

THE FLORIDA STATE UNIVERSITY
COLLEGE OF ARTS AND SCIENCES

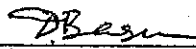
BAYESIAN SOLUTIONS TO SOME CLASSICAL
PROBLEMS OF STATISTICS

by


CARLOS ALBERTO DE BRAGANCA PEREIRA

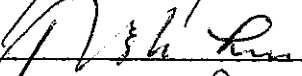
A Dissertation submitted to the
Department of Statistics
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

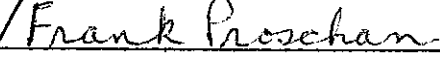
Approved:



Professor Directing Dissertation







December, 1980

BAYESIAN SOLUTIONS TO SOME CLASSICAL
PROBLEMS OF STATISTICS

(Publication No. _____)

Carlos Alberto de Bragança Pereira, Ph.D.

The Florida State University, 1980

Major Professor: Debrabrata Basu, Ph.D.

Three of the basic questions of Statistics may be stated as follows:

A. Which portion of the data X is actually informative about the parameter of interest θ ?

B. How can all the relevant information about θ provided by the data X be extracted?

C. What kind of information about θ do the data X possess?

The perspective of this dissertation is that of a Bayesian.

Chapter I is essentially concerned with question A. 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. The language of conditional independence is used in the remaining Chapters.

Chapter II deals with question B for the particular problem of analysing categorical data with missing entries. It is demonstrated how a suitably chosen prior for the frequency parameters can streamline the analysis in the presence of missing entries due to non-response or other causes. The two cases where the data follow the Multinomial or the Multivariate Hypergeometric model are treated separately. In the first case it is adequate to restrict the prior (for the cell probabilities) to the class of Dirichlet distributions. In the Hypergeometric case it is convenient to select a prior (for the cell population frequencies) from the class of Dirichlet-Multinomial (DM) distributions. The DM distributions are studied in detail.

Chapter III is directly related to question C. Conditions on the likelihood function and on the prior distribution are presented in order to assess the effect of the sample on the posterior distribution. More specifically, it is shown that under certain conditions, the larger the observations obtained, the larger (stochastically in terms of the posterior distribution) is the appropriate parameter.

Finally, Chapter IV deals with the characterization of distributions in terms of Blackwell comparison of experiments. It is shown that a result (for the Hypergeometric model) obtained in Chapter II is actually a consequence of a property of complete families of distributions.

REFERENCES

	Page
ABSTRACT	ii
ACKNOWLEDGMENTS	vi
CHAPTER I - CONDITIONAL INDEPENDENCE IN STATISTICS	1
1 - INTRODUCTION	1
2 - NOTATION AND PRELIMINARIES	5
3 - DEFINITION OF CONDITIONAL INDEPENDENCE	9
4 - THE DROP/ADD PRINCIPLES AND OTHER PROPERTIES OF CONDITIONAL INDEPENDENCE	12
5 - MARKOV CHAINS AND BAYESIAN INFERENCE	23
6 - ON MEASURABLE SEPARABILITY OF RANDOM OBJECTS	32
7 - BASU THEOREM	36
REFERENCES	41
CHAPTER II - ON THE BAYESIAN ANALYSIS OF CATEGORICAL DATA: THE PROBLEM OF NONRESPONSE	44
1 - INTRODUCTION	44
2 - NONRESPONSE: THE MULTINOMIAL MODEL	49
3 - THE DIRICHLET-MULTINOMIAL DISTRIBUTION: PROPERTIES	55
4 - THE DM DISTRIBUTION: A NATURAL FAMILY OF PRIORS FOR FINITE POPULATION STUDIES	61
5 - NONRESPONSE: THE MULTIVARIATE HYPERGEOMETRIC MODEL	64
6 - FINAL REMARKS	69

	Page
REFERENCES	71
APPENDIX	72
CHAPTER III - THE INFLUENCE OF THE SAMPLE ON THE POSTERIOR DISTRIBUTION	74
1 - INTRODUCTION	74
2 - PRELIMINARIES	75
3 - THEORETICAL RESULTS	78
4 - APPLICATIONS	85
5 - ACKNOWLEDGMENTS	90
REFERENCES	91
CHAPTER IV - ON THE CHARACTERIZATION OF DISTRIBUTIONS IN TERMS OF SUFFICIENCY AND COMPLETENESS	92
1 - INTRODUCTION	92
2 - CHARACTERIZATION OF THE HYPERGEOMETRIC MODELS	94
3 - CHARACTERIZATION OF OTHER DISTRIBUTIONS	98
REFERENCES	106
VITA	107

ACKNOWLEDGMENTS

I am very grateful to my major professor, Dr. D. Basu, for his guidance and encouragement prior to and throughout the preparation of this dissertation. His investments of time and energy made this research possible.

Special appreciation goes to Dr. F. Proschan who painstakingly read this work and offered his corrections. His expert assistance was vital in Chapter III. The other members of my committee, Drs. P. E. Lin and J. J. Andrews, receive thanks for their willing service.

