Journal of Statistical Research *ISSN 0256 - 422 X*
2007, Vol. 41, No. 2, pp. 1-20
Bangladesh

HIGHER ORDER ASYMPTOTICS: AN INTRINSIC DIFFERENCE BETWEEN UNIVARIATE AND MULTIVARIATE MODELS

Grace Y. Yi
Department of Statistics and Actuarial Science
University of Waterloo, Waterloo, Ontario, Canada N2L 3G1
Email: yyi@uwaterloo.ca

D.A.S. Fraser
Department of Statistics
University of Toronto, Toronto, Ontario, Canada M5S 3G3
Email: dfraser@utstat.toronto.edu

SUMMARY

Higher order asymptotic theory is targeted on the development of an asymptotic expansion for the distribution function of a statistic of interest. The asymptotic inference procedures are commonly based on simple characteristics of the density function at or near a data point of interest. In particular, exponential models are useful to provide accurate approximations to general statistical models. Typically, to the third order the exponential approximation has three primary parameters, two corresponding to pure model type and one for the departure from an exponential model (termed a non-exponentiality term). Andrews, Fraser and Wong (2005) discovered that to the third order, the observed significance function does not depend on the non-exponential term for univariate models. This finding has remarkable statistical implications for inference concerning univariate models. However, it is not clear whether this property holds for multivariate models. In this paper we address this question, and explore the intrinsic discrepancy between univariate and multivariate models.

Keywords and phrases: Exponential models; Likelihood formulation; Multivariate models; Observed significance functions; Transformations.

AMS Classification: 62F25

1 Introduction

In regular parametric models when the amount of information is large, first order asymptotic theory is widely used in statistical applications. For these models the Central Limit Theorem provides access to a range of statistical procedures. In particular, for models with independent random variables, the score function, being a sum of independent components, is asymptotically normal. Local linearization then relates the maximum likelihood estimate and likelihood ratio statistic to the score function. These three statistics provide important and powerful methods, often referred to as first order asymptotic theory. In some situations, however, first order asymptotic theory may not be adequate. Higher order asymptotics with more refined distributional approximations are required (e.g., Reid 1988); also and perhaps more importantly higher order asymptotics provides precise and often definite separation of interest parameters from nuisance parameters. For instance, when the number of nuisance parameters is large, first order theory may frequently fail to give reasonable and accurate approximations (Pierce and Peters 1992, Barndorff-Nielsen and Cox 1994).

For practical use in statistical inference, the primary interest often lies in approximating the cumulative distribution of a statistic in order to compute a p-value or confidence coefficients (e.g., Fraser 1990). Higher order asymptotic theory is targeted on the development of an asymptotic expansion for the distribution function of a statistic of interest. The asymptotic inference procedures are commonly based on simple characteristics of the density function at or near a data point of interest. These procedures include the Edgeworth expansion, the saddlepoint approximation, and the more recent approximations from likelihood theory (e.g., Lugannani and Rice 1980). For an overview, see Daniels (1987), Barndorff-Nielsen and Cox (1989), and Reid (1996).

For testing a scalar parameter, methods with third order accuracy are now available that are based on a reduction to the simple case having a scalar parameter and scalar variable. For such simple models on the real line, a canonical version that corresponds closely to an exponential model has been developed by Cakmak et al. (1998). The exponential approximation has three primary parameters, two corresponding to the pure model type and one for the departure from that model. This departure parameter is the measure of non-exponentiality. As the inference objective is to examine the observed significance function $p(\theta) = F(y_0; \theta) = P(y \leq y_0; \theta)$ at the data point y_0, it is important to investigate how these three model parameters may affect the observed significance function $p(\theta)$, especially how the non-exponentiality term may impact $p(\theta)$. Using only model derivatives at a data point and its corresponding maximum likelihood value, Andrews, Fraser and Wong (2005) explored this important problem and found that $p(\theta)$ is free of the non-exponentiality term to the third order, and it depends only on the two parameters which can be obtained from the observed likelihood and the gradient of the likelihood at the data point.

The finding in Andrews, Fraser and Wong (2005) has remarkable statistical implications. It provides a basis for understanding how nonnormality of the likelihood function affects related p-values. This casts some light on the familiar characteristic that the third order $p(\theta)$ is extremely accurate (Andrews, Fraser and Wong 2005). It also provides the basis for removing a computational singularity that commonly exists in the standard likelihood based higher order asymptotic methods, discussed in Daniels (1987), Robinson (1982) and Fraser et al. (2003). The investigation in Andrews, Fraser and Wong (2005) is, however, only addressed to univariate models to the third order inference. There are a couple of concerns remaining unclear. Does the property that $p(\theta)$ does not depend on the non-exponentiality term remain true beyond the third order? Or more importantly, does this property hold for multivariate models? The goal of this paper is to provide answers to these questions. Instead of directly working on the p-value as in Andrews, Fraser and Wong (2005) for univariate models, we follow a different route and use properties of moments of normal distributions. This allows us to transparently display the intrinsic discrepancy between univariate and multivariate models. As a by-product, the result in Andrews, Fraser and Wong (2005) is a special case of the current development.

The paper is organized as follows. Section 2 presents a number of properties of moments of normal distributions. In Section 3 we provide a brief overview of background results concerning asymptotic expansions of a univariate model when both parameter and variable are scalar. Such a simple model offers the basis for testing a scalar parameter of interest. The discussion is based directly on the Taylor series expansion of the model so that the exponential approximation is readily available. We also examine in this section the relationship between the p-value at the data point and the measure of the departure of a general model from the exponential family. In Section 4 we explore models with multi-dimensional parameter and variable. Exponential approximations to the models are derived, with special quantities involved to measure departure from the standard form of the exponential family. The dependence of the p-value on these parameters is discussed. The uniqueness of the transformations that are used for exponential approximations is discussed in Section 5. General discussion on asymptotic expansions is presented in the final Section 6.

2 Moments of Normal Distributions

In this section we discuss some properties for moments of normal distributions. These results are then applied for the development in the remaining sections.

Property 1. *If $Y \sim N(\theta, 1)$, then for any $n \in N$, there are constants a_i such that*

$$E(Y^{2n}) = \theta^{2n} + a_{2n-2}\theta^{2n-2} + ... + a_2\theta^2 + a_0,$$

and

$$E(Y^{2n+1}) = \theta^{2n+1} + a_{2n-1}\theta^{2n-1} + ... + a_3\theta^3 + a_1\theta.$$

Thus, the k^{th} moment of Y is a polynomial in θ of degrees of k, with coefficient for the highest order term being 1 and with successive terms having order reduced by 2.

