Conditional Independence In Statistics

by

D. Basu\(^1\) and Carlos A. B. Pereira\(^2\)

January, 1981
FSU Statistics Reprot No. M567

The Florida State University
Department of Statistics
Tallahassee, Florida 32306

\(^1\)Research partially supported by NSF Grant No. 79-04693

\(^2\)Research supported by CAPES and USP - Brazil
CONDITIONAL INDEPENDENCE IN STATISTICS

1 - INTRODUCTION

The notion of conditional independence is a central theme of Statistics. In a series of recent articles A. P. Dawid (1979a, b, 1980), J. P. Florens and M. Mouchart (1977), and M. Mouchart and J. M. Rolin (1978) have explained at length the grammar of conditional independence as a language of statistics. This article is a further elucidation on the subject and is generally of an expository nature. Several results that have already appeared elsewhere are amplified and their proofs simplified and unified. The only mathematical tool that is repeatedly used is that of conditioning operator.

The statistical perspective of this article is that of a Bayesian. A problem begins with a parameter (state of nature) \( \Theta \) with its prior probability model \( (\Theta, \mathcal{B}, \xi) \) that exists only in the mind of the investigator. There is an observable \( X \) with an associated statistical model \( (X, \mathcal{A}, \{P_\Theta: \Theta \in \Theta\}) \). Writing \( \omega = (\Theta, X) \), \( (\Theta, F) = (\Theta \times X, \mathcal{B} \times \mathcal{A}) \), and \( \Pi \) for the joint distribution of \( (\Theta, X) \), there exists then a subjective probability model \( (\Omega, F, \Pi) \) for \( \omega \). Hidden behind the wings of the Bayesian probability model \( (\Omega, F, \Pi) \) are the four models:
(i) The prior model \((\theta, B, \xi)\),

(ii) the statistical model \((X, A, \{P_\theta : \theta \in \Theta\})\)

(iii) the posterior model \((\Theta, B, \{\xi_x : x \in X\})\),

and (iv) the predictive model \((X, A, P)\), where \(P\) is the marginal or predictive distribution of \(X\).

In classical probability theory, the notion of conditional independence appears in a rather indirect fashion in the study of Markov chains and processes. A sequence of three random entities \((X, Y, Z)\) is said to possess the Markov property if, given \(X\) and \(Y\), the conditional distribution of \(Z\) depends on \((X, Y)\) only through \(Y\). An equivalent characterization of the Markov property may be stated in the symmetric form: \(X\) and \(Z\) are conditionally independent given \(Y\). In Section 3 we make precise these two definitions of conditional independence in terms of the conditioning operator.

In Statistics the phenomenon of conditional independence manifests itself in a much more direct and natural fashion. The statistical model that is most commonly in use is that of a sequence \(X = (X_1, X_2, \ldots)\) of observables that are independently and identically distributed (i.i.d.) for each given value of \(\theta\). It was DeFinetti (1937) who emphasized that, in view of the fact that \(\theta\) is not fully known, it is appropriate to regard the sequence of \(X_i\)'s not as i.i.d. random variables but as an exchangeable process. The fact that the \(X_i\)'s are conditionally i.i.d. implies that they are positively dependent.
Consider for example the particular case where $X_1, \ldots, X_n$ are i.i.d. with common distribution $N(\mu, \sigma^2)$ with $\theta = (\mu, \sigma^2)$ not fully known. In almost every textbook on Statistics it is proved that the statistic $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ is stochastically independent of $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$. Does it mean that $\overline{X}$, when observed, carries no information about $S^2$? That the answer cannot be "yes" is easily seen as follows. Suppose that the sample size $n = 25$ and that our partial knowledge about $\theta = (\mu, \sigma^2)$ is as follows: $\mu = 0$ or 1 and $\sigma^2 = 1$ or 100 (that is, $\theta = \{(0, 1), (1, 1), (0, 100), (1, 100)\}$). Suppose now that $\overline{X}$ is observed and is equal to 2.1. This observation generates the four likelihoods $L(0, 1)$, $L(1, 1)$, $L(0, 100)$, and $L(1, 100)$ where $L(0, 1) = \frac{5}{\sqrt{2\pi}} \exp\left(-\frac{25}{2} (2.1)^2\right)$ and so on. The relative likelihoods work out roughly as $10^{-17}$, $1$, $2(10)^5$, and $3(10)^5$ respectively. Thus, it is intuitive that the observation $\overline{X} = 2.1$ almost categorically rules out the points $(0, 1)$ and $(1, 1)$. Therefore, the observation of $\overline{X} = 2.1$ asserts that $\sigma^2 = 100$ with a lot of emphasis and so we may conclude that $S^2$ is of the order of 100. Then $\overline{X}$ and $S^2$, even though they are conditionally independent given $\theta$, are in effect highly dependent.

The three entities $\theta = (\mu, \sigma^2)$, $T = (\overline{X}, S^2)$, and $\mathcal{X} = (X_1, \ldots, X_n)$, in this order, have the Markov property in the sense that, given $\theta$
and $T$, the conditional distribution of $X$ depends on $(\theta, T)$ only through $T$. This is the sufficiency property of the statistic $T$ as recognized by R. A. Fisher (1920, 1922). A. N. Kolmogorov (1942) gave a Bayesian characterization of the notion of sufficiency by noting that irrespective of the choice of the prior distribution $\xi$ for the parameter $\theta$, the posterior distribution $\xi_X$ of $\theta$ depends on $X$ only through $T$. In other words, the sequence $(X, T, \theta)$ have the Markov property; that is $X$ and $\theta$ are conditionally independent given $T$. Note that the Fisher characterization of sufficiency is made only in terms of the statistical model for $X$ whereas the Kolmogorov characterization is made in terms of a large family of Bayesian models $(\Omega, F, \Pi)$ for $\omega = (\theta, X)$. (See Basu (1977) and Cheng (1978) for further details on these characterizations.)

Fisher regarded a sufficient statistic $T$ as one that summarizes in itself all the available relevant information in the sample $X$ about the parameter $\theta$. He called a statistic $Y = Y(X)$ ancillary if the conditional distribution of $Y$ given $\theta$, does not involve $\theta$ (is the same for all values of $\theta$). For example, the statistic $\sum(x_i - \overline{X})^4 / S^4$ is ancillary. In a series of articles D. Basu (1955, 1958, 1959, 1964, 1967) studied the phenomena of sufficiency, ancillarity, and conditional independence from various angles. In these articles, Basu's viewpoint was non-Bayesian in the sense that he did not introduce a prior distribution $\xi$ for the parameter $\theta$. 
M. Mouchart and J. M. Rolin (1978) studied in depth the familiar Basu theorems on sufficiency, ancillarity, and conditional independence from the viewpoint of a Bayesian model \((\Omega, F, \Pi)\). In Sections 5, 6, and 7 we review Basu's results from the Bayesian perspective. This is done mainly as an exercise in the use of the language of conditional independence developed earlier.

2 - NOTATION AND PRELIMINARIES

Let \((\Omega, F, \Pi)\) be the basic probability space. By a "random object" \(X\) we mean a measurable map \(\omega \rightarrow X(\omega)\) of \((\Omega, F)\) into another measurable space \((X, A)\). The sub-\(\sigma\)-algebra (to be called subfield) of \(X\)-events \(\{X^{-1}A; A \in A\}\) will be denoted by \(F_X\). The two probability spaces \((\Omega, F_X, \Pi)\), and \((X, A, \Pi^{-1})\) are indistinguishable in a sense, and so we shall, as a rule, identify a random object \(X\) with the induced subfield \(F_X\) of \(F\). In that way, one could say that random objects are generators of subfields. Examples of random objects include random variables, random vectors, and any collection of random variables (stochastic processes).

