Conditional Independence for Statistical Operations

A. Philip Dawid

Stable URL:
http://links.jstor.org/sici?sici=0090-5364%28198005%298%3A3%3C598%3ACIFSO%3E2.0.CO%3B2-D

CONDITIONAL INDEPENDENCE FOR STATISTICAL OPERATIONS

By A. Philip Dawid

University College London and The City University

A general calculus of conditional independence is developed, suitable for application to a wide range of statistical concepts such as sufficiency, parameter-identification, adequacy and ancillarity. A vehicle for this theory is the statistical operation, a structure-preserving map between statistical spaces. Concepts such as completeness and identifiability of mixtures arise naturally and play an important part. Some general theorems are exemplified by applications to ancillarity, including a study of a Bayesian definition of ancillarity in the presence of nuisance parameters.

1. **Introduction.** Concepts of conditional independence play an important rôle in unifying many seemingly unrelated ideas of statistical inference (Dawid, 1976, 1979a; Dawid and Dickey, 1977a, 1977b). If random variables X and Y are independent, given Z, we may write $X \perp\!\!\!\perp Y | Z$. The most fruitful intuitive interpretation of this statement is that the conditional distributions of X, given Y and Z, are in fact governed by the value of Z alone, further information about the value of Y being irrelevant. This intuitive property extends readily to statements such as $X \perp\!\!\!\perp \Theta | T$, in which X is a random variable with distributions governed by a parameter Θ, and T is (say) a function of X: a moment's reflection will show that this is just the requirement that T is *sufficient* for Θ based on data X. Similarly, $T \perp\!\!\!\perp \Theta$ (with the conditioning variable trivial) if and only if T is an *ancillary* statistic.

If T is sufficient, and Θ has any prior distribution, then the symmetry of conditional independence shows that $\Theta \perp\!\!\!\perp X | T$, which says that the posterior distribution of Θ is determined by T alone. This Bayesian characterization of sufficiency would appear to be due to Kolmogorov (1942). It has been taken up in a very general setting of conditional independence by Le Cam (1964, page 1439).

This approach suggests an obvious way to deal in general with the above extensions of conditional independence to cases involving parameters: if X, Y and Z are variables in a statistical model (which may be functions of the parameters as well as of the data), we might write $X \perp\!\!\!\perp Y | Z$ if this statement holds in the joint distribution generated by giving the parameter an arbitrary prior distribution. However, this natural approach is not entirely satisfactory. For example, in the above case of sufficiency, $\Theta \perp\!\!\!\perp X | T$ for all prior distributions implies only that T is *pairwise* sufficient (Martin, et al., 1973). Another difficulty arises in considering prediction sufficiency (Torgersen, 1977), where we consider data (X, Y) with joint

distribution depending on a parameter Θ, and wish to predict Y from X. From a Bayesian viewpoint, a function T of X is "sufficient" for this purpose if, for any prior distribution, $Y \perp\!\!\!\perp X \,|\, T$. Now this can arise in two essentially different ways: in both cases, $Y \perp\!\!\!\perp X \,|\, T$ in any of the sampling distributions, but in the first case we have T sufficient for Θ based on X (so that T is *adequate* in the sense of Skibinsky, 1967); while in the second, we have the conditional distribution of Y given T not depending on Θ. From the sampling theory point of view it seems worthwhile distinguishing these two states of affairs, and so we are led to seek a general definition of conditional independence which will be more sensitive than that motivated by the above Bayesian considerations.

In this paper we attack the problem of constructing such a rigorous general theory of conditional independence, which should be adequate for most applications. We shall need to deal rigorously and meaningfully with such heuristic notions as "the distributions of X given T and Θ", and to do this we shall use the idea of *statistical operation* (Morse and Sacksteder, 1966; Sacksteder, 1967). This is a slight generalization of *Markovian operator* (Neveu, 1965, V.4), which is like a conditional expectation operator and itself generalizes the notion of a *transition probability* (or *Markov kernel*). Sections 2 through 4 below are devoted to some background theory for such statistical operations.

An important consideration is that, as with conditional expectation, we may get several versions of the answer when we apply a statistical operation, and shall not wish to distinguish between them. In order to take account of this we introduce the idea of a *statistical space*, in which certain sets are ignorable, in Section 2. Section 3 defines general statistical operations and gives some important examples. Section 4 introduces a dual approach to statistical operations, analogous to the passage, in a parametric family of distributions, from a distribution over the parameter-space to the induced marginal distribution of the data. It also introduces the important concepts of bounded completeness and strong identification (identification of mixtures) for a statistical operation.

In Section 5 we introduce the general concept of conditional independence for a statistical operation, and show how it encompasses the various intuitive properties such as parametric sufficiency ($X \perp\!\!\!\perp \Theta \,|\, \Phi$), sufficiency ($X \perp\!\!\!\perp \Theta \,|\, T$), pointwise independence ($X \perp\!\!\!\perp Y \,|\, \Theta$), adequacy ($X \perp\!\!\!\perp (Y, \Theta) \,|\, T$), etc. Sections 6 and 7 deal with variations on the result ($X \perp\!\!\!\perp Z$ and $X \perp\!\!\!\perp Y \,|\, Z) \Leftrightarrow X \perp\!\!\!\perp (Y, Z)$ and some converses. Finally, in Section 8 we present applications of the theory: to rectify a well-known fallacious argument, and to consider the relationship between a classical and a Bayesian approach to ancillarity in the presence of nuisance parameters.

   2. Ignorable sets. This section introduces the concept of a *statistical space*, which takes into account the fact that many statistical concepts (such as density functions or conditional expectations) are not uniquely defined, but have several equivalent versions which we shall not be interested in distinguishing. Intuitively, the event that two such versions differ is in some sense ignorable. Because the

statistical operations to be introduced in Section 3 generalize the concept of conditional expectation, it is necessary first to investigate the sense in which different versions may be equivalent. The formalism presented here encompasses most of the technicalities which are needed in dealing with sets of measure zero.

2.1. *σ-ideals.* Let $\mathcal{S}$ be a σ-field of subsets of $\mathcal{X}$. For many statistical purposes, there will be sets in $\mathcal{S}$ which can be regarded as *ignorable.* For example, if we have a single probability measure P over $\mathcal{S}$, we can ignore $E \in \mathcal{S}$ if $P(E) = 0$; while if $\mathcal{S}$ is endowed with a family $\mathcal{P} = \{P_\lambda : \lambda \in \Lambda\}$ of distributions, $E \in \mathcal{S}$ is ignorable if $P_\lambda(E) = 0$ for all $\lambda \in \Lambda$. We shall denote the classes of sets obtained thus by $\mathcal{I}(P)$, $\mathcal{I}(\mathcal{P})$ respectively.

In the latter case it will often be useful to work with the product space $\mathcal{X} \times \Lambda$, having the induced σ-field $\mathcal{S}^*$, where $A \in \mathcal{S}^*$ if $A_\lambda \in \mathcal{S}$, all $\lambda \in \Lambda$. (Here $A_\lambda = \{x : (x, \lambda) \in A\}$, the *section* of A at $\lambda \in \Lambda$.) We would wish to ignore a set $A \in \mathcal{S}^*$ if $P_\lambda(A_\lambda) = 0$, all $\lambda \in \Lambda$. We shall denote the class of such sets by $\mathcal{I}^*(\mathcal{P})$.

A suitable abstract formulation of the concept of a "family of ignorable sets" (which covers the cases above) is the *σ-ideal,* as given by the following definition.

DEFINITION 2.1.   Let $\mathcal{S}$ be a σ-field, and $\mathcal{I} \subseteq \mathcal{S}$ a family of sets satisfying
   (i) $\varnothing \in \mathcal{I}$;
   (ii) $\mathcal{I}$ is closed under countable union;
   (iii) $I \in \mathcal{I}, E \in \mathcal{S} \Rightarrow I \cap E \in \mathcal{I}$.
Then $\mathcal{I}$ is termed a *σ-ideal (in $\mathcal{S}$).*

Clearly $\mathcal{S}$ itself is a σ-ideal in $\mathcal{S}$, but we shall exclude this case as of no interest; thus we may suppose $\mathcal{I}$ not to contain the whole space $\mathcal{X}$, in which case it may be termed a *proper σ-ideal.* We note that $\{\varnothing\}$ is a σ-ideal (the *trivial σ-ideal*) in $\mathcal{S}$. An arbitrary intersection of σ-ideals is again a σ-ideal. (It follows that any subclass of $\mathcal{S}$ generates a σ-ideal, the smallest σ-ideal containing it; however this may not be proper.)

If $\mathcal{I}$ is a σ-ideal in $\mathcal{S}$, we shall call $(\mathcal{S}, \mathcal{I})$ a *statistical space.*

2.2. *Equivalence.*   Let $(\mathcal{S}, \mathcal{I})$ be a statistical space. Denote by $\mathcal{L}(\mathcal{S})$ the space of all $\mathcal{S}$-measurable real functions (random variables), and by $\mathcal{L}^\infty(\mathcal{S})$ $[\mathcal{L}^\infty_+(\mathcal{S})]$ the subspace of all bounded [nonnegative bounded] functions.

If π is a proposition that may be true or false of points in the underlying space $\mathcal{X}$, we write $\pi[\mathcal{I}]$ to denote that π holds for all points outside some set in $\mathcal{I}$. (If $\mathcal{I} = \mathcal{I}(P)$ or $\mathcal{I}(\mathcal{P})$, we shall write $\pi[P]$ or $\pi[\mathcal{P}]$.) We can define a relation $\sim$ (to be explicit, $\sim_\mathcal{I}$, $\sim_P$ or $\sim_\mathcal{P}$) on $\mathcal{L}(\mathcal{S})$ by: $U \sim V$ if $U = V[\mathcal{I}]$, and this is readily seen to be an equivalence relation. Since events in $\mathcal{I}$ are to be construed as ignorable, we shall not wish to distinguish between equivalent random variables. Thus we are really only concerned with the *quotient space* $L(\mathcal{S}, \mathcal{I})$ consisting of the equivalence classes of $\mathcal{L}(\mathcal{S})$ under $\sim$, (written $L(\mathcal{S}, \mathcal{P})$ if $\mathcal{I} = \mathcal{I}(\mathcal{P})$, etc.) and its subspaces $L^\infty(\mathcal{S}, \mathcal{I})$, $L^\infty_+(\mathcal{S}, \mathcal{I})$ generated by $\mathcal{L}^\infty(\mathcal{S})$, $\mathcal{L}^\infty_+(\mathcal{S})$ respectively. (If $\tilde{U}$ is such an equivalence class, any $U \in \tilde{U}$ may be called a *version* of $\tilde{U}$). So long as we only

deal with a countable number of operations, these quotient spaces inherit much of the algebraic structure of their generating spaces. For if $U_i \sim V_i (i = 1, 2, \cdots)$, then there exists $I \in \mathcal{I}$ such that, for all i, and all $x \notin I$, $U_i(x) = V_i(x)$. So if we perform, pointwise, any arithmetic operation (summation, supremum, etc.) on the U's and again on the V's, we shall get equivalent answers. Henceforth we shall apply such operations without further comment.

We can extend the notion of equivalence to sets in $\mathcal{E}$. For $A, B \in \mathcal{E}$, we write $A \sim B$ if $\chi_A \sim \chi_B (\chi_A$ denoting the indicator variable taking value 1 in A and 0 elsewhere); that is to say, $A \sim B$ if the symmetric difference $A \triangle B \in \mathcal{I}$. Now consider two sub-$\sigma$-fields $\mathcal{F}, \mathcal{G}$ of $\mathcal{E}$. We write $\mathcal{F} \overset{\sim}{\subset} \mathcal{G}$ if, for all $F \in \mathcal{F}$, there exists $G \in \mathcal{G}$ such that $F \sim G$. If $\mathcal{F} \overset{\sim}{\subset} \mathcal{G}$ and $\mathcal{G} \overset{\sim}{\subset} \mathcal{F}$, we call $\mathcal{F}$ and $\mathcal{G}$ *equivalent* σ-fields, and write $\mathcal{F} \sim \mathcal{G}$.

Denote by $\tilde{\mathcal{F}}$ the σ-field generated by $\mathcal{F}$ and $\mathcal{I}$. It is easy to see that $\tilde{\mathcal{F}} = \{E \in \mathcal{E} : E \sim F, \text{ some } F \in \mathcal{F}\}$. Thus $\mathcal{F} \overset{\sim}{\subset} \mathcal{G}$ if and only if $\tilde{\mathcal{F}} \subseteq \tilde{\mathcal{G}}$, and $\mathcal{F} \sim \mathcal{G}$ if and only if $\tilde{\mathcal{F}} = \tilde{\mathcal{G}}$.

