**I N S T I T U T  D E**
**S T A T I S T I Q U E**

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

**D I S C U S S I O N**
**P A P E R**

**0511**

# IGNORABLE COMMON INFORMATION, NULL SETS AND BASUS FIRST THEOREM

SAN MARTIN, E., MOUCHART, M. and J.-M. ROLIN

http://www.stat.ucl.ac.be

# Ignorable Common Information, Null Sets and Basu's First Theorem

ERNESTO SAN MARTÍN[a], MICHEL MOUCHART[b] AND JEAN-MARIE ROLIN[b]
[a]*Departamento de Estadística, Pontificia Universidad Católica de Chile, Chile.*
[b]*Institut de statistique, Université catholique de Louvain, Belgium.*

April 11, 2005

**Abstract**

This paper deals with the Intersection Property, or Basu's First Theorem, which is valid under a condition of no common information, also known as measurable separability. After formalizing this notion, the paper reviews general properties and give operational characterizations in two topical cases: the finite one and the multivariate normal one. The paper concludes discussing the relevance of these characterizations for different fields as graphical models, zero entries in contingency tables, causal analysis and estimability in Markov processes.

*Keywords:* common information, conditional independence, forbidden states, measurable separability, structural zeros.

*AMS classification:* 62B99.

## 1 Introduction

Conditional independence is presently accepted as a fundamental concept not only in the theory of statistical inference (see, *e.g.*, Dawid, 1979a; Florens *et al.*, 1990; or Nogales *et al.*, 2000), but also in statistical modelling, particularly in structural modelling (see, *e.g.*, Novick, 1979; Speed and Kiiveri, 1986; Lauritzen and Wermuth, 1989; Pearl, 1995; and Mouchart and San Martín, 2003).

The use of graphical models to represent dependence relations among random variables, and therefore to represent conditional independence, has became very useful to model building since most dependencies and associations between variables can be *visualized* through graph representations. The key idea behind these specification schemes is to utilize the correspondence between *separation* in graphs and *conditional independence* in probability. A graphical representation is used to represent qualitative multivariate relationships, specify and visualize multivariate statistical models, determine statistical properties

of multivariate models, and develop computationally efficient algorithms for dealing with large multivariate models; for textbook expositions, see Whitakker (1990), Cox and Wermuth (1996) and Lauritzen (1998).

An aspect widely developed in the graphical literature consists in relating the properties of the conditional independence with algebraic structures satisfied by graph relationships. Thus, an operational link is established between conditional independence and graph representations in the sense that conditions obtained after manipulations with graphs can be translated in terms of conditional independence, and conversely; for details, see, *e.g.*, Pearl (1988), Geiger *et al.*(1988), Studený (1997) and Studený and Bouckaert (1998).

This mutual fertilization works when universally valid properties of the conditional independence are used; these ones can be found in, *e.g.*, Martin *et al.*(1973), Dawid (1979a), Döhler (1980) and Mouchart and Rolin (1984). Nevertheless, some specific problems in graphical models, or even some substantive considerations in models building (for instance, structural zeros in finite models), require to restrict the class of underlying probability distributions in order to obtain the desirable graphical property. To be more specific, and to introduce the problem analyzed in this paper, consider the following property, typically called *Intersection Property*:

$$\text{(i) } X_1 \perp\!\!\!\perp X_2 \mid X_3 \quad \text{and} \quad \text{(ii) } X_1 \perp\!\!\!\perp X_3 \mid X_2 \implies \text{(iii) } X_1 \perp\!\!\!\perp (X_2, X_3). \tag{1.1}$$

where $X_1$, $X_2$ and $X_3$ are random variables defined on a same probability space $(\Omega, \mathcal{F}, P)$. This condition is widely used in the graphical literature; see, among others, Frydenberg (1990, condition CI5), Spohn (1980, section 2; 1994, Definition 3), Pearl and Paz (1987, section 4), Cox and Wermuth (1993, section 2), Geiger and Pearl (1993, condition (7)), Kauermann (1996, section 2), Andersson, Madigan and Perlman (1997, p. 87; 2001, p. 45), Koster (1996, section 3; 1999, section 3) and Studený and Bouckaert (1998, p. 1438). It is, for instance, used to establish the equivalence between pairwise, local and global Markov properties for undirected graphs; for definitions and details, see Pearl and Paz (1987) and Frydenberg (1990).

Since Basu (1955, 1958), it is well known that the Intersection Property (1.1) does not hold universally, but only under additional conditions –essentially that there be no common information between $X_2$ and $X_3$. However, the implication is true under a stronger condition. Thus, for instance, when $\Omega$ is a finite set, Spohn (1994, Theorem 4) requires that $P$ be strictly positive in the sense that $P(A) = 0$ only for $A = \emptyset$. When $(X_1, X_2, X_3)$ is normally distributed, Cox and Wermuth (1993) require that the covariance matrix be definite positive. More in general, it is often required that $P$ has a positive joint probability density with respect to some product measure on $\Omega$; see, *e.g.*, Frydenberg (1990), Kauerman (1996), Anderson *et al.*(1997, p. 87) and Lauritzen (1998, Proposition 3.1). Nevertheless, as Andersson *et al.*(1997, Remark 3.3) pointed out, the strict positivity of the density of $P$ (w.r.t. some product measure on $\Omega$) is not a necessary condition under which the Intersection Property (1.1) is valid; and Hill (1993) asserts that "this positivity condition limits the possible applications [...] In particular, the theorem cannot be applied to Bayesian networks with functional constraints (Lauritzen and Spiegelhalter, 1988)

or to contingency tables with structural zeros or to statistical mechanics systems with forbidden states (Moussouris, 1974)" (p. 259).

Taking into account these considerations, the problem consists in looking for conditions much weaker than the positivity of the density of $P$ under which the Intersection Property (1.1) is valid. This is precisely the content of this paper. More specifically, in this paper we formalize the concept of "no common information", also known as "measurable separability", so as to provide a sufficient assumption to make the Intersection Property (1.1) valid. Next we closely examine the condition of no common information and provide equivalent characterizations in two particular cases, namely the cases of discrete random vector and of normally distributed random vector. We choose these two cases because they are the underlying structure of most of the graphical representations of conditional independence; see, *e.g.*, Spohn (1994, pp. 174s) for the first case, and Cox and Wermuth (1993) for the second case. In the finite case, we prove that the condition of no common information between $X_2$ and $X_3$ is equivalent to a condition restricting, but not excluding, the exact position of the null sets (or, sets of zero probability) in the matrix which represents the joint distribution of $(X_2, X_3)$. In the normal case, we prove that the no common information corresponds to an equality between the ranks of the covariance matrices of $X_2$ and of $(X_2 \mid X_3)$, respectively.

The problem addressed in this paper, as well as its contribution, are not only related with graphical models, but also with other fields such as Markov chains, causal inference and Basu's First Theorem in a Bayesian set-up. Let us also mention that Vantaggi (2001, 2002) establishes the Intersection Property under alternative definitions of stochastic conditional independence motivated by the De Finetti's (1949, 1970) critique of Kolmogorov axioms. The results obtained in this paper can, therefore, be considered as its counter-part in a purely Kolmogorovian set-up.

This paper is organized as follows. Section 2 introduces a formal definition of the concept of no common information. Thereafter, operational characterizations are discussed. This section ends with a review of results relevant to the problem considered in the present paper. The main results of this paper are contained in Sections 3 and 4. We complete the paper with some concluding remarks. The proofs of the main results are gathered in the Appendix.

# 2 A Formalization of the Concept of No Common Information

## 2.1 Definition

Let $X_1$ and $X_2$ be two random variables defined on a common probability space $(\Omega, \mathcal{F}, P)$ valued in $(N_1, \mathcal{N}_1)$ and $(N_2, \mathcal{N}_2)$, respectively. The information provided by the random variables $X_i$ may be represented by the generated $\sigma$-field $\mathcal{X}_i = X_i^{-1}(\mathcal{N}_i) = \{X_i^{-1}(B) \,:\, B \in \mathcal{N}_i\} \subset \mathcal{F}$, often denoted as $\sigma(X_i)$. As a matter of fact, $\mathcal{X}_i$ heuristically corresponds to the set of events that may be described in terms of that random variable (Florens and Mouchart, 1982, p. 588). The information thus defined does

not depend on the coordinate system chosen to represent the corresponding random variable because $\sigma(X_i) = \sigma[h(X_i)]$ for all injective function $h$.

