

CONDITIONAL INDEPENDENCE IN STATISTICS

By D. BASU¹

The Florida State University

and

CARLOS A. B. PEREIRA²

Universidade de São Paulo

SUMMARY. The theory of conditional independence is explained and the relations between ancillarity, sufficiency, and statistical independence are discussed in depth. Some related concepts like specific sufficiency, bounded completeness, and splitting sets are also studied in some details by using the language of conditional independence.

1. INTRODUCTION

The notion of conditional independence is a central theme in statistics. In a series of articles Dawid (1979a, 1979b; 1980), Florens and Mouchart (1977), and Mouchart and 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 in part of an expository nature.

The statistical perspective of this article is that of a Bayesian. A problem begins with a parameter (State of Nature) θ 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 statistics model $(\mathcal{X}, \mathcal{A}, \{P_\theta : \theta \in \Theta\})$. Writing $\omega = (\theta, X)$, $(\Omega, \mathcal{F}) = (\Theta \times \mathcal{X}, \mathcal{B} \times \mathcal{A})$, and Π for the joint distribution of (θ, X) , there then exists a subjective model $(\Omega, \mathcal{F}, \Pi)$ for ω . Hidden behind the wings of the Bayesian probability model $(\Omega, \mathcal{F}, \Pi)$ are the four models

- (i) the prior model $(\Theta, \mathcal{B}, \xi)$,
- (ii) the statistical model $(\mathcal{X}, \mathcal{A}, \{P_\theta : \theta \in \Theta\})$,
- (iii) the posterior model $(\Theta, \mathcal{B}, \{\xi_x : x \in \mathcal{X}\})$,

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

¹Research partially supported by NSF Grant No. 79-04693.

²Research supported by CNPq, CAPES and USP—Brazil.

Key Words : Conditional independence, ancillarity, sufficiency, Markov property, (strong) identification, splitting sets, measurable separability, specific sufficiency, variation independence.

AMS classification : Primary 62B05, Secondary 62B20.

In statistics the phenomenon of conditional independence manifests itself in a natural fashion. The statistical model that is most commonly in use is that of a sequence $\mathbf{X} = (X_1, X_2, \dots)$ of observables that are independently and identically distributed (i.i.d) for each given value of θ . It was DeFinetti (1937) who emphasized that, in view of the fact that θ 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 in the sense that the covariance (when exists) between any pair is non-negative. More specifically, $\text{cov}(X_i, X_j) = \text{var}(E(X_1|\theta))$.

Consider for example the particular case where X_1, X_2, \dots, X_n are i.i.d with common distribution $N(\mu, \sigma^2)$ with $\theta = (\mu, \sigma^2)$ not fully known. In almost every text book of statistics it is proved that $\bar{\mathbf{X}} = \frac{1}{n} \sum X_i$ is stochastically independent of $s^2 = \frac{1}{n} \sum (X_i - \bar{\mathbf{X}})^2$. Does it mean that $\bar{\mathbf{X}}$ 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 $\Theta = \{(0, 1), (0, 100), (1, 1), (1, 100)\}$). Suppose now that $\bar{\mathbf{X}}$ 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 $\bar{\mathbf{X}} = 2.1$ almost categorically rules out the points $(0, 1)$ and $(1, 1)$. Then $\bar{\mathbf{X}}$ and s^2 , even though they are conditionally independent given θ , are in effect highly dependent.

The three entities $\theta = (\mu, \sigma^2)$, $T = (\bar{\mathbf{X}}, s^2)$ and $\mathbf{X} = (X_1, \dots, X_n)$ in this order, have the Markov property in the sense that, given θ and T , the conditional distribution of \mathbf{X} depends on (θ, T) only through T . This is the sufficiency property of T as recognized by Fisher (1920, 1922). Kolmogorov (1942) gave a Bayesian characterization of the notion of sufficiency by noting that irrespective of the choice of the prior distribution ξ for the parameter θ , the posterior distribution $\xi_{\mathbf{X}}$ of θ depends on \mathbf{X} only through T . Note that the Fisher characterization of sufficiency is made only in terms of the statistical model for \mathbf{X} whereas the Kolmogorov characterization is made in terms of a large family of Bayesian models $(\Omega, \mathcal{F}, \Pi)$ 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 θ . He called a statistic $Y = Y(X)$ ancillary if the conditional distribution of Y given θ does not involve θ (is the same for all values of θ). For example, the statistic $\Sigma \frac{(X_i - \bar{X})^4}{s^4}$ is ancillary. In a series of articles 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 ξ for θ . Mouchart and Rolin (1978) studied in depth the familiar Basu theorems on sufficiency, ancillarity and conditional independence from the view point of the Bayesian model.