Property 2. *Let $\phi(t)$, $\Phi(t)$ denote respectively the probability density function and the cumulative distribution function for the $N(0,1)$, then for any $n \in N$,*

$$\int_{-\infty}^{0} y^n \phi(y-\theta)dy = r_n\phi(-\theta) + \mu_n\Phi(-\theta), \tag{2.1}$$

where $\mu_n = E(Y^n)$ denotes the n^{th} moment of Y with the $N(\theta,1)$ distribution, and the r's and μ's are polynomials that satisfy the recursive equations

$$r_{n+2} = (n+1)r_n + \theta r_{n+1} \text{ with } r_0 = 0, r_1 = -1,$$
$$\mu_{n+2} = (n+1)\mu_n + \theta\mu_{n+1} \text{ with } \mu_0 = 1, \mu_1 = \theta.$$

Proof. Integration by parts shows that $\int_{-\infty}^{a} x^n\phi(x)dx = p(a)\phi(a) + q(a)\Phi(a)$ where $p(a)$ and $q(a)$ are polynomials in a; and the transformation $y = x + \theta$ gives

$$\int_{-\infty}^{0} y^n\phi(y-\theta)dy = \int_{-\infty}^{-\theta} x^n\phi(x)dx + \int_{-\infty}^{-\theta} n\theta x^{n-1}\phi(x)dx + ... + \int_{-\infty}^{-\theta} \theta^n\phi(x)dx.$$

These indicate that there are polynomials μ_n and r_n in θ such that

$$\int_{-\infty}^{0} y^n\phi(y-\theta)dy = r_n\phi(-\theta) + \mu_n\Phi(-\theta).$$

It remains now to characterize the polynomials r_n and μ_n. Integrating $\int_{-\infty}^{0} y^n\phi(y-\theta)dy$ gives

$$\int_{-\infty}^{0} y^n\phi(y-\theta)dy = \int_{-\infty}^{0} \frac{1}{n+1}y^{n+2}\phi(y-\theta)dy - \int_{-\infty}^{0} \frac{\theta}{n+1}y^{n+1}\phi(y-\theta)dy.$$

Then using the above form of $\int_{-\infty}^{0} y^n\phi(y-\theta)dy$, we obtain

$$r_n\phi(-\theta) + \mu_n\Phi(-\theta) = \frac{1}{n+1}[r_{n+2}\phi(-\theta) + \mu_{n+2}\Phi(-\theta)] - \frac{\theta}{n+1}[r_{n+1}\phi(-\theta) + \mu_{n+1}\Phi(-\theta)],$$

and thus,

$$r_{n+2} = (n+1)r_n + \theta r_{n+1},$$
$$\mu_{n+2} = (n+1)\mu_n + \theta\mu_{n+1}.$$

It is then easily seen that $r_0 = 0, \mu_0 = 1,$ and $r_1 = -1, \mu_1 = \theta$. Furthermore, it can be shown that $\mu_n = E(Y^n)$ for $Y \sim N(\theta, 1)$. Indeed, let $E(Y^n) = m_n$. Using integration by parts, we obtain

$$m_n = \int_{-\infty}^{+\infty} y^n\phi(y-\theta)dy = \frac{1}{n+1}m_{n+2} - \frac{\theta}{n+1}m_{n+1},$$

i.e., $m_{n+2} = (n+1)m_n + \theta m_{n+1}$. It follows that m_n and μ_n have the same recursive equations. Since $m_0 = \mu_0$ and $m_1 = \mu_1$, it follows that $m_n = \mu_n$ for all n.

For convenience, we record some values for μ_n that will be used next and in Section 3.3:
$\mu_2 = \theta^2 + 1,\ \mu_3 = \theta^3 + 3\theta,\ \mu_4 = \theta^4 + 6\theta^2 + 3,\ \mu_5 = \theta^5 + 10\theta^3 + 15\theta.$

The following property discusses connections among the polynomials μ_n and r_n in (2.1).

Property 3. *(1). If $a_{s+t-2}, ..., a_3, a_1$ are constants that satisfy*

$$\theta^s\mu_t - \mu_{s+t} + a_{s+t-2}\mu_{s+t-2} + ... + a_3\mu_3 + a_1\mu_1 = 0,$$

where $s+t$ is odd with $s \geq 2$ and $t \geq 3$, then

$$\theta^s r_t - r_{s+t} + a_{s+t-2}r_{s+t-2} + ... + a_3r_3 + a_1r_1 = 0.$$

(2). If $a_{s+t-2}, ..., a_2, a_0$ are constants that satisfy

$$\theta^s\mu_t - \mu_{s+t} + a_{s+t-2}\mu_{s+t-2} + ... + a_2\mu_2 + a_0\mu_0 = 0,$$

where $s+t$ is even with $s \geq 2$ and $t \geq 2$, then

$$\theta^s r_t - r_{s+t} + a_{s+t-2}r_{s+t-2} + ... + a_2r_2 + a_0r_0 = 0.$$

Proof. First note that $\theta^2\mu_3 - \mu_5 + a_3\mu_3 + a_1\mu_1 = 0$ gives $a_3 = 7, a_1 = -6$. Now we claim $a_{2n+1} = 4n+3$, $a_{2n-1} = -2n(2n+1)$, $a_{2n-3} = ... = a_1 = 0$, if for $s = 2$ and $t = 2n+1 (n \geq 2)$,

$$\theta^2\mu_{2n+1} - \mu_{2n+3} + a_{2n+1}\mu_{2n+1} + a_{2n-1}\mu_{2n-1} + a_{2n-3}\mu_{2n-3} + ... + a_1\mu_1 = 0.$$

Indeed, let $I_2 = \theta^2\mu_{2n+1} - \mu_{2n+3} + a_{2n+1}\mu_{2n+1} + a_{2n-1}\mu_{2n-1} + a_{2n-3}\mu_{2n-3} + ... + a_1\mu_1$. Then repeatedly invoking the recursive equations in Property 2, we find that the first two terms $\theta^2\mu_{2n+1}$ and μ_{2n+3} have the same highest order term θ^{2n+3}, which then cancels; we thus obtain

$$\begin{aligned} I_2 &= [a_{2n+1} - (2n+2) - (2n+1)]\theta\mu_{2n} + [(a_{2n+1} - (2n+2))2n + a_{2n-1}]\mu_{2n-1} \\ &\quad + a_{2n-3}\mu_{2n-3} + ... + a_1\mu_1. \end{aligned}$$

From Property 1 we see that μ_t is a polynomial in θ of degrees t with successive terms having order reduced by 2. Thus, I_2 is a polynomial in θ of degrees $2n+1$ with successive terms having order dropping by 2. Consequently, we obtain that all coefficients in θ are 0 when $I_2 = 0$. That is, $a_{2n+1} - (2n+2) - (2n+1) = 0$, $(a_{2n+1} - (2n+2))2n + a_{2n-1} = 0$, and $a_{2n-3} = ... = a_1 = 0$, leading to

$$a_{2n+1} = 4n+3,\ a_{2n-1} = -2n(2n+1),\ a_{2n-3} = ... = a_1 = 0.$$

Now for r_t we have the same recursive equation as μ_t based on Property 2; thus for $J_2 = \theta^2 r_{2n+1} - r_{2n+3} + a_{2n+1}r_{2n+1} + ... + a_1 r_1$ we can conduct similar calculations and obtain

$$\begin{aligned} J_2 &= [a_{2n+1} - (4n+3)]\theta r_{2n} + [(a_{2n+1} - (2n+2))2n + a_{2n-1}]r_{2n-1} \\ &\quad + a_{2n-3}r_{2n-3} + ... + a_1 r_1, \end{aligned}$$

which gives 0 using the values of a_{2k+1} just obtained. That is, if $I_2 = 0$ holds, then $J_2 = 0$ holds as well. Analogously, the conclusions can be proved for the other cases with $s > 2$.

Note that the above result is true only for $s + t \geq 4$. For instance, when $s = 2, t = 1$, we have $I_2 = \theta^2\mu_1 - \mu_3 + 3\mu_1 = 0$, but $J_2 = \theta^2 r_1 - r_3 + 3r_1 = -1 \neq 0$.

3 Univariate Model: Observed Significance Function and Measure of Non-exponentiality

3.1 Exponential Model

For a real variable and real parameter consider a density function $f_n(y;\theta)$ that depends on a mathematical parameter n, often sample size. We assume that for each θ, y is $O_p(n^{-1/2})$ about a maximum density point and that $l(\theta; y) = \log f_n(y;\theta)$ is $O(n)$ and with either argument fixed has a unique maximum. Let y_0 be a data value of interest, and $\theta_0 = \hat{\theta}(y_0)$ be the corresponding maximum likelihood parameter value.

