

LIKELIHOOD AS APPROXIMATE PIVOTAL DISTRIBUTION

By
David Hinkley

University of Minnesota
Technical Report No. 360
January 1980

Research supported by NSF Grant MCS-7904558.

LIKELIHOOD AS APPROXIMATE PIVOTAL DISTRIBUTION

By D.V. Hinkley
University of Minnesota

SUMMARY

For a single-parameter curved exponential family, it is shown that the conditional distribution of the maximum likelihood estimator is the normalized likelihood function, except for a relative error proportional to inverse sample size.

Key Words: Ancillary statistic; Information; Edgeworth expansion; Maximum likelihood estimate; Sufficient statistic; Exponential family.

1. INTRODUCTION

Recently it has been shown (Efron & Hinkley, 1978; Peers, 1978) that there is a conditional normal approximation for the distribution of the maximum likelihood estimator in the one-parameter case. This conditional normal approximation uses an observed characteristic of the likelihood as a variance approximation. In principle the approximation is unnecessary for the special case of estimating a location parameter, because Fisher (1934) derived the exact, fully-informative conditional distribution of the maximum likelihood estimator. That distribution is the normalized likelihood function. In this paper we present theoretical evidence that Fisher's special case result is a good approximation in general.

Specifically, let $\text{lik}(\theta)$ denote the likelihood for a sample of size n from a curved exponential family distribution indexed by a single parameter θ and let $\hat{\theta}$ denote the maximum likelihood estimator. Then we find that the probability density of the pivot $\hat{\theta} - \theta$, conditional on a natural approximate ancillary a , is

$$h_{\theta}(\hat{\theta}|a) = \frac{\text{lik}(\theta)}{\int \text{lik}(u)du} \{1+O(n^{-1})\},$$

provided that θ is chosen to make the Fisher information constant. The proof is based on an Edgeworth expansion for the joint distribution of the sufficient statistic, which is developed in Section 2. The final step to the conditional distribution of $\hat{\theta}$ is outlined in Section 3, where we find it essential to choose θ so that the expected Fisher information is constant. Section 4 contains some comments on the result, its implications and its possible generalization.

2. JOINT DISTRIBUTION OF MAXIMUM LIKELIHOOD ESTIMATE AND ANCILLARIES

We suppose that X_1, \dots, X_N are independent, p -dimensional identically distributed vectors each with probability density

$$f_{\theta}(x) = \exp(\lambda_{\theta}^T x - \psi_{\theta}) \quad \theta \in \Omega,$$

where $\Omega \in \mathbb{R}$. Thus $f_{\theta}(x)$ is in a one-dimensional curved subset of the p -dimensional canonical exponential family. The minimal sufficient statistic is $S = \sum X_j$, whose probability density we denote by $g_{\theta}(s)$. Our first aim is to find an expansion for $g_{\theta}(s)$ in terms of the log likelihood $\ell_{\theta}(s) = \sum \log f_{\theta}(x_j)$ and its derivatives at the maximum.

Before proceeding we introduce the following definitions and notations. Let

$$\beta_{\theta} = E_{\theta}(X), \quad \Sigma_{\theta} = \text{Var}_{\theta}(X) \quad (1)$$

and

$$\Lambda_{\theta} = (\dot{\lambda}_{\theta}, \ddot{\lambda}_{\theta}, \dots, \lambda_{\theta}^{(\cdot p)}),$$

where dots refer to differentiation with respect to θ . Simple calculations show that

$$\dot{\ell}_{\theta}(s) = \dot{\lambda}_{\theta}^T(s - n\beta_{\theta}) \quad (2)$$

$$\ddot{\ell}_{\theta}(s) = \ddot{\lambda}_{\theta}^T(s - n\beta_{\theta}) - \mathcal{J}_{\theta} \quad (3)$$

and so on, where

$$E\{-\ddot{\ell}_{\theta}(s)\} = \mathcal{J}_{\theta} = \text{Var}_{\theta}\{\dot{\ell}_{\theta}(s)\} = n\dot{\lambda}_{\theta}^T\beta_{\theta}$$

is the expected Fisher information. We assume that the maximum likelihood estimate $\hat{\theta}$ is a solution of $\dot{\ell}_{\theta} = 0$, and we define the observed Fisher information $I = -\ddot{\ell}_{\hat{\theta}}$.

