SOME EXAMPLES OF ANCILLARY STATISTICS

AND THEIR PROPERTIES

by

Robert J. Buehler

Technical Report No. 364

University of Minnesota

December, 1979

ABSTRACT

Two lists are given. The first is a list of properties of ancillary statistics; the second is a list of examples of ancillary statistics. It is then indicated which of the properties are satisfied by each example. Many of the models have the property that the parameter θ is a location parameter for the MLE $\hat{\theta}$ in every conditional distribution determined by a fixed value of the ancillary statistic. In certain other models the same state of affairs is achieved by parameter transformation. In these cases it is reasonable to say that the value of the ancillary statistic determines the precision of the MLE. There exist irregular models, however, for which this is not the case.

1. Introduction

According to conventional definition, an ancillary statistic is one whose distribution is the same for all values of an unknown parameter θ . According to conventional wisdom: There are difficulties with existence and uniqueness of ancillary statistics; the principle of conditionality requires us to make inferences conditional on an ancillary statistic when one exists; and an ancillary statistic by itself carries no information about θ but when used together with the maximum likelihood estimator $\hat{\theta}$, the ancillary tells us the precision of $\hat{\theta}$. In the present paper we study this conventional wisdom through examples.

The conditionality principle as stated for example by Cox and Hinkley (1974) says (paraphrasing slightly): When there is an ancillary statistic, the conclusion about the parameter of interest is to be drawn as if the ancillary statistic were fixed at its observed value. Presumably a confidence interval would be an example of a conclusion. In the present paper we restrict our attention to confidence intervals for a single parameter and in particular to what might be called natural confidence intervals. These are solutions which arise from the distribution of the maximum likelihood estimator in models satisfying the regularity conditions in Definition 2 below.

The general plan of the paper is first to give a list of properties of ancillary statistics followed by a list of examples of ancillary statistics. We then indicate which examples satisfy which properties. In this way the examples are classified into main categories. Some implications for statistical inference are discussed in the conclusions.

2. Definitions

The assumed probability law will be represented by a density function $f(x,\theta)$ where x may be a vector but θ is a scaler with range $\Omega = \{\theta | \theta_L < \theta < \theta_U\}$.

D1: We will say the model $\{f(x,\theta), \theta \in \Omega\}$ (or more briefly $\{f(x,\theta)\}$) is B-regular if a unique maximum likelihood estimate $\hat{\theta}(x)$ exists for each x and if the distribution of $\hat{\theta}$ satisfies Lindley's (1958) "Condition B": The CDF $F(\hat{\theta}|\theta)$ has a derivative $\partial F/\partial\theta$ which is always negative, and $\lim F(\hat{\theta}\}\theta) = 0$ (or 1) as θ tends to θ_U (or θ_L).

Except for examples EM2 and EM4 in Section 7, the present paper considers only B-regular models.

D2: If $A = a(X)$ is any conditioning statistics (typically an ancillary) we will say that the pair $\{f(x,\theta), a\}$ is B-regular if the CDF's $F(\hat{\theta}|a,\theta)$ satisfy Lindley's Condition B for all values of a.

D3: The model $\{f(x,\theta), a\}$ is called AB-regular if it is B-regular and satisfies property P2T (transformed translation invariance) defined in Section 3 below (see also discussion in Section 4).

D4: The Fisher information is X is $i(\theta) = E[u(X,\theta)]^2 = \mathrm{Var}\ u(X,\theta)$ where $u(x,\theta)$ is the score function $\partial \log f(x,\theta) | \partial\theta$.

D5: When a is ancillary, the conditional Fisher information is $i(\theta,a) = \mathrm{Var}\ [u(X,\theta)|a]$.

We will want to consider distributions on the parameter space which may or may not be legitimate fiducial distributions. For these we will use the neutral term "induced distribution."

D6: For the unconditional B-regular model $\{f(x,\theta)\}$ the induced distribution of θ has density $g(\theta|x) = -\partial F(\hat{\theta}|\theta)/\partial\theta$. For the conditional B-regular model $\{f(x,\theta), a\}$ the induced distribution of θ has density of $g_a(\theta|x) = -\partial F(\hat{\theta}|a,\theta)/\partial\theta$. Equivalently the CDF's are $G(\theta|x) = 1 - F(\hat{\theta}|\theta)$ and $G_a(\theta|x) = 1 - F(\hat{\theta}|a,\theta)$.