For any two subfields \(F'\) and \(F''\) of \(F\), \(F' \vee F''\) denotes the smallest subfield of \(F\) that contains both \(F'\) and \(F''\). The smallest subfield that contains all null sets of \(F\) (a set \(N\) is null if \(\Pi(N) = 0\)) is denoted by \(\overline{F}_0\); that is,

\[
\overline{F}_0 = \{F; F \in F \text{ and } \Pi(F) = 0 \text{ or } \Pi(F) = 1\}
\]
and write $\mathcal{F}_0 = \{\emptyset, \Omega\}$, the trivial subfield.

A subfield of $\mathcal{F}$ is said to be completed if it contains $\overline{\mathcal{F}}_0$.

For any subfield $\mathcal{F}'$ of $\mathcal{F}$ its completion is defined by:

$$\overline{\mathcal{F}}' = \mathcal{F}' \vee \overline{\mathcal{F}}_0.$$

For a random object $X$, the notation $X \in \mathcal{F}'$ indicates that $\mathcal{F}_X \subset \overline{\mathcal{F}}'$ and $X$ is said to be essentially $\mathcal{F}'$-measurable or $X$ is \textit{ess-}$\mathcal{F}'$-measurable. A random variable is a random object with range $(\mathbb{R}_1, \mathcal{B}_1)$ where $\mathbb{R}_1$ is the real line and $\mathcal{B}_1$ is the Borel $\sigma$-algebra. A random variable $f$ is said to be bounded if $\exists \ a \in \mathbb{R}_1$ such that $\mathbb{P}\{\omega; |f(\omega)| \leq a\} = 1$. In the sequel, all random variables shall be regarded as bounded unless stated otherwise and the use of small letters shall be restricted to their representation. The notation $f \subset X$ indicates that the random variable $f$ is ess-$\mathcal{F}_X$-measurable. In the same spirit, for two random objects $X$ and $Y$, we write $X \leq Y$ to indicate that $\mathcal{F}_X \subset \mathcal{F}_Y$. If $\mathcal{F}_X = \mathcal{F}_Y$ we write $X \equiv Y$ to indicate the essential equivalence between $X$ and $Y$. The class of all bounded random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ is denoted by $L_\infty$ and $L_\infty(X)$ denotes the class of all ess-$\mathcal{F}_X$-measurable random variables. Here and for the rest of this article, equality of two random variables means essential equality; that is, $f = g$ means that $\{\omega; f(\omega) \neq g(\omega)\}$ is a null set.
DEFINITION 1

The conditional expectation of \( f \), given a random object \( X \), is a random variable \( f^*_X \in L_\infty(X) \) such that

\[
\int f g d\Pi = \int f^*_X g d\Pi \vee g \in L_\infty(X).
\]

Another notation for \( f^*_X \) is \( E(f|X) \). When the conditioning random object \( X \) is implicit in the context, \( f^* \) is substituted for \( f^*_X \).

The map \( f \mapsto f^* \) of \( L_\infty \) to \( L_\infty(X) \) is linear, constant preserving, monotone, idempotent, and is a contraction in the \( L_p \) norm if \( p \geq 1 \).

The following proposition, known as smoothing theorem, is widely used in this paper. Here, \( * \) is substituted for \( *X \) and \( + \) is substituted for \( *Y \).

PROPOSITION 1

If two random objects \( X \) and \( Y \) are such that \( X \subseteq Y \), then

\( \forall f \in L_\infty \)

(i) \( E(f^*|Y) = (f^*)^+ = f^* \)

(ii) \( E(f^+|X) = (f^+)^* = f^* \)

(iii) \( f^+ \subseteq X \rightarrow f^+ = f^* \)

The following result which is a restatement of the property of self-adjointness of the \( * \)-operator will be repeatedly used in the sequel.

PROPOSITION 2

If \( f \in L_\infty \), \( g \in L_\infty \), and \( h \in L_\infty(X) \), then
\[ E(f^*gh) = E(fg^*h) = E(f^*g^*h). \text{ (*) stands for } ^*X. \]

The proof follows from the observation that \((f^*gh)^* = f^*g^*h\) and that \(E(f) = E(f^*)\) for every \(f \in L_\infty\).

This proposition together with the fact that the \(^*\)-operator is idempotent (that is, \((f^*)^* = f^*\)) implies that the \(^*\)-operator is a projection of \(L_\infty\) in \(L_\infty(X)\) when the \(L_2\) norm is considered.

Given two random objects \(X\) and \(Y\), the random object \((X, Y)\): \(\Omega \to X \times Y\) generates the subfield \(F_X \vee F_Y\) and may be identified with its completion; that is, \(\overline{F_{(X,Y)}} = \overline{F_X \vee F_Y}\). A random object that essentially generates the subfield \(F_X \cap F_Y\) will be denoted in this chapter by \(X \wedge Y\) despite the fact that it does not have a neat representation in terms of \(X\) and \(Y\) as in the case of \((X, Y)\).

**REMARK**

Given any two subfields \(F'\) and \(F''\) of \(F\), the following are well known relations among completed subfields:

(i) \(\overline{F' \vee F''} = \overline{F'} \vee \overline{F''} = \overline{F'} \vee F''\).

(ii) \(\overline{F' \cap F''} \subset \overline{F' \cap F''} = \overline{F'} \cap \overline{F''}\).

The following definition and theorem due to Dynkin are of great importance. They enable us to present simple proofs of some of the results stated in the sequel.

**DEFINITION 2**

Let \(\mathcal{D}\) be a class of subsets of \(\Omega\). \(\mathcal{D}\) is said to be a \(D\)-system (\(D\) for Dynkin) if the following conditions hold:
(i) \( \Omega \in \mathcal{D} \).

(ii) If \( B, A \in \mathcal{D}, B \subseteq A \) then \( A - B \in \mathcal{D} \).

(iii) If \( A_1, A_2, \ldots \in \mathcal{D} \) and \( A_n \uparrow A \) then \( A \in \mathcal{D} \).

**Theorem 1**

Let \( \mathcal{C} \) be a class of subsets of \( \Omega \) and assume that \( \mathcal{C} \) is closed under finite intersections. If \( \mathcal{D} \) is a \( \mathcal{D} \)-system such that \( \mathcal{C} \subseteq \mathcal{D} \) then \( \sigma(\mathcal{C}) \subseteq \mathcal{D} \). (\( \sigma(\mathcal{C}) \) is the smallest \( \sigma \)-field that contains \( \mathcal{C} \).)

For a proof of this result we refer to Ash (1972) pp. 168-169. For applications see Basu (1967).

In the next section we discuss the concept of conditional independence.

**3 - Definition of Conditional Independence**

In this section, the two most popular definitions of conditional independence (c.i.) are discussed. They are called here Intuitive and Symmetric. A simple proof of the equivalence between them is presented. Further characterization of the concept of c.i. will be presented in Section 4.

Three random objects \( X, Y, \) and \( Z \) are being considered and, in this section, \( * \) stands for the \( *Z \)-operator.

**Definition 3 - (Intuitive)**

The random objects \( X \) and \( Y \) are **conditionally independent given** \( Z \) (in symbols \( X \perp \! \! \! \! \! \! \perp Y \mid Z \) if for any \( f \in L_\infty(X) \)
\[ E\{f(Y, Z)\} = f^*(Y, Z) = f^* \]

Note that if \( X, Y, \) and \( Z \) are random variables, then to say that \( X \perp\!
\!
\!
\!
\!
\perp Y \mid Z \) is equivalent to say that \( X \mid (Y, Z) \) has the same conditional distribution as does \( X \mid Z \). This is the intuition behind Definition 3. Frequently we will use the notation \( X \mid(Y, Z) \sim X \mid Z \) for \( X \perp\!
\!
\!
\!
\!
\perp Y \mid Z \).

