



**HAL**  
open science

## A note on marginal and conditional independence

Nicola Loperfido

► **To cite this version:**

Nicola Loperfido. A note on marginal and conditional independence. *Statistics and Probability Letters*, 2010, 80 (23-24), pp.1695. 10.1016/j.spl.2010.07.007 . hal-00691791

**HAL Id: hal-00691791**

**<https://hal.science/hal-00691791>**

Submitted on 27 Apr 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Accepted Manuscript

A note on marginal and conditional independence

Nicola Loperfido

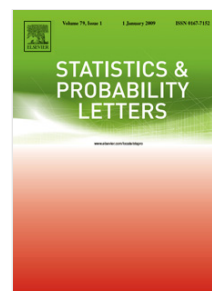
PII: S0167-7152(10)00201-4  
DOI: 10.1016/j.spl.2010.07.007  
Reference: STAPRO 5745

To appear in: *Statistics and Probability Letters*

Received date: 1 April 2010

Revised date: 7 July 2010

Accepted date: 9 July 2010



Please cite this article as: Loperfido, N., A note on marginal and conditional independence. *Statistics and Probability Letters* (2010), doi:10.1016/j.spl.2010.07.007

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# A note on marginal and conditional independence

BY NICOLA LOPERFIDO

*Dipartimento di Economia e Metodi Quantitativi,*

*Università degli Studi di Urbino “Carlo Bo”*

*Via Saffi, 42 – 61029 (PU), Italy*

nicola.loperfido@uniurb.it

## SUMMARY

Some statistical models imply that two random vectors are marginally independent as well as being conditionally independent with respect to another random vector. When the joint distributions of the vectors is normal, Canonical Correlation Analysis may lead to relevant simplifications of the dependence structure. A similar application can be found in elliptical models, where linear independence does not imply statistical independence. Implications for Bayes analysis of the general linear model are discussed.

*Some key words:* Bayes linear analysis; Canonical Correlation Analysis; Elliptical distributions; Sylvester law of nullity; Unrelated parameters.

## 1. INTRODUCTION

Data modelling and interpretation are greatly simplified when independence relations can be assumed or derived. In order to achieve the simplest structure from the least of

constraints, it is necessary to derive all implications of the assumed independence relations. Properties in Dawid (1979) are very useful to this purpose and can be regarded as axioms for a logical system. Unfortunately, no finite set of axioms can completely characterize conditional independence (Studeny, 1992). Other implications of conditional independence relations require more probabilistic structure than that embodied in Dawid's axioms, as it happens when dealing with marginally independent random vectors  $X$  and  $Y$  which are also conditionally independent with respect to a random vector  $Z$  (Dawid, 1998). We shall write  $X \perp Y$  and  $X \perp Y | Z$  to denote the former and the latter independence relation, respectively.

Statistical implications of  $X \perp Y$  and  $X \perp Y | Z$  were first investigated within the framework of three-way tables. Yule (1903) showed that  $X \perp Y$  and  $X \perp Y | Z$  imply either  $(Z, X) \perp Y$  or  $X \perp (Y, Z)$ , when  $X, Y, Z$  are dichotomous random variables. The result was independently rediscovered by Simpson (1951) while considering situations where  $X \perp Y$  may be a reasonable hypothesis while  $X \perp Y | Z$  is irrelevant, and vice versa. Birch (1963) generalized the result to  $r \times s \times 2$  tables, but considered doubtful whether there were any practical situation in which both  $X \perp Y$  and  $X \perp Y | Z$  might be expected to hold. According to Darroch (1962), a relevant example of such a situation was given by perfect tables, where  $X \perp Y$  holds if and only if  $X \perp Y | Z$  does.