I also wish to acknowledge Universidade de Sao Paulo and CAPES for providing the financial support. Mrs. Kathy Strickland deserves my thanks for the precious typing job.

I am thankful to my Florida State colleagues, especially Wai and Tiwari, for the stimulating discussions with me. The moral support of my Brazilian friends, Julio, Severo, and Wagner, will never be forgotten.

Finally, my deepest gratitude goes to my parents, my wife, and my son for their love and encouragement throughout these years in graduate school.

CHAPTER I. 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 chapter 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 language of conditional independence developed in this chapter will be fully utilized in Chapter 2.

The statistical perspective of this dissertation 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 statistical model $(X, \mathcal{A}, \{P_\theta : \theta \in \Theta\})$. Writing $\omega = (\theta, X)$, $(\Omega, \mathcal{F}) = (\Theta \times X, \mathcal{B} \times \mathcal{A})$, and Π for the joint distribution of (θ, X) , there exists then a subjective probability model $(\Omega, \mathcal{F}, \Pi)$ for $\tilde{\omega}$. 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 $(X, \mathcal{A}, \{P_\theta: \theta \in \Theta\})$
 - (iii) the posterior model $(\Theta, \mathcal{B}, \{\xi_x: x \in X\})$,
- and (iv) the predictive model (X, \mathcal{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 $\underline{X} = (X_1, X_2, \dots)$ 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 - if we consider the (predictive) conditional distributions,

X_2 is stochastically increasing with X_1 , X_3 is stochastically increasing with (X_1, X_2) , and so on. (See Chapter 3 for details in some concepts of dependence.)

Consider for example the particular case where X_1, \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 textbook on Statistics it is proved that the statistic $\bar{X} = n^{-1} \sum X_i$ is stochastically independent of $S^2 = n^{-1} \sum (X_i - \bar{X})^2$. Does it mean that \bar{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 \bar{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 $L(0, 1) = \frac{5}{\sqrt{2\pi}} \exp\{-\frac{25}{2} (2.1)^2\}$ 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 $\bar{X} = 2.1$ almost categorically rules out the points $(0, 1)$ and $(1, 1)$. Therefore, the observation of $\bar{X} = 2.1$ 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 \bar{X} and S^2 , even though they are conditionally independent given θ , are in effect highly dependent.

The three entities $\theta = (\mu, \sigma^2)$, $T = (\bar{X}, S^2)$, and $\underline{X} = (X_1, \dots, X_n)$, in this order, have the Markov property in the sense that, given θ

and T , the conditional distribution of \underline{X} depends on (θ, 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 ξ for the parameter θ , the posterior distribution $\xi_{\underline{X}}$ of θ depends on \underline{X} only through T . In other words, the sequence $(\underline{X}, T, \theta)$ have the Markov property; that is \underline{X} and θ are conditionally independent given T . Note that the Fisher characterization of sufficiency is made only in terms of the statistical model for \underline{X} whereas the Kolmogorov characterization is made in terms of a large family of Bayesian models (Ω, F, Π) for $\omega = (\theta, \underline{X})$. (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 \underline{X} about the parameter θ . He called a statistic $Y = Y(\underline{X})$ 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 $\sum (x_i - \bar{X})^4 / S^4$ 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 ξ for the parameter θ .

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 $(\Omega, \mathcal{F}, \Pi)$. 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 $(\Omega, \mathcal{F}, \Pi)$ 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 (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 $(X, \mathcal{A}, \Pi^{-1})$ are undistinguishable in a 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, and any collection of random variables (stochastic processes).

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 that contains all null sets of \mathcal{F} (a set N is null if $\Pi(N) = 0$) is denoted by $\overline{\mathcal{F}}_0$; that is,

$$\overline{\mathcal{F}}_0 = \{F; F \in \mathcal{F} \text{ and } \Pi(F) = 0 \text{ or } \Pi(F) = 1\}$$

and write $F_0 = \{\phi, \Omega\}$, the trivial subfield.

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

For any subfield F' of F its completion is defined by:

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

For a random object X , the notation $X \in F'$ indicates that $F_X \subset \bar{F}'$ and X is said to be essentially F' -measurable or X is $\text{ess-}F'$ -measurable. A random variable is a random object with range (R_1, B_1) where R_1 is the real line and B_1 is the Borel σ -algebra. A random variable f is said to be bounded if $\exists a \in R_1$ such that $\Pi\{\omega; |f(\omega)| \leq a\} = 1$. 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 $\text{ess-}F_X$ -measurable. In the same spirit, for two random objects X and Y , we write $X \subset Y$ to indicate that $\bar{F}_X \subset \bar{F}_Y$. If $\bar{F}_X = \bar{F}_Y$ we write $X \equiv Y$ to indicate the essential equivalence between X and Y . The class of all bounded random variables on (Ω, F, Π) is denoted by L_∞ and $L_\infty(X)$ denotes the class of all $\text{ess-}F_X$ -measurable random variables. Here and for the rest of this chapter, 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

$$\int fg d\Pi = \int f^{*X} g d\Pi \quad \forall g \in L_{\infty}(X).$$

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, 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 chapter. Here, $*$ is substituted for $*X$ and \dagger is substituted for $*Y$.

