Separating error, nuisance effect and main effect: Tangent models and third order inference

D.A.S. Fraser and N. Reid
Department of Statistics
University of Toronto
Toronto, Canada, M5S 1A1

SUMMARY

Normal linear regression provides a standardized procedure for separating error, nuisance effect, and main effect. A parallel procedure is discussed for the third order asymptotic context; this provides a resolution of apparently contradictory procedures found with exponential and transformation models. The analysis uses tangent location and tangent exponential models.

1. INTRODUCTION

The procedure for separating error, nuisance effect, and main effect is well standardized in the context of normal linear regression models. More generally, with exponential linear models the procedure is to *marginalize* to separate effects and *condition* to get the main effect, with transformation models is to *condition* to separate effects and *marginalize* to get main effect. Recent asymptotic and ancillary methods provide clear procedures that correspond closely to the second pattern but also provide an alternative and compatible interpretation for the first pattern.

The recent asymptotic methods of inference are available primarily for the case with variable and parameter of the same dimension. This special situation arises with exponential and transformation models but typically not more generally. The methods can however be applied easily in the presence of an approximate third order ancillary.

A general construction procedure for approximate third order ancillaries is surveyed in Sections 2, 3 and 4.

Tangent location models are introduced in Section A and their use to construct ancillaries and separate error from effect is outlined in Sections 2, 3, 4, 5. Tangent exponential models are introduced in Section B and their use to separate main effect from nuisance effects and to approximate the related p values is surveyed in Sections 6, 7, 8, 9, 10.

A. TANGENT LOCATION MODELS

Location and transformation models lead to a marginal distribution for error and a conditional distribution for effects. A central technique for the asymptotic context uses a tangent location model for a scalar parameter change. This model fits the given model for all values of the variable and for θ at θ_0 and to first derivative at θ_0 where $\theta_0 = \hat{\theta}(y^0)$ is the observed maximum likelihood estimate. The model is called the tangent location model at θ_0. The tangent location-model technique leads to the construction of an approximate ancillary for third order inference.

2. FIRST DERIVATIVE ANCILLARIES

A first derivative ancillary $a(y)$ at the parameter value θ_0 has the property that its distribution has a zero derivative with respect to θ at θ_0: $\partial/\partial\theta|_{\theta_0} f(a;\theta) = 0$. Of particular interest is such an ancillary of dimension $n-p$ where the parameter has dimension p.

For the case of a scalar parameter $p=1$, the construction of a first derivative ancillary of dimension $p-1$ was developed in [1], as a basis for approximate conditional inference. This provides a basic technique for obtaining the approximate ancillaries needed for general asymptotic inference.

For the case of a vector parameter $p>1$ first derivative ancillaries are typically not available, a consequence of integrability conditions not generally being fulfilled with compound vector fields: an exception, however, arises with certain types of linear structure as with generalized linear models [19]. More generally, the methods for the $p=1$ case can be applied to scalar or directional departures from θ_0 and produce second order accurate first derivative ancillaries; see [24]; these are ancillaries where the condition $\partial/\partial\theta|_{\theta_0} f(a;\theta) = 0$ holds to the second order.

3. SECOND ORDER ANCILLARIES

If θ is examined to the second order about θ_0 we obtain a second order description of the model when $\theta_0 = \hat{\theta}(y^0)$. The analysis in [24] and [31] shows that the first derivative ancillary just described can be adjusted to give a second order ancillary and that this can be done without changing the ancillary's tangent plane at the data point y^0. The method is to modify by a pure quadratic term at the data point; this modification has of course zero first derivative at the data point, thus not altering the tangent plane.

We thus have that the tangent plane to the first derivative ancillary is in fact the tangent plane to a second order ancillary.

4. THIRD ORDER ANCILLARY INFORMATION

Skovgaard (1986) has shown that a second order ancillary can be upgraded to give a third order ancillary; the techniques mentioned in the preceding section can allow also be used to show this. The actual upgrading however is not needed for third order statistical inference.

We will see in Sections 6-10 that a tangent exponential model can be used to give third order p values for scalar and vector parameters. For this, it is shown in [7], [8], [24] that a third order tangent exponential model at the data point can be constructed using only (i) the observed likelihood at the data point and (ii) the gradient of the likelihood in the tangent plane of a third order ancillary; also in [24] it is shown that such gradient is in fact determined to third order accuracy using only the tangent plane to a related second order ancillary at the data point y^0.

