The Annals of Statistics
1983, Vol. 11, No. 1. 104–113

THE GENERALISED PROBLEM OF THE NILE: ROBUST CONFIDENCE SETS FOR PARAMETRIC FUNCTIONS

BY G. A. BARNARD AND D. A. SPROTT

University of Waterloo

The pivotal model is described and applied to the estimation of parametric functions $\phi(\theta)$. This leads to equations of the form $H(x; \theta) = G\{p(x, \theta)\}$. These can be solved directly or by the use of differential equations. Examples include various parametric functions $\phi(\theta, \sigma)$ in a general location-scale distribution $f(p)$, $p = (x - \theta)/\sigma$ and in two location-scale distributions. The latter case includes the ratio of the two scale parameters σ_1/σ_2, the difference and ratio of the two location parameters $\theta_1 - \theta_2$ and the common location θ when $\theta_1 = \theta_2 = \theta$. The use of the resulting pivotals to make inferences is discussed along with their relation to examples of non-uniqueness occurring in the literature.

1. Introduction. The "Problem of the Nile" was formulated by Fisher (1936) in the context of a "likelihood model" for parametric inference. For such a model the inferential structure is fully specified by three elements $\{S, \Omega, f\}$, the sample space $S = \{x\}$, the parameter space $\Omega = \{\theta\}$, and the probability function f: $Sx\,\Omega \to R$. The Problem of the Nile then consisted in finding a "maximal ancillary", a function $a(x)$ whose probability distribution does not involve θ, maximal in the sense that any other such function $a^*(x)$ is expressible as a function of $a(x)$. The importance of the problem lies in the fact that inferences about θ are then based on the conditional distribution of the observations given $a(x)$ (Fisher, 1934). The widespread notion that such conditional inferences are inefficient was shown by Barnard (1976) to be based on a misapprehension.

Difficulties arose, however, concerning the possible absence of maximal ancillaries—see e.g. Basu (1964). Such difficulties are, however, overcome by the use of more structured models, such as those of Fraser (1979), the "functional model" of H. Bunke (1975) and O. Bunke (1976)(see also Dawid and Stone, 1982), and the "pivotal model" of Barnard (1977). Here we use the pivotal model to present a more constructive formulation of the Nile Problem, and a generalisation of it.

The notion of a pivotal quantity was introduced by Fisher (1945) as a function $p(x, \theta)$ of the observations and parameters whose distribution does not involve the unknown parameters. The pivotal model as described in some detail by Barnard (1977) has five elements, $\{S, \Omega, P, p, D\}$. The S and Ω are as with the likelihood model, $\mathbb{P}$ is a (measure) space of values of the function p: $Sx\Omega \to \mathbb{P}$, and D is a set of (completely specified) densities on $\mathbb{P}$, these densities usually being thought of as close to each other in some sense, for example all being "approximately standard normal." The function p is called the basic pivotal. Adopting a pivotal model for a given experimental situation amounts to asserting that the function p has a distribution specified by one of the densities in D. It was shown by Segal (1938) that any likelihood model involving continuous distributions can (by an extension of the probability integral transformation) be reformulated as a pivotal model. What is special about Barnard's pivotal model is that the set of densities D is supposed to contain at least a small neighbourhood of an exactly specified density, just as an observation "x", being known only to finite precision, really denotes a small neighbourhood of x.

Received February 1982; revised August 1982

AMS 1980 *subject classification.* Primary 62A99; secondary 62F35.

Key words and phrases. Ancillary statistics; conditional inferences; confidence intervals for parametric functions; pivotal quantities; robust.

It was shown in Barnard (1977) that the small amount of uncertainty concerning the exact distribution allowed by the set D is sufficient to ensure that all other pivotals $H(x, \theta)$ must be expressible as functions of the basic pivotal p:

$$H(x, \theta) = G\{p(x, \theta)\}. \tag{1}$$

Functions of the basic pivotal will have (approximately) known distributions, and the only quantities having (approximately) known distributions are functions of the basic pivotal. Thus the situation considered here differs markedly from that considered by Owen (1948). Owen starts with a likelihood model, with the density functions f supposed exactly known; as is natural in such a case, Owen considers f to belong to the exponential family, and in particular the case of a scalar parameter θ for which two functions $U(x)$, $V(x)$ are jointly sufficient. He attempts to find a pair of pivotals $p_1(U, V, \theta)$ and $p_2(U, V)$, uniquely related to the pivotals derivable from U and V by Segal's argument, such that p_2, being constant on the parameter space, can be used to condition the distribution of p_1. His analysis is somewhat related to ours and, in the case of the normal distribution with θ a function of the location and scale parameters, his results parallel some of ours. However, our results for location and scale place no restrictions on the form of the distributions; and, as not infrequently happens, this generalisation helps to overcome some of the difficulties met with in Owen's treatment. For example, in our case, as shown in Barnard (1977), the existence and uniqueness of the maximal ancillary can be guaranteed; it is the maximal solution to the functional equation

$$A(x) = G\{p(x, \theta)\}. \tag{2}$$

We adopt Birnbaum's Conditionality Principle (C), basing inferences about θ on conditional distributions given the observed value $A(x_0)$ of $A(x)$; and because the pivotal model fails to satisfy Birnbaum's Mathematical Principle (M) (Birnbaum, 1972) we escape the argument leading to the Likelihood Principle (L) with all the difficulties that L entails.