D7: The γ percentiles of the induced distributions of D6 will be denoted by $\theta_\gamma(\theta)$ $(F(\hat{\theta}|\theta_\gamma(\hat{\theta})) = 1 - \gamma)$ and $\theta_\gamma(\hat{\theta},a)$ $(F(\hat{\theta}|a,\theta_\gamma(\theta,a)) = 1 - \gamma)$.

Thus $\theta_\gamma(\hat{\theta})$ and $\theta_\gamma(\hat{\theta},a)$ are respectively unconditional and conditional upper confidence limits for θ.

3. Properties of Ancillary Statistics

In this section we list properties which may or may not be satisfied by B-regular models $\{f(x,\theta), a(x)\}$.

P1: $(\hat{\theta},a)$ is minimal sufficient.

P2: Translation invariance. $(\theta_L,\theta_U) - (-\infty , \infty)$, and for all a and all $-\infty < c < \infty$, $F(\hat{\theta} + c|a,\theta+c) = F(\hat{\theta}|a,\theta)$.

P3: $Var(\hat{\theta}|a,\theta)$ depends on a but not on θ .

P4: For two values a_1, a_2 and any two values θ_1, θ_2. $Var(\hat{\theta}|a_1,\theta_1) > Var(\hat{\theta}|a_2,\theta_1)$ implies $Var(\hat{\theta}|a_1,\theta_2) > Var(\hat{\theta}|a_2,\theta_2)$.

P5: $i(\theta,a)$ depends on a but not on θ .

P6: For any fixed γ, and fixed a, the sign of $\theta_\gamma(\hat{\theta},a) - \theta_\gamma(\hat{\theta})$ is the same for al $\hat{\theta}$.

P7: $P\{\theta \leq \theta_\gamma(\hat{\theta})|a,\theta\}$ depends on a but not on θ.

P8: There exists an improper prior $\pi(\theta)$ such that for all a, the induced density $g_a(\theta|x)$ equals the posterior density.

PkT $(k = 2,3,\ldots,8)$: There exists a transformation $\tau(\theta)$ (the same for all a) such that Pk holds with τ substituted for θ .

4. Discussion of Properties

P1 is Fisher's requirement for obtaining the fiducial distribution of θ by a conditional pivotal argument.

P2 could also be expressed as $F(\hat{\theta}|a,\theta) = F_o(\hat{\theta} - \theta|a)$ for some $F_o(\cdot|\cdot)$. P2T states that the distribtuion $F(\hat{\theta}|a,\Theta)$ satisfies Lindley's (1958) Condition A for all a with the transformation $\tau(\theta)$ the same for all a.

P3 and P4 attempt to formalize the statement that the ancillary statistic determines the precision of $\hat{\theta}$.

P6 and P7, when they hold, show us how to determine relevant reference sets for the natural unconditional confidence limits, in the sense of Buehler (1959).

5. Implications among the Properties

Proposition 1: The following implications hold:

(i) Pk $\Rightarrow$ Pkt (all k).

(ii) P1T $\Rightarrow$ P1, P6T $\Rightarrow$ P6, P7T $\Rightarrow$ P7, P8T $\Rightarrow$ P8.

(iii) P2 $\Rightarrow$ Pk and P2T $\Rightarrow$ PkT for k = 3,4,5,6,7,8.

(iv) P3 $\Rightarrow$ P4

(v) P2T $\Rightarrow$ P5T

(vi) P8 $\Rightarrow$ P2T

(i) is trivial.

(ii) holds because P1,6,7,8 are invariant under parameter transformation.

(iii), which is straightforward to prove, states that the invariance condition P2 implies all the following ones.

(iv) and (v) are evident.

(vi) can be proved by Lindley's (1958) argument.

6. Notation for Distributions

$N(\mu,\sigma^2)$ denotes normal with mean μ, variance σ^2.

$G(\alpha,p)$ denotes a gamma distribution with density $\alpha^p e^{-\alpha x} x^{p-1}/\Gamma(p)$.

$\mathrm{Exp}(\theta)$ denotes an expontential distribution with density $\theta e^{-\theta x}$ $\theta e^{-\theta x}$ $(x > 0,\ \theta > 0)$. $\mathrm{Exp}(\theta) = G(\theta,1)$.

$\mathrm{Lind}(\theta)$ denotes what we will call a Lindley (1958) distribution with density $f(x,\theta) = \theta^2(\theta+1)^{-1}(x+1)e^{-\theta x}$ $(x > 0,\ \theta > 0)$ (see Appendix E).