Thus for third order inference we need only the observed likelihood

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

and an appropriate gradient

$$\begin{aligned}\varphi(\theta) &= \frac{d}{dV}\ell(\theta; y)\Big|_{y^0} \\ &= \left\{ \frac{d}{dv_1}\ell(\theta; y)\Big|_{y^0}, \ldots, \frac{d}{v_p}\ell(\theta; y)\Big|_{y^0} \right\}\end{aligned}$$

of the likelihood at the data point. The vectors $v_1, \ldots, v_p$ are tangent to first derivative ancillaries for parameter changes in linearly independent directions at $\theta_0 = \hat{\theta}(y^0)$. In some generality these directions can be obtained from an invertible pivotal quantity where differentiation is for fixed value of the pivotal quantity;

$$V = (v_1, \ldots, v_p) = \frac{\partial y}{\partial \theta'}\Big|_{(y^0;\theta_0)} .$$

This can be called parameter forcing. For examples, see [19], [25], [28], [30].

5. A NOTE ON ANCILLARIES

An ancillary has a fixed or an approximately fixed distribution. This is common to all the various definitions for the concept. Fisher, it seems clear, attributed further properties to his definition of ancillaries and there have been many divergent views on what is best. In some central way, the choice of ancillary should be sensible. For a view from this direction see [2].

The first derivative ancillaries in Section 2 are not unique, using only the fixed distribution requirement. However, from a stronger viewpoint that the component variables in a basic sense are measuring the parameters, there is essential uniqueness in some generality.

This measurement approach can be applied in certain contexts where the fixed distribution requirement does not hold; see [15]. The central idea then is that component processes are measuring the parameter of interest.

B. TANGENT EXPONENTIAL MODELS

Exponential models lead to a conditional distribution for testing a component canonical parameter; they also have close ties to moment and cumulant generating functions and to saddlepoint approximation methods.

We use an approximating exponential model to summarize basic model information at and in a neighbourhood of a data point. This summary information can lead to third order p-values for testing component parameters. The approximating model is defined for the case of variable and parameter of the same dimension, as is obtained say after the dimension reduction described in the preceding section.

The tangent exponential model approximates the given model for the parameter nominally for all values and for variable only at both the data point y^0 and to first derivative at

y^0. It is a curious event that this approximating model locally defined on the sample space is able to provide sufficient information for third order p values. This however becomes clearer when we examine an asymptotic expansion of the statistical model about the point $(y^0, \hat{\theta}(y^0))$.

6. ASYMPTOTIC EXPANSIONS OF STATISTICAL MODELS

A statistical model with scalar variable y and scalar parameter θ can have asymptotic properties as a result of marginalization or conditioning from a large sample context. The local form of an asymptotic exponential model was examined in [12], [17], [14]. The latter reference also examines the asymptotic connections among the likelihood ratio, the standardized score and the standardized maximum likelihood variable.

The local form of a general asymptotic model was then examined in [22] from a parameter centered viewpoint and in [23] from a data centered viewpoint; formulas in each case were presented for the connections among the likelihood ratio, score, and maximum likelihood quantities. The vector variable and parameter case was discussed in [17] and [20].

The asymptotic expansions for the general asymptotic model in [20], [22], [23] show that it can be closely approximated by a location model or by an exponential model. A data-point related measure of departure from exponentiality is developed which relates closely to Efron's measure of departure; a similar measure of departure from location form is also obtained.

The exponential approximation is most useful for obtaining approximate distribution function values at the data point. Also, the location approximation provides an informative presentation of likelihood, as it gives p-values when integrated.