Marginal and conditional independence also occur in Bayesian analysis, when the distribution of the sample  $Y$  is indexed by the parameter  $\Theta = (A, \Psi)$ , with a prior distribution on it. Subparameters  $A$  and  $\Psi$  are said to be unrelated (Basu, 1977) if they are independent in the prior distribution ( $A \perp \Psi$ ) as well as in the posterior one ( $A \perp \Psi | Y$ ). Dawid (1980) introduced the strong identifiability condition and showed its usefulness in deriving further independence relations when a statistic  $A$  is marginally ancillary for the

parameter  $A$  (i.e.  $A \perp\!\!\!\perp A$ ), as well as being specific ancillary for it (i.e.  $A \perp\!\!\!\perp A | \mathcal{P}$ ). Basu & Pereira (1983) also assumed strong identifiability, when dealing with the case of a statistic  $T$ , ancillary for  $\Theta$  (i.e.  $T \perp\!\!\!\perp \Theta$ ), and a statistic  $V$ , sufficient for  $\Theta$  (leading to  $T \perp\!\!\!\perp \Theta | V$ ).

Statistical analysis in graphical models sometimes focuses on the conditional distribution of  $Y_S$  given  $Y_C$ , before and after marginalization over  $Y_M$ , where  $Y_S$ ,  $Y_C$  and  $Y_M$  are the random vectors associated to the sets of nodes  $S$ ,  $C$  and  $M$ , respectively. In particular, it may be of interest to know whether independence relations between  $Y_S$  and  $Y_C$ , like  $Y_S \perp\!\!\!\perp Y_C | Y_M$ , also holds after marginalization over  $Y_M$ , i.e.  $Y_S \perp\!\!\!\perp Y_C$ . Cox & Wermuth (1996, section 8.5) deal with the issue using summary graphs and collisionless paths. Wermuth & Cox (2004) state conditions preventing simultaneous occurrence of  $Y_S \perp\!\!\!\perp Y_C | Y_M$  and  $Y_S \perp\!\!\!\perp Y_C$ . Cox & Wermuth (1993) discuss the issues with respect to chordless four-cycles in the covariance and in the concentration matrix.

In econometric theory, independence often models Granger noncausality or one of its variants (Granger, 1969; 1980; 1988). Both  $X \perp\!\!\!\perp Y$  and  $X \perp\!\!\!\perp Y | Z$  might hold when spurious and indirect causal relationships are identified by adding new variables to the information set or removing from it already included variables, an approach proposed by Hsiao (1982) and later developed by Eichler (2009). Unfortunately, independence properties often lead to inferential difficulties, especially within complex, nonparametric models (Dahlhaus & Eichler, 2003). Linear Granger causality avoids the problem by focusing on linear dependencies (Hosoya, 1977; Florens & Mouchart, 1985; Dahlhaus, 2000; Eichler, 2007). As a direct consequence, focus shifts from marginal and conditional independence to ordinary uncorrelation and partial uncorrelation, respectively.

If  $Z=(Z_1, Z_2)$  and  $(X, Z_1)\perp(Y, Z_2)$ , then both  $X\perp Y$ ,  $X\perp Y|Z$  hold, as a trivial application of axioms in Dawid (1979). On the other hand,  $X\perp Y$  and  $X\perp Y|Z$  do not imply that there exists a one-to-one function of  $Z$  into  $(Z_1, Z_2)$  such that  $(X, Z_1)\perp(Y, Z_2)$  holds ( a simple counterexample can be found in Dawid, 1975). A natural question is then whether  $X\perp Y$ ,  $X\perp Y|Z$  imply  $(X, Z_1)\perp(Y, Z_2)$ , at least in relevant statistical models. The paper approaches the problem using correlations and partial correlations, in order to apply well-known results of linear algebra. Independence and conditional independence easily follow when dealing with special families of distributions, i.e. the normal ones. Other kind of independencies are also considered. The paper is structured as follows: § 2 contains the main results; § 3 shows their relations with Canonical Correlation Analysis; § 4 applies them to Bayes analysis of the general linear model; § 5 discusses the results. Proof of Proposition 1 is in the Appendix.

## 2. MAIN RESULTS

The statement “ $Y$  is irrelevant to  $X$  once  $Z$  is given” can be formalized in several ways , other than  $X\perp Y|Z$  (Dawid, 1998), and uncorrelation is a natural choice. We shall write  $X\perp Y$  to denote two uncorrelated random vectors and  $X\perp Y||Z$  to denote that residuals of linear regression of  $X$  and  $Y$  on  $Z$  are uncorrelated (it is implicitly assumed that all necessary moments exist). The following proposition shows an implication of their joint occurrence.