PROPOSITION 1

If two random objects X and Y are such that $X \subset Y$, then
 $\forall f \in L_{\infty}$

- (i) $E\{f^*|Y\} = (f^*)^{\dagger} = f^*$
- (ii) $E\{f^{\dagger}|X\} = (f^{\dagger})^* = f^*$
- (iii) $f^{\dagger} \subset X \rightarrow f^{\dagger} = f^*$

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

$$E\{f*gh\} = E\{fg*h\} = E\{f*g*h\}. \quad (* \text{ stands for } *X.)$$

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_∞ in $L_\infty(X)$ when the L_2 - norm is considered.

Given two random objects X and Y , the random object $(X, Y): \Omega \rightarrow X \times Y$ 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 $\overline{F}_X \cap \overline{F}_Y$ 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:

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

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

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 Ω . \mathcal{D} is said to be a D-system (D for Dynkin) if the following conditions hold:

- (i) $\Omega \in \mathcal{D}$.
- (ii) If $B, A \in \mathcal{D}$, $B \subset A$ then $A - B \in \mathcal{D}$.
- (iii) If $A_1, A_2, \dots \in \mathcal{D}$ and $A_n \uparrow A$ then $A \in \mathcal{D}$.

THEOREM 1

Let C be a class of subsets of Ω and assume that C is closed under finite intersections. If \mathcal{D} is a \mathcal{D} -system such that $C \subset \mathcal{D}$ then $\sigma(C) \subset \mathcal{D}$. ($\sigma(C)$ is the smallest σ -field that contains 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 $X \perp\!\!\!\perp Y|Z$) 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 $X \perp\!\!\!\perp Y|Z$ is equivalent to say that $X|(Y, Z)$ has the same conditional distribution as does $X|Z$. This is the intuition behind Definition 3. Frequently we will use the notation $X|(Y, Z) \sim X|Z$ for $X \perp\!\!\!\perp Y|Z$.

An equivalent way to define c.i. is to say that the map $f \rightarrow f^*(Y, Z)$ from L_∞ 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 $X \perp\!\!\!\perp Y|Z$ implies $Y \perp\!\!\!\perp X|Z$ which is not clear by looking only at Definition 3.

THEOREM 2

Definitions 3 and 3a are equivalent.

PROOF3 \rightarrow 3a

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

$$\begin{aligned} (fg)^* &= E\{E\{fg|(Y, Z)\}|Z\} = E\{gE\{f|(Y, Z)\}|Z\} \\ &= E\{gE\{f|Z\}|Z\} = (gf^*)^* = f^*g^* \end{aligned}$$

3a \rightarrow 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 $E = \{E; E \in \bar{F}_Y \vee \bar{F}_Z \text{ and } \int_E fd = \int_E f^*d\pi \vee f \in X\}$. 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 $A \subset B$ we have $B - A \in E$.

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

$$\int_{CD} fd\pi = E\{I_C I_D f\} = E\{(I_C I_D f)^*\} = E\{I_D (I_C f)^*\}.$$

But by Proposition 2 and by hypothesis, we have

$$E\{I_D (I_C f)^*\} = E\{I_D I_C^* f^*\} = E\{I_D I_C f^*\} = \int_{CD} f^*d\pi.$$

Thus, $E' \subset E$ where

$$E' = \{CD; C \in \bar{F}_Y \text{ and } D \in \bar{F}_Z\}.$$

Since E' is closed under finite intersections, and $\sigma(E') = \bar{F}_Y \vee \bar{F}_Z$ we conclude, by Theorem 1, that $\bar{F}_Y \vee \bar{F}_Z \subset E$; that is, $f^* = E\{f|(Y, Z)\} \vee f \subset X$. \square

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

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

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 , and Z_1 are random objects such that $X \perp\!\!\!\perp Y|Z, X_1 \subset X$, and $Z_1 \subset Z$. What can be said about the relation $\perp\!\!\!\perp$ 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 , 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 $\perp\!\!\!\perp$ does not hold we write $\not\perp\!\!\!\perp$.

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 $A \subset \Omega$, $I_A(\omega)$ 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, $X \perp\!\!\!\perp Y$ and $X \perp\!\!\!\perp W$ but $X \not\perp\!\!\!\perp (Y, W)$. \square

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 Ω , then $X \perp\!\!\!\perp Y$ and $X \perp\!\!\!\perp Y|W$. This is clear since we obtain the following probability functions (p.f.):

		Y		
		0	1	
X	0	.3	.5	.8
	1	0	.2	.2
		.3	.7	1

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

		Y		
		0	1	
X	0	.2	0	.2
	1	.5	.3	.8
		.7	.3	1

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

		Y		
		0	1	
X	0	.25	.25	.5
	1	.25	.25	.5
		.5	.5	1

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, $(X - Y) \perp\!\!\!\perp (X + Y) | Y$.