7. Examples of Ancillary Statistics

The following notation is convenient. E = example, L = location parameter model, S = scale parameter model, I = irregular model, M = miscellaneous model, g = generalized, n = sample of size n. Thus ES1gn denotes scale parameter example number one, generalized, with sample size n.

EL1: Two measuring instruments (Cox, 1958). $P(A = o) = P(A = 1) = \frac{1}{2}$; $X|a \sim N(\theta,\sigma_a^2)$, $\hat{\theta} = x$.

EL1n: Sample $(x_1,a_1),\ldots,(x_n,a_n)$ from EL1. $\hat{\theta} = (\Sigma\, \sigma_{a_j}^{-2} x_j)(\Sigma\, \sigma_{a_j}^{-2})^{-1}$, (Efron and Hinkley, 1978).

EL1g: In EL1 replace normal densities by two arbitrary location models $f(x|A = o,\theta) = f_o(x - \theta)$, $f(x|A = 1,\theta) = f_1(x - \theta)$.

EL1gn: n observations from EL1g.

EL2: Fisher-Pitman location model. $f(\underset{\sim}{x};\theta) = \prod_{i=1}^{n} f(x_i-\theta)$. $a = (x_{(1)} - x_{(2)}, x_{(1)} - x_{(3)},\ldots, x_{(1)} - x_{(n)})$, the spacings of the ordered observations $x_{(i)}$.

EL2g: Location model assuming neither independence nor identical distributions. $f(\underset{\sim}{x}_i\theta) = f(x_1 - \theta,\ldots, x_n - \theta)$. $a = (x_1 - x_2,\ldots, x_{n-1} - x_n)$, spacings of unordered observations.

EL3: One-parameter normal regression. $A \sim N(o,1)$, $X|a \sim N(\theta a, 1)$. Observe $(a_1, x_1),\ldots,(a_n, x_n)$, $\hat{\theta} = \Sigma\, a_i x_i/\Sigma\, a_i^2$, $\{\hat{\theta}|(a_1,\ldots, a_n)\} \sim N(\theta,(\Sigma\, a_i^2)^{-1})$.

EL4: Sprott's (1961) ancillary. $X_1 \sim N(n\theta,n)$, $X_2 \sim G(m, ce^{k\theta})$, $a = x_1/n - (\log x_2)/k$.

ES1: Two-valued ancillary with reciprocal expoentials.

$P(A = 0) = P(A = 1) = \frac{1}{2}$. $X|0 \sim \text{Exp}(\theta)$, $X|1 \sim \text{Exp}(\theta^{-1})$.

ES1n: n observations from ES1.

ES1g: $f(x|A = 0, \theta) = \theta f(\theta x)$, $f(x|A = 1, \theta) = \theta^{-1} f(x/\theta)$.

ES2: Fisher-Pitman scale model. $f(x;\theta) = \Theta^{-n} \prod_{i=1}^{n} f(x_i/\theta)$, $\Theta > 0$, $x_i > o$, a = quotients of ordered observations.

ES2g: $f(x,\theta) = \theta^{-n} f(x_1/\theta,\ldots, x_n/\theta)$, a = quotients of x_i's (unordered).

ES3: Fisher's gamma hyperbola (Fisher, 1973, p. 169; Efron and Hinkley, 1978, example 3.2). $X_1 \sim \text{Exp}(\theta)$, $X_2 \sim \text{Exp}(\theta^{-1})$. Observe n pairs X_1, X_2. If $S_j = \sum_{i=1}^{n} X_{ji}$, $j = 1, 2$, then $\hat{\theta} = \frac{1}{2} \log (s_1/S_2)$ and $A = S_1 S_2$.

ES3g: Sample of size n from $f(x_1, x_2; \theta) = f(\theta x_1, x_2/\theta)$, $\theta > 0$, where $f(z_1, z_2)$ is a density on $0 < z_1 < \infty$, $0 < z_2 < \infty$. The ancillary statistic can be represented by the n product $x_{1i} X_{2i}$ together with $n - 1$ quotients of the ordered x_1's: $x_{(1)}/x_{(2)},\ldots, x_{(n-1)}/x_{(n)}$.

ES4: Normal with known coefficient of variation (Hinkley, 1977). $X \sim N(\theta, c^2 \theta^2)$ $(\theta > 0$, c known). $a = S_2^{\frac{1}{2}}/S_1$, $S_k = n^{-1} \Sigma x_j^k$, $\hat{\theta} = \frac{1}{2} S_1\{(1 + 4a^2)^{\frac{1}{2}} - 1\}$.

