Statistica Sinica Preprint No: SS-2022-0016	
Title	Asymptotic Behaviour of the Modified Likelihood Root
Manuscript ID	SS-2022-0016
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0016
Complete List of Authors	Yanbo Tang and Nancy Reid
Corresponding Authors	Yanbo Tang
E-mails	yanbotang17@gmail.com
Notice: Accepted version subject to English editing.	

Statistica Sinica

Asymptotic Behaviour of the Modified Likelihood Root

Yanbo Tang and Nancy Reid

Imperial College London and University of Toronto

Abstract: We examine the normal approximation to the distribution of the modified likelihood root, an inferential tool of higher-order asymptotic theory, for the linear exponential and location-scale families. We show that the modified likelihood root, $r^\star$, can be expressed as a location and scale adjustment to the likelihood root, r, to $O_p(n^{-3/2})$, and more generally can be expressed as a polynomial in r. We discuss some extensions of these results to the high-dimensional regime.

Key words and phrases: Statistical Inference, Higher-Order Asymptotics, High-Dimensional Statistics, Location-Scale Families, Linear Exponential Families.

1. Introduction

The use of p-values, although sometimes controversial, has become a key part of modern statistical science, for example as a building block in various multiple testing procedures used in statistical genetics, where large numbers of hypotheses are simultaneously considered. In most circumstances

p-values are not exact but are calculated from the limiting distribution of a test statistic. The usual test statistics provided in statistical software, such as the likelihood ratio test, Wald test and score test, all have a common known limiting distribution and are accurate to the first order, meaning that the approximation error is $O(n^{-1/2})$. However, in the small sample setting or when the number of nuisance parameters is high relative to the number of observations, this trio of tests may perform poorly. An improved test statistic, $r^{\star}$, a modified version of the likelihood root, can be used for likelihood-based inference for scalar parameters of interest. It produces more accurate p-values than the first order approximations of the test statistics. The accuracy of the p-values generated by $r^{\star}$ can be quite remarkable as demonstrated in Brazzale et al. (2007, §3.2) and the references within, see also Pierce and Peters (1992) for a discussion focused on the linear exponential family.

Given the importance that p-values play in statistical inference, the exact mechanism through which $r^{\star}$ generates more accurate p-values warrants a careful examination. We provide insight into the behaviour of $r^{\star}$ by expressing it as a formal asymptotic expansion, showing that it is asymptotically linear in the likelihood root, which we introduce below.

We assume the data $y = (y_1, \cdots, y_n)^{\top}$ are generated independently from a model parametrized by $\theta = (\psi, \lambda)$ where ψ is a scalar parameter of interest, and λ is the nuisance parameter. We denote the log-likelihood function by $l(\psi, \lambda; y)$, and the data generating parameter by $\theta_0 = (\psi_0, \lambda_0)$. Let $\hat{\lambda}_\psi$ denotes the constrained maximum likelihood estimate; i.e. the value of λ that maximises the log-likelihood function for fixed ψ. The profile log-likelihood function,

$$l_{\mathrm{p}}(\psi; y) := \sup_{\lambda} l(\psi, \lambda; y) = l(\psi, \hat{\lambda}_\psi; y),$$

accounts for the presence of nuisance parameters through constrained maximization. Under suitable regularity conditions (Barndorff-Nielsen and Cox, 1994, §3.4),

$$w(\psi_0; y) := 2\{l_{\mathrm{p}}(\hat{\psi}) - l_{\mathrm{p}}(\psi_0)\} \xrightarrow{d} \chi^2_1,$$

where χ^2_1 is a random variable distributed as chi-squared with one degree of freedom. Equivalently, the log-likelihood root

$$r(\psi_0; y) := \mathrm{sign}(\hat{\psi} - \psi_0)[2\{l_{\mathrm{p}}(\hat{\psi}; y) - l_{\mathrm{p}}(\psi_0; y)\}]^{\frac{1}{2}} \xrightarrow{d} Z, \tag{1.1}$$

where the random variable Z has the standard normal distribution (Barndorff-Nielsen and Cox, 1994, §2.3).