In our approach, separate inferences about a function $\phi(\theta)$ are possible iff a suitable pivotal can be found such that

$$H\{x, \phi(\theta)\} = G\{p(x, \theta)\}. \tag{3a}$$

What is required to make such an H "suitable" is discussed below.

If in (3a) we take ϕ to be a trivial function, taking a single constant value, we obtain (2). We therefore call (3a), and a slightly more general version (3b) to be given later, the generalised Problem of the Nile. The rest of this paper is devoted to examination of various situations that can arise in attempting to solve (2) or (3) for a specified $\phi(\theta)$.

2. Direct method. We illustrate with the case $S = \{\mathbf{x}\} = \mathbb{R}^n, \Omega = \{\boldsymbol{\lambda}\} = \mathbb{R}^n, P = \{\mathbf{p}\} = \mathbb{R}^n, p_i = x_i - \lambda_i (i = 1, 2, \cdots, n)$, and D any subset of the set of all densities on $\mathbb{R}^n$. If we take $\phi(\lambda) = \boldsymbol{\lambda}'\boldsymbol{\lambda}$, (3) requires

$$H(\mathbf{x}, \boldsymbol{\lambda}'\boldsymbol{\lambda}) = G(\mathbf{x} - \boldsymbol{\lambda}).$$

Thus, given any two pairs $(\mathbf{x}_1, \boldsymbol{\lambda}_1)$, $(\mathbf{x}_2, \boldsymbol{\lambda}_2)$, and setting $\mathbf{x}_2 - \mathbf{x}_1 = \mathbf{a} + \mathbf{b}$, $-(\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2) = \mathbf{a} - \mathbf{b}$, we have

$$\begin{aligned} G(\mathbf{x}_1, \boldsymbol{\lambda}_1) &= \mathrm{G}(\mathbf{x}_1 + \mathbf{a}, \boldsymbol{\lambda}_1 + a) = G(\mathbf{x}_1 + \mathbf{a}, -\boldsymbol{\lambda}_1 - \mathbf{a}) \\ &= G(\mathbf{x}_1 + \mathbf{a} + \mathbf{b}, -\boldsymbol{\lambda}_1 - \mathbf{a} + \mathbf{b}) = G(x_2, \lambda_2), \end{aligned}$$

since H is invariant under $\boldsymbol{\lambda} \to -\boldsymbol{\lambda}$. Thus G must be constant and there is no non-trivial solution in this case. Other cases of non-existence of non-trivial solutions can be dealt with similarly, showing that no essential differentiability assumptions need be made. However the theory of partial differential equations (with or without interpretation in terms of Schwartzian "distributions") is a developed body of techniques for such problems and we now indicate how our problems can be dealt with using this theory.

3. Use of differential equations. Differentiating (3a) with respect to the components of θ leads to

$$\frac{\partial H}{\partial \phi}\frac{\partial \phi}{\partial \theta_i} = \sum_\alpha \frac{\partial G}{\partial p_\alpha}\frac{\partial p_\alpha}{\partial \theta_i}. \tag{4}$$

If ϕ is not a function of θ_i, this is

$$\sum_\alpha \frac{\partial G}{\partial p_\alpha}\frac{\partial p_\alpha}{\partial \theta_i} = 0. \tag{5a}$$

Otherwise $\partial H/\partial \phi$ can be eliminated by division of all such equations by one of them to give

$$\sum_\alpha (\partial G/\partial p_\alpha)(\partial p_\alpha/\partial \theta_1)/(\partial \phi/\partial \theta_1) = \sum_\alpha (\partial G/\partial p_\alpha)(\partial p_\alpha/\partial \theta_i)/(\partial \phi/\partial \theta_i). \tag{5b}$$

These are simultaneous linear equations in G. For a solution of the form $G(p)$ not involving θ, to exist, it is necessary that (5a), or respectively (5b) should not involve θ. The required "maximal" solution will then correspond to the general solution of the corresponding complete set of equations associated with (5a) or (5b). If there are k equations in the complete set, the solution will be an arbitrary function of $(n - k)$ independent functions of the p's, and hence will be equivalent to $(n - k)$ independent functions of the p's. These can be found by the usual method of auxiliary equations. References may be made to any standard book on differential equations, e.g. Goursat (1959). It is more convenient to treat examples separately than to derive general results from (5).

It may be noted that (5a), (5b) make no distinction between (3a) and the more general form

$$H(x, \phi) = G(p, \phi) \tag{3b}$$

where ϕ is allowed to occur explicitly in G. The equations (5a), (5b) follow as well from (3b) as from (3a). However, if solutions of the form (3b) are allowed, then (5a), (5b) may be allowed to involve ϕ.