**Proposition 1:** *Let  $Y_a$ ,  $Y_b$  and  $Y_c$  be three random vectors satisfying  $Y_a \perp Y_b$ ,  $Y_a \perp Y_b || Y_c$ .*

*Then  $(Y_a, Y_A) \perp (Y_b, Y_B)$ , where  $(Y_A, Y_B)$  is a one-to-one linear function of  $Y_c$ .*

The relation  $(Y_a, Y_A) \perp (Y_b, Y_B)$  implies  $(Y_a, HY_A) \perp (Y_b, KY_B)$ , where  $H$  and  $K$  are two square, full rank matrices of appropriate dimensions. In particular, if  $\text{cov}(Y_b, Y_c)$  is a null matrix,  $Y_A=Y_c$ ,  $Y_B$  is a constant and  $(Y_a, Y_c) \perp Y_b$ .

In the normal case, (partial) uncorrelation implies (conditional) independence. The following corollary relies on this very convenient property.

**Corollary 1:** *Let  $Y_a$ ,  $Y_b$  and  $Y_c$  be three random vectors whose joint distribution is multivariate normal and such that  $Y_a \perp\!\!\!\perp Y_b$ ,  $Y_a \perp\!\!\!\perp Y_b | Y_c$ . Then  $(Y_a, Y_A) \perp\!\!\!\perp (Y_b, Y_B)$ , where  $(Y_A, Y_B)$  is a one-to-one linear function of  $Y_c$ .*

Unfortunately, Corollary 1 may not hold when the density of the data  $Y=(Y_1, \dots, Y_p)^T$  is elliptical, i.e. it depends on the outcome  $y=(y_1, \dots, y_p)^T$  only through the quadratic form  $(y-\xi)^T \Omega^{-1} (y-\xi)$ ,  $\xi \in \mathfrak{R}^p$ ,  $\Omega = \{\omega_{ij}\}$ ,  $\Omega^{-1} = \{\omega^{ij}\} \in \mathfrak{R}^p \times \mathfrak{R}^p$ . Components of an elliptical, nonnormal vector are always dependent, so that another formalization of the statement “ $Y$  is irrelevant to  $X$  once  $Z$  is given” is called for. We shall write  $X \perp\!\!\!\perp Y$  and  $X \perp\!\!\!\perp Y | Z$  to denote that, almost surely,  $E(X|Y)=E(X)$  and  $E(X|Y,Z)=E(X|Z)$ , respectively (it is implicitly assumed that all necessary moments exist). When  $\omega_{ij}=0$  and first (second) moments exist,  $Y_i \perp\!\!\!\perp Y_j$  ( $Y_i \perp Y_j$ ). Similarly, if  $\omega^{ij}=0$  and first (second) moments exist,  $Y_i \perp\!\!\!\perp Y_j | Y_{ij}$  ( $Y_i \perp Y_j | Y_{ij}$ ), where  $Y_{ij}$  is the vector of all components of  $Y$  but  $Y_i$  and  $Y_j$ . Given constraints on matrices  $\Omega$  and  $\Omega^{-1}$  lead to relevant simplifications of the dependence structure, as shown in the following corollary.

**Corollary 2:** *Let  $Y_a$ ,  $Y_b$  and  $Y_c$  be three random vectors whose joint distribution is elliptical and such that  $Y_a \perp\!\!\!\perp Y_b$ ,  $Y_a \perp\!\!\!\perp Y_b | Y_c$ . Then  $(Y_a, Y_A) \perp\!\!\!\perp (Y_b, Y_B)$ , where  $(Y_A, Y_B)$  is a one-to-one linear function of  $Y_c$ .*

The proof easily follows from the expectation  $E(X|Y)$  being linear in  $Y$ , when the joint distribution of  $X$  and  $Y$  is elliptical and  $E(X)$  exists finite.