If $a_{ij} = (\partial^{i+j}/\partial\theta^i\partial y^j)l(\theta;y)|_{(\theta_0,y_0)}$, then the log density $l(\theta;y)$ has a Taylor series expansion in parameter and variable about the data,

$$l(\theta;y) = \sum_{i,j} \frac{a_{ij}}{i!j!}(\theta-\theta_0)^i(y-y_0)^j.$$

To examine the model more easily, we simply record the coefficients of the expansion in a matrix form; we then standardize the variable and the parameter using location and scale properties with the transformations

$$\phi = (-a_{20})^{1/2}(\theta-\theta_0),$$

$$x = (-a_{20})^{-1/2}a_{11}(y-y_0).$$

For the particular case of a standard exponential model $f(y,\theta) = \exp\{y\theta - c(\theta)\}h(y)$ we obtain the following coefficient array

$$\begin{pmatrix} a_{00} & a_{01} & a_{02} & a_{03} & a_{04} & \cdots \\ 0 & 1 & 0 & 0 & 0 & \cdots \\ -1 & 0 & 0 & 0 & 0 & \cdots \\ a_{30} & 0 & 0 & 0 & 0 & \cdots \\ a_{40} & 0 & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}, \tag{3.1}$$

where a_{ij} denotes the typical nonzero coefficient after the transformation. The third derivatives are of order $O(n^{-1/2})$, the fourth derivatives are of order $O(n^{-1})$, and in general, the r^{th} $(r \geq 5)$ derivatives are of order $O(n^{-r/2+1})$. When examining a coefficient array after standardization we will again use y and θ for variable and parameter, these being then the current variable and parameter.

Of special importance we see that all the coefficients in the matrix outside the first row and the first column are zero except for the coefficient of the cross term $y\theta$. As $a_{20} = -1, a_{11} = 1$, we can, therefore, write

$$l(\theta;y) = \log\phi(y-\theta) + r(\theta) + s(y);$$

thus, the log density can be expressed as the log density of the $N(\theta,1)$ plus a series $r(\theta)$ in θ and a series $s(y)$ in y, where from the integration property of the density function, we have that the coefficients in $r(\theta)$ determine those in $s(y)$ and vice versa. In particular, to order $O(n^{-3/2})$, Cakmak et al. (1998) described the relationship between these coefficients by examining terms up to the fourth derivative. Specifically if we write $a_{30} = -\alpha_3/\sqrt{n}, a_{40} = -\alpha_4/n$, then the first row coefficients are given as

$$a_{00} = -\log(2\pi)/2 + (3\alpha_4 - 5\alpha_3^2)/24n,$$
$$a_{01} = -\alpha_3/2\sqrt{n},$$

$$a_{02} = -1 + (\alpha_4 - 2\alpha_3^2)/2n,$$
$$a_{03} = \alpha_3/\sqrt{n},$$
$$a_{04} = (\alpha_4 - 3\alpha_3^2)/n.$$

We may examine the coefficients of $r(\theta)$ and $s(y)$ more closely by considering a higher order expansion. For instance, with order $O(n^{-2})$ we need the fifth derivative a_{50} when expressing $r(\theta)$, and there is an $O(n^{-3/2})$ term in the third derivative a_{30}. It is known then, from the Taylor series expansion, that the coefficients of $s(y)$ have the form

$$a_{00} = -\log(2\pi)/2 + a_0/n,$$
$$a_{01} = -a_1/\sqrt{n} + b_1/n^{3/2},$$
$$a_{02} = -1 + a_2/n,$$
$$a_{03} = a_3/\sqrt{n} + b_3/n^{3/2},$$
$$a_{04} = a_4/n,$$
$$a_{05} = a_5/n^{3/2}$$

for constants $a_i (i = 0, 1, ..., 5), b_1$ and b_3. Thus, the r^{th} $(r \geq 3)$ derivative a_{0r} is of order $O(n^{-r/2+1})$, and the first derivative a_{01} and the third derivative a_{03} now include $O(n^{-3/2})$ terms. If we write $a_{30} = -\alpha_3/\sqrt{n} - \beta/n^{3/2}$ and $a_{40} = -\alpha_4/n,\quad a_{50} = -\alpha_5/n^{3/2}$, where $\alpha_i (i = 3, 4, 5)$ and β are constants, then the coefficients a_{0j} are determined by the coefficients a_{i0}:

$$a_0 = (3\alpha_4 - 5\alpha_3^2)/24,$$
$$a_1 = -\alpha_3/2,$$
$$a_2 = -(\alpha_4 - 2\alpha_3^2)/2,$$
$$a_3 = \alpha_3,$$
$$a_4 = \alpha_4 - 3\alpha_3^2,$$
$$a_5 = \alpha_5 + 15\alpha_3^3 - 10\alpha_3\alpha_4,$$
$$b_1 = -(8\beta - 2\alpha_5 + 29\alpha_3^3 - 22\alpha_3\alpha_4)/16,$$
$$b_3 = (24\beta - 12\alpha_5 + 21\alpha_3^3 - 14\alpha_3\alpha_4)/24.$$

3.2 General Model

Cakmak et al. (1998) described a series of transformations to approximate general models by the exponential families. Essentially they employed the following transformations:

$$\begin{aligned} \phi &= (-a_{20})^{1/2} a_{11}^{-1} [a_{11}(\theta - \theta_0) + \frac{a_{21}}{2}(\theta - \theta_0)^2 + \frac{a_{31}}{6}(\theta - \theta_0)^3], \\ x &= (-a_{20})^{-1/2} [a_{11}(y - y_0) + \frac{a_{12}}{2}(y - y_0)^2 + \frac{a_{13}}{6}(y - y_0)^3], \end{aligned} \quad (3.2)$$

which lead to the coefficient matrix with the missing terms being of order $O(n^{-3/2})$,

$$\begin{pmatrix} A_{00} & A_{01} & A_{02} & A_{03} & A_{04} \\ 0 & 1 & 0 & 0 & - \\ -1 & 0 & A_{22} & - & - \\ A_{30} & 0 & - & - & - \\ A_{40} & - & - & - & - \end{pmatrix}. \tag{3.3}$$

Here the transformed coefficients A_{ij} and the original coefficients a_{ij} are linked as follows:

$A_{30} = (-a_{20})^{-3/2}a_{30} + 3(-a_{20})^{-1/2}a_{11}^{-1}a_{21}$,

$A_{40} = (-a_{20})^{-2}a_{40}+4(-a_{20})^{-1}a_{11}^{-1}a_{31}-6(-a_{20})^{-2}a_{11}^{-1}a_{30}a_{21}-15(-a_{20})^{-1}a_{11}^{-2}a_{21}^2$,

$A_{22} = a_{11}^{-3}(a_{11}a_{22} - a_{21}a_{12})$,

and $A_{0j}, j = 0, 1, 2, 3, 4$, are determined by A_{30}, A_{40}, and A_{22}.

