The Annals of Statistics
1981, Vol. 9, No. 6, 1328–1333

A COMPARISON OF THE EFRON-HINKLEY ANCILLARY AND THE LIKELIHOOD RATIO ANCILLARY IN A PARTICULAR EXAMPLE

BY BO V. PEDERSEN

Aarhus University

Two approximate ancillaries, the Efron-Hinkley ancillary and the (signed) likelihood ratio, are compared in a specific example by means of their level curves, their marginal densities and the conditional and unconditional likelihood functions. It is shown that in the present model the latter of the two ancillaries is the better. It is almost exactly distribution constant and has more stable conditional likelihood functions.

1. Introduction. In an example with a single observation from a model having a one-dimensional parameter, Barndorff-Nielsen (1980, Figure 1) indicates that remarkably different conclusions might be drawn depending on whether the inference is based on one or the other of the two main, approximate, ancillaries of that paper, namely the Efron-Hinkley (1978) ancillary, a, and a signed version of the log likelihood ratio, $\pm r$. For a comparison of the two it is natural to look at a and $\pm\sqrt{2r}$, as these are asymptotically equivalent for n, the number of observations, tending to infinity.

In some work connected with the former paper, it was seen that the conditional distributions given the two ancillaries would be alike for observations yielding positive or negative but numerically small values of a and $\pm\sqrt{2r}$, while the difference would become more and more striking for observations corresponding to decreasing negative values of a and $\pm\sqrt{2r}$. This is a consequence of the different shapes of the level curves for the two ancillaries, cf. Figure 1 below.

In this note it will be shown that for the cited example, with $n = 1$, $\pm\sqrt{2r}$ is the better of the two ancillaries. It is almost exactly distribution constant and leads to more reasonably behaved conditional likelihoods. This conclusion seems also to be true for moderate n, as illustrated for $n = 5$ by comparing conditional maximum likelihood estimates with unconditional maximum likelihood estimates. As n tends to infinity the difference between the two ancillaries becomes negligible—but so does the effect on conditioning.

2. The model. Let u and $v - 1$ be independent and exponentially distributed with parameters χ and ψ, i.e. with density

$$p(u, v; \chi, \psi) = \chi\psi e^{\psi} e^{-\chi u - \psi v}, \qquad u > 0, v > 1. \tag{1}$$

Consider the submodel, a curved exponential family, determined by $\chi(\psi)\psi e^{\psi} = 1$. This gives

$$p(u, v; \psi) = e^{-\chi(\psi)u - \psi v}, \tag{2}$$

with mean value curve $T_0 = \{\tau(\psi) : \psi > 0\}$, where

$$\tau(\psi) = E_\psi(u, v) = (\psi e^{\psi}, 1 + \psi^{-1}).$$

The maximum likelihood estimator $\hat{\psi}$ in the submodel is the unique solution of

$$v = -\chi'(\psi)u,$$

Received November, 1980; revised April, 1981.

AMS 1970 *subject classification*. Primary 62F99; secondary 62A10.

Key words and phrases. Conditional inference, curved subfamily, Efron-Hinkley ancillary, exponential family, likelihood ratio statistic.

so it is seen that the level curves of $\hat{\psi}$ in the (u, v)-plane are straight lines through (0, 0) with slopes

$$-\chi'(\hat{\psi}) = (1 + \hat{\psi}^{-1})\hat{\psi}^{-1}e^{-\hat{\psi}} > 0.$$

