Austral. J. Statist., **27** (3), 1985, 264–272

2

ANCILLARIES SUFFICIENT FOR THE SAMPLE SIZE[1]

C. J. LLOYD

Department of Statistics, The University of Melbourne

Summary

The process of recovery of information proceeds by conditioning on ancillaries which are part of the minimal sufficient statistic (MSS), so-called internal ancillaries. This is motivated by the notion that ancillaries function as precision indices. In some cases the ancillary can itself replace the sample size entirely in the model. It is decided that such ancillaries must be conditioned on if one of the sufficiency or conditionality principles is not to be violated.

Key words: Ancillarity; Information recovery; Sufficiency

1. Introduction

The notion of ancillarity was originally introduced by Fisher (1925) in discussing the appropriate analysis of models for which the maximum likelihood estimator (MLE), T, was not sufficient. When an ancillary complement A, such that the MSS may be written as (T, A), exists then it was recommended that the precision of T be quoted relative to the family of measures conditional on A. Apart from the "information recovery" to which this gives rise, the process was seen as analogous to the usual practice of quoting the precision of estimators based on an i.i.d. sample $X_1, \ldots, X_n$ conditional on the sample size N.

> "Their (ancillaries') function is, in fact, analogous to the part which the size of our sample is always expected to play, in telling us what reliance to place on the result" (1935, p. 39)

The details of this process were first elucidated in a series of papers by Basu (1955, 1958, 1959, 1964) which focused initially on some technical aspects of ancillarity and later on a detailed study of the schema of Fisher just described. Briefly the problems encountered were

(i) more than one ancillary complement to T may exist,

(ii) no ancillary complement to T may exist.

[1] Manuscript received September 11, 1984; revised June 6, 1985.

In either case the appropriate precision index for the MLE is unclear. Basu, and later Kalbfleisch (1975), attempted to resolve the dilemma by permitting only those ancillaries corresponding to a performable experiment, and thus more closely mimicking the sample size, to be considered as conditioning variables. More recently Bartlett (1983) gives some interesting examples where quoting the variance of the estimator conditional on an appropriate ancillary is strongly indicated. This variance will often, but not always, be lower than the unconditional variance bound (on average it is the same) and it is clear that the fundamental choice of which, if any, ancillaries to condition on must be made before the imposition of properties like minimum variance or unbiasedness. One aspect of these models, not stressed in that paper but noted in Williams (1982), is that the distribution of the estimator, conditional on the ancillary, no longer depends on the sample size. With the sample size playing such an important role in the conditionality argument we will examine some models for which the sample size, in some sense, is contained in an available ancillary component of the MSS.

2. Some Examples of *N*-Sufficient Ancillaries

The following is an extension of one of the best known examples of an internal ancillary and was in fact used by Fisher to illustrate the recovery of information process. The ancillary involved however has the property that conditioning on it removes the sample size from the joint density of the MSS, a property which forms the main subject matter of this paper.

Example 1. Let X, Y be exponential variables with joint density given by

$$f(x, y, \theta) = e^{-\theta x + y/\theta} \quad \text{for} \quad x, y, \theta > 0.$$

For an independent sample of pairs $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ the loglikelihood is given by

$$\log L(\theta) = \theta^{-1}\Sigma\, y_i + \theta \Sigma\, x_i$$

from which it is clear that $S = (X, Y) = (\Sigma\, X_i, \Sigma\, Y_i)$ is the minimal sufficient statistic for θ with respectively, $\Gamma(n, 1/\theta)$. $\Gamma(n, \theta)$ distributions and joint density function

$$f_{X,Y}(x, y) = (xy)^{n-1} e^{-(x\theta + y/\theta)}/(n-1)!^2.$$

Then $T = Y/X$ is the MLE for θ and $U = XY$ is distributed free of θ. The determinant of the Jacobian of this transformation turns out to be $1/2UT$ so that the joint density of U, T is given by

$$f_{U,T}(t, u) = \frac{1}{2ut} u^{n-1} e^{-u(\theta/t + t/\theta)}/(n-1)!^2.$$

Integrating out t and changing variables gives the marginal, θ-free, density of U as

$$f_U(u) = u^{n-2}K(u)/[2(n-1)!^2]$$

where $K(u)$ is a normalising constant, being that of the product of two independent $\Gamma(n, 1)$ variates. We hence obtain the U-conditional density

