1. INTRODUCTION

Fisher (1934) introduced the conditional analysis of the location scale model. Peisakoff (1951) discussed the more general transformation model from a decision theoretic viewpoint. A general discussion may be found for example in Fraser (1979).

Exponential models may be found in Fisher and provide key examples for his concept of sufficiency. For a recent survey see Barndorff-Nielsen (1978).

The steps in the process of statistical inference for these two model types are opposites in the sense of switching marginal and conditional methods. For a parameter decomposition \( \theta = (\psi, \lambda) \) into an interest and nuisance parameter, let \( y = (y_1, y_2, y_3)' \) be a variable decomposition where in an appropriate way \( y_1 \) is error or parameter-free, \( y_2 \) measures departure from \( \psi \) free of \( \lambda \), and \( y_3 \) measures \( \lambda \) for given \( \psi \). Then for a transformation model, we use \( y_2 | y_1 \) and for an exponential model we use \( y_2 | y_3 \); this switches marginalization and conditioning for each of \( y_1 \) and \( y_3 \). See Fraser (1987) for some general discussion.

Highly accurate third order approximations seem to have originated with the Bartlett (1955) corrections for the likelihood ratio quantities.

For an exponential model, saddlepoint methods (Daniels, 1954; Barndorff-Nielsen & Cox, 1979) gave density approximations and gave (Lugannani & Rice) a distribution function approximation

\[
F(\hat{\theta}; \theta) = \Phi(r) + \varphi(r) \left\{ \frac{1}{r} - \frac{1}{q} \right\}
\]

(1.1)

for testing within a real parameter canonical exponential model; \( r \) is the signed likelihood ratio quantity, \( q \) is the standardized maximum likelihood departure,

\[
r = \text{sgn}(\hat{\theta} - \theta) \cdot \left[ 2(\ell(\hat{\theta}; y) - \ell(\theta; y)) \right]^{1/2}, \quad q = (\hat{\theta} - \theta)^{1/2},
\]

(1.2)

and \( \hat{j} = -\partial^2 \ell(\theta; y)/\partial \theta^2 |_{\theta} \) is the observed information.

For transformation models Taylor series expansion of density functions (DiCiccio, Field and Fraser, 1990) led to a third order approximation for testing a component canonical parameter.
For a general asymptotic context, initial development centred on density approximations (Barndorff-Nielsen, 1983, 1986) provided by the $p^*$ formula, conditional on some third order ancillary statistic. Use of the approximation, however, was hampered by the lack of general methods for constructing the ancillary. For the case of a real parameter, a distribution function approximation was obtained (Barndorff-Nielsen, 1986, 1990) in terms of the $r^*$ adjusted likelihood ratio or a related modification of (1.1); third order accuracy for the modification of (1.1) was indicated in Barndorff-Nielsen (1991). Also for a real parameter model, Taylor series expansions of the model gave (Fraser & Reid, 1989, 1993a) a third order formula for testing the parameter. For the case of a component real parameter in an asymptotic model, an approximation using a modified $q$ for the formula (1.1) was developed (Barndorff-Nielsen 1986, 1991). The formula involved determinants of score related variables; third order accuracy of the modification (1.1) was indicated in Barndorff-Nielsen(1991). The preceding approximations all require a preliminary reduction to a third order ancillary.

Fraser & Reid (1993b) showed that only the tangent directions of a second order ancillary are needed for inference based on a third order ancillary and gave a general construction procedure for the needed ancillary directions. Tangent exponential models were then used to produce a third order approximation for testing a real parameter component.

We survey generally this third order procedure for testing a component real parameter. In Section 2 we describe the overall ancillary giving third order inference. In Section 3 we describe the conditional inference model and the needed directions $V$ for third order inference. Section 4 describes component parameter inference and an example is discussed in Section 5.