Finally let

$$nV_{\theta} = \text{Var}_{\theta}\{\dot{\ell}_{\theta}(s), \ddot{\ell}_{\theta}(s), \dots, \ell_{\theta}^{(p)}(s)\} = n\Lambda_{\theta}^T\Sigma_{\theta}\Lambda_{\theta}. \quad (4)$$

For stylistic convenience we often drop the subscript θ , and use a circumflex to denote evaluation at $\hat{\theta}$; for example, $V = V_{\theta}$ and $\hat{V} = V_{\hat{\theta}}$.

The first step in expanding the joint density $g_{\theta}(s)$ is to note that sufficiency implies

$$g_{\theta}(s) = g_{\tilde{\theta}}(s) \exp\{\ell_{\theta}(s) - \ell_{\tilde{\theta}}(s)\} \quad (5)$$

for any $\tilde{\theta} \in \Omega$. A saddle-point expansion of $g_{\theta}(s)$ corresponds to the optimal choice $\tilde{\theta} = \hat{\theta}(s)$ used in the Edgeworth expansion

$$g_{\tilde{\theta}}(s) = \frac{|n\Sigma_{\tilde{\theta}}|^{-\frac{1}{2}}}{(2\pi)^{\frac{1}{2}p}} [\exp\{-\frac{1}{2}(s - n\beta_{\tilde{\theta}})^T(n\Sigma_{\tilde{\theta}})^{-1}(s - n\beta_{\tilde{\theta}})\}] (1 + r_n), \quad (6)$$

where

$$r_n = \frac{1}{6\sqrt{n}} (H^T\rho')^{[3]}(s') + O(n^{-1}), \quad (7)$$

as described in detail by Barndorff-Nielsen & Cox (1979). The term $(H^T\rho')^{[3]}(s')$ in (7) involves Hermite polynomials of degree up to 3 with argument s' which is an orthogonalised, standardized version of s . An important property is that $(H^T\rho')^{[3]}(s')$ is invariant under affine transformation of s .

With the choice $\tilde{\theta} = \hat{\theta}(s)$ in (6), an initial simplification is obtained by

$$\hat{s} = \hat{\Lambda}^T (s - n\beta). \quad (8)$$

Note that \hat{s}_j is the deviation of $\dot{\ell}_{\hat{\theta}}^{(\cdot j)}$ from $E_{\theta}\{\dot{\ell}_{\theta}^{(\cdot j)}\}$ evaluated at $\theta = \hat{\theta}$, and that in particular $\hat{s}_1 = 0$. We shall write $\hat{s}^T = (0, \hat{u})$, and partition V in (4) accordingly as

$$V = \begin{pmatrix} v_{11} & v_{12} \\ v_{12}^T & v_{22} \end{pmatrix},$$

where $nv_{11} = J$. Then writing

$$W = v_{22} - v_{12}^T v_{12} / v_{11}, \quad (9)$$

and using (8), expression (6) yields

$$g_{\hat{\theta}}(s) = \frac{|n\hat{\Sigma}|^{-\frac{1}{2}}}{(2\pi)^{\frac{1}{2}p}} \exp \left\{ -\frac{1}{2} \hat{u}^T (nW)^{-1} \hat{u} \right\} \{1 + r_n(s')\}. \quad (10)$$

The next step is to rewrite (10) in terms of approximately ancillary statistics, which are transformations of $\ddot{\ell}_{\hat{\theta}}, \dots, \dot{\ell}_{\hat{\theta}}^{(\cdot p)}$. To do this we choose the unique lower triangular matrix $B = B_{\hat{\theta}}$ such that

$$BWB^T = \text{identity} \quad (11)$$

and define

$$a = (a_2, \dots, a_p)^T = \frac{1}{\sqrt{n}} \hat{B} \hat{u}. \quad (12)$$

The latter statistic is asymptotically ancillary, having a normal limiting distribution with zero mean and identity variance matrix. This can be proved by first applying a central limit theorem to the linear Taylor expansion of $n^{-\frac{1}{2}} \hat{u}$, whose limiting variance matrix is

$$n^{-1} \text{Var}_{\theta}(\ddot{\ell}_{\theta}, \dots, \ell_{\theta}^{(p)} | \dot{\ell}_{\theta} = 0) = W_{\theta};$$

