond the scope of Theorem 2, but Corollary 1 holds. For details of these distributions, see Johnson, Kotz & Balakrishnan (1997) and Kotz, Balakrishnan & Johnson (2000).

We have discussed the equivalence of the partial correlation and the conditional correlation. For the covariance, we have the following corollary as a direct consequence of Theorem 1.

Corollary 2 of Theorem 1. *For any random vectors $\boldsymbol{X} = (X_1, X_2)$ and $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_p)$, the following two conditions are equivalent.*

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

Kelker (1970 Theorem 6 and Theorem 7) showed that the multivariate normal distribution is the only distribution that satisfies condition (i) of Corollary 2 in the class of elliptical distributions. On the other hand, it is also true that we can construct various families of distributions which satisfy this condition other than the normal, by introducing various distributions as H in Example 1.

3. Conditional independence

We have seen that the partial covariance and the conditional covariance become equivalent even for non-normal distributions. One of the aims of this paper is to determine the effectiveness of such quantities in specifying conditional independence. However, it is important to check if conditional independence is well defined or not. For example, X_1 and X_2 in Theorem 2 cannot be conditionally independent. Also for elliptical distributions, if a pair of variables is required in a family of elliptical distributions, then the family becomes no more than that of multivariate normal distributions.

Theorem 3. *Assume that* $(X_1, X_2, Y_1, Y_2, \ldots, Y_{n-2})$ *is distributed as* $\mathrm{EC}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ *with* $\boldsymbol{\Sigma} > 0$ *and that the conditional variance of* $\boldsymbol{X}$ *given* $\boldsymbol{Y}$ *exists and is finite.* X_1 *and* X_2 *are conditionally independent given* $\boldsymbol{Y}$ *if and only if the family of distributions is a multivariate normal with zero partial correlation.*

Proof. It is known that two variables with distributions in the class of elliptical distributions are independent if and only if they are normally distributed. This implies that conditional independence of X_1 and X_2 is equivalent to the normality of the conditional distribution. It is sufficient to note that normality of the conditional distribution is equivalent to the normality of joint distributions in a family of elliptical distributions.

This theorem implies that conditional independence is quite restrictive; it does not exist even in a family of elliptical distributions which is believed to be a natural extension of the normal distribution.

Next, we consider $\boldsymbol{X} = (X_1, X_2)$ which has a bivariate normal distribution, and let $\boldsymbol{Z} = \psi(\boldsymbol{X}) = (\psi_1(X_1), \psi_2(X_2))$ be transformed variables by monotone increasing or decreasing functions ψ_1 and ψ_2.

Theorem 4. *Zero covariance, i.e.* $\mathrm{cov}(Z_1, Z_2) = 0$, *is equivalent to the independence of* Z_1 *and* Z_2 *for any strictly monotone increasing (decreasing) transformations* ψ_1 *and* ψ_2.

Proof. It is sufficient to prove the theorem for monotone increasing transformations ψ_1 and ψ_2. We can assume that the variances of X_1 and X_2 are 1 without loss of generality. Since the conditional expectation of Z_2 given $Z_1 = z_1$ is

$$\begin{aligned}\mathrm{E}(Z_2 \mid Z_1 = z_1) &= \frac{1}{\sqrt{1-\rho^2}} \int \psi_2(x_2)\varphi\Big(\frac{x_2 - \rho\psi_1^{-1}(z_1)}{\sqrt{1-\rho^2}}\Big)\, dx_2 \\ &= \frac{1}{\sqrt{1-\rho^2}} \int \psi_2\big(x + \rho\psi_1^{-1}(z_1)\big)\varphi\Big(\frac{x}{\sqrt{1-\rho^2}}\Big)\, dx\,,\end{aligned}$$

where φ is the standard normal density function, and the conditional expectation is strictly increasing in z_1 if $\rho > 0$, and decreasing in z_1 if $\rho < 0$. Therefore, $\mathrm{cov}(Z_1, Z_2)$ should be positive or negative according to $\rho > 0$ or $\rho < 0$. In fact, for the case when $\rho > 0$, the conditional expectation is strictly increasing and

$$(z - \mu)\big(\mathrm{E}(Z_2 \mid Z_1 = z) - \mathrm{E}(Z_2 \mid Z_1 = \mu)\big) > 0 \quad \text{for any } z \neq \mu\,.$$