ES4g: Arbitrary shape with known coefficient of variation. $x = \theta y$ where y has density $g(y)$ for $-\infty < y < \infty$. Then $f(x) = \theta^{-1} g(x/\theta)$. The ancillary statistic gives the number of negative observations and the quotients of ordered positive and negative observations separately.

EI1: $P(A = 0) = P(A = 1) = \frac{1}{2}$. $X|0 \sim N(\theta,1)$ $X|1 \sim N(\theta^3,1)$.

EI2: $P(A = 0) = P(A = 1) = \frac{1}{2}$ $X|0 \sim \text{Exp}(\theta)$, $X|1 \sim \text{Lind}(\theta)$.

EM1: Fisher's normal circle (Fisher, 1973, p. 138, Efron and Hinkley, 1978, p. 464). $X_1 - \rho \cos \theta$, $X_2 - \rho \sin \theta$ are independent $N(0,1)$, where ρ is known. The ancillary is $a = (x_1^2 + x_2^2)^{\frac{1}{2}}$.

EM2: One observation (x_1, x_2) from a bivariate normal distribution with zero means, unit variances and correlation θ. $a = x_1$.

EM2n: Sample of size n from EM2.

EM3: (Basu, 1959) Two observations (x_1, x_2) from $N(\theta,1)$. $a = x_1 - x_2$ if $x_1 + x_2 < c$ and $a = x_2 - x_1$ if $x_1 + x_2 \geq c$.

EM4: (Basu, 1964, Cox 1971). Multinomial distribution with four cells whose frequencies are $X_1, \ldots, X_4$ and whose probabilities are $p_1 = (1 - \theta)/6$, $p_2 = (1 + \theta)/6$, $p_3 = (2 - \theta)/6$, $p_4 = (2 + \theta)/6$, where $-1 \leq \theta \leq 1$. $a = x_1 + x_2$ and $b = x_1 + x_4$ are separately but not jointly ancillary.

8. Properties of the Examples

We give two general results before discussing the models individually.

Proposition 2: Subject to B-regularity (see D1, Section 2), all of the EL models satisfy property Pk for $k = 2, 3, \ldots, 8$.

Proof. Verfication of P2 is reasonably straightforward in each case, and the rest follows from Proposition 1(iii).

Proposition 3: Subject to B-regularity, all of the ES models satisfy PkT for $k = 2, 3, \ldots, 8$.

Proof. The transformation $\tau = \log \theta$ reduces each ES model to a location model.

8.1. Location Models

EL1 is occasionally put forward in support of the principle of conditionality -- if we know which of two measuring instruments was used, our inference about θ should be conditional on this information.

EL1n has been discussed by Efron and Hinkley as an example of combining information and determination of the relevant conditional variability of the combined estimator.

In EL1g the induced density (see D6, Section 2) is $g_a(\theta|x) = f_a(x - \theta)$.

In EL1gn suppose $a = (a_1, \ldots, a_n)$ contains r zeros and $s = n - r$ ones. For fixed r we have r observations from location model f_o and s from model f_1. This falls within the generalized Fisher-Pitman model (Appendix A). The induced distribution is a posterior distribution for a uniform prior conditionally for each fixed r and hence also unconditionally.

EL2 and EL2g are discussed in Appendix A.

In EL3 the distribution of A need not be normal. The conditional aspects of the more usual two-parameter model have been discussed by Fisher (1973), pp. 86-89.

Sprott's example, EL4, falls within the general location model theory as indicated in Appendix D.

8.2. Scale Models

ES1 is clearly a variant of Cox's example EL1. Unconditionally x has the mixture distribution $f(x;\theta) = \frac{1}{2}(\theta e^{-\theta x} + \theta^{-1}e^{-x/\theta})$, but this distribution should not be used, even by disbelievers in the conditionally principle, as it is not the most natural procedure. One reason is that $f(x;\theta)$ is not B-regular. But more basically we want the analog of the procedure used in more subtle models where the ancillary may be hard to recognize. In such cases the natural procedure is to find first the MLE $\hat{\theta}$ and then its distribution. Thus we have $\hat{\theta} = x^{-1}$ if $A = 0$ and $\hat{\theta} = x$ if $A = 1$, and the MLE has unconditional CDF $F(\hat{\theta}|\theta) = \frac{1}{2}(1 - e^{-\hat{\theta}/\theta} + e^{-\theta/\hat{\theta}})$, which is seen to be a scale model. An alternative analysis leading to the same result, would consist in transforming first to a location model, as in the following paragraph. The situation here differs from that in Cox's example, EL1, in an interesting way. In EL1 the induced distributions for θ for $a = 0, 1$ ($\hat{\theta}$ fixed) differ in variance but not in mean. In EL2 the induced distributions for $\log \theta$ for $a = 0, 1$ ($\hat{\theta}$ fixed) are stochastically ordered (because f_o and f_1 defined in the next paragraph are). Thus lower and upper confidence limits shift <u>in the same direction</u> as a varies (and the unconditional limits of course take intermediate values).