The distinction between (3a) and (3b) is sometimes important. We shall see that in location and scale problems with $\theta = (\lambda, \sigma)$ we can find $G_1(\mathbf{p}) = (\bar{x} - \lambda)\sqrt{n}/s_x$ as a pivotal for λ and $G_2(\mathbf{p}) = s_x/\sigma\sqrt{n}$ as a pivotal for σ, and (G_1, G_2) will have an (approximately) known conditional joint distribution, given the observed value of the maximal ancillary. In the absence of other knowledge of σ, we can infer the distribution of G_1 by taking its marginal distribution, and use this for inferences about λ alone. But if G_2 were of the form $G_2(\mathbf{p}, \sigma)$, with σ unknown, the conditional joint distribution of (G_1, G_2) would not be approximately known and the marginalisation step might not be justifiable. We do not in this paper enter further into these logical considerations.

4. Examples. In the following examples, we take $S = \{\mathbf{x}\} = R^n$, $\Omega = (\lambda, \sigma) = R^1 \times R^+$, $P = \{\mathbf{p}\} = R^n$, $p_i = (x_i - \lambda)/\sigma$, $i = 1, 2, \cdots, n$, and D any subset of the set of all densities on R^n, sometimes narrowed to a neighbourhood of the standard spherical normal density on R^n. Unless otherwise specified, λ and σ are taken to be unknown.

EXAMPLE 1. *Ancillary statistics in location-scale model.* Here $\partial p_i/\partial \lambda = -1/\sigma$, $\partial p_i/\partial \sigma = -p_i/\sigma$, so that (5a) is

$$\sum \partial G/\partial p_i = 0, \quad \sum p_i \partial G/\partial p_i = 0.$$

This is a complete set of two linear differential equations which are equivalent to the auxiliary equations

$$dp_1 = dp_2 = \cdots = dp_n, \quad dp_1/p_1 = dp_2/p_2 = \cdots = dp_n/p_n.$$

The general solution is equivalent to the $(n - 2)$ independent functions $(p_i - p_1)/(p_2 - p_1) = (x_n - x_1)/(x_2 - x_1)$. This is the maximal ancillary in the form used by Fisher (1934). We will use the more usual form

$$a_i = (p_i - \bar{p})/s_p = (x_i - \bar{x})/s_x$$

where $n(n-1)s_p^2 = \sum(p_i - \bar{p})^2$, so that $\sum a_i = 0$, $\sum a_i^2 = n(n-1)$.

We now make $1-1$ transformations $\{p\}$ to $\{t, z, A\}$ where

$$t = \bar{p}/s_p = (\bar{x} - \lambda)/s_x, \quad z = s_p = s_x/\sigma, \quad A = \{a_i\}.$$

The basic pivotal p is thus equivalent to $\{t, z, A\}$. Inferences about (λ, σ) will be based on the conditional distribution $f(t, z \mid k)$ of $q = (t, z)$ given $A = k$ as described above. This procedure was outlined by Fisher (1934). In particular, the joint conditional distribution of $(t, z \mid A = k)$ yields a complete set of efficient, robust confidence regions for (λ, σ), efficient in the sense that no information has been lost through making $1-1$ transformations and by conditioning on A, and robust in the sense of being exact confidence intervals for every distribution in D. Integration with respect to z gives the marginal density of t and so will give separate inferences for λ, and integration with respect to t will similarly give separate inferences about σ.

The above reduction to $q = (t, z)$ with distribution conditional on A will be assumed in the following examples; q will be called the reduced pivotal and will replace p in equations (5a), (5b).

EXAMPLE 2a. *Division of a location-scale distribution in a given ratio.* The reduced pivotal as above is $q = (t, z)$,

$$t = (\bar{x} - \lambda)/s, \ z = s/\sigma, \ \phi = \lambda + k\sigma, \ (k \text{ known})$$

where $s = s_x$. Since $\partial t/\partial\lambda = -1/s$, $\partial z/\partial\sigma = -s/\sigma^2$, $\partial\phi/\partial\lambda = 1$ and $\partial\phi/\partial\sigma = k$, it can be readily verified that equations (5b) are

$$(\partial G/\partial t)(-1/s) = (\partial G/\partial z)(-\,s/\sigma^2)(1/k)$$

or

$$k\partial G/\partial t = z^2 \partial G/\partial z.$$

The auxiliary equations are $dt/k = -dz/z^2$, giving $t - k/z =$ constant. This yields a solution of (3a): $(t - k/z) = \{\bar{x} - (\lambda + k\sigma)\}/s$. Thus $\phi = \lambda + k\sigma$ can be estimated via $\{\bar{x} - (\lambda + k\sigma)\}/s = t - k/z$ with distribution conditional on $\{a_i\} = \{(x_i - \bar{x})/s\}$.

EXAMPLE 2b. *Coefficient of variation.* Define $\sigma/\lambda = 1/\phi$. Equations (5b) are now

$$\partial G/\partial t = (-s^2/\lambda\sigma)\partial G/\partial z,$$

giving

$$\phi(\partial G/\partial t) = -z^2(\partial G/\partial z)$$

which involves ϕ. There is thus no solution of the form (3a), but there will be of the more general form (3b). The auxiliary equations are $dt/\phi = dz/z^2$ with solution $t + \phi/z =$ constant. This yields a solution of the form $(t + \phi/z) = (\bar{x}/s)$. Thus ϕ, and hence $\sigma/\lambda = 1/\phi$ can be estimated via $t + \phi/z = \bar{x}/s$, for an arbitrary joint density of the observations.