However, it is well known that $(X - Y) \not\perp\!\!\!\perp (X + Y)$. □

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 Π 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 $X \perp\!\!\!\perp Y | W$, but $X \not\perp\!\!\!\perp Y$. The probability functions in this case are:

		Y		
		0	1	
X	0	.2	.2	.4
	1	.3	.3	.6
		.5	.5	1

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

		Y		
		0	1	
X	0	.2	.3	.5
	1	.2	.3	.5
		.4	.6	1

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

		Y		
		0	1	
X	0	.20	.25	.45
	1	.25	.30	.55
		.45	.55	1

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, $(X - Y + Z) \perp\!\!\!\perp (X + Y + Z) | Z$ but $(X - Y + Z) \not\perp\!\!\!\perp (X + Y + Z)$. □

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, δ 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 $X' \subset X$ we have:

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

PROOF

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

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

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

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, $\forall g \in Y$ $E\{g|(Z, X')\} = E\{g|(Z, X', X)\}$. \square

COROLLARY 1

For any $Z' \subset Z$,

$$X \perp\!\!\!\perp Y|Z \text{ if and only if } X \perp\!\!\!\perp (Y, Z')|Z.$$

COROLLARY 2

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

$$(i) \quad W_1 \perp\!\!\!\perp W_2|Z$$

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

By way of explanation, if $X \perp\!\!\!\perp Y|Z$ 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 $X \perp\!\!\!\perp Y|X'$:

$$\forall f \subset X, E\{f^*|Y\} = E\{f|Y\}, \text{ where } f^* = E\{f|X'\}.$$

PROOF

Here, $*$ stands for $*X'$ and \dagger for $*Y$.

(i) Necessity.

Since $\forall f \subset X$, $f^* = E\{f|(Y, X')\}$, by Proposition 1 we conclude that $\forall f \subset X$, $(f^*)^\dagger = f^\dagger$.

(ii) Sufficiency.

Let $f \subset X$, $g \subset Y$, and $f' \subset X'$. Clearly $ff' \subset X$. Note that

$$(fgf')^\dagger = g(ff')^\dagger = g(ff')^{*\dagger} = g(f^*f')^\dagger = (f^*gf')^\dagger.$$

Since $E\{(fgf')^\dagger\} = E\{fgf'\}$, by Proposition 2 we can write

$$E\{fgf'\} = E\{f^*gf'\} = E\{f^*g^*f'\}$$

Then $(fg)^* = f^*g^*$. \square

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 $X \perp\!\!\!\perp Y|Z$:

$$\forall f \subset (X, Z), E\{f^*|Y\} = E\{f|Y\}, \text{ where } f^* = E\{f|Z\}.$$

The equivalence of this result with Proposition 4 follows directly from Corollary 1.

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, τ 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 (τ_1, τ_2) of τ , 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

$$(X_1, X_2) \perp\!\!\!\perp X_3|Z \rightarrow X_1 \perp\!\!\!\perp X_3|(Z, X_2).$$

A useful result in statistical applications by Dawid (1979a), is stated as follows:

PROPOSITION 5

The following properties are equivalent:

(i) $X \perp\!\!\!\perp Y|Z$ and $X \perp\!\!\!\perp W|(Y, Z)$.

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

PROOF

(i) \rightarrow (ii)

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

(ii) \rightarrow (i)

By Proposition 3, we conclude that $X \perp\!\!\!\perp Y|Z$ and $X \perp\!\!\!\perp (Y, W)|(Z, Y)$ which implies $X \perp\!\!\!\perp W|(Z, Y)$. \square

Note that since property (ii) is symmetric (Y and W may commute), the inclusion of the following property is implicit: (iii) $X \perp\!\!\!\perp W|Z$ and $X \perp\!\!\!\perp Y|(Z, W)$. 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 $X \perp\!\!\!\perp Y|(Z, W)$ and $T \perp\!\!\!\perp W|Z$, then $T \perp\!\!\!\perp Y|Z$ and $T \perp\!\!\!\perp W|(Y, Z)$.

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

a) $\Omega \in E$

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

$$\begin{aligned} (I_{(E_2-E_1)} f)^* &= (I_{E_2} f - I_{E_1} f)^* = I_{E_2}^* f^* - I_{E_1}^* f^* = (I_{E_2} - I_{E_1})^* f^* \\ &= I_{(E_2-E_1)}^* f^*. \end{aligned}$$

That is, $E_2 - E_1 \in E$.

c) For any monotone sequence E_1, E_2, \dots , of E , we have that

$I_{\lim E_n} = \lim I_{E_n}$ and by the dominated convergence theorem for conditional expectation, $(\lim I_{E_n} f)^* = \lim (I_{E_n} f)^*$. Since $E_n \in E$, $(\lim I_{E_n} f)^* = \lim I_{E_n}^* f^* = I_{\lim E_n}^* f^*$. That is, $\lim E_n \in E$.