The Efron-Hinkley ancillary, also termed the affine ancillary, is given by $a = (\hat{j}/\hat{i} - 1)/\hat{\gamma}$, where $\hat{j}$, $\hat{i}$ and $\hat{\gamma}$ denote respectively the observed and expected information and the curvature of the model, all evaluated at $\psi = \hat{\psi}$. The log likelihood ratio for testing (2) versus (1) is given by $r = \hat{\hat{\ell}} - \hat{\ell}$, where $\hat{\hat{\ell}}$ and $\hat{\ell}$ denote the maxima of the log likelihood function under (1) and (2), respectively. Defining

$$\Psi_{\pm} = \{1 + (1 + \psi)^{\pm 2}\}^{-1/2}$$

one obtains

$$a = \{v - 1 - \hat{\psi}^{-1}\}\hat{\psi}\hat{\Psi}_{-}^{-1}.$$

Further an expression of r as a function of $(\hat{\psi}, a)$ is given by

$$r = \{\hat{\Psi}_{-} + \hat{\psi}_{+}\}a - \ln\{1 + \hat{\Psi}_{-}a\} - \ln\{1 + \hat{\Psi}_{+}a\}.$$

It is seen that a and r are both zero for an observation on the mean curve T_0. Further, a is positive for an observation above T_0, see Figure 1, and negative in the opposite case. The sign of $\pm r$ and $\pm\sqrt{2r}$ is defined so as to coincide with the sign of a. With this choice of sign, $\pm\sqrt{2r}$ and a are equivalent near T_0 in the sense that the ratio $\pm\sqrt{2r}/a$ tends to one as a or r tends to zero, i.e., as (u, v) tends to a point on T_0. The difference between the two approximate ancillaries a and $\pm\sqrt{2r}$ is illustrated in Figure 1 which shows their level curves in the (u, v)-plane.

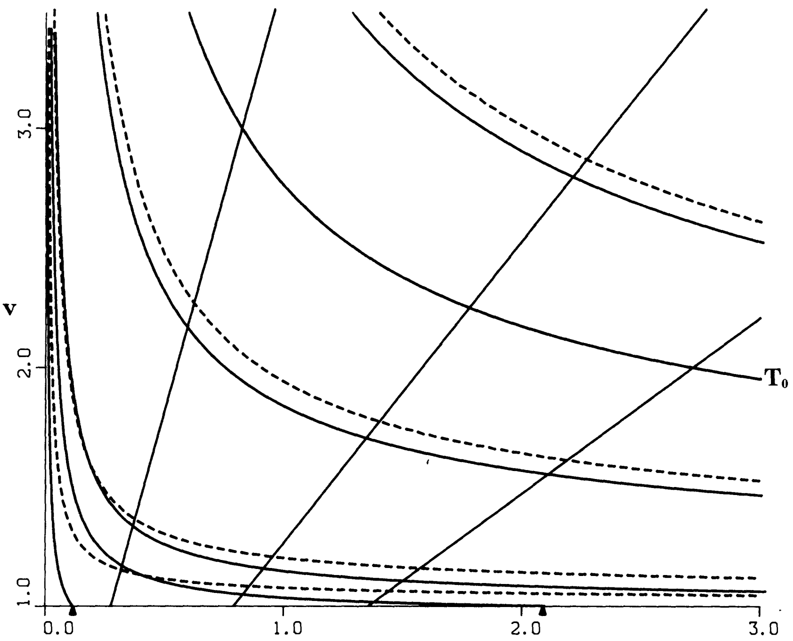

FIG. 1 *Level curves for* $\hat{\psi}$, a *and* $\pm\sqrt{2r}$. *The three straight lines represent, from left to right,* $\hat{\psi} = .5, .8, 1.0$. *The mean curve* T_0 *corresponds to* $a = \pm\sqrt{2r} = 0$. *Further level curves from above correspond to* $a = .5, -.5, -1.0, -1.1, -1.25$, *solid curves, and* $\pm\sqrt{2r} = .5, -.5, -1.5, -2.0$, *broken curves. The arrow heads indicate the points where the level curves for a smaller than* -1 *hit the u-axis, giving* $\hat{\psi} = \psi_a$. *Here* ψ_a *is the upper bound of the possible values of* $\hat{\psi}$ *for the particular a, i.e.* $\psi_a = (a^2 - 1)^{-1/2} - 1$.