To analyze ES1, first tranform by $\tau = \log \theta$, $u = -\log x$ if $A = 0$ and $u = \log x$ if $A = 1$. This reduces the problem to a location model already considered in EL1gn, with $f_o(y) = \exp(-y - e^{-y})$ and $f_1(y) = \exp\{y - e^{y}\}$.

Similarly ES2 and ES2g are transformed into EL2 and EL2g by taking logs.

ES3 and ES3g are discussed in Appendix B.

ES4 and ES4g are discussed in Appendix C.

8.3. Irregular Models

EI1 and EI2 are deliberately pathological counterexamples. In EI1, different functions of θ serve as location parameter depending on the value of a (compare the definition of property PkT in Section 3. This suffices to violate all the properties in Section 3 except P1. In EI2 we mix a scale model, Exp (θ), with the Lindley model Lind(θ) which is known not to transform to a location or scale model. Some details of the analysis are given in Appendix E. On a log-log plot the contours of $F(\hat{\theta}|A = 0, \theta)$ $(A = 0$ gives Exp $(\theta))$ are parellel lines having unit slope. For $F(\hat{\theta}|A = 1, \theta)$ (Lind (θ) case) the countours are curvilinear, concave upward. The crossing of the straight and curvilinear contours violates properties like P6.

8.4. Miscellaneous Models

EM1 fails to fall in the EL (location) category only because θ defines points on a circle rather than on $(-\infty, \infty)$. EM1 does exhibit all the desirable properties of the EL models suitably restated for the circle.

EM2 is a standard example in which x_1, x_2 are separately but not jointly ancillary. The ancillary x_1 (or x_2) is of little help for inference for two reasons: (i) $(\hat{\theta}, x_1)$ is not sufficient, so that P1 is violated, and (ii) $\{f(x,\theta), x_1\}$ is not B-regular.

EM2n has been considered by Efron and Hinkley (1978), Section 6. It is not known whether there exists an ancillary a such that $(\hat{\theta},a)$ is sufficient.

EM3 is of interest in exhibiting nonuniqueness of ancillaries, but in fact it has little implication for inference. The MLE $\hat{\theta}$ alone is sufficient so that the conditional induced distribution would not differ from the unconditional one for any ancillary a (any value of c).

EM4 falls outside the primary framework of this paper because the distribution of $X_1, \ldots, X_4$ is discrete, so that we cannot obtain an induced distribution by a pivotal argument, except perhaps in some large-sample approximation. A second difficulty however is that $(\hat{\theta},A)$ is not sufficient for general $n = \Sigma X_i$, so that P1 is violated. (Cox (1971) points out that A is a component of a minimal sufficient statistic, but does not consider the sufficiency of $(\hat{\theta},A)$). To see this take $n = 4$. Then $(X_1, \ldots, X_4) = (0,0,3,1)$ and $(0,0,4,0)$ both give $(\hat{\theta},A) = (-1,0)$ but have different likelihoods. For $n = 1$, $(\hat{\theta},A)$ is minimal sufficient, as Basu (1964) pointed out.

9. Discussion and conclusions

In estimating θ by $\hat{\theta}$ (the MLE, or for that matter any other estimator) it is almost universally considered desirable to give not only the value of $\hat{\theta}$ but also some indication of its precision. Fiducial theory and the theory of confidence intervals are two attempts to accomplish this. Should the estimated precision be conditional or unconditional? If conditional, on what should one condition? No satisfactory general definition has been given the relevant reference set, the set on which one ought to condition.

The principle of conditionality tells us to condition on an ancillary statistic when one is available. It has long been recognized that the principle is less than satisfactory because of problems of existence and uniqueness. It is not known how to determine whether ancillary statistics exist (the "problem of the Nile"), and when an ancillary exists it may not be unique. Thus the principle of conditionality fails to determine a unique reference set.