An equivalent way to define c.i. is to say that the map \( f \to f^*(Y, Z) \) from \( L_\infty \) to \( L_\infty(Y, Z) \) has its range restricted to \( L_\infty(Z) \). Particularly, if \( Z \) is essentially a generator of \( F_0 \) (the trivial subfield), then \((Y, Z) \subset Y\) and the usual concept of independence is attained since \( L_\infty(Z) \) becomes the class of all essentially constant functions. In this case the notation is \( X \perp\!
\!
\!
\!
\!
\perp Y \).

**Definition 3a (Symmetric)**

The random objects \( X \) and \( Y \) are conditionally independent given \( Z \) if for any \( f \in L_\infty(X) \) and \( g \in L_\infty(Y) \),

\[(fg)^* = f^*g^* \]

The following theorem gives the equivalence of the two definitions showing that \( X \perp\!
\!
\!
\!
\!
\perp Y \mid Z \) implies \( Y \perp\!
\!
\!
\!
\!
\perp X \mid Z \) which is not clear by looking only at Definition 3.

**Theorem 2**

Definitions 3 and 3a are equivalent.
PROOF

3 → 3a

By using Proposition 1 and the linearity of the *-operator we have:

\[(fg)^* = E\{E[f|Y, Z]|Z\} = E\{gE\{f|Y, Z\}|Z\} = E\{gE\{f|Z\}|Z\} = (gf^*)^* = f^*g^*\]

3a + 3

We wish to prove that for any \( f \in X \) and \( g \in Y \), \((fg)^* = f^*g^*\) implies \( E[f|(Y, Z)] = f^* \).

Let \( E \) be a class of subsets defined as \( E = \{E \in \mathcal{F}_Y \vee \mathcal{F}_Z \mid \int_E fd\Omega = \int f^*d\Omega \ \forall \ f \in X\} \). Clearly \( E \) is a D-system since \( \Omega \in E \), \( E \) is a monotone class (by monotone convergence theorem) and for \( A, B \in E \) with \( A \subseteq B \) we have \( B - A \in E \).

Now take any two sets \( C \) and \( D \) with \( C \in \mathcal{F}_Y \) and \( D \in \mathcal{F}_Z \). Clearly \( CD \in \mathcal{F}_Y \vee \mathcal{F}_Z \) and

\[\int_{CD} fd\Omega = E\{I_CI_Df\} = E\{(I_CI_D)f^*\} = E\{I_D(I_Cf)^*\}.\]

But by Proposition 2 and by hypothesis, we have

\[E\{I_D(I_Cf)^*\} = E\{I_DI_Cf^*\} = E\{I_Df_C^*\} = \int_{CD} f^*d\Omega.\]

Thus, \( E' \subset E \) where

\[E' = \{CD; \ C \in \mathcal{F}_Y \text{ and } D \in \mathcal{F}_Z\}.\]
Since $E'$ is closed under finite intersections, and
$\sigma(E') = F_Y \vee F_Z$ we conclude, by Theorem 1, that $F_Y \vee F_Z \subset E'$; that is, $f^* = \mathbb{E}(f|Y, Z) \vee f \subset X$. □

An important case of c.i. is $X \Perp Y|X'$ where $X' \subset X$. Note that the meaning of this relation is better understood when stated as

$$\forall g \subset Y, \mathbb{E}(g|X) = \mathbb{E}(g|X')$$

since $X \equiv (X, X')$. In Bayesian inference, if $X$ represents the sample, and $Y$ the parameter then $X'$ is said to be sufficient for $X$.

Some applications of the concept of c.i. are presented in the sequel and emphasis is given to the Bayesian framework.

4 - THE DROP/ADD PRINCIPLES AND OTHER PROPERTIES OF CONDITIONAL INDEPENDENCE

The concept of c.i. gives rise to many questions. Among them are questions involving the DROP and ADD (DROP/ADD) principles. Suppose that $X, Y, Z, W, X_1, \text{ and } Z_1$ are random objects such that $X \Perp Y|Z, X_1 \subset X, \text{ and } Z_1 \subset Z$. What can be said about the relation $\Perp$ if $X_1$ is substituted for $X, Z_1$ for $Z, (Y, W)$ for $Y$, or $(Z, W)$ for $Z$? In other words, can $F_X, F_Y, \text{ or } F_Z$ be essentially reduced or enlarged without destroying the c.i. relation? In general, the answer is no. However, for certain kinds of reductions and enlargements, the relationship will be preserved. To indicate that the relation $\Perp$ does not hold we write $\not\Perp$. 
The following simple examples show that arbitrary enlargements of $F_X$, $F_Y$, or $F_Z$ may destroy the c.i. property. For a set $A \subseteq \Omega$, $I_A(\omega)$ is the indicator function of $A$.

EXAMPLE 1

Let $\Omega = \{1, 2, 3, 4\}$, $F$ be the power set, and $\Pi(i) = 1/4$. Let $X = I_{\{1,2\}}$, $Y = I_{\{1,3\}}$, $Z = \text{constant}$, and $W = I_{\{1,4\}}$. Clearly, $X \parallel Y$ and $X \parallel W$ but $X \not\parallel (Y, W)$. □

EXAMPLE 2

Let $\Omega = \{0, 1\} \times \{0, 1\} \times \{0, 1\}$, $F$ be the power set, and for $i \neq j$ $(i, j = 0, 1)$ $\Pi\{(i, i, i)\} = .15$, $\Pi\{(i, i, j)\} = .10$ and $\Pi\{(i, j, i)\} = .25$. If $X$, $Y$, $Z$, and $W$ are such that $X(x, y, w) = x$, $Y(x, y, w) = y$, $W = (x, y, w) = w$, and $Z$ is a constant in $\Omega$, then $X \parallel Y$ and $X \not\parallel Y|W$. This is clear since we obtain the following probability functions (p.f.):

\[
\begin{array}{ccc}
  & 0 & 1 \\
 X \quad 0 & .3 & .5 \\
 1 & .2 & .2 \\
\end{array}
\]

\[
\begin{array}{ccc}
  & 0 & 1 \\
 Y \quad 0 & .2 & 0 \\
 1 & .5 & .3 \\
\end{array}
\]

p.f. of $(X, Y)$ given $W = 0$

\[
\begin{array}{ccc}
  & 0 & 1 \\
 X \quad .3 & .7 \\
\end{array}
\]

p.f. of $(X, Y)$ given $W = 1$
Y

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>.25</td>
<td>.25</td>
</tr>
</tbody>
</table>

X

<table>
<thead>
<tr>
<th>1</th>
<th>.25</th>
<th>.25</th>
<th>.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>.5</td>
<td>.5</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

p.f. of \((X, Y)\). □

**EXAMPLE 3**

Suppose that \(X\) and \(Y\) are two independent random variables with the same distribution \(N(0, 1)\). Clearly, \((X - Y) \uparrow (X + Y)\).\(\mid Y\).

However, it is well known that \((X - Y) \perp (X + Y)\). □

Looking at the problem from the opposite direction, we present the following similar examples which show that arbitrary reductions of the conditioning subfield may destroy the c.i. relation.

**EXAMPLE 4**

In Example 2 consider \(\Pi\) as follows:

\[
\Pi\{(0, 0, 0)\} = \Pi\{(0, 0, 1)\} = \Pi\{(0, 1, 0)\} = \Pi\{(1, 0, 1)\} = .10, \text{ and}
\]

\[
\Pi\{(1, 1, 1)\} = \Pi\{(1, 1, 0)\} = \Pi\{(1, 0, 0)\} = \Pi\{(0, 1, 1)\} = .15.
\]

Here, we conclude that \(X \perp Y \mid W\), but \(X \uparrow Y\). The probability functions in this case are:
\[
\begin{array}{c|cc|c|cc|c}
Y & \text{0} & \text{1} & \text{Y} & \text{0} & \text{1} & \text{Y} \\
\hline
\text{0} & .2 & .2 & .4 & \text{0} & .2 & .3 & .5 \\
\text{1} & .3 & .3 & .6 & \text{1} & .2 & .3 & .5 \\
\hline
.5 & .5 & 1 & .4 & .6 & 1
\end{array}
\]

p.f. of \((X, Y)\) given \(W = 0\).  
p.f. of \((X, Y)\) given \(W = 1\).

\[
\begin{array}{c|cc|c}
Y & \text{0} & \text{1} & \text{Y} \\
\hline
\text{0} & .20 & .25 & .45 \\
\text{1} & .25 & .30 & .55 \\
\hline
.45 & .55 & 1
\end{array}
\]

p.f. of \((X, Y)\).  