To conclude the proof recall that, by hypothesis,

$\bigcup_{n=1}^{\infty} F_n \subset E$ and then by Theorem 1,

$$\bigcap_{n=1}^{\infty} F_n \subset E. \quad \square$$

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

PROPOSITION 7

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

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 $X_1 \perp\!\!\!\perp (X_2, X_3) | Z$. \square

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

$$(X_1, \dots, X_k) \perp\!\!\!\perp X_{k+1} | Z$$

for every $k = 1, 2, \dots, n - 1$.

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, \dots be random objects such that $X \perp\!\!\!\perp (Y_1, Y_2, \dots, Y_n) | Z$ for every $n = 1, 2, \dots$. Then, $X \perp\!\!\!\perp (Y_1, Y_2, \dots) | Z$ where (Y_1, Y_2, \dots) essentially is the generator of $\bigvee_{n=1}^{\infty} F_n = \sigma(\bigcup_{n=1}^{\infty} F_n)$. (Here, $F_n \equiv F_{Y_n}$.)

PROOF

Since $\bigcup F_n$ is a field, it is closed under finite intersections.

Let $*$ stand for $*Z$ and consider the set

$$E = \{E; E \in \bigvee_{n=1}^{\infty} F_n \text{ and } (I_E f)^* = I_E^* f^* \forall f \in X\}.$$

The following conditions show that E is a D-system:

PROOF

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

$(X_{i_1}, X_{i_2}, \dots) \perp\!\!\!\perp (X_{j_1}, X_{j_2}, \dots) | Z$ holds. Note that, for any

$k, \ell \in \{1, 2, \dots\}$ the finite relation

$(X_{i_1}, X_{i_2}, \dots, X_{i_k}) \perp\!\!\!\perp (X_{j_1}, X_{j_2}, \dots, X_{j_\ell}) | Z$ holds. This

follows from the discussion after Proposition 6 and from the fact

that $(X_1, \dots, X_m) \perp\!\!\!\perp X_{m+1} | Z \forall m = 1, 2, \dots, v$, where

$v = \max(i_1, \dots, i_k, j_1, \dots, j_\ell)$. By Theorem 3 it follows that

$(X_{i_1}, \dots, X_{i_k}) \perp\!\!\!\perp (X_{j_1}, X_{j_2}, \dots) | Z$. Finally, applying again

Theorem 3 we prove our claim. \square

We write $X_1 \perp\!\!\!\perp X_2 \perp\!\!\!\perp \dots | Z$ or $\prod_{n=1}^{\infty} X_n | Z$ to indicate that the sequence (X_1, X_2, \dots) 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, \dots is said to form a Markov Chain if,

$$(5.1) \quad \forall n \geq 1, (X_1, \dots, X_n) \perp\!\!\!\perp X_{n+2} | X_{n+1}.$$

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

$$(5.2) \quad \forall n \geq 1, (X_1, \dots, X_n) \perp\!\!\!\perp (X_{n+2}, X_{n+3}, \dots) | X_{n+1}.$$

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, \dots 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

$$\forall n \geq 1, (X_1, \dots, X_n) \perp\!\!\!\perp (X_{n+2}, X_{n+3}) | X_{n+1}.$$

But, using DROP/ADD principles, (5.1) implies that

$\forall n \geq 1, (X_1, \dots, X_n) \perp\!\!\!\perp X_{n+2} | X_{n+1}$ and

$(X_1, \dots, X_n) \perp\!\!\!\perp X_{n+3} | (X_{n+1}, X_{n+2})$.

The conclusion of step 1 follows now directly from Proposition 5.

Step 2 - Now we wish to prove that

$\forall n \geq 1$ and $k \geq 2, (X_1, \dots, X_n) \perp\!\!\!\perp (X_{n+2}, \dots, X_{n+k}) | X_{n+1}$.

By induction (in k), suppose that

$\forall n \geq 1, (X_1, \dots, X_n) \perp\!\!\!\perp (X_{n+2}, \dots, X_{n+k-1}) | X_{n+1}$ and

$(X_1, \dots, X_{n+1}) \perp\!\!\!\perp (X_{n+3}, \dots, X_{n+k}) | X_{n+2}$.

With the same argument as in step 1, we conclude step 2.

Step 3 - Finally, we wish to prove (5.2). But (5.2) follows directly from step 2 and Theorem 3. \square

Having established the concept of Markov Chains, the following properties are immediately stated:

A - If (X_1, X_2, \dots) forms a Markov Chain, so does (\dots, X_2, X_1) . [From Definition 3a.]

B - Any subsequence of a Markov Chain is a Markov Chain. [From DROP/ADD principles.]