2. ELIMINATING ERROR

We will refer to a component variable as error if it has a distribution free of the parameter. Such a variable can be thought of as measuring some basic randomness of the
distribution and as such could be used for tests of the overall model structure. A simple example from location model theory has $y_1, y_2$ with distribution $f(y - \theta)$ and can be written $y_1 = \theta + e_1$, $y_2 = \theta + e_2$ with $e_1, e_2$ distributed as $f(e)$. We see that $y_1 - y_2 = e_1 - e_2$ is 'measuring' an aspect of the error distribution.

As mentioned earlier, error arises in a different manner with transformation and exponential models. Let $y = (y_1, y_2)$ where $y_1$ describes error and $y_2$ in some sense measures the full parameter, say $\theta$. With a transformation model, the marginal of $y_1$ is $\theta$-free and $y_2 | y_1$ is used to examine $\theta$; with an exponential model, $y_2$ is a sufficient statistic and $y_1 | y_2$ is $\theta$-free. Another way of looking at the second model comes by defining a variable $a$ be the probability integral transform of $y_1 | y_2$; this $a$ has conditionally a uniform distribution and thus has marginally a uniform distribution. We could then analyze the exponential model using $(a, y_2)$ with $a$ having a $\theta$-free distribution and $y_2 | a$ being used to measure $\theta$. In this form, the second case becomes a special version of the first, where the distribution of $y_2 | a$ is in fact, independent of $a$. Thus we see that one of the differences between the two types of model can be ignored and we choose here to follow the transformation model pattern.

Consider a continuous model $f(y; \theta) = f_n(y_1, \ldots, y_n; \theta)$ with parameter $\theta$ of dimension $p$ and with asymptotic properties as $n$ becomes large; for details see Fraser & Reid (1993b). From the transformation model viewpoint, we seek an ancillary variable $a$ of dimension $n - p$ for measuring error and a conditional variable $s$ of dimensional $p$ for assessing the parameter $\theta$. The discussion (ibid) indicates that many possibilities exist for $a$ but each has the same distribution when projected to the observed maximum likelihood surface $\hat{\theta} = \hat{\theta}^0$. We first describe this distribution and then in the next section discuss the conditional distribution for assessing $\theta$.

With the conditional variable $s$ measuring $\theta$, we would have a one-one correspondence (at least asymptotically) between $s$ and $\hat{\theta}$. It follows that the surface $\hat{\theta}(y) = \hat{\theta}^0$ for some $\hat{\theta}^0$ is cross sectional to the surfaces $a(y)$ and accordingly we can use the points on $\hat{\theta}(y) = \hat{\theta}^0$.
to index the surfaces $a(y)$. In fact we'll exhibit the distribution of $a(y)$ by projecting it to
the surface $\hat{\theta}(y) = \hat{\theta}^0$.

Let $dy = dy_t + dy_p$ be a vector increment represented as $dy_t$ on the surface $\hat{\theta}(y) = \hat{\theta}^0$
and $dy_p$ perpendicular to the surface. Then the score equation for $\hat{\theta}$ gives the volume relationship

$$dy_p = |\ell_{\theta;y}(\hat{\theta}; y)|^{-1} |\hat{\theta}| d\hat{\theta}$$  \hspace{1cm} (2.1)

where $\ell_{\theta;y}$ is the $p \times n$ gradient $\partial^2 \ell / \partial \theta \partial y'$ of the score and the determinant corresponds
to $|X| = |X'X|^{1/2}$ as used in regression analysis. Some marginal and conditional analysis
(ibid) then gives the distribution

$$\frac{(2\pi)^{p/2}}{c} \exp\{\ell(\hat{\theta}; y)\} |\ell_{\theta;y}(\hat{\theta}; y)|^{-1} |\hat{\theta}|^{1/2} dy_t$$  \hspace{1cm} (2.2)

on the surface $\hat{\theta}(y) = \hat{\theta}^0$. The distribution is determined to order $O(n^{-3/2})$ and the
constant $c = 1$ to order $O(n^{-1})$. Note that it does not depend on the choice of the
ancillary $a(y)$.