The level curves for a and $\pm\sqrt{2r}$ are nearly identical for numerically small values of the two ancillaries. Moreover, the level curves are similar in shape for all positive values. In contrast, the level curves deviate considerably for a slightly smaller than -1. The effect of this on the conditional inference given a or $\pm\sqrt{2r}$ is that for observations yielding positive or negative but numerically small values of the two ancillaries, similar conclusions will be drawn, whereas observations below T_0, yielding negative values of a and $\pm\sqrt{2r}$, may result in quite different conclusions. For instance, for $(u, v) = (0.049, 1.199)$, giving $a = -1.25$ and $\pm\sqrt{2r} = -2.48$, the conditional distribution of $\hat{\psi}$ given a is concentrated on $(0, \psi_a) = (0, .33)$ while if $\psi = 1.0$ most of the mass of the conditional distribution of $\hat{\psi}$ given $\pm\sqrt{2r}$ is concentrated beyond .33.

3. Marginal densities of the ancillaries. Using the general expression for the probability density function of $(\hat{\psi}, a)$ given in Barndorff-Nielsen (1980), one obtains in the present example

$$P(\hat{\psi}, a; \psi) = i(\hat{\psi})^{1/2}\{1 + \hat{\Psi}_{+}a\}e^{\hat{\psi}}e^{-\chi(\psi)/\chi(\hat{\psi})\{1+\hat{\Psi}_{+}a\}-\psi/\hat{\psi}\{1+\hat{\psi}+\hat{\Psi}_{-}a\}}. \tag{3}$$

The probability density function for $(\hat{\psi}, \pm\sqrt{2r})$ is found by noting that for a fixed value of $\hat{\psi}$, the numerical value of the partial derivative of $\pm\sqrt{2r}$ with respect to a is

$$c(\hat{\psi}, a)^{-1} = \frac{|a|}{\sqrt{2r}}\left|\frac{\hat{\Psi}_{+}^{2}}{1 + \hat{\Psi}_{+}a} + \frac{\hat{\Psi}_{-}^{2}}{1 + \hat{\Psi}_{-}a}\right|;$$

here $|a|/\sqrt{2r}$ is defined, by continuity, to be 1 for $(u, v) \in T_0$ i.e. for $a = \pm\sqrt{2r} = 0$. (Notice that for $(u, v) \in T_0$ one has $c(\hat{\psi}, a) = 1$, in accordance with the equivalence of a and $\pm\sqrt{2r}$ for numerically small values.) This gives

$$P(\hat{\psi}, \pm\sqrt{2r}; \psi) = c(\hat{\psi}, \hat{a}_r)\cdot p(\hat{\psi}, \hat{a}_r; \psi),$$

where $\hat{a}_r$ is the unique solution to $\pm r(\hat{\psi}, a) = \pm r$.

The marginal densities of a and $\pm\sqrt{2r}$ are found by numerical integration. For $\psi = .1$ and $\psi = 2.5$ these densities are shown in Figure 2. The two values of ψ selected correspond to rather extreme mean vectors $\tau(\psi)$, namely (0.1, 11) respectively (30.5, 1.4). For ψ varying in between, the marginal densities vary gradually from the former to the latter form.

It is seen that the densities for $\pm\sqrt{2r}$ are almost identical and normal shaped, though not standard normal as the mode varies from $-.6$ to $-.52$. On the other hand the densities for a vary quite a lot and they are not at all normal shaped.

4. Conditional likelihoods. To further illustrate the difference between the two approximate ancillaries the conditional log likelihoods will be considered. Great differences are expected for observations yielding values of $a < -1$. Let us consider observations corresponding to a value of a in this area but not too extreme, say $a = -1.1$. Indexing the observation by $\hat{\psi} \le \psi_a = 1.18$ and comparing with the maxima of the two conditional likelihoods, we get Table 1.