In this paper we too review Basu's results and also the two-parameter problem from the Bayesian perspective. Many results are stated without proof since the proofs involve standard measure theoretic arguments and can be found for instance in Loeve (1977).

2. NOTATION AND PRELIMINARIES

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

For any two subfields \mathcal{F}' and \mathcal{F}'' of \mathcal{F} , $\mathcal{F}' \vee \mathcal{F}''$ denotes the smallest subfield of \mathcal{F} that contains both \mathcal{F}' and \mathcal{F}'' . The smallest subfield of \mathcal{F} that contains all null sets of \mathcal{F} (a set N is null if $\pi(N) = 0$) is denoted by $\bar{\mathcal{F}}_0$, and write $\mathcal{F}_0 = \{\emptyset, \Omega\}$, the trivial subfield.

A subfield of \mathcal{F} is said to be completed if it contains $\bar{\mathcal{F}}_0$. For any subfield \mathcal{F}' of \mathcal{F} its completion $\bar{\mathcal{F}}'$ is defined by

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

For a random object X , the notation $X \in \mathcal{F}'$ indicates that $\mathcal{F}_X \subset \mathcal{F}'$ and X is said to be essentially \mathcal{F}' measurable. A random variable is a random object with range (R_1, \mathcal{B}_1) where R_1 is the real line and \mathcal{B}_1 is the Borel

σ -algebra. A random variable f is said to be bounded if there exists $a \in R_1$ such that $\pi\{\omega : |f(\omega)| \leq a\} = 1$. In the sequel, all random variables will be regarded as bounded unless stated otherwise and the use of small letters shall be restricted to their representation. The notation $f \subseteq X$ indicates that the random variable f is $\text{ess-}\mathcal{F}_X$ measurable. In the same spirit, for two random objects X and Y , we write $X \subseteq Y$ to indicate that $\bar{\mathcal{F}}_X \subseteq \bar{\mathcal{F}}_Y$. The class of all bounded random variables on $(\Omega, \mathcal{F}, \Pi)$ is denoted by L_∞ and $L_\infty(X)$ denotes the class of all $\text{ess-}\mathcal{F}_X$ measurable random variables. Here and for the rest of this article, equality of two random variable means essential equality; that is $f = g$ means $\{\omega : f(\omega) \neq g(\omega)\}$ is a null set.

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

$$\int fgd\pi = \int f^{*X}gd\pi \quad \forall g \in L_\infty(X)$$

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

3. CONDITIONAL INDEPENDENCE : DEFINITION, PROPERTIES AND THE DROP/ADD PRINCIPLES

In this section the definition and properties of conditional independence are briefly discussed.

Three random objects X, Y and Z are being considered and in this section $*$ stands for $*Z$ operator.

Definition 1. (Intuition) : The random objects X and Y are conditionally independent given Z (in symbols $X \amalg Y|Z$) if for any $f \in L_\infty(X)$

$$E(f|Y, Z) = f^{*YZ} = f^*$$

Note that to say $X \amalg Y|Z$ is equivalent to say that $X|(Y, Z)$ has the same conditional distribution as $X|Z$. This is the intuition behind the definition. In the case where Z is essentially a generator of \mathcal{F}_0 , we obtain the independence of X and Y in the usual sense. In this case the notation is $X \amalg Y$.

Definition 1a. (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 well known theorem gives the equivalence of the two definitions showing that $X \amalg Y|Z$ implies $Y \amalg X|Z$ which is not clear by looking at definition 1.

Theorem 1 : *Definitions 1 and 1a are equivalent.*

The concept of conditional independence (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, Z_1 , are random objects such that $X \amalg Y|Z$, $X_1 \subset X$, $Z_1 \subset Z$. What can be said about the relationship \amalg if X_1 is substituted for X , Z_1 for Z , (Y, W) for Y or (Z, W) for Z ? In other words, can \mathcal{F}_X , \mathcal{F}_Y or \mathcal{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 relationship \amalg does not hold we write $\text{not } \amalg$.

