Austral. J. Statist., **30(2)**, 1988, 163–173

APPLICATIONS AND EXTENSIONS OF BASU'S RESULTS ON MAXIMAL ANCILLARITY

CHRISTOPHER J. LLOYD

Department of Statistics, University of Melbourne

Summary

Maximality of ancillaries is important in both conditional inference without nuisance parameters and marginal inference with nuisance parameters. We extend results of Basu (1959) to the more classical former case and discuss the different nature of ancillaries in these two contexts. We apply the results to a general ancillary independent of a sufficient statistic. Finally, we discuss difficulties in finding necessary conditions for maximality.

Key words: conditional completeness; conditional inference.

1. Introduction

Ancillary statistics have two rather distinct roles in statistical inference. Let us first describe and contrast these dual roles. In the first context we have a model for data X involving an unknown parameter θ. Ancillaries are statistics which are distributed free of θ and it is usually held appropriate to condition on the observed values of such irrelevant statistics. The arguments for such conditioning will not be elucidated here. What should be mentioned, however, is that when the ancillary is neither independent of, nor a part of, the minimal sufficient statistic (MSS) for θ then conditioning will violate the sufficiency principle as applied to the full unconditional model. This being the case it is useful to introduce some terminology to identify ancillaries of different logical types.

Let S^θ be MSS for θ. An ancillary will be called *internal* (to S^θ) if it is a function of S^θ, *external* (to S^θ) if it is independent of S^θ and *neutral* if it is neither internal nor external. In the present context, external ancillaries are of no relevance to inference since conditioning on them is essentially achieved by sufficiency reduction. Neutral ancillaries lead to the conflict between the conditionality and sufficiency principles described above. Internal ancillaries may be effectively conditioned on without violating sufficiency.

The second context embraces models involving nuisance parameters, ϕ, as well as interest parameters θ. Let $S^{(\theta,\phi)}$ be the MSS for (θ,ϕ) and $T^{\phi}(\theta)$ the MSS for ϕ for a given known value of θ. Sufficiency requires us to base inference on some function of $S^{(\theta,\phi)}$. Typically, a set

$$\{\theta < \psi(S^{(\theta,\phi)})\}$$

with ϕ-free mass $1-\alpha$ is required for some function ψ since this generates a confidence interval for θ as a critical region for testing θ. If we treat $\theta = \theta_0$ as temporarily known then the indicator of this set is ancillary for ϕ. It may be either internal, external or neutral to $T^{\phi}(\theta)$. In contrast to the first context, the ancillary here is marginalised to, rather than conditioned on. It forms the θ-informative summary of the data. Of course, these quantities are a function of θ_0 and ought more properly to be called pivotal quantities. Nevertheless, it is clear in the present context that using ancillaries (or pivotal quantities) which are not internal to $T^{\phi}(\theta)$ does not violate the sufficiency principle as applied to the model with both θ and ϕ unknown.

The notion of ancillarity was first introduced by Fisher (1925, 1934) in the first context. Although he defined ancillaries without explicit reference to their relation with the MSS, it is clear that his discussion concerned internal ancillaries since they were to be characteristics of the likelihood function. Cox (1971) does explicitly define ancillaries to be part of the MSS (internal in our terminology) in contrast to earlier work by Basu (1959) where only freedom from θ in the distribution is required. Barnard (1963) and Fraser (1973) define ancillaries in terms of transformation invariance in an attempt to bring some (group) structure to the problem. More recently, Cox (1980) considers the use of statistics which are internal approximate ancillaries for conditioning in asymptotics. However, it is clear that in the second context, Basu's relaxed definition is appropriate since only with such a definition may the treatment of ancillaries in the two contexts be unified.

Lastly, why study maximal ancillaries (i.e. ancillaries not a function of any other)? In the first context maximality corresponds to carrying the conditioning process as far as it can go; in the second context maximality provides a largest pivotal quantity to marginalise to. Again in both contexts maximality of the ancillary is complementary to the minimality of the sufficient statistic. In the first case we reduce the sample to its minimum dimension and further reduce by conditioning on the maximal ancillary. In the second we pack all the ϕ-relevant data into as small a statistic as possible, $T^{\phi}(\theta)$, so that we can base inference on as much remaining data as possible.