LEMMA 2.1.
  (i) $\mathcal{F} \vee \mathcal{G} \sim \tilde{\mathcal{F}} \vee \tilde{\mathcal{G}} \sim \tilde{\mathcal{F}} \vee \tilde{\mathcal{G}}$;
  (ii) $\mathcal{F} \cap \mathcal{G} \overset{\sim}{\subset} \tilde{\mathcal{F}} \cap \tilde{\mathcal{G}}$;
  (iii) $\mathcal{F} \cap \tilde{\mathcal{G}} \sim \tilde{\mathcal{F}} \cap \tilde{\mathcal{G}}$.

PROOF. (i) and (ii) are trivial. For (iii), take $S \in \tilde{\mathcal{F}} \cap \tilde{\mathcal{G}}$. Thus $S \in \tilde{\mathcal{G}}$, and for some $F \in \mathcal{F}, S \sim F$. It follows that $F \in \tilde{\mathcal{G}} = \tilde{\mathcal{G}}$, and so $F \in \mathcal{F} \cap \tilde{\mathcal{G}}$, so that $\tilde{\mathcal{F}} \cap \tilde{\mathcal{G}} \overset{\sim}{\subset} \mathcal{F} \cap \tilde{\mathcal{G}}$.

In general it is not *not* true that $\mathcal{F} \cap \mathcal{G} \sim \tilde{\mathcal{F}} \cap \tilde{\mathcal{G}}$.

The following lemma is essentially the same as Lemma 7.1 of Bahadur (1954).

LEMMA 2.2. $U \in \mathcal{L}(\tilde{\mathcal{F}})$ *if and only if* $U \sim V$ *for some* $V \in \mathcal{L}(\mathcal{F})$.

It follows that we could, without loss of generality, consider only *completed* sub-σ-fields of $\mathcal{E}$ (that is to say those containing $\mathcal{I}$), and random variables measurable with respect to these.

3. Statistical operations.

3.1. *Conditional expectation.* Let $(\mathcal{X}, \mathcal{S}, P)$ be a probability space, and $\mathcal{E}$ a subfield of $\mathcal{S}$. We denote by $P^{\mathcal{E}}$ the restriction of P to $\mathcal{E}$, and introduce the σ-ideals $\mathcal{I} = \mathcal{I}(P)$ in $\mathcal{S}$, and $\mathcal{I} = \mathcal{I}(P^{\mathcal{E}})$ in $\mathcal{E}$.

For any $U \in \mathcal{L}^{\infty}(\mathcal{S})$ we can define the *conditional expectation* of U given $\mathcal{E}$: $\hat{U} = E(U|\mathcal{E})$, with $\hat{U} \in \mathcal{L}^{\infty}(\mathcal{E})$. Such a $\hat{U}$ is not uniquely determined: any variable V such that $V \sim_{\mathcal{I}} \hat{U}$ also serves as a version of $E(U|\mathcal{E})$, and only such variables so serve. Thus the conditional expectation operation may be regarded as a map $\Pi : \mathcal{L}^{\infty}(\mathcal{S}) \rightarrow L^{\infty}(\mathcal{E}, \mathcal{I})$. Furthermore, if $U \sim_{\mathcal{I}} U'$, then $E(U|\mathcal{E}) \sim_{\mathcal{I}} E(U'|\mathcal{E})$, so that an alternative representation is as a map from $L^{\infty}(\mathcal{S}, \mathcal{I})$ into $L^{\infty}(\mathcal{E}, \mathcal{I})$. (In fact this extends to a map from $L^1(\mathcal{S}, P)$ into $L^1(\mathcal{E}, P^{\mathcal{E}})$, which has

the properties of a *Markovian operator* (Neveu, 1965, V.4).)

The map $\Pi : \mathcal{L}^{\infty}(\mathbb{S}) \to L^{\infty}(\mathbb{S}, \mathcal{I})(or \ \Pi : L^{\infty}(\mathbb{S}, \mathcal{I}) \to L^{\infty}(\mathbb{S}, \mathcal{I}))$ has the following properties.

(P1). *Linearity.* $\Pi(a_1 U_1 + a_2 U_2) = a_1.\Pi U_1 + a_2.\Pi U_2$.

(P2). *Positivity.* $U \geqslant 0 \Rightarrow \Pi U \geqslant 0$.

(P3). *Normalization.* $\Pi 1 = 1$.

(P4). *Continuity.* If (U_i) is a countable sequence decreasing monotonically to 0, then so is (ΠU_i). (In fact (P4) plays no essential rôle in the sequel, and could probably be removed).

In many statistical problems we deal with operations which behave very much like conditional expectation above. We present below some important examples, which will recur throughout this paper, involving a family $\mathcal{P} = \{\mathcal{P}_\lambda : \lambda \in \Lambda\}$ of probability distributions over $\mathbb{S}$. We may suppose Λ to have been endowed with a σ-field $\mathcal{H}$ such that the map $\lambda \rightsquigarrow P_\lambda(A)$ is $\mathcal{H}$-measurable for all $A \in \mathbb{S}$. (We may equivalently regard $\mathcal{P}$ as a *Markov kernel* from $(\Lambda, \mathcal{H})$ to $(\mathcal{X}, \mathbb{S})$.)

EXAMPLE I(a). If $U \in \mathcal{L}^{\infty}(\mathbb{S})$, we may define a function $\hat{U}$ on Λ by: $\hat{U}(\lambda) = E_\lambda(U)$. Then $\hat{U} \in \mathcal{L}^{\infty}(\mathcal{H})$ (Meyer, 1966, IX, 3), and is *uniquely* defined; equivalently, $\hat{U} \in L^{\infty}(\mathcal{H}, \mathcal{I}_0)$ where $\mathcal{I}_0 = \{\varnothing\}$, the trivial σ-ideal. Writing $\hat{U} = \Pi_1 U$, $\Pi_1 : \mathcal{L}^{\infty}(\mathbb{S}) \to \mathcal{L}^{\infty}(\mathcal{H})$ satisfies (P1) through (P4). In this case, Π_1 is just a representation of $\mathcal{P}$ itself. Since $U_1 = U_2[\mathcal{P}] \Rightarrow \hat{U}_1 = \hat{U}_2$, we can alternatively consider $\Pi_1 : L^{\infty}(\mathbb{S}, \mathcal{P}) \to \mathcal{L}^{\infty}(\mathcal{H})$.

EXAMPLE II(a). Let $\mathcal{E}$ be a sub-σ-field of $\mathbb{S}$, and construct the statistical space $(\mathcal{E}^*, \mathcal{I}^*)$ in $\mathcal{X} \times \Lambda$, with $\mathcal{I}^* = \mathcal{I}^*(\mathcal{P})$, as in Section 2.1. For $U \in \mathcal{L}^{\infty}(\mathbb{S})$ we can consider the "conditional expectation of U given $\mathcal{E}$ and λ": $\hat{U}(x, \lambda) = E_\lambda(U|\mathcal{E})(x)$. Then $\hat{U}$ is in $\mathcal{L}^{\infty}(\mathcal{E}^*)$, but is only defined *modulo* $\mathcal{I}^*$. The induced map $\Pi_2 : \mathcal{L}^{\infty}(\mathbb{S}) \to L^{\infty}(\mathcal{E}^*, \mathcal{I}^*)$ satisfies (P1) through (P4). Again we could take $\Pi_2 : L^{\infty}(\mathbb{S}, \mathcal{P}) \to L^{\infty}(\mathcal{E}^*, \mathcal{I}^*)$.

REMARK 3.1. It would often be convenient if $\mathcal{E}^*$ in Example II(a) above could be replaced be the ordinary tensor product σ-field $\mathcal{E} \otimes \mathcal{H}$. This will be the case if, for any $U \in \mathcal{L}^{\infty}(\mathbb{S})$, we could choose versions of $\hat{U}(x, \lambda)$ to be jointly $\mathcal{E} \otimes \mathcal{H}$-measurable. Then we will call the σ-field $\mathcal{E}$ *regular* (with respect to $\mathcal{P}$). If so, we can take $\Pi_2 : L^{\infty}(\mathbb{S}, \mathcal{P}) \to L^{\infty}(\mathcal{E} \otimes \mathcal{H}, \mathcal{I})$ where $\mathcal{I} = \mathcal{I}^* \cap (\mathcal{E} \otimes \mathcal{H})$.

If Λ is countable, then any σ-field $\mathcal{E}$ is clearly regular. If $\mathcal{E}$ is separable (countably generated) then $\mathcal{E}$ is regular with respect to any $\mathcal{P}$; this may be shown by an argument based on Meyer (1966, VIII, 10). It now follows that $\mathcal{E}$ will be regular with respect to $\mathcal{P}$ if $\mathcal{E} = \mathcal{F} \vee \mathcal{I}(\mathcal{P})$ for some separable $\mathcal{F}$, since we can then take $E_\lambda(U|\mathcal{E})(x) = E_\lambda(U|\mathcal{F})(x)$. In particular, if $\mathbb{S}$ is separable and $\mathcal{P}$ is dominated, this property holds for any completed subfield $\mathcal{E}$ of $\mathbb{S}$, as follows on applying Lemma 3 of Bahadur (1955) to the probability space $(\mathcal{X}, \mathbb{S}, \mu)$, where μ is such that $\mathcal{I}(\mu) = \mathcal{I}(\mathcal{P})$.

EXAMPLE III(a). Suppose $\mathcal{E}$, $\mathcal{F}$ are sub-σ-fields of $\mathcal{S}$, such that $\mathcal{E}$ is *sufficient* for $\mathcal{P}$ over $\mathcal{F}$ (and thus over $\mathcal{E} \vee \mathcal{F}$) : that is, for any $U \in \mathcal{L}^{\infty}(\mathcal{F})$ there exists $\hat{U} \in \mathcal{L}^{\infty}(\mathcal{E})$ which serves as a version of $E_{\lambda}(U|\mathcal{E})$ for all $\lambda \in \Lambda$. We shall write $\hat{U} = E_{\mathcal{P}}(U|\mathcal{E})$. Then $E_{\mathcal{P}}(U|\mathcal{E})$ is defined *modulo* $\mathcal{P}^{\mathcal{E}} = \{ P^{\mathcal{E}} : P \in \mathcal{P} \}$, and we thus get a map $\Pi_3 : \mathcal{L}^{\infty}(\mathcal{F}) \rightarrow L^{\infty}(\mathcal{E}, \mathcal{P}^{\mathcal{E}})$, or $\Pi_3 : L^{\infty}(\mathcal{F}, \mathcal{P}^{\mathcal{F}}) \rightarrow L^{\infty}(\text{FE}, \mathcal{P}^{\mathcal{E}})$, again satisfying (P1) through (P4).

3.2. *Statistical operations.* The above examples motivate the general concept of a statistical operation. Let $(\mathcal{E}, \mathcal{I})$, $(\mathcal{F}, \mathcal{J})$ be two statistical spaces.

DEFINITION 3.1. A map $\Pi : L^{\infty}(\mathcal{F}, \mathcal{J}) \rightarrow L^{\infty}(\mathcal{E}, \mathcal{I})$ satisfying (P1) through (P4) is termed a *statistical operation* (s.o.) *over* $(\mathcal{F}, \mathcal{J})$, *given* $(\mathcal{E}, \mathcal{I})$.

If, as seems reasonable, we regard statistical spaces as the principal objects of statistical study, then statistical operations are the structure-preserving maps. We thus have a statistical *category* of some interest (Sacksteder, 1967).

If $\mathcal{F} \subseteq \mathcal{E}$ and $\mathcal{J} \subseteq \mathcal{I}$, then $\mathcal{L}^{\infty}(\mathcal{F}) \subseteq \mathcal{L}^{\infty}(\mathcal{E})$, and $U_1 = U_2[\mathcal{J}] \Rightarrow U_1 = U_2[\mathcal{I}]$. Thus the identity mapping of $\mathcal{L}^{\infty}(\mathcal{F})$ into $\mathcal{L}^{\infty}(\mathcal{E})$ induces a trivial s.o., the *natural injection*, over $(\mathcal{F}, \mathcal{J})$ given $(\mathcal{E}, \mathcal{I})$. The case $\mathcal{F} = \mathcal{E}$, $\mathcal{J} \subset \mathcal{I}$ is of some special interest.