EXAMPLE 3. *Two location-scale distributions.* Consider two samples $(\mathbf{x}, \mathbf{y})$ of sizes (m, n) from two location-scale distributions with parameters (λ_1, σ), $(\lambda_2, \rho\sigma)$. The reduced pivotals arising from both distributions are $(t_1\ z_1)$, (t_2, z_2), where $t_1 = (\bar{x} - \lambda_1)/s_1$, $z_1 = s_1/\sigma$, $t_2 = (\bar{y} - \lambda_2)/s_2$, $z_2 = s_2/\rho\sigma$, and $s_1 = s_x$, $s_2 = s_y$. Inferences about $\lambda_1, \lambda_2, \rho, \sigma$ will then be based on the conditional distribution of t_1, t_2, z_1, z_2 conditional on $\{a_i = (x_i - \bar{x})/s_1,\ b_j = (y_j - \bar{y})/s_2\}$, so that $q = (t_1, t_2, z_1, z_2)$ will replace p in equations (5a), and (5b).

EXAMPLE 3a. *Ratio of scale parameters, $\phi = \rho$.* Equation (3a) is $H(\mathbf{x}, \mathbf{y}, \rho) = G(t_1,$

$t_2, z_1, z_2)$. Differentiation with respect to λ_1, λ_2 shows that $G = G(z_1, z_2)$. Differentiation with respect to σ gives (5a) as

$$0 = \partial G/\partial z_1(-s_1/\sigma^2) + \partial G/\partial z_2(-s_2/\rho\sigma^2),$$

or

$$z_1\partial G/\partial z_1 = z_2\partial G/\partial z_2.$$

The auxiliary equation is $dz_1/z_1 = dz_2/z_2$ with a solution of the form (3a), z_2/z_1. Thus $G(z_1, z_2) = z_2/z_1 = (s_2/\rho)/s_1 = r/\rho$ where $r = s_2/s_1$. Thus ρ can be estimated via $z_2/z_1 = r/\rho$ conditional on $\{a_i\}$, $\{b_i\}$. For a normal parent distribution, ρ^2 is proportional to the variance ratio, and the result is the usual F distribution.

EXAMPLE 3b. *Difference in location.* Let $\phi = \lambda_1 - \lambda_2$ and ρ be assumed known. Then from the results of example 3a, r, having distribution depending only on ρ, is ancillary. Therefore, in addition to conditioning on $\{a_i\}$, $\{b_j\}$, we now also condition on r or equivalently on $v = r/\rho$. Thus r will be regarded as constant. Now (3a) is $H(\mathbf{x}, \mathbf{y}, \phi) = G(t_1, t_2, z_1, z_2)$, and differentiation with respect to ρ and to σ shows that G is independent of z_1, z_2, so that $G = G(t_1, t_2)$. Since $\partial t_i/\partial\lambda_i = -1/s_i$ $(i = 1, 2)$, and $\partial\phi/\partial\lambda_1 = -\partial\phi/\partial\lambda_2 = 1$, (5b) is

$$(1/s_1)\partial G/\partial t_1 = -(1/s_2)\partial G/\partial t_2,$$

or

$$r\partial G/\partial t_1 + \partial G/\partial t_2 = 0, \; r = s_2/s_1. \tag{6}$$

Since r is considered constant, the equation is of the required form, with auxiliary equation $dt_1/r = dt_2$, the solution being $t_1 - rt_2$.

Thus $\phi = \lambda_1 - \lambda_2$ can be estimated via $(\bar{x} - \bar{y} - \phi)/s_1 = t_1 - rt_2$ using the known distribution of $t_1 - rt_2$ conditional on $\{a_i\}$, $\{b_j\}$, and r. This will give a family of confidence intervals for $\phi = \lambda_1 - \lambda_2$ in terms of ρ.

As a referee has pointed out, since ρ is known we can allow G to be a function of ρ, so that $H(\underset{\sim}{x}, \underset{\sim}{y}, \phi) = G(t_1, t_2, z_1, z_2, \rho)$. The preceding method then yields the solution

$$(\bar{x} - \bar{y} - \phi)/s_1 = t_1 - (\rho z_1/z_2)t_2$$

and since $\rho z_2/z_1 = r$ is known and hence ancillary, we condition as before on this known value, and the two solutions are equivalent.

It is clear that in general $k_1t_1 + k_2rt_2$ similarly estimates any linear combination $k_1\lambda_1 + k_2\lambda_2$.

EXAMPLE 3c. *Length of mean vector.* Let $\phi = \lambda_1^2 + \lambda_2^2$ and assume ρ is known. Here $\partial\phi/\partial\lambda_1 = 2\lambda_1$, $\partial\phi/\partial\lambda_2 = 2\lambda_2$, so that (2b) is

$$(r/\lambda_1)\partial G/\partial t_1 = (1/\lambda_2)\partial G/\partial t_2$$

which involves λ_1, λ_2 explicitly, and so has no solution of the form (3b). It is thus not possible to estimate $\phi = \lambda_1^2 + \lambda_2^2$ via robust pivotals.