### 3. CANONICAL CORRELATION ANALYSIS

Linear transformations leading to  $Y_A$  and  $Y_B$  can be obtained through basic results in Canonical Correlation Analysis. However, unlike Canonical Correlation Analysis, here interest focuses on linear functions minimizing, rather than maximizing, squared correlations. Moreover, standard Canonical Correlation Analysis does not impose any structure on the covariance matrix, whereas here some structure is required in order to satisfy assumptions of Proposition 1. The procedure for obtaining  $Y_A$  and  $Y_B$  can be outlined as follows. Let

$$U_1 = \begin{pmatrix} Y_a \\ Y_A \end{pmatrix} \in \mathfrak{R}^b \quad U_2 = \begin{pmatrix} Y_b \\ Y_B \end{pmatrix} \in \mathfrak{R}^k \quad X_1 = \begin{pmatrix} Y_a \\ Y_c \end{pmatrix} \in \mathfrak{R}^p \quad X_2 = \begin{pmatrix} Y_b \\ Y_c \end{pmatrix} \in \mathfrak{R}^q.$$

Without loss of generality we can assume that  $p \geq q$  and that  $Y_a, Y_b, Y_c$  are linearly independent. Consider now the following matrices

$$\text{var} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \quad R = \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2} = A \begin{pmatrix} D \\ O \end{pmatrix} B^T,$$

where the columns  $a_1, \dots, a_p$  ( $b_1, \dots, b_p$ ) of  $A$  ( $B$ ) are the eigenvectors of  $RR^T$  ( $R^T R$ ). The matrix  $D = \text{diag}(\lambda_1, \dots, \lambda_q)$  is a diagonal matrix with  $\lambda_1 \geq \dots \geq \lambda_q \geq 0$ .

Proposition 1 and basic results in Canonical Correlation Analysis imply existence of two random vectors  $Z = (Z_1, \dots, Z_h)^T$  and  $W = (W_1, \dots, W_h)^T$ , where

$$Z_i = \alpha_i^T \Sigma_{11}^{-1/2} X_1 \quad \alpha_i \in \{a_1, \dots, a_p\}; \quad W_j = \beta_j^T \Sigma_{22}^{-1/2} X_2 \quad \beta_j \in \{b_1, \dots, b_q\}$$

satisfying



$$V(Z_i) = V(W_j) = 1 \quad \text{cov}(Z_i, W_j) = 0 \quad i = 1, \dots, h \quad j = 1, \dots, k.$$

The constraint  $\text{cov}(Z, W) = O_{hk}$  implies that  $Z$  ( $W$ ) is linear one-to-one transformation of  $U_1$  ( $U_2$ ), i.e.  $Z = A_1 U_1$  ( $W = A_2 U_2$ ), so that  $U_1 = A_1^{-1} Z$  ( $U_2 = A_2^{-1} W$ ).

As a numerical example, let  $Y_1, Y_2, Y_3, Y_4$  be four standardized random variables whose variance and concentration matrix are

$$\begin{pmatrix} 1 & 0 & 0.12 & 0.16 \\ 0 & 1 & 0.48 & 0.36 \\ 0.12 & 0.48 & 1 & 0.96 \\ 0.16 & 0.36 & 0.96 & 1 \end{pmatrix} \text{ and } \begin{pmatrix} 1.042 & 0 & 0.447 & -0.595 \\ 0 & 1.563 & -2.679 & 2.009 \\ 0.447 & -2.679 & 17.538 & -15.944 \\ -0.595 & 2.009 & -15.944 & 15.679 \end{pmatrix},$$

respectively. It follows that  $Y_1 \perp Y_2$  and  $Y_1 \perp Y_2 \mid (Y_3, Y_4)$ , but no other independence relation is apparent from the above matrices. Proposition 1 shows that the above covariance structure can be simplified by choosing an appropriate linear transformation of  $(Y_3, Y_4)$ , while the method described in this section leads to  $(Y_1, 3Y_3 + 4Y_4) \perp (Y_2, 4Y_3 + 3Y_4)$ .