cf. Lemma 2 of Efron & Hinkley (1978) for the case $p = 2$. The particular choice of a in (12), in which a_j depends only on $\ddot{\ell}_{\hat{\theta}}, \dots, \ell_{\hat{\theta}}^{(j)}$, corresponds to a relevance ordering with a_2 the dominant ancillary. If γ_{θ} denotes the statistical curvature per observation (Efron, 1975), one can see from (12) that

$$a_2 = \sqrt{n} (I - \mathcal{J}_{\hat{\theta}}) / (\mathcal{J}_{\hat{\theta}} \gamma_{\hat{\theta}}^2)^{1/2}, \quad (13)$$

which was denoted by q in §5 of Efron & Hinkley (1978).

Returning to expression (10), we now substitute from (12) to obtain

$$g_{\hat{\theta}}(s) = \frac{|n\hat{\Sigma}|^{-1/2}}{(2\pi)^{1/2p}} \{ \exp(-1/2 a^T a) \} \{ 1 + \hat{r}_n(a) \}, \quad (14)$$

where in $\hat{r}_n(a)$ we have used the affine invariance property of the Hermite polynomials to replace s' by a . The constant coefficients ρ' in (7) are evaluated at $\hat{\theta}$, hence the circumflex on r_n .

In order to find an expansion for the joint distribution of $(\hat{\theta}, a)$ it remains only to derive the Jacobian $J = |\partial(\hat{\theta}, a) / \partial s|$. This can be obtained by a geometric argument, or by direct calculus, and is

$$J = |\hat{\Lambda}| \begin{vmatrix} I^{-1} & 0 \\ 0 & \frac{1}{\sqrt{n}} \hat{B}^T \end{vmatrix}; \quad (15)$$

see Appendix for details. Using (15) together with (11), (9), and (4) in (14) leads to

$$g_{\hat{\theta}}(\hat{\theta}, a) = (2\pi)^{-1/2p} \frac{I}{\mathcal{J}_{\hat{\theta}}^{1/2}} \{ \exp(-1/2 a^T a) \} \{ 1 + r_n(a) \}.$$

This can be simplified, by substitution for I from (13), to give

$$g_{\hat{\theta}}(\hat{\theta}, a) = (2\pi)^{-\frac{1}{2}p} \int_{\hat{\theta}}^{\frac{1}{2}} \{\exp(-\frac{1}{2}a^T a)\} \{1 + \hat{r}_n^*(a)\}, \quad (16)$$

where

$$\begin{aligned} \hat{r}_n^*(a) &= 1 + \frac{1}{\sqrt{n}} \left\{ \frac{1}{6} (H^T \rho)^{[3]}(a) + \hat{\gamma} a_2 \right\} + O(n^{-1}) \\ &= 1 + \frac{1}{\sqrt{n}} \hat{\eta}_1(a) + O(n^{-1}), \end{aligned} \quad (17)$$

say. Finally, substituting (16) in (5), we have the general density expansion

$$g_{\hat{\theta}}(\hat{\theta}, a) = \frac{\int_{\hat{\theta}}^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}}} [\exp\{\ell_{\theta}(s) - \ell_{\hat{\theta}}(s)\}] \left\{ \prod_{j=2}^p \phi(a_j) \right\} \left\{ 1 + \frac{1}{\sqrt{n}} \hat{\eta}_1(a) + O(n^{-1}) \right\}, \quad (18)$$

where ϕ is the standard normal density.

Of course by carrying further terms in the Edgeworth expansion, as in Barndorff-Nielsen & Cox (1979), one can obtain a more detailed expansion of the final term in (18). However, such a finer development would be of no value in the approximate analysis of the next section.

3. CONDITIONAL DISTRIBUTION OF THE MAXIMUM LIKELIHOOD ESTIMATE

The main result we seek involves the expansion for the conditional density $h_{\hat{\theta}}(\hat{\theta} | a)$ obtained from (18). In Fisher's (1934) analysis for the location model, where $x_j = \theta + e_j$ with known distribution for the e_j , the ancillary a is the ordered set of residuals $x_j - \hat{\theta}$. Fisher showed that, because $\ell_{\theta}(s) = \ell_{\hat{\theta}}(\hat{\theta}, a) = \ell_{\hat{\theta}-\hat{\theta}}(0, a)$,

$$h_{\theta}(\hat{\theta}|a) = \frac{\exp\{\ell_{\theta}(\hat{\theta}, a)\}}{\int \exp\{\ell_{\theta}(t, a)\} dt} = \frac{\exp\{\ell_{\theta}(s)\}}{\int \exp\{\ell_t(s)\} dt}. \quad (19)$$

This is the normalized likelihood function, a function only of the pivot $\hat{\theta}-\theta$ and ancillary a . We wish to show that (19) is correct, to within a $O(n^{-1})$ relative error, for the curved exponential family model.