Let $\Pi : L^{\infty}(\mathcal{F}, \mathcal{J}) \rightarrow L^{\infty}(\mathcal{E}, \mathcal{I})$ be a s.o., and define $\mathcal{J}_{\Pi}$ as $\{ F \in \mathcal{F} : \Pi \tilde{\chi}_F = 0 \}$, where $\tilde{\chi}_F$ denotes the equivalence class of χ_F under $\sim_{\mathcal{J}}$. Then $\mathcal{J}_{\Pi}$ is a σ-ideal containing $\mathcal{J}$, and it is easy to see that Π factors through the natural injection $\Pi_0 : L^{\infty}(\mathcal{F}, \mathcal{J}) \rightarrow L^{\infty}(\mathcal{F}, \mathcal{J}_{\Pi})$. The induced s.o. form $L^{\infty}(\mathcal{F}, \mathcal{J}_{\Pi})$ into $L^{\infty}(\mathcal{E}, \mathcal{I})$ may also be denoted by Π. We obtain the same induced s.o. if we start with the s.o. from $\mathcal{L}^{\infty}(\mathcal{F})$ into $L^{\infty}(\mathcal{E}, \mathcal{I})$ got by composing $\Pi : L^{\infty}(\mathcal{F}, \mathcal{J}) \rightarrow L^{\infty}(\mathcal{E}, \mathcal{I})$ with the natural injection from $\mathcal{L}^{\infty}(\mathcal{F})$ into $L^{\infty}(\mathcal{F}, \mathcal{J})$. Thus we may restrict attention to s.o.'s which have the form $\Pi : \mathcal{L}^{\infty}(\mathcal{F}) \rightarrow L^{\infty}(\mathcal{E}, \mathcal{I})$, equivalent to $\Pi : L^{\infty}(\mathcal{F}, \mathcal{J}_{\Pi}) \rightarrow L^{\infty}(\mathcal{E}, \mathcal{I})$. We call $\mathcal{J}_{\Pi}$ the σ-ideal *induced* by Π.

3.3. *Construction by sufficiency.* Statistical operations which arise from the sufficiency construction of Example III(a) are of particular interest. We shall say that such a s.o. is *constructed by sufficiency* from $(\mathcal{P}; \mathcal{E}, \mathcal{F})$; it is a s.o. over $(\mathcal{F}, \mathcal{I}(\mathcal{P}^{\mathcal{F}}))$ given $(\mathcal{E}, \mathcal{I}(\mathcal{P}^{\mathcal{E}}))$.

By means of an appropriate auxiliary construction, it will frequently be possible to regard a given s.o. as constructed by sufficiency. For instance, in the set-up of the examples of Section 3.1 above, let $\mathfrak{M}$ be the family of all probability distributions on $\mathcal{K}$. We note that $\mathcal{I}(\mathfrak{M})$ is trivial, since, if $H_0 \in \mathcal{K}$ is nonempty, we can choose $\lambda_0 \in H_0$, and construct $M_0 \in \mathfrak{M}$ such that $M_0(H) = 1$ for $\lambda_0 \in H$, $M_0(H) = 0$ otherwise; and thus $M_0(H_0) > 0$. For $M \in \mathfrak{M}$ denote by Q_M the distribution over $\mathcal{S} \otimes \mathcal{K}$ given on $\mathcal{S} \times \mathcal{K}$ by : $Q_M(S \times H) = \int_H P_{\lambda}(S) dM(\lambda)$. Then identifying $\mathcal{S}$ with $\mathcal{S} \times \{\Lambda\} \subseteq \mathcal{S} \otimes \mathcal{K}$, etc., it is clear that $\mathcal{K}$ is sufficient for $\mathfrak{Q} = \{Q_M : M \in \mathfrak{M}\}$ over $\mathcal{S}$ and, if Π^* is the s.o. constructed by sufficiency from $(\mathfrak{Q}; \mathcal{K}, \mathcal{S})$, then $\Pi^* = \Pi_1$.

For Example II, suppose $\mathcal{E}$ is a *regular* sub-σ-field of $\mathcal{S}$ with respect to $\mathcal{P}$. Then, taking $\Pi_2 : \mathcal{L}^\infty(\mathcal{S}) \to L^\infty(\mathcal{E} \otimes \mathcal{K}, \mathcal{I})$, for $U \in \mathcal{L}^\infty(\mathcal{S})$ any version of $\Pi_2 U$ will clearly do as a conditional expectation of U given $\mathcal{E} \otimes \mathcal{K}$, for any $Q \in \mathcal{Q}$. Thus Π_2 may itself be regarded as constructed by sufficiency from $(\mathcal{Q}; \mathcal{E} \otimes \mathcal{K}, \mathcal{S})$, so long as $\mathcal{I} = \mathcal{I}(\mathcal{Q}^{\mathcal{E} \otimes \mathcal{K}})$, which is in fact so (Kudō, 1978).

In general, let $(\mathcal{E}, \mathcal{I})$ be a statistical space. If M is a finite measure on $\mathcal{E}$ such that $M(I) = 0$ for all $I \in \mathcal{I}$, we shall call M *absolutely continuous with respect to* $\mathcal{I}$, and write $M \ll \mathcal{I}$. We denote by $\mathfrak{M}(\mathcal{E}, \mathcal{I})$ the family $\{M : M \ll \mathcal{I}\}$, and by $\mathfrak{M}_1(\mathcal{E}, \mathcal{I})$ the subfamily of probability measures.

We always have $\mathcal{I} \subseteq \mathcal{I}(\mathfrak{M}(\mathcal{E}, \mathcal{I}))$. If $\mathcal{I} = \mathcal{I}(\mathfrak{M}(\mathcal{E}, \mathcal{I}))$, we shall call $\mathcal{I}$ *reflexive*. Any σ-ideal of the form $\mathcal{I}(\mathfrak{M})$ is reflexive. Conversely, suppose $\mathcal{I}$ is reflexive, and let $\Pi : \mathcal{L}^\infty(\mathcal{F}) \to L^\infty(\mathcal{E}, \mathcal{I})$ be a s.o. For $M \in \mathfrak{M}_1(\mathcal{E}, \mathcal{I})$, define Q_M on $\mathcal{F} \times \mathcal{E}$ by : $Q_M(F \times E) = \int_E U dM$, where $U \in \Pi(\chi_F)$. Although Q_M is finitely additive, it may fail to be σ-additive. We can show σ-additivity under further weak restrictions: see, for example, Caillot et Martin (1972, Proposition I-8). If, for all M, Q_M is σ-additive, then each Q_M extends to a distribution on $\mathcal{F} \otimes \mathcal{E}$, and Π may then be constructed by sufficiency in the same way as for Π_1 above.

4. Duality.

This section generalizes the familiar notions of (i) (bounded) completeness of a family of distributions, and (ii) identification of mixtures of distributions. These concepts are seen to arise naturally in the study of statistical operations, are closely related, and have important consequences.

Let $(\mathcal{E}, \mathcal{I})$ be a statistical space. Then the spaces $L^\infty(\mathcal{E}, \mathcal{I})$ and $\mathfrak{M}(\mathcal{E}, \mathcal{I})$ are in duality with respect to the bilinear product $\langle M, \tilde{U} \rangle = \int U dM$, where U is an arbitrary version of $\tilde{U}$. If $\Pi : L^\infty(\mathcal{F}, \mathcal{J}) \to L^\infty(\mathcal{E}, \mathcal{I})$ is a s.o., we can define its *transpose* $\Pi' : \mathfrak{M}(\mathcal{E}, \mathcal{I}) \to \mathfrak{M}(\mathcal{F}, \mathcal{J})$ by the requirement $\langle \Pi'M, \tilde{U} \rangle = \langle M, \Pi\tilde{U} \rangle$. Note that Π' maps $\mathfrak{M}_1(\mathcal{E}, \mathcal{I})$ into $\mathfrak{M}_1(\mathcal{F}, \mathcal{J})$, and is completely determined by its restriction to $\mathfrak{M}_1(\mathcal{E}, \mathcal{I})$.

If Π is constructed by sufficiency from $(\mathcal{P}; \mathcal{E}, \mathcal{F})$, then $\mathcal{I} = \mathcal{I}(\mathcal{P}^{\mathcal{E}})$, $\mathcal{J} = \mathcal{I}(\mathcal{P}^{\mathcal{F}})$, and, for any $P \in \mathcal{P}$, $\Pi'P^{\mathcal{E}} = P^{\mathcal{F}}$. If $\Pi : L^\infty(\mathcal{F}, \mathcal{J}) \to L^\infty(\mathcal{E}, \mathcal{I})$ is the natural injection, then $\Pi' : \mathfrak{M}(\mathcal{E}, \mathcal{I}) \to \mathfrak{M}(\mathcal{F}, \mathcal{J})$ is given by restriction: $\Pi'M = M^{\mathcal{F}}$.

In general the maps $\Pi : L^\infty(\mathcal{F}, \mathcal{J}) \to L^\infty(\mathcal{E}, \mathcal{I})$ and $\Pi' : \mathfrak{M}(\mathcal{E}, \mathcal{I}) \to \mathfrak{M}(\mathcal{F}, \mathcal{J})$ are many-to-one. We obtain important properties by requiring that they be one-to-one.

DEFINITION 4.1. If the s.o. $\Pi : L^\infty(\mathcal{F}, \mathcal{I}_\Pi) \to L^\infty(\mathcal{E}, \mathcal{I})$ is one-to-one, we call Π *boundedly complete*. If Π is constructed by sufficiency from $(\mathcal{P}; \mathcal{E}, \mathcal{F})$, we also say that $\mathcal{F}$ is *boundedly complete for* $\mathcal{E}$ (*with respect to* $\mathcal{P}$).

Clearly Π is boundedly complete if $\Pi U = 0 \Rightarrow U = 0$; or, regarding Π as a many-valued map from $\mathcal{L}^\infty(\mathcal{F})$ to $\mathcal{L}^\infty(\mathcal{E})$, Π is boundedly complete if $\Pi U \sim_{\mathcal{I}} 0 \Rightarrow U \sim_{\mathcal{I}_\Pi} 0$. In the context of Example I(a), this recovers the definition of Lehmann and Scheffé (1950, 1955).

The above concept can be strengthened if Π is constructed by sufficiency from $(\mathscr{P}; \mathscr{E}, \mathscr{F})$. In this case, denote by $\mathcal{L}^1(\mathscr{E}, \mathscr{P})$ the set of $\mathscr{E}$-measurable variables that are P-integrable for all $P \in \mathscr{P}$, and by $L^1(\mathscr{E}, \mathscr{P})$ the set of equivalence classes under $\sim_{\mathscr{P}\mathscr{E}}$. Then the sufficiency construction extends Π to a map from $L^1(\mathscr{F}, \mathscr{P})$ into $L^1(\mathscr{E}, \mathscr{P})$. If this extended map is one-to-one we shall say that $\mathscr{F}$ is *complete* for $\mathscr{E}$ (with respect to $\mathscr{P}$). Equivalently we require that, for $U \in \mathcal{L}^1(\mathscr{F}, \mathscr{P})$, $E_{\mathscr{P}}(U|\mathscr{E}) = 0[\mathscr{P}^{\mathscr{E}}]$ if and only if $U = 0[\mathscr{P}^{\mathscr{F}}]$.

For the transpose map we introduce the following definition.

DEFINITION 4.2. If the map $\Pi' : \mathfrak{M}(\mathscr{E}, \mathscr{I}) \to \mathfrak{M}(\mathscr{F}, \mathscr{J})$ is one-to-one, we say Π is *strongly identifying*. If Π is constructed by sufficiency from $(\mathscr{P}; \mathscr{E}, \mathscr{F})$, we say $\mathscr{F}$ *strongly identifies* $\mathscr{E}$ (*with respect to* $\mathscr{P}$).

This property is normally studied under the title "identifiability of mixtures" (Teicher, 1960, 1961, 1967; Barndorff-Nielsen, 1973; Chandra, 1977). It is only necessary to check that the restricted map $\Pi: \mathfrak{M}_1(\mathscr{E}, \mathscr{I}) \to \mathfrak{M}_1(\mathscr{F}, \mathscr{J})$ is one-to-one. Note that if Π_1 in Example I(a) is strongly identifying, then *a fortiori* any parameter-function which induces $\mathfrak{K}$ is identified in the usual sense.

4.1. *Two theorems.* The following two results, which will be needed for Theorem 7.3 below, are of some interest in their own right.

First, consider a single probability distribution P on a σ-field $\mathscr{S}$, and let $\mathscr{E}, \mathscr{F}$ be sub-σ-fields of $\mathscr{S}$. We can consider the statistical operations which arise by taking conditional expectations over $\mathscr{F}$ given $\mathscr{E}$, and over $\mathscr{E}$ given $\mathscr{F}$.