More generally, under the following transformations we can re-express the coefficient matrix as "exponential-like", i.e., it has a form like (3.1) except for the right lower corner. Let

$$\phi = (-a_{20})^{1/2}a_{11}^{-1}\sum_i \frac{a_{i1}}{i!}(\theta - \theta_0)^i,$$

$$x = (-a_{20})^{-1/2}\sum_j \frac{a_{1j}}{j!}(y - y_0)^j,$$

then the coefficient matrix for the model $l(\theta; y) = \log f(y; \theta)$ becomes:

$$\begin{pmatrix} a_{00} & a_{01} & a_{02} & a_{03} & a_{04} & \cdots \\ 0 & 1 & 0 & 0 & 0 & \cdots \\ -1 & 0 & a_{22} & a_{23} & a_{24} & \cdots \\ a_{30} & 0 & a_{32} & a_{33} & a_{34} & \cdots \\ a_{40} & 0 & a_{42} & a_{43} & a_{44} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}. \tag{3.4}$$

Then to avoid excessive notation we use the same a_{ij} to denote the typical coefficient after the transformation and use y, θ again to denote correspondingly the revised variable and parameter; the new expansion point accordingly is $(y_0, \theta_0) = (0, 0)$.

Note that $a_{ij} = O(n^{-(i+j)/2+1})$ for $i \geq 2, j \geq 2$; thus to order $O(n^{-3/2})$ the general model differs from the exponential family by the term a_{22}, which is the coefficient of the quadratic-quadratic term $y^2\theta^2$. In general, terms a_{ij} measure the departure of a general model from the standard exponential model when $i \geq 2, j \geq 2$. If $a_{ij} = 0$ for all $i \geq 2, j \geq 2$ in (3.4), then a_{i0} and a_{0j} in (3.4) should be the respective coefficients of $r(\theta)$ and $s(y)$ in (3.1), and their expressions are given in Section 3.1.

3.3 The Observed Significance Function

As illustrated in Section 3.1, if $a_{i'j'} = 0$ for $i' \geq 2, j' \geq 2$, then the first row and the first column in (3.4) correspond to $s(y)$ and $r(\theta)$, respectively. Denoting the corresponding density function for this exponential model as $g(y;\theta)$, we write

$$\log g(y;\theta) = \log \phi(y-\theta) + (\frac{1}{2}\log(2\pi) + r(\theta)) + s(y). \tag{3.5}$$

We now explore the relationship between the observed significance function and the terms a_{ij} for $i \geq 2, j \geq 2$, which measure the departure of a general model from the standard exponential model. To get insight into this relationship, we consider the case where only one of the $a_{ij} \neq 0$ for $i \geq 2, j \geq 2$. The discussion on more general cases can proceed in the same manner.

Theorem 1. *If there is only one non-exponentiality term a_{ij} with $i \geq 2, j \geq 2$ in (3.4) that is not zero, then to order $O(n^{-(i+j)/2+1/2})$, the observed significance function $F(0;\theta) = \int_{-\infty}^{0} f(y;\theta)dy$ is free of the non-exponentiality term a_{ij}. Furthermore,*

$$F(0;\theta) = \int_{-\infty}^{0} g(y;\theta)dy.$$

This result has an important implication concerning the accuracy of higher order asymptotic inferences. It says that if there is only one non-exponentiality term a_{ij} with $i \geq 2, j \geq 2$ present in (3.4), then the observed significance function $F(0;\theta)$ for the initial model is, to order $O(n^{-(i+j)/2+1/2})$, completely determined by that from its approximate exponential model. In particular, the observed significance function $F(0;\theta)$ is free of a_{22} to order $O(n^{-3/2})$, which is obtained in Andrews, Fraser and Wong (2005).

Proof. In the same spirit of the proof of Property 3 in Section 1, we can show for any $i \geq 2, j \geq 2$, that there exists a unique polynomial in y, say $h_{ij}(y)$, such that

$$\int_{-\infty}^{+\infty} \phi(y-\theta)(\theta^i y^j + h_{ij}(y))dy = 0. \tag{3.6}$$

For an illustration we examine the case with $i = 2$ and $j = 2$. Since $\int_{-\infty}^{+\infty} \phi(y-\theta) \cdot \theta^2 y^2 dy = \theta^2 \mu_2 = \theta^4 + \theta^2$, it suffices to find a polynomial function $h_{22}(y)$ such that (3.6) holds. By Property 1 we assume $h_{22}(y) = b_4 y^4 + b_3 y^3 + b_2 y^2 + b_1 y + b_0$, and it remains to check the existence and uniqueness of such constants $b_k (k = 4, 3, ..., 0)$. Indeed, by Property 2,

$$\begin{aligned}
& \int_{-\infty}^{+\infty} \phi(y-\theta) \cdot h_{ij}(y)dy \\
= \;& b_4\mu_4 + b_3\mu_3 + b_2\mu_2 + b_1\mu_1 + b_0 \\
= \;& b_4(\theta^4 + 6\theta^2 + 3) + b_3(\theta^3 + 3\theta) + b_2(\theta^2 + 1) + b_1\theta + b_0 \\
= \;& b_4\theta^4 + b_3\theta^3 + (6b_4 + b_2)\theta^2 + (3b_3 + b_1)\theta + (3b_4 + b_2 + b_0).
\end{aligned}$$

Note that this is a polynomial in θ having the number of unknown coefficients equal to the highest degree of θ, thus equating it to $-\int_{-\infty}^{+\infty}\phi(y-\theta)\cdot\theta^2y^2dy = -\theta^4-\theta^2$ would lead to the unique solution $b_k's$. That is,

$$b_4 = -1,\ b_3,\ b_2 = 5,\ b_1 = 0,\ b_0 = -2,$$

which shows the result.

Accordingly, for a model (3.4) with $a_{ij} \neq 0$ for some $i \geq 2$ and $j \geq 2$, there would be a polynomial in y that is the adjustment to the first row $s(y)$ based on the integration property of the density. That is, the log density for the model with $a_{ij} \neq 0$ is

$$\log f(y;\theta) = \log g(y;\theta) + \frac{a_{ij}}{i!j!}h_{ij}(y) + \frac{a_{ij}}{i!j!}\theta^i y^j \tag{3.7}$$

to order $O(n^{-(i+j)/2+1/2})$, where $h_{ij}(y)$ satisfies (3.6).

The following justifies that such a form is the log density function. Note that $(\frac{1}{2}\log(2\pi)+r(\theta))+s(y)$ is of order $O(n^{-1/2})$ and $a_{ij} = O(n^{-(i+j)/2+1})$ for $i \geq 2, j \geq 2$. Using the Taylor series expansion $e^t = 1+t+t^2/2+\cdots$, we obtain to order $O(n^{-(i+j)/2+1/2})$ from equation (3.6) that

$$\begin{aligned}
&\int_{-\infty}^{+\infty} g(y;\theta)\cdot\exp\{\frac{a_{ij}}{i!j!}h_{ij}(y) + \frac{a_{ij}}{i!j!}\theta^i y^j\}dy \\
=\ &\int_{-\infty}^{+\infty} g(y;\theta)\cdot\{1+\frac{a_{ij}}{i!j!}(\theta^i y^j + h_{ij}(y))\}dy \qquad (3.8)\\
=\ &1+\int_{-\infty}^{+\infty}\phi(y-\theta)\cdot\frac{a_{ij}}{i!j!}(\theta^i y^j + h_{ij}(y))dy \qquad (3.9)\\
=\ &1.
\end{aligned}$$

We now examine the observed significance function $F(0;\theta)$ at the data point. In the above calculation the integral is on the full real line, while for $F(0;\theta)$ the integral is on the half line left the origin. We then can do similar calculations and obtain, to order $O(n^{-(i+j)/2+1/2})$,

$$\begin{aligned}
F(0;\theta) &= \int_{-\infty}^{0} f(y;\theta)dy \\
&= \int_{-\infty}^{0} g(y;\theta)dy + \int_{-\infty}^{0}\phi(y-\theta)\frac{a_{ij}}{i!j!}(\theta^i y^j + h_{ij}(y))dy \\
&= \int_{-\infty}^{0} g(y;\theta)dy,
\end{aligned}$$

where the second integral disappears by Property 3. Therefore, we conclude that $F(0;\theta)$ is free of a_{ij} to order $O(n^{-(i+j)/2+1/2})$.