In this paper we extend Basu's (1955, 1959) results which identify maximal ancillarity as complementary to complete sufficiency. Briefly stated, completeness is first relaxed to conditional completeness. Secondly, when considering only external ancillaries, the requirement of completeness may be removed entirely. These results may be used to demonstrate maximality of internal, external or neutral ancillaries.

2. Statistics and Sub-Fields

The results of this paper are given in the language of measure theory although they will be reexpressed in the language of statistics where appropriate. Those unfamiliar with the measure theory approach should consult Chung (1968) or a standard text. We give here only a brief description of the mathematical ideas employed.

We deal throughout with a σ-field (X, Σ) representing the sample space of an experiment together with all events of possible interest. In addition there is a family of probability measures defined on (X, Σ) indexed by a possibly vector valued parameter θ, $\{P_\theta : \theta \in \Omega\}$. These represent possible models for the data. We denote the integral of a Σ-measurable function f with respect to P_θ by $E_\theta(f)$. For a given set $x \in \Sigma$ and σ-field $\mathcal{S} \subset \Sigma$ we define the 'conditional probability of x given $\mathcal{S}$' to be an $\mathcal{S}$-measurable function $P(x \mid \mathcal{S})$ which satisfies

$$\int_s P(x \mid \mathcal{S})\, dP = P(s \cap x) \qquad \text{for every } s \in \mathcal{S}\ ,$$

where θ has been suppressed. Clearly, $P(x \mid \mathcal{S})$ can be defined arbitrarily on sets of measure zero, however any two versions of $P(x \mid \mathcal{S})$ are equal P a.s.

Just as the σ-field Σ represents the whole data, so one can define a σ-field to represent an arbitrary statistic. If T is any Σ-measurable function taking values in R^k then

$$S_T = \{x \in \Sigma : T(x) \quad \text{is a Borel subset of } R^k\}$$

is called the 'σ-field induced by T'. The σ-fields induced by statistics have been studied by Bahadur (1954, 1955). The relationships of statistics are mirrored by the relationships of their induced σ-fields. For example if T_1 is expressible as a function of T_2 then $S_{T_1} \subseteq S_{T_2}$. If T_1 and T_2 are equal (P a.s.) then S_{T_1} and S_{T_2} are 'essentially equal' i.e. to every set in one there corresponds a (P a.s.) equal set in the other. The intersection of two statistics (T_1, T_2) induces the σ-field $S_{T_1} v S_{T_2}$ where 'v' denotes Borel extension of the union of two σ-fields. Finally, a maximal σ-field is one which is contained in no other and if a maximal σ-field is induced by a statistic then that statistic is maximal.

3. Conditional Sufficiency, Completeness and Independence

We give conditional versions of some common statistical notions in terms of properties of σ-fields.

Definitions.

(i) Λ and Γ are independent σ-fields given Ξ if, for every set $\lambda \in \Lambda$

$$P_\theta(\lambda \mid \Gamma v \Xi) = P_\theta(\lambda \mid \Xi) \quad \text{almost surely } (P_\theta)\,.$$

(ii) Λ is a complete σ-field given Ξ if the only $\Lambda v \Xi$-measurable functions with identically zero expectation over every set $\xi \in \Xi$ are zero a.s. (P_θ).

(iii) Λ is a sufficient σ-field for θ given Ξ if, for every set $X \in \Sigma$, the conditional probability $P(X \mid \Lambda v \Xi)$ may be chosen free of θ.

(iv) A is an ancillary σ-field for θ given Ξ if, for every $\alpha \in A$, the conditional probability $P(\alpha \mid \Xi)$ may be chosen free of θ.

