Ancillarity

E.L. Lehmann
Department of Statistics
University of California
Berkeley, California 94720

F.W. Scholz
Boeing Computer Services
P.O. Box 24346, MS: 7L-22
Seattle, Washington 98124-0346

Technical Report No. 191
January 1989

Department of Statistics
University of California
Berkeley, California 94720

Ancillarity

By E.L. LEHMANN [1]
Department of Statistics, University of California,
Berkeley, California 94720

and F.W. SCHOLZ
Boeing Computer Services, P.O. Box 24346, MS: 7L-22,
Seattle, Washington 98124-0346

1. Introduction.

A statistic is ancillary if its distribution does not depend on the parameters of the model. It might appear at first sight as if ancillary statistics could make no contribution to inference about these parameters. However, as was pointed out by Fisher who first defined and named the concept (1925, 1934, 1935, 1936), this appearance is deceptive. By themselves ancillaries of course carry no information about the parameters, but they may be very useful in conjunction with other parts of the data.

Ancillarity has connections with many other statistical concepts, among them sufficiency, group families, conditionality, completeness, information, pre–randomization and mixtures. Its most important impact on statistical methodology comes from the suggestion that inference should be carried out conditionally given an ancillary statistic rather than unconditionally. For small samples, the resulting conditional procedures could be less efficient than their unconditional counterparts; however, they have the advantage of greater relevance to the situation at hand and frequently are simpler. Typically, the efficiency difference tends to disappear as the sample size becomes large (see for example Barndorff–Nielsen (1983) and Liang (1984)).

Since ancillaries typically are not unique, the recommendation to condition on an ancillary is not sufficiently specific. Conditioning comes closest to its purpose of making the inference relevant to the situation at hand if the ancillary is maximal, i.e. if there exists no other (nonequivalent) ancillary of which it is a function. The concept of maximal ancillary, which is

Research was supported in part by National Science Foundation grant DMS-8401388.

basic to the theories of ancillarity and conditioning, was introduced by Basu (1959) who showed that maximal ancillaries always exist[1], but noted that even they may not be unique. In the same paper he also pointed out some measure theoretic complications which require the slightly weaker definition of essential maximality for their resolution. Further results and some basic examples were given in Basu (1964) and some additional generalizations in Basu (1967).

Ancillarity is in a certain sense the dual to sufficiency. If T is a sufficient statistic, then any inference can be based solely on T and the conditional distribution of the full data set X given T is independent of the parameters. Conversely, if V is ancillary, inference may be based entirely on the conditional distribution of X given V, while the distribution of V is independent of the parameters. In this duality, a maximal ancillary corresponds to a minimal sufficient statistic. They differ however in that a minimal sufficient statistic is essentially unique and that explicit methods for its construction are available, neither of which is the case for maximal ancillaries.

Systems including sufficient and ancillary statistics as special cases are discussed in Basu (1967). Another common generalization of both sufficiency and ancillarity are the corresponding concepts (partial sufficiency and partial ancillarity) in the presence of nuisance parameters. Discussions of these concepts can be found, for example, in Dawid (1975), Basu (1977), and Barndorff-Nielsen (1978).

General discussions of various aspects of ancillarity are given by Cox and Hinkley (1974), Hinkley (1980b), Buehler (1982), Kalbfleisch (1982), and Lehmann (1986). A recent important development is the extension to asymptotic ancillarity, i.e. statistics with limit distribution independent of the parameters, and from that to higher order and local ancillaries. In the present paper, we shall restrict attention to exact ancillaries with respect to all unknown parameters, i.e. in the original sense considered by Fisher and Basu. However, work on both partial and approximate ancillaries will

[1] For a more precise statement see Theorem 4.1

be included in the bibliography.

2. Relation to other concepts.

(i) Group families.

A *group family* or *transformation model* is obtained by subjecting a random variable with a fixed distribution to a group $\mathcal{G}$ of transformations. Any statistic $V(X)$ that is invariant under $\mathcal{G}$ is ancillary. Thus in particular a maximal invariant with respect to $\mathcal{G}$ is ancillary.

Example 2.1. Location family.

Let $X = (X_1, \ldots, X_n)$ be distributed according to a location family with density

$$f(x_1 - \theta) \cdots f(x_n - \theta) .$$

This is a group family obtained by subjecting a random variable $X = (X_1, \ldots, X_n)$ with density $f(x_1, \ldots, x_n)$ to the group of transformations

$$X_i' = X_i + c, \quad i = 1, \ldots, n, \quad -\infty < c < \infty .$$

A maximal invariant is the set of differences

$$Y = (X_1 - X_n, \ldots, X_{n-1} - X_n) .$$

This is the example with which Fisher introduced the concept of ancillarity.

For some general results for the case of group families see Barndorff-Nielsen (1980).

(ii) Mixture experiments.

Suppose a family of experiment $\mathcal{E}_z$, $z \in \mathcal{Z}$ is available, each experiment consisting of a family of distributions $\mathcal{P}_z = \{P_{z,\theta}, \ \theta \in \Omega\}$, labeled by the same parameter θ, i.e. corresponding to the same states of nature. A value of z is selected according to a known distribution Π and the experiment $\mathcal{E}_z$ is performed, resulting in the observation of a random quantity X with distribution $P_{z,\theta}$. For the final result X of such a mixture experiment, Z is ancillary since its distribution Π is known.

Example 2.2. Two workers.

Let

$$\mathcal{E}_0 = (X, \mathcal{P}), \quad \mathcal{P} = \{P_\theta, \ \theta \in \Omega\}$$
$$\mathcal{E}_1 = (Y, \mathcal{Q}), \quad \mathcal{Q} = \{Q_\theta, \ \theta \in \Omega\}$$

be two experiments, corresponding for example to two different workers A and B performing a needed experimental task. One of the workers is chosen at random (with probability 1/2 each) and is assigned to perform the experiment. Here a random variable taking on the values 0 and 1 as worker A or B is chosen plays the role of Z. —The example, which was first discussed in this context by Cox (1958), makes clear the appeal of conditioning on the experiment actually performed.