THEOREM 4.1. $\mathscr{F}$ *is complete for* $\mathscr{E}$ *with respect to* P *if and only if* $\mathscr{E}$ *strongly identifies* $\mathscr{F}$ *with respect to* P.

PROOF. Let the induced statistical operations be $\Pi_1 : L^1(\mathscr{F}, P) \to L^1(\mathscr{E}, P)$, and $\Pi_2 : L^1(\mathscr{E}, P) \to L^1(\mathscr{F}, P)$. Suppose first that Π_1 is complete, and let P_1, P_2 be measures on $\mathscr{F}$ such that $P_i \ll P^{\mathscr{F}}$. We must show that $\Pi_2'P_1 = \Pi_2'P_2 \Rightarrow P_1 = P_2$. Now it is easy to see that $\Pi_2'P_i = P_i'$, say, is given by $P_i'(A) = \int_A U_i dP = \int_A E(U_i|\mathscr{E})dP^{\mathscr{E}}(A \in \mathscr{E})$, where $U_i \in dP_i/dP^{\mathscr{F}} \in L^1(\mathscr{F}, P)$. If now $P_1' = P_2'$, then $E(U_1 - U_2|\mathscr{E}) = 0[P^{\mathscr{E}}]$ whence, by completeness of Π_1, $U_1 = U_2[P^{\mathscr{F}}]$, and so $P_1 = P_2$.

Conversely, suppose Π_2 is strongly identifying, and let $U \in \mathcal{L}^1(\mathscr{F}, P)$. We have to show that $E(U|\mathscr{E}) = 0[P^{\mathscr{E}}] \Rightarrow U = 0[P^{\mathscr{F}}]$. Set $U_1 = \max\{U, 0\}$, $U_2 = \max\{-U, 0\}$, so that $U = U_1 - U_2$ and $U_i \in \mathcal{L}^1_+(\mathscr{F}, P)$. We can define finite measures M_1, M_2 over $\mathscr{S}$ by: $M_i(S) = \int_S U_i dP$, and it may be seen that $\Pi_2'M_i^{\mathscr{F}} = M_i^{\mathscr{E}}$. Hence if $E(U|\mathscr{E}) = 0[P^{\mathscr{E}}]$ then $M_1^{\mathscr{E}} = M_2^{\mathscr{E}}$ and so, by the strong identification $M_1^{\mathscr{F}} = M_2^{\mathscr{F}}$, so that $U = 0[P^{\mathscr{F}}]$.

For the next theorem we recall (Halmos and Savage, 1949) that a σ-field $\mathscr{D}$ is termed *pairwise sufficient* for $\mathscr{P}$ over $\mathscr{E}$ if, for all $P_1, P_2 \in \mathscr{P}$, $\mathscr{D}$ is sufficient for the pair $\{P_1, P_2\}$ over $\mathscr{E}$. In this case $\mathscr{D}$ is sufficient over $\mathscr{E}$ for any subfamily of $\mathscr{P}$ which is dominated on $\mathscr{D} \vee \mathscr{E}$. Furthermore, under mild regularity conditions,

pairwise sufficiency implies sufficiency more generally (see, e.g., Kusama and Yamada, 1972).

THEOREM 4.2. *Let $\mathcal{D}$, $\mathcal{E}$, $\mathcal{F}$ be sub-σ-fields of $\mathcal{S}$, with $\mathcal{D} \subseteq \mathcal{E}$. Suppose $\mathcal{D}$ and $\mathcal{E}$ are each sufficient for $\mathcal{P}$ over $\mathcal{F}$, and that $\mathcal{F}$ strongly identifies $\mathcal{E}$ (with respect to $\mathcal{P}$). Then $\mathcal{D}$ is pairwise sufficient for $\mathcal{P}$ over $\mathcal{E}$.*

PROOF. Take $P_1, P_2 \in \mathcal{P}$, and let $W_1 \in \mathcal{L}^{\infty}_{+}(\mathcal{D})$ be a version of $dP_1^{\mathcal{D}}/d(P_1 + P_2)^{\mathcal{D}}$. Then $W_2 = 1 - W_1$ is a version of $dP_2^{\mathcal{D}}/d(P_1 + P_2)^{\mathcal{D}}$. Define events A_i : "$W_i = 0$" ($i = 1, 2$) and A_3 : "$0 < W_1 < 1$". Then A_1, A_2, A_3 form a $\mathcal{D}$-measurable partition.

Take $D \in \mathcal{D}$ with $D \subseteq A_3$, and define a finite measure Δ over $\mathcal{D}$ by: $\Delta(S) = P_1(S \cap D) + P_2(S \cap D)$. Then $\Delta \ll P_i^{\mathcal{D}}$ ($i = 1, 2$), and so we can unambiguously define measures M_i ($i = 1, 2$) over $\mathcal{S}$ by: $M_i(S) = \int P_i(S|\mathcal{D})d\Delta$. By sufficiency of $\mathcal{D}$ over $\mathcal{F}$, $M_1^{\mathcal{F}} = M_2^{\mathcal{F}}$. Also, note that for any $U \in \mathcal{L}^{\infty}(\mathcal{S})$, $\int UdM_i = \int E_i(U|\mathcal{D})d\Delta$, where E_i denotes (conditional) expectation for P_i. Thus $M_i(S) = \int \chi_S dM_i = \int E_i(\chi_S|\mathcal{D})d\Delta = \int E_i[E_i(\chi_S|\mathcal{E})|\mathcal{D}]d\Delta$ (since $\mathcal{D} \subseteq \mathcal{E}$) $= \int E_i(\chi_S|\mathcal{E})dM_i$. In particular, for $F \in \mathcal{F}$, $M_i(F) = \int E_{\mathcal{P}}(\chi_F|\mathcal{E})dM_i^{\mathcal{E}}$ whence, since $M_1^{\mathcal{F}} = M_2^{\mathcal{F}}$, by strong identification $M_1^{\mathcal{E}} = M_2^{\mathcal{E}}$. We thus have, for $U \in \mathcal{L}^{\infty}(\mathcal{E})$, $\int UdM_i$ does not depend on i. But $\int UdM_i = \int E_i(U|\mathcal{D})d\Delta = \int_D E_i(U|\mathcal{D}) \cdot d(P_1 + P_2)$. Now letting D vary, we deduce that $E_1(U|\mathcal{D}) = E_2(U|\mathcal{D})[P_1 + P_2]$ on A_3.

Take particular versions of $E_i(U|\mathcal{D})$ ($i = 1, 2$), and define $V \in \mathcal{L}^{\infty}(\mathcal{D})$ by : $V = E_1(U|\mathcal{D})$ on $A_1 \cup A_3$, $V = E_2(U|\mathcal{D})$ on A_2. Then V serves as a version of the conditional expectation of U given $\mathcal{D}$, for both P_1 and P_2, and so the result is proved.

5. Conditional independence.

5.1. *Definitions and examples.* Let P be a probability distribution over a σ-field $\mathcal{S}$, and $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ sub-σ-fields of $\mathcal{S}$. The familiar concept of *conditional independence of $\mathcal{A}$ and $\mathcal{B}$ given $\mathcal{C}$* (with respect to P) can be characterized as follows: for any $U \in \mathcal{L}^{1}(\mathcal{A}, P)$, there exists a $\mathcal{C}$-measurable version of $E_P(U|\mathcal{B} \vee \mathcal{C})$ (see for example Meyer, 1966, II, T51).

This motivates the following general definition. Let Π be a statistical operation over $\mathcal{F}$ given $(\mathcal{S}, \mathcal{I})$, and suppose $\mathcal{A}$ is a sub-σ-field of $\mathcal{F}$, and $\mathcal{B}$, $\mathcal{C}$ sub-σ-fields of $\mathcal{S}$ satisfying $\mathcal{B} \vee \mathcal{C} = \mathcal{S}$.

DEFINITION 5.1. We say *$\mathcal{A}$ is independent of $\mathcal{B}$ given $\mathcal{C}$* (*with respect to Π*), and write $\mathcal{A} \perp\!\!\!\perp \mathcal{B} |\mathcal{C}[\Pi]$, if, for all $U \in \mathcal{L}^{\infty}(\mathcal{A})$, there exists a version of ΠU which is $\mathcal{C}$-measurable. If $\mathcal{C}$ is trivial, we may say *$\mathcal{A}$ is independent of $\mathcal{B}$* (*with respect to Π*), and write $\mathcal{A} \perp\!\!\!\perp \mathcal{B} [\Pi]$.

In checking that $\mathcal{A} \perp\!\!\!\perp \mathcal{B} |\mathcal{C}[\Pi]$, it is enough to verify the definition for all indicator functions of sets in $\mathcal{A}$. Note that $\mathcal{B}$ plays no essential rôle in Definition 5.1; nevertheless it is helpful to have it in the notation.

LEMMA 5.1. $\mathcal{A} \perp \mathcal{B} \,|\, \mathcal{C}[\Pi]$ *if and only if, for* $U \in L^\infty(\mathcal{A})$, $\Pi U \in \mathcal{L}^\infty(\tilde{\mathcal{C}}, \mathcal{I})$, *where* $\tilde{\mathcal{C}} = \mathcal{C} \vee \mathcal{I}$.

PROOF. From Lemma 2.2.

If $\mathcal{A} \perp \mathcal{B} \,|\, \mathcal{C}[\Pi]$, we can regard Π as yielding a s.o. over $\mathcal{A}$ given $(\tilde{\mathcal{C}}, \mathcal{I})$ (or alternatively given $(\mathcal{C}, \mathcal{I} \cap \mathcal{C})$). Such an induced s.o. will be denoted by Π^*. Then $\Pi : \mathcal{L}^\infty(\mathcal{A}) \to L^\infty(\mathcal{S}, \mathcal{I})$ is given by $\Pi = \Pi_0 \Pi^*$, where Π_0 is a natural injection. Conversely, the existence of such Π^*, satisfying $\Pi = \Pi_0 \Pi^*$, implies $\mathcal{A} \perp \mathcal{B} \,|\, \mathcal{C}[\Pi]$.

Now suppose $\mathcal{P}$ is a family of distributions over $\mathcal{S}$, and let $\mathcal{A}, \mathcal{B}, \mathcal{C}$ be sub-σ-fields of $\mathcal{S}$. We shall be interested in conditional independence for s.o.'s constructed by sufficiency.

DEFINITION 5.2. We say $\mathcal{A}$ *is independent of* $\mathcal{B}$ *given* $\mathcal{C}$ (*with respect to* $\mathcal{P}$), and write $\mathcal{A} \perp \mathcal{B} \,|\, \mathcal{C}[\mathcal{P}]$ if
   (i) $\mathcal{B} \vee \mathcal{C}$ is sufficient for $\mathcal{P}$ over $\mathcal{A}$; and
   (ii) $\mathcal{A} \perp \mathcal{B} \,|\, \mathcal{C}[\Pi]$, where Π is the s.o. over $\mathcal{A}$ given $(\mathcal{B} \vee \mathcal{C}, \mathcal{I})$ constructed by sufficiency from $(\mathcal{P}; \mathcal{B} \vee \mathcal{C}, \mathcal{A})$.

Equivalent to Definition 5.2 is the following: for $U \in \mathcal{L}^\infty(\mathcal{A})$, there exists $V \in \mathcal{L}^\infty(\mathcal{C})$ such that $E_P(U|\mathcal{B} \vee \mathcal{C}) = V[P]$, all $P \in \mathcal{P}$. It follows that $\mathcal{C}$ is sufficient for $\mathcal{P}$ over $\mathcal{A}$, and $E_\mathcal{P}(U|\mathcal{C}) = E_\mathcal{P}(U|\mathcal{B} \vee \mathcal{C})[\mathcal{P}]$.

At this point we may note that Theorem 4.2 continues to hold if the condition $\mathcal{D} \subseteq \mathcal{E}$ is replaced by $\mathcal{F} \perp \mathcal{D} \,|\, \mathcal{E}[\mathcal{P}]$. The proof is virtually unchanged.