It is not difficult to find examples showing that arbitrary reductions or enlargements of \mathcal{F}_Z = the conditioning subfield, may destroy the c.i. property. With the example of the normal distribution presented in the introduction, we have $\bar{X} \amalg s^2|\theta$ but not $\bar{X} \amalg s^2$. In yet another statistical context, suppose that θ is the unknown (real) parameter of interest and let X and Y be two i.i.d. random variables with common uniform distribution on the interval $(\theta - \frac{1}{2}, \theta + \frac{1}{2})$. Since θ is unknown, we can only say that $X \in \mathcal{R}$, $Y \in \mathcal{R}$. However, after X has been observed equal to x , we would for sure know that $x-1 \leq Y \leq x+1$. This shows clearly that $X \amalg Y|\theta$ but not $X \amalg Y$. To show that we can have $X \amalg Y$ and $X \text{ not } \amalg Y|Z$, let W and Z be two i.i.d. $N(0, 1)$ variables and take $X = Z - W$ and $Y = Z + W$.

The above examples may be viewed as cases of Simpson's paradox (see Dawid, 1979a). The paradox, however, is much stronger. For instance, let W and Z be two independent normal variables with zero means. As before 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 is clearly equal to -1 . On the other hand, δ may be taken very small to make $\rho(X, Y)$ close to 1. This shows that we can have cases 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 problem of dependence inversion is discussed in depth in Lindley and Novick (1981). The following example may be of relevance for applied statisticians.

Example : Suppose that an urn contains θ (unknown) white balls in a total of N (known) balls. A sample, without replacement, of n balls is selected

from this urn. Let X be the number of white balls in the sample which implies that $Y = \theta - X$ is the number of white balls that remain in the urn. Since $\rho(X, Y | \theta) = -1$, we have X not $\Pi Y | \theta$. It can be proved (see Whitt, 1979) that $X \Pi Y$ iff θ has binomial (prior) distributions with fixed parameters N and $p \in (0, 1)$. On the other hand, if a priori, $\Pr\{\theta = 0\} = \Pr\{\theta = N\} = \frac{1}{2}$, then $\rho(X, Y) = 1$ showing an extreme inverted dependence.

The essence of Drop/Add principles for c.i. is contained in the following propositions.

Proposition 1: If $X \Pi Y | Z$, then for every $X' \subset X$, we have :

- (i) $X' \Pi Y | Z$
- (ii) $X \Pi Y | (Z, X')$
- (iii) $(X, Z) \Pi (Y, Z) | Z$.

By way of explanation, if $X \Pi Y | Z$, then the relation Π is preserved when (i) X and Y are increased (Add) by any essential part of Z , (ii) Z is increased (Add) by any essential part of X or Y , and (iii) X and Y are arbitrarily reduced (Drop).

To end this section we present an extreme case of Drop/Add principles for the conditioning random object. It appears in Dawid (1980) and it was originally introduced by G. Udny Yule in terms of collapsibility of contingency tables. It must clarify the problems with Simpson's paradox for contingency tables.

Proposition 2: Let X, Y and Z be three random objects such that $\mathcal{F}_Z = \{\phi, \Omega, A, A^c\}$ with $0 < \Pi(A) < 1$. If $X \Pi Y$ and $X \Pi Y | Z$, then either $X \Pi Z$ or $Y \Pi Z$.

The proof becomes simple when we recognize the following result.

Lemma: If $X \Pi Y$ and $X \Pi Y | Z$, then for every atom A of Z with $\Pi(A) > 0$, we have

$$E(I_A | (X, Y)) = \frac{1}{\Pi(A)} E(I_A | X)E(I_A | Y).$$

Proof of lemma: Let B, C , be two sets such that $I_B \subset X, I_C \subset Y$

$$\begin{aligned} \int_{BC} E(I_A | X)E(I_A | Y)d\Pi &= \int E(I_{AB} | X)E(I_{AC} | Y)d\Pi \\ &= \int E(I_{AB} | X)d\Pi \int E(I_{AC} | Y)d\Pi = \Pi(AB)\Pi(AC) \\ &= [\Pi(A)]^2\Pi(B | A)\Pi(C | A) = [\Pi(A)]^2\Pi(BC | A) \\ &= \Pi(A)\Pi(ABC) = \Pi(A) \int_{BC} E(I_A | (X, Y))d\Pi. \end{aligned}$$

Since sets of the form BC generate $\mathcal{F}_{X,Y}$ a standard argument completes the proof.