The defining property of an ancillary statistic is that its distribution be free of θ. But in conditional inference it is really our hope that the ancillary statistic determines in some sense the precision of the MLE $\hat{\theta}$ (or other estimator). In the present paper we have attempted to study through examples the sense in which this is the case. The examples are found to fall in two broad categories which might be called regular and irregular. In the regular models either θ is a location parameter for $\hat{\theta}$ in every conditional distribution (location, or EL, models), or there is a transformation

$\tau(\theta)$ such that the same relationship holds for τ and $\hat{\tau}$ (primarily scale, or ES, models). These regular models are ones for which fiducial theory is considered to apply, and for which the induced (fiducial) distributions are posterior distributions for an appropriate improper prior. For these regular models there is a clear sense in which the ancillary statistic determines the precision of estimation -- namely translation invariance of the conditional distribution:
$f(\hat{\tau}|a, \tau) = f_a(\hat{\tau} - \tau)$.

The second category contains artificially constructed ancillaries which are intended to show that not every ancillary need relate in a simple way to the precision of $\hat{\theta}$. The conditionality principle is perhaps less compelling in the irregular cases than in the regular ones.

Many questions remain unanswered. If new ancillaries are discovered, will they be regular or irregular? How should approximate ancillarity be defined? Can we have an approximate ancillary which is approximately regular? If so, what is a suitable principle of approximate conditionality? It is hoped that the properties studied in this paper may eventually be brought to bear on these questions.

Appendix A

Generalized Fisher-Pitman Fiducial Distributions

The Fisher-Pitman theory of location and scale models applies to random samples from location models, scale models, and joint location-scale models having respectively the densities $f(x-\theta)$, $\sigma^{-1}f(x/\sigma)$ and $\sigma^{-1}f((x-\theta)/\sigma)$. There is no difficulty in generalizing the main results of this theory to the case of nonidentically distributed observations (as with likelihood $\Pi f_i(x_i-\theta)$ replacing $\Pi f(x_i-\theta)$) and moreover to dependent observation, as with likelihood $f(x_1-\theta,\ldots, x_n-\theta)$. Indeed these cases fall within the scope of the invariant models considered by Fraser (1961 a,b) and Hora and Buehler (1966). The principal results needed for our present purposes are that the fiducial distribution is a posterior distribution corresponding to prior measure equal to right Haar measure ($d\theta$, $d\sigma/\sigma$ and $d\theta d\sigma/\sigma$ in the three cases cited) and the fiducial limits are confidence limits obtained from a pivotal quantity conditional on an appropriate ancillary statistic. For the model $f(x_1-\theta,\ldots, x_n-\theta)$, $a = (x_1-x_2, x_1- x_2,\ldots, x_1-x_n)$ is an appropriate ancillary. For the model $\Pi f(x_i-\theta)$ the order statistic $(x_{(1)},\ldots x_{(n)})$ is sufficient and it is possible to make a sufficiency reduction of a to $a^* = (x_{(1)}-x_{(2)},\ldots, x_{(1)}-x_{(n)})$, but this is not essential since we get the same induced distribution either way. Similar considerations apply to intermediate models such as

$$\text{(A.1)} \qquad \prod_{i=1}^{r} f_o(x_i-\theta) \prod_{i=r+1}^{n} f_1(x_i-\theta).$$

Appendix B

Fisher's Gamma Hyperbola and Generalizations

Fisher (1973) considers the joint density $f(x,y;\theta) = \exp(-\theta x - y/\theta)$. Efron and Hinkley (1978) call this "Fisher's Gamma Hyperbola." Under the transformation $\tau = \log \theta$, $u = -\log x$, $v = \log y$ we find

$$\text{(B.1)} \qquad g(u,v;\tau) = g_1(u-\tau)\ g_2(v-\tau)$$

where $g_1(y) = \exp\{-y-e^{-y}\}$ and $g_2(y) = \exp\{y-e^{y}\}$. This falls within the scope of generalized Fisher-Pitman models (Appendix A). For one bivariate observation (u,v) the ancillary is $u-v = -\log (xy)$, or equivalently xy. For n observations the statistic $(u_1-u_2,\ldots, u_{n-1}-u_n, u_n-v_1, v_1-v_2,\ldots, v_{n-1}-v_n)$ is ancillary, but a minimal sufficient reduction brings this down to $(\Sigma x_i)(\Sigma y_i)$.