The conditional distribution for assessing $\theta$ given some third order ancillary $a(y)$ will
then exist on the surface $a(y) = a^0$ through the observed $y^0$ and has element

$$\frac{c}{(2\pi)^{p/2}} \exp\{\ell(\theta; y) - \ell(\hat{\theta}; y)\}^{-1} |\hat{\theta}|^{1/2} d\hat{\theta}$$  \hspace{1cm} (2.3)

where the density is evaluated along the particular ancillary surface. This is Barndorff-Nielsen's (1983,1986) $p^*$ formula for the conditional distribution of the maximum likelihood estimator, and for most uses needs a determination of the ancillary $a(y)$. We discuss the
construction of the ancillary in the next section.

3. ASSESSING THE FULL PARAMETER

Given an ancillary $a(y)$, we have an asymptotic model (2.3) with $p$ dimensional parameter $\theta$ and $p$ dimensional variable $y|a = a^0$, and we have an observed value $y|a^0 = y^0|a^0$. 

31
The discussions in Fraser & Reid (1993a) and in Abebe, Cakmak, Cheah, Fraser, Kuhn, Reid (1993) show how such a model can be approximated by a location model or by an exponential model, to the third order in a first derivative neighbourhood of the data and to second order generally. The analysis in Fraser (1988, 1990), Fraser & Reid (1989, 1993a) shows that key probabilities for statistical inference can be calculated from the exponential model approximation, with full third order accuracy using versions of (1.1). From Fraser (1988) we can determine the approximating exponential model using only the likelihood at the data point and the derivative of the likelihood at the data point. As likelihood marginally and conditionally given an ancillary is the same, it follows that we only need the observed likelihood and its gradient tangent to the ancillary surface at the data point in order to calculate the approximating exponential model which in turn produces third order probabilities for key inference quantities. Such an approximating model is called the tangent exponential model and its inference role is discussed in Fraser & Reid (1993b).

This then would seem to require that we need the tangent directions to an appropriate third order ancillary at the observed $y^0$. Analysis in Fraser & Reid (1993b) shows that the same accuracy for likelihood and likelihood gradient can be obtained from the tangent directions to a second order ancillary.

For ancillaries in the continuous case, the location and transformation models provide the prime construction procedures. For example, with $y_1, y_2$ from $f(y - \theta)$ or with $y_1 = \theta + e_1, y_2 = \theta + e_2$ and $e_1, e_2$ from $f(e)$, we see that a change of $\theta$ from $\theta^0$ to $\theta^0 + d\theta$ can be thought of as moving a $(y_1, y_2)$ value to $(y_1 + d\theta, y_2 + d\theta)$. This notion was extended in Fraser (1964) to sampling from stochastically monotone distributions. More generally (Fraser & Reid 1993b) we can have a probability movement or transport corresponding to a shift in the parameter from $\theta^0$ to $\theta^0 + d\theta^0$. Under mild conditions, this probability movement corresponds to a first order ancillary. The procedure gives a tangent direction for a first order ancillary but the discussion (ibid) shows that there is a second order ancillary with the same tangent directions. We thus have the needed directions in which
to calculate the likelihood gradient and obtain third order inference as discussed earlier in this section.

Consider now the methodology connected with the tangent directions $V = (v_1, \ldots, v_p)$ to an ancillary at a data point $y^0$. For the simple case where independent coordinates $y_i$ each have a monotone distribution in a real parameter $\eta_i$ the data movement can be calculated from Fraser (1964). Let $\hat{\eta}_i^0$ be the maximum likelihood estimate for $\eta_i$ from within the full model. Then a change $\hat{\eta}_i^0$ to $\hat{\eta}_i^0 + \delta$ gives a displacement

$$
dy_i = \left. \frac{-\partial F_i(y_i; \eta)}{\partial F_i(y_i; \eta) / \partial \eta} \right|_{\hat{\eta}_i^0} \cdot \delta = d(y_i; \hat{\eta}_i^0) \cdot \delta;
$$

where $\hat{\eta}_i^0 = \hat{\eta}_i(\hat{\theta}^0)$; this is obtained from the total differential of the distribution function $F_i(y_i; \eta)$ and can be interpreted in terms of the shift $dy_i$ of the distribution corresponding to the parameter change $\hat{\eta}_i^0$ to $\hat{\eta}_i^0 + \delta$.