**EXAMPLE 5**

In example 3 consider an additional random variable \(Z\) such that \(Z \perp \perp (X - Y)\) and \(Z \perp \perp (X + Y)\). Obviously, \((X - Y + Z) \perp \perp (X + Y + Z)\). \(\Box\)

Examples 2 to 5 can be viewed as cases of Simpson's paradox (Dawid \[1979a\]). The paradox, however, is much stronger. For instance, let \(Z\) and \(W\) be two independent normal variables with zero means. Define \(X = Z + W\) and \(Y = Z - W\). The correlation between \(X\) and \(Y\) is given by \(\rho(X, Y) = \frac{1 - \delta}{1 + \delta}\) where \(\delta = \frac{\text{Var}(W)}{\text{Var}(Z)}\). Given \(Z\), the
conditional correlation $\rho(X, Y|Z)$ is clearly equal to $-1$. On the other hand, $\delta$ may be taken very small in order to make $\rho(X, Y)$ close to 1. This shows that we can have a case where $X$ and $Y$ are strongly positive (negative) dependent but, when $Z$ is given, $X$ and $Y$ turn to be strongly negative (positive) dependent.

The essence of DROP/ADD principles for conditional independence is contained in the following proposition and corollaries.

**PROPOSITION 3**

If $X \!\! \perp \!\! \perp Y|Z$ then for every $X' \subseteq X$ we have:

(i) $X' \!\! \perp \!\! \perp Y|Z$.

(ii) $X \!\! \perp \!\! \perp Y|(Z, X')$.

**PROOF**

(i) Since $X' \subseteq X$, $\forall f \subseteq X' \Rightarrow f \subseteq X$. Then, for every $f \subseteq X'$, since $X \!\! \perp \!\! \perp Y|Z$, $E\{f|Y, Z\} = E\{f|Z\}$.

(ii) Clearly, $(Z, X', X) \equiv (Z, X)$ then, for every $g \subseteq Y$,

$$E\{g|(Z, X', X)\} = E\{g|(Z, X)\} = E\{g|Z\} = g^*.$$ 

On the other hand, by Proposition 1,

$$E\{g|(Z, X')\} = E\{E\{g|(Z, X', X)\}|(Z, X')\} = E\{g^*|(Z, X')\} = g^*$$

Thus, $\forall g \subseteq Y$ $E\{g|(Z, X')\} = E\{g|(Z, X', X)\}$. $\Box$
COROLLARY 1

For any $Z' \subset Z$, $X \perp\!
\!
\!
\perp Y \mid Z$ if and only if $X \perp\!
\!
\!
\perp (Y, Z') \mid Z$.

COROLLARY 2

If $X \perp\!
\!
\!
\perp Y \mid Z$ then, for any $W_1 \subset (X, Z)$ and $W_2 \subset (Y, Z)$, we have:

(i) $W_1 \perp\!
\!
\!
\perp W_2 \mid Z$

(ii) $X \perp\!
\!
\!
\perp Y \mid (Z, W_1, W_2)$.

By way of explanation, if $X \perp\!
\!
\!
\perp Y \mid Z$ then the relation $\perp\!
\!
\!
\perp$ is preserved when (i) $X$ and $Y$ is increased (ADD) by any essential part of $Z$, (ii) $Z$ is increased (ADD) by any essential part of $X$ or of $Y$, and (iii) $X$ and $Y$ are arbitrarily reduced (DROP).

The following interesting result, in one direction, has its version in classical statistics. If $X_0$ is sufficient for $X$ then, for every statistic $f$, there is a corresponding function $g$ of $X_0$ with the same mean of $f$.

PROPOSITION 4

Let $X'$, $X$, and $Y$ be three random objects such that $X' \subset X$. The following condition is necessary and sufficient to have $X \perp\!
\!
\!
\perp Y \mid X'$:

$\forall f \subset X, E(f^* \mid Y) = E(f \mid Y)$, where $f^* = E(f \mid X')$. 
PROOF

Here, * stands for *X' and † for *Y.

(i) Necessity.

Since ∀f ∈ X, f* = E{f|Y, X'}, by Proposition 1 we conclude that ∀f ∈ X, (f*)† = f†.

(ii) Sufficiency.

Let f ∈ X, g ∈ Y, and f' ∈ X'. Clearly ff' ∈ X. Note that

(fgf')† = g(ff')† = g(ff')*† = g(f*f')† = (f*g'f)†.

Since E{(fgf')†} = E{fgf'}, by Proposition 2 we can write

E{fgf'} = E{f*g'f} = E{f*g*Y}

Then (fg)* = f*g*. □

An equivalent result introduced by Mouchart and Rolin (1978), which is stated below, is a characterization of c.i.

COROLLARY 3

The following condition is necessary and sufficient to have

X ⊥⊥ Y|Z:

∀f ∈ (X, Z), E{f*|Y} = E{f|Y}, where f* = E{f|Z}.

The equivalence of this result with Proposition 4 follows directly from Corollary 1.
A useful result in statistical applications by Dawid (1979a), is stated as follows:

**PROPOSITION 5**

The following properties are equivalent:

(i) \( X \perp Y \mid Z \) and \( X \perp W \mid (Y, Z) \).

(ii) \( X \perp (Y, W) \mid Z \).

**PROOF**

(i) \( \rightarrow \) (ii)

From (i) we have that \( X \mid (W, Y, Z) \sim X \mid (Y, Z) \sim X \mid Z \). Then, \( X \mid (W, Y, Z) \sim X \mid Z \) or equivalently, \( X \perp (W, Y) \mid Z \).

(ii) \( \rightarrow \) (i)

By Proposition 3, we conclude that \( X \perp Y \mid Z \) and \( X \perp (Y, W) \mid (Z, Y) \) which implies \( X \perp W \mid (Z, Y) \). \( \Box \)

Note that since property (ii) is symmetric (\( Y \) and \( W \) may commute), the inclusion of the following property is implicit: (iii) \( X \perp W \mid Z \) and \( X \perp Y \mid (Z, W) \). The corollary below is an example of a kind of result we may prove by using the equivalence between (i) and (iii).

**COROLLARY 4**

For \( T \subset (X, Z, W) \), if \( X \perp Y \mid (Z, W) \) and \( T \perp W \mid Z \), then \( T \perp Y \mid Z \) and \( T \perp W \mid (Y, Z) \).

This result is better understood in the Bayesian context when \( X \) represents the sample, \( (T, W) \subset X \), \( Y \) represents the parameter, and \( Z \) is essentially a constant (\( Z \) is a generator of \( F_0 \)). We might say
that if $W$ is a sufficient statistic and $T$ is a statistic "marginally independent" of $W$ ($T \perp \perp W$), then $T$ is ancillary ($T \perp \perp Y$) and is "independent" of $W(T \perp \perp W|Y)$.

Now we extend the concept of conditional independence for a set of random objects. Let $Z$ be a random object, $\tau$ be a set of indices, and $\{X_t; t \in \tau\}$ be a collection of random objects.

**DEFINITION 4**

The set $\{X_t; t \in \tau\}$ is said to be mutually conditionally independent given $Z$ if, for any partition $(\tau_1, \tau_2)$ of $\tau$, the two random objects $\{X_t; t \in \tau_1\}$ and $\{X_t; t \in \tau_2\}$ are conditionally independent given $Z$.