By adding a correction term to r, we obtain the modified likelihood root

$$r^\star(\psi_0; y) = r(\psi_0; y) + \frac{1}{r(\psi_0; y)} \log \left\{ \frac{q(\psi_0; y)}{r(\psi_0; y)} \right\}; \tag{1.2}$$

the specific form of q depends on the model. It has been shown under regularity conditions that the normal approximation to the distribution of $r^\star$ is accurate to $O(n^{-3/2})$ (Barndorff-Nielsen and Cox, 1994), whereas the normal approximation to the distribution of r is only accurate to $O(n^{-1/2})$. For a vector parameter of interest, the Bartlett correction is used instead of $r^\star$. It is defined by dividing the likelihood ratio:

$$w_{corr} := w(\psi_0; y)/\{1 + B_{corr}(\psi_0)\}$$

such that $E[w_{corr}(\psi_0; y)] = \dim(\psi_0)\{1+O(n^{-2})\}$. When the model contains no nuisance parameters and ψ is one-dimensional, the Bartlett correction simultaneously adjusts for the bias in location and scale of r by correcting for the non-centrality parameter of the limiting chi-squared distribution and can be related to $r^\star$, specifically if the location bias of r is $\tilde{A}/n^{1/2}$ and the scale bias is $\{1 + \tilde{B}/n\}$, then the Bartlett correction will be $(1 + \tilde{B}/n)^2 + \tilde{A}^2/n$.

The mechanism through which $r^\star$ achieves this accuracy is not entirely transparent. Cakmak et al. (1998) show that in models with no nuisance parameters, $r^\star$ corrects for the location and scale bias present in r up to the third order. We generalize this result to models with nuisance parameters, showing that even for such models the modified likelihood root is equivalent to a location and scale correction to r up to the third order, see Remark

2. We then further show that the adjustment factor can be expressed as a polynomial in r up to arbitrary order. We discuss the asymptotic behaviour of $r^\star$ when the number of parameters increases with the number of observations. We show that the adjustment factor $(1/r)\log(q/r)$ is potentially of the same asymptotic order as r in the high-dimensional regime, agreeing with results in Tang and Reid (2020), where results for the modified likelihood root was obtained for general models. The expansions obtained in this paper provides a more precise characterization of the asymptotic behaviour of the adjustments than what is available in Tang and Reid (2020), as we derive a polynomial asymptotic series for the adjustments up to an arbitrary order. These expansions also allow us to extend the results of Cakmak et al. (1998) to models with nuisance parameters whereas the results in Tang and Reid (2020) do not allow for this extension. We also show that $r = q + q^2A/n^{1/2} + q^3B/n + O_p(n^{-3/2})$, in the linear exponential and location scale families, where A and B are $O_p(1)$, and may depend on y and θ_0. This result provides a simple proof that the normal approximation to the distribution of $r^\star$ has relative error $O(n^{-3/2})$ in the presence of nuisance parameters (Brazzale et al., 2007; Davison and Reid, 2021, §8.5), which to our knowledge has not yet been established.

We focus our analysis on the location-scale and linear exponential family, as the expression for q in (1.2) is explicitly available. We begin with background details on $r^\star$, the linear exponential family and the location-scale family in §2. We present our main theorem in §3 showing that the adjustment factor can be expressed as a polynomial in r up to arbitrary order, and following that a corollary that r can be expressed as a third order polynomial in q in §3.1. Most of the discussions involved will be in the fixed dimensional case, where p is held constant and $n \to \infty$. The discussion for the high-dimensional case is contained in §4, while certain technical details are deferred to §6.4 and §6.5. We present some simulations in §5 which illustrate the accuracy of the approximations to $r^\star$ and conclude with some additional proof details in §6.

2. Background and Assumptions

We begin by introducing some notation. Derivatives of the log-likelihood function l are denoted by subscripts, for example $l_{\psi\lambda\lambda}(\theta)$ represents the matrix whose components are $[l_{\psi\lambda\lambda}(\theta)]_{rs} = \partial^3 l(\theta)/\partial\psi\partial\lambda_r\partial\lambda_s$. We use j to denote the observed information matrix, $j(\theta) = -l_{\theta\theta}(\theta)$; subscripts placed on j denote sub-matrices of j and we let $j_{\mathrm{p}} = -d^2 l_{\mathrm{p}}(\psi)/d\psi^2$. A tilde on any quantity denotes evaluation at the constrained maximum likelihood estimate, $(\psi, \hat{\lambda}_\psi)$, for example $\tilde{j}_{\lambda\lambda} = j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)$, and a hat denotes that it is evaluated at the global maximum likelihood estimate, $\hat{\theta}$, thus $\hat{\jmath} = j(\hat{\psi}, \hat{\lambda}) = j(\hat{\theta})$.