Now suppose the components $\eta_i$ are expressed in terms of the overall parameter $\theta$ by a link function

$$
\eta = g(\theta)
$$

expressing an n-vector in terms of a p vector. A change from $\hat{\theta}^0$ to $\hat{\theta}^0 + \delta$ thus gives the change

$$
dy = D(y_i; \hat{\eta}_i^0)g'(\hat{\theta}^0)\delta = V\delta
$$

where $D$ is a diagonal matrix with elements $d(y_i; \hat{\eta}_i^0)$ and $g'(\theta) = \partial g(\theta) / \partial \theta'$ is the link Jacobian. The array $V = (v_1, \ldots, v_p)$ gives the tangent vectors for the location ancillary generated at the maximum likelihood value $\hat{\theta}^0$.

The tangent exponential model approximating the conditional model given the third order ancillary related to the preceding location ancillary can be expressed in terms of the observed likelihood

$$
\ell(\theta) = \ell(\theta; y^0),
$$

and the gradient of the likelihood in the directions tangent to the location ancillary

$$
\varphi(\theta) = \ell_V(\theta; y^0) = \left[ \frac{d}{dv_1} \ell(\theta; y), \ldots, \frac{d}{dv_p} \ell(\theta; y) \right]_{y^0}
$$

33
The saddlepoint expression for the exponential model is

\[
\frac{c}{(2\pi)^{p/2}} \exp \{ \ell(\theta) - \ell(\hat{\theta}^0) + (\varphi - \varphi^0)s \} j_{\varphi}^{-1/2} ds = \frac{c}{(2\pi)^{p/2}} \exp \{ \ell(\theta) - \ell(\hat{\theta}^0) + (\varphi - \varphi^0)s \} j_{\varphi}^{-1/2} d\varphi
\] (3.6)

where \( j \) is observed information function calculated from the tilted likelihood in the exponent. The data value corresponds to \( s = 0 \) with maximum likelihood value \( \varphi^0 = \varphi(\hat{\theta}^0) \) for the canonical parameter.

Various forms of third order inference can be obtained from this tangent model.

4. INFERENCE FOR AN INTEREST PARAMETER

In the preceding section we discussed the model reduction that gives a variable of the same dimension \( p \) as that of the parameter; the approximating model at the data point is given by (3.6).

Now consider an interest parameter \( \psi \) of dimension \( r \). For fixed \( \psi \), the constrained parameter \( \theta_\psi \) describes the nuisance parameter and we find it convenient to use the notation \( \theta = (\lambda, \psi) \) with \( \lambda \) providing coordinates for \( \theta_\psi \) for each \( \psi \) value.

The transformation and exponential models provide quite different patterns for inference concerning \( \psi \). We assume that the overall error has been eliminated. Then for fixed \( \psi \), let \( y_2 \) describe the error or \( \lambda \) free variation and \( y_3 \) measure in some sense the nuisance parameter \( \lambda \). For the transformation model, the marginal of \( y_2 \) is \( \lambda \)-free and \( y_3 | y_2 \) is available for \( \lambda \); for the exponential model, \( y_3 \) is a sufficient statistic and \( y_2 | y_3 \) is \( \lambda \)-free. The situation here is similar to that at the beginning of Section 2, and in a similar way we can recast the second pattern as a special case of the first. We are thus able to conform to both inference patterns by seeking a marginal variable that is \( \lambda \)-free.