#### 4. BAYESIAN LINEAR MODEL

Bayesian analysis of the normal linear model provides a statistical application of results in § 2. Let

$$Y \mid \beta_1, \beta_2 \sim N(X_1 \beta_1 + X_2 \beta_2, \sigma^2 I_n), \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \sim N(m, \sigma^2 V),$$

where  $Y \in \mathfrak{R}^n$  is a vector of observations,  $X_1 \in \mathfrak{R}^n \times \mathfrak{R}^h$  and  $X_2 \in \mathfrak{R}^n \times \mathfrak{R}^k$  are matrices of known coefficients,  $\beta_1 \in \mathfrak{R}^h$  and  $\beta_2 \in \mathfrak{R}^k$  are the parameters (O'Hagan, 1994, § 9). The matrix  $V$ , the vector  $m$  and the scalar  $\sigma$  are assumed to be known (the latter assumption will be relaxed later in this section). Also, let the parameters  $\beta_1$  and  $\beta_2$  be unrelated

(Basu, 1977), i.e.  $\beta_1 \perp\!\!\!\perp \beta_2$  and  $\beta_1 \perp\!\!\!\perp \beta_2 | Y$ . As a direct consequence of Corollary 1, there exists a one-to-one linear function  $(A_1Y, A_2Y)$  of  $Y$  satisfying  $(\beta_1, A_1Y) \perp\!\!\!\perp (\beta_2, A_2Y)$ . Axioms in Dawid (1979) lead to  $Y \perp\!\!\!\perp \beta_2 | (\beta_1, A_2Y)$ ,  $A_1Y \perp\!\!\!\perp \beta_2$  and  $Y \perp\!\!\!\perp \beta_1 | (\beta_2, A_1Y)$  and  $A_2Y \perp\!\!\!\perp \beta_1$ . Equivalently, we can say that  $A_1Y$  and  $A_2Y$  are partially sufficient for  $\beta_1$  and  $\beta_2$ , respectively (Basu, 1977).

In the general case, the parameter  $\sigma$  is assumed to be the realization of a random variable, so that:

$$Y | \beta_1, \beta_2, \sigma \sim N(X_1\beta_1 + X_2\beta_2, \sigma^2 I_n); \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} | \sigma \sim N(m, \sigma^2 V); \quad \sigma \sim F.$$

After marginalization over  $\sigma$ , the joint distribution of  $\beta_1, \beta_2, Y$  is elliptical (as all scale mixtures of centered normal distributions are) and  $\beta_1$  and  $\beta_2$ , are dependent, as well as being dependent conditionally on  $Y$ . If appropriate moments exist, however,  $\beta_1 \perp\!\!\!\perp \beta_2$  and  $\beta_1 \perp\!\!\!\perp \beta_2 | Y$ . Hence Corollary 2 implies that there exists a one-to-one linear function  $(A_1Y, A_2Y)$  of  $Y$  satisfying  $(\beta_1, A_1Y) \perp\!\!\!\perp (\beta_2, A_2Y)$ . The statistics  $A_1Y$  and  $A_2Y$  are no longer partially sufficient. However, there is no way that a linear estimate of  $\beta_1$  ( $\beta_2$ ), based on statistics  $A_1Y$  ( $A_2Y$ ) can be improved by using the full data  $Y$ , under mean squared loss. This property is useful in Bayes linear analysis, where observational data are combined with prior judgements using expectation rather than probability as the primitive expression for the judgements themselves. Moreover, there is also a close similarity between the same property and Bayes linear sufficiency (Goldstein & O'Hagan, 1996),

a weakening of the general notion of probabilistic sufficiency when only second-order prior specifications are made.

## 5. DISCUSSION

Implications of marginal and conditional independence simplify the dependence structure of linear systems with independent components, which include the multivariate normal distribution as a special case. When components are dependent but conditional expectations are linear in the conditioning variables, other simplifications of the dependence structure can be achieved.