For example, $X_1$, $X_2$, and $X_3$ are mutually conditionally independent given $Z$ if $X_1 \perp \perp (X_2, X_3)|Z$, $X_2 \perp \perp (X_1, X_3)|Z$, and $X_3 \perp \perp (X_1, X_2)|Z$.

The next result is called here the transfer principle for c.i.. It shows that, for finite sets of random objects, to check Definition 4 we do not have to study all partitions.

**PROPOSITION 6**

If $X_1 \perp \perp X_2|Z$ and $(X_1, X_2) \perp \perp X_3|Z$, then $X_1 \perp \perp (X_2, X_3)|Z$.

**PROOF**

By DROP/ADD principles

$$(X_1, X_2) \perp \perp X_3|Z \Rightarrow X_1 \perp \perp X_3|(Z, X_2).$$
By Proposition 5, \( X_1 \perp \perp X_2 | Z \) and \( X_1 \perp \perp X_3 | (Z, X_2) \) hold if and only if \( X_1 \perp \perp (X_2, X_3) | Z \).

It is clear now that to check Definition 4 for a finite set of random objects, say \( X_1, \ldots, X_n \), we need only check that

\[
(X_1, \ldots, X_k) \perp \perp X_{k+1} | Z
\]

for every \( k = 1, 2, \ldots, n - 1 \).

To extend this result to the countable case, we prove the following theorem which is called the limiting property of c.i.

It will be applied in a characterization of Markov Chains presented in Section 5.

**Theorem 3**

Let \( Z, X, Y_1, Y_2, \ldots \) be random objects such that

\[
X \perp \perp (Y_1, Y_2, \ldots, Y_n) | Z \text{ for every } n = 1, 2, \ldots.
\]

Then,

\[
X \perp \perp (Y_1, Y_2, \ldots) | Z \text{ where } (Y_1, Y_2, \ldots) \text{ essentially is the generator of } \bigvee_{n=1}^{\infty} F_n = \sigma(\bigvee_{n=1}^{\infty} F_n). \text{ (Here, } F_n = F_{y_n}. )
\]

**Proof**

Since \( \bigvee_{n=1}^{\infty} F_n \) is a field, it is closed under finite intersections.

Let \( * \) stand for \( \perp \perp Z \) and consider the set

\[
E = \{ f; E \in \bigvee_{n=1}^{\infty} F_n \text{ and } (L_E)^* = \bigvee_{n=1}^{\infty} F_n \} \subset X.
\]

The following conditions show that \( E \) is a D-system:
a) $\Omega \in \mathcal{E}$

b) For $E_1, E_2 \subset \mathcal{E}$, if $E_1 \subset E_2$ then

$$(I_{E_2 - E_1} f^*)^* = (I_{E_2} f - I_{E_1} f)^* = I_{E_2}^* f^* - I_{E_1}^* f^* = (I_{E_2} - I_{E_1})^* f^*$$

$$= I_{(E_2 - E_1)}^* f^*. $$

That is, $E_2 - E_1 \in \mathcal{E}$.

c) For any monotone sequence $E_1, E_2, \ldots$, of $\mathcal{E}$, we have that $\lim_{n} I_{E_n} = I_{\lim E_n}$ and by the dominated convergence theorem for conditional expectation, $(\lim I_{E_n} f)^* = \lim(I_{E_n} f)^*$. Since $E_n \subset \mathcal{E}$,

$$(\lim I_{E_n} f)^* = \lim I_{E_n}^* f^* = I_{\lim E_n}^* f^*. $$

That is, $\lim E_n \in \mathcal{E}$.

To conclude the proof recall that, by hypothesis,

$$\bigvee_{n=1}^{\infty} F_n \subset \mathcal{E}$$

and then by Theorem 1,

$$\bigvee_{n=1}^{\infty} F_n \subset \mathcal{E}. \blacksquare$$

To conclude this section we extend Proposition 6 to the countable case.

**PROPOSITION 7**

Let $Z, X_1, X_2, \ldots$ be a sequence of random objects such that, for each $n = 1, 2, \ldots$, $(X_1, \ldots, X_n) \perp\!
\!
\perp X_{n+1} \mid Z$. Then $(X_1, X_2, \ldots)$ is mutually conditionally independent given $Z$. 
PROOF

Let \((i_1, i_2, \ldots), (j_1, j_2, \ldots)\) be a partition of the set \(\{1, 2, \ldots\}\). We wish to prove that the relation

\[(X_{i_1}, X_{i_2}, \ldots) \perp \perp (X_{j_1}, X_{j_2}, \ldots) | Z\]

holds. Note that, for any \(k, \ell \in \{1, 2, \ldots\}\) the finite relation

\[(X_{i_1}, X_{i_2}, \ldots, X_{i_k}) \perp \perp (X_{j_1}, X_{j_2}, \ldots, X_{j_\ell}) | Z\]

holds. This follows from the discussion after Proposition 6 and from the fact that \((X_1, \ldots, X_m) \perp \perp X_{m+1} | Z \forall m = 1, 2, \ldots, v, \text{ where} v = \max(i_1, \ldots, i_k, j_1, \ldots, j_\ell)\). By Theorem 3 it follows that

\[(X_{i_1}, \ldots, X_{i_k}) \perp \perp (X_{j_1}, X_{j_2}, \ldots) | Z.\]

Finally, applying again Theorem 3 we prove our claim. \(\square\)

We write \(X_1 \perp \perp X_2 \perp \perp \ldots | Z\) or \(\bigwedge_{n=1}^{\infty} X_n | Z\) to indicate that the sequence \((X_1, X_2, \ldots)\) is mutually conditionally independent given \(Z\).

The next section presents some applications of c.i. in Bayesian statistics and in a characterization of Markov chains.

5 - MARKOV CHAINS AND BAYESIAN INFERENCE

As discussed in Dawid (1979a, 1980), many of the important statistical concepts are simply manifestations of the concept of conditional independence. In this section we use the framework of c.i. to study a well known characterization of the Markov Chain property and to describe the Bayesian version of those statistical concepts and their properties.

The following is the usual definition of Markov Chain.
DEFINITION 5

A sequence of random objects, $X_1, X_2, \ldots$ is said to form a Markov Chain if,

\[(5.1) \quad \forall n \geq 1, (X_1, \ldots, X_n) \upharpoonright X_{n+2} | X_{n+1}.\]

This concept is better understood when the relations (5.1) are replaced by,

\[(5.2) \quad \forall n \geq 1, (X_1, \ldots, X_n) \upharpoonright (X_{n+2}, X_{n+3}, \ldots) | X_{n+1}.\]

Here, if the indices represent time we might say that the past is independent of the future given the present. The following proposition states the equivalence among (5.1) and (5.2).

PROPOSITION 8

The sequence $X_1, X_2, \ldots$ of random objects forms a Markov Chain if and only if the relations (5.2) are satisfied.

PROOF

(5.2) $\Rightarrow$ (5.1) Follows directly from Proposition 3.