LEMMA 5.2. *The following properties hold.*
   (i) $\mathcal{A} \perp \mathcal{B} \,|\, \mathcal{C}[\Pi] \Leftrightarrow \mathcal{A} \perp (\mathcal{B} \vee \mathcal{C}) \,|\, \mathcal{C}[\Pi]$.
   (ii) $\mathcal{A} \perp \mathcal{B} \,|\, \mathcal{C}[\Pi], \mathcal{D} \subseteq \mathcal{A} \Rightarrow \mathcal{D} \perp \mathcal{B} \,|\, \mathcal{C}[\Pi]$.
   (iii) $\mathcal{A} \perp \mathcal{B} \,|\, \mathcal{C}[\mathcal{P}] \Leftrightarrow (\mathcal{A} \vee \mathcal{C}) \perp \mathcal{B} \,|\, \mathcal{C}[\mathcal{P}]$.
   (iv) $\mathcal{A} \perp \mathcal{B} \,|\, \mathcal{C}[P] \Leftrightarrow \mathcal{B} \perp \mathcal{A} \,|\, \mathcal{C}[P]$.
   (v) $\mathcal{A} \perp \mathcal{B} \,|\, \mathcal{C}[\mathcal{P}], \mathcal{P}' \subseteq \mathcal{P} \Rightarrow \mathcal{A} \perp \mathcal{B} \,|\, \mathcal{C}[\mathcal{P}']$.

PROOF. For (iv), see, e.g., Meyer, loc. cit. The rest are trivial.

We now illustrate with our standard examples the scope of the general concept of conditional independence.

EXAMPLE I(b). *Sufficient parameters.* Suppose that, in the labelled family of distributions $\{P_\lambda : \lambda \in \Lambda\}$, the parameter λ is *not identified*: that is, there exists a function, $\psi : \Lambda \to \Psi$ say, such that $\psi(\lambda_1) = \psi(\lambda_2) \Rightarrow P_{\lambda_1} = P_{\lambda_2}$. Let $\mathcal{K}$ be the sub-σ-field of $\mathcal{H}$ generated by ψ: $\mathcal{K} = \{K \in \mathcal{H} : K = \psi^{-1}(G), \text{ some } G \subseteq \Psi\}$. Then, for any $A \in \mathcal{S}$, the map $\lambda \rightsquigarrow P_\lambda(A)$ is $\mathcal{K}$-measurable, and so the dependence of the distributions on ψ alone may be expressed as: $\mathcal{S} \perp \mathcal{H} \,|\, \mathcal{K}[\Pi_1]$. In this case the induced map $\Pi_1^* : \mathcal{L}^\infty(\mathcal{S}) \to \mathcal{L}^\infty(\mathcal{K})$ represents the same family of distributions, but reparametrized by ψ.

If $\mathcal{K}$ is trivial, $\mathcal{S} \perp \mathcal{H}[\Pi_1]$ is equivalent to: P_λ does not depend on λ: that is, $\mathcal{S}$ is an *ancillary* σ-field.

When $\mathbb{S} \perp\!\!\!\perp \mathcal{K} | \mathcal{K} [\Pi_1]$, we may call $\mathcal{K}$ a *sufficient parametric σ-field*, and, if $\mathcal{K}$ is generated by ψ, we may term ψ a *sufficient parameter* (Barankin, 1961).

EXAMPLE II(b). *Sufficiency.* Consider the property: $\mathbb{S} \perp\!\!\!\perp \mathcal{E}^* | \mathcal{E} [\Pi_2]$, where $\mathcal{E}$ is regarded as a sub-σ-field of $\mathcal{E}^*$ in the natural way. This says that, given $U \in \mathcal{L}^\infty(\mathbb{S})$, we can find $V \in \mathcal{L}^\infty(\mathcal{E})$ such that, for all $\lambda \in \Lambda$, $E_\lambda(U|\mathcal{E}) = V[P_\lambda]$, and this is just the requirement that $\mathcal{E}$ be sufficient for $\mathcal{P}$ over $\mathbb{S}$. We have therefore rigorized the intuitive identity between sufficiency and conditional independence discussed in Section 1. The induced s.o. Π_2^* is essentially the same as Π_3 of Example III(a).

If $\mathcal{E}$ is regular, the above property is equivalent to: $\mathbb{S} \perp\!\!\!\perp \mathcal{K} | \mathcal{E} [\Pi_2]$ (regarding $\mathcal{E}$, $\mathcal{K}$ as sub-σ-fields of $\mathcal{E} \otimes \mathcal{K}$): i.e., the observation is independent of the parameter, given a sufficient statistic (cf. $X \perp\!\!\!\perp \Theta | T$, Section 1). Clearly, if $\mathcal{E}$ is sufficient then $\mathcal{E}$ is regular.

EXAMPLE II(c). *Pointwise independence.* Now let $\mathcal{F} \subseteq \mathbb{S}$ and restrict Π_2 to act on $\mathcal{L}^\infty(\mathcal{F})$. Consider: $\mathcal{F} \perp\!\!\!\perp \mathcal{E}^* | \mathcal{K} [\Pi_2]$. For $U \in \mathcal{L}^\infty(\mathcal{F})$, we get $W \in \mathcal{L}^\infty(\mathcal{K})$ satisfying: for all λ, $E_\lambda(U|\mathcal{E}) = W[P_\lambda]$. Since W is nonrandom, this merely asserts the probabilistic independence of $\mathcal{F}$ and $\mathcal{E}$ for every P_λ, more intuitively expressed as $\mathcal{F} \perp\!\!\!\perp \mathcal{E} | \mathcal{K} [\Pi_2]$ (or, in random variable terms, $X \perp\!\!\!\perp Y | \Theta$).

5.2. *Adequacy.* Let $\mathcal{P}$ be a family of distributions over $\mathbb{S}$, and $\mathcal{Q}, \mathcal{B}, \mathcal{C}$ sub-σ-fields of $\mathbb{S}$. Following Skibinsky (1967) (who, however, took $\mathcal{C} \subseteq \mathcal{Q}$) we call $\mathcal{C}$ *adequate for* $\mathcal{Q}$ *with respect to* $\mathcal{B}$ *and* $\mathcal{P}$ if

(i) $\mathcal{C}$ is sufficient for $\mathcal{P}$ over $\mathcal{Q}$; and
(ii) $\mathcal{B} \perp\!\!\!\perp \mathcal{Q} | \mathcal{C} [P]$, for all $P \in \mathcal{P}$.

THEOREM 5.1. $\mathcal{C}$ *is adequate for* $\mathcal{Q}$ *with respect to* $\mathcal{B}$ *and* $\mathcal{P}$ *if and only if* $\mathcal{Q} \perp\!\!\!\perp \mathcal{B} | \mathcal{C} [\mathcal{P}]$.

PROOF. (Skibinsky, 1967, Theorem 1). That conditional independence implies adequacy is trivial. So suppose the adequacy condition holds, and let $U \in \mathcal{L}^\infty(\mathcal{Q})$, $P \in \mathcal{P}$. By (ii) and the symmetry property (Lemma 5.2 (iv)) $E_P(U|\mathcal{B} \vee \mathcal{C}) = E_P(U|\mathcal{C})[P] = E_{\mathcal{P}}(U|\mathcal{C})[P]$ by (i). Thus $E_{\mathcal{P}}(U|\mathcal{C})$ serves as a version of $E_P(U|\mathcal{B} \vee \mathcal{C})$ for all $P \in \mathcal{P}$, whence $\mathcal{Q} \perp\!\!\!\perp \mathcal{B} | \mathcal{C} [\mathcal{P}]$.

COROLLARY. *If* $\mathcal{Q} \perp\!\!\!\perp \mathcal{B} | \mathcal{C} [\mathcal{P}]$, *and* $\mathcal{C}$ *is sufficient for* $\mathcal{P}$ *over* $\mathcal{B}$, *then* $\mathcal{B} \perp\!\!\!\perp \mathcal{Q} | \mathcal{C} [\mathcal{P}]$.

Theorem 5.1 is important in three different ways. First, it exhibits adequacy as a special case of conditional independence. Secondly, since many instances of conditional independence may be expressed in the form of Definition 5.2 by using an auxiliary construction by sufficiency, conditional independence may be studied by means of known results on adequacy. Finally, the two-stage definition of adequacy proves useful in verifying conditional independence in two separate steps.

6. Extension of the future. The title of this section refers to the following simple well-known result for ordinary conditional independence in a probability space $(\mathcal{S}, P)$: $\mathcal{A} \perp (\mathcal{B} \vee \mathcal{C})|\mathcal{D} \Leftrightarrow (\mathcal{A} \perp \mathcal{C}|\mathcal{D}$ and $\mathcal{A} \perp \mathcal{B}|(\mathcal{C} \vee \mathcal{D}))$. Here we investigate generalizations.

6.1. *Direct implication.*

THEOREM 6.1. $\mathcal{A} \perp (\mathcal{B} \vee \mathcal{C})|\mathcal{D}[\Pi] \Rightarrow \mathcal{A} \perp \mathcal{B}|(\mathcal{C} \vee \mathcal{D})[\Pi]$ *and* $\mathcal{A} \perp \mathcal{C}|\mathcal{D}[\Pi^*]$.

The proof is trivial. Note that $\Pi^* : \mathcal{L}^\infty(\mathcal{A}) \to L^\infty(\mathcal{C} \vee \mathcal{D}, \mathcal{I}^*)$, say, is the s.o. induced by Π through the property $\mathcal{A} \perp \mathcal{B}|(\mathcal{C} \vee \mathcal{D})[\Pi]$. Theorem 6.1 holds in particular if $[\Pi]$ and $[\Pi^*]$ are both replaced by $[\mathcal{P}]$ or (recovering the standard result) $[P]$. The case that $\mathcal{D}$ is trivial is of some special importance.

In the general case, the symmetry of simple conditional independence is absent. We must therefore investigate separately the implications of $(\mathcal{B} \vee \mathcal{C}) \perp \mathcal{A}|\mathcal{D}[\Pi]$. The first result is trivial.

THEOREM 6.2. $(\mathcal{B} \vee \mathcal{C}) \perp \mathcal{A}|\mathcal{D}[\Pi] \Rightarrow \mathcal{C} \perp \mathcal{A}|\mathcal{D}[\Pi]$.

THEOREM 6.3.   *Suppose* $(\mathcal{B} \vee \mathcal{C}) \perp \mathcal{A}|\mathcal{D}[\mathcal{P}]$, *and further that* $\mathcal{C} \vee \mathcal{D}$ *is sufficient for* $\mathcal{P}$ *over* $\mathcal{B}$. *Then* $\mathcal{B} \perp \mathcal{A}|(\mathcal{C} \vee \mathcal{D})[\mathcal{P}]$.

REMARK 6.1.   The sufficiency condition is a weak one. For if, as is supposed, $\mathcal{D}$ is sufficient for $\mathcal{P}$ over $\mathcal{B} \vee \mathcal{C}$, then $\mathcal{D}$ is sufficient over $\mathcal{B} \vee \mathcal{C} \vee \mathcal{D}$. Since $\mathcal{D} \subseteq \mathcal{C} \vee \mathcal{D} \subseteq \mathcal{B} \vee \mathcal{C} \vee \mathcal{D}$, it follows that $\mathcal{C} \vee \mathcal{D}$ is at least pairwise sufficient over $\mathcal{B} \vee \mathcal{C} \vee \mathcal{D}$, and thus over $\mathcal{B}$. However, sufficiency may fail to hold in pathological circumstances (Burkholder, 1961).

PROOF OF THEOREM 6.3.   By Theorem 5.1, we only need to show that $\mathcal{B} \perp \mathcal{A}|(\mathcal{C} \vee \mathcal{D})[P]$, for each $P \in \mathcal{P}$. But $(\mathcal{B} \vee \mathcal{C}) \perp \mathcal{A}|\mathcal{D}[P]$, and so the standard result applies.

COROLLARY. $(\mathcal{B} \vee \mathcal{C}) \perp \mathcal{A}[\mathcal{P}] \Rightarrow \mathcal{B} \perp \mathcal{A}|\mathcal{C}[\mathcal{P}]$.

PROOF.   Taking $\mathcal{D}$ trivial, we have that $P^{\mathcal{B} \vee \mathcal{C}}$ does not depend on $P \in \mathcal{P}$, and so the sufficiency property holds trivially.

6.2. *Converse implication.*

THEOREM 6.4.   *If* $\mathcal{A} \perp \mathcal{B}|(\mathcal{C} \vee \mathcal{D})[\Pi]$ *and* $\mathcal{A} \perp \mathcal{C}|\mathcal{D}[\Pi^*]$, *then* $\mathcal{A} \perp (\mathcal{B} \vee \mathcal{C})|\mathcal{D}[\Pi]$.

The proof is straightforward. Π^* is again interpreted as in Theorem 6.1.