We conclude this section with the comment that the result in Theorem is not true to an order higher than $O(n^{-(i+j)/2+1/2})$. The reason is that to an order higher than

$O(n^{-(i+j)/2+1/2})$, function $f(y;\theta)$ in (3.7) is not a density function any more. For example, considering order $O(n^{-(i+j)/2})$, we may repeat the calculation in the foregoing proof until step (3.8), but not step (3.9). When raising an additional order $O(n^{-1/2})$, there would be a non-zero additional term $\int_{-\infty}^{+\infty}[(\frac{1}{2}\log(2\pi)+r(\theta))+s(y)]\cdot\frac{a_{ij}}{i!j!}(\theta^i y^j+h_{ij}(y))dy$ included according to the form of (3.5) and that $(\frac{1}{2}\log(2\pi)+r(\theta))+s(y)$ is of order $O(n^{-1/2})$.

To be more specific, we examine a simple case that $F(0;\theta)$ is not free of a_{22} to order $O(n^{-2})$. For simplicity, suppose that the coefficients of $r(\theta)$ expressed in Section 3.1 take the values $\beta=0, \alpha_4=0, \alpha_5=0$, and we write $a_{22}=c/n$, then the first row in (3.4) corresponds to $s(y)+h(y)$, where $h(y)$ is a polynomial in y. From the integration property of the density function, we obtain that $h(y)$ is of the form

$$h(y)=c_5y^5/n^{3/2}+c_3y^3/n^{3/2}+c_1y/n^{3/2}+d_4y^4/n+d_2y^2/n+d_0/n,$$

and the coefficients c_i, d_i are determined by α_3 and c as follows:

$$c_5=\alpha_3c/4, c_3=-13\alpha_3^3/16-41\alpha_3c/4, c_1=-77\alpha_3^3/16+113\alpha_3c/4,$$

$$d_4=-c/4, d_2=5c/4, d_0=-c/2.$$

By Property 2 and direct calculation, we can see that $F(0;\theta)$ is of the form

$$F(0;\theta)=\Phi(-\theta)+\phi(-\theta)\cdot k(\theta),$$

where $k(\theta)$ is a polynomial in θ. Simply examining the constant term in $k(\theta)$, we can see that it involves $-37\alpha_3c/4$, and it follows that $F(0;\theta)$ is not free of a_{22} to order $O(n^{-2})$.

4 Multivariate Models: Observed Marginal Significance Functions and Non-exponentiality Terms

In this section we investigate if the same property in Section 3.3 holds for multivariate models. It turns out that the observed marginal significance functions depend on the non-exponentiality terms in the case of the multivariate models. This demonstrates intrinsic difference between univariate and multivariate models when using the third order approximations. The discussion starts with the p dimensional exponential approximations to the multivariate models with p dimensional variable and parameter.

4.1 Location-Scale Standardization

Consider a statistical model $f(y;\theta)$ with a p dimensional variable and p dimensional parameter and suppose the model is asymptotic as some mathematical parameter $n\to\infty$: that for each θ, y is $O_p(n^{-1/2})$ about the maximum density value; and that $l(\theta;y)=\log f(y;\theta)$ with either argument fixed is $O(n)$ and has a unique maximum. For some background, see Cakmak, Fraser and Reid (1994).

Let y^0 be a data point of interest and $\theta^0=\hat{\theta}(y^0)$ be the corresponding maximum likelihood estimate. We consider the Taylor series expansion of $l(\theta; y)$ about (θ^0, y^0):

$$\begin{aligned} l(\theta; y) &= l(\theta^0; y^0) + \sum a^\alpha (y_\alpha - y_\alpha^0) \\ &+ \sum a_{ij}(\theta_i - \theta_i^0)(\theta_j - \theta_j^0)/2! + \sum a_i^\alpha (\theta_i - \theta_i^0)(y_\alpha - y_\alpha^0) \\ &+ \sum a^{\alpha\beta}(y_\alpha - y_\alpha^0)(y_\beta - y_\beta^0)/2! + ..., \end{aligned}$$

where for example $a_{ij}^\alpha = (\partial/\partial\theta_i)(\partial/\partial\theta_j)(\partial/\partial y_\alpha) l(\theta; y)|_{(\theta^0, y^0)}$.

To avoid excessive notation, in each of the following steps, let θ, y, a_{ij} denote the respective initial parameter, variable and coefficients, and let ϕ, x, A_{ij} denote the respective transformed parameter, variable and coefficients. At the final step we let a and y be used again for the final model. The indices of the coefficients run from 1 to p and summation over repeated indices will be used and implied (McCullagh 1987).

First recenter the model at (θ^0, y^0) by the transformations

$$\phi = \theta - \theta^0, \; x = y - y^0,$$

and rescale the parameter using $\theta_i = c_{ij}\phi_j$ to give an identity observed information array at the expansion point. Next, rescale the variable so that the cross Hessian a_i^α becomes an identity array I_i^α: let $y_\alpha = d_\alpha^\beta x_\beta$ define a new variable x so that $I_i^\alpha = a_i^\beta d_\beta^\alpha$ has an identity array. The coefficient array then becomes

$$\begin{pmatrix} a & a^\alpha & a^{\alpha\beta} & a^{\alpha\beta\gamma} & a^{\alpha\beta\gamma\delta} \\ 0 & I_i^\alpha & a_i^{\alpha\beta} & a_i^{\alpha\beta\gamma} & - \\ -I_{ij} & a_{ij}^\alpha & a_{ij}^{\alpha\beta} & - & - \\ -a_{ijk} & a_{ijk}^\alpha & - & - & - \\ -a_{ijkl} & - & - & - & - \end{pmatrix}, \tag{4.1}$$

where elements with three indices are $O(n^{-1/2})$, those with four are $O(n^{-1})$, and the lower right elements that are missing are $O(n^{-3/2})$. In general, elements with r indices are $O(n^{-r/2+1})$.

4.2 Exponential Approximation to The Multivariate Model

In this section we examine an exponential approximation to the multivariate model $f(y; \theta)$. First note that for the location-scale standardization of the p dimensional exponential model, the coefficient array takes the form

$$\begin{pmatrix} a & a^{\alpha} & a^{\alpha\beta} & a^{\alpha\beta\gamma} & a^{\alpha\beta\gamma\delta} \\ 0 & I_i^{\alpha} & 0 & 0 & - \\ -I_{ij} & 0 & 0 & - & - \\ -a_{ijk} & 0 & - & - & - \\ -a_{ijkl} & - & - & - & - \end{pmatrix},$$

where the elements in the first row are in fact determined by the elements in the first column except for the first element, and vice versa. In fact the elements in the first row are available from the formula in Section 4 of Fraser and Reid (1993).