Mixture models appear to represent a rather special case of models admitting ancillaries but in fact, unlike group families, they cover all cases. To see this, suppose that X is distributed according to one of the distributions P_θ, $\theta \in \Omega$ and that V is ancillary for X. For each value v, let $\mathcal{E}_v$ be the experiment consisting in observing a random quantity X', distributed according to the conditional distribution of X given v. Then X' is the outcome of a mixture experiment and its distribution is the same as that of X.

Some authors have introduced distinctions between real and conceptual (Basu, 1964) or experimental and mathematical (Kalbfleisch, 1975, 1982) mixtures. However, these distinctions require going outside the postulated models and are based on considerations involving other models.

(iii) Conditionality; pre–randomization.

Fisher's suggestion that inference should be conditional on an ancillary is called the ***principle of conditionality.*** As was discovered by A. Birnbaum (1962), conditionality has surprisingly strong consequences for the foundations of statistics since in conjunction with sufficiency it implies the likelihood principle. For discussions of this result and its consequences see Rao (1971), Basu (1975), Joshi (1983), Berger and Wolpert (1984), and Evans, Fraser and Monette (1986).

Typically, conditioning on ancillaries seems reasonable. However, it runs into difficulty when the design involves deliberate randomization (e.g.

random selection of a sample, random assignment of subjects, or random choice of a Latin square). Since the random selection process with known probabilities is ancillary, the conditionality principle would require conditioning on the selected arrangement, thus largely vitiating the purposes of randomization. This difficulty is discussed, for example, in Basu (1969, 1978, 1980), Berger and Wolpert (1984), and Finch (1986).

(iv) Sufficiency.

Sufficient statistics provide data reduction without loss of information. The amount of reduction that can be achieved in this way depends on the situation.

Example 2.1. Location family (continued).

If the density f in Example 2.1 is the standard normal density, sufficiency reduces the full n–dimensional sample $X_1, \ldots, X_n$ to the single statistic $\bar{X} = \sum_{i=1}^{n} X_i/n$, regardless of the size of n. On the other hand, if f is, for example, the logistic, Cauchy, or double exponential density, the minimal sufficient statistic is the set of order statistics $X_{(1)} \leq \ldots, \leq X_{(n)}$, so that there is hardly any reduction. As discussed in Lehmann (1981), the amount of reduction depends essentially on how much of the ancillary information the minimal sufficient statistic retains.

(v) Completeness.

The most favorable situation for reduction by means of a sufficient statistic T is that in which all ancillaries are independent of T. A sufficient condition for this to occur is given by the following result which (together with a converse) is known as Basu's theorem (Basu, 1955, 1958, 1982 and Koehn and Thomas, 1975).

Theorem 2.1. (Basu).

If T is boundedly complete, then every ancillary is independent of T.

That bounded completeness is not necessary for every ancillary to be independent of T can be seen for instance from examples in which the constants are the only ancillaries. A condition that is necessary, but not sufficient, is provided by the concept of *weak completeness*, introduced by Basu and Ghosh (1969), and independently in the present context by Lehmann (1981) under the term $\mathcal{F}_0$–completeness.

Definition 2.1. A statistic T is weakly complete with respect to a family $\mathcal{P}^T = \{P_\theta,\ \theta \in \Omega\}$ of distributions of T if

$$E_\theta f(T) = 0 \ \text{ for all } \ \theta \in \Omega \implies f(t) = 0 \ \ (a.e.\mathcal{P}^T)$$

for all two–valued functions f.

As we shall see later, this concept is central to the study of maximal ancillaries.

Note. A (not very useful) completeness condition that is both necessary and sufficient for every ancillary to be independent of T is given by Lehmann (1981).

(vi) Conditionality and sufficiency in conflict.

The principles of conditionality and sufficiency may conflict, as in the following example of Becker and Gordon (1983), which is essentially equivalent to one considered in a different context by Fisher (1956 p. 47).

Example 2.3. Quadrinomial.

Consider n quadrinomial trials with the probabilities of the four outcomes being

$$p_1 = \frac{1+\theta}{5},\ \ p_2 = \frac{1-\theta}{5},\ \ p_3 = \frac{1-\theta}{5},\ \ p_4 = \frac{2+\theta}{5}, \quad -1 < \theta < 1\,,$$

and with $N_1, \ldots, N_4$ denoting the numbers of the trials resulting in these outcomes. Then $T = (N_1,\ N_2 + N_3,\ N_4)$ is minimal sufficient and it appears that there are no ancillaries based on T. On the other hand, $A = (N_1{+}N_2, N_3{+}N_4)$ is clearly ancillary, and so is $B = (N_1{+}N_3,\ N_2{+}N_4)$.

It seems clear to the present authors that here sufficiency should be given priority over ancillarity, and inference should be based on T. For otherwise, given a trinomial situation with probabilities $((1+\theta)/5,\ (1-2\theta)/5,\ (2+\theta)/5)$, (the distribution of T), we would prefer a procedure that would require dividing the trials in the middle category, each with probability 1/2 between two artificial subcategories. This seems very unappealing.

(vii) Similar regions and regions of Neyman structure.

A set S in the sample space is a similar region with respect to a family $\mathcal{P} = \{P_\theta,\ \theta \in \Omega\}$ if $P_\theta(X \in S)$ does not depend on θ, i.e. if its indicator is ancillary. The set S is said to have Neyman structure with respect to a sufficient statistic T if the conditional probability

$$P(X \in S|t) \quad \text{is independent of } t \ a.e.$$

Suppose now that T is boundedly complete. Then by Theorem 2.1 every ancillary — and therefore the indicator I_S of any similar region — is independent of T and therefore has Neyman structure. The characterization of all similar regions as having Neyman structure in the presence of a complete sufficient statistic is therefore mathematically (although not in its interpretation) equivalent to Theorem 2.1.

(viii) Information.

Fisher's primary interest in introducing ancillary statistics was the "recovery of information." If $I_X(\theta)$ and $I_{\hat{\theta}}(\theta)$ denote the amount of Fisher information in the sample X and the maximum likelihood estimator $\hat{\theta}$ respectively, then it will often happen that[2] $I_{\hat{\theta}}(\theta) < I_X(\theta)$, so that $\hat{\theta}$ is not fully informative. Fisher discovered that the lost information can be recovered if there exists an ancillary statistic V such that $(\hat{\theta},\ V)$ is sufficient, in the following sense. If $I_{\hat{\theta}|v}(\theta)$ is the information carried by $\hat{\theta}$ in the conditional distribution given $V = v$, then

$$E\, I_{\hat{\theta}|V}(\theta) = I_X(\theta)\,. \tag{2.1}$$

For a discussion of the implementation of this program in two important classes of models, see Barndorff-Nielsen (1980). When (2.1) holds, the average conditional information equals the whole information in the sample; for particular values of v, the conditional information of $\hat{\theta}$ given v may be smaller or larger than $I_X(\theta)$.