As we do not want to distinguish two $P$-a.s. equal events, we rather consider as the relevant information the completed $\sigma$-fields $\overline{\mathcal{X}_i} = \mathcal{X}_i \vee \overline{\mathcal{F}_0}$, where $\overline{\mathcal{F}_0}$ is the completed trivial $\sigma$-field, namely $\overline{\mathcal{F}_0} = \{A \in \mathcal{F} : P(A) \in \{0,1\}\}$ (where $\mathcal{A}_1 \vee \mathcal{A}_2$ is the smallest $\sigma$-field containing $\mathcal{A}_1 \cup \mathcal{A}_2$). We use the measurable completion rather than the Lebesgue completion not only to avoid loosing the countability generated character of completed $\sigma$-fields (this condition might be viewed as a "technicality"), but also to avoid introducing events *not* generated by the random variables (this condition is directly related with our concern, namely the information provided by a random variable). Note that the completed trivial $\sigma$-field $\overline{\mathcal{F}_0}$ is the same for equivalent probability measures (*i.e.*. probability measures having the same null sets as $P$).

The common information provided by $X_1$ and $X_2$ can be accordingly described as $\overline{\mathcal{X}_1} \cap \overline{\mathcal{X}_2}$. Therefore, $X_1$ *and* $X_2$ *don't share common information* if and only if

$$\overline{\mathcal{X}_1} \cap \overline{\mathcal{X}_2} = \overline{\mathcal{F}_0}, \tag{2.1}$$

and we denote this property as $X_1 \parallel X_2$. When (2.1) holds, we also say that $X_1$ *and* $X_2$ *are measurably separated*; see Florens *et al.*(1990, section 5.2).

Let $X_3$ be a random variable from $(\Omega, \mathcal{F}, P)$ to $(N_3, \mathcal{N}_3)$. The previous concept can be extended to the case of no common information between $X_1$ and $X_2$ conditionally on $X_3$, as follows:

$$\overline{\mathcal{X}_1 \vee \mathcal{X}_3} \cap \overline{\mathcal{X}_2 \vee \mathcal{X}_3} = \overline{\mathcal{X}_3}. \tag{2.2}$$

We denote this property as $X_1 \parallel X_2 \mid X_3$. When (2.2) holds, we also say that $X_1$ *and* $X_2$ *are measurably separated conditionally on* $X_3$. Clearly condition (2.2) reduces to condition (2.1) when $\overline{\mathcal{X}_3} = \overline{\mathcal{F}_0}$. If we want to make explicit the role of the probability $P$ in this concept, we write $X_1 \parallel X_2 \mid X_3;\, P$.

## 2.2 Equivalent characterizations

Heuristically, the concept of measurable separability, or no common information, means that the information common to $X_1$ and $X_2$ is either trivial (formulation (2.1)) or "already known" through $X_3$ (formulation (2.2)). A deeper understanding of the concept may be obtained by considering equivalent conditions. This is the objective of next theorem:

**Theorem 2.1** *Let $X_i$ ($i = 1, 2, 3$) be random variables defined on a fixed probability space $(\Omega, \mathcal{F}, P)$ and valued on the measurable spaces $(N_i, \mathcal{N}_i)$. The following conditions are equivalent:*

(i) *$X_1$ and $X_2$ are measurably separated conditionally on $X_3$.*

(ii) *If $f(X_1, X_3) = g(X_2, X_3)$ a.s. for some $f$, bounded Borel function defined on $(N_1 \times N_3, \mathcal{N}_1 \otimes \mathcal{N}_3)$, and some $g$, bounded Borel function defined on $(N_2 \times N_3, \mathcal{N}_2 \otimes \mathcal{N}_3)$, then $f(X_1, X_3) = h(X_3)$ a.s. for some $h$, bounded Borel function defined on $(N_3, \mathcal{N}_3)$.*

(iii) *If $V[f(X_1, X_3) \mid X_2, X_3] = 0$ a.s. for some $f$, bounded Borel function defined on $(N_1 \times N_3, \mathcal{N}_1 \otimes \mathcal{N}_3)$, then $V[f(X_1, X_3) \mid X_3] = 0$ a.s.*

The equivalence between statements 2 and 3 in Theorem 2.1 is straightforward. The equivalence between statements 1 and 3 follows from the following relationship:

$$A \in (\mathcal{X}_1 \vee \mathcal{X}_3) \cap \overline{(\mathcal{X}_2 \vee \mathcal{X}_3)} \Longleftrightarrow A \in (\mathcal{X}_1 \vee \mathcal{X}_3) \text{ and } E(\mathbb{1}_A \mid \mathcal{X}_2 \vee \mathcal{X}_3) = \mathbb{1}_A \quad \text{a.s.}$$

For additional details, see Florens *et al.*(1990, section 5.2). From condition (2.2), it should be clear that the concept of measurable separability is symmetric between $X_1$ and $X_2$. Thus, assertions 2 and 3 in Theorem 2.1 may also be symmetrized by permuting the indexes 1 and 2.

The property of measurable separability is meant to exclude joint distributions with a support such as that one depicted in Figures 1 and 2. Indeed, in such cases $X_1$ and $X_2$ are not measurably separated because the event $\{X_1 \in A_1\}$ is a.s. equal to the event $\{X_2 \in B_1\}$: these events represent a non-trivial information common to $X_1$ and $X_2$.

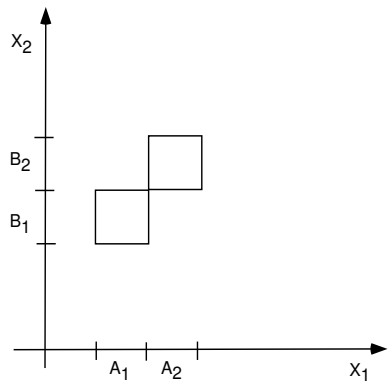


Figure 1: Counter-example to measurable separability

It should be clear from Theorem 2.1 and from these remarks that measurable separability depends on the probability $P$ through its null sets only. Thus, if $P$ and $P'$ are equivalent probabilities, then $X_1 \parallel X_2 \mid X_3;\ P \Longleftrightarrow X_1 \parallel X_2 \mid X_3;\ P'$.

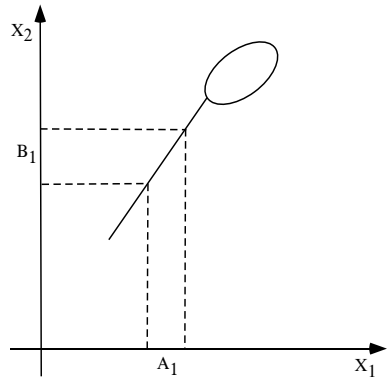


Figure 2: Counter-example to measurable separability

## 2.3 Measurable separability and conditional independence

Two relationships between measurable separability and conditional independence are relevant for our discussion about the Intersection Property (1.1). The first one tells us that measurable separability is a (much) weaker property than conditional independence. More precisely,

**Proposition 2.1** *If* $X_1 \perp\!\!\!\perp X_2 \mid X_3$*, then* $X_1 \parallel X_2 \mid X_3$.

For a proof, see Florens *et al.*(1990, Theorem 5.2.7). Thus, a sufficient, but not necessary, condition for measurable separability is that the joint density of $X_2$ and $X_3$ be equivalent to a distribution making $X_2$ and $X_3$ independent or that the support of the joint distribution be a rectangle.

A second relevant property is contained in the following proposition:

**Proposition 2.2** *The following properties are equivalent:*

(i) $X_1 \perp\!\!\!\perp X_2 \mid X_3$ *and* $X_1 \perp\!\!\!\perp X_3 \mid X_2$.

(ii) $X_1 \perp\!\!\!\perp (X_2, X_3) \mid M$ *where* $M$ *represent the information common to* $X_2$ *and* $X_3$*, namely* $\overline{\mathcal{X}_2} \cap \overline{\mathcal{X}_3}$.

For a proof, see Dawid (1980, Theorem 7.1), Mouchart and Rolin (1984, Corollary 3.6) or Florens *et al.*(1990, Corollary 2.2.13). This proposition provides us with a condition under which the Intersection Property (1.1) is true. As a matter of fact, if in Proposition 2.2, the information common to $X_1$ and $X_2$ reduces to the completed trivial $\sigma$-field $\overline{\mathcal{F}_0}$, then implication (1.1) becomes true. This is the content of the following theorem:

**Theorem 2.2** *Under the condition of no common information between $X_2$ and $X_3$, namely $\overline{\mathcal{X}_2} \cap \overline{\mathcal{X}_3} = \overline{\mathcal{F}_0}$, the Intersection Property (1.1) is true.*

As mentioned in the introduction, the literature of graphical models often use a condition of strict positivity of the joint density of $(X_2, X_3)$. This condition assume that the joint probability distribution is dominated by the Lebesgue measure (on $\mathbb{R}^2$), implying that the interior of the support is not empty and excluding situations with mixed distributions composed of a discrete component and a continuous one, as illustrated in Figure 2. Under this restriction, the condition of positive density does indeed imply measurable separability, but is actually much stronger and not necessary, as will be shown in next section.