$$f(t \mid u) = \frac{1}{t} e^{-u(\theta/t + t/\theta)}/K(u).$$

The usual definition of sufficiency relates the information of a statistic to a parameter. A useful extension of this concept is to say a statistic S_1 is sufficient for a statistic S_2 if conditioning on S_1 removes S_2 from the model. Certainly, if S_2 is a function of S_1 then this is the case. However, this extended sufficiency is a more relevant relation for conditional inference than mere functional inclusion. In the above example, we see that the ancillary statistic has completely replaced the sample size as a precision index. We thus say that U is "sufficient for the sample size" and we term U "N-sufficient". Such variables are, in some sense, better than or equal to N as conditioning variables. Another example, which beautifully brings out this nature of an ancillary as a precision index, is found in Bartlett (1983).

Example 2. Consider n independent observations on the $R[\theta - \frac{1}{2}, \theta + \frac{1}{2}]$ distribution for which the minimal sufficient statistic is $(\max X_i, \min X_i)$ and, since θ is a location parameter, $\max X_i - \min X_i = A(X)$ is an internal ancillary by invariance arguments. The estimator

$$T = \tfrac{1}{2}(\max X_i + \min X_i)$$

has the $R[\theta - \frac{1}{2}(1-a), \theta + \frac{1}{2}(1-a)]$ distribution, conditional on $A = a$, so again A has replaced n as a precision index. This is in accord with intuition since it is the interval width $1 - a$ which tells us how closely we can estimate θ, not the sample size. Unfortunately, the following example of Fisher (1956), though not considered in the same context, shows us that ancillaries are not, in general, N-sufficient.

Example 3. This example is found in Fisher (1956)'s discussion of the existence of internal ancillaries.

For $i = 1, \ldots, n$ let (X_i, Y_i) be independent normal variates with unit variance and expectation on the unit circle. Letting $E(X) = \cos\theta$ and $E(Y) = \sin\theta$, the minimal sufficient statistic for θ is just

$$(X, Y) = \left(\sum_{i=1}^{n} X_i/\sqrt{n}, \sum_{i=1}^{n} Y_i/\sqrt{n}\right)$$

being independent normal variates with unit variance and means on the circle of radius $\sqrt{n}$. Transform from (X, Y) to (R, T) via

$$X = R \sin T,\ Y = R \cos T$$

giving joint density

$$f_{R,T}(r, t) = \frac{r}{2\pi} e^{-1/2(r^2+1)} e^{r\sqrt{n} \cos (t-\theta)}$$

so that the conditional density of t is

$$f(t \mid r) = \kappa(r\sqrt{n}) e^{r\sqrt{n} \cos (t-\theta)}$$

where $\kappa(r\sqrt{n})$ is the appropriate normalising constant. It will be observed that this density depends on both n and r.

Example 4. Let us take a sequence of independent observations $x_1, \ldots, x_N$ on a variable X with density $f(x, \theta)$ where the sample size N is not fixed but a random variable with distribution not depending on θ. The likelihood for this data is then

$$\log L(\theta) = \sum_{i=1}^{n} \log f(x_i, \theta) + \log P(N = n)$$

from which it is clear that a sufficient statistic for θ is given by

$$S = (X_1, X_2, \ldots, X_N, N).$$

When the family $f(x, \theta)$ is not exponential (ex. Cauchy) this will also be necessary (i.e. minimal) sufficient. The statistic N, however, is an ancillary statistic internal to the MSS and is of course also N-sufficient.

How common then is the phenomenon of N-sufficiency? The following is a generalisation of Example 1.

Example 5 (Set of exponential variates with related scales). For $i = 1, \ldots, k$ and $j = 1, \ldots, n_i$ let X_{ij} be independent exponential variates with rates $1/\theta^{b_i}$ where no two b_i are equal or zero. Then the minimal sufficient statistic for the parameter θ has ith component

$$(X_i) = X_{i1} + \ldots + X_{in_i} \quad \text{for} \quad i = 1, \ldots, k$$

and $X_i\theta^{b_i}$ has the gamma distribution with convolution parameter n_i and unit rate. Denote $\mathbf{b} = (b_1, \ldots, b_k)$, $\mathbf{n} = (n_1, \ldots, n_k)$, $\mathbf{1} = (1, \ldots, 1)$ and

$$\alpha = \mathbf{b} \cdot \mathbf{1} = \sum_{i=1}^{k} b_i, \quad \Phi = \boldsymbol{b} \cdot \boldsymbol{n} = \sum_{i=1}^{k} b_i n_i.$$