It turns out to be not only convenient but also crucial to choose θ so that $\mathcal{J}_{\theta} \equiv n$. We now assume this choice, and attempt to duplicate Fisher's result for the pivot $\hat{\theta}-\theta$. The first step is to obtain the expansion

$$h_{\theta}(\hat{\theta}|a) = \frac{\exp\{\ell_{\theta}(\hat{\theta}, a) - \ell_{\hat{\theta}}(\hat{\theta}, a)\}}{\int \exp\{\ell_{\theta}(t, a) - \ell_t(t, a)\} dt} \{1 + O(n^{-1})\} \quad (20)$$

from (18). This follows by noting that $\hat{\eta}_1(a) = \eta_1(a) + O(n^{-1/2})$ for $\hat{\theta}-\theta = O(n^{-1/2})$, and then formally computing the integral of (18). The term $\eta_1(a)$ cancels in the resulting expression for $h_{\theta}(\hat{\theta}|a)$, leaving the $O(n^{-1})$ term in (20).

Notice that in Fisher's problem $\ell_{\hat{\theta}}(\hat{\theta}, a)$ is ancillary and therefore has cancelled out in (19). But this is not possible in our problem. Nor is the last step in (19) valid for us. Instead we must resort to approximations using the relationships

$$I = n + \sqrt{n} \gamma a_2 = n + \sqrt{n} \gamma a_2 + O_p(1)$$

$$\ell_{\theta}^{(\cdot, j)} = E_{\theta}\{\ell_{\theta}^{(\cdot, j)}\} + O_p(n^{-1/2}) \quad (j \geq 3), \quad (21)$$

where the first relationship depends on the fact that $\mathcal{J}_{\theta} \equiv n$. Now for $\theta-t = O(n^{-1/2})$ and $\hat{\theta}-\theta = O(n^{-1/2})$, (21) implies that the Taylor expansion for the exponent in the integral in (20) may be written

$$\begin{aligned}
\ell(t, a) - \ell_t(t, a) &= -\frac{1}{2}(t-\theta)^2 (n + \sqrt{n}\gamma_t a_2) - \frac{n}{6} (t-\theta)^3 \mu_t + O(n^{-1}) \\
&= -\frac{1}{2}(t-\theta)^2 (n + \sqrt{n}\gamma_\theta a_2) - \frac{n}{6} (t-\theta)^3 \mu_\theta + O(n^{-1}) \\
&= -\frac{1}{2}(t-\theta)^2 (n + \sqrt{n}\gamma_{\hat{\theta}} a_2) - \frac{n}{6} (t-\theta)^3 \mu_{\hat{\theta}} + O(n^{-1}) \\
&= \ell_{\hat{\theta}-t+\theta}(\hat{\theta}, a) - \ell_{\hat{\theta}}(\hat{\theta}, a) + O(n^{-1}), \tag{22}
\end{aligned}$$

where $n\mu_\theta = E_\theta(\ddot{\ell}_\theta)$. But the integral in (20) may be truncated at $t-\theta = 0(n^{-\frac{1}{2}})$ with relative error $O(n^{-1})$, so that we may substitute (22) in the integrand and obtain

$$h_\theta(\hat{\theta}|a) = \frac{\exp\{\ell_\theta(s)\}}{\int \exp\{\ell_u(s)\} du} \{1+O(n^{-1})\}, \tag{23}$$

the analog of Fisher's result (19), valid for $\hat{\theta}-\theta = O(n^{-\frac{1}{2}})$.