At this point a reinterpretation of Theorem 5.1 is instructive, along the lines of Example II. Introducing the parametric σ-field $\mathcal{K}$, and taking, for simplicity, all σ-fields to be regular, the two conditions for adequacy may be reexpressed as: (i) $\mathcal{A} \perp \mathcal{K}|\mathcal{C}$, and (ii) $\mathcal{A} \perp \mathcal{B}|(\mathcal{C} \vee \mathcal{K})$ (where the omitted s.o.'s are obvious). Theorems 6.1 and 6.4 show these to be equivalent to $\mathcal{A} \perp (\mathcal{B} \vee \mathcal{K})|\mathcal{C}$, a reexpression of $\mathcal{A} \perp \mathcal{B}|\mathcal{C}[\mathcal{P}]$.

THEOREM 6.5. *If* $\mathcal{B} \perp\!\!\!\perp \mathcal{A}|(\mathcal{C} \vee \mathcal{D})[\mathcal{P}]$ *and* $\mathcal{C} \perp\!\!\!\perp \mathcal{A}|\mathcal{D}[\mathcal{P}]$, *then* $(\mathcal{B} \vee \mathcal{C}) \perp\!\!\!\perp \mathcal{A}|\mathcal{D}[\mathcal{P}]$.

PROOF. The result holds for a single distribution. Thus by Theorem 5.1 we need only show that $\mathcal{D}$ is sufficient for $\mathcal{P}$ over $\mathcal{B} \vee \mathcal{C}$. But $\mathcal{D}$ is sufficient over $\mathcal{C}$, and thus over $\mathcal{C} \vee \mathcal{D}$; and likewise $\mathcal{C} \vee \mathcal{D}$ is sufficient over $\mathcal{B} \vee \mathcal{C} \vee \mathcal{D}$. The result follows (Bahadur, 1954, Corollary 5.1).

7. Variations on a theme. We already know that $(*)\mathcal{A} \perp\!\!\!\perp (\mathcal{B} \vee \mathcal{C})|\mathcal{D} \Rightarrow$ (a) $\mathcal{A} \perp\!\!\!\perp \mathcal{B}|(\mathcal{C} \vee \mathcal{D})$ and (b) $\mathcal{A} \perp\!\!\!\perp \mathcal{C}|\mathcal{D}$; similarly (*) implies (c) $\mathcal{A} \perp\!\!\!\perp \mathcal{C}|(\mathcal{B} \vee \mathcal{D})$ and (d) $\mathcal{A} \perp\!\!\!\perp \mathcal{B}|\mathcal{D}$. In the converse direction, we have (a) and (b) $\Rightarrow$ (*), (c) and (d) $\Rightarrow$ (*). We now turn to investigate the consequences of assuming (a) and (c), or (a) and (d), and in particular find further conditions under which these assumptions imply that (*) holds.

We first consider (a) and (c) together.

LEMMA 7.1. *Let* $\mathcal{B} \subseteq \mathcal{S}$, $\mathcal{C} \subseteq \mathcal{S}$. *Suppose* $\mathcal{A} \perp\!\!\!\perp \mathcal{S}|\mathcal{B}[\Pi]$ *and* $\mathcal{A} \perp\!\!\!\perp \mathcal{S}|\mathcal{C}[\Pi]$. *Then* $\mathcal{A} \perp\!\!\!\perp \mathcal{S}|(\tilde{\mathcal{B}} \cap \tilde{\mathcal{C}})[\Pi]$, *and conversely*.

The proof follows from Lemma 5.1.

COROLLARY. *An equivalent condition to the above is*: $\mathcal{A} \perp\!\!\!\perp \mathcal{S}|(\mathcal{B} \cap \tilde{\mathcal{C}})[\Pi]$ (*or* $\mathcal{A} \perp\!\!\!\perp \mathcal{S}|(\tilde{\mathcal{B}} \cap \mathcal{C})[\Pi]$).

PROOF. From Lemmas 2.1 (iii) and 5.1. Note that $\mathcal{B} \cap \tilde{\mathcal{C}} \subseteq \mathcal{B}$, $\tilde{\mathcal{B}} \cap \mathcal{C} \subseteq \mathcal{C}$. The above results immediately yield the following.
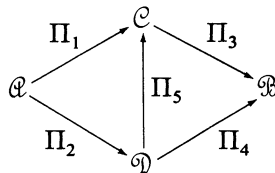
THEOREM 7.1. *Suppose* $\mathcal{A} \perp\!\!\!\perp \mathcal{B}|(\mathcal{C} \vee \mathcal{D})[\Pi]$ *and* $\mathcal{A} \perp\!\!\!\perp \mathcal{C}|(\mathcal{B} \vee \mathcal{D})[\Pi]$. *Then* $\mathcal{A} \perp\!\!\!\perp (\mathcal{B} \cap \mathcal{C})|(\widetilde{\mathcal{B} \vee \mathcal{D}}) \cap (\widetilde{\mathcal{C} \vee \mathcal{D}})[\Pi]$ *and conversely. Equivalently*, $\mathcal{A} \perp\!\!\!\perp (\mathcal{B} \vee \mathcal{C})|(\mathcal{B} \vee \mathcal{D}) \cap (\widetilde{\mathcal{C} \vee \mathcal{D}})[\Pi]$.

COROLLARY. *Suppose* $(\widetilde{\mathcal{B} \vee \mathcal{D}}) \cap (\widetilde{\mathcal{C} \vee \mathcal{D}}) \sim \mathcal{D}$. *Then* $\mathcal{A} \perp\!\!\!\perp \mathcal{B}|(\mathcal{C} \vee \mathcal{D})[\Pi]$ *and* $\mathcal{A} \perp\!\!\!\perp \mathcal{C}|(\mathcal{B} \vee \mathcal{D})[\Pi] \Leftrightarrow \mathcal{A} \perp\!\!\!\perp (\mathcal{B} \vee \mathcal{C})|\mathcal{D}$.

Note that $(\mathcal{B} \cap \mathcal{C}) \vee \mathcal{D} \subseteq (\widetilde{\mathcal{B} \vee \mathcal{D}}) \cap (\widetilde{\mathcal{C} \vee \mathcal{D}})$ so that, if the condition of the corollary is to hold, it is necessary (but not sufficient) that $\mathcal{B} \cap \mathcal{C} \tilde{\subseteq} \mathcal{D}$. The case $\mathcal{D}$ trivial is of some importance.

We now turn to consider the joint conditions (a) and (d).

LEMMA 7.2. *Consider the following diagram of statistical spaces and statistical operations between them, where for example* $\mathcal{A} \xrightarrow{\Pi_1} \mathcal{C}$ *denotes a s.o.* $\Pi_1 : L^\infty(\mathcal{A}, \mathcal{J}^{\mathcal{A}}) \to L^\infty(\mathcal{C}, \mathcal{J}^{\mathcal{C}})$. *We require* $\mathcal{J}^{\mathcal{C}} = \mathcal{J}^{\mathcal{B}}_{\Pi_3}$.

Suppose $\Pi_3\Pi_1 = \Pi_4\Pi_2$, $\Pi_4 = \Pi_3\Pi_5$, *and* Π_3 *is boundedly complete. Then* $\Pi_1 = \Pi_5\Pi_2$ (*and so the whole diagram is commutative*).

PROOF. $\Pi_3\Pi_1 = \Pi_4\Pi_2 = \Pi_3\Pi_5\Pi_2$. Thus $\Pi_3(\Pi_1 - \Pi_5\Pi_2) = 0$. Since bounded completeness is equivalent to Π_3 being one-to-one, the result follows.

THEOREM 7.2. *Suppose* $\mathcal{C} \perp\!\!\!\perp \mathcal{B} \mid \mathcal{D} [\mathcal{P}]$ *and* $\mathcal{C} \perp\!\!\!\perp \mathcal{B} \mid (\mathcal{C} \vee \mathcal{D})[\mathcal{P}]$. *Suppose further that* $\mathcal{B} \vee \mathcal{D}$ *is sufficient for* $\mathcal{P}$ *over* $\mathcal{C}$ (*and thus over* $\mathcal{C} \vee \mathcal{D}$) *and that* $\mathcal{C} \vee \mathcal{D}$ *is boundedly complete for* $\mathcal{B} \vee \mathcal{D}$ *with respect to* $\mathcal{P}$. *Then* $\mathcal{C} \perp\!\!\!\perp (\mathcal{B} \vee \mathcal{C})\mid \mathcal{D} [\mathcal{P}]$.

PROOF. In the diagram of Lemma 7.2, replace $\mathcal{B}$ by $\overline{\mathcal{B}} = \mathcal{B} \vee \mathcal{D}$, $\mathcal{C}$ by $\overline{\mathcal{C}} = \mathcal{C} \vee \mathcal{D}$, and let the σ-ideals be given by the $\mathcal{P}$-null sets in the corresponding σ-fields. Take Π_4, Π_5 to be the natural injections, and Π_1, Π_2 and Π_3 to be constructed by sufficiency. Also let $\Pi_6 : L^\infty(\mathcal{C}, \mathcal{P}) \to L^\infty(\mathcal{B}, \mathcal{P})$ be constructed by sufficiency. Then $\mathcal{C} \perp\!\!\!\perp \mathcal{B} \mid \mathcal{D} [\mathcal{P}] \Rightarrow \mathcal{C} \perp\!\!\!\perp \overline{\mathcal{B}} \mid \mathcal{D} [\mathcal{P}] \Rightarrow \Pi_6 = \Pi_4\Pi_2$, and $\mathcal{C} \perp\!\!\!\perp \mathcal{B} \mid (\mathcal{C} \vee \mathcal{D})[\mathcal{P}] \Rightarrow \mathcal{C} \perp\!\!\!\perp \overline{\mathcal{B}} \mid \overline{\mathcal{C}} [\mathcal{P}] \Rightarrow \Pi_6 = \Pi_3\Pi_1$. It is clear that $\Pi_4 = \Pi_3\Pi_5$. The conditions of the lemma therefore hold, and so $\Pi_1 = \Pi_5\Pi_2$, equivalent to $\mathcal{C} \perp\!\!\!\perp \overline{\mathcal{C}}\mid \mathcal{D} [\mathcal{P}]$, or $\mathcal{C} \perp\!\!\!\perp \mathcal{C}\mid \mathcal{D} [\mathcal{P}]$. On combining this with $\mathcal{C} \perp\!\!\!\perp \mathcal{B} \mid (\mathcal{C} \vee \mathcal{D})[\mathcal{P}]$, using Theorem 6.4, the result follows.

COROLLARY 1. *Suppose* $\mathcal{C} \perp\!\!\!\perp \mathcal{B} \mid \mathcal{D} [\mathcal{P}]$ *and* $\mathcal{C} \perp\!\!\!\perp \mathcal{B} \mid (\mathcal{C} \vee \mathcal{D})[\mathcal{P}]$, *and, for each* $P \in \mathcal{P}$, $\mathcal{C} \vee \mathcal{D}$ *is boundedly complete for* $\mathcal{B} \vee \mathcal{D}$ *with respect to* P. *Then* $\mathcal{C} \perp\!\!\!\perp (\mathcal{B} \vee \mathcal{C})\mid \mathcal{D} [\mathcal{P}]$.

The proof follows from the theorem, applied for a single P, and the fact that $\mathcal{D}$ is sufficient for $\mathcal{P}$ over $\mathcal{C}$. In the case that $\mathcal{B} \vee \mathcal{D}$ *is* sufficient for $\mathcal{P}$ over $\mathcal{C}$, the condition of the theorem is, in general, weaker than that of Corollary 1.

COROLLARY 2. *Suppose* $\mathcal{B} \perp\!\!\!\perp \mathcal{C}\mid \mathcal{D} [\mathcal{P}]$ *and* $\mathcal{B} \perp\!\!\!\perp \mathcal{C}\mid (\mathcal{C} \vee \mathcal{D})[\mathcal{P}]$. *Suppose further that* $\mathcal{D}$ *is sufficient for* $\mathcal{P}$ *over* $\mathcal{C}$, *and, for each* $P \in \mathcal{P}$, $\mathcal{C} \vee \mathcal{D}$ *is boundedly complete for* $\mathcal{B} \vee \mathcal{D}$ *with respect to* P. *Then* $(\mathcal{B} \vee \mathcal{C})\perp\!\!\!\perp \mathcal{C}\mid \mathcal{D} [\mathcal{P}]$.

The proof is straightforward.