Recall now the other motive for conditioning on ancillaries: to make the inference more relevant to the situation at hand. Cox (1971) points out that

[2] We have here assumed for the sake of simplicity that θ is real valued.

ancillaries are therefore most useful when the amount $I_v(\theta)$ of information in the conditional distribution of X given v varies widely with v, so that some values of v are much more informative than others. This point is nicely illustrated by Example 2.2, where conditioning on the chosen worker seems particularly important when there is a big difference in the quality of their work.

In the light of this remark, Cox suggests that when the maximal ancillary is not unique, that ancillary should be preferred for which $I_v(\theta)$ is most variable, e.g. for which the variance $\mathrm{var}[I_V(\theta)]$ is the largest.

3. Weak completeness.

The central concept for the characterization of maximal ancillaries is weak completeness. It is easy to see that the definition of weak completeness given in the preceding section is equivalent to the following statement.

(3.1) The family $\mathcal{P} = \{P_\theta,\ \theta \in \Omega\}$ is weakly complete if any measurable set A with probability independent of θ has probability 0 or 1.

This is the form in which the definition was given by Basu and Ghosh (1969).

A simple restatement of (3.1) yields

Theorem 3.1. A family $\mathcal{P}$ admits no nontrivial ancillaries (i.e. any ancillary statistic is almost surely constant) if and only if $\mathcal{P}$ is weakly complete.

To illustrate the situation of no ancillaries consider the following examples.

Example 3.1. No ancillaries.

Let X_i be independent $N(\theta_i, 1)$, $i = 1, \ldots, n$. Then $X = (X_1, \ldots, X_n)$ is complete, hence weakly complete, and so there are no ancillaries.

Example 3.2. Sequential binomial sampling.

Consider a sequence of binomial trials, with success probability p and a stopping rule (i.e. with probability 1 of eventually stopping). This can be represented by a random walk in the plane starting at the origin, with a unit step to the right for a success and a unit step up for a failure. The stopping rule is represented by a set of stopping points. The observation is a path starting at $(0,0)$ and ending at some stopping point (a,b). Since every path ending at (a,b) has probability $p^a(1-p)^b$, it follows that the coordinates (a,b) of the stopping point constitute a sufficient statistic, which may or may not be complete (necessary and sufficient conditions for completeness are given in Lehmann and Stein, 1950). The path itself is of course not complete except in the rare cases in which there is only one path to each stopping point.

(i) In light of this it is very surprising that not only the endpoint but also the path itself is weakly complete, provided the stopping rule has a finite boundary point on the x or the y axis. To see this let S be a set of paths with $P_p(S) = c \quad \forall\ p \in (0,1)$. Suppose the stopping rule has a finite boundary point $(0,k)$ for some $k \geq 1$. Then the path π_0 from $(0,0)$ to $(0,k)$ is either contained in S or in its complementary set of paths S^c. It follows that either $c = P_p(S) \geq P_p(\pi_0) = (1-p)^k \to 1$ or $1 - c = P_p(S^c) \geq P_p(\pi_0) = (1-p)^k \to 1$ as $p \to 0$ so that either $c = 1$ or $c = 0$. The case of finite boundary point $(k,0)$ is treated similarly. Hence there are no ancillaries.

(ii) If there is no bound on the stopping rule along the $x-$ or $y-$axis then weak completeness may not obtain as the following example shows. Perform the binomial trials in pairs until the first time that either (success, failure) or (failure, success) is observed. Then the set S of paths that end in (failure, success) has probability 1/2 for all $p \in (0,1)$.

Note. Exactly the same result as in Example 3.2 with the same proof applies to sequential sampling from trinomial (or any multinomial) trials.

The following example is due to Basu and Ghosh (1969) where many additional examples can be found.

Example 3.3. Two-point location families.

Let X take on the two values θ and $\theta + c$ with probabilities

$$P(X = \theta) = \pi, \;\; P(X = \theta + c) = 1 - \pi, \;\; -\infty < \theta < \infty,$$

π and c known. Then X is weakly complete provided $\pi \neq 1/2$, but not when $\pi = 1/2$. In the latter case any set A whose complement is $A + c$ has probability 1/2, independent of θ.

It turns out that Theorem 3.1 is a special case of a general characterization of maximality for an ancillary statistic V, given in its proper setting in Theorem 5.1. Loosely, this characterization finds V to be maximal if and only if the family of conditional distributions of X given V is weakly complete. In the situation of Theorem 3.1, where V is constant, this family of conditional distributions coincides with the family $\mathcal{P}$ of distributions for X.

In the case when the only ancillary statistics are the a.s. constant functions there (usually) does not exist a maximal ancillary (due to null set problems) but a maximal ancillary σ–field $\mathcal{A}_m$ does exists, see Theorem 4.1. The reason is that not every σ–field is induced by a statistic. Since the σ–field induced by an a.s. constant function is essentially equivalent to $\mathcal{A}_m$ (to be made precise below) it makes sense to call such an a.s. constant function essentially maximal ancillary; the alternative would be to admit that there are no maximal ancillary statistics due to null set problems. This state of affairs carries over to the general case and the above loosely stated characterization is that of essential maximal ancillarity. Bearing this in mind one may want to accept that characterization and skip or skim Sections 4 and 5.

4. Notation and definitions.

Let $(\mathcal{X}, \mathcal{B})$ be an arbitrary measurable space and $\{P_\theta,\ \theta \in \Omega\}$ be a family of probability measures on $\mathcal{B}$. Considering $\mathcal{X}$ as the sample space we denote the random element in $\mathcal{X}$ by X and write $P_\theta(X \in B) = P_\theta(B)$ $\forall B \in \mathcal{B}$. We now give some definitions and a theorem taken from Basu (1959).