EXAMPLE 3d. *The ratio of means.* Let $\phi = \lambda_1/\lambda_2$ and ρ be assumed known. Now $\partial\phi/\partial\lambda_1 = 1/\lambda_2$, $\partial\phi/\partial\lambda_2 = -\lambda_1/\lambda_1^2$, so that (5b) is

$$(\lambda_2/s_1)\partial G/\partial t_1 + (\lambda_2^2/\lambda_1 s_2)\partial G/\partial t_2 = 0,$$

or

$$r\phi\partial G/t_1 + \partial G/\partial t_2 = 0$$

which depends on $\lambda = (\lambda_1, \lambda_2)$ and so has no solution of the form (3a). However, as in

Example 2b, λ occurs only through ϕ, so that there is a solution of the more general form (3b). The auxiliary equations are

$$dt_1/r\phi = dt_2$$

with solution of the form $t_1 - r\phi t_2 = (\bar{x} - \phi\bar{y})/s_1$. Thus $(\bar{x} - \phi\bar{y})/s_1 = t_1 - r\phi t_2$ has a distribution conditional on $\{a_i\}$, $\{b_j\}$, r, that is a function of ϕ and ρ only, and so can be used to set up confidence intervals for ϕ in terms of ρ.

EXAMPLE 3e. *The common mean.* Let $\lambda_1 - \lambda_2 = 0$, so that $\lambda_1 = \lambda_2 = \lambda$ and let ρ be assumed known. Here $t_1 - rt_2 = (\bar{x} - \bar{y})/s_1$ is parameter free and so is ancillary (that is, has a distribution depending only on the known ρ). There is one remaining functionally independent pivotal. Its distribution conditional as before on $\{a_i\}$, $\{b_j\}$, r and, in addition, on $t_1 - rt_2$, will yield confidence intervals for λ in terms of ρ. Conditional on $t_1 - rt_2$, the remaining pivotal can be taken as a linear function of t_1, t_2. It is convenient to take the "orthogonal complement"

$$rt_1 + t_2 = (\hat{\lambda} - \lambda)(s_1^2 + s_2^2)/s_1^2 s_2^2$$

where

$$\hat{\lambda} = (s_2^2\bar{x} + s_1^2\bar{y})/(s_1^2 + s_2^2).$$

EXAMPLE 4. *Predictive inference.* Future observations may be regarded as unknown parameters to which the foregoing methods can be applied. Let $p(x, \theta) = (p_1, \cdots, p_m)$, where $p_i = (x_i - \lambda)/\sigma$, and suppose it is required to predict the mean of n future observations $\bar{y} = \sum y_i/n$, the y_i occurring in the pivotals $q_i = (y_i - \lambda)/\sigma$. The reduced pivotals are, as before, $t = (\bar{x} - \lambda)/s$ and $z = s/\sigma$. Now $\phi = \bar{y}$ is to be estimated, and (3a) is $H(y, \phi) = G(t, z, q_1, \cdots, q_n)$. The equations (5) are

$$\partial H/\partial y_i = (\partial H/\partial \phi)/n = (\partial G/\partial q_i)/\sigma, \quad i = 1, 2, \cdots, n$$

$$\partial H/\partial \sigma = 0 = (\partial G/\partial z)(-s/\sigma^2) - \sum (\partial G/\partial q_i)(y_i - \lambda)/\sigma^2$$

$$\partial H/\partial \lambda = 0 = (\partial G/\partial t)(-1/s) - \sum (\partial G/\partial q_i)/\sigma.$$

The solution of the first n equations above is $\bar{q} = (\bar{y} - \lambda)/\sigma$. The next equation is then

$$z\partial G/\partial z + \bar{q}\partial G/\partial \bar{q} = 0,$$

with solution $u = \bar{q}/z = (\bar{y} - \lambda)/s$. Substituting $\bar{q} = uz$, the last equation is

$$\partial G/\partial t + \partial G/\partial u = 0$$

with solution $t - u = (\bar{x} - \bar{y})/s$. Thus $\bar{y}$ can be estimated by the known distribution of $(\bar{x} - \bar{y})/s = t - \sum q_i/(nz)$. For a specific future observation y_i, by setting $m = 1$ the associated robust pivotal can easily be seen to be $(\bar{x} - \bar{y}_i)/s$. This yields m pivotals for the estimation of the entire future sample $y_1, \cdots, y_m$. Note that these pivotals are not independent.