Following steps similar to those in the one dimensional case, we introduce a series of transformations to convert the location-scale standardized model (4.1) towards an exponential model. First we define a new parameter ϕ to obtain appropriate zeros in the second column of the coefficient array (4.1):

$$\phi_{\alpha} = \theta_{\alpha} + a_{ij}^{\alpha}\theta_i\theta_j/2! + a_{ijk}^{\alpha}\theta_i\theta_j\theta_k/3!.$$

This transformation changes many coefficients in the array as indicated by the scalar case in Section 3.2. Using the same notation for the new coefficients to avoid notational growth we then obtain the array

$$\begin{pmatrix} a & a^{\alpha} & a^{\alpha\beta} & a^{\alpha\beta\gamma} & a^{\alpha\beta\gamma\delta} \\ 0 & I_i^{\alpha} & a_i^{\alpha\beta} & a_i^{\alpha\beta\gamma} & - \\ -I_{ij} & 0 & a_{ij}^{\alpha\beta} & - & - \\ -a_{ijk} & 0 & - & - & - \\ -a_{ijkl} & - & - & - & - \end{pmatrix}.$$

Secondly, we define a new variable x to obtain appropriate zeros in the second row:

$$x_i = y_i + a_i^{\alpha\beta}y_{\alpha}y_{\beta}/2! + a_i^{\alpha\beta\gamma}y_{\alpha}y_{\beta}y_{\gamma}/3!,$$

the resulting array then has the form

$$\begin{pmatrix} a & a^{\alpha} & a^{\alpha\beta} & a^{\alpha\beta\gamma} & a^{\alpha\beta\gamma\delta} \\ 0 & I_i^{\alpha} & 0 & 0 & - \\ -I_{ij} & 0 & a_{ij}^{\alpha\beta} & - & - \\ -a_{ijk} & 0 & - & - & - \\ -a_{ijkl} & - & - & - & - \end{pmatrix}.$$

If the coefficients $a_{ij}^{\alpha\beta} = 0$, then we have an exponential model to $O(n^{-3/2})$. Accordingly, the terms $a_{ij}^{\alpha\beta}$ measure the non-exponentiality of the model, and report the departure of the model from the exponential case. In particular, for bivariate models, we list all the non-exponentiality terms:

$$a_{11}^{11}, a_{11}^{12}, a_{11}^{22}, a_{12}^{11}, a_{12}^{12}, a_{12}^{22}, a_{22}^{11}, a_{22}^{12}, a_{22}^{22}$$

to order $O(n^{-3/2})$.

4.3 The Observed Marginal Significance Functions

To get insight into the dependence of the observed marginal significance functions on the non-exponentiality terms, we focus the discussion on the bivariate case. Let $p_1(\theta) = P(y_1 \leq 0, -\infty < y_2 < +\infty; \theta)$ and $p_2(\theta) = P(-\infty < y_1 < +\infty, y_2 \leq 0; \theta)$. Typically, we will consider the case where only one non-exponentiality term $a_{ij}^{\alpha\beta}$ is not zero, with $i, j, \alpha, \beta \in \{1, 2\}$.

Let $g(y;\theta)$ denote the exponential model obtained with all $a_{i'j'}^{\alpha'\beta'} = 0$. When there is one non-exponentiality term $a_{ij}^{\alpha\beta} \neq 0$, we can show that there is an adjustment $h(y_1, y_2)$ to the first row elements of the coefficient array such that to order $O(n^{-3/2})$,

$$l(\theta_1, \theta_2) = \log g(y;\theta) + h(y_1, y_2) + a_{ij}^{\alpha\beta}\theta_i\theta_j y_\alpha y_\beta$$

is the log density for the model. Here note that $a_{ij}^{\alpha\beta}$ absorbs the factorial coefficient $1/i!j!\alpha!\beta!$.

Indeed, according to the Taylor series expansion $e^t = 1 + t + t^2/2 + ...$ and that $a_{ij}^{\alpha\beta}$ is of order $O(n^{-1})$, it suffices to determine whether there exists an adjustment $h(y_1, y_2)$ to order $O(n^{-1})$ such that

$$\int_{-\infty}^{+\infty} dy_1 \int_{-\infty}^{+\infty} \{1 + h(y_1, y_2) + a_{ij}^{\alpha\beta}\theta_i\theta_j y_\alpha y_\beta\} \cdot g(y;\theta) dy_2 = 1 \tag{4.2}$$

to order $O(n^{-3/2})$.

In the same spirit of the proof of Theorem in Section 3.3, we can show that $h(y_1, y_2) = a_{ij}^{\alpha\beta} h_{ij}^{\alpha\beta}(y_1, y_2)$ is a unique polynomial of order $O(n^{-1})$ that satisfies (4.2), where $h_{ij}^{\alpha\beta}(y_1, y_2)$ satisfies

$$\int_{-\infty}^{+\infty} dy_1 \int_{-\infty}^{+\infty} \phi(y_1 - \theta_1)\phi(y_2 - \theta_2)(\theta_i\theta_j y_\alpha y_\beta + h_{ij}^{\alpha\beta}(y_1, y_2)) dy_2 = 0.$$

Now set $h(y_1, y_2) = a_{ij}^{\alpha\beta} h_{ij}^{\alpha\beta}(y_1, y_2)$. Accordingly, we can examine $p_1(\theta)$ by integrating the density on the half plane and obtain

$$\begin{aligned} p_1(\theta) &= \int_{-\infty}^{0} dy_1 \int_{-\infty}^{+\infty} \exp\{a_{ij}^{\alpha\beta}(h_{ij}^{\alpha\beta}(y_1, y_2) + \theta_i\theta_j y_\alpha y_\beta)\} g(y;\theta) dy_2 \\ &= \int_{-\infty}^{0} dy_1 \int_{-\infty}^{+\infty} \{1 + a_{ij}^{\alpha\beta}(h_{ij}^{\alpha\beta}(y_1, y_2) + \theta_i\theta_j y_\alpha y_\beta)\} g(y;\theta) dy_2 \\ &= \int_{-\infty}^{0} dy_1 \int_{-\infty}^{+\infty} g(y;\theta) dy_2 + a_{ij}^{\alpha\beta} \int_{-\infty}^{0} dy_1 \int_{-\infty}^{+\infty} (h_{ij}^{\alpha\beta}(y_1, y_2) + \theta_i\theta_j y_\alpha y_\beta) g(y;\theta) dy_2. \end{aligned}$$

The first iterated integral depends only on the pure exponential model. To see if $p_1(\theta)$ is free of the non-exponentiality terms $a_{ij}^{\alpha\beta}$, we need only to inspect whether or not the second iterated integral is zero. Since the non-exponentiality terms $a_{ij}^{\alpha\beta}$ is of order $O(n^{-1})$, we need $g(y;\theta)$ only to order $O(n^{-1/2})$ which is the $N(\theta, I)$, it then suffices to examine whether or not

$$R = \int_{-\infty}^{0} dy_1 \int_{-\infty}^{+\infty} \phi(y_1-\theta_1)\phi(y_2-\theta_2)(\theta_i\theta_j y_\alpha y_\beta + h_{ij}^{\alpha\beta}(y_1,y_2))dy_2 = 0.$$

By straightforward calculations, we obtain that, to order $O(n^{-3/2})$, $p_1(\theta)$ is free of a_{12}^{11}, a_{12}^{12}, a_{22}^{11}, a_{22}^{12}, a_{22}^{22}; but $p_1(\theta)$ depends on the other non-exponentiality terms $a_{11}^{11}, a_{11}^{12}, a_{11}^{22}, a_{12}^{22}$.