Definition 4.1. A σ–field $\mathcal{A} \subset \mathcal{B}$ is said to be *ancillary* if $P_\theta(A)$ is constant in $\theta \in \Omega\ \forall A \in \mathcal{A}$.

Comment. One easily sees that $\mathcal{A}$ is ancillary iff $\int f(x)\, dP_\theta(x)$ is constant in $\theta \in \Omega$ for all integrable and $\mathcal{A}$-measurable functions $f : \mathcal{X} \to R$.

Definition 4.2. If $V : (\mathcal{X}, \mathcal{B}) \to (\mathcal{Y}, \mathcal{C})$ is a statistic ($\mathcal{A}_V := V^{-1}(\mathcal{C}) \subset \mathcal{B}$) then V is said to be ancillary if $\mathcal{A}_V$ is ancillary.

Comment. Rather than dealing with (ancillary) statistics we follow Basu's example and continue the following theoretical exposition in terms of (ancillary) σ–fields. When dealing with concrete examples we will use the more intuitive term "statistic" in place of σ–field. Hence it is understood that the following definitions in terms of σ–fields have analogous counterparts in terms of "statistics."

Definition 4.3. An ancillary σ–field $\mathcal{A} \subset \mathcal{B}$ is said to be *maximal ancillary* if there exists no other ancillary σ–field $\mathcal{A}^\star \subset \mathcal{B}$ such that $\mathcal{A} \subset \mathcal{A}^\star$.

Theorem 4.1. (Basu, 1959). Given an ancillary σ–field $\mathcal{A} \subset \mathcal{B}$ there exists a maximal ancillary σ–field $\mathcal{A}_m \subset \mathcal{B}$ such that $\mathcal{A} \subset \mathcal{A}_m$.

Definition 4.4. Two σ–fields $\mathcal{A}_1$, $\mathcal{A}_2$ $\subset \mathcal{B}$ are said to be *essentially equivalent* if for any $A_1 \in \mathcal{A}_1$ ($A_2 \in \mathcal{A}_2$) there exists an $A_2 \in \mathcal{A}_2$ ($A_1 \in \mathcal{A}_1$) such that

$$P_\theta(A_1 \Delta A_2) = 0 \quad \forall\, \theta \in \Omega\,.$$

Definition 4.5. Any ancillary σ–field that is essentially equivalent to a maximal ancillary σ–field is called *essentially maximal ancillary.*

Comment. Although Theorem 4.1 guarantees the existence of a maximal ancillary σ–field $\mathcal{A}_m$ containing any given ancillary σ–field $\mathcal{A}$ the same does not necessarily hold for statistics. The reason is that $\mathcal{A}_m$ is usually too rich to be generated by any statistic V.

The following definition of conditional weak completeness is a direct adaptation of the concept of weak completeness to the conditioned case.

Definition 4.6. X given $\mathcal{A}$ is said to be *conditionally weakly complete* if for any given function

$$g(x) = a(x)I_B(x) + b(x)I_{B^c}(x)$$

with $B \in \mathcal{B}$, $a(\cdot)$ and $b(\cdot)$ $\mathcal{A}$–measurable and such that

$$\forall\theta \in \Omega \quad E_\theta(g(X)|\mathcal{A}) = 0 \quad a.s.(P_\theta)$$

we have

$$\forall\theta \in \Omega \quad P_\theta(g(X) = 0|\mathcal{A}) = 1 \quad a.s.(P_\theta)\,,$$

i.e. $P_\theta(g(X) = 0) = 1\ \forall\theta \in \Omega$.

An equivalent formulation of Definition 4.6, without the "*a.s.*" qualifiers, is

Definition 4.6.′ X given $\mathcal{A}$ is said to be *conditionally weakly complete* if for any given function

$$g(x) = a(x)I_B(x) + b(x)I_{B^c}(x)$$

with $B \in \mathcal{B}$, $a(\cdot)$ and $b(\cdot)$ $\mathcal{A}$-measurable and such that

$$E_\theta(I_A(X)g(X)) = 0 \quad \forall\, \theta \in \Omega \quad \text{and } \forall\, A \in \mathcal{A}$$

we have $P_\theta(g(X) = 0) = 1 \quad \forall\, \theta \in \Omega$.

Note that Definitions 4.6 and 4.6′ are not contingent on the existence of regular conditional distributions. However, if X admits regular conditional distributions given $\mathcal{A}$ a natural question is: how does weak completeness of a family of regular conditional probability distributions relate to the conditional weak completeness of X given $\mathcal{A}$ defined above? Lemma 4.1 will provide a partial answer under certain regularity conditions. These conditions are as follows:

i) Ω is a separable topological space,

ii) $\mathcal{A}$ is generated by the ancillary statistic $V : (\mathcal{X}, \mathcal{B}) \rightarrow (\mathcal{Y}, \mathcal{C})$,

iii) $\forall\, v \in \mathcal{Y}$: $\{f_\theta(\cdot|v),\ \theta \in \Omega\}$ is a family of conditional densities for X given $V = v$ with respect to a σ-finite dominating measure μ on $(\mathcal{X}, \mathcal{B})$,

iv) $\exists\, N \in \mathcal{C}$ with $P(V \in N) = 0$ so that $\forall\, v \in N^c$ we have $f_\theta(x|v) \rightarrow f_{\theta_0}(x|v)$ *a.s.* in $x[\mu]$ whenever $\theta \rightarrow \theta_0$,

Lemma 4.1. Under conditions i)–iv) the weak completeness of the families $\{f_\theta(\cdot|v),\ \theta \in \Omega\}$ $\forall\, v \in N_1^c$ with $P(N_1) = 0$ implies the conditional weak completeness (Definition 4.6) of X given $\mathcal{A}$.

Proof: Let g be as in Definition 4.6, then for any $\theta \in \Omega$ we have

$$0 = E_\theta(g(X)|V) = \int g(x) f_\theta(x|V)\, d\mu(x) \quad a.s.\ P\,. \tag{4.1}$$

Since the exceptional null set may depend on θ (through f_θ) we invoke (4.1) for all θ in a countably dense subset of Ω. Using Scheffé's theorem in conjunction with iv) it follows that there exists a set $N_0 \in \mathcal{C}$ such that for $v \in N_0^c$ we have

$$0 = \int g(x) f_\theta(x|v)\, d\mu(x) \quad \forall\, \theta \in \Omega$$

which by weak completeness of the conditional densities entails for all $v \in N_0^c$

$$\int I_{\{g(X)=0\}} f_\theta(x|v)\, d\mu(x) = 1 \quad \forall\, \theta \in \Omega\,,$$

hence $P_\theta(g(X) = 0) = 1 \;\; \forall\, \theta \in \Omega$.