We claim that there exist exactly $k-1$ functionally independent ancillaries. Consider the equation

$$U_j = \prod_{i=1}^{k} X_i^{c_{ji}},$$

Then the $k-1$ linearly independent vectors $\mathbf{c}_j$ satisfying

$$\mathbf{c}_j \,.\, \boldsymbol{b} = 0$$

correspond to $k-1$ functionally independent statistics each with unit scale. Since the M.L.E is the solution of a polynomial of degree $\max(b_i) - \min(b_i)$ take as a reasonable estimate of θ

$$T = \prod_{i=1}^{k} X_i^{1/b_i},$$

with scale θ^k. We will transform $(X_1, \ldots, X_k) \to (U_1, \ldots, U_{k-1}, T)$. Inverting this transformation and taking the appropriate powers of the U_i and T we may write

$$X_i = g_i(\mathbf{U})T^{b_i},$$

where $g_i(\mathbf{U})$ is a product of appropriate powers of the u_i. The Jacobian of this transformation has terms

$$\frac{\partial}{\partial u_j}(x_i) = \frac{\partial g_i(\mathbf{u})}{\partial u_j} t^{b_i} = g_{ij}(\mathbf{u})t^{b_i} \quad i = 1, \ldots, k \quad j = 1, \ldots, k-1$$

$$\frac{\partial}{\partial t}(x_i) = b_i g_i(\mathbf{u})t^{b_i - 1} = g_{ik}(\mathbf{u})t^{b_i - 1} \quad i = 1, \ldots, k \quad j = 1, \ldots, k-1$$

with determinant of the form $h(\mathbf{u})t^{\alpha-1}$. The density of $X_1, \ldots, X_k$ is thus

$$H(\mathbf{n})\theta^{\Phi} \prod_{i=1}^{k} x_i^{n_i - 1} \exp\left[-\sum_{i=1}^{k} x_i \theta^{b_i}\right] \sum_{i=1}^{k} dx_i,$$

where $H(\mathbf{n}) = \prod_{i=1}^{k} 1/\Gamma(n_i)$. The density of $U_1, \ldots, U_{k-1}, T$ is hence

$$H(\mathbf{n})g(\mathbf{u})h(\mathbf{u}) \prod_{i=1}^{k-1} du_i \theta^{\Phi} t^{\Phi-1} \exp\left[-\prod_{i=1}^{k} g_i(\mathbf{u})(t\theta)^{b_i}\right] dt$$

where $g(\mathbf{u}) = \prod_{i=1}^{k} g_i(\mathbf{u})$ while the joint marginal density of $U_1, \ldots, U_{k-1}$ is

$$\left[H(\mathbf{n})g(\mathbf{u})h(\mathbf{u}) \prod_{i=1}^{k-1} du_i\right]\left[\int \theta^{\Phi} t^{\Phi-1} \exp\left[-\sum_{i=1}^{k} g_i(\mathbf{u})(t\theta)^{b_i}\right] dt\right]$$

$$= H(\mathbf{n})g(\mathbf{u})h(\mathbf{u}) \prod_{i=1}^{k-1} du_i I(\Phi, \mathbf{b}, \mathbf{u})$$

where

$$I(\Phi, \mathbf{b}, \mathbf{u}) = \int v^{\Phi-1} \exp\left[-\sum_{i=1}^{k} g_i(\mathbf{u})v^{b_i}\right] dv.$$

Hence the density of $X = \theta T$, conditional on $U = u$, is given by

$$dF(x) = \frac{\sqrt{x^{\Phi-2}} \exp\left[-\sum_{i=1}^{k} g_i(\mathbf{u})x^{b_i}\right]}{I(\Phi, \mathbf{b}, \mathbf{u})} dx$$

which depends on the sample sizes $n_1, \ldots, n_k$ only through the term Φ.