C - If (X_1, X_2, \dots) forms a Markov Chain, then $\forall n \geq m + 2, m \geq 1$

$(X_1, \dots, X_m) \perp\!\!\!\perp (X_{m+2}, \dots, X_n) \perp\!\!\!\perp (X_{n+2}, \dots) | (X_{m+1}, X_{n+1})$.

To prove property C, it is enough to have

PROPOSITION 9

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

$$X_1 \perp\!\!\!\perp X_3 \perp\!\!\!\perp X_5 \mid (X_2, X_4).$$

PROOF

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

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 $\forall n \geq 1, (X_1, \dots, X_n) \perp\!\!\!\perp X_{n+2} \mid (Z, X_{n+1})$. In this case, we say that (X_1, X_2, \dots) 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, \dots) 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, \dots) 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 Θ is the usual

parameter "space". In addition, the Bayesians consider 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 $P_\theta(A)$.

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

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

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

$$\Pi(B \times A) = \int_B P_\theta(A) \xi(d\theta).$$

The uniqueness of Π and the fact that Π 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, \mathcal{A}) in the following way:

$$P(A) = \Pi(\Theta \times A)$$

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

$$F_X \equiv \{\theta \times A; A \in \mathcal{A}\} \text{ and}$$

$$F_Y \equiv \{B \times X; B \in \mathcal{B}\}.$$

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 $X \perp\!\!\!\perp Y|X_1$ we say that X_1 is sufficient for X with respect to Y .

b) If $X_2 \perp\!\!\!\perp Y$ 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) $X_1 \perp\!\!\!\perp X_2|Y$.

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 chapter 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 $X_1 \perp\!\!\!\perp X_2|Y$ we have $X \perp\!\!\!\perp Y|(X_1, X_2)$, then $X \perp\!\!\!\perp Y|X_1$.

COROLLARY 4

Let X , Y , and Z be three random objects such that $X \perp\!\!\!\perp Y|Z$ and $X \perp\!\!\!\perp Z|Y$. Then, $X \perp\!\!\!\perp (Y, Z)|Y \wedge Z$.

Note that Corollary 4 shows that relations a) and c) imply b) if $X_1 \wedge Y$ is essentially constant on Ω (that is, essentially generates F_0). This condition will be studied in Section 6 in connection with Basu's second result.

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. Udney Yule in terms of collapsibility of contingency tables. It must clarify the problems with Simpson's paradox in Examples 2 and 4.

PROPOSITION 12

Let X , Y , and Z be three random objects such that $F_Z \equiv \{\phi, \Omega, A, A^c\}$ with $0 < \Pi(A) < 1$. If $X \perp\!\!\!\perp Y$ and $X \perp\!\!\!\perp Y|Z$, then either $X \perp\!\!\!\perp Z$ or $Y \perp\!\!\!\perp Z$.

The proof becomes simple when we recognize the following general result:

LEMMA 1

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

$$E\{I_A|(X, Y)\} = [\Pi(A)]^{-1}E\{I_A|X\}E\{I_A|Y\}.$$

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 $(X_2, X) \equiv X$. \square

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 of this chapter.

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 $X \perp\!\!\!\perp Y|X_0 \wedge X_1$. (If X_0 and X_1 are sufficient for X , then so is $X_0 \wedge X_1$.)

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 \dagger 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,

$$\int_{BC} I_A^* I_A^\dagger d\Pi = \int_{I_B} I_A^* I_C I_A^\dagger d\Pi = \int I_{AB}^* I_{AC}^\dagger d\Pi =$$

$$[\int I_{AB}^* d\Pi][\int I_{AC}^\dagger d\Pi] = [\int I_{AB} d\Pi][\int I_{AC} d\Pi].$$

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

$$\int_{BC} I_A^* I_A^\dagger d\Pi = \Pi(AB)\Pi(AC) = [\Pi(A)]^2 \Pi(B|A)\Pi(C|A),$$

where $\Pi(B|A) = \frac{\Pi(AB)}{\Pi(A)}$. We notice now that on the atom A , the functions I_B^{*Z} and I_C^{*Z} are constants and equal respectively to $\Pi(B|A)$ and $\Pi(C|A)$. Analogously, the function I_{BC}^{*Z} is equal to $\Pi(BC|A)$ on A . On the other hand since $X \perp\!\!\!\perp Y|Z$, $I_B^{*Z} I_C^{*Z} \equiv I_{BC}^{*Z}$; thus $\Pi(B|A)\Pi(C|A) = \Pi(BC|A)$. This shows that

$$\int_{BC} I_A^* I_A^\dagger d\Pi = \Pi(A)\Pi(ABC).$$

To conclude the proof we must prove that $\Pi(A) \int_D I_A d\Pi = \int_D I_A^* I_A^\dagger d\Pi$ 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. \square

PROOF OF PROPOSITION 12

Let $p = \Pi(A)$, $I^* = E\{I_A | X\}$, and $I^\dagger = E\{I_A | Y\}$. From Lemma 1 we have that

$$E\{I_A | (X, Y)\} = \frac{I^* I^\dagger}{p} \text{ and } E\{I_{A^c} | (X, Y)\} = \frac{(1 - I^*)(1 - I^\dagger)}{1 - p}.$$