In Section 2 we examined the ancillary distribution as projected to the observed maximum likelihood surface; in computational form (Fraser & Reid, 1993b) we can obtain this by factoring out the conditional distribution for the maximum likelihood estimate of
the parameter; see (2.2), (2.3). We follow the same procedure here but now work within
the reduced model (2.3) using \( \theta_{\psi} \) or \( \lambda \) as the operative parameter with \( \psi \) fixed; the full
distribution concerning \( \hat{\theta} \) is

\[
\frac{c}{(2\pi)^{p/2}} \exp\{\ell(\theta_{\psi}; y) - \ell(\hat{\theta}; y)\} |j|^{1/2} d\hat{\theta} \\
= \frac{c}{(2\pi)^{p/2}} \exp\{\ell(\theta_{\psi}; y) - \ell(\hat{\theta}; y)\} |j|^{-1/2} |\ell_{\theta; V}(\hat{\theta}; y)| ds
\]

(4.1)

where the second expression uses coordinates \( s \) with respect to the vector \( V \).

The conditional distribution for measuring \( \lambda \) is available by appropriate replacement
in (2.3),

\[
\frac{c'}{(2\pi)^{(p-r)/2}} \exp\{\ell(\theta_{\psi}; y) - \ell(\hat{\theta}_{\psi}; y)\} |\lambda|^{1/2} d\lambda \\
= \frac{c'}{(2\pi)^{(p-r)/2}} \exp\{\ell(\theta_{\psi}; y) - \ell(\hat{\theta}_{\psi}; y)\} |\lambda_{\psi}(\hat{\theta}_{\psi})|^{-1/2} |\ell_{\lambda; V}(\hat{\theta}_{\psi}; y)| ds_p
\]

(4.2)

where \( ds_p \) is \( p-r \) dimensional volume (in the coordinates re \( V \)) as calculated perpendicular
to the maximum likelihood surface \( \lambda_{\psi} = \lambda_{\psi}^0 \) constant.

The marginal distribution for the error with \( \psi \) fixed is then obtained by dividing (4.1)
by (4.2)

\[
\frac{c''}{(2\pi)^{r/2}} \exp\{\ell(\hat{\theta}_{\psi}; y) - \ell(\hat{\theta}; y)\} |j|^{-1/2} |\ell_{\theta; V}(\hat{\theta}; y)| \\
\cdot \frac{|\lambda|^{-1/2} |\ell_{\lambda; V}(\hat{\theta}_{\psi}; y)|}{|\lambda_{\psi}(\hat{\theta}_{\psi})|^{-1/2} |\ell_{\lambda; V}(\hat{\theta}_{\psi}; y)|} ds_t
\]

(4.3)

where \( ds_t \) is volume on the maximum likelihood surface \( \lambda_{\psi} = \lambda_{\psi}^0 \) examined in the \( V \)-based
coordinates.

Now consider the use of a tangent exponential approximation to calculate probabilities
for the distribution (4.3). We view (4.3) as having a fixed value for \( \psi \) and yet in developing
the approximation we will use an exponential type parameter for which a particular value
gives the distribution (4.3) of interest.

Consider the reparameterization \( \varphi(\theta) \) at the observed \( y^0 \) as given by (3.5). Examining
parameter change near \( \hat{\theta}_{\psi}^0 \) we obtain

\[
d\varphi = (d\lambda, d\psi) \frac{\partial \varphi(\theta)}{\partial \theta} \bigg|_{\hat{\theta}_{\psi}^0} = (d\lambda, d\psi) \varphi(\hat{\theta}_{\psi}^0) \\
d\psi = d\varphi \cdot \psi(\hat{\theta}_{\psi}^0)
\]

(4.4)
where the inverse matrix $\varphi_\psi^{-1}$ is written $\varphi^0 = (\varphi^\lambda, \varphi^\psi)$. Thus specifying $\psi$ near $\hat{\theta}_\psi^0$ is equivalent to $d\psi = 0$ or to $d\varphi \cdot \varphi^\psi(\hat{\theta}^0) = 0$ as seen from (4.4).