It is not clear whether the converse of Lemma 4.1 is true under the stated conditions.

5. Characterization of maximal ancillarity.

The following theorem will give necessary and sufficient conditions for an ancillary σ–field $\mathcal{A} \subset \mathcal{B}$ to be essentially maximal ancillary. A special case of Theorem 5.1 was proved by Basu and Ghosh (1969) for the case of the dominated location family.

Theorem 5.1. If $\mathcal{A} \subset \mathcal{B}$ is ancillary, then the following statements are equivalent:

- i) $\mathcal{A}$ is essentially maximal ancillary.
- ii) $\nexists\, B \in \mathcal{B}$ such that $P_\theta(B|\mathcal{A})$ admits a version ψ_B ($\mathcal{A}$–measurable) independent of $\theta \in \Omega$ with $P(0 < \psi_B(X) < 1) > 0$.
- iii) X given $\mathcal{A}$ is conditionally weakly complete.

Proof. i) $\Rightarrow$ ii). $\mathcal{A}$ be ancillary and let $B \in \mathcal{B}$ be such that $P_\theta(B|\mathcal{A})$ admits a version ψ_B ($\mathcal{A}$–measurable) independent of $\theta \in \Omega$. First note that the smallest σ–field $\mathcal{A}_B$ containing both $\mathcal{A}$ and B is ancillary, since

$$P_\theta(A \cap B) = \int_A \psi_B(x)\, dP_\theta(x) \quad A \in \mathcal{A}$$

is independent of $\theta \in \Omega$ ($\mathcal{A}$ is ancillary and ψ_B is $\mathcal{A}$–measurable and independent of $\theta \in \Omega$) and since this property extends to all of $\mathcal{A}_B$ by the usual unique measure extension.

Next let $A_0 = \{x \in \mathcal{X} : 0 < \psi_B(x) < 1\} \in \mathcal{A}$. Assuming $\mathcal{A}$ to be essentially maximal ancillary we can find $A_1 \in \mathcal{A}$ such that $N = A_1 \Delta (A_0 \cap B) \in \mathcal{A}_B$ has probability zero for all $\theta \in \Omega$. Then

$$I_{A_1}(X) = I_{A_0}(X) I_B(X) \quad \forall\; X \in N^c$$

and taking conditional expectation given $\mathcal{A}$ we have

$$I_{A_1}(X) = I_{A_0}(X)\psi_B(X) \quad a.s.\ P_\theta \quad \theta \in \Omega$$

which implies $P(0 < \psi_B < 1) = 0$, thus i) $\Rightarrow$ ii).
ii) $\Rightarrow$ iii). Let

$$g(x) = a(x)I_B(x) + b(x)I_{B^c}(x) = (a(x) - b(x))I_B(x) + b(x)$$

$B \in \mathcal{B}$, $a(\cdot)$ and $b(\cdot)$ $\mathcal{A}$–measurable such that

$$\forall\, \theta \in \Omega \quad E_\theta(g(X)|\mathcal{A}) = 0 \quad a.s.\ P_\theta\,. \tag{5.1}$$

Let $C_0 = \{x \in \mathcal{X} : a(x) \neq b(x)\}$, $B_0 = C_0 \cap B$ and

$$\begin{aligned} \psi_{B_0}(x) \ &= b(x)/(b(x) - a(x)) \quad && x \in C_0 \\ &= 0 && x \in C_0^c \end{aligned}$$

The condition (5.1) on g implies that ψ_{B_0} may serve as a θ–independent version of $P_\theta(B_0|\mathcal{A})$ $\forall\, \theta \in \Omega$, since

$$0 = E_\theta(I_{C_0}(X)g(X)|\mathcal{A}) = (a(X) - b(X))P_\theta(B_0|\mathcal{A}) + b(X)I_{C_0}(X) \quad \text{a.s. } P_\theta,$$

i.e.

$$X \in C_0 \ \Rightarrow \ P_\theta(B_0|\mathcal{A}) = b(X)/(b(X) - a(X)) = \psi_{B_0}(X) \ \text{ a.s. } P_\theta$$

and for $X \in C_0^c \ \Rightarrow P_\theta(B_0|\mathcal{A}) = 0 = \psi_{B_0}(X)$ a.s. P_θ.
Condition (5.1) also implies

$$0 = E_\theta(g(X)I_{C_0^c}(X)|\mathcal{A}) = g(X)I_{C_0^c}(X) \quad a.s.\ P_\theta \ \forall\, \theta \in \Omega\,. \tag{5.2}$$

Since $P(0 \leq \psi_{B_0} \leq 1) = 1$

$$ii) \Rightarrow P(\psi_{B_0} \in \{0,1\}) = 1$$

$$\Rightarrow I_{B_0}(X) = \psi_{B_0}(X) \quad a.s. \quad P_\theta \ \forall\, \theta \in \Omega$$

$$\Rightarrow g(X)I_{C_0}(X) = 0 \quad a.s. \quad P_\theta \ \forall\, \theta \in \Omega$$

since

$$0 = E_\theta(g(X)I_{C_0}(X)|\mathcal{A}) = (a(X) - b(X))\psi_{B_0}(X) + b(X)I_{C_0}(X)$$

$$= (a(X) - b(X))I_{B_0}(X) + b(X)I_{C_0}(X) = I_{C_0}(X)g(X) \quad \text{a.s. } P_\theta \ \forall\, \theta \in \Omega\,.$$