Expressed in pivotal form, the result for $T = \hat{\theta}-\theta$ is

$$h_\theta(t|a) = \frac{\text{lik}_s(\hat{\theta}-t)}{\int \text{lik}_s(u) du} \{1+O(n^{-1})\}. \tag{24}$$

In principle the normalized likelihood is, then, a better approximation than Efron & Hinkley's (1978) normal approximation, since the latter involves a relative error of order $n^{-\frac{1}{2}}$. The result (24) also more closely resembles the Bayes posterior distribution obtained for the pivot $\hat{\theta}-\theta$ when using the Jeffreys' prior density proportional to $J_\theta^{-\frac{1}{2}}$.

4. COMMENTS

The result linking normalized likelihood to the conditional distribution of $\hat{\theta}$ is very similar to a result obtained by Cox (1980) for the "confidence distribution" based on the score statistic $\dot{\ell}_\theta$; Cox's second-order ancillary seems to be an expansion of our a_2 .

A referee has rightly questioned the value of our theoretical result, particularly in view of the restriction on parameterization. To take the theoretical implications first, it would appear that likelihood per se has a stronger interpretation than was previously believed, and that the oft-advocated information-stabilizing choice of θ does indeed correspond to a (local) location parameter. I believe that (23) can be used to show that the log likelihood ratio statistic $2(\ell_{\hat{\theta}} - \ell_{\theta})$ is equal to $\chi_1^2 + o_p(n^{-1})$ conditional on a , generalizing Lemma 3 of Efron & Hinkley (1978). For the particular class of problems considered, $a^T a$ as defined by (12) is the usual chi-squared goodness of fit statistic, whose distribution can be expanded in asymptotic series by use of (18).

The practical use of (23) requires choosing θ so that \mathcal{J}_{θ} is constant. Section 6 of Efron & Hinkley (1978) gives details of one example where this is relatively easy. Cox's (1980) Poisson process example is also straightforward. But the example discussed by Hinkley (1979) appears intractable.

One suggestion to be drawn from the restriction $\mathcal{J}_{\theta} = n$ is that the result (23), and Efron & Hinkley's (1978) conditional normal approximation, will not hold for the multiparameter case in general. To shed a little more light on this point, consider the following example due to George Cobb. Let $Y_i = \theta_2 Z_i + \varepsilon_i$, where the ε_i are independent $N(0,1)$ and the Z_i are independent $N(\theta_1, 1)$. The minimal sufficient statistic is $(\sum Z_i, \sum Z_i^2, \sum Z_i Y_i)$, and $A = \sum Z_i^2 - (\sum Z_i)^2/n$ is ancillary with respect to $\theta^T = (\theta_1, \theta_2)$. The expected Fisher information matrix is $\mathcal{J}_{\theta} = \text{diag}(n, n+n\theta_1^2)$, and the observed Fisher information matrix is $I = \text{diag}(n, a+n\hat{\theta}_1^2)$. Evidently the conditional distribution of $\hat{\theta}_2$ given a is not approximately $N(\theta_2, I^{22})$, which would generalize Efron & Hinkley's (1978) result, but this is the conditional distribution given $a + n\hat{\theta}_1^2$. Note, however, that

$$(\hat{\theta}-\theta)^T I(\hat{\theta}-\theta) = 2(\ell_{\hat{\theta}}-\ell_{\theta}) = \chi_2^2 \text{ given } a. \quad (25)$$

One might speculate that (25) holds approximately in general, for an appropriate approximate "simultaneous ancillary" a , so that a simple approximate conditional theory exists for simultaneous estimation.

ACKNOWLEDGEMENTS

I thank Patricia Grambsch for discussions relating to Section 4. The research described here was supported by a grant from the National Science Foundation.

APPENDIX

To evaluate the Jacobian given in (15), we begin with (8) and (13) which combine to give

$$a = B_{\hat{\theta}} \Lambda_{\hat{\theta}}^T (s - n\beta_{\hat{\theta}}),$$

if it is understood that the zero first element of $\Lambda_{\hat{\theta}}^T (s - n\beta_{\hat{\theta}})$ is deleted; the statistic $\hat{\theta} = \hat{\theta}(s)$ satisfies the likelihood equation

$$\dot{\lambda}_{\hat{\theta}}^T (s - n\beta_{\hat{\theta}}) = 0.$$

Consider a small perturbation δs of s , with corresponding perturbations $\delta \hat{\theta}$ and δa . Then first expanding the likelihood equation we find that, retaining first-order terms,

$$\{\dot{\lambda}_{\hat{\theta}}^T + (\delta \hat{\theta}) \ddot{\lambda}_{\hat{\theta}}\} \{s - n\beta_{\hat{\theta}} + \delta s - n(\delta \hat{\theta}) \dot{\beta}_{\hat{\theta}}\} = 0.$$