To see this, we examine two cases for illustrations. For example, for the non-zero non-exponentiality term a_{12}^{12}, the adjustment function is $h_{12}^{12}(y_1,y_2) = -(y_1^2-1)(y_2^2-1)$, and

$$\begin{aligned} R &= \int_{-\infty}^{0} dy_1 \int_{-\infty}^{+\infty} \phi(y_1-\theta_1)\phi(y_2-\theta_2)\{\theta_1\theta_2 y_1 y_2 - (y_1^2-1)(y_2^2-1)\}dy_2 \\ &= \int_{-\infty}^{0} [\theta_1\theta_2^2 y_1 - \theta_2^2(y_1^2-1)]\phi(y_1-\theta_1)dy_1. \end{aligned}$$

By Property 2,

$$\begin{aligned} &\textstyle\int_{-\infty}^{0} \phi(y_1-\theta_1)dy_1 = \Phi(-\theta_1), \\ &\textstyle\int_{-\infty}^{0} y_1\phi(y_1-\theta_1)dy_1 = -\phi(-\theta_1) + \theta_1\Phi(-\theta_1), \\ &\textstyle\int_{-\infty}^{0} y_1^2\phi(y_1-\theta_1)dy_1 = -\theta_1\phi(-\theta_1) + (\theta_1^2+1)\Phi(-\theta_1), \end{aligned}$$

we then obtain $R = 0$, therefore, $p_1(\theta)$ is free of a_{12}^{12} to order $O(n^{-3/2})$. However, For the non-zero non-exponentiality term a_{12}^{22}, the adjustment function $h_{12}^{22}(y_1,y_2) = -y_1y_2^3 + 2y_1y_2$, and then

$$\begin{aligned} R &= \int_{-\infty}^{0} dy_1 \int_{-\infty}^{+\infty} \phi(y_1-\theta_1)\phi(y_2-\theta_2)(\theta_1\theta_2 y_2^2 - y_1y_2^3 + 2y_1y_2)dy_2 \\ &= \int_{-\infty}^{0} [\theta_1\theta_2^3 + \theta_1\theta_2 - (\theta_2^3+\theta_2)y_1]\phi(y_1-\theta_1)dy_1 \\ &= (\theta_2^3+\theta_2)\phi(-\theta_1) \\ &\neq 0, \end{aligned}$$

therefore, $p_1(\theta)$ depends on a_{12}^{22}.

For the observed marginal significance function $p_2(\theta)$ we can conduct a similar discussion and obtain that $p_2(\theta)$ is not free of all non-exponentiality terms to order $O(n^{-3/2})$.

5 Discussion on Transformations

In Section 4.2 we established the exponential approximation by a series of transformations of the parameter and variable. In each stage we considered the transformations such that the model is converted to the one that has partial characteristics of an exponential model, and at the final stage the terms (called the non-exponentiality terms) that cannot be transformed to the form of those in an exponential model are used to measure the departure of the model from an exponential family. An important questions then arises: are those non-exponentiality terms unique?

In this section, we investigate this problem and find that the exponential type approximations have a unique form for non-exponentiality terms if the transformations that are used are polynomials in the variable and parameter. Notice that the procedure in Section 4.2 is exactly the same as that in Section 3.2. For simplicity we discuss only the form of the transformations for the model with a scalar parameter and scalar variable.

Consider the model given in Section 3.1 and its Taylor series expansion around the point (θ_0, y_0). Suppose that the coefficients of $l(\theta; y)$ are $a_{ij} = (\partial^{i+j}/\partial\theta^i \partial y^j) l(\theta; y)|_{(\theta_0, y_0)}$. We intend to transform the parameter and variable such that the new coefficient matrix is that of an exponential model as much as possible. Let the transformations be

$$\begin{aligned} \theta - \theta_0 &= A_1\phi + A_2\phi^2 + A_3\phi^3, \\ y - y_0 &= B_1 x + B_2 x^2 + B_3 x^3, \end{aligned} \qquad (5.1)$$

and the transformed model has coefficients $A_{ij} = (\partial^{i+j}/\partial\phi^i \partial x^j) l(\phi; x)|_{(0,0)}$ with the expansion point being (0,0); we then have the following:

Property 4. *If transformations (5.1) produce to order $O(n^{-3/2})$,*

$A_{11} = 1,$

$A_{12} = 0, A_{13} = 0,$

$A_{21} = 0, A_{31} = 0,$

then

$$A_{22} = a_{11}^{-3}(a_{11}a_{22} - a_{12}a_{21}).$$

Proof. First note that for the log density functions of the initial model and the transformed model we have $l(\phi; x) = l(\theta; y) + c$, where c is free of θ or ϕ. By differentiation we can then express the new coefficients A_{ij} in terms of the initial coefficients a_{ij}.

By taking the second mixed derivative

$$\begin{aligned} \frac{\partial^2 l(\phi; x)}{\partial\phi\partial x} &= \frac{\partial^2 l(\theta; y)}{\partial\theta\partial y} \cdot \frac{\partial y}{\partial x} \cdot \frac{\partial\theta}{\partial\phi} \\ &= \frac{\partial^2 l(\theta; y)}{\partial\theta\partial y} \cdot (B_1 + 2B_2 x + 3B_3 x^2) \cdot (A_1 + 2A_2\phi + 3A_3\phi^2), \end{aligned}$$

we obtain $A_{11} = a_{11}A_1B_1$.

Taking the third mixed derivative

$$\begin{aligned}
\frac{\partial^3 l(\phi;x)}{\partial\phi^2\partial x} &= \frac{\partial^3 l(\theta;y)}{\partial\theta^2\partial y}\cdot\frac{\partial y}{\partial x}\cdot\left(\frac{\partial\theta}{\partial\phi}\right)^2 + \frac{\partial^2 l(\theta;y)}{\partial\theta\partial y}\cdot\frac{\partial y}{\partial x}\cdot\frac{\partial^2\theta}{\partial\phi^2}\\
&= \frac{\partial^3 l(\theta;y)}{\partial\theta^2\partial y}\cdot(B_1+2B_2x+3B_3x^2)\cdot(A_1+2A_2\phi+3A_3\phi^2)^2\\
&\quad+\frac{\partial^2 l(\theta;y)}{\partial\theta\partial y}\cdot(B_1+2B_2x+3B_3x^2)\cdot(2A_2+6A_3\phi),
\end{aligned}$$

we then obtain $A_{21}=a_{21}B_1A_1^2+2a_{11}B_1A_2$.

Now consider the fourth mixed derivative

$$\begin{aligned}
\frac{\partial^4 l(\phi;x)}{\partial\phi^3\partial x} &= \frac{\partial^4 l(\theta;y)}{\partial\theta^3\partial y}\cdot\frac{\partial y}{\partial x}\cdot\left(\frac{\partial\theta}{\partial\phi}\right)^3 + 3\frac{\partial^3 l(\theta;y)}{\partial\theta^2\partial y}\cdot\frac{\partial y}{\partial x}\cdot\frac{\partial\theta}{\partial\phi}\cdot\frac{\partial^2\theta}{\partial\phi^2}\\
&\quad+\frac{\partial^2 l(\theta;y)}{\partial\theta\partial y}\cdot\frac{\partial y}{\partial x}\cdot\frac{\partial^3\theta}{\partial\phi^3}\\
&= \frac{\partial^4 l(\theta;y)}{\partial\theta^3\partial y}\cdot(B_1+2B_2x+3B_3x^2)\cdot(A_1+2A_2\phi+3A_3\phi^2)^3\\
&\quad+3\frac{\partial^3 l(\theta;y)}{\partial\theta^2\partial y}\cdot(B_1+2B_2x+3B_3x^2)\cdot(A_1+2A_2\phi+3A_3\phi^2)\\
&\quad\cdot(2A_2+6A_3\phi)+\frac{\partial^2 l(\theta;y)}{\partial\theta\partial y}\cdot(B_1+2B_2x+3B_3x^2)\cdot 6A_3,
\end{aligned}$$

accordingly, we obtain $A_{31}=a_{31}B_1A_1^3+6a_{21}B_1A_1A_2+6a_{11}B_1A_3$.