Theorem 9.2.1 of Chung (1968) shows that (i) is equivalent to

$$P_\theta(\lambda\gamma \mid \Xi) = P_\theta(\lambda \mid \Xi) P_\theta(\gamma \mid \Xi) \quad \text{for every } \gamma \in \Gamma,\ \lambda \in \Lambda,\ \theta \in \Omega\,.$$

An alternative to (ii) is that for every Λ-measurable function g

$$E_\theta(g \mid \Xi) = 0 \quad \text{for every } \theta \in \Omega \text{ implies } g = 0\ (P_\theta \text{ a.s.})\,.$$

This weaker definition turns out to be inadequate for our purposes. Of course, (iii) is more simply stated as sufficiency of $\Lambda v \Xi$.

Lemma 1. *If, for every set α in a σ-field A, $P(\alpha \mid \Xi) = 1_\alpha P_\theta$ a.s. then A is essentially contained in Ξ.*

Proof. The set $\{P(\alpha \mid \Xi) = 1\}$ is in Ξ and essentially equal to α.

Lemma 2. *If Ξ is an ancillary σ-field contained in the ancillary σ-field A then A is ancillary given Ξ.*

Proof. For arbitrary $\theta_1, \theta_2 \in \Omega$, $\alpha \in \Xi$ let $S = \{x : P_{\theta_1}(\alpha \mid \Xi) > P_{\theta_2}(\alpha \mid \Xi)\}$ which belongs to Ξ. Now for any Ξ-measurable function f, $E_{\theta_1}(f) = E_{\theta_2}(f)$. In particular,

$$E_{\theta_1}\big(P_{\theta_2}(\alpha \mid \Xi)1_S\big) = E_{\theta_2}\big(P_{\theta_2}(\alpha \mid \Xi)1_S\big) .$$

We hence obtain

$$\int_S \big[P_{\theta_1}(\alpha \mid \Xi) - P_{\theta_2}(\alpha \mid \Xi)\big]\, dP_{\theta_1} = 0$$

so that $P_{\theta_1}(\alpha \mid \Xi) = P_{\theta_2}(\alpha \mid \Xi)$ (P_{θ_1} a.s.) on S. It follows that the same is true on all of X and that $P(\alpha \mid \Xi)$ may be chosen free of θ.

4. Sufficient Conditions for Maximality

The theoretical results of the paper are stated and explained in this section. Some applications are given later. The first result extends the result of Basu (1955) and the second and third his Theorem 7 (1959).

Theorem 1. *Let Λ be complete and sufficient for θ given Ξ and A be ancillary given Ξ. Then A and Λ are independent given Ξ.*

Proof. Let $\alpha \in A$. By conditional sufficiency, $P(\alpha \mid \Lambda v \Xi)$ can be chosen free of θ. By definition of conditional probability

$$\int_\sigma P(\alpha \mid \Lambda v \Xi)\, dP_\theta = P_\theta(\alpha \cap \sigma) \quad \text{for every } \theta \in \Omega,\ \sigma \in \Lambda v \Xi \tag{1}$$

$$\int_\xi P(\alpha \mid \Xi)\, dP_\theta = P_\theta(\alpha \cap \xi) \quad \text{for every } \theta \in \Omega,\ \xi \in \Xi\, . \tag{2}$$

Since $\Lambda v \Xi$ contains Ξ, this implies that for every $\theta \in \Omega$, $\xi \in \Xi$

$$\int_\xi \big[P(\alpha \mid \Lambda v \Xi) - P(\alpha \mid \Xi)\big]\, dP_\theta = 0 \tag{3}$$

and by conditional completeness the integrand is zero almost surely (P_θ). By definition (i) this is just Ξ-conditional independence of Λ and A.

Theorem 2. *Let Ξ be ancillary and Λ complete and sufficient for θ given Ξ. Suppose $\Lambda v \Xi$ is essentially equivalent to Σ. Then Ξ is maximal ancillary.*

Proof. Let A be ancillary and contain Ξ. Then A is ancillary given Ξ by Lemma 2 and so, by Theorem 1, for every set $\alpha \in A$

$$P(\alpha \mid \Xi) = P(\alpha \mid \Lambda v \Xi) \quad \text{almost surely } (P_\theta)\ . \tag{4}$$

But since $\Lambda v \Xi$ is essentially equivalent to Σ, $P(\alpha \mid \Xi) = 1_\alpha$ almost surely (P_θ) so by Lemma 1, A is essentially contained in Ξ.