Theorem 2.2 can be extended to a conditional version. More precisely,

**Theorem 2.3** *If $X_1 \perp\!\!\!\perp X_2 \mid (X_4, X_3)$ and $X_1 \perp\!\!\!\perp X_4 \mid (X_2, X_3)$, then $X_1 \perp\!\!\!\perp (X_2, X_4) \mid X_3$ provided that $X_2 \parallel X_4 \mid X_3$.*

For a proof, see Florens *et al.*(1990, Theorem 5.2.10).

**Remark 1** As pointed out in the introductory section, Theorem 2.2 corresponds, in a Bayesian set-up, to the First Basu's Theorem as correctly established in Basu (1958). In a sampling-theory framework, Koehn and Thomas (1975) have proved Basu's (1958) result under a condition of the non-existence of a splitting set. In a Bayesian framework, if the prior distribution $\mu$ is such that the predictive distribution dominates all the sampling probabilities $\{P^\theta : \theta \in \Theta\}$, then measurable separability implies the non-existence of splitting sets. If furthermore the prior distribution $\mu$ is such that $P^\theta(A) \in \{0, 1\}$ $\mu$-a.s. implies $P^\theta(A) \in \{0, 1\}$ for all $\theta \in \Theta$, then measurable separability is equivalent to the non-existence of splitting sets; for details, see Florens *et al.*(1990, section 5.3.3).

The main conclusion of this section is that in order to establish the Intersection Property (1.1) in particular cases, it is necessary to characterize the condition $\overline{\mathcal{X}_2} \cap \overline{\mathcal{X}_3} = \overline{\mathcal{F}_0}$ in more operational terms. This is precisely the content of the next two sections.

# 3 Measurable Separability in the Finite Case

## 3.1 Common information in the finite case: An example

Before characterizing measurable separability in the finite case, let us introduce an example which shows that such a condition is necessary to establish implication (1.1). Consider so a finite distribution defined on $\{0, 1\} \times \{1, 2, 3\} \times \{1, 2, 3\}$ with a support containing 8 points only, defined as follows:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $X_1 = 0$<br>$X_2 = 1$<br>$X_3 = 1$ | $\alpha q_1$ | $X_1 = 0$<br>$X_2 = 1$<br>$X_3 = 2$ | $\alpha q_2$ | $X_1 = 0$<br>$X_2 = 2$<br>$X_3 = 3$ | $\beta q_3$ | $X_1 = 0$<br>$X_2 = 3$<br>$X_3 = 3$ | $\beta q_4$ |
| $X_1 = 1$<br>$X_2 = 1$<br>$X_3 = 1$ | $(1-\alpha)q_1$ | $X_1 = 1$<br>$X_2 = 1$<br>$X_3 = 2$ | $(1-\alpha)q_2$ | $X_1 = 1$<br>$X_2 = 2$<br>$X_3 = 3$ | $(1-\beta)q_3$ | $X_1 = 1$<br>$X_2 = 3$<br>$X_3 = 3$ | $(1-\beta)q_4$ |

where $q_1 q_2 q_3 q_4 \alpha(1-\alpha)\beta(1-\beta)(\alpha-\beta) > 0$ and $q_1 + q_2 + q_3 + q_4 = 1$.

It is easily checked, by direct computations, that $X_1 \perp\!\!\!\perp X_2 \mid X_3$ and, by symmetry between $X_2$ and $X_3$, that $X_1 \perp\!\!\!\perp X_3 \mid X_2$. Nevertheless, $X_1 \perp\!\!\!\perp (X_2, X_3)$ is false (except in the excluded case $q_1 = 1$). Consequently, implication (1.1) does not hold although the probability distribution of $(X_1, X_2, X_3)$ satisfies conditions (i) and (ii) of the Intersection Property.

This example provides a key for an easy understanding of the concept of measurable separability in the finite case. As a matter of fact, the joint distribution of $(X_2, X_3)$ is given by

| | $X_3 = 1$ | $X_3 = 2$ | $X_3 = 3$ |
|---|---|---|---|
| $X_2 = 1$ | $q_1$ | $q_2$ | $0$ |
| $X_2 = 2$ | $0$ | $0$ | $q_3$ |
| $X_2 = 3$ | $0$ | $0$ | $q_4$ |

(3.1)

So, the support of $(X_2, X_3)$ has 4 points which satisfy the following relationship, as Figure 3 shows:

$$\{X_2 = 1\} = \{X_3 \neq 3\} \qquad \text{a.s.} \tag{3.2}$$

Condition (3.2) represents an information common to $X_2$ and $X_3$ (*i.e.*, the event $\{X_2 = 1\}$ is the *same* as the event $\{X_3 \neq 3\}$ for the joint probability distribution). Thus, "no common information" between $X_2$ and $X_3$ can be expressed saying that if there exists two functions $f$ and $g$ such that $f(X_1) = g(X_2)$ a.s. for the joint probability, then there exists a constant $c$ such that $f(X_1) = c$ a.s.; see Theorem 2.1, statement (ii). In other words, "no common information" means that the only common information is the trivial one, *i.e.*, the class of measurable null sets.

## 3.2 Characterization of measurable separability in the finite case

To characterize measurable separability in the finite case, let $N_r$ (with $r = 2, 3, 4$) be finite sets and $X_r : \Omega \longrightarrow N_r$ be random variables under the not restrictive condition that $P[X_r = i] > 0$ for all $i \in N_r$. We define

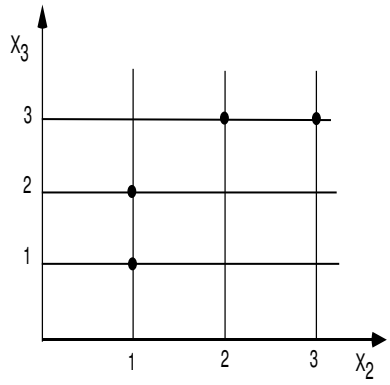


Figure 3: Support of $(X_2, X_3)$

$$N_2^{(k)} \;=\; \{i \in N_2 \;:\; P[X_2 \,=\, i \mid X_4 \,=\, k] \,>\, 0\} \quad \text{for } k \in N_4,$$

$$N_3^{(k)} \;=\; \{j \in N_3 \;:\; P[X_3 \,=\, j \mid X_4 \,=\, k] \,>\, 0\} \quad \text{for } k \in N_4.$$

For $k \in N_4$, define the $|N_2^{(k)}| \times |N_3^{(k)}|$ matrix $\mathbf{P}^{(k)}$ by

$$p_{ij|k} \;\equiv\; (\mathbf{P}^{(k)})_{ij} \;=\; P[X_2 \,=\, i, X_3 \,=\, j \mid X_4 \,=\, k] \quad \text{for } (i,j) \in N_2^{(k)} \times N_3^{(k)}.$$

Finally, for $k \in N_4$, let

$$N_{3i}^{(k)} \;=\; \{j \in N_3^{(k)} \;:\; P[X_2 \,=\, i, X_3 \,=\, j \mid X_4 \,=\, k] \,>\, 0\} \quad \text{for } i \in N_2^{(k)}.$$

The following theorem characterizes the measurable separability in the finite case:

**Theorem 3.1** *The following statements are equivalent:*

(i) $X_2 \parallel X_3 \mid X_4$*;*

(ii) $(\forall\, k \in N_4)\; (\forall\, I \subset N_2^{(k)})$ *with* $I \neq \emptyset$ *and* $I \neq N_2^{(k)}$*:*

$$\left( \bigcup_{i \in I} N_{3i}^{(k)} \right) \;\cap\; \left( \bigcup_{i' \in N_2^{(k)} \setminus I} N_{3i'}^{(k)} \right) \;\neq\; \emptyset;$$

(iii) $(\forall\, k \in N_4)\ (\forall\, I \subset N_2^{(k)})$ *with* $I \neq \emptyset$ *and* $I \neq N_2^{(k)}$ $[\exists\, (i, i', j) \in I \times (N_2^{(k)} \setminus I) \times N_3^{(k)}]$ *such that*

$$p_{ij|k} \cdot p_{i'j|k} > 0,$$

*where* $A \setminus B$ *denotes the difference between sets* $A$ *and* $B$*.*

For a proof, see Appendix A.

**Remark 2** Since the measurable separability condition $X_2 \parallel X_3 \mid X_4$ is symmetric in $X_2$ and $X_3$, one could formally add conditions to Theorem 3.1, which would be obtained by interchanging $(I, i, 2)$ with $(J, j, 3)$.