Proof of proposition : Let $p = \Pi(A)$. From lemma we have

$$E(I_A | (X, Y)) = \frac{E(I_A | X)E(I_A | Y)}{p}$$

and
$$E(I_{A^c} | (X, Y)) = \frac{[1 - E(I_A | X)][1 - E(I_A | Y)]}{1 - p}$$

and consequently

$$\left(1 - \frac{E(I_A | X)}{p}\right) \left(1 - \frac{E(I_A | Y)}{p}\right) = 0$$

Since $X \perp Y$, this equation holds only if $\frac{E(I_A | X)}{p} = 1$ or $\frac{E(I_A | Y)}{p} = 1$.

4. BAYESIAN INFERENCE : SUFFICIENCY, ANCILLARITY AND INDEPENDENCE

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 conditional independence to describe the Bayesian version of those statistical concepts and their properties. First we review some of the structures involved.

Let $(\mathcal{X}, \mathcal{A})$ be the usual Sample Space and $\{P_\theta : \theta \in \Theta\}$ be a family of probability measures on $(\mathcal{X}, \mathcal{A})$ where Θ is the usual parameter "Space". In addition the Bayesian considers a (prior) probability space $(\Theta, \mathcal{B}, \xi)$ where \mathcal{B} is a σ -algebra of subsets of Θ such that $P_\theta(A)$ is a \mathcal{B} -measurable function for every fixed $A \in \mathcal{A}$. 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 $P_\theta(A)$.

We then consider the probability space $(\Omega, \mathcal{F}, \Pi)$, where now $\Omega = \Theta \times \mathcal{X}$, $\mathcal{F} = \mathcal{B} \times \mathcal{A}$ and Π is defined by

$$\Pi(F) = \int_{\Theta} P_\theta(F^\theta) d\xi(d\theta)$$

where $F^\theta = \{x : (\theta, x) \in F\}$. The marginal on \mathcal{X} is defined by

$$P(A) = \Pi(\Theta \times A) \text{ for every } A \in \mathcal{A}.$$

Let X and Y be two random objects on (Ω, \mathcal{F}) . We say that X represents the sample and Y represents the parameter if

$$\mathcal{F}_X = \{\Theta \times A, A \in \mathcal{A}\}$$

and

$$\mathcal{F}_Y = \{B \times \mathcal{X}, B \in \mathcal{B}\}.$$

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

Definition 2 :

(a) If $X \amalg Y | X_1$ we say X_1 is sufficient for X with respect to Y .

(b) If $X_2 \amalg Y$ 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) $X_1 \amalg X_2 | Y$.

Basu (1955, 1958) speculates under what conditions two of the three relation (a), (b) and (c) imply the third. In this section we 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 3 : If in addition to $X_2 \amalg Y$ and $X_1 \amalg X_2 | Y$ we have $X \amalg Y | (X_1, X_2)$ then $X \amalg Y | X_1$.

Proof : Note that $X_2 \amalg Y$ and $X_1 \amalg X_2 | Y$ implies $X_2 \amalg Y | X_1$, also $X_2 \amalg Y | X_1$ and $X \amalg Y | (X_1, X_2)$ implies $X \amalg Y | X_1$.

Arguing similarly it is easy to see that if $X_1 \amalg X_2$, then (a) implies (b) and (c). The meaning of $X_1 \amalg X_2$ in classical statistics however is void.

Note that Proposition 3 gives conditions for reducing (Drop) the conditioning random object. Actually all of Basu's theorems are cases of Drop/Add principles.

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.

The following theorem is a Bayesian version of the result of Koehn and Thomas (1975).