Lemma 7.2 and Theorem 7.2 may be considered as generalizing Basu's theorem (1955): if, in a statistical model with parameter Θ, T is a boundedly complete sufficient statistic and S an ancillary statistic, then S and T are independent (given Θ). This follows from the lemma, on letting $\mathcal{D}$ be trivial, $\mathcal{B}$ be the parametric σ-field and $\mathcal{C}$, $\mathcal{C}$ the σ-fields in the sample-space corresponding to S, T respectively. The proof parallels that of the theorem, on making natural choices for the Π's (in fact the theorem itself can be applied since $\mathcal{C}$ is regular, so that the Π's may be constructed by sufficiency).

Theorem 7.2 and its corollaries are somewhat difficult to understand if $\mathcal{D}$ is nontrivial. The bounded completeness condition may be roughly interpreted as follows. Imagine $\mathcal{D}$ to be generated by a random variable D, and denote by P_d the distribution P conditioned on $D = d$ (assumed possible), and let $\mathcal{P}_d = \{P_d\}$. Then we require that, for almost all d, $\mathcal{C}$ is boundedly complete for $\mathcal{B}$ with respect to $\mathcal{P}_d$ (or w.r.t. P_d, all $P \in \mathcal{P}$). Regularity conditions could make this precise, effectively

reducing the case of general $\mathcal{D}$ (by looking at conditional distributions) to the case $\mathcal{D}$ trivial. We shall not go into further details here.

The following theorem, again exploring the consequences of the joint conditions (a) and (d), does not appear to be related to known results.

THEOREM 7.3. *Suppose* $\mathcal{B} \perp\!\!\!\perp \mathcal{C} \,|\, \mathcal{D}\,[\mathcal{P}]$ *and* $\mathcal{B} \perp\!\!\!\perp \mathcal{C} \,|\, (\mathcal{C} \vee \mathcal{D})[\mathcal{P}]$. *Suppose further that* $\mathcal{B} \vee \mathcal{D}$ *strongly identifies* $\mathcal{C} \vee \mathcal{D}$ *with respect to* $\mathcal{P}$. *Then, so long as* $\mathcal{D}$ *is sufficient for* $\mathcal{P}$ *over* $\mathcal{C}$, $(\mathcal{B} \vee \mathcal{C}) \perp\!\!\!\perp \mathcal{C} \,|\, \mathcal{D}\,[\mathcal{P}]$.

REMARK 7.1. The sufficiency condition will normally hold. For by Theorem 4.2, taking $\mathcal{E} = \mathcal{C} \vee \mathcal{D}$, $\mathcal{F} = \mathcal{B} \vee \mathcal{D}$, $\mathcal{D}$ will certainly be pairwise sufficient over $\mathcal{C}$. In particular, sufficiency will hold in the case that $\mathcal{D}$ is trivial.

PROOF OF THEOREM 7.3. The strong identification clearly holds with respect to any $P \in \mathcal{P}$ and hence, by Theorem 4.1, $\mathcal{C} \vee \mathcal{D}$ is complete, and so boundedly complete, for $\mathcal{B} \vee \mathcal{D}$ with respect to P. The result now follows from Corollary 2 to Theorem 7.2.

Once again the strong identification condition can be roughly interpreted as requiring that $\mathcal{B}$ strongly identify $\mathcal{C}$ in the distributions conditional on $\mathcal{D}$.

8. Applications. In this section we illustrate some of the uses to which the calculus of conditional independence may be put. For further examples see Dawid (1976, 1979a, 1979b, 1979c), Dawid and Dickey (1977a, 1977b).

8.1 *A fallacious argument.* (Dawid 1979b). Consider a statistical model with parameter Θ. Let T be a sufficient statistic, and suppose that the statistic S is independent of T for all values of Θ.

Basu (1955) argued that S must be ancillary, as follows: the distribution of S given Θ is the same as that of S given T and Θ (since $S \perp\!\!\!\perp T | \Theta$), and this does not depend on Θ since T is sufficient. He later (Basu, 1958) pointed out that this appealing result was false, and provided a counterexample and correction. The problem has been further discussed by Koehn and Thomas (1975).

We can set up the above problem within the framework of Example II. Let $\mathcal{K}$ be the parametric σ-field, and $\mathcal{E}$ and $\mathcal{F}$ the σ-fields in the sample space generated by T and S. Define the statistical space $(\mathcal{E}^{*}, \mathcal{G}^{*})$ as for Example II(a), and let $\Pi_2 : \mathcal{L}^\infty(\mathcal{F}) \to L^\infty(\mathcal{E}^{*}, \mathcal{G}^{*})$ be the s.o. corresponding to "conditional distributions for S given (T, Θ)". Then, by sufficiency of T, $\mathcal{F} \perp\!\!\!\perp \mathcal{E}^{*} | \mathcal{E}\,[\Pi_2]$, and by the pointwise independence of S and T, $\mathcal{F} \perp\!\!\!\perp \mathcal{E}^{*} | \mathcal{K}[\Pi_2]$. Taking these conditions together we deduce, from Lemma 7.1, the equivalent conditions: (a) $\mathcal{F} \perp\!\!\!\perp \mathcal{E}^{*} | \tilde{\mathcal{E}} \cap \tilde{\mathcal{K}}[\Pi_2]$; (b) $\mathcal{F} \perp\!\!\!\perp \mathcal{E}^{*} | \bar{\mathcal{E}}[\Pi_2]$; or (c) $\mathcal{F} \perp\!\!\!\perp \mathcal{E}^{*} | \bar{\mathcal{K}}[\Pi_2]$; where $\bar{\mathcal{E}} = \mathcal{E} \cap \tilde{\mathcal{K}} \subseteq \mathcal{E}$, $\bar{\mathcal{K}} = \tilde{\mathcal{E}} \cap \mathcal{K} \subseteq \mathcal{K}$. (Note that (b) is equivalent to: $\bar{\mathcal{E}}$ is adequate for $\mathcal{E}$ with respect to $\mathcal{F}$.)

Now $E \in \bar{\mathcal{E}}$ if and only if $E \in \mathcal{E}$ and $E \triangle \hat{E} \in \mathcal{G}^{*}$ for some $\hat{E} \in \mathcal{K}$. This last requirement is equivalent to: $P_\theta(E) = 1$ for $\theta \in \hat{E}$, $P_\theta(E) = 0$ for $\theta \notin \hat{E}$. That is, E is a *strong zero-one set*, as defined by Breiman, et al. (1964), and $\bar{\mathcal{E}}$ is the σ-field

of such sets. Each $E \in \overline{\mathcal{E}}$ determines the corresponding $\hat{E} \in \mathcal{K}$ uniquely: $\hat{E} = \{ \theta : P_\theta(E) = 1 \}$; and it is easily seen that $\overline{\mathcal{K}} = \{ \hat{E} : E \in \overline{\mathcal{E}} \}$ (compare Skibinsky, 1969). Basu's original argument holds if and only if $\overline{\mathcal{E}}$ (and thus $\overline{\mathcal{K}}$) is trivial; that is to say, if there does not exist a *splitting set* (Koehn and Thomas, 1975) in $\mathcal{E}$: a set of $E \in \mathcal{E}$ for which $P_\theta(E) = 0$ or 1 for all θ, both values being taken.

In the general case, suppose $\overline{\mathcal{E}}$ is induced by a statistic W (a function of T), and $\overline{\mathcal{K}}$ by a parameter Φ (a function of Θ). Although W may not be constant overall, W will be constant with probability one for each value θ of Θ, and the value of this constant may be taken as $\Phi(\theta)$. We can interpret (b) above in terms of adequacy, as implying that the conditional distribution of S given W may be chosen independently of Θ. Likewise, (c) shows that the sampling distribution of S can only depend on the value of Φ. These twin properties are the appropriate generalization of ancillarity for rectifying Basu's argument.

As an example, let Θ take values in $(-1, 0) \cup (0, 1)$, and let the data be X_1, X_2, independently distributed, being uniform over the interval $(0, \Theta)$ for $\Theta > 0$, or $(\Theta, 0)$ for $\Theta < 0$. Let $T = \max(|X_1|, |X_2|) \cdot \text{sign}(X_1)$, $U = (X_1 - X_2)/T$. Then T is sufficient and U ancillary. Take $S = U + \text{sign}(X_1)$. Then S is independent of T for all θ, but is not ancillary. Here we can take $W = \text{sign}(X_1)$, $\Phi = \text{sign}(\Theta)$.

8.2. *Marginal ancillarity.*

Let $\mathcal{P}$ be a family of distributions for data X over $\mathcal{E}$, parametrized by Θ. Recall that a statistic or σ-field is said to be *ancillary* if its distribution is the same for all values of Θ. The importance of ancillary statistics in the theory of inference is that no information relevent to inference about Θ should be lost if one modifies the distributions in $\mathcal{P}$ by conditioning on the observed value of an ancillary statistic (Birnbaum, 1962).

Now suppose that $\mathcal{P} = \{ P_{\theta, \phi} \}$ is parametrized by a pair of parameters (Θ, Φ), where Θ is the parameter of interest and Φ a nuisance parameter. There is no universally acceptable definition of a statistic being "ancillary for Θ" (which would again justify conditioning for purposes of inference about Θ). In this section we examine two attempts at such a definition, one classical, the other Bayesian, and use conditional independence to investigate the relationship between the two. For further background, criticism, and extensions to sufficiency, see Dawid and Dickey (1977b), Dawid (1979c).

Let $\mathcal{K}$ be the full parametric σ-field, and $\mathcal{T}$, $\mathcal{F}$ the sub-σ-fields of $\mathcal{K}$ induced by Θ and Φ respectively. Let $\mathcal{S}$ be the sub-σ-field of $\mathcal{E}$ induced by the statistic S. The s.o. over $\mathcal{E}$ given $\mathcal{K}$ corresponding to the family of distributions over $\mathcal{E}$ (constructed as in Example I(a)) is denoted by Π_1.

DEFINITION 8.1.   (Basu, 1977). We call S *specific ancillary for* Θ if $S \perp\!\!\!\perp \Theta | \Phi$, that is: $\mathcal{S} \perp\!\!\!\perp \mathcal{T} | \mathcal{F}[\Pi_1]$.

If $\mathcal{F}$ is trivial, so that there are no nuisance parameters, this recovers the standard definition of ancillarity. In general, it requires that the distributions of S "do not depend on Θ" but only on the nuisance parameter Φ.

If conditioning on S is to provide a useful simplification of the problem of inference for Θ, it would be convenient if the following property were also to hold.

DEFINITION 8.2.   We call S Θ-*inducing* if $X \perp\!\!\!\perp \Phi|(S, \Theta)$, so that the conditional sampling distributions of X given S are governed by Θ alone.

When S is both specific ancillary for Θ and Θ-inducing, it is termed *S-ancillary* for Θ (Barndorff-Nielsen, 1978). This has been taken as a justification, from a sampling theory viewpoint, for basing inference for Θ on the simpler conditional sampling distributions of X given S.

A possible Bayesian approach is the following. If we had a prior distribution over $\mathfrak{X}$, we would get a single joint distribution over $\mathfrak{X} \otimes \mathfrak{S}$, and we might consider S to be ancillary for Θ if, marginally, $S \perp\!\!\!\perp \Theta$ (that is, $\mathbb{S} \perp\!\!\!\perp \mathfrak{T}$) in this joint distribution.

Now since the property $S \perp\!\!\!\perp \Theta$ may be expressed in terms of the distribution for S *given* Θ, it may be investigated as soon as the conditional prior distribution for Φ given Θ is determined.

We therefore introduce a family $\mathfrak{Q} = \{Q_\theta\}$ of distributions for Φ, labelled by Θ, to be interpreted as the prior conditional distributions for Φ given Θ. The marginal prior distribution for Θ need not be specified. Each Q_θ induces a joint distribution Q_θ^* for (Φ, X) given by $Q_\theta^*(\Phi \in F, X \in A) = \int_F P_{\theta, \phi}(X \in A) dQ_\theta(\phi)$. Thus $\mathfrak{Q}^* = \{Q_\theta^*\}$ gives the family of conditional distributions for (Φ, X) jointly, given Θ. Restricted to $\mathfrak{S}$, the family $\mathfrak{Q}^*$ gives the distributions for X given Θ, the nuisance parameter Φ having been marginalized out, and our Bayesian can confine himself to considering these distributions.

DEFINITION 8.3.   We call S *marginally ancillary for* Θ if S is ancillary for the family $\mathfrak{Q}^*$.

This definition depends, of course, on the choice of $\mathfrak{Q}$. If there are no nuisance parameters there is nothing to choose, and again the standard definition is recovered.