This together with (5.2) implies $P_\theta(g(X) = 0) = 1 \ \forall\, \theta \in \Omega$, i.e. ii) $\Rightarrow$ iii).
iii) $\Rightarrow$ i). By theorem 4.1 there exists a maximal ancillary σ–field $\mathcal{A}_m \supset \mathcal{A}$. Let $D_0 \in \mathcal{A}_m$ and for some fixed $\theta_0 \in \Omega$ and some version $P_{\theta_0}(D_0|\mathcal{A})$ let $\psi_{D_0}(x) := P_{\theta_0}(D_0|\mathcal{A})$, then for $A \in \mathcal{A}$:

$$\int_A P_\theta(D_0|\mathcal{A})\, dP_\theta = P_\theta(A \cap D_0) = P_{\theta_0}(A \cap D_0)$$

$$= \int_A \psi_{D_0}(x)\, P_{\theta_0}(x) = \int_A \psi_{D_0}(x)\, dP_\theta(x)\,,$$

i.e. ψ_{D_0} may serve as a θ–independent version of $P_\theta(D_0|\mathcal{A}) \ \forall\, \theta \in \Omega$. Let

$$g(x) = I_{D_0}(x) - \psi_{D_0}(x)$$

then $\forall\ \theta \in \Omega \quad E_\theta(g(X)|\mathcal{A}) = 0 \quad a.s. \ P_\theta$, which under iii) implies $P_\theta(g(X) = 0) = 1 \ \forall\, \theta \in \Omega$, i.e.

$$\psi_{D_0}(X) = I_{D_0}(X) \ \ a.s. \ P_\theta \ \forall\, \theta \in \Omega$$

which shows $\mathcal{A}$ and $\mathcal{A}_m$ to be essentially equivalent, i.e. iii) $\Rightarrow$ i) q.e.d.

6. Examples.

In the examples that follow it is understood that when claiming maximal ancillarity what is really meant is essential maximal ancillarity. However, these two concepts coincide when the null set idiosyncracies do not arise, as in situations when that ancillary is discrete.

Example 6.1. With probability $\frac{1}{2}$ let $X_1, \ldots, X_n$ be i.i.d. from $N(\theta, 1)$, and with probability $\frac{1}{2}$ from $N(\theta, 2)$. Let $I = 1$ or 0 as the first or the second case obtains. Then $V = (I, X_1 - X_n, \ldots, X_{n-1} - X_n)$ is maximal ancillary since $(I, X_1, \ldots, X_n)$ is equivalent to $(\bar{X}, V)$ and the conditional distribution of $\bar{X}$ given V is complete.

Example 6.2. Let $X_1, \ldots, X_n$ be i.i.d. with continuous and strictly increasing c.d.f. F. This model is invariant under the group G of common, continuous, strictly increasing transformations $X_i' = g(X_i), \ i = 1, \ldots, n$. Maximal invariant is the vector of ranks $(R_1, \ldots, R_n)$ of the n X's. Since the group G is transitive, the maximal invariant is ancillary. Is it maximal ancillary? Since the conditional distribution of the X's given the ranks is the same as the joint distribution of the rank permuted order statistics and since the distribution of the latter is complete, hence weakly complete, it follows that the ranks are maximal ancillary.

Example 6.3. In Example 6.2, suppose attention is restricted to F with median 0. Now the ranks are no longer maximal ancillary since the ranks together with the number of positive observations are ancillary. This latter ancillary is maximal since the order statistic given the number of positive and negative observations are complete. (We are dealing with n_+ and n_- observations from arbitrary continuous and strictly increasing distribution functions on $(0, \infty)$ and $(-\infty, 0)$ respectively. Note: This maximal ancillary is a maximal invariant under a smaller group than in Example 6.2, namely the group G of transformations g which are continuous, strictly increasing and satisfy $g(0) = 0$.)

Example 6.4. Let $X_1, \ldots, X_n$ be i.i.d. $N(\theta, 1)$. Here of course the vector of differences $(X_1 - X_n, \ldots, X_{n-1} - X_n)$ is maximal ancillary since the distribution of $\bar{X}$ is complete.

As has been pointed out by Basu (1959) and others, maximal ancillarity does not mean that there are no other maximal ancillaries. As a well known example, in the present case with $n = 2$, we have that $V = (X_2 - X_1)\text{sign}(\bar{X})$ is ancillary. To see that it is also maximal note that (X_1, X_2) is equivalent to $(\bar{X}, V)$ and that $\bar{X}$ and V are independent. Now the completeness of $\bar{X}$ entails the conditional weak completeness of $(\bar{X}, V)$ given V.

Another maximal ancillary is $V' = X_2 - X_1$. Which of these two ancillaries is preferable? The Cox criterion discussed in Section 2 (viii) does not distinguish between them; however, a criterion advanced by Barnard and Sprott (1971) applies and gives preference to $X_2 - X_1$ since it is invariant under translations (see Padmanabhan, 1977).

Example 6.5. Let $X_1, \ldots, X_n$ be i.i.d. uniform on $[\theta, \theta + 1)$. Here (denoting by $[x]$ the integer part of x) $(X_1 - X_n, \ldots, X_{n-1} - X_n)$ together with $X_n - [X_n]$ are ancillary and are easily seen to be maximal ancillary since the conditional distribution of $[X_n]$ (all that is left of the data for any fixed θ) given that ancillary is just a one point distribution which is complete.

Basu (1964) treats this example in the case $n = 1$; Basu and Ghosh (1969) treat the same example for the case of arbitrary n for which they determine the maximal ancillary σ–field.

Basu and Ghosh (1969) show that a sufficient condition for weak completeness of the location family of densities $\{f(x - \theta) : \theta \in R\}$ is that the characteristic function $\hat{f}(t) = \int \exp(-itx) f(x)\, dx$ of f has at most a finite number of roots on the real line.

Example 6.6. (Basu and Ghosh). Let X have density $f(x - \theta)$ with $f(x) = x^2 \exp(-x^2/2)/\sqrt{2\pi}$. Since $\hat{f}(t) = (1 - t^2)\exp(-t^2/2)$ which has only two roots it follows that X is weakly complete and hence admits only the *a.s.* constant functions as ancillaries.

Example 6.7. The general location family. Let $X_1, \ldots, X_n$ be i.i.d. $\sim f(x-\theta)$ where $f(x)$ is a density with respect to Lebesgue measure on R. The differences $V = (V_2, \ldots, V_n) = (X_1 - X_2, \ldots, X_1 - X_n)$ are ancillary and the question is for which f may one claim also maximal ancillarity? Examples 6.4 and 6.5 show that the answer depends on f. The conditional density of $U = X_1$ given $V = v = (v_2, \ldots, v_n)$ is $h_\theta(u|v) = c\, f(u-\theta) f(u-\theta-v_2) \cdots f(u-\theta-v_n)$ with c being the appropriate normalizing constant. Since this yields a univariate location family $\{h_\theta(u|v) = h_v(u-\theta) : \theta \in R\}$ with $h_v(x) = c\, f(x) f(x - v_2) \cdots f(x - v_n)$ one could appeal to the above sufficient criterion of Basu and Ghosh to establish weak completeness for this family by showing that $\hat{h}_v(t)$ has only a finite number of roots.