Example 6 (Set of normal variates with related scales). For $i = 1, \ldots, k$ and $j = 1, \ldots, n_i$ let X_{ij} be independent normal variates with zero means and variances $1/\sigma^{-2b_i}$ where no two b_i are equal or zero. Then the MSS for the parameter σ has ith component

$$(X_i) = X_{i1}^2 + \ldots + X_{in_i}^2 \quad \text{for} \quad i = 1, \ldots, k.$$

$X_i\sigma^{2b_i}$ has the chi-squared distribution with n_i degrees of freedom. By proceeding analogously to Example 5 it may be shown that the distribution of the estimator $\sigma^2 T$, conditional on the ancillary U, is

$$dF(x) = \frac{\sqrt{x^{\Phi-2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{k} g_i(\mathbf{u}) x^{b} i\right]}{I(\Phi, \mathbf{b}, \mathbf{u})} dx$$

where $I(\Phi, \mathbf{b}, \mathbf{u})$ is the integral of the numerator with respect to x and which depends on the sample sizes $n_1, \ldots, n_k$ again, only through the terms Φ. Suppose that the sampling scheme is such that $n_i = n$ for $i = 1, \ldots, k$. Then for the conditional density to be free of n it is necessary and sufficient that $\Sigma\, b_i = 0$. On the other hand if the sampling scheme is such that each n_i is chosen independently then a necessary and sufficient condition for the conditional density to be free of n_i is that $b_i = 0$. However in this case X_i will not be part of the minimal sufficient statistic.

We conclude from the above examples that N-sufficiency of an ancillary is the exception rather than the rule. Be that as it may, what is the significance of this property when it is present?

3. *N*-Sufficiency and the Conditionality/Sufficiency Principles

Gordon (1983) gives a multinomial example where no ancillary contained in the MSS (which we call an internal ancillary) appears to exist but where another intuitively appealing ancillary is available. Another well known, and compelling continuous example, is the model of correlated unit normal variables in Basu (1964). Such examples appear to bring the sufficiency and conditionality principles into conflict since, if the sufficiency principle is first applied, then non-internal ancillaries are unavailable. Why the insistence that only ancillaries contained in the MSS be considered? Apart from the fact that any other ancillaries utilise aspects of the model besides the likelihood function, non-internal ancillaries include all the statistics which are independent of the MSS and consequently useless as conditioning variables. We will term such ancillaries external and will arbitrarily include constant ancillaries under this heading. Ancillaries which are neither a function of, not independent of, the MSS we may term neutral and it is these which, paradoxically, often appear to be appropriate conditioning variables.

The next lemma establishes the exclusiveness of the definitions of internal/external. Neutral ancillaries are clearly exclusive of the other two classes by definition.

Lemma. *Let A be an internal ancillary and denote by T the, possibly vector valued, residual of the minimal sufficient statistic S for θ. Then A and T are not independent.*

Proof. Suppose the minimal sufficient statistic can be written $S=(T, A)$. T may contain still more ancillary components. We may then write the likelihood function as

$$L(\theta)=f_1(t(x), a(x);\theta)f_2(x)$$

where $A(X)$ has θ-free distribution. Denoting the density of T, conditional on $A=a$, by g and the θ-free density of A by h, we may rewrite this as $L(\theta)=g(t \mid a;\theta)h(a)$. Then, if A is independent of T then $g(t \mid a;\theta)$ is free of "a" so that T alone is the MSS for θ contradicting the internal ancillarity of A.

The converse of this proposition, namely that if A and T are independent then A is not an internal ancillary, may be strengthened to a conclusion of external ancillarity under regularity conditions which exclude A from being not ancillary at all (See Basu (1959)).

Before we dismiss external ancillaries as having the same status as constant ancillaries it must be stressed that only in the present context are they "useless"; in the presence of nuisance parameters they can be of considerable value. Even in simple cases a set of external ancillaries can, when put together, yield the whole data (Basu (1959)). This is hardly what we would expect of an informationless statistic.

Perhaps the worst case of conflict between the conditionality and sufficiency principles we could imagine is if the sample size itself were non-internal (i.e. was neutral). This is indeed possible; consider the likelihood function in Example 1:

$$L(\theta;u,t,n)=\frac{1}{2ut}u^{n-1}e^{-u(\theta/t+t/\theta)}/(n-1)!^2.$$

Then the minimal sufficient statistic for θ is (U, T) and so N is not internal. In this case, however, we are saved by conditioning on U since this leaves us with a model entirely free of N. It turns out that N-sufficient ancillaries are always available in such situations.