7. TANGENT EXPONENTIAL MODEL

Consider an asymptotic model where the variable and parameter have the same dimension p. It was shown in [7] and [8] that the likelihood $\ell^0(\theta) = \ell(\theta; y^0)$ and the gradient of likelihood $(\partial/\partial y)\ell(\theta; y)|_{y_0}$ uniquely determine the model if it is exponential and more generally uniquely determine the best exponential model approximation at and near the data point y^0. The saddlepoint version of the approximating exponential model can then be written

$$\begin{aligned} g(y;\theta)dy &= \frac{c}{(2\pi)^{p/2}} \ \exp\{\ell^0(\theta) + \varphi' s\}|\tilde{\jmath}|^{-1/2} ds \\ &= \frac{c}{(2\pi)^{p/2}} \ \exp\{\ell^0(\theta) + \varphi' s\}|\tilde{\jmath}|^{1/2} d\hat{\varphi} \end{aligned}$$

to third order; the constant $c = 1 + O(n^{-1})$ is determined by $\ell^0(\theta)$ and $\varphi(\theta)$; the observed information is calculated using φ tilts of $\ell^0(\theta)$. The tangent exponential model determined by just $\ell^0(\theta)$ and $\varphi(\theta)$ contains sufficient information to provide third order distribution function values at the data point y^0; see [20], [22], and [24].

The preceding tangent exponential model provides an interesting generalization of Barndorff-Nielsen's p^* formula: the p^* formula provides an approximation at a data point; the tangent exponential model provides third order approximation in the neighbourhood of a data point (first derivative), second order accuracy in a compact range for the standardized variable, and third order accuracy for distribution function values at the data point. For a detailed development see [24].

8. DISTRIBUTION FUNCTION FOR THE TANGENT EXPONENTIAL MODEL

Consider a general asymptotic model $f(y;\theta)$ with scalar variable and parameter; and let $g(y;\theta)$ be the corresponding exponential model (Section 7) that is tangent at the data point y^0. Then it was shown in [12] and [17] that the distribution functions coincide to third order at y^0. The proof given is long but a simple proof is available by integration of the discrepancy; this uses only the standard normal density and distribution functions together with integration by parts.

9. TESTING A SCALAR PARAMETER

Consider a general asymptotic model with p dimensional variable and parameter, and suppose that interest centers on testing a scalar parameter component $\psi = \psi(\theta)$ where θ can be rewritten say as (λ, ψ) with $p-1$ dimensional nuisance parameter λ.

With ψ fixed the theory in Sections 2-5 show that there is a third order ancillary of dimension 1. The values of this ancillary are most easily indexed by the points on the curve $\hat{\lambda}_\psi = \hat{\lambda}^0_\psi$ where the notation designates the constrained maximum likelihood estimate for λ.

It is straighforward to obtain the p variate tangent exponential model, and to obtain a tangent exponential model for a conditional model corresponding to the nuisance parameter. The quotient of these two gives the tangent exponential model describing the marginal ancillary variable. For details see [24].

The results in the preceding section then give p values for testing the value of the parameter ψ. Appropriate formulas are recorded in [24] and examples given in [10], [19], [25], [28], and [30].

10. OVERVIEW OF TANGENT EXPONENTIAL MODELS

The tangent location model theory in Sections 2-5 gives the tangent direction V for a third order ancillary, and these lead directly to the gradient $\varphi(\theta)$ of likelihood in the directions V at the data point. We then have that $\ell^0(\theta)$ and $\varphi(\theta)$ define the tangent exponential model (Section 7) for the conditional distribution given the ancillary.

For the scalar parameter case we then obtain (Section 8) the distribution function values at the data point to third order ancillary. More generally for a scalar component of a vector parameter, we obtain the distribution function value at the data point for the scalar ancillary variable recorded on the curve $\hat{\lambda}_\psi = \hat{\lambda}^0_\psi$.

In the first case we are working with an exponential model, and in the second case with an exponential model with an adjustment factor that is equal to 1 to first order. A detailed analysis of such adjusted exponential models is given in [26].

For a scalar variable y and parameter θ, left tail probabilities at a data point have been available since Edgeworth expansions. The saddlepoint methods are known to provide much improved accuracy and the Lugannani and Rice (1980) formula

$$\Phi(R,Q) = \Phi(R) + \varphi(R)\ \{R^{-1} - Q^{-1}\}$$

provides a widely used third order approximation. For this R is the signed likelihood root

$$R = \text{sgn}(\hat{\theta}^0 - \theta)[2\{\ell^0(\hat{\theta}^0) - \ell(\theta)\}]^{1/2}\ ,$$

and Q is the data standardized maximum likelihood estimate

$$Q = (\hat{\varphi}^0 - \varphi)\hat{\jmath}^{1/2}_{\varphi}\ .$$

An alternative way of packaging R and Q is provided by the Barndorff-Nielsen (1991) formula

$$\Phi(R,Q) = \Phi(R - R^{-1}\log R/Q)\ .$$

The first formula often gives better values but can give anomalous values, outside the $[0,1]$ range for probabilities. For some discussion see [26] and [29].