Results in the paper rely on Sylvester law of nullity and singular value decomposition, providing another example of tools from linear algebra applied to the statistical analysis of independence. The prominent role of linear algebra in conditional independence theory is well established, to the point that it is not uncommon to devote first sections in papers on graphical models to review of known results in matrix form (Cox & Wermuth, 1993; Wermuth & Cox, 2004). Moreover, linear algebra can benefit from results obtained within a statistical framework, as remarked by Rao (2006). As an example, let the vector spaces  $V_1, V_2, V_3$  and write  $V_1 \perp V_2, V_1 \perp V_2 \parallel V_3$  to denote that that vector spaces  $V_1$  and  $V_2$  are orthogonal to each other and that projections of  $V_1$  and  $V_2$  onto the vector space  $V_3$  are orthogonal to each other, too. Minor modifications to Proposition 1 and its proof can be made in order to derive implications of the orthogonality relations  $V_1 \perp V_2, V_1 \perp V_2 \parallel V_3$ .

## ACKNOWLEDGEMENTS

The author would like to thank Giovanni Marchetti and Elena Stanghellini for the very useful advices about literature on marginal and conditional independence with

application to graphical models. The author would also like to thank an anonymous referee for her/his advices which greatly helped to improve the final manuscript.

#### APPENDIX: PROOF OF PROPOSITION 1

We shall denote with  $n_a$ ,  $n_b$  and  $n_c$  the dimensions of  $Y_a$ ,  $Y_b$  and  $Y_c$ , respectively. Let  $Y$  be the vector obtained by stacking  $Y_a$ ,  $Y_b$  and  $Y_c$  on top of each other. Without loss of generality we can assume that the covariance  $\Sigma$  of  $Y$  is positive definite, so that its inverse  $\Sigma^{-1}$  exists. Moreover, let

$$\text{cov}(Y) = \text{cov} \begin{pmatrix} Y_a \\ Y_b \\ Y_c \end{pmatrix} = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} & \Sigma_{ac} \\ \Sigma_{ba} & \Sigma_{bb} & \Sigma_{bc} \\ \Sigma_{ca} & \Sigma_{cb} & \Sigma_{cc} \end{pmatrix}, \quad \text{con}(Y) = \text{con} \begin{pmatrix} Y_a \\ Y_b \\ Y_c \end{pmatrix} = \begin{pmatrix} \Sigma^{aa} & \Sigma^{ab} & \Sigma^{ac} \\ \Sigma^{ba} & \Sigma^{bb} & \Sigma^{bc} \\ \Sigma^{ca} & \Sigma^{cb} & \Sigma^{cc} \end{pmatrix}.$$

By definition  $\Sigma \Sigma^{-1}$  is an identity matrix, so that

$$\Sigma_{aa} \Sigma^{ab} + \Sigma_{ab} \Sigma^{bb} + \Sigma_{ac} \Sigma^{cb} = O_{ab}$$

where  $O_{ab}$  is a  $n_a \times n_b$  matrix of zeros. Assumptions  $Y_a \perp Y_b$  and  $Y_a \perp Y_b \mid Y_c$  imply that  $\Sigma_{ab}$  and  $\Sigma^{ab}$  are null matrices, so that  $\Sigma_{ac} \Sigma^{cb} = O_{ab}$ . By Sylvester law of nullity, there exists a non-singular  $n_c \times n_c$  matrix  $M$  such that the last  $k$  columns of  $\Sigma_{ac} M$  and the first  $n_c - k$  rows of  $M^{-1} \Sigma^{cb}$  are null vectors, where  $k$  is an integer such that  $0 \leq k \leq n_c$ . Equivalently, there exists a linear transformation  $MY_c = (Y_{\bullet}^T, Y_{*}^T)^T$  such that