To develop the distribution theory it is notionally convenient to assume that the tangent vectors for the ancillary have been replaced by a rotated set so that the canonical parameter $\varphi = (\varphi_1, \varphi_2)$ separates like $\theta = (\lambda, \psi)$ in the neighbourhood of $\hat{\theta}_\psi^0$, in particular so that $d\varphi_2 = 0$ is equivalent to $d\psi = 0$ at $\hat{\theta}_\psi^0$. It follows that the surface $\hat{\theta}_\psi = \hat{\theta}_\psi^0$ is the same as $\hat{\theta}_\psi = \hat{\theta}_\varphi^0$, for the particular $\psi$ value and its corresponding $\varphi_2$ value say $\varphi_{20}$.

The tangent exponential model for the full distribution is given by (3.6),

$$
\frac{c}{(2\pi)^{p/2}} \exp\{\ell(\theta) - \ell(\hat{\theta}^0) + (\varphi - \varphi^0)s\} |\bar{J}_{\psi}|^{-1/2} \, ds,
$$

where $\theta$ is viewed as a function of $\varphi$, $\bar{J}$ is calculated from the tilted likelihood, and $s$ gives coordinates with respect to $V$ at $\gamma^0$.

In the sample space of the reduced model described by (4.1) let $C$ be the curve or surface along which $\lambda_\psi = \hat{\lambda}_\psi^0$. For the exponential approximation (4.5) in terms of the special $\varphi$ parameterization, the surface $C$ is the profile contour supporting profile likelihood for the component $\varphi_2$; for some background details see Cheah, Fraser & Reid (1991).

Now consider some reparameterization $\alpha = (\alpha_1, \alpha_2)$ such that $\alpha_2 = \alpha_{20}$ gives the same contour as the tested $\psi$ value and other contours of $\alpha_2$ coincide to first derivative with those for $\varphi_2$ at points on the profile curve $C$.

The conditional density concerning $\alpha_1$ at its maximum value on the profile curve $C$ with $\alpha_1 = \hat{\alpha}_1 \alpha_2$ is

$$
\frac{c'}{(2\pi)^{(p-r)/2}} \exp\{\ell(\alpha_1 \alpha_2, \alpha_2) - \ell(\alpha_1 \alpha_2, \alpha_2)^0\} |\bar{J}(\alpha_1)|^{-1/2} \, ds_1,
$$

where the information is calculated for the nuisance parameter expressed as $\varphi_1$ with $\alpha_2$ fixed.

The marginal density projected to the profile curve $C$ is obtained as the quotient of (4.5) by (4.6) using the parameter value $(\hat{\varphi}_{2\varphi_2}, \varphi_2)$ on the profile curve $C$:

$$
\frac{c''}{(2\pi)^{r/2}} \exp\{\ell(\theta_\psi) - \ell(\hat{\theta}^0) + (\varphi_2 - \varphi^0_2)s_2\} \frac{|\bar{J}_1(\varphi)|^{-1/2}}{|\bar{J}_{\psi}(-1/2) \, ds_2}
$$

36
where component information is calculated with respect to \( \varphi_1 \) with \( \alpha_2 \) constant. The quotient of informations in (4.7) has the quotient form in (4.3) with \( y = y^0 \).

For the special case \( r = 1 \) of a real interest parameter \( \psi \), the significance \( p(\psi) \) at the observed data joint \( y^0 \) can be calculated from the generalization of (1.1) discussed in Fraser & Reid (1993b) for a modulation of an exponential model; this uses the signed profile likelihood ratio

\[
r = \text{sgn}(\dot{\psi}^0 - \dot{\psi}) [2 \{ \ell(\hat{\theta}^0) - \ell(\hat{\theta}_\psi) \}]^{1/2},
\]

and a maximum likelihood departure standardized by a quotient-calculated profile information

\[
q = Q = \frac{(\dot{\varphi}^0 - \varphi) \varphi^\psi(\dot{\theta}^0_\varphi)}{|\varphi(\dot{\theta}^0_\varphi)|} \cdot \frac{|f|^{1/2} |\ell_\psi, V(\hat{\theta}; y^0)|^{-1}}{|f(\hat{\theta}, V(\hat{\theta}; y^0)|^{-1}}; \tag{4.9}
\]

see (4.3) and (4.4). This provides \( O(n^{-3/2}) \) accuracy.