Theorem 2 : *Let $X_1 \subset X$ be a sufficient random object (i.e. $X \amalg Y | X_1$). The random object $Y \wedge X_1$ is essentially a constant (i.e. $F_{Y \wedge X_1} = F_0$) iff $X_2 \amalg Y$ whenever $X_2 \subset X$ and $X_1 \amalg X_2 | Y$ (i.e. X_2 is ancillary if X_1 and X_2 are statistically independent).*

Proof : $E(I_A | Y) = E(I_A | X_1, Y) = E(I_A | X_1)$ by $X \amalg Y | X_1$. Now since, $X_1 \wedge Y = F_0$ it follows that $E(I_A | Y)$ is a constant. Take X_2 such that $X_2 \equiv Y \wedge X_1$. Since $X_2 \subset Y$, $X_1 \amalg X_2 | Y$. Then by hypotheses $X_2 \amalg Y$, which implies that $X_2 \amalg X_2$ since $X_2 \subset Y$; that is $X_2 \equiv Y \wedge X_1$ is essentially a constant.

Remarks : The condition introduced by Koehn and Thomas (1975) is the non existence of a splitting set. A set A in the sample space is a 'splitting' set if $P_\theta(A) = 0$ or 1 for all θ , and at least for a pair $\{\theta_1, \theta_2\} \in \Theta$, $P_{\theta_1}(A) = P_{\theta_2}(A^c) = 1$. In the Bayesian framework, 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 $E(I_A | Y) = E^2(I_A | Y)$. It is easy to see that if A is a splitting set then $I_A \subset Y \wedge X$. We conclude that the non existence of a splitting set is equivalent to $Y \wedge X$ being essentially a constant.

Basu (1955) proved that any ancillary statistic is statistically independent of any bounded complete sufficient statistic. The Bayesian analogue of the 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.

Definition 3 : The random objects X and Y are said to be measurably separated conditionally on Z if $(X, Z) \wedge (Y, Z) \equiv Z$. 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).

Let X and Y be two random objects. We shall study some aspects of the linear maps $L_x(Y) \xrightarrow{*} L_\infty(X)$ and $L_\infty(X) \xrightarrow{+} L_\infty(Y)$ where $*$ is for $E(\cdot | X)$ and $+$ for $E(\cdot | Y)$.

Definition 4 : We say that X is strongly identified by Y and write $X \ll Y$ if the map $L_\infty(X) \xrightarrow{+} L_\infty(Y)$ is essentially one-one.

Proposition 4: If the map $L_\infty(Y) \xrightarrow{*} L(X)$ is essentially onto then $X \ll Y$.

Proof: Let $(f, h) \subset X$ and $f^+ = 0$. Since $*$ is essentially onto $\exists g \subset Y$ such that $g^* = h$. Then

$$E(fh) = E(fg^*) = E(f^+g) = 0$$

since h is arbitrary $f = 0$.

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

- Proposition 5:* (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: The proof is easy and hence omitted.

Remark: From Proposition 5 it is easy to see that a Burkholder type theorem on intersection of sufficient subfields is true in the Bayesian framework. Precisely, if $X \perp\!\!\!\perp Y | X_1$ and $X \perp\!\!\!\perp Y | X_2$ then $X \perp\!\!\!\perp Y | X_1 \wedge X_2$.

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 6: If $X \ll Y$ then $X_{[Y]} \equiv X$.

Proof: Note that $X \perp\!\!\!\perp Y | X_{[Y]}$. Thus $\forall f \subset X$

$$E\{E(f | Y, X_{[Y]} | Y)\} = E\{E(f | X_{[Y]}) | Y\}.$$

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 $X = X_{[Y]}$.

The Bayesian version of Basu's theorem is contained in the result below.

Theorem 3: Let X, Y and Z be three random objects. If $X \perp\!\!\!\perp Y, X \perp\!\!\!\perp Y | Z$ and $Z \ll Y$ then $X \perp\!\!\!\perp Z | Y$.

Proof: Since $X \perp\!\!\!\perp Y | Z$ we have, for any $f \subset X$

$$E(f | Y, Z) = E(f | Z)$$

and since

$$X \perp\!\!\!\perp Y, E(f | Y) = E(f)$$

Therefore

$$E[\{E(f | Z) - E(f)\} | Y] = 0.$$

Now since $Z \ll Y$ $E(f|Z) = E(f)$ we thus have $E(f|Y, Z) = E(f)$.