$$\text{cov} \begin{pmatrix} Y_a \\ Y_b \\ Y_{\bullet} \\ Y_{*} \end{pmatrix} = \begin{pmatrix} \Sigma_{aa} & O_{ab} & \Sigma_{a\bullet} & O_{a*} \\ O_{ba} & \Sigma_{bb} & \Sigma_{b\bullet} & \Sigma_{b*} \\ \Sigma_{\bullet a} & \Sigma_{\bullet b} & \Sigma_{\bullet\bullet} & \Sigma_{\bullet*} \\ O_{*a} & \Sigma_{*b} & \Sigma_{*\bullet} & \Sigma_{**} \end{pmatrix}, \quad \text{con} \begin{pmatrix} Y_a \\ Y_b \\ Y_{\bullet} \\ Y_{*} \end{pmatrix} = \begin{pmatrix} \Sigma^{aa} & O^{ab} & \Sigma^{a\bullet} & \Sigma^{a*} \\ O^{ba} & \Sigma^{bb} & O^{b\bullet} & \Sigma^{b*} \\ \Sigma^{\bullet a} & O^{\bullet b} & \Sigma^{\bullet\bullet} & \Sigma^{\bullet*} \\ \Sigma^{*a} & \Sigma^{*b} & \Sigma^{*\bullet} & \Sigma^{**} \end{pmatrix}.$$

Consider now the linear transformation

$$\begin{pmatrix} Y_A \\ Y_B \end{pmatrix} = \begin{pmatrix} I_{\cdot} & -\Sigma_{\cdot*} \Sigma_{**}^{-1} \\ O_{\cdot*} & I_{*} \end{pmatrix} \begin{pmatrix} y_{\cdot} \\ y_{*} \end{pmatrix}.$$

The covariance and concentration matrices of  $Y_a, Y_b, Y_A$  and  $Y_B$  are

$$\text{cov} \begin{pmatrix} Y_a \\ Y_b \\ Y_A \\ Y_B \end{pmatrix} = \begin{pmatrix} \Sigma_{aa} & O_{ab} & \Sigma_{aA} & O_{aB} \\ O_{ba} & \Sigma_{bb} & \Sigma_{bA} & \Sigma_{bB} \\ \Sigma_{Aa} & \Sigma_{Ab} & \Sigma_{AA} & O_{AB} \\ O_{Ba} & \Sigma_{Bb} & O_{BA} & \Sigma_{BB} \end{pmatrix}, \quad \text{con} \begin{pmatrix} Y_a \\ Y_b \\ Y_A \\ Y_B \end{pmatrix} = \begin{pmatrix} \Sigma^{aa} & O^{ab} & \Sigma^{aA} & \Sigma^{aB} \\ O^{ba} & \Sigma^{bb} & O^{bA} & \Sigma^{bB} \\ \Sigma^{Aa} & O^{Ab} & \Sigma^{AA} & \Sigma^{AB} \\ \Sigma^{Ba} & \Sigma^{Bb} & \Sigma^{BA} & \Sigma^{BB} \end{pmatrix}.$$

Since the two matrices are inverses, the following equality holds:

$$\Sigma_{Aa} O^{ab} + \Sigma_{Ab} \Sigma^{bb} + \Sigma_{AA} O^{Ab} + O_{AB} \Sigma^{Bb} = O_{ab}$$

By assumption  $\Sigma$  is a positive definite matrix, so that  $\Sigma^{bb}$  is a positive definite matrix and  $\Sigma_{Ab}$  is a null matrix. By rearranging elements in the above vectors and matrices we obtain

$$\text{cov} \begin{pmatrix} Y_a \\ Y_A \\ Y_b \\ Y_B \end{pmatrix} = \begin{pmatrix} \Sigma_{aa} & \Sigma_{aA} & O_{ab} & O_{aB} \\ \Sigma_{Aa} & \Sigma_{AA} & O_{Ab} & O_{AB} \\ O_{ba} & O_{bA} & \Sigma_{bb} & \Sigma_{bB} \\ O_{Ba} & O_{BA} & \Sigma_{Bb} & \Sigma_{BB} \end{pmatrix}$$

Hence  $(Y_a, Y_A) \perp (Y_b, Y_B)$ . □

## REFERENCES

- BASU, D. (1977). On the Elimination of Nuisance Parameters. *J. Amer Statist. Assoc.* **72**, 355-366.
- BASU, D. & PEREIRA, C. A. B. (1983). Conditional Independence in Statistics. *Sankhya Ser. A* **45**, 324-337.