Similarly, we can express A_{12} and A_{13} in terms of the initial coefficients a_{ij}:

$A_{12}=a_{12}A_1B_1^2+2a_{11}A_1B_2,$

$A_{13}=a_{13}A_1B_1^3+6a_{12}A_1B_1B_2+6a_{11}A_1B_3.$

If the transformed model is "exponential-like", that is, we have

$A_{11}=1, A_{12}=A_{13}=0, A_{21}=A_{31}=0,$

then, by solving the equations above, we obtain

$$\begin{aligned}
A_2 &= -\frac{1}{2}a_{11}^{-1}a_{21}A_1^2,\quad A_3=\frac{1}{6}a_{11}^{-2}(3a_{21}^2-a_{11}a_{31})A_1^3,\\
B_1 &= a_{11}^{-1}A_1^{-1},\quad B_2=-\frac{1}{2}a_{11}^{-3}a_{12}A_1^{-2},\quad B_3=\frac{1}{6}a_{11}^{-5}(3a_{12}^2-a_{11}a_{13})A_1^{-3};
\end{aligned}\tag{5.2}$$

that is, the coefficients in the transformations (5.1) are partly determined by the initial coefficients a_{ij}.

Finally, we examine the expression of A_{22} in terms of a'_{ij}s when the transformed model is forced to be "exponential-like". Consider the fourth mixed derivative of the log density functions

$$
\begin{aligned}
\frac{\partial^4 l(\phi;x)}{\partial\phi^2\partial x^2} &= \frac{\partial^4 l(\theta;y)}{\partial\theta^2\partial y^2}\cdot\left(\frac{\partial y}{\partial x}\right)^2\cdot\left(\frac{\partial\theta}{\partial\phi}\right)^2+\frac{\partial^3 l(\theta;y)}{\partial\theta^2\partial y}\cdot\frac{\partial^2 y}{\partial x^2}\cdot\left(\frac{\partial\theta}{\partial\phi}\right)^2\\
&\quad+\frac{\partial^3 l(\theta;y)}{\partial\theta\partial y^2}\cdot\left(\frac{\partial y}{\partial x}\right)^2\cdot\frac{\partial^2\theta}{\partial\phi^2}+\frac{\partial^2 l(\theta;y)}{\partial\theta\partial y}\cdot\frac{\partial^2 y}{\partial x^2}\cdot\frac{\partial^2\theta}{\partial\phi^2}\\
&= \frac{\partial^4 l(\theta;y)}{\partial\theta^2\partial y^2}\cdot(B_1+2B_2x+3B_3x^2)^2\cdot(A_1+2A_2\phi+3A_3\phi^2)^2\\
&\quad+\frac{\partial^3 l(\theta;y)}{\partial\theta^2\partial y}\cdot(2B_2+6B_3x)\cdot(A_1+2A_2\phi+3A_3\phi^2)^2\\
&\quad+\frac{\partial^3 l(\theta;y)}{\partial\theta\partial y^2}\cdot(B_1+2B_2x+3B_3x^2)^2\cdot(2A_2+6A_3\phi)\\
&\quad+\frac{\partial^2 l(\theta;y)}{\partial\theta\partial y}\cdot(2B_2+6B_3x)\cdot(2A_2+6A_3\phi),
\end{aligned}
$$

we obtain

$$A_{22}=a_{22}B_1^2A_1^2+2a_{21}B_2A_1^2+2a_{12}A_2B_1^2+4a_{11}B_2A_2.$$

From equations (5.2), we then obtain $A_{22}=a_{11}^{-3}(a_{11}a_{22}-a_{12}a_{21})$, which is the same as that in (3.3).

The calculation shows to $O(n^{-3/2})$, that we cannot force the model $l(\theta;y)$ to be an exact exponential model, that is, there is one term A_{22} left to measure the departure of the model from an exponential model.

It is easy to show that the transformations (5.1) are unique if we need the coefficients A_{ij} to be the same as those in (3.3) in Section 3.2. That is, if we need further that $A_{20}=-1$, then $A_1=(-a_{20})^{-1/2}$, and A_i, B_i are identical to those listed in (3.3) in Section 3.2.

6 Discussion

In higher order asymptotic inferences, exponential models play a remarkable role due to their ability of offering tractable but accurate approximations to general statistical models. A further advantage of applying exponential approximations is due to the separatability of model type parameters from the measure of departure (i.e., non-exponentiality terms) from the exponential model. As a major objective of higher order asymptotic inference concerns computing a p-value or an observed significance function at the data point, it is fundamental to understand how the non-exponentiality terms would impact observed significant functions. Andrews, Fraser and Wong (2005) investigated this important problem for univariate models, and found that observed significant functions do not depend on the non-exponentiality term to the third order. However, it is not clear whether this property holds for multivariate models. In contrast to univariate case with the third order approximation, there are more than one non-exponentiality term involved to facilitate departure of a multivariate model from the exponential model, and this may significantly change the nature of observed marginal significance functions. In this paper, we explored the relationship between observed marginal significant functions and the non-exponentiality terms, and revealed the intrinsic difference between univariate and multivariate models. This finding is of significant importance in understanding higher order asymptotic inference on multivariate data. It sheds light on the complexity of handling multivariate models with higher order inference techniques.

Acknowledgement

The research was supported by the Natural Sciences and Engineering Research Council of Canada.

References

[1] Andrews, D.F., Fraser, D.A.S., and Wong, A. (2005). Computation of distribution functions from likelihood information near observed data. *Journal of Statistical Planning and Inference*, **134**, *180-193.*

[2] Barndorff-Nielsen, O.E. and Cox, D.R. (1989). *Asymptotic Techniques for Use in Statistics.* Chapman and Hall, London.

[3] Barndorff-Nielsen, O.E. and Cox, D.R. (1994). *Inference and Asymptotics.* Chapman and Hall, London.

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

[5] Cakmak, S., Fraser, D.A.S., McDunnough, P., Reid, N., and Yuan, X. (1998). Likelihood centered asymptotic model: exponential and location model versions. *Journal Statistical Planning and Inference*, **66**, *211-222.*

[6] Daniels, H.E. (1987). Tail probability approximations. *Internal Statistical Review*, **54**, *37-48.*

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

[8] Fraser, D.A.S. and Reid, N. (1993). Simple asymptotic connections between density and cumulant generating functions leading to accurate approximations for distribution functions. *Statistica Sinica*, **3**, *67-82.*

[9] Fraser, D.A.S., Reid, N., Li, R., and Wong, A. (2003). On bridging the singularities of p-value formulas from likelihood analysis. *Journal of Statistical Research*, **37**, *1-15.*

[10] Lugannani, R. and Rice, S. (1980). Saddlepoint approximation for the distribution of the sum of independent random variables. *Advanced Applied Probability*, **12**, *475-490.*

[11] McCullagh, P. (1987). *Tensor Methods in Statistics.* Chapman & Hall, London.

[12] Pierce, D.A. and Peters, D. (1992). Practical use of higher order asymptotics for multiparameter exponential families (with discussion). *Journal of the Royal Statistical Society, Series B* , **54**, *701-738.*

[13] Reid, N. (1988). Saddlepoint methods and statistical inference. *Statistical Science*, **3**, *213-237.*

[14] Reid, N. (1996). Likelihood and higher order approximations to tail areas: A review and annotated bibliography. *Canadian Journal of Statistics*, **24**, *141-166.*

[15] Robinson, J. (1982). Saddlepoint approximations for permutation tests and confidence intervals. *Journal of the Royal Statistical Society, Series B*, **44**, *91-101.*