Note that to obtain Basu's theorem, we consider X as the sample, Y as the parameter, and X_0 and X_1 two random objects such that $(X_0, X_1) \subset X$, $X_0 \perp\!\!\!\perp Y$, $X \perp\!\!\!\perp Y|X_1$ and $X_1 \ll Y$. Clearly $X_0 \perp\!\!\!\perp Y|X_1$ and the result $X_0 \perp\!\!\!\perp X_1|Y$ follows.

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

Proposition 7: Suppose $X_1 \subset X$ and $X \perp\!\!\!\perp Y|X_1$. If $X_1 \ll Y$ then $X_1 = X_{\{Y\}}$.

Proof: From Proposition 6 $X_{\{Y\}} \subset X_1$ and $X \perp\!\!\!\perp Y|X_{\{Y\}}$.

Let $f \subset X_1$ then $E(f|Y) = E[E(f|X_{\{Y\}})|Y]$ or

$$E(f|Y) = E[E(f|X_{\{Y\}}, Y)|Y]$$

$$= E(E(f|X_{\{Y\}})|Y)$$

since $X_1 \ll Y$ we conclude that $f = E(f|X_{\{Y\}}) \subset X_{\{Y\}}$.

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

$$(X, Z)_{\{Y, z\}} = (X, Z).$$

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).

5. THE TWO PARAMETER PROBLEM

We now briefly discuss sufficiency in the presence of a nuisance parameter.

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 $X \perp\!\!\!\perp Y_1|(X_2, Y_2)$. (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, X_2) imply the sufficiency of (X_1, X_2) ?

Proposition 8: If $(X_1, Y_1) \wedge (X_2, Y_2) \subset (X_1, X_2)$ then $X \perp\!\!\!\perp Y_2|(X_1, Y_1)$ and $X \perp\!\!\!\perp Y_1|(X_2, Y_2)$ imply $X \perp\!\!\!\perp Y|(X_1, X_2)$.

Proof: We have

$$X \perp\!\!\!\perp Y | (X_1, Y_1)$$

and

$$X \perp\!\!\!\perp Y | (X_2, Y_2)$$

and a simple argument yields

$$X \perp\!\!\!\perp Y | (X_1, Y_1) \wedge (X_2, Y_2).$$

The following related result may also be of interest.

Proposition 9: If $X \perp\!\!\!\perp Y_2 | (X_1, Y_1)$ and $X \perp\!\!\!\perp Y_1 | (X_2, Y_2)$ then

$$X \perp\!\!\!\perp Y | (X_1, X_2) \text{ if and only if } X \perp\!\!\!\perp Y_1 | (X_1, X_2).$$

Note the condition $X \perp\!\!\!\perp Y_1 | (X_1, X_2)$ does not have an interpretation in classical statistics since distributions depend on both parameters Y_1 and Y_2 .

The following example is again relevant. Note that the parameter space θ is variation independent (if the parameter space is the cartesian product of the domain Y_1 by the domain of Y_2 ; see Basu, 1977 and Barndorff-Nielsen, 1978).

Take

$$\Theta = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$$

Then $\Theta = \Theta_1 \times \Theta_2$ where $\Theta_1 = \Theta_2 = \{0, 1\}$

$$X = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$$

$$P_\theta = \delta_\theta \text{ the point mass at } \theta.$$

Then $T = I_{\{(0, 1), (1, 0)\}}(x)$ is specific sufficient for θ_1 and for θ_2 but is not sufficient. This example shows that variation independence on (θ_1, θ_2) and specific sufficiency of X_1 , and X_2 does not imply that (X_1, X_2) being sufficient.

Acknowledgement. Thanks are due to the referee for helping in improving presentation of the paper.

REFERENCES

ASH, R. B. (1972): *Real Analysis and Probability*, Academic Press, New York.
 BHADUR, R. R. (1955): Measurable subspaces and subalgebras. *Proc. Amer. Math. Soc.*, 6, 565-70.
 BARNDORFF-NIELSEN, O. (1978): *Information and Exponential Families in Statistical Theory*, John Wiley, New York.
 BASU, D. (1955): On statistics independent of a complete sufficient statistics. *Sankhyā*, A, 15, 377-80
 ——— (1958): On statistics independent of a sufficient statistics. *Sankhyā*, A, 20, 223-26.
 ——— (1959): The family of ancillary statistics. *Sankhyā*, A, 21, 247-56
 ——— (1964): Recovery of ancillary information. *Sankhyā*, A, 26, 3-16.