- BIRCH, M. W. (1963). Maximum likelihood in three-way contingency tables. *J. Roy. Statist. Soc. Ser. B* **25**, 220-233.
- COX, D. R. & WERMUTH, N. (1993). Linear dependencies represented by Chain Graphs. *Statist. Science* **8**, 204-218.
- COX, D. R. & WERMUTH, N. (1996). *Multivariate Dependencies: Models, Analysis and Interpretation*. Chapman & Hall, London, UK.
- DAHLHAUS, R. (2000). Graphical interaction models for multivariate time series. *Metrika* **51**, 157-172.
- DAHLHAUS, R & EICHLER, M. (2003). Causality and graphical models for time series. In: P. Green, N. Hjort and S. Richardson (eds.), *Highly structured stochastic systems*. University Press, Oxford, 115-137.
- DARROCH, J. N. (1962). Interactions in multi-factor contingency tables. *J. Roy. Statist. Soc. Ser. B* **24**, 251-263.
- DAWID, A. P. (1975). On the concepts of sufficiency and ancillarity in the presence of nuisance parameters. *J. Roy. Statist. Soc. Ser. B* **37**, 248-258.
- DAWID, A. P. (1979). Conditional independence in statistical theory. *J. Roy. Statist. Soc. Ser. B* **41**, 1-31.
- DAWID, A. P. (1980). Conditional independence for statistical operations. *Ann. Statist.*, **8**, 598-617.
- DAWID, A. P. (1998). Conditional independence. In: *Encyclopedia of Statistical Sciences*, Update Volume 2, edited by S. Kotz, C. B. Read & D. L. Banks. Wiley-Interscience, 146-155.
- EICHLER, M. (2007). Granger-causality and path diagrams for multivariate time series. *J. Econometr.* **137**, 334-353.

- EICHLER, M. (2009). Causal inference from multivariate time series: What can be learned from Granger causality. In: C. Glymour, W. Wang, D. Westerthal (eds), *Proceedings of the 13<sup>th</sup> International Congress of Logic, Methodology and Philosophy of Science*, King's College Publications, London.
- FLORENS, J.P. & MOUCHART, M. (1985). A linear theory for noncausality. *Econometrica* **53**, 157-175.
- GOLDSTEIN, M. & O'HAGAN, A. (1996). Bayes linear sufficiency and systems of expert posterior assessments. *J. R. statist. Soc.* **B 58**, 301-316.
- GRANGER, C.W.J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**, 424-438.
- GRANGER, C.W.J. (1980). Testing for causality, a personal viewpoint. *J. Econ. Dyn. Control* **2**, 329-352.
- GRANGER, C.W.J. (1988). Some recent developments in a concept of causality. *J. Econometr.* **39**, 199-211.
- O'HAGAN, A. (1994). *Kendall's Advanced Theory of Statistics, Volume 2B: Bayesian Inference*. Wiley, New York.
- HSIAO, C. (1982). Autoregressive modelling and causal ordering of econometric variables. *J. Econ. Dyn. Control* **4**, 243-259.
- HOSOYA, Y. (1977). On the Granger condition for non-causality. *Econometrica* **45**, 1735-1736.
- RAO, C. R. (2006). Statistical proofs of some matrix theorems. *International Statistical Review* **74**, 169-185.
- SIMPSON, E. H. (1951). The interpretation of interaction in contingency tables. *J. Roy Statist. Soc. Ser. B* **13**, 238-241.

STUDENY, M. (1992). Conditional independence relations have no finite complete characterization. *Transactions of the Eleventh Prague Conference on Information Theory, Statistical Decision Functions and Random Processes B*, 377-396. Academia: Prague.

YULE, G. U. (1903). Notes on the theory of association of attributes in Statistics. *Biometrika* **2**, 121-134.

WERMUTH, N. & COX, D. R. (2004). Joint response graphs and separation induced by triangular systems. *J. Roy. Statist. Soc. Ser. B* **66**, 687-717.