TABLE 1 *Likelihood estimates compared with the conditional maximum likelihood estimates for observations on the curve $a = -1.1$. The quantities $\hat{\psi}_{|a}$ and $\hat{\psi}_{|r}$ are the maxima of the conditional likelihood functions given a and $\pm\sqrt{2r}$, respectively.*

$\hat{\psi}$	.1	.3	.5	.8	1.0
$\hat{\psi}_{\mid a}$	.090	.26	.48	7.1	38.3
$\hat{\psi}_{\mid r}$	.098	.30	.51	.84	1.1

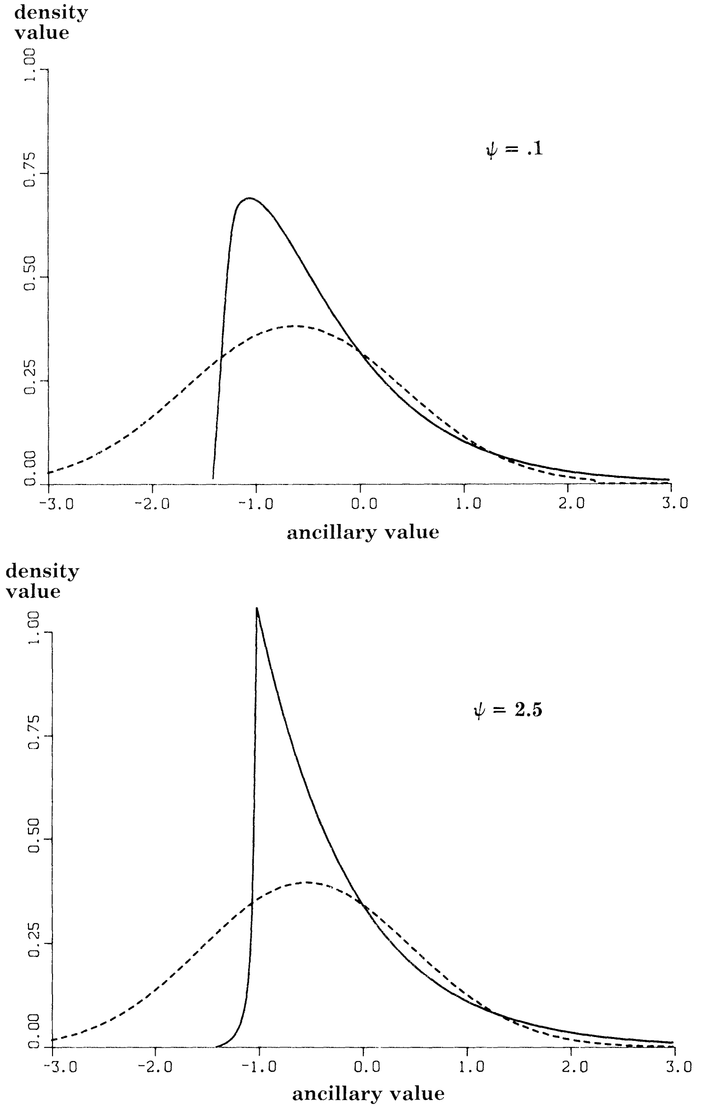

FIG. 2 *The marginal probability densities of a and $\pm\sqrt{2r}$ for $\psi = .1$ and 2.5; a solid curves, $\pm\sqrt{2r}$ broken curves.*

As can be seen from the table, the difference between $\hat{\psi}$ and $\hat{\psi}_{|a}$ becomes drastic as $\hat{\psi}$ approaches ψ_a (for $\boldsymbol{a} = -1.1$, $\psi_a = 1.18$) whereas $\hat{\psi}$ and $\hat{\psi}_{|a}$ are close over the whole range considered. As a further illustration we present in Figure 3 the two conditional log likelihood functions corresponding to $(\hat{\psi}, a) = (1.0, -1.1)$, together with the unconditional log likelihood function. Note that the functions are drawn as functions of $\log \psi$ rather than ψ.