EXAMPLE 5. *Linear functional relationships.*

$$p_i = (x_i - \lambda_i)/\sigma, \; q_i = (y_i - \alpha - \beta\lambda_i)/\rho\sigma, \quad i = 1, \cdots, n,$$

where ρ is assumed known. Here we require $H(\mathbf{x}, \mathbf{y}, \beta) = G(p_1, \cdots, p_n, q_1, \cdots, q_n)$. Differentiation with respect to λ_i gives

$$\frac{\partial G}{\partial p_i} + \frac{\beta}{\rho}\frac{\partial G}{\partial q_i} = 0,$$

the solution of which is $G\{u_1, \cdots, u_n\}$ where $u_i = \beta p_i - \rho q_i = (\beta x_i - y_i - \alpha)/\sigma$. The elimination of α, σ is as in Example 1, yields

$$a_i = \sqrt{n(n-1)}\,[\beta(x_i - \bar{x}) - (y_i - \bar{y})]/\{\sum [\beta(x_i - \bar{x}) - (y_i - \bar{y})]^2\}^{1/2}.$$

This yields the maximal β-pivotal as n-dimensional.

Differentiating $H(\mathbf{x}, \mathbf{y}) = G(u_1, \cdots, u_n)$ with respect to β gives

$$\sum \frac{\partial G}{\partial u_i} x_i = 0$$

which has no solution of the form (2) independent of x. There are thus no pivotals independent of β and so no ancillary statistics. The above n dimensional pivotal in β cannot therefore be reduced without loss of information to a one dimensional pivotal on which to base confidence intervals.

5. Explicit expressions for Examples 3b, 3d, and 3e. Let g_1, g_2 be the parent density functions for the pivotals arising from the first and second samples respectively. Then the conditional distribution of t_1, t_2, z_1, z_2 given $\{a_i\}$, $\{b_j\}$ is proportional to

$$z_1^{m-1} z_2^{n-1} \prod_{i=1}^m g_1\{(a_i + t_1)z_1\} \prod_{j=1}^n g_2\{(b_j + t_2)z_2\}.$$

Letting $v = z_2/z_1$, the resulting distribution t_1, t_2, v is proportional to

$$v^{n-1} \int_{z_1} z_1^{m+n-1} \prod g_1\{(a_i + t_1)z_1\} \prod g_2\{(b_j + t_2)vz_1\}\, dz_1. \tag{7}$$

The conditional distribution of t_1, t_2 given r (i.e. v) is proportional to (7) with v regarded as constant.

Let $w_1 = t_1 - ct_2$, where $c = r$ in Example 3b and $r\phi$ in Example 3d. Let $w_2 = ct_1 + t_2$. From (7), the conditional distribution of w_1, w_2 given the $\{a_i\}$, $\{b_i\}$, and v is proportional to

$$\int_{z_1} z_1^{m+n-1} \prod g_1\{[a_i + (w_1 + cw_2)/(1 + c^2)]z_1\} \cdot \prod g_2\{[b_j + (w_2 - cw_1)/(1 + c^2)]vz_1\}\, dz_1. \tag{8}$$

Integrating with respect to w_2 gives the required distribution of w_1 for Example 3b, 3d.

If g_1 and g_2 are standard normal, the integrand of (8) is proportional to

$$z_1^{m+n-1} \exp(-\tfrac{1}{2} z_1^2 \{\textstyle\sum_i [a_i + (w_1 + cw_2)/(1 + c^2)]^2 + v^2 \sum_j [b_j + (w_2 - cw_1)/(1 + c^2)]^2\}).$$

Integrating this with respect to w_2 and recalling that $\sum a_i^2 = m(m-1)$, $\sum b_j^2 = n(n-1)$, $\sum a_i = \sum b_j = 0$, and that c and v are held constant, the result is porportional to

$$z_1^{m+n-1} \exp[-\tfrac{1}{2} z_1^2 \{m(m-1) + n(n-1)v^2 + mnv^2 w_1^2/(mc^2 + nv^2)\}].$$

Integrating this with respect to z_1 gives the required conditional density of w_1 as proportional to

$$[m(m-1) + n(n-1)v^2 + mnv^2 w_1^2/(mc^2 + nv^2)]^{-(m+n-1)/2}$$

which is proportional to

$$[1 + t^2/(m + n - 2)]^{-(m+n-1)/2},$$

where

$$t^2 = \frac{mnv^2 w_1^2 (m + n - 2)}{(mc^2 + nv^2)[m(m-1) + n(n-1)v^2]}.$$

Setting $c = r$ and $r\phi$ gives

$$t^2 = mn(\bar{x} - \bar{y} - \phi)^2/(m\rho^2 + n)\hat{\sigma}^2,\ mn(\bar{x} - \phi\bar{y})/(m\phi^2\rho^2 + n)\hat{\sigma}^2,$$

where $(m + n - 2)\hat{\sigma}^2 = m(m-1)s_1^2 + n(n-1)s_2^2/\rho^2$ for Examples 3b, 3d respectively. In each case this gives the usual $t = t_{(m+n-2)}$. The former is the usual test for the difference of normal means, usually derived under the assumption of equal variances ($\rho = 1$). The latter is the Fieller solution to the problem of the ratio of normal means. It is interesting to note that this solution has the form (3a). Thus for a normal distribution the explicitly-occurring $\phi = \lambda_1/\lambda_2$ on the right hand side of (3b) and in (8) is absorbed to form the t variate above, thus for this special case yielding the simpler solution (3a).