Clearly

$$\frac{I^* I^\dagger}{p} + \frac{(1 - I^*)(1 - I^\dagger)}{1 - p} = 1;$$

that is,

$$\left(1 - \frac{I^*}{p}\right) \left(1 - \frac{I^\dagger}{p}\right) = 0.$$

Since $X \perp\!\!\!\perp Y$, this last equation holds if and only if either $\frac{I^*}{p} \equiv 1$ or $\frac{I^\dagger}{p} \equiv 1$ 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., $Y_1 \perp\!\!\!\perp Y_2$ and $Y_1 \perp\!\!\!\perp Y_2 | X$), then from Lemma 1,

$$(5.3) \quad E\{I_A | Y\} = [\pi(A)]^{-1} E\{I_A | Y_1\} E\{I_A | Y_2\},$$

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 $X_2 \perp\!\!\!\perp X_1$ since $X_2 \subset Y$; that is, $X_2 \equiv Y \wedge X_1$ is essentially a constant. \square

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., $A \in \mathcal{A}$) 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

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

PROPOSITION 13

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

From DROP/ADD principles we have that $X \perp\!\!\!\perp Y | (X_1, Y_1)$ and $X \perp\!\!\!\perp Y | (X_2, Y_2)$. Thus, by Proposition 11,

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

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 $X \perp\!\!\!\perp Y_2 | (X_1, Y_1)$ and $X \perp\!\!\!\perp Y_1 | (X_2, Y_2)$, then $X \perp\!\!\!\perp Y | (X_1, X_2)$ if and only if $X \perp\!\!\!\perp Y_1 | (X_1, X_2)$ [equivalently $X \perp\!\!\!\perp Y_2 | (X_1, X_2)$].

Note that 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 . 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 $(X, Z) \wedge (Y, Z) \equiv Z$. 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).

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 $L_{\infty}(Y) \xrightarrow{*} L(X)$ and $L_{\infty}(X) \xrightarrow{\dagger} L_{\infty}(Y)$, where $*$ is for $*X$, and \dagger 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 $L_{\infty}(X) \xrightarrow{\dagger} L_{\infty}(Y)$ is essentially one-one if $f_1^{\dagger} \neq f_2^{\dagger}$ whenever $(f_1, f_2) \in X$ 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 $f \in X$, $f^{\dagger} = 0$ 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 $L_{\infty}(Y) \xrightarrow{*} L_{\infty}(X)$ is essentially onto if for every $f \in X$ there is a $g \in Y$ such that $g^* = f$.

The following result relates these two definitions.

PROPOSITION 15

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

PROOF

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

$$E\{fh\} = E\{fg^*\} = E\{fg\} = E\{f^{\dagger}g\} = 0.$$

Since h is arbitrary, $f = 0$. \square

Let $X_{[Y]}$ be the random object that generates the smallest subfield that contains all functions g^* where $g \in Y$. 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) $\forall g \in Y$, $E\{g | (X, X_{[Y]})\} = g^* \in X_{[Y]}$ by definition.(ii) $\forall g \in Y$, $E\{g | (X, X_1)\} = E\{g | X\} = E\{g | X_1\}$.

Then for every $g \in Y$, $g^* \in X_1$. 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 17If $X \ll Y$, then $X_{[Y]} \equiv X$.PROOFNote that $X \perp\!\!\!\perp Y | X_{[Y]}$. Thus,

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

For $f^\dagger = E\{f | X_{[Y]}\}$ since $X \ll Y$, we have that
$$E\{(f - f^\dagger) | Y\} = 0 \rightarrow f = f^\dagger. \quad \text{Then } \forall f \in X, f \in X_{[Y]} \text{ and } X \equiv X_{[Y]}. \quad \square$$

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

PROOF

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

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

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

REMARK

4 - 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 $f \in (X, Z)$, $E\{f|(Y, Z)\} = 0$ 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).

THEOREM 5

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 $X \perp\!\!\!\perp Y|Z \forall f \in X$, $E\{f|(Y, Z)\} = E\{f|Z\}$. On the other hand, since $X \perp\!\!\!\perp Y$, $E\{f|Y\} = E\{f\}$ but, by Proposition 1, $E\{f|Y\} = E\{E\{f|(Y, Z)\}|Y\} = E\{E\{f|Z\}|Y\}$. Then $E\{f\} = E\{E\{f|Z\}|Y\}$ 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 $X \perp\!\!\!\perp Y$ and $X \perp\!\!\!\perp Y|Z$ implies $X \perp\!\!\!\perp Z$. Now, by Proposition 5 we have that $X \perp\!\!\!\perp Y|Z$ and $X \perp\!\!\!\perp Z$ is equivalent to $X \perp\!\!\!\perp Z|Y$ and $X \perp\!\!\!\perp Y$. \square

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

Lehmann and Scheffé (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 $X_1 \in X$ and $X \perp\!\!\!\perp Y|X_1$. If $X_1 \ll Y$ then $X_1 \equiv X_{[Y]}$.