A special case is when Λ is complete sufficient unconditionally. This is Basu's Theorem 7 (1959). In terms of statistics Theorem 2 says that if the data may be written as (S, A) where A is ancillary and S is complete sufficient for θ given A then A is maximal ancillary. The proof of this is based on the independence of ancillary and complete sufficient statistics. Not surprisingly then there is an analogous result with ancillarity replaced by independence.

Theorem 3. *Let Ξ and Λ be independent σ-fields and suppose $\Lambda v \Xi$ is essentially equivalent to Σ. Then Ξ is essentially maximal among σ-fields independent of Λ.*

Proof. Let A be independent of Λ and contain Ξ. Since $Av\Xi = A$ is independent of Λ it follows that for any $\alpha \in A$,

$$P(\alpha \mid \Xi) = P(\alpha \mid \Lambda v \Xi) \quad \text{almost surely } (P_\theta)$$

(see Chung (1968), p.308). i.e. A is independent of Λ given Ξ. Finally, since $\Lambda v \Xi$ is essentially equivalent to Σ we have $P(\alpha \mid \Xi) = 1_\alpha$ almost surely (P_θ) so that A essentially contains Ξ.

This result can be used where Λ is minimal sufficient but not complete.

Corollary. *Let Λ be minimal sufficient for θ but not necessarily complete. Any external ancillary σ-field Ξ such that $\Lambda v \Xi$ is essentially equivalent to Σ is maximal external ancillary.*

Basu's Theorem 7 only identifies maximal external ancillaries in the special case when the minimal sufficient statistic is complete, whence all ancillaries are external. Our results are applicable to all three logical types as we show in the following well-known example.

Example 1. Consider n independent pairs of observations on the double exponential distribution of Fisher (1948) with density

$$f_{X,Y}(x, y) = e^{-x\theta - y/\theta} \qquad x, y, \theta > 0\ .$$

The data may be written as

$$(XY, (Y/X)^{1/2}, X_1/X_n, \ldots, X_{n-1}/X_n, Y_1/Y_n, \ldots, Y_{n-1}/Y_n)$$

where $X = \sum X_i$, $Y = \sum Y_i$. Consider the MSS for θ $(XY, (Y/X)^{1/2})$. Since, conditional on XY, the maximum likelihood estimator $(Y/X)^{1/2}$ is complete sufficient for θ it follows by Theorem 2 that the ancillary XY is maximal internal. Conditional on the neutral ancillary

$$\mathbf{A}_1 = (XY, X_1/X_n, \ldots, X_{n-1}/X_n, Y_1/Y_n, \ldots, Y_{n-1}/Y_n)$$

$(Y/X)^{1/2}$ is, of course, still complete and so $\mathbf{A}_1$ is a functionally maximal ancillary statistic by Theorem 2.

Finally, the ancillary

$$\mathbf{A}_2 = (X_1/X_n, \ldots, X_{n-1}/X_n, Y_1/Y_n, \ldots, Y_{n-1}/Y_n)$$

is external to $S = (XY, (Y/X)^{1/2})$. Since $(\mathbf{X}, \mathbf{Y}) = (S, \mathbf{A}_2)$, $\mathbf{A}_2$ is maximal external ancillary by the Corollary to Lemma 3. None of these facts follows from Basu's original Theorem 7.

5. Application to the Conditional Distribution Transform

Williams (1982) and Lloyd (1987) have studied a method of constructing external ancillaries. We study the maximality of a generalised version below and also consider its relation to the sufficiency principle.