The property of marginal ancillarity is again most useful when S is also Θ-inducing, in which case S is termed *D-ancillary* for Θ (with respect to $\mathfrak{Q}$). In this case, a Bayesian whose conditional prior distributions for Φ given Θ are given by $\mathfrak{Q}$ can obtain his marginal posterior distribution for Θ by combining his prior marginal distribution with the likelihood generated from the simplified conditional sampling distributions of X given S (Dawid, 1979c).

Definitions 8.1 and 8.3 have different motivations, and it is interesting to investigate the extent to which they overlap. We shall proceed by supposing that S is simultaneously specific ancillary and marginally ancillary for Θ. The results below state, roughly, that if the prior dependence between Θ and Φ is sufficiently strong, this conjunction can occur only if S is ancillary for (Θ, Φ) in the usual sense; while if the dependence of the distributions of S on the parameter Φ is sufficiently strong, then it can only occur if Θ and Φ are *a priori* independent.

Suppose that Definition 8.1 holds. Then $P_{\theta,\phi}(S \in A) = P_\phi(S \in A)$, say, serves as a version of $Q_\theta^*(S \in A|\Phi = \phi)$ for all ϕ; that is, Φ is sufficient for $\mathfrak{Q}^*$ over S. If Definition 8.3 holds also, then S is ancillary for $\mathfrak{Q}^*$, and we can therefore apply Basu's theorem (Section 7) to yield the following result.

THEOREM 8.1. *Suppose S is both specific ancillary and marginally ancillary for Θ, and suppose further that, in the family $\mathfrak{Q}$ of conditional prior distributions, Φ is boundedly complete. Then $S \perp\!\!\!\perp \Phi[\mathfrak{Q}^*]$.*

The conclusion of Theorem 8.1 is that we can choose a common version for $Q_\theta^*(S \in A|\Phi = \phi)$, for all (θ, ϕ). But since $P_{\theta,\phi}(S \in A)$ is one version of this quantity, it follows that S is almost ancillary, in the sense that there exists $P(A)$, say, such that $P_{\theta,\phi}(S \in A) = P(A)$ so long as $\phi \notin N_\theta$, where $Q_\theta(\Phi \in N_\theta) = 0$, all θ. In particular, for any Bayesian whose conditional distributions for Φ given Θ are given by $\mathfrak{Q}$, $S \perp\!\!\!\perp (\Theta, \Phi)$, and he can treat S as a totally ancillary statistic.

The condition of bounded completeness in Theorem 8.1 imposes a strong relationship between Φ and Θ in the prior distribution: for example it cannot occur if $\Phi \perp\!\!\!\perp \Theta$. The conclusion implies that the dependence of S on the parameters is extremely weak. In the following result these properties are effectively interchanged.

THEOREM 8.2. *Suppose S is both specific ancillary and marginally ancillary for Θ, and suppose further that, in the sampling distributions, S strongly identifies Φ. Then it must hold that $\Phi \perp\!\!\!\perp \Theta(\mathcal{F}$ is ancillary with respect to $\mathfrak{Q})$ in the prior distribution.*

Note that, since $S \perp\!\!\!\perp \Theta|\Phi$, the sampling distributions of S do depend on Φ alone. The proof proceeds by introducing the family of all prior distributions over $\mathcal{F}$ and combining with $\mathfrak{Q}^*$ to yield a family $\mathfrak{Q}^\dagger$ over $\mathcal{K} \otimes \mathcal{E}$. We then find that $\mathcal{S} \perp\!\!\!\perp \mathcal{F}[\mathfrak{Q}^\dagger]$, $\mathcal{S} \perp\!\!\!\perp \mathcal{F}|\mathcal{F}[\mathfrak{Q}^\dagger]$, and $\mathcal{S}$ strongly identifies $\mathcal{F}$ with respect to $\mathfrak{Q}^\dagger$. The result follows from Theorem 7.3.

8.3. *Two-point mixtures.* Finally, as a very simple application of Theorem 7.3, we prove the following result. It is probably due originally to G. Udny Yule, although it is continually being rediscovered, particularly in relation to the collapsibility of contingency tables.

THEOREM 8.3. *Let (X, Y, Z) be random variables, with Z having just two possible values. Suppose X and Y are independent, both conditionally on Z and marginally. Then* either X or Y *must be independent of Z.*

The proof follows from the following lemma, on applying Theorem 7.2 with $\mathcal{Q} = \sigma(X)$, $\mathcal{B} = \sigma(Y)$, $\mathcal{C} = \sigma(Z)$, and $\mathcal{D}$ trivial, and $\mathcal{P} = \{P\}$, where P is given by the joint distribution of (X, Y, Z).

LEMMA 8.1. *With notation as above, if Z has a two-point distribution then exactly one of the following properties holds.*

(i) $\mathcal{B} \perp\!\!\!\perp \mathcal{C}[P]$.

(ii) $\mathcal{C}$ *is complete for* $\mathcal{B}$ *with respect to* P.

PROOF. We may code the values of Z as 0 and 1. Then any $\mathcal{C}$-measurable variable has the form $aZ + b$. If $E(Z|\mathcal{B})\{= P(Z = 1|\mathcal{B})\}$ is nondegenerate, then (i) does not hold, while (ii) does, since only if $a = b = 0$ will we have $E(aZ + b|\mathcal{B}) = 0[P]$. Conversely, if $E(Z|\mathcal{B})$ is degenerate, then (i) holds while (ii) does not.

9. Extensions. The abstract framework of statistical operations and conditional independence introduced in this paper has been motivated by the wish to rigorize the intuitive concept of conditional independence in such a way as to be applicable in a very wide range of statistical problems. For most purposes it should prove adequate, but it is possible to bring in still further abstraction, and with it relax some of the regularity conditions imposed. One direction is to replace statistical operations by *statistical morphisms*, introduced by Martin, et al, (1971). Roughly speaking, these bear the same relationship to pairwise sufficiency as statistical operations do to sufficiency. Their use would enable pairwise sufficiency to replace sufficiency throughout this paper, and render unnecessary the sufficiency conditions of Theorems 6.3 and 7.3, as well as the several requirements that a σ-field be regular. A further small step to abstraction retrieves the formalism of Le Cam (1964). However, while mathematically very well behaved, this is rather far removed from everyday statistical structures.

Indeed, the framework presented here may seem somewhat more abstract than needed, but I believe it to be necessary if it is to cover all the desired applications. In practice it will be important to know, for example, when statistical operations can be treated as regular conditional distributions (see, for example, Neveu, 1965, Proposition V.4.4, for a special case). Such will often be so, and ease considerably the interpretation of the results.

REFERENCES

[1] BAHADUR, R. R. (1954). Sufficiency and statistical decision functions. *Ann. Math. Statist.* **25** 423–462.

[2] BAHADUR, R. R. (1955). Statistics and subfields. *Ann. Math. Statist.* **26** 490–497.

[3] BARANKIN, E. W. (1961). Sufficient parameters: solution of the minimal dimensionality problem. *Ann. Inst. Statist. Math.* **12** 91–118.

[4] BARNDORFF-NIELSEN, O. (1965). Identifiability of mixtures of exponential families. *J. Math. Anal. Appl.* **12** 115–121.

[5] BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families in Statistical Theory.* Wiley, Chichester, New York, Brisbane.

[6] BASU, D. (1955). On statistics independent of a complete sufficient statistic. *Sankhyā* **15** 377–380.

[7] BASU, D. (1958). On statistics independent of sufficient statistics. *Sankhyā* **20** 223–226.

[8] BASU, D. (1977). On the elimination of nuisance parameters. *J. Amer. Statist. Assoc.* **72** 355–366.

[9] BIRNBAUM, A. (1962). On the foundations of statistical inference (with discussion). *J. Amer. Statist. Assoc.* **57** 269–326.

CONDITIONAL INDEPENDENCE 617

[10] BREIMAN, L., LE CAM, L. and SCHWARTZ, L. (1964). Consistent estimates and zero-one sets. *Ann. Math. Statist.* **35** 157–161.
[11] BURKHOLDER, D. L. (1961). Sufficiency in the undominated case. *Ann. Math. Statist.* **32** 1191–1200.
[12] CAILLOT, C. et MARTIN, F. (1972). Le modèle bayésien. *Ann. Inst. H. Poincaré, Sec. B* **8** 183–210.
[13] CHANDRA, S. (1977). On the mixture of probability distributions. *Scand. J. Statist.* **4** 105–112.
[14] DAWID, A. P. (1976). Properties of diagnostic data distributions. *Biometrics* **32** 647–658.
[15] DAWID, A. P. (1979a). Conditional independence in statistical theory (with discussion). *J. Roy. Statist. Soc. Ser. B* **41** 1–31.
[16] DAWID, A. P. (1979b). Some misleading arguments involving conditional independence. *J. Roy. Statist. Soc. Ser. B* **41** 249–252.
[17] DAWID, A. P. (1979c). A Bayesian look at nuisance parameters. Paper presented at the International Meeting on Bayesian Statistics, Valencia, May 28–June 2, 1979. *Trabajos de Estadistica.* To appear.
[18] DAWID, A. P. and DICKEY, J. M. (1977a). Likelihood and Bayesian inference from selectively reported data. *J. Amer. Statist. Assoc.* **72** 845–850.
[19] DAWID, A. P. and DICKEY, J. M. (1977b). Problems with nuisance parameters—traditional and Bayesian concepts. Mimeographed manuscript, Univ. College London.
[20] HALMOS, P. R. and SAVAGE, L. J. (1949). Application of the Radon-Nikodym theorem to the theory of sufficient statistics. *Ann. Math. Statist.* **20** 225–241.
[21] KOEHN, U. and THOMAS, D. L. (1975). On statistics independent of a sufficient statistic: Basu's lemma. *Amer. Statist.* **29** 40–42.
[22] KOLMOGOROV, A. N. (1942). Sur l'estimation statistique des paramètres de la loi de Gauss. *Bull. Acad. Sci. U.R.S.S. Ser. Math.* **6** 3–32.
[23] KUDŌ, H. (1978). Notes on the almost measurability. In *Statistical Decision Functions and Measure Theory*. (Lecture Notes 327), Inst. Math. Sci., Kyoto Univ.
[24] KUSAMA, T. and YAMADA, S. (1972). On compactness of the statistical structure and sufficiency. *Osaka J. Math.* **9** 11–18.
[25] LE CAM, L. (1964). Sufficiency and approximate sufficiency. *Ann. Math. Statist.* **85** 1419–1455.
[26] LEHMANN, E. L. and SCHEFFÉ, H. (1950). Completeness, similar regions and unbiased tests. Part I. *Sankhyā* **10** 305–340.
[27] LEHMANN, E. L. and SCHEFFÉ, H. (1955). Completeness, similar regions and unbiased tests. Part II. *Sankhyā* **15** 219–236.
[28] MARTIN, F., PETIT, J-L. et PETIT-LITTAYE, M. (1971). Comparison des expériences. *Ann. Inst. H. Poincaré, Section B* **7** 145–176.
[29] MARTIN, F., PETIT, J-L. et LITTAYE, M. (1973). Independence conditionnelle dans le modèle statistique bayésien. *Ann. Inst. H. Poincare, Section B* **9** 19–40.
[30] MEYER, P. A. (1966). *Probability and Potentials*. Blaisdell, Waltham, Mass.
[31] MORSE, N. and SACKSTEDER, R. (1966). Statistical isomorphism. *Ann. Math. Statist.* **37** 203–213.
[32] NEVEU, J. (1965). *Mathematical Foundations of the Calculus of Probability*. Holden-Day, San Francisco.
[33] SACKSTEDER, R. (1967). A note on statistical equivalence. *Ann. Math. Statist.* **38** 787–794.
[34] SKIBINSKY, M. (1967). Adequate subfields and sufficiency. *Ann. Math. Statist.* **38** 155–161.
[35] SKIBINSKY, M. (1969). Some known results concerning zero-one sets. *Ann. Inst. Statist. Math.* **21** 541–545.
[36] TEICHER, H. (1960). On the mixture of distributions. *Ann. Math. Statist.* **31** 55–73.
[37] TEICHER, H. (1961). Identifiability of mixtures. *Ann. Math. Statist.* **32** 244–248.
[38] TEICHER, H. (1967). Identifiability of mixtures of product measures. *Ann. Math. Statist.* **38** 1300–1302.
[39] TORGERSEN, E. N. (1977). Prediction sufficiency when the loss function does not depend on the unknown parameter. *Ann. Statist.* **5** 155–163.

DEPARTMENT OF MATHEMATICS
THE CITY UNIVERSITY
NORTHAMPTON SQUARE
LONDON EC1V OHB ENGLAND