For an exact ancillary, the likelihood function, and in particular the maximum likelihood estimate, would be unaffected by conditioning. This is almost true when conditioning on numerically small or positive values of $\boldsymbol{a}$ and $\pm\sqrt{2r}$. The similarity of the two stems from the equal shaped level curves in this region, cf. Figure 1. Table 1 and Figure 3 illustrate that the effect on the likelihood by conditioning on a small negative value of $\boldsymbol{a}$ may be very considerable.

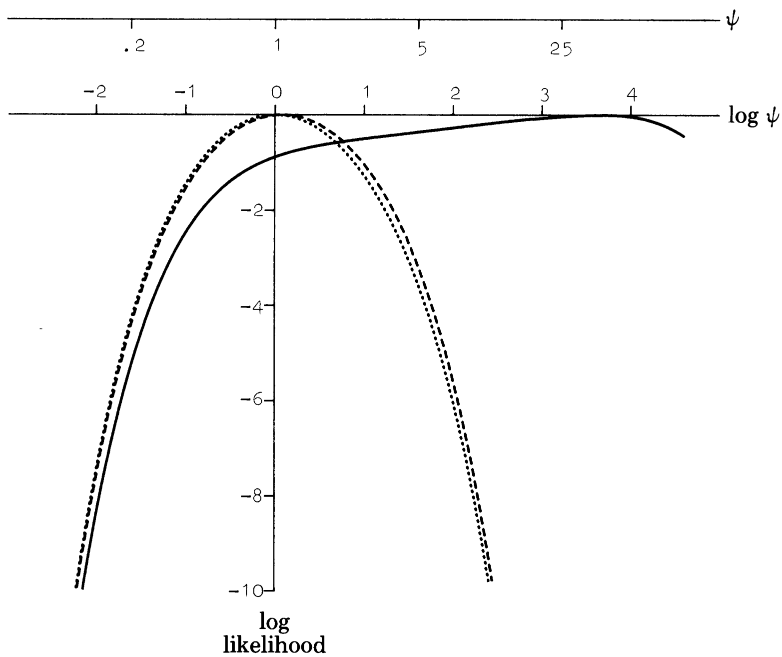

FIG. 3 *The conditional* log *likelihood functions for* log ψ *given a, solid curve, and* $\pm\sqrt{2r}$, *broken curve, together with the unconditional* log *likelihood function, dotted curve* (*natural logarithms*). *The observation* $(u, v) = (1.38, 1.02)$ *corresponds to* $\hat{\psi} = 1.0$, $a = -1.1$, $\pm\sqrt{2r} = -2.58$.

5. Replicated observations. As the choice of a is based on asymptotic theory, let us consider what happens when n independent replicates of (u, v) are available. It is clear that the two ancillaries are functions of $(\bar{u}, \bar{v}) = (n^{-1}\Sigma u_i, n^{-1}\Sigma v_i)$ only. Further it is easily seen that the Efron-Hinkley ancillary and the likelihood ratio statistic satisfy

$$a_n(\bar{u}, \bar{v}) = \sqrt{n}\, a_1(\bar{u}, \bar{v}) \tag{4}$$

and

$$r_n(\bar{u}, \bar{v}) = n\, r_1(\bar{u}, \bar{v}), \tag{5}$$

so that the level curves of $\hat{\psi}$ and the two ancillaries in Figure 1 are left unchanged when (u, v) is replaced by $(\bar{u}, \bar{v})$. The level curve given by $a_1 = a'$ (a' constant) is identical to the level curve $a_n = \sqrt{n}\, a'$, and similarly for $\pm\sqrt{2r_n}$.

The distributions of $(\hat{\psi}, a_n)$ and $(\hat{\psi}, \pm\sqrt{2r_n})$, the marginal distributions of the ancillaries, and the conditional likelihoods are found from the distribution of $(\Sigma u_i, \Sigma(v_i - 1))$ by transformation and by numerical integration exactly as before. The probability density function of the latter distribution is the product of two independent gamma density functions.