- (1967): Problems relating to the existence of maximal and minimal elements in some families of statistics (Subfields). *Proc. Fifth Berkeley Sym. Math. Statist. Prob.*, **1**, 41–50.
- (1977): On the elimination of nuisance parameters. *Jour. Amer. Statist. Assoc.*, **72**, 355–66.
- (1978): On partial sufficiency: A review. *J. Statist. Plan. Inf.*, **2**, 1–13.
- BASU, D. and CHENG, S. C. (1979): A note on sufficiency in coherent models. *Int. J. Math. Math. Sci.* To appear.
- BURKHOLDER, D. L. (1961): Sufficiency in the undominated case. *Ann. Math. Statist.*, **32**, 1191–200.
- CHENG, S. C. (1978): A mathematical study of sufficiency and adequacy in statistical theory. Ph.D. Dissertation, FSU, Florida.
- DAWID, A. P. (1979a): Conditional independence in statistical theory. *J. Roy. Statist. Soc.*, **B**, **41**, 1–31.
- (1979b): Some misleading arguments involving conditional independence. *J. Roy. Statist. Soc.*, **B**, **41**, 249–52.
- (1979c): A Bayesian look at nuisance parameters. *Trabajos de Estadística*, To appear.
- (1980): Conditional independence for statistical operations. *Ann. Statist.*, **8**, 598–617.
- DE FINETTI, B. (1937): Foresight: Its logical laws, its subjective sources. Translated edition 1964 in *Studies in Subjective Probability* (H. E. Kyburg and H. E. Smokler, editors). John Wiley, New York.
- (1970): *Theory of Probability*, Vols. 1 and 2. Translated edition, 1974. John Wiley, London.
- DOOB, J. L. (1953): *Stochastic Processes*, John Wiley, New York.
- FERREIRA, P. E. (1980): Comments on Berkson's Paper "In dispraise of ...". Unpublished report.
- FISHER, R. A. (1920): A mathematical examination of the methods of determining the accuracy of an observation by the mean error, and by the mean square error. *Mon. Not. Roy. Ast. Soc.*, **80**, 758–70.
- (1922): On the mathematical foundations of theoretical statistics. *Phil. Trans.*, **A**, **222**, 309–68.
- FLORENS, J. P. and MOUCHAR, M. (1977): Reduction of Bayesian experiments. *CORE*, Discussion Paper 7737.
- GODAMBE, V. P. (1979): On sufficiency and ancillarity in presence of nuisance parameter. Unpublished report.
- HALL, W. J., WIJSMAN, R. A. and GHOSH, J. K. (1965): The relationship between sufficiency and invariance. *Ann. Math. Statist.*, **36**, 375–614.
- KOEHN, U. and THOMAS, D. L. (1975): On statistics independent of a sufficient statistics: Basu's lemma. *Amer. Statist.*, **29**, 40–2.
- KOLMOGOROV, A. N. (1942): Determination of the center of dispersion and degree of accuracy for a limited number of observations (in Russian). *Izvestija Akademii Nauk, Ser. Mat.*, **6**, 3–32.

- LEHMAN, E. L. and ACHEFFE, H. (1950) : Completeness, similar regions and unbiased estimation, Part I. *Sankhyā, A*, **10**, 305-40.
- LINDLEY, D. V. and NOVICK, M. R. (1981) : The role of exchangeability in inference. *Ann. Statist.*, **9**, 45-58.
- LOEVE, M. (1977) : *Probability Theory*, 4th ed. Springer-Verlag, NY.
- MOUCHART, M. and ROLIN, J. M. (1978) : A note on conditional independence. Unpublished report.
- MOY, S. T. C. (1954) : Characterization of conditional expectation as a transformation on function spaces. *Pacific J. Math.*, **4**, 47-63.
- PATHAK, P. K. (1975) : Note on Basu's lemma. Unpublished report.
- PICCI, G. (1977) : Some connections between the theory of sufficient statistics and the identifiability problem. *SIAM J. Appl. Math.*, **33**, 383-98.
- WHITT, W. (1979) A note on the influence of the sample on the posterior distribution. *Jour. Amer. Statist. Assoc.* **74**, 424-426.

Paper received : March, 1981.

Revised : January, 1983.