(5.1) $\Rightarrow$ (5.2)

Step 1 - First we wish to prove that

\[\forall n \geq 1, (X_1, \ldots, X_n) \upharpoonright (X_{n+2}, X_{n+3}) | X_{n+1}.\]

But, using DROP/ADD principles, (5.1) implies that
To prove property (a), it is enough to have

**Proposition 9**

If \((X_1, X_2, X_3, X_4, X_5)\) forms a Markov Chain, then

\[ X_1 \perp \perp X_5 \mid (X_2, X_4). \]

**Proof**

By hypothesis, \( X_1 \perp \perp (X_3, X_4) \mid X_2 \) and \( (X_1, X_2, X_3) \perp \perp X_5 \mid X_4. \)
By PROP/ADD principles this implies that \( X_1 \perp \perp (X_3, X_4) \mid (X_2, X_3) \) and \( X_3 \perp \perp X_5 \mid (X_2, X_4). \) The conclusion follows directly from Proposition 6. \( \Box \)

To conclude our discussion on the concept of Markov Chains, we notice that Definition 5 can be generalized by considering an additional random object \( Z \) in the conditioning random objects of (5.1). That is, in the place of (5.1) consider the relations

\[ \forall n \geq 1, (X_1, \ldots, X_n) \perp \perp X_{n+2} \mid (Z, X_{n+1}). \]

In this case, we say that \((X_1, X_2, \ldots)\) form a conditional Markov Chain given \( Z \). It is clear that we could have a similar discussion for this general concept. Finally, we notice that if \((X_1, X_2, \ldots)\) forms a conditional Markov Chain given \( Z \) and \( \forall n \geq 1, X_{n+2} \perp \perp Z \mid X_{n+1} \) then, \((X_1, X_2, \ldots)\) forms a Markov Chain. This is a direct application of Proposition 5.

In order to focus our attention on applications in Bayesian statistics, it is important to review some of the structures involved.

Let \((X, A)\) be the usual sample space and \(\{P_\theta, \theta \in \Theta\}\) be a family of probability measures on \((X, A)\) where \(\Theta\) is the usual
parameter "space". In addition, the Bayesians consider a (prior) probability space \((\Theta, \mathcal{B}, \xi)\) where \(\mathcal{B}\) is a \(\sigma\)-algebra of subsets of \(\Theta\) such that \(P_{\theta}(A)\) is a \(\mathcal{B}\)-measurable function for every fixed \(A \in A\). Clearly, the choice of the prior model is not completely arbitrary, since it has to match the statistical structure on the \(\mathcal{B}\)-measurability of \(P_{\theta}(A)\).

After all these considerations, it becomes clear that we can restrict ourselves to the probability space \((\Omega, F, \Pi)\), where now \(\Omega = \Theta \times X\), \(F = \mathcal{B} \times A\) and \(\Pi\) is defined as

\[
\Pi(F) = \int_{\Theta} P_{\theta}(F[\theta]) \xi(d\theta)
\]

for every \(F \in F\) where \(F[\theta] = \{x \in X; (\theta, x) \in F\}\). Note that if \(A \in A\) and \(B \in \mathcal{B}\), then

\[
\Pi(B \times A) = \int_{B} P_{\theta}(A) \xi(d\theta).
\]

The uniqueness of \(\Pi\) and the fact that \(\Pi\) is a probability measure are included in Theorem 2.6.2 of Ash (1972). Now, we can define a ( marginal ) probability measure \(P\) on \((X, A)\) in the following way:

\[
P(A) = \Pi(\Theta \times A)
\]

for every \(A \in A\).

Let \(X\) and \(Y\) be two random objects on \((\Omega, F)\). We say that \(X\) represents the sample and \(Y\) represents the parameter if
\[ F_X \equiv \{ \emptyset \times A; A \in A \} \quad \text{and} \quad F_Y \equiv \{ B \times X; B \in B \} \].

In addition to \( X \) and \( Y \) as defined above, consider two random objects \( X_1 \) and \( X_2 \) such that \( (X_1, X_2) \subset X \). The Bayesian version of the concepts of sufficiency and ancillarity is contained in the following.

**DEFINITION 6**

a) If \( X \perp \!
\!
\!\perp Y \mid X_1 \) we say that \( X_1 \) is **sufficient** for \( X \) with respect to \( Y \).

b) If \( X_2 \perp \!
\!
\!\perp Y \mid \) we say that \( X_2 \) is **ancillary** with respect to \( Y \).

The classical concept of statistical independence between \( X_1 \) and \( X_2 \) has its Bayesian version as:

c) \( X_1 \perp \!
\!
\!\perp X_2 \mid Y \).

Basu (1955, 1958) speculates under what conditions two of the three relations a), b), and c) imply the third. One of the objectives of this **article** is to study Basu's theorems under the Bayesian framework. The next result which is Basu's first conjecture presents conditions to have b) and c) implying a).

**PROPOSITION 10**

If in addition to \( X_2 \perp \!
\!
\!\perp Y \) and \( X_1 \perp \!
\!
\!\perp X_2 \mid Y \) we have \( X \perp \!
\!
\!\perp Y \mid \left(X_1, X_2\right) \), then \( X \perp \!
\!
\!\perp Y \mid X_1 \).
PROOF

By Proposition 5 we have that:

i) \( X_2 \perp \perp Y \) and \( X_1 \perp \perp X_2 | Y \) if and only if \( X_1 \perp \perp X_2 \) and \( X_2 \perp \perp Y | X_1 \).

ii) \( X_2 \perp \perp Y | X_1 \) and \( X \perp \perp Y | (X_1, X_2) \) if and only if \( X \perp \perp Y | X_1 \) since \( (X_2, X) \equiv X. \)

Looking at the above proof, we see that if \( X_1 \perp \perp X_2 \), then a) implies b) and c). The meaning of the relation \( X_1 \perp \perp X_2 \) in classical statistics, however, is void.

Note that Proposition 10 gives conditions for reducing (DROP) the conditioning random object. Actually, all of Basu's theorems are cases of DROP/ADD principles. Basu's other theorems are discussed in the next sections.

Another type of reduction of the conditioning random object is presented in the proposition below which is a Bayesian version of a theorem introduced by Burkholder (1961).

PROPOSITION 11

Let \( X_0 \) and \( X_1 \) be two random objects such that \( (X_0, X_1) \subset X \), \( X \perp \perp Y | X_0 \) and \( X \perp \perp Y | X_1 \). Then \( X \perp \perp Y | X_0 \land X_1 \). (If \( X_0 \) and \( X_1 \) are sufficient for \( X \), then so is \( X_0 \land X_1 \).)

The proof follows directly from the definition of c.i.. As an important consequence of this proposition we have the following result which was introduced by Dawid (1979b).
PROOF OF LEMMA 1

Here we use * for X and † for Y. Let B and C be two sets such that \( I_B \subset X \) and \( I_C \subset Y \). Using the properties of conditional expectation and the fact that \( X \perp \perp Y \) we have that,

\[
\int_{BC} I^* I^{†} d\Pi = \int_{B} I^* I_{AC}^{†} d\Pi = \int_{I_{AB}} I^* I_{AC}^{†} d\Pi =
\]

\[
[\int_{I_{AB}} ^* d\Pi ][\int_{I_{AC}} ^{†} d\Pi ] = [\int_{I_{AB}} ^* d\Pi ][\int_{I_{AC}} d\Pi ].
\]

That is, since \( \Pi(A) > 0 \),

\[
\int_{BC} I^* I^{†} d\Pi = \frac{\Pi(AB) \Pi(AC)}{\Pi(A)} = [\Pi(A)]^2 \Pi(B \mid A) \Pi(C \mid A),
\]

where \( \Pi(B \mid A) = \frac{\Pi(AB)}{\Pi(A)} \). We notice now that on the atom A, the functions \( I^*_B \) and \( I^*_C \) are constants and equal respectively to \( \Pi(B \mid A) \) and \( \Pi(C \mid A) \). Analogously, the function \( I^*_Z \) is equal to \( \Pi(BC \mid A) \) on A. On the other hand since \( X \perp \perp Y \), \( I^*_B I^*_C = I^*_Z \); thus \( \Pi(B \mid A) \Pi(C \mid A) = \Pi(BC \mid A) \). This shows that

\[
\int_{BC} I^* I^{†} d\Pi = \Pi(A) \Pi(ABC).
\]

To conclude the proof we must prove that \( \Pi(A) \int_{D} I_{A} d\Pi = \int_{D} I^* I^{†} d\Pi \) for every D such that \( I_D \subset (X, Y) \). Following the same technique used in Theorem 2 and 3 we obtain this as a consequence of Theorem 1.