Unfortunately, the Basu–Ghosh criterion of a finite number of roots frequently is not satisfied and then does not provide an answer concerning maximality. Examples for which this is the case are the Cauchy and double exponential distributions with $n = 2$.

Example 6.8. Cox and Hinkley (p.33, 1974) give the following simplified version of an example due to Basu (1964) which points out the dilemma of multiple ancillaries. Consider N quadrinomial trials with probabilities

$$\frac{1}{6}(1-\theta),\ \frac{1}{6}(1+\theta),\ \frac{1}{6}(2-\theta),\ \frac{1}{6}(2+\theta)\,.$$

If the number of outcomes in the four categories are X, U, Y, V, respectively, then $X+U$ is ancillary, as is $X+V$. The question is whether either is maximal ancillary. The answer is somewhat surprising and still mostly a conjecture.

(i) First consider the case of $X+U$. The conditional distribution of (X,Y) given $X+U=m$, $Y+V=n$, $(n+m=N)$ is that of two independent binomial random variables, distributed respectively as $b(p_1,m)$ and $b(p_2,n)$ with $p_1=(1-\theta)/2$ and $p_2=(2-\theta)/4$. Since the conditional expectation of $X/m-2Y/n+1/2$ vanishes for all θ we do not have conditional bounded completeness, whenever $m\geq 1$ and $n\geq 1$. If $m=0$ or $n=0$ completeness follows easily.

To establish weak completeness (conditionally) one needs to show that for any indicator function $f(x,y)$ with constant conditional expectation for all θ it follows that f is either identically one or zero with conditional probability one. For $0\leq\alpha\leq 1$ consider therefore the following identity for all θ:

$$\sum_{x=0}^{m}\sum_{y=0}^{n} f(x,y)\binom{m}{x}\binom{n}{y}\left(\frac{1-\theta}{2}\right)^{x}\left(\frac{1+\theta}{2}\right)^{m-x}\left(\frac{2-\theta}{2}\right)^{y}\left(\frac{2+\theta}{2}\right)^{n-y}\equiv\alpha\,.$$

Show that $f\equiv 0$ and $f\equiv 1$, or equivalently that $\alpha=0$ and $\alpha=1$, are the only solution. Reparametrizing $\lambda=(1-\theta)/(1+\theta)$ the identity becomes

$$\sum_{x=0}^{m}\sum_{y=0}^{n} f(x,y)\binom{m}{x}\binom{n}{y}\lambda^{x}(1+3\lambda)^{y}(3+\lambda)^{n-y}\equiv\alpha 4^{n}(1+\lambda)^{m+n}\,.$$

Comparing the coefficients of λ^{i} and λ^{m+n-i} for $i=0,\ 1,\ 2$ on both sides of the identity and exploiting the binary nature of f it is easy yet tedious to show weak completeness for the following cases: 1) $n=1$ and $m=1$, $m\geq 3$ and 2) $n=2$ and $m\geq 1$. For the case $(m,n)=(2,1)$ we don't have weak completeness as can easily be seen by using $f(0,1)=f(2,0)=1$ and $f(x,y)=0$ otherwise.

Using the reparametrization $\lambda=(2-\theta)/(2+\theta)$ one can show weak completeness for all (n,m) with 3) $m=1$ and $n\geq 1$ and 4) $m=2$ and $n\geq 1$ (no counter example here). The above approach does not appear promising for the situations $n\geq 3$ and $m\geq 3$.

(ii) Similar results can be obtained when considering the other ancillary, $X + V$, except that the above counter example does not obtain, i.e. the conditional distribution of (X, Y) given $X + V = m$ is weakly complete for (m, n) in the following cases 1) $n = 1$ and $m \geq 1$, 2) $n = 2$ and $m \geq 1$, 3) $m = 1$ and $n \geq 1$, 4) $m = 2$ and $n \geq 1$.

What does this mean with respect to maximal ancillarity of $X + U$ and $X + V$? For $N = m + n \leq 5$ the latter is maximal ancillary whereas the former is maximal ancillary for $N = 1, 2, 4, 5$ but not for $N = 3$. The maximality in the cases $N > 5$ at this point can only be conjectured.

References

Amari, S. (1982). Geometrical theory of asymptotic ancillarity and conditional inference. *Biometrika* **69**, 1–17

Andersen, E.B. (1967). On partial sufficiency and partial ancillarity. *Skand. Aktuar. Tidskr.* **50**, 137–152.

Barnard, G.A. and Sprott, D.A. (1971). A note on Basu's examples of anomalous ancillary statistics. In: *Foundations of Statistics* (Godambe and Sprott, eds), Holt, Rinehart and Winston. Toronto.

Barndorff–Nielsen, O. (1978). *Information and exponential families in statistical theory.* John Wiley. New York

Barndorff–Nielsen, O. (1980). Conditionality resolutions. *Biometrika* **67**, 293–310.

Barndorff–Nielsen, O. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70**, 343–365.

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

Basu, D. (1958). On statistics independent of a sufficient statistic. *Sankhyā* **20**, 223–226.

Basu, D. (1959). The family of ancillary statistics. *Sankhyā* (A) **21**, 247–256.

Basu, D. (1964). Recovery of ancillary information. *Sankhyā* (A) **26**, 3–16.

Basu, D. (1967) Problems relating to the existence of maximal and minmal elements in some families of statistics (subfields). In *Proc. Fifth Berkeley Symp. on Math. Statistics and Probability* Vol. 1, 41–50, Univ. of Calif. Press.

Basu, D. (1969). Role of the sufficiency and likelihood principles in sample survey theory. *Sankhyā* **31**, 441–454.

Basu, D. (1971). An essay on the logical foundations of survey sampling. In: *Foundations of Statistics* (Godambe and Sprott, eds), Holt, Rinehart and Winston. Toronto.