Consider data $X_1, X_1, \ldots, X_n$ with X_0 degenerate and a MSS $T(X_1, \ldots, X_n)$ for the parameter ϕ. Define the statistic W with ith component given by

$$W_i = F_i(X_i \mid T, X_{i-1}, \ldots, X_0) \quad i = 1, \ldots, n-1 .$$

Here $F_i(\ \mid t, x_{i-1}, \ldots, x_0)$ is a conditional distribution function (CDF). Provided each CDF is continuous, each W_i is defined free of ϕ and together they have independent uniform distributions jointly independent of T. For W to be maximal (external) we must be able to solve for the data from (T, W). By monotonicity of each CDF we can solve for X_1 from T, W_1 and for X_2 from X_1, T, W_2. Thus we can solve for $X_1, \ldots, X_{n-1}$ and the last variable X_n may be obtained if and only if, T is monotonic in X_n. This gives necessary and sufficient conditions for construction of a maximal external ancillary whose maximality follows by Theorem 3. However, W

may, or may not, be maximal when monotonicity of T does not hold. A typical case when it is not is if $T = \sum \psi_i(X_i)$ and $\psi_n(X_n)$ is equivalent to $f(X_n), g(X_n)$ with $g(X_n)$ ancillary and $f(X_n)$ monotonic for fixed $g(X_n)$. If T is complete given $g(X_n)$ then the maximal ancillary is

$$W_1, \ldots, W_{n-1}, g(X_n) .$$

If we now introduce an interest parameter θ into the model then W becomes a pivotal quantity external to $T^\phi(\theta)$. In order to satisfy sufficiency we choose functions of this pivotal quantity involving only the MSS for (θ, ϕ). The same pivotal quantities apparently arise if we first marginalise to this MSS and then apply the CDF transform to the reduced experiment. A study of this will not be made here. We illustrate these ideas with a simple example.

Example 2. Consider $Z_1, \ldots, Z_n$ i.i.d. on $N(\theta, \phi^2)$. Let $X_i = Z_i - \theta$ 'absorb' the interest parameter θ. Since the X_i are independent

$$W_i = F(X_i \mid T^2) \qquad i = 1, \ldots, n-1$$

where $T^2 = \sum X_i^2$ is sufficient for ϕ. It is easy to show that W_i is a 1–1 function of X_i/T. However, we cannot deduce it is maximal since T^2 is not monotonic. In fact, it is not maximal as is seen by decomposing X_n^2 into $|X_n|$ and $\text{sign}(X_n)$. Since $\text{sign}(X_n)$ is ancillary for ϕ and $|X_n|$ is 1-1 given $\text{sign}(X_n)$ the maximal (external) pivotal quantity is

$$A = \left((Z_1 - \theta)/T, \ldots, (Z_{n-1} - \theta)/T, \text{sign}(Z_n - \theta)\right) .$$

If we initially marginalise to the sample mean and variance $\overline{Z}$ and S^2 then, with θ known, $T^2 = (n-1)S^2 + n\overline{Z}^2$ is still sufficient for ϕ and

$$F(\overline{X} \mid T^2) = \Phi(\sqrt{n}(\overline{Z} - \theta)/T) ,$$

a monotonic function of the student t pivotal, $\sqrt{n}(\overline{Z} - \theta)/S$. This is the only function of the pivotal quantity A which is a function of $\overline{Z}, S^2$ only.

6. Difficulties with Necessary Conditions for Maximality

Given the solubility conditions of Theorems 2 and 3, is conditional completeness also necessary for maximality? The fact that it is not rests on the observation that not every incomplete statistic contains an ancillary component. More problematically, we cannot even tell whether or not certain incomplete statistics contain an ancillary. For example, observations on standard normal variables (X, Y) with correlation ρ admit a minimal sufficient statistic $\big(\sum(X^2+Y^2), \sum XY, N\big)$, where the sample size N is considered random. Conditional on $N = n$, $\big(\sum(X^2+Y^2), \sum XY\big)$ is incomplete yet it is the consensus of the literature that N is maximal internal ancillary for ρ. There is also a host of non-regular examples where the support set depends on the parameter θ. In such examples, conditioning on the ancillary seems intuitively compelling since only by doing so do we take into account the restricted range for θ which different configurations of the sample imply. A typical example is

Example 3. Consider i.i.d. observations (X_i, Y_i) on the distribution

$$f_{X,Y}(x,y) = 2\theta y \qquad 0 \le x \le \theta, \quad 0 \le y \le \theta^{-1} .$$