Next corollary makes explicit the particular case where $X_4$ is a constant random variable (equivalently, $\overline{\mathcal{X}_4} = \overline{\mathcal{F}_0}$:

**Corollary 3.1** *The following statements are equivalent:*

(i) $X_2 \parallel X_3$*;*

(ii) $(\forall\, I \subset N_2)$ *with* $I \neq \emptyset$ *and* $I \neq N_2$*, it follows that*

$$\left(\bigcup_{i \in I} N_{3i}\right) \cap \left(\bigcup_{i' \in N_2 \setminus I} N_{3i'}\right) \neq \emptyset;$$

(iii) $(\forall\, I \subset N_2)$ *with* $I \neq \emptyset$ *and* $I \neq N_2$ $[\exists\, (i, i', j) \in I \times (N_2 \setminus I) \times N_3]$ *such that*

$$p_{ij} \cdot p_{i'j} > 0.$$

In Corollary 3.1, the sets $N_2$ and $N_3$ only contain points of positive probability. As mentioned before, this is not a restrictive assumption since if there exists $i_0 \in N_2$ such that $P[X_2 = i_0] = 0$, then $P[X_2 = i_0, X_3 = j] = 0$ for all $j \in N_3$. So, $N_{3i_0} = \emptyset$ and, consequently, the corresponding column in the joint probability distribution of $(X_2, X_3)$ can be eliminated.

Considering condition (ii) of Corollary 3.1, it can be noticed that the measurable separability between $X_2$ and $X_3$ not only depends on each marginal distribution of $X_2$ and $X_3$ through the sets $N_2$ and $N_3$, but also on the joint distribution of $(X_2, X_3)$ through the sets $N_{3i}$ for each $i \in N_2$. Moreover, $N_{3i} \subsetneq N_3$ for some (possible all) $i \in N_2$. Therefore, the case where there exists $(i, j) \in N_2 \times N_3$ such that

$P[X_2 = i, X_3 = j] = 0$ is not excluded. In other words, condition (ii) of Corollary 3.1 tells us in what position must be the non-zeros (and so the zeros) probabilities for the joint distribution of $(X_2, X_3)$: for each $(i, i') \in I \times (N_2 \setminus I)$, there exists at least one column $j \in N_3$ such that $p_{ij} \cdot p_{i'j} > 0$, as illustrated in Figure 4.

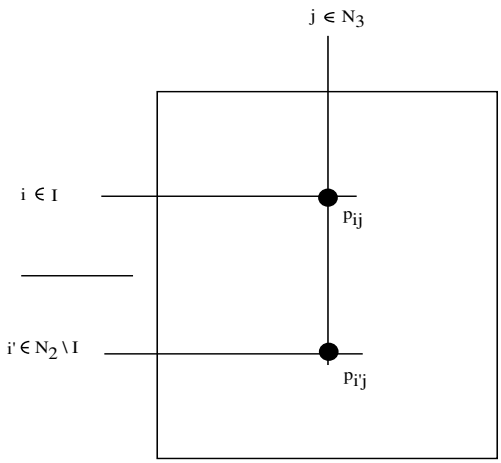


Figure 4: Condition (iii) of Corollary 3.1

**Example 1** As an example, consider $X_2 \in \{1, \ldots, 5\}$ and $X_3 \in \{1, \ldots, 4\}$ and the following joint probability distribution:

| $X_2 \backslash X_3$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | $p_1$ | 0 | 0 | $p_2$ |
| 2 | 0 | $p_3$ | $p_4$ | 0 |
| 3 | $p_5$ | 0 | 0 | $p_6$ |
| 4 | 0 | $p_7$ | 0 | $p_8$ |
| 5 | $p_9$ | 0 | $p_{10}$ | 0 |

where $p_i > 0$ for all $i = 1, \ldots, 10$ and $\sum_{i=1}^{10} p_i = 1$. Although the support of this distribution has only 10 points (so, there are 10 points of zero probability), it is possible to verify condition (ii) or (iii) of Corollary 3.1, and therefore $X_2 \parallel X_3$.

Condition (iii) in Corollary 3.1 shows that the condition $p_{ij} > 0$ for all $(i, j)$ is sufficient but far from necessary for obtaining measurable separability between $X_2$ and $X_3$. The literature on graphical models repeatedly mention the non-necessity of the strict positivity of all $p_{ij}$ (see the references mentioned in Section 1): condition (iii) in Corollary 3.1 gives, for the finite case, a necessary *and* sufficient condition.

**Remark 3** In the more general case, namely $X_2 \parallel X_3 \mid X_4$, with $X_4$ a non trivial random variable (*i.e.*, $\overline{\mathcal{F}_0} \subsetneq \overline{\mathcal{X}_4}$), the measurable separability between $X_2$ and $X_3$ conditionally on $X_4$ should be verified for all $k \in N_4$. More precisely, the condition (iii) of Theorem 3.1 should be verified for each matrix $\mathbf{P}^{(k)}$, with $k \in N_4$.

### 3.3 Common information in the finite case: A general condition

In which cases $X_2$ and $X_3$ are not measurably separated? Using the equivalence between conditions (i) and (iii) of Corollary 3.1, it follows that $X_2$ and $X_3$ are not measurably separated if and only if $(\exists\, I \subset N_2)$ with $I \neq \emptyset$ and $I \neq N_2$ $[\forall\, (i, i', j) \in I \times (N_2 \setminus I) \times N_3]$ such that $p_{ij} \cdot p_{i'j} = 0$, *i.e.*, $p_{ij} = 0$, or $p_{i'j} = 0$, or both. This condition is equivalent to the following one: the matrix $\mathbf{P}$ representing the joint probability distribution of $(X_2, X_3)$ can, after permuting (if necessary) rows and/or columns, be put in the form of a block-diagonal matrix. This is a standard issue in the non-decomposability of finite Markov-chain in which case the probability matrix is square.

**Example 2** Consider the case discussed in section 3.1: the joint probability distribution of $(X_2, X_3)$ is represented by the matrix (3.1). Such a matrix is a block-diagonal one, so $X_2 \not\parallel X_3$. This explains why $X_1$ is not independent of $(X_2, X_3)$, although $X_1 \perp\!\!\!\perp X_2 \mid X_3$ and $X_1 \perp\!\!\!\perp X_3 \mid X_2$.

**Example 3** Consider the counter-example provided by Hill (1993, p. 259), namely to assume a trivariate discrete distribution such that $P(X_1 = 0, X_2 = 0, X_3 = 0) = P(X_1 = 1, X_2 = 1, X_3 = 1) = 0.5$ and $P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = 0$ otherwise. As pointed out by Hill (1993), this distribution satisfies conditions (i) and (ii) of the Intersection Property, but not the conclusion. This situation can be explained using the result established above: it can indeed be verified that the matrix representing the joint distribution of $(X_2, X_3)$ can be putted in the form of a block-diagonal matrix.

**Remark 4** In the more general case, namely when $X_4$ is a non trivial random variable, $X_2 \not\parallel X_3 \mid X_4$ if and only if *there exists at least one* $k_0 \in N_4$ such that, after permuting (if necessary) rows and/or columns, the conditional distribution $\mathbf{P}^{(k_0)}$ of $(X_2, X_3)$ given $X_4 = k_0$ can be put in the form of a block-diagonal matrix.

### 3.4 Measurable separability and independence in the finite case

Let us now re-examine Proposition 2.1 in the discrete case. By definition $X_2 \perp\!\!\!\perp X_3$ if and only if $\forall\, (i, j) \in N_2 \times N_3$

$$P[X_2 = i, X_3 = j] = P[X_1 = i] \cdot P[X_2 = j],$$

which is equivalent to $r[\mathbf{P}] = 1$, where $\mathbf{P}$ represents the joint distribution of $(X_2, X_3)$. Consequently, there are no non-null entries in the matrix $\mathbf{P}$; so, condition (iii) of Corollary 3.1 is trivially satisfied. This clearly shows that the condition of independence is sufficient but far from necessary to obtain the measurable separability.

**Remark 5** Again, when $X_4$ is a non trivial random variable, condition $X_2 \perp\!\!\!\perp X_3 \mid X_4$ is equivalent to $(\forall\, k \in N_4)\; r[\mathbf{P}^{(k)}] = 1$, and implies condition (iii) in Theorem 3.1.

# 4 Measure Separability in the Normal Case

A second case in which we want to analyze measurable separability, or no common information, is the normal one. In such a case, the null sets are well described through the null space of a $p \times q$ matrix $A$, namely Ker$(A) = \{x \in \mathbb{R}^q : Ax = 0\}$.