REFERENCES

1. ASH, R. B. (1972). Real Analysis and Probability. Academic Press, New York.
2. BAHADUR, R. R. (1955). Measurable Subspaces and Subalgebras. Proc. Amer. Math. Soc., 6, 565-70.
3. BARNDORFF-NIELSEN, O. (1978). Information and Exponential Families in Statistical Theory. John Wiley, New York.
4. BASU, D. (1955). On Statistics Independent of a Complete Sufficient Statistic. Sankhya A, 15, 377-80.
5. BASU, D. (1958). On Statistics Independent of a Sufficient Statistic. Sankhya A, 20, 223-26.
6. BASU, D. (1959). The Family of Ancillary Statistics. Sankhyā A, 21, 247-56.
7. BASU, D. (1964). Recovery of Ancillary Information. Sankhyā A, 26, 3-16.
8. BASU, D. (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.
9. BASU, D. (1977). On the Elimination of Nuisance Parameters. JASA, 72, 355-66.
10. BASU, D. (1978). On Partial Sufficiency: A Review. J. Statist. Plan. Inf., 2, 1-13.
11. BASU, D. and CHENG, S. C. (1979). A Note on Sufficiency in Coherent Models, Int. J. Math. Math. Sci. To appear.
12. BURKHOLDER, D. L. (1961). Sufficiency in the Undominated Case. Ann. Math. Statist., 32, 1191-200.
13. CHENG, S. C. (1978). A Mathematical Study of Sufficiency and Adequacy in Statistical Theory. Ph.D. Dissertation, FSU, Florida.

14. DAWID, A. P. (1979a). Conditional Independence in Statistical Theory. JRSS, B, 41, 1-31.
15. DAWID, A. P. (1979b). Some Misleading Arguments Involving Conditional Independence. JRSS, B, 41, 249-52.
16. DAWID, A. P. (1979c). A Bayesian Look at Nuisance Parameters. Trabajos de Estadística. To appear.
17. DAWID, A. P. (1980). Conditional Independence for Statistical Operations. Ann. Statist., 8, 598-617.
18. 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.
19. DE FINETTI, B. (1970). Theory of Probability, Vols. 1 and 2. Translated edition, 1974. John Wiley, London.
20. DOOB, J. L. (1953). Stochastic Processes. John Wiley, New York.
21. FERREIRA, P. E. (1980). Comments on Berkson's Paper "In Dispraise of ...". Unpublished report.
22. FISCHER, 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.
23. FISCHER, R. A. (1922). On the Mathematical Foundations of Theoretical Statistics. Phil. Trans. A, 222, 309-68.
24. FLORENS, J. P. and MOUCHART, M. (1977). Reduction of Bayesian Experiments. CORE, Discussion Paper 7737.
25. GODAMBE, V. P. (1979). On Sufficiency and Ancillarity in Presence of Nuisance Parameter. Unpublished report.
26. HALL, W. J., WIJSMAN, R. A., and GOSH, J. K. (1965). The Relationship Between Sufficiency and Invariance. Ann. Math. Statist., 36, 375-614.
27. KOEHN, U. and THOMAS, D. L. (1975). On Statistics Independent of a Sufficient Statistic: Basu's Lemma. Amer. Statist., 29, 40-2.

28. 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.
29. LEHMAN, E. L. and SCHEFFÉ, H. (1950). Completeness, Similar Regions, and Unbiased Estimation, Part I. Sankhya A, 10, 305-40.
30. MOUCHART, M. and ROLIN, J. M. (1978). A Note on Conditional Independence. Unpublished report.
31. MOY, S. T. C. (1954). Characterization of Conditional Expectation as a Transformation on Function Spaces. Pacific J. Math, 4, 47-63.
32. PATHAK, P. K. (1975). Note on Basu's Lemma. Unpublished report.
33. PICCI, G. (1977). Some Connections Between the Theory of Sufficient Statistics and the Identifiability Problem. SIAM J. Appl. Math., 33, 383-98.

CHAPTER II. ON THE BAYESIAN ANALYSIS OF CATEGORICAL DATA:
THE PROBLEM OF NONRESPONSE

1 - INTRODUCTION

The simplest case of the problem of nonresponse is as follows. Let π_1 be the unknown proportion of individuals in a certain population, P , that belong to a particular category A_1 . With π_1 as the only parameter of interest, a survey is conducted using a simple random sample of size n . Of the n individuals surveyed, n_1 respond to the question "Do you belong to category A_1 ?" with a yes/no answer, but $n_2 = n - n_1$ individuals do not respond. Denoting the category of respondents by R , and the complementary category by R' , the survey data may be summarized as:

(1.1)

	R	R'	
A_1	x_1	n_2	
A_2	x_2		
	n_1	n_2	n

with A_2 being the complement of A_1 .