Therefore,

$$\mathrm{cov}(Z_1, Z_2) = \mathrm{E}(Z_1 Z_2) - \mathrm{E}(Z_1)\mathrm{E}(Z_2) > 0\,.$$

The same discussion applies when $\rho < 0$. This implies that $\mathrm{cov}(Z_1, Z_2) \neq 0$ as far as $\rho \neq 0$, and the result is obtained.

This theorem is essentially known. For example, Sibuya (1960 Remark 2) showed the dependence function is the constant 1 if and only if $\rho = 0$. The same result can be found in Joe (1997 p. 141) in the context of a copula. As a direct consequence of this theorem, we have the following corollary.

Corollary 3. *Assume that the conditional distribution of* $\boldsymbol{X} = (X_1, X_2)$ *given* $\boldsymbol{Y}$ *is bivariate normal. Then for any monotone increasing (decreasing) transforms* ψ_1 *and* ψ_2, $\mathrm{cov}(Z_1, Z_2 \mid \boldsymbol{Y}) = 0$ *a.s. is equivalent to the conditional independence of* $Z_1 = \psi_1(X_1)$ *and* $Z_2 = \psi_2(X_2)$.

Example 4 (Lognormal distribution). If $\psi_1(t) = \psi_2(t) = \exp(t)$, then the conditional distribution of $\boldsymbol{Z} = (Z_1, Z_2)$ given $\boldsymbol{Y}$ is lognormal. From Corollary 3, the zero conditional covariance is equivalent to the conditional independence of Z_1 and Z_2 given $\boldsymbol{Y}$. In fact,

$$\begin{aligned}\operatorname{cov}(Z_1, Z_2 \mid \boldsymbol{Y}) = \exp\bigl(&\mathrm{E}(X_1 \mid \boldsymbol{Y}) + \mathrm{E}(X_2 \mid \boldsymbol{Y}) + \tfrac{1}{2}\operatorname{var}(X_1 \mid \boldsymbol{Y}) + \tfrac{1}{2}\operatorname{var}(X_2 \mid \boldsymbol{Y})\bigr) \\ &\times \bigl(\exp\bigl(\operatorname{cov}(X_1, X_2 \mid \boldsymbol{Y})\bigr) - 1\bigr).\end{aligned}$$

However, in this example the partial correlation and the covariance do not coincide with the conditional correlation and the covariance, and the zero partial correlation does not necessarily imply the conditional independence. For example, if (X_1, X_2, Y) is normally distributed with covariance matrix $\boldsymbol{\Sigma}$, then

$$\begin{aligned}\sigma_{Z_1 Z_2 \cdot Y} = &\frac{(e^{\sigma_{12}} - 1)(e^{\sigma_{YY}} - 1) - (e^{\sigma_{1Y}} - 1)(e^{\sigma_{2Y}} - 1)}{e^{\sigma_{YY}} - 1} \\ &\times \exp\bigl(\mathrm{E}(X_1 \mid Y) + \mathrm{E}(X_2 \mid Y) + \tfrac{1}{2}\operatorname{var}(X_1 \mid Y) + \tfrac{1}{2}\operatorname{var}(X_2 \mid Y)\bigr),\end{aligned}$$

where σ_{ij} is the (i, j)th element of $\boldsymbol{\Sigma}$. Since $\operatorname{cov}(X_1, X_2 \mid Y) = \sigma_{12} - \sigma_{1Y}\sigma_{2Y}/\sigma_{YY}$, the conditional independence is equivalent to $\sigma_{12} = \sigma_{1Y}\sigma_{2Y}/\sigma_{YY}$, but this does not imply $\sigma_{Z_1 Z_2 \cdot Y} = 0$.

References

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

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

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

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

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

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

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

Lawrance, A.J. (1976). On conditional and partial correlation. *Amer. Statist.* **30**, 146–149.

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

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

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