The generalization ES3g assumes $f(x,y;\theta) = f(\theta x, y/\theta)$ where $f(\cdot,\cdot)$ is any suitably regular bivariate density on the first quadrant. Then with the same transformation the joint density of (u,v) is

$$\text{(b.2)} \qquad g(u,v;\tau) = e^{v-u} f(e^{\tau-u},e^{v-\tau}) = e^{z}e^{-y} f(e^{-y},e^{z}),\ y = u-\tau,\ z = v-\tau.$$

with the previously mentioned ancillary, which would not in general be reducible.

A second generalization, mentioned by Fisher (1973), p. 175, (but omitted in Section 7 above) takes $f(x,y;\theta) = \theta\phi e^{-\theta x-\phi y}$ with $\phi = \theta^{s}$. This reduces to a location model under the transformation $\tau = \log \theta$, $u = -\log x$ and $v = -(1/s) \log y$. Evidently this model itself generalizes as above to $\theta^{s+1} f(\theta x, \theta^{s} y)$.

Appendix C

Inference With Known Coefficient of Variation

Hinkley (1977) has considered inference about μ when the parent population is $N(\mu, c^2\mu^2)$ where the coefficient of variation c is known. If we assume with Hinkley that $\mu > 0$ then the density and c.d.f. are

(C.1) $$f(x;\mu) = (1/c\mu)\ \phi\ [(x-\mu)/c\mu]$$

and

(C.2) $$F(x;\mu) = \Phi[(x-\mu)/c\mu] = G(x/\mu)$$

where ϕ and Φ are the standard normal density and c.d.f. and where

(C.3) $$G(\lambda) = \Phi(\frac{\lambda}{c} - \frac{1}{c}).$$

Let us consider a generalization in which φ, Φ, G are replaced by ψ, Ψ, H, where ψ is an arbitrary density with support $(-\infty, \infty)$. If X has density (1) with ψ substituted for ϕ then

$$P(X \leq 0) = \Psi(-1/c) = H(0) = q, \text{ say.}$$

Thus the indicator function $I(x)$ which equals 1 for $x \leq 0$, 0 for $x > 0$ is an ancillary statistic. For n observations the corresponding indicators $I_1, \ldots, I_n$ are jointly ancillary.

Returning to one observation, given that $x > 0$ the c.d.f. of x is

$$F_+(x;\mu) = \frac{1}{p}(F(x;\mu) - q)$$

where $p - 1-q$, from which we get the fiducial density

$$\text{(C.5)} \qquad \varphi(\mu|x) = -\frac{\partial}{\partial\mu} F(x;\mu) = \frac{x}{pc\mu^2}\,\psi\left(\frac{x}{c\mu} - \frac{1}{c}\right) = \frac{x}{p\mu}\, f(x;\mu).$$

Similarly, given that $x < 0$, the fiducial density is

$$\text{(C.6)} \qquad \varphi(\mu|x) = -\frac{x}{p\mu}\, f(x;\mu).$$

The last two expressions agree with a posterior corresponding to the prior $d\mu/\mu$.

Next consider two observations, $x_1\ x_2$. If both are positive we have two independent observations from F_+, that is two observations from a scale family, and it is known from the Fisher-Pitman theory that the fiducial distribution obtained by conditioning on the ancillary x_1/x_2 equals the posterior for prior $d\mu/\mu$. A similar argument applies if both are negative. If $> x_1\ 0$ and $x_2<0$ then x_1/x_2 is again ancillary and we again get a posterior corresponding to the same prior.

Finally suppose that of n observations, r are negative and $s = n-r$ are positive. Without loss of generality we may suppose the first r are negative. If $y_1 = \log|x_i|$, $\tau = \log\mu$, the joint density of the y's has the form

$$\text{(C.7)} \qquad \prod_{i=1}^{r} h_1(y_i-\tau) \prod_{i=r+1}^{n} h_2(y_i-\tau)$$

a generalized location model which falls within the generalized

Fisher-Pitman theory described in Appendix A. The spacings $(y_1-y_2, y_2-y_3, \ldots, y_r-y_{r+1}, \ldots, y_{n-1}-y_n)$ are ancillary, and the fiducial distribution of τ is a posterior corresponding to a uniform prior, the separate cases need not be distinguished in stating that the fiducial distribution of μ in simply the posterior distribution corresponding to prior $d\mu/\mu$.