Let us consider a random vector $X = ({X_2}', {X_3}', {X_4}')' \in \mathbb{R}^{p_2+p_3+p_4}$. Let

$$\begin{aligned} \text{Ker}\,[V(X_2 \mid X_4)] &= \{a \in \mathbb{R}^{p_2} : a'X_2 = E(a'X_2 \mid X_4) \quad \text{a.s.}\} \\ \text{Ker}\,[C(X_3, X_2 \mid X_4)] &= \{a \in \mathbb{R}^{p_2} : C(X_3, a'X_2 \mid X_4) = 0 \quad \text{a.s.}\} \end{aligned}$$

where $V(\cdot \mid \cdot)$ and $C(\cdot, \cdot \mid \cdot)$ are the conditional variance and the conditional covariance operators, respectively; for details, see, *e.g.*, Drygas (1970) or Eaton (1989).

Suppose that $({X_2}', {X_3}' \mid {X_4}')' \sim \mathcal{N}_{p_2+p_3}(\mu(X_4), \Sigma(X_4))$, where $\Sigma(X_4)$ can be a positive or semi-positive definite symmetric matrix. The following lemma establishes a simple result which provides an easy key to characterize the measurable separability in the normal case; for a proof, see Appendix B.

**Lemma 4.1** *If* $({X_2}', {X_3}' \mid {X_4}')' \sim \mathcal{N}_{p_2+p_3}(\mu(X_4), \Sigma(X_4))$, *then*

$$\text{Ker}\,[V(X_2 \mid X_4)] = \text{Ker}\,[V(X_2 \mid X_3, X_4)] \cap \text{Ker}\,[C(X_3, X_2 \mid X_4)] \qquad \text{a.s.} \tag{4.1}$$

The following theorem (see Florens *et al.*, 1993, Lemma 1.7) characterizes the measurable separability in the normal case; for a proof, see Appendix C.

**Theorem 4.1** *If* $({X_2}', {X_3}' \mid {X_4}')' \sim \mathcal{N}_{p_2+p_3}(\mu(X_4), \Sigma(X_4))$, *then the following propositions are equivalent:*

(i) $X_2 \parallel X_3 \mid X_4$;

(ii) $\text{Ker}\,[V(X_2 \mid X_4)] = \text{Ker}\,[V(X_2 \mid X_3, X_4)]$ *a.s.;*

(iii) $\text{Ker}\,[V(X_2 \mid X_3, X_4)] \subset \text{Ker}\,[C(X_3, X_2 \mid X_4)]$ *a.s.;*

(iv) $r[V(X_2 \mid X_4)] = r[V(X_2 \mid X_3, X_4)]$ *a.s.*

Lemma 4.1 and Theorem 4.1 are valid whether the conditional covariance matrix $\Sigma(X_4)$ is singular or regular. The singular case is of particular interest since our concern is to examine the role of the null sets for characterizing measurable separability.

**Remark 6** Since the measurable separability condition $X_2 \parallel X_3 \mid X_4$ is symmetric in $X_2$ and $X_3$, one could formally add conditions to Theorem 4.1 by interchanging $X_2$ and $X_3$.

If we consider the case $X_4 = c$ a.s. ($c \in N_4$), then the following corollary characterizes the measurable separability between $X_2$ and $X_3$ as follows:

**Corollary 4.1** *If $({X_2}', {X_3}')' \sim \mathcal{N}_{p_2+p_3}(\mu, \Sigma)$, with $\Sigma$ is a positive or semi-positive definite symmetric matrix, then the following propositions are equivalent:*

(i) $X_2 \parallel X_3$;

(ii) $\text{Ker}\,[V(X_2)] = \text{Ker}\,[V(X_2 \mid X_3)]$;

(iii) $\text{Ker}\,[V(X_2 \mid X_3)] \subset \text{Ker}\,[C(X_3, X_2)]$;

(iv) $r[V(X_2)] = r[V(X_2 \mid X_3)]$.

Clearly, when $\Sigma > 0$, the density of $(X_2, X_3)$ exists and is strictly positive, trivially ensuring the measurable separability between $X_3$ and $X_4$; see Cox and Wermuth (1993, p. 206). Corollary 4.1 gives a necessary and sufficient condition of measurable separability far weaker than the existence of a strictly positive density. Corollary 4.1 also provides operational conditions to verify when $X_2$ and $X_3$ are not measurably separated.

The following lemma is useful to illustrate Proposition 2.1 in the normal case:

**Lemma 4.2** *If $({X_2}', {X_3}' \mid {X_4}')' \sim \mathcal{N}_{p_2+p_3}(\mu(X_4), \Sigma(X_4))$, with $\Sigma(X_4)$ a positive or semi-positive definite symmetric matrix, then*

$$X_2 \perp\!\!\!\perp X_3 \mid X_4 \Longleftrightarrow r[C(X_3, X_2 \mid X_4)] = 0 \Longleftrightarrow \text{Ker}\,[C(X_3, X_2 \mid X_4)] = \mathbb{R}^{p_2}.$$

Using Lemmas 4.1 and 4.2, it follows that

$$X_2 \perp\!\!\!\perp X_3 \mid X_4 \Longrightarrow \mathrm{Ker}\,[V(X_2 \mid X_4)] = \mathrm{Ker}\,[V(X_2 \mid X_3, X_4)].$$

By statement (ii) of Theorem 4.1, we again conclude that measurable separability is much weaker than conditional independence.

**Example 4** The following example illustrates very simply that the *a.s.* positivity of the density (or, in the normal case, the regularity of the covariance matrix) is not a necessary condition for measurable separability, and that the singularity of the covariance matrix is not a sufficient condition for non separability. Indeed, consider a trivariate normal distribution with covariance matrix:

$$\Sigma = \begin{bmatrix} 1 & 1 & .5 \\ 1 & 1 & .5 \\ .5 & .5 & 1 \end{bmatrix}$$

It may be checked that $V(Y_3 \mid Y_1, Y_2) = .75 > 0$. Thus, in view of Corollary 4.1, $Y_3$ and $(Y_1, Y_2)$ are measurably separated, whereas $V(Y_1 \mid Y_2, Y_3) = 0$ and, therefore, $Y_1$ and $(Y_2, Y_3)$ are not measurably separated. It may be noticed that, in this example, the singularity of the covariance matrix implies that $Y_1 - Y_2$ is *a.s.* a constant, there is accordingly common information between $Y_1$ and $Y_2$ and, therefore, between $Y_1$ and $(Y_2, Y_3)$, whereas there is no common information between $Y_3$ and $(Y_1, Y_2)$.

# 5 Concluding remarks

The concept of "no common information", also called "measurable separability" or absence of "splitting sets", appears in different contexts in the statistical literature. This paper has endeavored to enhance a better understanding of this concept by characterizing and illustrating what it is and what it is not in two topical cases: the finite one and the multivariate normal one. An important issue was to analyze the role of the null sets. In particular, in the finite case, the no common information was obtained even if the corresponding contingency table has some zeros.

Another way for getting a deeper understanding is to examine the role of that property in different contexts:

*Basu's First Theorem:* The condition of measurable separability appears, in the Introduction of this paper, as a supplementary condition for making the implication embodied in the Intersection Property (1.1) valid. This condition has been met in the First Basu's Theorem; see Basu and Pereira (1983, Theorem 2). Interestingly enough, the first "proof" without the supplementary condition in Basu (1955) was wrong because of mistreating null sets associated with conditional densities and the corrected proof, Basu (1958)

and Koehn and Thomas (1975), also shows that the supplementary condition aims at avoiding somewhat trivial pathologies, such as two independent observations of the exact value of the parameter. Situations formally similar to Basu's first theorem are frequently met in the literature on statistical inference; for examples, see Dawid (1979b). This condition is also relevant in the literature on graphical models where that supplementary condition is weaker than the too strong condition of *a.s.* positive density and this weakening is recognized as providing a considerably more useful, and operational, condition.