Now, as n tends to infinity, the probability of observing $(\bar{u}, \bar{v})$ far from T_0 tends to zero, and so numerically small values of $a_1(\bar{u}, \bar{v})$ and $r_1(\bar{u}, \bar{v})$ will generally be observed, giving almost equivalent conditioning results.

As an illustration take $n = 5$. For moderate numerical values of a_n and $\pm\sqrt{2r_n}$ nice results are obtained – for example for $(\hat{\psi}, a_n) = (1.0, -1.1)$, corresponding to $a_1 = -.49$, the two conditional log likelihoods cannot be distinguished from the unconditional log likelihood, on the scale of Figure 3. Even for $(\hat{\psi}, a_n) = (1.0, -1.96)$ the three curves are almost identical, though the log likelihood conditional on a is slightly more dispersed. However,

TABLE 2 *Maximum likelihood estimates compared with the conditional maximum likelihood estimates for observations* $(\bar{u}, \bar{v})$ *on the curve* $a_n = -2.46$, *with* $n = 5$; *or, equivalently, with* $(\bar{u}, \bar{v})$ *on the curve of Table* 1, *where* $a_1 = -1.1$. *The quantities* $\hat{\psi}_{|a}$ *and* $\hat{\psi}_{|r}$ *are defined as before.*

$\hat{\psi}$	.1	.3	.5	.8	1.0
$\hat{\psi}_{\mid a}$	.11	.40	.97	7.9	39.8
$\hat{\psi}_{\mid r}$	.099	.30	.51	.83	1.0

the troubles start again as a_n becomes smaller than $-\sqrt{5} = -2.24$ or $a_1 < -1$. Taking again $a_1 = -1.1$ or $a_n = -2.46$, which, thinking of a_n as a standard normal variate, is not that extreme, we obtain Table 2 as an analogue of Table 1.

The same picture as that of Table 1 emerges. The difference between $\hat{\psi}_{|a}$ and $\hat{\psi}$ grows for $\hat{\psi}$ approaching $\psi_a = 1.18$, while $\hat{\psi}$ and $\hat{\psi}_{|r}$ are close over the whole range considered. Further it appears that conditioning on a small value of a_n yields a more dispersed likelihood function, especially for $\hat{\psi}$ large; this implies a loss of information.

6. Discussion. In the case of a single observation (u, v) from the model considered, the signed log likelihood ratio statistic is very near to being exactly ancillary whereas the Efron-Hinkley statistic a is not. The shape of the marginal probability density function of a varies substantially in its dependence on ψ, especially for negative values of a indicating loss of information in this region, cf. Figure 2. This is in accordance with the fact that for small values of a $(a < -1)$ the conditional likelihood function given a in many cases takes its maximum far away from the unconditional maximum likelihood estimate and is, moreover, considerably more dispersed than the unconditional likelihood function.

For sample sizes $n > 1$ similar problems are encountered with observations $(\bar{u}, \bar{v})$ yielding small values of a_n $(a_n < -\sqrt{n})$. The import of this non-ancillarity of a_n disappears quickly because the probability of such observations tends to zero as n increases and is small already for $n = 5$.

In conclusion, it can be said that except for very small n, one is likely to obtain observations in the region where the two different ancillaries yield equivalent conditional inferences. However, drawing the inference conditional on a_n when observing $a(\bar{u}, \bar{v})$ for which $a_1(\bar{u}, \bar{v}) < -1$ one will generally be misled with respect to the value and precision of the maximum likelihood estimate.

Acknowledgement. I should like to thank Ole Barndorff-Nielsen for helpful discussions and advice.

REFERENCES

BARNDORFF-NIELSEN, O. (1980): Conditionality resolutions. *Biometrika* **67** 293–310.

EFRON, B. and HINKLEY, D. V. (1978): Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika* **65** 457–482.

DEPARTMENT OF THEORETICAL STATISTICS
INSTITUTE OF MATHEMATICS
UNIVERSITY OF AARHUS
NY MUNKEGADE
DK-8000 AARHUS C, DENMARK