PROOF OF PROPOSITION 12

Let \( p = \Pi(A) \), \( I^* = E(I_A \mid X) \), and \( I^{†} = E(I_A \mid Y) \). From Lemma 1 we have that
\[ E(I_A^* | (X, Y)) = \frac{I^* I^+}{p} \text{ and } E(I_A^{ac} | (X, Y)) = \frac{(1 - I^*)(1 - I^+)}{1 - p}. \]

Clearly
\[ \frac{I^* I^+}{p} + \frac{(1 - I^*)(1 - I^+)}{1 - p} = 1; \]

that is,
\[ (1 - \frac{I^*}{p})(1 - \frac{I^+}{p}) = 0. \]

Since \( X \perp Y \), this last equation holds if and only if either \( \frac{I^*}{p} \equiv 1 \) or \( \frac{I^+}{p} \equiv 1 \) almost surely. \( \square \)

**Remark**

1. Let \( Y \equiv (Y_1, Y_2) \) represent the parameter and \( X \) represent the sample. If \( Y_1 \) and \( Y_2 \) are independent a priori and a posteriori (i.e., \( Y_1 \perp Y_2 \) and \( Y_1 \perp Y_2 | X \)), then from Lemma 1,
\[
E(I_A | Y) = \left[ \Pi(A) \right]^{-1} E(I_A | Y_1) E(I_A | Y_2),
\]
where \( A \) is a positive atom of \( X \). Note that if \( Y_1 \) and \( Y_2 \) are independent a priori, and \( X \) is a discrete random variable, then \( Y_1 \) and \( Y_2 \) are independent a posteriori if and only if (5.3) holds and (5.3) defines the likelihood function. This result is the discrete case of the theorem introduced in Section 9 of Basu (1977).

6 - **ON MEASURABLE SEPARABILITY OF RANDOM OBJECTS**

Basu (1955) stated that any statistic independent of a sufficient statistic is ancillary. Later on Basu (1958) presented a
counter-example and recognized the necessity of an additional condition (connectedness) on the family \( \{P_\theta : \theta \in \Theta\} \) of probability measures. Koehn and Thomas (1975) strengthened this result by introducing a necessary and sufficient condition on the family.

More recently Basu and Cheng (1979), generalizing results of Pathak (1975), showed the equivalence between these two conditions in Coherent Models.

In the scope of the present work, this question will be stated in terms of random objects. Suppose that \( X \) represents the sample and \( Y \) the parameter. The following theorem is a Bayesian version of the result of Koehn and Thomas (1975).

**THEOREM 4**

Let \( X_1 \subset X \) be a sufficient random object (i.e., \( X \perp\!
\!
\perp Y | X_1 \)). The random object \( Y \wedge X_1 \) is essentially a constant (i.e., \( F_{Y \wedge X_1} \equiv F_0 \)) if and only if \( X_2 \perp\!
\!
\perp Y \) whenever \( X_2 \subset X \) and \( X_1 \perp\!
\!
\perp X_2 | Y \) (i.e., \( X_2 \) is ancillary if \( X_1 \) and \( X_2 \) are statistically independent).

**PROOF**

→ See discussion following Corollary 4.

→ Take \( X_2 \) such that \( X_2 \equiv Y \wedge X_1 \). Since \( X_2 \subset Y \), \( X_1 \perp\!
\!
\perp X_2 | Y \).

Then by hypothesis \( X_2 \perp\!
\!
\perp Y \), which implies that \( X_2 \perp\!
\!
\perp X_2 \) since \( X_2 \subset Y \); that is, \( X_2 \equiv Y \wedge X_1 \) is essentially a constant. □
REMARKS

2 - The condition introduced by Koehn and Thomas (1975) is the non-existence of a splitting set. A set \( A \) in the sample space (i.e., \( A \subset X \)) is a splitting set if \( P_{\theta}(A) = 0 \) or 1 for all \( \theta \in \Theta \) and at least for a pair \( \{\theta_1, \theta_2\} \subset \Theta \), \( P_{\theta_1}(A) = P_{\theta_2}(A^C) = 1 \). In the Bayesian framework, since \( X \) represents the sample and \( Y \) the parameter, an analogous definition is as follows: A set \( A \) such that \( I_A \subset X \) is a splitting set if \( 0 < \Pi(A) < 1 \) and 
\[
E(I_A | Y) = E^2(I_A | Y).
\]
Let \( I_A^* = E(I_A | Y) \) and note that
\[
\{(I_A - I_A^*)^2\} = I_A^* - (I_A^*)^2.
\]
Thus, if \( A \) is a splitting set,
\[
E((I_A - I_A^*)^2) = 0;
\]
that is, \( I_A = I_A^* \). Then \( I_A \subset Y \) or equivalently \( I_A \subset Y \wedge X \). We conclude that the non-existence of a splitting set is equivalent to \( Y \wedge X \) being essentially a constant.

3 - Let \( X, Y, \) and \( Z \) be three random objects such that \( X \perp\!
\!
\!\perp Y | Z \). Since this is equivalent to \( (X, Z) \perp\!
\!
\!\perp (Y, Z) | Z \), with the same argument we use in the proof of Theorem 4, we can easily show that \( (X, Z) \wedge (Y, Z) \equiv Z \). Intuitively we would say that if \( X \perp\!
\!
\!\perp Y | Z \), then \( Z \) possesses all common information contained in both \( X \) and \( Y \).

The following result is a Bayesian solution for a two-parameter problem in inference. Suppose that the parameter \( Y \) is such that \( Y \equiv (Y_1, Y_2) \). Let \( X \) represent the sample, \( X_1 \subset X \) be specific sufficient with respect to \( Y_2 \), and \( X_2 \subset X \) be specific sufficient with respect to \( Y_1 \). That is, \( X \perp\!
\!
\!\perp Y_2 | (X_1, Y_1) \) and
X \perp Y \mid (X_2, Y_2). (See Basu (1978) for details on the notion of specific sufficiency.) The question here is under what conditions does the specific sufficiency of \( X_1 \) and \( X_2 \) imply the sufficiency of \( (X_1, X_2) \)?

**Proposition 13**

If \( (X_1, Y) \wedge (X_2, Y) \subset (X_1, X_2) \), then \( X \perp Y \mid (X_1, Y) \) and \( X \perp Y \mid (X_2, Y) \) imply \( X \perp Y \mid (X_1, X_2) \).

**Proof**

From DROP/ADD principles we have that \( X \perp Y \mid (X_1, Y) \) and \( X \perp Y \mid (X_2, Y) \). Thus, by Proposition 11,

\[
X \perp Y \mid (X_1, Y) \wedge (X_2, Y),
\]

and since \( (X_1, X_2) \subset X \), the result follows. \( \square \)

The following related result is a direct consequence of Proposition 5.

**Proposition 14**

If \( X \perp Y_2 \mid (X_1, Y) \) and \( X \perp Y_1 \mid (X_2, Y_2) \), then \( X \perp Y \mid (X_1, X_2) \) if and only if \( X \perp Y_1 \mid (X_1, X_2) \) [equivalently \( X \perp Y_2 \mid (X_1, X_2) \)].

Note that the condition \( X \perp Y_1 \mid (X_1, X_2) \) does not have an interpretation in classical statistics since distributions depend on both parameters \( Y_1 \) and \( Y_2 \). Our conjecture for a future work is that specific sufficiency of \( X_1 \) and \( X_2 \) implies sufficiency of
(X_1, X_2) if Y_1 and Y_2 are variation independent (i.e., the parameter space is the cartesian product of the domain of Y_1 by the domain of Y_2). (See Basu (1977) and Barndorff-Nielsen (1978) for details on the notion of variation independence.) Dawid (1979b) presented an example where (X_1, X_2) is not sufficient even though X_1 and X_2 are specific sufficient. In this example, however, the parameters are not variation independent.