If $\lambda_1 = \lambda_2 = \lambda$, then as discussed in Example 3e, $w_1 = t_1 - rt_2 = (\bar{x} - \bar{y})/s_1$ is ancillary. We use the distribution of w_2 conditional on w_1 for inferences about λ. This distribution is proportional to (8) with $\{a_i\}$, $\{b_i\}$, v and w_1 all held constant. For g_1 and g_2 standard normal, this is proportional to

$$\left\{m(m-1) + n(n-1)v^2 + \frac{[m(w_1 + rw_2)^2 + nv^2(w_2 - rw_1)^2]}{(1+r^2)^2}\right\}^{-(m+n)/2}$$

$$= \left\{m(m-1) + n(n-1)v^2 + \frac{mnw_1^2}{m\rho^2 + n} + \left(m + \frac{n}{\rho^2}\right)\left(\frac{\hat{\lambda} - \lambda}{s_1}\right)^2\right\}^{-(m+n)/2}$$

where

$$\hat{\lambda} = (m\rho^2\bar{x} + n\bar{y})/(m\rho^2 + n).$$

Since v and w_1 are held constant, this is proportional to

$$\left[1 + \frac{t^2}{m+n-1}\right]^{-(m+n)/2}$$

where $t^2 = (m + n/\rho^2)(\hat{\lambda} - \lambda)^2/\hat{\sigma}^2$,

$$(m+n-1)\hat{\sigma}^2 = m(m-1)s_1^2 + n(n-1)s_2^2/\rho^2 + mn(\bar{x} - \bar{y})^2/(m\rho^2 + n)$$

and $\hat{\lambda}$ is given above. Thus t has the $t_{(m+n-1)}$ distribution.

6. The Behrens Fisher problem. This is the same as Examples 3b, 3e with ρ regarded as completely unknown. Thus r is not ancillary. Eliminating r from (6) gives

$$\rho z_2 \partial G/\partial t_1 + z_1 \partial G/\partial t_2 = 0 \tag{9}$$

which involves the unknown ρ. There is thus no solution of the required form (3a). In the absence of knowledge of ρ, $\phi = \lambda_1 - \lambda_2$ cannot be estimated via robust pivotals.

7. Discussion. In using a pivotal $G(p, \phi) = H(t(x), \phi)$ to make inferences about $\phi(\theta)$ given observed values $x = x_0$, we need to make use of the fact that the supposition that ϕ is in a set $\mathscr{C}_0$ is "equivalent, when $x = x_0$ is known" to G being in a set

$$\mathscr{C} = \{G: G = H(t(x_0), \phi) \quad \text{for some} \quad \phi \in \mathscr{C}_0\}$$

in the following sense: Denoting the proposition "$x = x_0$" by "E", we must have

$$E \text{ known and } \phi \in \mathscr{C}_0 \to G \in \mathscr{C}, \text{ and } \Pr\{G \in \mathscr{C}\} = \Gamma$$

$$E \text{ known and } \phi \notin \mathscr{C}_0 \to G \notin \mathscr{C}, \text{ and } \Pr\{G \notin \mathscr{C}\} = 1 - \Gamma.$$

This enables us to attach the probability that $G(p) \in \mathscr{C}$ to that of $\phi \in \mathscr{C}_0$, though it does not thereby specify ϕ as a random variable—only functions of p are random variables. $\mathscr{C}_0$ will be a conditional confidence set; by appealing to the law of large numbers, regarding $\mathscr{C}$ as fixed, in a series on independent experiments in which the ancillaries $A(x)$ take the same values as in the current experiment, $\mathscr{C}_0$ will be a set-valued function of x and ϕ which will cover ϕ with frequency Γ. If, as the ancillaries vary, we change $\mathscr{C}$ (if necessary) so as to keep Γ constant, the conditioning on $A(x)$ becomes irrelevant and $\mathscr{C}_0$ is an unconditional confidence set; however since variation of ancillaries will often imply variation in the informativeness of the data, such forcing of Γ to be constant may lead to unnecessarily long confidence intervals—see Barnard (1976, 1982a).

R. A. Fisher noted that substituting $t(x_0)$ for $t(x)$ in $H(t, \phi)$ generates from the distribution of H a distribution of ϕ which he called the fiducial distribution of ϕ. He seemed to suggest (Fisher, 1973, page 169) for example, in the case of location λ and scale σ, with $\phi_1 = (\lambda, \sigma)$, that it would be possible to derive, for any (measurable) function ϕ_2 of (λ, σ), a distribution which would have the same status as that of (λ, σ); that is, to any supposition of the form $\phi(\lambda, \sigma) \in \mathscr{D}$, there would correspond, given the observations, a proposition about the pivotals (t, z), whose maximally conditioned probability could be evaluated using the density $\eta(t, z \mid \mathbf{k})$. Bartlett (1937) queried this, and suggested that $\phi(\lambda, \sigma) = \lambda + 3\sigma$ would provide a counter-example. Fisher then showed, in reply to Bartlett, that the pivotal $u = t - 3/z$ would provide a basis for statements about this ϕ of the form required, as in Example 2b. However, although Bartlett's "counter-example" was unhappily chosen, he was right in principle, as shown in Section 2. There it is clear that $\phi(\lambda, \sigma) = \lambda^2$ has no corresponding pivotal $u' = G(t, z)$, such that $\phi \in \mathscr{D}$ implies, and is implied by $H_0 \in \mathscr{D}_0$.