For a vector parameter θ with scalar interest parameter ψ we need only minor modifications of the definition for R and Q: the first becomes the profile likelihood root

$$R = \text{sgn}(\hat{\psi}^0 - \psi)\cdot[2\{\ell^0(\hat{\theta}^0_\psi) - \ell(\hat{\theta}_\psi)\}]^{1/2}$$

and the second becomes a generalized standardized maximum likelihood departure

$$Q = (\hat{\varphi}^0_p - \varphi_p)\left[\frac{\hat{\jmath}_{(\theta\theta)}}{\hat{\jmath}_{(\lambda\lambda)}(\hat{\theta}_\psi)}\right]^{1/2}\ .$$

For this φ_p is a scalar component of φ that measures departure from the value ψ at $\hat{\theta}^0_\psi$, $\hat{\jmath}_{(\theta\theta)}$ is the observed information for θ recalibrated on the φ scale, and $\hat{\lambda}_{(\lambda\lambda)}(\hat{\theta}_\psi)$ is the observed information for the nuisance parameter with ψ fixed recalibrated on the φ scale. These computations are easily organized for computer evaluation; see [24] and the examples in [25], [28], [30].

11. ADDENDUM

The preceding sections outlined the steps involved in using tangent location models to eliminate error and tangent exponential models to eliminate nuisance effect. Third order p-values for a scalar interest parameter are obtained. The methods can be generalized to vector interest parameters by testing scalar components in sequence; see [6] and Barndorff-Nielsen (1986). In general such tests would depend on the order chosen for the components, unlike normal analysis of variance.

The limiting normality of the conditional distribution given the ancillary has been examined in [4], [5]; these methods are being extended to the third order context.

The tangent exponential model methods in Section 7 can be used to obtain flat priors a scalar interest parameter; see [31].

Some computation methods for obtaining p-values are discussed in [9], [11], [13], [16], [17].

Recent asymptotic methods of inference have evolved from the saddlepoint method in statistics, (Daniels, 1954; Barndorff-Nielsen and Cox, 1979) and its extension to distribution functions (Lugannani and Rice, 1980). The applications typically involve exponential models and distributions with moment generating functions.

The extension of these methods to more general statistical models is largely due to Barndorff-Nielsen, for example Barndorff-Nielsen (1980, 1983, 1986, 1990, 1991). The tangent exponential model provides an alternative approach to many of these results and leads to a simpler formula (Section 10) for significance for a scalar parameter; the computation is straightforward and uses only likelihood, maximum likelihood estimates, information, and reparameterization. In some generality the formula for R and Q can be shown to be equivalent to formulas in Barndorff-Nielsen.

The preceding theory was however restricted by the need to obtain an $n-p$ dimension ancillary. The first derivative location model theory provides the theory and methods for constructing such ancillaries.

The following references relate to the tangent model approach to significance values, and in turn cite the related literature.

REFERENCES

[1] Fraser, D.A.S. (1964). Local conditional sufficiency, *J. Royal Statist. Soc.* **B26**, 52-62.

[2] Fraser, D.A.S. (1973). The elusive ancillary, *Multivariate Statistical Inference* (Ed: D.G. Kabe, R.P. Gupta), North Holland Publishing Company, Amsterdam, 41-48.

[3] Fraser, D.A.S. and MacKay, J. (1975). Parameter factorization and inference based on significance likelihood and objective posterior, *Annals Statist.* **3**, 559-572.

[4] Brenner, D., Fraser, D.A.S. and McDunnough, P. (1982). On asymptotic normality of likelihood and conditional analysis, *Can. J. Statist.* **10**, 163-172.

[5] Fraser, D.A.S. and McDunnough, P. (1984). Further remarks on asymptotic normality and conditional analysis, *Can. J. Statist.* **12**, 183-190.

[6] Fraser, D.A.S. and McDunnough, P. (1988). On generalization of analysis of variance, *Annals Inst. Statist. Math.* **40**, 353-66.