The title of this section was motivated by the following:

**DEFINITION 7**

The random objects X and Y are said to be **measurably separated conditionally** on Z if (X, Z) \& (Y, Z) \equiv Z. When Z is essentially a constant we simply say that X and Y are measurably separated.

A large list of results related with this concept appears in Mouchart and Rolin (1978).

**7 - BASU THEOREM**

Basu (1955) proved that any ancillary statistic is statistically independent of any bounded complete sufficient statistic. The Bayesian analogous concept of boundedly completeness is the concept of strong identifiability (Dawid [1980] and Mouchart, and Rolin [1978]). The main objective of this section is to study this concept and present Basu's result under the Bayesian framework.

Let X and Y be two random objects. As before, we study some aspects of the linear maps L_\infty(Y) \& L(X) and L_\infty(X) \& L_\infty(Y), where \& is for \&X, and \& is for \&Y. Recall that for two random variables f_1 and f_2, by f_1 \neq f_2 we mean that \Pi(\omega; f_1(\omega) \neq f_2(\omega)) > 0.
DEFINITION 8

The map $L_\infty(X) \overset{\dagger}{\rightarrow} L_\infty(Y)$ is essentially one-one if $f_1^\dagger \neq f_2^\dagger$ whenever $(f_1, f_2) \subset X$ and $f_1 \neq f_2$. In this case we say that $X$ is strongly identified by $Y$ and write $X \ll Y$.

Clearly, $X \ll Y$ if and only if for $f \subset X$, $f^\dagger = 0$ implies $f = 0$. This shows intuitively that when $Y$ represents the parameter and $X$ the sample, Definition 8 is the Bayesian version of the concept of bounded completeness.

DEFINITION 9

The map $L_\infty(Y) \overset{*}{\rightarrow} L_\infty(X)$ is essentially onto if for every $f \subset X$ there is a $g \subset Y$ such that $g^* = f$.

The following result relates these two definitions.

PROPOSITION 15

If the map $L_\infty(Y) \overset{*}{\rightarrow} L_\infty(X)$ is essentially onto, then $X \ll Y$.

PROOF

Let $(f, h) \subset X$ and $f^\dagger = 0$. Since $*$ is essentially onto $\forall g \subset Y$ s.t. $g^* = h$. Then

$$E(fh) = E(fg^*) = E(fg) = E(f^\dagger g) = 0.$$  

Since $h$ is arbitrary, $f = 0$. □

Let $X_{[Y]}$ be the random object that generates the smallest subfield that contains all functions $g^*$ where $g \subset Y$. Note that $X_{[Y]} \subset X$. The following result shows that $X_{[Y]}$ may be viewed as the Bayesian minimal sufficient statistic.
PROPOSITION 16

(i) \( X \perp\!\!\!\!\!\!\!\!\!\!\perp Y | X_{[Y]} \)

(ii) If \( X_1 \subset X \) is such that \( X \perp\!\!\!\!\!\!\!\!\!\!\perp Y | X_1 \), then \( X_{[Y]} \subset X_1 \).

PROOF

(i) \( \forall g \subset Y, E\{g|X, X_{[Y]}\} = g^* \subset X_{[Y]} \) by definition.

(ii) \( \forall g \subset Y, E\{g|X, X_1\} = E\{g|X\} = E\{g|X_1\} \).

Then for every \( g \subset Y \), \( g^* \subset X_1 \). Since \( X_{[Y]} \) is the generator of the smallest subfield containing the functions \( g^* \), \( X_{[Y]} \subset X_1 \). \( \square \)

When \( X_{[Y]} \equiv X \), \( X \) is said to be identified by \( Y \) (Dawid [1980], and Mouchart and Rolin [1978]). The name strong identification was motivated by the following result:

PROPOSITION 17

If \( X \ll Y \), then \( X_{[Y]} \equiv X \).

PROOF

Note that \( X \perp\!\!\!\!\!\!\!\!\!\!\perp Y | X_{[Y]} \). Thus,

\[ \forall f \subset X, \quad E\{E\{f|Y, X_{[Y]}\}|Y\} = E\{E\{f|X_{[Y]}\}|Y\}. \]

For \( f^+ = E\{f|X_{[Y]}\} \) since \( X \ll Y \), we have that

\[ E\{(f - f^+)|Y\} = 0 \rightarrow f = f^+ \). Then \( \forall f \subset X, f \subset X_{[Y]} \) and \( X \equiv X_{[Y]} \). \( \square \)

The Bayesian version of the Basu Theorem is contained in the result below.
THEOREM 5

Let X, Y, and Z be three random objects. If $X \bot\bot Y$, $X \bot\bot Y | Z$, and $Z \ll Y$, then $X \bot\bot Z | Y$.

PROOF

Since $X \bot\bot Y | Z \iff X$, $E(f | Y, Z) = E(f | Z)$. On the other hand, since $X \bot\bot Y$, $E(f | Y) = E(f)$ but, by Proposition 1,


which implies that $E((E(f | Z) - E(f)) | Y) = 0$. Since $Z \ll Y$, $E(f) = E(f | Z)$ for every $f \in X$. That is, if $Z \ll Y$, then $X \bot\bot Y$ and

$X \bot\bot Y | Z$ implies $X \bot\bot Z$. Now, by Proposition 5 we have that

$X \bot\bot Y | Z$ and $X \bot\bot Z$ is equivalent to $X \bot\bot Z | Y$ and $X \bot\bot Y$. □

Note that to obtain the Basu Theorem we consider $X$ as the sample, $Y$ as the parameter, and $X_0$ and $X_1$ two random objects such that

$(X_0, X_1) \in X$, $X_0 \bot\bot Y$, $X \bot\bot Y | X_1$ and $X_1 \ll Y$. Clearly $X_0 \bot\bot Y | X_1$ and the result $X_0 \bot\bot X_1 | Y$ follows.

Lehmann and Scheffe (1950) proved that if a sufficient statistic is boundedly complete, then it is a minimal sufficient statistic. The Proposition below is the Bayesian version of this result.

PROPOSITION 18

Let $X_1$, $X$, and $Y$ be three random objects such that $X_1 \subset X$ and $X \bot\bot Y | X_1$. If $X_1 \ll Y$ then $X_1 \equiv X_{[Y]}$. 
PROOF

From Proposition 16, \( X_{[Y]} \subset X_1 \) and \( X \perp\!
\perp Y \mid X_{[Y]} \). Using Proposition 3 we can write (i) \( X_{[Y]} \perp\!
\perp Y \mid X_1 \), (ii) \( X_1 \perp\!
\perp Y \mid X_{[Y]} \) and (iii) \( X_1 \subset Y \). Let \( f \in X_1 \), and note that from (ii) and Proposition 1 we have

\[ E(f \mid Y) = E(E(f \mid X_{[Y]} ) \mid Y). \]

Since \( X_{[Y]} \subset X_1 \), \( E(f \mid X_{[Y]} ) \subset X_1 \). From (iii) we conclude that \( f = E(f \mid X_{[Y]} ) \subset X_{[Y]} \). Then every \( f \in X_1 \) implies \( f \subset X_{[Y]} \) which implies that \( X_1 \subset X_{[Y]} \). □

REMARK

4 - The concept of strong identifiability may be generalized as follows: \( X \) is strongly identified by \( Y \) conditionally on \( Z \) \((X \subset Y \mid Z)\) if for every \( f \in (X, Z) \), \( E(f \mid (Y, Z)) = 0 \) implies \( f = 0 \). Analogously, \( X \) is identified by \( Y \) conditionally on \( Z \) if

\[ (X, Z)_{[Y, Z]} \equiv (X, Z). \]

All the results of this section may be easily generalized by introducing a conditioning random object \( Z \) to each relation stated. For our future work we intend to relate these general results with the work of Dawid (1979c), Ferreira (1980), and Godambe (1980).
REFERENCES