This bears on another objection to Fisher's fiducial argument which was raised by Stein (1959), who considered a model which, in pivotal terms, has $S = R^n = \{\mathbf{x}\}$, $\Omega = R^n = \{\boldsymbol{\lambda}\}$, $P = R^n = \{\mathbf{p}\}$, $\mathbf{p}$ a vector with ith component $p_i = x_i - \lambda_i$, $D = \{\xi\}$ with ξ the spherical unit normal density in R^n. Stein showed that the fiducial distribution of $\boldsymbol{\lambda}$, given $x = x_0$, would assign probability ½ to the set for which

$$\lambda'\lambda > w + \mathbf{x}'\mathbf{x}$$

though for sufficiently large n the confidence coefficient to be attached to this set can be made arbitrarily small. But the results of Section 2 imply that in the pivotal model for Stein's example, $\phi = \lambda'\lambda$ has no corresponding pivotal $u = G(p)$. Stein's paradox is thus avoided.

Example 3 was phrased in terms of σ, $\rho\sigma$ to emphasize that the scale parameters are not required to be equal, as seems commonly assumed in textbooks. It is only necessary that their ratio ρ be known. Fisher's solution to the Behrens-Fisher problem assumes ρ to be **wholly** unknown, apart from such information about ρ as is contained in the data. He represents this ignorance by assuming that, when r is known, the pivotal r/ρ retains its marginal distribution; such complete ignorance of ρ is unlikely to occur often in practice. Thus in practice, when this ratio is uncertain, it would be preferable to examine how the resulting confidence intervals for $\lambda_1 - \lambda_2$ are affected by variations in ρ.

To obtain an overall average figure for the confidence coefficient to be attached to any specified interval, the coefficient $\Gamma(\rho)$, for a specified ρ, can be averaged over a distribution intermediate between the Fisher distribution and the singular distribution $\rho = \rho_0$. Further details are given in Barnard (1982b). It is also shown in Barnard (1982b) that, in requiring a test in the Behrens-Fisher problem to be similar, giving a constant α-level independent of ρ, Neyman's approach to the problem rules out all reasonable solutions. The proof, using results given above, is simpler than Linnik's related result which applies only to the case of normal distributions.

REFERENCES

BARNARD, G. A. (1976). Conditional inference is not inefficient. *Scand. J. Statist.* **3** 132–134.

BARNARD, G. A. (1977). Pivotal inference and the Bayesian controversy. *Bull. Inter. Statist. Inst.* **47** 543–551.

BARNARD, G. A. (1982a). Conditionality versus similarity in the analysis of 2 × 2 tables. *Statistics and Probability: Essays in Honour of C. R. Rao.* North-Holland, Amsterdam. 59–65.

BARNARD, G. A. (1982b). A new approach to the Behrens-Fisher problem. *Utilitas Mathematica,* to appear.

BARTLETT, M. S. (1937). Properties of sufficiency and statistical tests. *Proc. Roy. Statist. Soc. Ser. A* **160** 268–282.

BASU, D. (1964). Recovery of ancillary information. *Sankhya Ser. A* **26** 3–16.

BIRNBAUM, A. (1972). More on concepts of statistical evidence. *J. Amer. Statist. Assoc.* **67** 858–861.

BUNKE, H. (1975). Statistical inference: fiducial and structural vs. likelihood. *Math. Operationsforsch. u. statistik* **6** 667–676.

BUNKE, O. (1976). Conditional probability in incompletely specified stochastic equations and statistical inference. *Math. Operationsforsch. u. statist.* **7** 673–678.

DAWID, A. P. and STONE, M. (1982). The functional model basis of fiducial inference. *Ann. Statist.* **10** 1054–1067.

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

FISHER, R. A. (1936). Uncertain inference. *Proc. Amer. Acad. Arts and Sciences* **71** 245–258.

FISHER, R. A. (1945). The logical inversion of the notion of the random variable. *Sankhya* **7** 129–132.

FRASER, D. A. S. (1979). *Inference and Linear Models.* McGraw-Hill, New York.

GOURSAT, E. (1959). *Differential Equations.* Translated by E. R. Hedrich and O. Dunkel. Dover, New York.

SEGAL, I. E. (1938). Fiducial distributions of several parameters with application to a normal system. *Proc. Camb. Phil. Soc.* **34** 41–47.

STEIN, C. (1959). An example of wide discrepancy between fiducial and confidence limits. *Ann. Math. Statist.* **30** 887–880.

MILL HOUSE, HURST GREEN
BRIGHTLINGSEA COLCHESTER
ESSEX C070EH
ENGLAND

DEPARTMENT OF STATISTICS
UNIVERSITY OF WATERLOO
WATERLOO, ONTARIO
CANADA N2L 3G1