[7] Fraser, D.A.S. (1988). Normed likelihood as saddlepoint approximation, *J. Mult. Anal.* **27**, 181-193.

[8] Fraser, D.A.S. (1990). Tail probabilities from observed likelihoods, *Biometrika* **77**, 65-76.

[9] Fraser, D.A.S., Reid, N. and Wong, A. (1991). Exponential linear models: a two pass procedure for saddle point approximation. *J. Royal Statist. Soc.* **B53**, 483-492.

[10] Fraser, D.A.S., Reid, N. and Wong, A. (1991) Simple and accurate inference for the mean parameter of the gamma model, in revision.

[11] Fraser, D.A.S. and Reid, N. (1991). Converting observed likelihoods to tail probabilities, *Computational Statistics and Data Analysis* **12**, 179-185.

[12] Fraser, D.A.S. and Reid, N. (1990). From multiparameter likelihood to tail probabilities for a scalar parameter. Technical Report, University of Toronto, Department of Statistics.

[13] Fraser, D.A.S. (1991). Statistical inference: Likelihood to significance, *J. Amer. Statist. Assoc.* **86**, 258-265.

[14] Cheah, P.K., Fraser, D.A.S., Reid, N. and Tapia, A. (1992). Third order asymptotics: connections among test quantities, *Comm. Statist Th. Meth.* **21(8)**, 2127-33.

[15] Fraser, D.A.S. (1993). Directional analysis and statistical frames, *Statistical Papers* **34**, 213-236.

[16] Fraser, D.A.S. and Wong, A. (1993). Approximate Studentization with marginal and conditional inference, *Can. J. Statist.* **21**, 313-320.

[17] Fraser, D.A.S. and Reid, N. (1993). Third order asymptotic models: likelihood functions leading to accurate approximations for distribution functions, *Statist. Sinica* **3**, 67-82.

[18] Cheah, P.K., Fraser, D.A.S. and Reid, N. (1994). Multipliparameter testing in exponential models: third order approximations from likelihood, *Biometrika* **81**, 271-8.

[19] Fraser, D.A.S., Monette, G., Ng, K.W., and Wong, A. (1994). Higher order approximations with generalized linear models. *Multivariate Anaylsis and Its Applications* (eds Anderson, T.W., Fang, K.T., and Olkin, I.), *IMS Lecture Notes, Monograph Series*, **24**, 253-262.

[20] Cakmak, S. and Fraser, D.A.S. (1994). Multivariate asymptotic model: exponential and location approximations, submitted *Utilitas Mathematica* **46**, 21-31.

[21] Fraser, D.A.S. (1994). Bayes Posteriors for Scalar Interest Parameters,Bayesian International Conference at Valencia, June 6.

[22] Abebe, F., Cakmak, S., Cheah, P.K., Fraser, D.A.S., Kuhn, J., Reid, N. (1995). Third order asymptotic model: Exponential and location type approximations, accepted *Parisankhyan Samikkha* **1**, 1-5.

[23] Cakmak, S., Fraser, D.A.S., McDunnough, P., Reid, N. and Yuan, X. (1995). Likelihood centered asymptotic model: exponential and location model versions, accepted *Festschrift for A.M. Mathai*, Ed: S. Provost.

[24] Fraser, D.A.S. and Reid, N. (1995). Ancillaries and third order significance, *Utilitas Mathematica* **47**, 33-53.

[25] Abebe, F., Fraser, D.A.S., Reid, N. and Wong, A. Nonlinear regression: third order significance, ot appear, Utilitas Mathematica

[26] Cheah, P.K., Fraser, D.A.S, and Reid, N. Adjustment to likelihood and densities; calculating significance, *J.Statist. Research.* , to appear.

[27] Fraser, D.A.S. and Wong, A. On the accuracy of approximate Studentization, submitted *Can. J. Statist.*.

[28] Brenner, D., Fraser, D.A.S. and Zeng, Q. The Behrens Fisher problem by third order asymptotics;

[29] Fraser, D.A.S. and Reid, N. Evolution in statistical inference: from sufficiency to likelihood asymptotics; to appear, J. Statist. Research.

[30] Fraser, D.A.S. and Yuan, X. The Box and Cox problem: Asymptotic significance levels.

[31] Fraser, D.A.S. and Reid, N. Construction of third order ancillaries using measurement principles, in revision.