5. EXAMPLES

Consider independent variables \( y_1, \ldots, y_n \) where \( y_i \) has density \( f_i(y_i; \theta_i) \) which is stochastically increasing in terms of a real parameter \( \theta_i \). We assume that \( \theta_i = g_i(\beta_1, \ldots, \beta_p) \) is linked to the parameter \( \beta = (\beta_1, \ldots, \beta_p) \) by a link function \( g_i \). As a special case we could have \( f_i = \exp\{\theta_i y_i - c(\theta_i)\} h(y_i) \) and \( \theta_i = g(X_i \beta) \) where \( X_i = (x_{i1}, \ldots, x_{ip}) \) records values for concomitant variables; this is a generalized linear model with noncanonical link function \( g \).

The methods from Fraser (1964) and Fraser, Monette, Ng (1992) give the directions \( V = (v_1, \ldots, v_p) \) for the ancillary described in Sections 2, 3. Let \( F_i \) be the distribution function for \( f_i \); then

\[
d y_i = \frac{-\partial F_i(y_i; \theta)}{\partial F_i(y_i; \theta)/d y_i} d \theta = d(y_i; \theta) \ d \theta
\]

is the change in \( y_i \) induced by a change \( d \theta \) in \( \theta \). The directions \( V \) at the maximum likelihood value \( \hat{\beta}^0 \) are

\[
V = \text{diag}[d_i(y_i^0, g_i(\hat{\beta}^0))] \frac{\partial \theta^r}{\partial \beta} (\hat{\beta}^0). \tag{5.2}
\]
The canonical parameters for the approximating exponential model are

$$
\varphi = \frac{d}{dV} \ell(\beta; y)|_{y^0}.
$$

(5)

Significance $p(\psi)$ for a scalar inference parameter $\psi$ is then obtained from (1.1) with (4.1) and (4.9).

For a numerical example, consider Example 1 from Cox & Snell (1981). The response $y$ is lifetime in weeks for leukemia patients and the concomitant variable is the logarithm $x$ of the initial white blood cell count. The indicated model is exponential

$$
f_i(y_i) = \exp\{-\theta_i y_i + \log \theta_i\}
$$
on $y_i > 0$ and the link is given by

$$
E(y_i) = \theta_i^{-1} = \exp\{\alpha + \beta(x_i - \bar{x})\}
$$

where the canonical parameters $\theta_i$ are nonlinear in $(\alpha, \beta)$.

The example has special features which allow an exact analysis for $\alpha$ and $\beta$. $w_i = \log y_i$; then $w_i = \alpha + \beta(x_i - \bar{x}) + t_i$ where $t_i$ has the extreme value distribution $\exp\{t - e^t\}$: the model is a location model in $\alpha$ and $\beta$; see Fraser (1979) and Law (1982).

The conditional density for $(\hat{\alpha}, \hat{\beta})$ can then be obtained from location model theory (Fraser, 1979)

$$
f(\hat{\alpha}, \hat{\beta}; \alpha, \beta) = cL^0(\alpha - \hat{\alpha} + \alpha^0, \beta - \hat{\beta} + \beta^0);
$$

the exact significance for $\beta$ is

$$
p(\beta) = P(\hat{\beta} \leq \beta^0; \beta) = \int_{-\infty}^{\infty} \int_{\beta}^{\infty} cL^0(\alpha, \gamma) \, d\alpha \, d\gamma
$$

where the norming constant $c$ and the probabilities can be obtained by numerical integration.

38
The third order approximation for $p(\beta)$ obtained from (1.1) with (4.8) and (4.9) superimposes the exact value to the graphing accuracy; see Fraser, Monette, Ng (ibid).

REFERENCES


Cheah, P.K. Fraser, D.A.S., Reid, N. (1991) Multiparameter testing in exponential models: Third order approximations from likelihood, in revision for *Biometrika*.