Therefore, because $\hat{\theta} = n \dot{\lambda}_{\hat{\theta}}^T \dot{\beta}_{\hat{\theta}}$, and because of (3),

$$\delta \hat{\theta} = \frac{1}{I} \dot{\lambda}_{\hat{\theta}}^T (\delta s)$$

which implies

$$\partial \hat{\theta} / \partial s = I^{-1} \dot{\lambda}_{\hat{\theta}}.$$

A similar linear expansion for $a + \delta a$ leads to

$$\begin{aligned} \frac{\delta a}{\delta s} &= \frac{1}{\sqrt{n}} \Lambda_{\hat{\theta}} B_{\hat{\theta}}^T + \frac{1}{\sqrt{n}} I^{-1} \dot{\lambda}_{\hat{\theta}} \{ (s - n\beta_{\hat{\theta}})^T (B_{\hat{\theta}} \dot{\Lambda}_{\hat{\theta}}^T + \dot{B}_{\hat{\theta}} \Lambda_{\hat{\theta}}^T) - n \dot{\beta}_{\hat{\theta}}^T \Lambda_{\hat{\theta}} B_{\hat{\theta}}^T \} \\ &= \frac{1}{\sqrt{n}} \Lambda_{\hat{\theta}} B_{\hat{\theta}}^T + \dot{\lambda}_{\hat{\theta}} D_{\hat{\theta}}, \end{aligned}$$

say, wherein $\Lambda_{\hat{\theta}}$ has its first column $\dot{\lambda}_{\hat{\theta}}$ removed. Combining the expressions for $\partial \hat{\theta} / \partial s$ and $\partial a / \partial s$ we have

$$\begin{aligned}
|\delta(\hat{\theta}, a)/\delta s| &= |I^{-1} \dot{\lambda}_{\hat{\theta}}, \frac{1}{\sqrt{n}} \Lambda_{\hat{\theta}} B_{\hat{\theta}}^T + \dot{\lambda}_{\hat{\theta}}^T D_{\hat{\theta}}| \\
&= |I^{-1} \dot{\lambda}_{\hat{\theta}}, \frac{1}{\sqrt{n}} \Lambda_{\hat{\theta}} B_{\hat{\theta}}^T| \\
&= |\Lambda_{\hat{\theta}}^T \text{diag} (I^{-1}, \frac{1}{\sqrt{n}} B_{\hat{\theta}}^T)|,
\end{aligned}$$

where the last expression $\Lambda_{\hat{\theta}}$ now includes its first column $\dot{\lambda}_{\hat{\theta}}$.

This result can be achieved simply by decomposing δs into a perturbation δs_1 in the plane $\{s : \dot{\lambda}_{\hat{\theta}}^T (s - n\beta_{\hat{\theta}}) = 0\}$ and a perturbation δs_2 parallel to the curve $\{n\beta_{\hat{\theta}}\}$, and noting that $\delta a = \frac{1}{\sqrt{n}} \Lambda_{\hat{\theta}} B_{\hat{\theta}}^T (\delta s_1)$, $\delta \hat{\theta} = \dot{\lambda}_{\hat{\theta}}^T (\delta s_2)$, where $\Lambda_{\hat{\theta}}$ has its first column removed.

REFERENCES

- Barndorff-Neilsen, O. and Cox, D.R. (1979). Edgeworth and saddle-point approximations with statistical applications. J. Roy. Statist. Soc. B, 41 (to appear).
- Cox, D.R. (1980). Local ancillarity. Biometrika, 67 (to appear).
- Efron, B. (1975). Defining the curvature of a statistical problem (with applications to second order efficiency) (with Discussion). Ann. Statist. 3, 1189-242.
- Efron, B. and Hinkley, D.V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. Biometrika, 65, 457-82.
- Fisher, R.A. (1934). Two new properties of mathematical likelihood. Proc. R. Soc. A, 144, 285-307.
- Hinkley, D.V. (1979). Likelihood inference for a simple growth model. Biometrika, 66 (to appear).
- Peers, H.W. (1978). Second-order sufficiency and statistical invariants. Biometrika, 65, 489-96.