Appendix D

Sprott's Ancillary

In Sprott's (1961) example, $X_1 \sim N(n\theta,n)$ and $X_2 \sim \Gamma(m,ce^{k\theta})$. Thus $X_1/n \sim N(\theta,1)$ and $Y_2 \equiv ce^{k\theta} X_2 \sim \Gamma(m,1)$. We get $\log Y_2 = \log c + k(\theta + (1/k) \log X_2)$. Since the distribution of Y_2 is free of θ we see that θ is a location parameter for $Z_2 \equiv (1/k) \log X_2$. But θ is also a location parameter for $Z_1 \equiv X_1/n$. Therefore by the location parameter theory of Appendix A, an ancillary statistic is $Z_1 - Z_2 = X_1/n-(1/k) \log X_2$ (as Sprott showed by a different argument).

Appendix E

The Lindley Distribution

The Lindley distribution was originally presented (Lindley, 1958) as an example satisfying Condition B (B-regularity; see D1 in Section 2) but not Condition A (see remark at the end of Section 3).

We write $X \sim \text{Lind}(\theta)$ if X has density

$$\text{(E.1)} \qquad f(x;\theta) = \theta^2(\theta+1)^{-1}(x+1)e^{-\theta x} \qquad x > 0,\ 0 > 0.$$

The Lindley distribution is incidentally a mixture of two gamma distribtuions with weights depending on θ:

$$\text{(E.2)} \qquad \text{Lind}(\theta) = \frac{\theta}{\theta+1} G(\theta,1) + \frac{1}{\theta+1} G(\theta,2).$$

The CDF is

$$\text{(E.3)} \qquad F(x,\theta) = 1 - e^{-\theta x}\left[1 + \theta x/(\theta+1)\right].$$

Given one observation x, the MLE $\hat{\theta}$ is the value of θ satisfying

$$\text{(E.4)} \qquad x = \psi(\theta) = \frac{2}{\theta} - \frac{1}{\theta+1} = \frac{\theta+2}{\theta(\theta+1)}$$

Thus the MLE has CDF

$$\text{(E.5)} \qquad \begin{aligned} P\{\hat{\theta} \leq u\} &= P\{\psi(\hat{\theta}) \geq \psi(u)\} \\ &= P\{X \geq \psi(u)\} \\ &= 1 - P\{X \leq \psi(u)\} \\ &= \{\exp(-\theta\psi(u))\}\{1 + \theta\psi(u)/(\theta+1)\}. \end{aligned}$$

With θ as abscissa and u as ordinate, vertical sections (θ fixed) of this last function give values of the CDF of $\hat{\theta}$, horizontal sections (u fixed) give one minus the induced CDF of θ, and thus conditional confidence limits for model EI1 of Section 7 when $A = 1$.

REFERENCES

[1] Basu, D. (1959). The family of ancillary statistics. Sankhya, 21, 247-256.

[2] Basu, D. (1964). Recovery of ancillary information. Sankhya Series A, 26, 3-16.

[3] Buehler, Robert J. (1959). Some validity criteria for statistical inferences. Ann. Math. Statist., 30, 845-863.

[4] Cox, D. R. (1958). Some problems connected with statistical inference. Ann. Math. Statist., 29, 357-372.

[5] Cox, D. R. (1971). The choice between ancillary statistics. J. Roy. Statist., Soc. Ser. B, 33, 251-255.

[6] Cox, D. R. and Hinkley, D. V. (1974). Theoretical Statistics. London: Chapman and Hall.

[7] Efron, Bradley and Hinkley, David V. (1978). Assessing the accuracy of the maximum likelihood estimates: Observed versus expected Fisher information. Biometrika, 65, 457-487.

[8] Fisher, Sir Ronald A. (1973). Statistical methods and scientific inference (third edition). New York: Hafner.

[9] Fraser, D. A. S. (1961a). On fiducial inference. Ann. Math. Statist., 32, 661-676.

[10] Fraser, D. A. S. (1961b). The fiducial method and invariance. Biometrika, 48, 261-280.

[11] Hinkley, D. V. (1977). Conditional inference about a normal mean with known coefficient of variation. Biometrika, 64, 105-108.

[12] Hora, R. B. and Buehler, R. J. (1966). Fiducial theory and invariant estimation. Ann. Math. Statist., 37, 643-656.

[13] Lindley, D. V. (1958). Fiducial distributions and Bayes' theorem. J. Roy. Statist., Ser. B, 20, 102-197.

[14] Sprott, D. A. (1961). An example of an ancillary statistic and the combination of two samples by Bayes' theorem. Ann. Math. Statist., 32, 616-618.