*Causal Inference:* The relevance of a suitable understanding of the concept, and the role, of measurable separability is provided by an interesting paper on "The assumptions on which causal inferences rest", namely Stone (1993). Thus, let us consider the following random variables: $X$ for treatment, $Z$ for observed covariates, $U$ for unobserved covariates and $Y$ for responses, under the assumption that $U$ is comprehensive enough to make the response determined by $X$, $Z$ and $U$, namely $Y = f(X, Z, U)$. The no-causation hypothesis may be written as $Y \perp\!\!\!\perp X \mid (Z, U)$, but is not directly testable because $U$ is not observed. A testable version could be $Y \perp\!\!\!\perp X \mid Z$ and hopefully equivalent under a further assumption of covariate sufficiency, namely $Y \perp\!\!\!\perp U \mid (Z, X)$. Stone (1993) paper raises two interesting issues. A first issue regards the role of measurable separability. As mentioned in Remark 2.3, if $X$ and $U$ are measurably separated conditionally on $Z$ (*i.e.*, $X \parallel U \mid Z$), then $Y \perp\!\!\!\perp X \mid (Z, U)$ and $Y \perp\!\!\!\perp U \mid (Z, X)$ imply $Y \perp\!\!\!\perp (X, U) \mid Z$ and, therefore, $Y \perp\!\!\!\perp X \mid Z$. In other words, the desired no-causation hypothesis along with covariate sufficiency imply the testable version of no-causation only under an hypothesis of measurable separability. But, the equivalence asserted in Stone (1993) is misleading because Theorem 2.2.10 in Florens *et al.*(1990) says that $Y \perp\!\!\!\perp X \mid Z$ and $Y \perp\!\!\!\perp U \mid Z, X$ is actually equivalent to $Y \perp\!\!\!\perp (X, U) \mid Z$ which implies $Y \perp\!\!\!\perp X \mid Z$, so that the testable version of no-causation along with the covariate sufficiency imply the desired no-causation *without* requiring a condition of measurable separability. The other issue regards the meaning of measurable separability which is *not* the hypothesis that the support of the conditional distribution of $(X \mid Z, U)$ does not depend on $U$, as asserted in Stone (1993): this is made clear in Example 4 after reminding that (using the notation of the example) always $P[(Y_1, Y_2, Y_3)' \in \mathrm{Im}\,(\Sigma) + \mu] = 1$, *i.e.*, $\mathrm{supp}[(Y_1, Y_2, Y_3)'] \subset \mathrm{Im}\,(\Sigma) + \mu$ with probability 1. Interestingly enough, Stone (1993) correctly noticed that the measurable separability is actually part of the *definition* of unobserved covariates. Indeed, $U$ may be *defined* by the properties $Y = f(X, Z, U)$ and $X \parallel U \mid Z$: if there were common information between $X$ and $U$ conditionally on $Z$, it would be difficult to interpret $U$ as being both unobserved and comprehensive; for more details, see Mouchart (2004).

*Estimability in Markov Processes:* Somewhat different is the role of measurable separability in problems of exact estimability. Thus Florens *et al.*(1990, Proposition 9.3.24) shows that the sampling measurable separability of the first two observations in a stationary Markovian process is sufficient to ensure the exact estimability of the minimal sufficient parameter: these authors also mention that the condition of measurable separability is slightly too strong but easier to handle than Doeblin's condition; see, *e.g.*, Stout (1974, Section 3.6) or Breiman (1968, Section 7.3).

*Identification of ATE:* In recent unpublished works, for the analysis of identification of the Average Treatment Effect (ATE) in non parametric models, Florens *et al.*(2003) have repeatedly used the condition of measurable separability (see, *e.g.*, their Theorem 3.5 for the equivalence between an exclusion condition

and a Local Instrumental Variable condition, or Theorem 3.6 for the identification of the ATE).

# Appendix

## A Proof of Theorem 3.1

The equivalence between statements (ii) and (iii) follows from the definition of the sets $N_{3i}^{(k)}$ and $N_{3i'}^{(k)}$. Before proving that statement (i) implies statement (ii), note that by Theorem 2.1, $X_2 \parallel X_3 \mid X_4$ is equivalent to asserts that if there exist two functions $f$ and $g$ such that $f(i,k) = g(j,k)\ \forall\,(i,j,k)$ such that $p_{ijk} > 0$, then there exists a function $h$ such that $f(i,k) = h(k)\ \forall\,(i,k)$ such that $p_{i|k} > 0$. By the definition of the sets $N_2^{(k)}$, $N_3^{(k)}$ and $N_{3i}^{(k)}$, this last implication is equivalent to the following one:

$$\begin{aligned} f(i,k) = g(j,k) \quad \forall\, k \in N_4 \quad \forall\,(i,j) \in N_2^{(k)} \times N_{3i}^{(k)} \\ \implies f(i,k) = h(k) \quad \forall\, k \in N_4 \quad \forall\, i \in N_1^{(k)}. \end{aligned} \tag{A.1}$$

**Proof of (i) $\implies$ (ii):** Indeed, if the condition (ii) is not satisfied, it follows that $(\exists\, k \in N_4)\ (\exists\, I \subset N_2^{(k)})$ with $I \neq \emptyset$ and $I \neq N_2^{(k)}$ such that

$$\left(\bigcup_{i\in I} N_{3i}^{(k)}\right) \cap \left(\bigcup_{i\in N_2^{(k)}\setminus I} N_{3i}^{(k)}\right) = \emptyset.$$

Denoting $\bigcup_{i\in I} N_{3i}^{(k)}$ as $J(I)$, it follows that

$$\begin{aligned} &\text{(i)} \quad J(I) \subset \left(\bigcup_{i\in N_2^{(k)}\setminus I} N_{3i}^{(k)}\right)^c = \bigcap_{i\in N_2^{(k)}\setminus I} \left(N_3^{(k)} \setminus N_{3i}^{(k)}\right), \\ &\text{(ii)} \quad J(I)^c = \bigcap_{i\in I} \left(N_3^{(k)} \setminus N_{3i}^{(k)}\right). \end{aligned} \tag{A.2}$$

Thus,

$$\text{(A.2.i)} \implies [\forall\, j \in J(I)] \;\; (\forall\, i \in N_2^{(k)} \setminus I) \qquad p_{ij|k} \;=\; 0$$

$$\text{(A.2.ii)} \implies [\forall\, j \in J(I)^c] \;\; (\forall\, i \in I) \qquad p_{ij|k} \;=\; 0.$$

Consequently, $P[X_2 \in N_2^{(k)} \backslash I, X_3 \in J(I) \mid X_4 = k] \;=\; 0$ and $P[X_2 \in I, X_3 \in J(I)^c \mid X_4 = k] \;=\; 0$. Therefore

$$\{X_2 \in I\} \cap \{X_4 \;=\; k\} \;=\; \{X_3 \in J(I)\} \cap \{X_4 \;=\; k\} \quad \text{a.s.} \tag{A.3}$$

Moreover,

$$0 < P[X_2 \in I, X_4 \;=\; k] \;<\; 1 \tag{A.4}$$

since $P[X_2 \in I \mid X_4 \;=\; k] < 1$. If not, *i.e.*, if $P[X_2 \in I \mid X_4 \;=\; k] = 1$, then:

$$\sum_{i \in I} P[X_2 \;=\; i \mid X_4 \;=\; k] \;=\; 1 \quad \text{and} \quad I \subsetneq N_2^{(k)}.$$

Hence, $(\exists\, i_0 \in N_2^{(k)} \setminus I) \quad P[X_2 \;=\; i_0 \mid X_4 \;=\; k] = 0$. This is a contradiction with the definition of the set $N_2^{(k)}$. Therefore, (A.3) and (A.4) jointly imply that $X_2 \;\|\; X_3 \mid X_4$ is violated (see (A.1)).

**Proof of (iii) $\implies$ (i):** Assume that there exist two functions $f$ and $g$ such that

$$f(i,k) \;=\; g(j,k) \quad \forall\, k \in N_4 \quad \forall\, (i,j) \in N_2^{(k)} \times N_{3i}^{(k)}. \tag{A.5}$$

Condition (iii) implies that $(\forall\, k \in N_4)$ $(\forall\, I \subset N_2^{(k)})$ with $I \neq \emptyset$ and $I \neq N_2^{(k)}$ $(\exists\, (i,i') \in I \times N_2^{(k)} \setminus I)$ such that

$$N_3^{(k)} \;\supset\; N_{3i}^{(k)} \;\cap\; N_{3i'}^{(k)} \;\neq\; \emptyset.$$

Let $j_0 \in N_{3i}^{(k)} \;\cap\; N_{3i'}^{(k)}$. By (A.5) it follows that $f(i,k) \;=\; f(i',k)$. Therefore, we have that $(\forall\, k \in N_4)$ $(\forall\, I \subset N_2^{(k)})$ with $I \neq \emptyset$ and $I \neq N_2^{(k)}$ $(\exists\, (i,i') \in I \times N_2^{(k)} \setminus I)$ such that $f(i,k) \;=\; f(i',k)$. Applying inductively this condition we obtain that $\forall\, i, i' \;\in\; N_2^{(k)} \quad f(i,k) \;=\; f(i',k)$. Consequently, taking $i_0 \in N_2^{(k)}$ fixed, this last equality is equivalent to $f(i,k) \;=\; f(i_0,k) \;\equiv\; h(k) \quad \forall\, i \in N_2^{(k)}$.

□

## B Proof of Lemma 4.1

In general, $V(a'X_2 \mid X_4) = E[V(a'X_2 \mid X_3, X_4) \mid X_4] + V[E(a'X_2 \mid X_3, X_4) \mid X_4]$. Under normality, $V[E(a'X_2 \mid X_3, X_4) \mid X_4] = 0$ a.s. is equivalent to $C(X_3, a'X_2 \mid X_4) = 0$ a.s. Therefore, the nullity of each member of the equality corresponds to $a$ pertaining to the respective null spaces of (4.1).