Generally speaking, the densities that we initially work with in any particular problem are conditional on the sample sizes n. The case we have been considering is the one where, conditional on $\mathbf{N}$, considered random, $\mathbf{U}$ is an ancillary complement to T and that further conditioning on $\mathbf{U}$ removes $\mathbf{N}$ from the model. Letting

n-subscripts denote densities conditional on $\mathbf{N}=\mathbf{n}$ the joint density may be factorized as

$$dF(\theta, t, \mathbf{u}, \mathbf{n}) = dF^{\theta}_{\mathbf{n}}(t \mid \mathbf{u})\, dF_{\mathbf{n}}(u)\, dF(\mathbf{n})$$

with obvious notation. Thus

$$L(\theta) \alpha\, dF^{\theta}_{\mathbf{n}}(t \mid \mathbf{u})$$

and so $(T, \mathbf{U}, \mathbf{N})$ is, most generally at least, minimal sufficient. The case of $\mathbf{U}$ being N-sufficient (or the first term being free of n) corresponds exactly to the case of $L(\theta)$ being free of n. Thus U is N-sufficient iff $\mathbf{N}$ is not an internal ancillary (i.e. neutral). We thus have the pleasing result that when $\mathbf{N}$ is neutral, and the sufficiency principle does not allow us to condition on it, conditioning on the N-sufficient ancillary $\mathbf{U}$, which is internal, effectively eliminates $\mathbf{N}$ from the argument. We have in fact reconciled the model conditional on $\mathbf{N}=n$

$$f^{\theta}_{\mathbf{n}}(\mathbf{u}, t) = f^{\theta}(t \mid \mathbf{u}) f_{\mathbf{n}}(\mathbf{u})$$

with the marginal model, $f^{\theta}(\mathbf{u}, t)$, recommended by the sufficiency principle and obtained by integrating out $\mathbf{n}$. Further conditioning on $\mathbf{U}$ reduces both models to $f^{\theta}(t \mid \mathbf{u})$. In such cases, therefore, conditioning on $\mathbf{U}$ saves us from having to decide between these two basic principles of inference. Since the motivation for conditioning on ancillaries is to treat as fixed all aspects of the experiment not of direct relevance to θ, if conditioning on a particular statistic leaves the sample size free then we have failed in this aim. The above discussion ensures that only when the sample size is internal ancillary, and itself available for direct conditioning, do such ancillaries appear.

The above factorization of the likelihood also provides the simplest characterization of the N-sufficiency of an n-conditional ancillary $\mathbf{U}$. A necessary and sufficient condition is that the $\mathbf{n}$-conditional density of $(T, \mathbf{U})$ have the form

$$g(\theta, t, \mathbf{u}) h(\mathbf{n}, \mathbf{u})$$

for some functions g and h.

Finally note that, given $\mathbf{U}$, T is complete and $(\mathbf{U}, \mathbf{N})$ is jointly ancillary. It is shown (Becker & Gordon (1982)) that this implies $\mathbf{U}$ is a better conditioning variable than $\mathbf{N}$ according to the criterion of Cox (1970) and this may be generalised (Lloyd (1985)) to all choice algorithms satisfying certain natural axioms.

Acknowledgement

The author would like to thank Professor E. J. Williams for suggesting the topic as well as for numerous helpful discussions. This research was partially conducted under an A.R.G.S grant which is gratefully acknowledged.

References

BARTLETT, M. S. (1983). Examples of minimum variance estimation. *Austral. J. Statist.* **25**, 174–181.

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

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

BASU, D. (1959). The family of ancillaries. *Sankhya* **21**, 247–256.

BASU, D. (1964). Recovery of ancillary information. *Sankhya* **26**, 3–15.

COX, D. R. (1970). The choice between alternative ancillary statistics. *J. R. Statist. Soc. B* **33**, 251–255.

FISHER, R. A. (1925). The theory of estimation. *Proc. Cam. Phil. Soc* **22**, 700–725 or CP42.

FISHER, R. A. (1935). Logic of inductive inference. *J. R. Statist. Soc.* **98**, 39–54.

FISHER, R. A. (1956). Statistical Methods and Scientific Inference. London: Oliver & Boyd.

BECKER, N. & GORDON, I. (1983). On Cox's criterion for discriminating between alternative ancillary statistics. *Int. Stat. Rev.* **51**, 89–92.

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

KALBFLEISH, J. D. (1975). Sufficiency and conditionality. *Biometrika* **62**, 251–259.

LLOYD, C. J. (1985). Ancillary statistics in statistical inference. Unpublished Ph.D. thesis, University of Melbourne.

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