The MSS is $(X_{(n)}, Y_{(n)})$ with $T = 1/Y_{(n)}$ the maximum likelihood estimate. The statistic $A = X_{(n)}, Y_{(n)}$ is ancillary for θ and the conditional density is

$$f_{T|a}(t;\theta) = n\theta^n(1-a^n)^{-1}t^{-(n+1)} \qquad t \in [\theta, \theta/a]$$

where $0 \le a \le 1$. As a is close to 1 so is the conditional density of T close to θ. However, T is not conditionally complete since

$$t^n \cos(2\pi \log t / \log a)$$

has conditional expectation 0. On the other hand, A is a maximal invariant in the transformation model

$$(X/\theta, Y\theta) \quad \text{has joint density } 2y \text{ on } [0,1]\times[0,1]$$

and so is maximal ancillary. However, its maximality cannot be deduced from Theorem 2.

Similarly, the configuration ancillaries of Pitman (1939) are all maximal invariants in a location or scale transformation model but only in special but important cases will the conditional distribution be complete. On the other hand, there are models without a transformational structure where the completeness condition seems the only check of maximality. The following example demonstrates this as well as providing a wide class of non-transformational models which admit an internal ancillary. The author knows of no non-trivial examples of this in the literature.

Example 4. Let $X_1, \ldots, X_k$ be k random variables each with densities f_i of exponential form and unknown parameter θ_i. Partition the real line into k intervals $J_i = (a_{i-1}, a_i)$ with $a_0 = -\infty, a_k = \infty$ and transform $X_i = \phi_i(Y_i)$ where ϕ_i is 1-1 and Y_i varies over J_i. Then for arbitrary probability vector $p = (p_1, \ldots, p_k)$ we can generate a distribution depending on $\theta_1, \ldots, \theta_k$:

$$f(y;\theta) = f_i(\phi_i(y);\theta_i)\phi_i'(y) \qquad y \in J_i\ .$$

An observation on Y can be interpreted as a transformed observation on X_i, 'i' being identified by the interval J_i containing Y and occuring with probability p_i. For an i.i.d. sample Y^j, $j = 1, \ldots, n$ the variable

$$A(Y^i, \ldots, Y^n) = \left\{\#(Y^j \in J_i),\, i = 1, \ldots, k\right\}$$

is a k-dimensional ancillary statistic for θ with multinomial distribution and parameter p. The MSS for θ comprises A together with sums $S_i = \sum s_i(Y^j)$ where s_i comes from the exponent of f_i and the sum is taken over those Y^j in J_i. Conditional on A, the distribution is complete and hence A is the maximal internal ancillary statistic.

Acknowledgement

The author thanks Gopalan Nair, R.M. Huggins, E.J. Williams and H. Cohn for helpful advice and gratefully acknowledges the financial support of an A.R.G.S. grant. Similar unpublished results by M. Mouchart and J.M. Rolin have recently been made known to the author.

References

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

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

BARNARD, G.A. (1962). Some logical aspects of the fiducial argument. *J. Roy. Statist. Soc. B* **25**, 111–114.

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

BASU, d. (1959). The family of ancillaries. *Sankhya* **21**, 245–256.

CHUNG, K.L. (1968). *A Course in Probability Theory.* Edition 2. New York: Academic Press.

FISHER, R.A. (1925). Theory of statistical estimation. *Proc. Cam. Phil. Soc.* **22**, 700–725.

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

FISHER, R.A. (1948). Conclusions fiducaires. *Ann. Inst. Henri Poincare* **10**, 191–213.

FRASER, D.A.S. (1973). The elusive ancillary. In *Multivariate Statistical Inference*. Ed. D.G. Kabe and R.P. Gupta. North Holland.

LLOYD, C.J. (1987). Pivotal quantities based on the conditional distribution function transform in models with accessory parameters. *Comm. Statist. A.* **16**, 361–374.

PITMAN, E.J.G. (1939). The estimation of the location and scale parameters of a continuous population of any given form. *Biometrika* **30**, 391–421.

WILLIAMS, E.J. (1982). Some classes of conditional inference procedures. *Adv. Appl. Prob.* **19A**, 293–303.

Received March 1987; revised August 1987