□

## C Proof of Theorem 4.1

The equivalence between (ii) and (iii) is an immediate consequence of Lemma 4.1, whereas the equivalence between (ii) and (iv) is a consequence of the rank theorem in linear algebra (see, *e.g.*, Halmos, 1974, Theorem 1, section 50). The proof of the equivalence between (i) and (iii) is based on the following lemma:

**Lemma C.1** *Let $({Z_1}', {Z_2}')' \in \mathbb{R}^{p_1+p_2}$ be a random vector such that $r[V(Z_1 \mid Z_2)] = q_1 \leq p_1$. Then there exists a $q_1 \times p_1$ matrix $T$ with $r(T) = q_1$ such that*

(i) $r[V(TZ_1 \mid Z_2)] = q_1$;

(ii) $\overline{\sigma(TZ_1, Z_2)} = \overline{\sigma(Z_1, Z_2)}$.

**Proof of Lemma C.1:** If $q_1 = p_1$, take $T = I_{p_1}$. Assume therefore that $q_1 < p_1$. Then there exists an orthogonal matrix $\begin{pmatrix} T \\ Q \end{pmatrix}$ with $T$ a $q_1 \times p_1$ matrix and $Q$ a $(p_1 - q_1) \times p_1$ matrix, such that $V(Z_1 \mid Z_2) = T'\Delta T$, where $\Delta$ is the diagonal matrix with the positive eigenvalues of $V(Z_1 \mid Z_2)$. Therefore $V(TZ_1 \mid Z_2) = \Delta$; this proves (i). Moreover, $V(QZ_1 \mid Z_2) = 0$, $C(TZ_1, QZ_1 \mid Z_2) = 0$, $TT' = I_{q_1}$ and $T'T + Q'Q = I_{p_1}$. It follows that

$$QZ_1 = E(QZ_1 \mid Z_2) \qquad Z_2\text{-a.s.}$$

and consequently $Z_1 = T'TZ_1 + Q'QZ_1 = T'TZ_1 + Q'E(QZ_1 \mid Z_2) \quad Z_2$-a.s. Therefore $\sigma(Z_1, Z_2) \subset \overline{\sigma(TZ_1, Z_2)}$. The inverse inclusion $\sigma(TZ_1, Z_2) \subset \sigma(Z_1, Z_2)$ being trivial, we obtain (ii).

□

**Proof of (i) $\Longleftrightarrow$ (iii):** To verify this equivalence, Lemma C.1 is used, on the one hand, to characterize the $\sigma$-fields $\overline{\sigma(X_2, X_3)}$ and $\overline{\sigma(X_4, X_3)}$, and, on the other hand, to find a non-degenerate normal distribution $(Y_2', Y_3' \mid X_4)$, where $Y_2$ and $Y_3$ are suitable transformations of $X_2$ and $X_3$, respectively.

As a matter of fact, assume that $r[V(X_3 \mid X_4)] = q_3 < p_3$. By Lemma C.1, there exists a full rank $q_3 \times p_3$ matrix $T_3$ such that $Y_3 = T_3 X_3$, $r[V(Y_3 \mid X_4)] = q_3$ and

$$\overline{\sigma(X_3, X_4)} = \overline{\sigma(Y_3, X_4)}. \tag{C.1}$$

It follows that

$$(X_2 \mid Y_3, X_4) \sim \mathcal{N}_{p_2}(g(X_4) + B_3(X_4)Y_3,\, V(X_2 \mid Y_3, X_4)), \tag{C.2}$$

where $r[V(X_2 \mid Y_3, X_4)] = q_2 < p_2$. By Lemma C.1, there exists a full rank $q_2 \times p_2$ matrix $T_2$ such that $Y_2 = T_2 X_2$ and

$$(Y_2 \mid Y_3, X_4) \sim \mathcal{N}_{q_2}(T_2 g(X_4) + T_2 B_3(X_4)Y_3,\, V(Y_2 \mid Y_3, X_4)), \quad r[V(Y_2 \mid Y_3, X_4)] = q_2. \tag{C.3}$$

Moreover, using arguments similar to that used in the proof of Lemma C.1, it follows that

$$X_2 = T_2' Y_2 + Q_2' Q_2\, E(X_2 \mid Y_3, X_4) \quad (Y_3, X_4)\text{-a.s.}$$

where $Q_2$ is a full rank $(p_2 - q_2) \times p_2$ matrix such that $Q_2 Q_2' = I_{p_2 - q_2}$ and $Q_2 T_2' = T_2 Q_2' = 0$. Therefore, $Q_2 X_2 = Q_2\, E(X_2 \mid Y_3, X_4)$ $(Y_3, X_4)$-a.s. and, consequently,

$$\sigma(X_2, X_4) = \sigma(T_2 X_2, Q_2 X_2, X_4) \subset \overline{\sigma(Y_2, Q_2\, E(X_2 \mid Y_3, X_4), X_4)}.$$

Since the other inclusion is trivial, it follows that

$$\overline{\sigma(X_2, X_4)} = \overline{\sigma(Y_2, Q_2\, E(X_2 \mid Y_3, X_4), X_4)}. \tag{C.4}$$

Thus, from (C.2) and (C.3) it follows that $(Y_2', Y_3' \mid X_4)$ has a non-degenerate probability distribution. Therefore, there exists a probability distribution $P' \sim P$ such that

$$\sigma(Y_2) \perp\!\!\!\perp \sigma(Y_3) \mid \sigma(X_4);\, P', \tag{C.5}$$

where $P$ corresponds to the probability distribution of $(Y_2', Y_3', X_4')$. Since $\sigma(Q_2\, E(X_2 \mid Y_3, X_4)) \subset \sigma(Y_3, X_4)$, condition (C.5) implies that

$$\sigma(Y_2) \perp\!\!\!\perp \sigma(Y_3) \mid \sigma(Q_2\, E(X_2 \mid Y_3, X_4), X_4);\ P'.$$

Since $P' \sim P$, Proposition 2.1 implies that

$$\sigma(Y_2) \ \parallel\ \sigma(Y_3) \mid \sigma(Q_2\, E(X_2 \mid Y_3, X_4), X_4);\ P.$$

Thus, using conditions (C.1) and (C.4), it follows that

$$\begin{aligned}
\overline{\sigma(X_2, X_4)} \cap \overline{\sigma(X_3, X_4)} &= \overline{\sigma(Y_2, Q_2\, E(X_2 \mid Y_3, X_4), X_4)} \cap \overline{\sigma(Y_3, X_4)} \\
&= \overline{\sigma(Q_2\, E(X_2 \mid Y_3, X_4), X_4)}.
\end{aligned}$$

Therefore, $\sigma(X_2) \ \parallel\ \sigma(X_3) \mid \sigma(X_4)$ if and only if $\sigma(Q_2\, E(X_2 \mid Y_3, X_4)) \subset \overline{\sigma(X_4)}$ or, equivalently, $V[Q_2\, E(X_2 \mid Y_3, X_4)] = 0$, *i.e.*,

$$Q_2' Q_2\, C(X_2, Y_3 \mid X_4) = 0. \tag{C.6}$$

Since $Q_2' Q_2$ is an orthonormal projection on Ker $[V(X_2 \mid Y_3, X_4)]$, condition (C.6) is equivalent to Im $[C(X_2, Y_3 \mid X_4)] \subset \{\text{Ker}\,[V(X_2 \mid Y_3, X_4]\}^{\perp}$. This last relation can equivalently be rewritten as Ker $[V(X_2 \mid Y_3, X_4] \subset$ Ker $[C(Y_3, X_2 \mid X_4)]$, which is equivalent to condition (iii) of Theorem 4.1 because of condition (C.1) and $C(X_3, X_2 \mid X_4) = T_3'\, C(Y_3, X_2 \mid X_4)$.

□

**Acknowledgments:** This research is financially supported by different grants: the FONDECYT Project No. 1030801 grant from Fondo Nacional de Ciencia y Tecnología de Chile; a grant to the first author from the COIMBRA Group for visiting the Department of Psychology, K. U. Leuven, Belgium; the "Projet d'Actions de Recherche Concertées" No. 98/03-217; and the "Interuniversitary Attraction Pole", Phase V (No. P5/24) from the Belgian Government. The first author acknowledges several discussions held with I. Meilijson (during his visit to Chili in 2001), P. Suppes (during his visit to Chili in 2003) and B. Vantaggi (during the ISBA Meeting, Valparaíso, Chili, May 2004). The first author also wants specially to acknowledge several discussions held with J. Kadane in 2004, during his visit to the Department of Statistics, Pontificia Universidad Católica de Chile. The last two authors gratefully acknowledge years-long discussions with J. P. Florens.

## References


ANDERSSON, S. A., MADIGAN, D. AND PERLMAN, M. (1997). On the Markov Equivalence of Chain Graphs, Undirected Graphs, and Acyclic Digraphs. *Scand. J. Statist.* **24**, 81–102.