Basu, D. (1975). Statistical information and likelihood (with discussion). *Sankhyā* (A) **37**, 1–71.

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

Basu, D. (1978). Relevance of randomization in data analysis. In *Survey Sampling and Measurement* (Namboodiri, ed.) 267–292.

Basu, D. (1980). Randomization analysis of experimental data: The Fisher randomization test (with discussion). *J. Amer. Statist. Assoc.* **75**, 575-595.

Basu, D. (1982). Basu Theorems. In *Encycl. Statist. Sci.* **1**, 193–196.

Basu, D. and Ghosh J.K. (1969). Invariant sets for translation–parameter families of measures. *Ann. Math. Statist.* 40, 162–174.

Becker, N. and Gordon, I. (1983). On Cox's criterion for discriminating between alternative ancillary statistics. *Int. Statist. Rev.* **51**, 89–92.

Berger, J.O. and Wolpert, R.L. (1984). *The Likelihood Principle*, IMS Lecture Notes–Monograph Series, Vol. 6.

Birnbaum, A. (1962). On the foundations of statistical inference (with discussion). *J. Amer. Statist. Assoc.* **57**, 269–306.

Buehler, R.J. (1982). Some ancillary statistics and their properties. *J. Amer. Statist. Assoc.* **77**, 581–589.

Cox, D.R. (1958). Some problems connected with statistical inference. *Ann. Math. Statist.* **29**, 357–372.

Cox, D.R. (1971). The choice between alternative ancillary statistics. *J. Roy. Statist. Soc.* (B) **33**, 251–255.

Cox, D.R. (1980). Local ancillarity. *Biometrika* **67**, 279–286.

Cox, D.R. (1984). Discussion of a paper by F. Yates. *J. Roy. Statist. Soc.* (A) **147**, 451.

Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*, Chapman and Hall. London.

Dawid, A.P. (1975). On the concepts of sufficiency and ancillarity in the presence of nuisance pasrameters. *J. Roy. Statist. Soc.–B* **37**, 248–258.

Efron, B. and Hinkley, D.V. (1978). Assessing the accuracy of the maximum likelihood estimators: observed versus expected Fisher information. *Biometrika* **65**, 457–487.

Evans, M., Fraser, D.A.S. and Monette, G. (1986). On principles and arguments to likelihood. *Can. J. Statist.* **14**, 181–200.

Finch, P.D. (1986). Randomization I. In *Encycl. Statist. Sci.* **7**, 516–519.

Fisher, R.A. (1925). Theory of statistical estimation. *Proc. Camb. Philos. Soc.* **22**, 700-725.

Fisher, R.A. (1934). Two new properties of mathematical likelihood. *Proc. Roy. Soc.* (A) **144**, 285–307.

Fisher, R.A. (1935). The logic of inductive inference. *J. Roy. Statist. Soc.* **98**, 39–54.

Fisher, R.A. (1936). Uncertain inference. *Proc. Am. Acad. Arts and Sci.* **71**, 245–258.

Fisher, R.A. (1948). Conclusions fiduciaires. *Ann. Inst. Henri Poincaré* **10**, 191–213.

Fisher,. R.A. (1956). *Statistical Methods and Scientific Inference*, Oliver and Boyd. Edinburgh.

Fraser, D.A.S. (1973). The elusive ancillary. In: *Multivariate Statistical Inference*, edited by D.G. Kabe and R.P. Gupta, North Holland, American Elsevier.

Fraser, D.A.S. and Reid, N. (1988). On conditional inference for a real parameter: A differential approach on the sample space. *Biometrika* **75**, 251–264.

Gordon, I. (1983). Ancillarity and minimal sufficiency. *Austral. J. Statist.* **25**, 273–277.

Hinkley, D.V. (1980a). Likelihood as approximate pivotal distribution. *Biometrika* **67**, 287–292.

Hinkley, D.V. (1980b). Fisher's development of conditional inference. In *R.A. Fisher: An Appreciation* (101–108). Springer. New York.

Hosoya, Y. (1988). The second order Fisher information. *Biometrika* **75**, 265–274.

Joshi, V.M. (1983). Likelihood principle. *Encycl. Statist. Sci.* **4**, 644–647.

Kalbfleisch, J.D. (1975). Sufficiency and conditionality (with discussion). *Biometrika* **62**, 251–268.

Kalbfleisch, J.D. (1982). Ancillary Statistics. *Encycl. Statist. Sci.* **1**, 77–81.

Kempthorne, O. (1986). Randomization II. *Encycl. Statist. Sci.* **7**, 519–524.

Koehn, U. and Thomas, D.L. (1975). On statistics independent of a sufficient statistic: Basu's Lemma. *Amer. Statist.* **29**, 40–42.

Lehmann, E.L. (1981). An interpretation of completeness and Basu's theorem. *J. Amer. Statist. Assoc.* **76**, 335–340.

Lehmann, E.L. (1986). *Testing Statistical Hypotheses.* 2^{nd} *Ed.* John Wiley. New York.

Lehmann, E.L. and Stein, C. (1950). Completeness in the sequential case. *Ann. Math. Statist.* **21**, 376–385.

Liang, K.Y. (1984). The asymptotic efficiency of conditional likelihood methods. *Biometrika* **71**, 305–313.

Padmanabhan, A.R. (1977). Ancillary statistics which are not invariant. *Amer. Statistician* **31**, 124.

Rao, C.R. (1971). Some aspects of statistical inference in problems of sampling from finite populations (with discussion). In: *Foundations of Statistics* (Godambe and Sprott, eds), Holt, Rinehart and Winston. Toronto.

Sandved, E. (1967). A principle for conditioning on an ancillary statistic. *Skand. Aktuar. Tidskr.* 50, 39–47.

Sandved, E. (1972). Ancillary statistics in models without and with nuisance parameters. *Skand. Aktuar. Tidskr.* **55**, 81–91.

Skovgaard, I. (1986). Successive improvement of the order of ancillarity. *Biometrika* **73**, 516–519.

Sverdrup, E. (1966). The present state of the decision theory and the Neyman–Pearson theory. *Rev. Int. Statist. Inst.* **34**, 309–333.

Young, A. (1986). Conditioned data–based simulations: Some examples from geometrical statistics. *Rev. Int. Statist. Inst.* **54**, 1–13.