ANDERSSON, S. A., MADIGAN, D. AND PERLMAN, M. D. (2001). Alternative Markov Properties for Chain Graphs. *Scand. J. Statist.* **28**, 33-85.

BASU, D. (1955). On Statistics Independent of a Complete Sufficient Statistics. *Sankhyā* **15**, 377-380.

BASU, D. (1958). On Statistics Independent of Sufficient Statistics. *Sankhyā* **20**, 223-226.

BASU, D. AND PEREIRA, C. B. (1983). Conditional Independence in Statistics. *Sankhyā, Series A* **45**, 324-337.

BREIMAN, L. (1968). *Probability*. Addison Wesley, Reading Massachusetts.

COX, D. R. AND WERMUTH, N. (1993). Linear Dependencies Represented by Chain Graphs (with Discussion). *Statist. Sc.* **8**, 204-283.

COX, D. R. AND WERMUTH, N. (1996). *Multivariate Dependencies. Models, analysis and interpretation*. Chapman and Hall, London.

DAWID, A. P. (1979a). Conditional Independence in Statistical Theory (with Discussion). *J. R. Statist. Soc., Series B* **41**, 1-31.

DAWID, A. P. (1979b). Some Misleading Arguments Involving Conditional Independence. *J. R. Statist. Soc., Series B* **41**, 249-252.

DAWID, A. P. (1980). Conditional independence for statistical operations. *Ann. Statist.* **8**, 598–617.

DAWID, A. P. (1985). Invariance and Independence in Multivariate Distributions Theory. *J. Multivariate Anal.* **17**, 304–315.

DE FINETTI, B. (1949). Sull'impostazione assiomatica del calcolo delle probabilit. *Annali Univ. Trieste* bf 19, 3–55.

DE FINETTI, B. (1970). *Theory of Probability, vols. I and II*. Wiley, London.

DÖHLER, R. (1980). On the Conditional Independence of Random Events. *Theory of Probability and Its Applications* **25**, 628-634.

DRYGAS, H. (1970). *The Coordinate-Free Approach to Gauss-Markov Estimation*. Lecture Notes in Operation Research and Mathematical Systems **40**. Berlin: Springer.

EATON, M. L. (1989). *Multivariate Statistics. A Vector Space Approach*. John Wiley and Sons, New York.

FLORENS, J. -P. AND MOUCHART, M. (1982). A note on non causality. *Econometrica* **50**, 583–591.

FLORENS, J. -P., MOUCHART, M. AND ROLIN, J. -M. (1990). *Elements of Bayesian Statistics*. Marcel Dekker, New York.

FLORENS, J. -P., MOUCHART, M. AND ROLIN, J. -M. (1993). Noncausality and marginalization of Markov processes. *Econometric Theory* **9**, 241–262.

FLORENS, J. -P., HECKMAN, J. J., MEGHIR, C. AND VYTLACIL, E. (2003). Instrumental Variables, Local Instrumental Variables and Control Functions. Unpublished discussion paper. Available at

`http://idei.fr/doc/by/florens/nonpar_iv_april1.pdf.`

FRYDENBERG, M. (1990). The Chain Graph Markov Property. *Scand. J. Statist.* **17**, 333-353.

GEIGER, P., PAZ, A. AND PEARL, J. (1988). Axioms and Algorithms for Inferences Involving Probabilistic Independence. *Information and Computation* **91**, 128–141.

GEIGER, D. AND PEARL, J. (1993). Logical and Algorithmic Properties of Conditional Independence and Graphical Models. *Ann. Statist.* **21**, 2001-2021.

HALMOS, P. (1974). *Finite-Dimensional Vector Space*. Springer, New York.

HILL, J. R. (1993). Comments on "Linear Dependencies Represented by Chain Graphs ". *Statist. Sc.* **8**, 258–261.

KAUERMANN, G. (1996). On a Dualization of Graphical Gaussian Models. *Scand. J. Statist.* **23**, 105-116.

KOEHN, U. AND THOMAS, D. L. (1975). On Statistics Independent of a Sufficient Statistic: Basu's Lemma. *Amer. Statistician* **29**, 40-42.

KOSTER, J. A. T. (1996). Markov Properties of Nonrecursive Causal Models. *Ann. Statist.* **24**, 2148-2177.

KOSTER, J. A. T. (1999). On the Validity of the Markov Interpretation of Path Diagrams of Gaussian Structural Equation Systems with Correlated Errors. *Scand. J. Statist.* **26**, 413-431.

LAURITZEN, S. (1998). *Graphical Models*. Clarendon Press, Oxford.

LAURITZEN, S. AND SPIEGELHALTER, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert system (with discussion). *J. Roy. Statist. Soc., Series B* **50**, 157–224.

LAURITZEN, S. AND WERMUTH, N. (1989). Graphical models for associations between variables, some of which are qualitative and some quantitative. *Ann. Statist.* **17**, 31–57.

MARTIN, F., PETIT, J. -L. AND LITTAYE, M. (1973). Indépendance conditionnelle dans le modèle statistique bayésien. *Ann. Inst. Henri Poincaré* **9**, 19–40.

MOUCHART, M. (2004), Ignorable Missingness, Discussion Paper. Available at http://www.stat.ucl.ac.be.

MOUCHART, M. AND ROLIN, J. -M. (1984). A Note on Conditional Independence with Statistical Applications. *Statistica* **44**, 557-584.

MOUCHART, M. AND SAN MARTÍN, E. (2003). Specification and identification issues in models involving a latent hierarchical structure. *J. Statist. Plann. Inf.* **111**, 143–163.

MOUSSOURIS, J. (1974). Gibbs and Markov random systems with constraints. *J. Statist. Phys.* **10**, 11–33.

NOGALES, A. G., OYOLA, J. A. AND PÉREZ, P. (2000). On conditional independence and the relationship between sufficiency and invariance under the Bayesian point of view. *Statist. Prob. Lett.* **46**, 75-84.

NOVICK, M. R. (1979). Comments on "Conditional Independence in Statistical Theory". *J. R. Statist. Soc., Series B* **41**, 27.

PEARL, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufman, San Mateo, Ca.

PEARL, J. (1995). Causal Diagrams for Empirical Research (with Discussion). *Biometrika* **82**, 669-710.

PEARL, J. AND PAZ, A. (1987). Graphoids: Graph-based Logic for Reasoning about Relevance Relations. In: B. Du Boulay, D. Hogg and L. Steels (Eds.), *Advance in Artificial Inteligence-II*, pp. 357-363. Elsevier, Amsterdam.

SPEED, T. P. AND KIIVERI, H. T. (1986). Gaussian Markov Distributions over Finite Graphs. *Ann. Statist.* **14**, 138-150.

SPOHN, W. (1980). Stochastic Independence, Causal Independence, and Shieldability. *J. Phil. Logic* **9**, 73-99.

SPOHN, W. (1994). On the Properties of Conditional Independence. In: P. Humphreys, *Patrick Suppes: Scientific Philosopher. Volumen 1. Probability and Probabilistic Causality*, pp. 173-196. Kluwer, Dordrecht.

STONE, R. (1993). The Assumptions on which Causal Inference Rest. *J. R. Statist. Soc., Series B* **55**, 455-466.

STOUT, W.F. (1974). *Almost Sure Convergence*, Academic Press, New York.

STUDENÝ, M. (1997). Semigraiphoids and structures of probabilistic conditional independence. *Ann. Math. Art. Intell.* **21**, 71–98.

STUDENÝ, M. AND BOUCKAERT, R. R. (1998). On Chain Graph Models for Description of Conditional Independence Structures. *Ann. Statist.* **26**, 1434-1495.

VANTAGGI, B (2001). Conditional independence in a coherent setting. *Ann. Math. Artif. Intell.* **32**, 287–313.

VANTAGGI, B. (2002). The $L$-separation criterion for description of cs-independence models. *International Journal of Approximate Reasoning* **29**, 291–316.

WHITAKKER, J. (1990). *Graphical Models in Applied Multivariate Statistics*. Wiley, Chichester.

ERNESTO SAN MARTÍN
DEPARTAMENTO DE ESTADÍSTICA
PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
CASILLA 306, SANTIAGO 22, CHILE
E-MAIL: ESANMART@MAT.PUC.CL

MICHEL MOUCHART AND JEAN-MARIE ROLIN
INSTITUT DE STATISTIQUE
UNIVERSITÉ CATHOLIQUE DE LOUVAIN
20 VOIE DU ROMAN PAYS, B-1348 LOUVAIN-LA-NEUVE, BELGIUM
E-MAIL: MOUCHART@STAT.UCL.AC.BE
ROLIN@STAT.UCL.AC.BE