We use $d/d\psi$ to denote the total derivative with respect to ψ and $\partial/\partial\psi$ to denote the partial derivative with respect to ψ. The k-th derivative of the profile log-likelihood is $\zeta_k(\psi) = d^k l_{\mathrm{p}}(\psi)/d\psi^k$ and k-th total derivative of the log determinant of the information matrix is

$$\gamma_k(\psi) = d^k \log\{|j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|\}/d\psi^k.$$

We define the k-th quasi-cumulant of the profile log-likelihood function as:

$$\kappa_k(\psi) = \frac{\zeta_k(\psi)}{\{-\zeta_2(\psi)\}^{k/2}}. \tag{2.3}$$

In the sequel we suppress the dependence of functions on the data y in the notation, for example $r(\psi_0; y)$ will simply be written as $r(\psi_0)$. We write $\sigma_{\max}(A)$ for the maximum singular value of a matrix A. The following inequalities will prove useful, for a square matrices A, and a positive definite square matrix B of compatible dimensions:

$$|\operatorname{tr}(AB)| \le \|A\|_F \, \|B\|_F \,, \quad |\operatorname{tr}(AB)| \le \sigma_{\max}(A) \operatorname{tr}(B),$$

where $\|A\|_F = (\sum_{i,j} A_{ij}^2)^{1/2}$ is the Frobenius norm and the latter inequality is a consequence of the von Neumann trace inequality (Mirsky, 1975). For a vector v, we let $\|v\|_p$ denote the L^p norm.

We assume the following conditions on the model:

Assumption 1. The k-th order partial derivatives of $l(\theta; y)$ with respect to the elements of θ are $O_p(n)$ for all integers $k > 1$ when evaluated at $\hat{\theta}$.

Assumption 2. $\hat{\theta} - \theta_0 = O_p(n^{-1/2})$

Assumption 3. The eigenvalues of $\hat{\jmath}/n$, $j(\hat{\theta}_{\psi_0})/n$, $n\hat{\jmath}^{-1}$ and $nj^{-1}(\hat{\theta}_{\psi_0})$ are positive and $O_p(1)$.

Assumption 1 ensures that the likelihood derivatives grow at the usual rate when evaluated at the maximum likelihood estimate. Assumption 2 states that the maximum likelihood estimate is $n^{1/2}$-consistent for the true parameter value: this rate of consistency is typically achieved for most well behaved parametric models (van der Vaart, 1998, §5). Finally Assumption 3 ensures that the asymptotic covariance matrix for the maximum likelihood estimate is well defined. For the regression problems that we consider, Assumption 3 is equivalent to a restriction on the eigenvalues on the Gramian matrix, $X^\top X$, as well as a lower bound on the variance of the fitted values.

2.1 Linear Exponential Family

Let X be a $n \times p$ matrix of covariates with (i, j) entry, $x_{i,j}$, and i-th row $x_i^\top$. We assume the density of y_i is a member of the full exponential family with log-likelihood function

$$l(\psi, \lambda; y_i, x_i) = \psi u(x_{i,p}, y_i) + \lambda^\top v(x_i, y_i) - c_i(\psi, \lambda) + h(x_i, y_i), \tag{2.4}$$

where $u(x_{i,p}, y_i)$ is a scalar sufficient statistic associated with ψ and

$$v(x_i, y_i) = \{v_1(x_{i,1}, y_i), \cdots, v_{p-1}(x_{i,p-1}, y_i)\}^\top,$$

is the vector of sufficient statistics associated with the nuisance parameters.

In this model, q in (1.2) takes the form

$$q(\psi_0) = t(\psi_0)\rho(\psi_0), \text{ where}$$

$$t(\psi_0) = (\hat{\psi} - \psi_0) j_{\mathrm{p}}^{1/2}(\hat{\psi}), \tag{2.5}$$

is the Wald statistic for testing $\psi = \psi_0$ and

$$\rho(\psi_0) = \{|j_{\lambda\lambda}(\hat{\theta})|/|j_{\lambda\lambda}(\hat{\theta}_{\psi_0})|\}^{1/2},$$

where $j_{\lambda\lambda}(\theta)$ is the $(p-1) \times (p-1)$ sub-matrix of $j(\theta)$ associated with the nuisance parameters (Brazzale et al., 2007, §8.6.1). We follow Pierce and Peters (1992) and write

$$r^\star = r + r_{np} + r_{inf},$$

where r_{np} is a nuisance parameter adjustment and r_{inf} is an information adjustment. This partitioning of the adjustments will be helpful to the analysis of the asymptotic behaviour of $r^\star$. For this model,

$$r_{np} = \frac{1}{r}\log(\rho), \quad r_{inf} = \frac{1}{r}\log\left(\frac{t}{r}\right). \tag{2.6}$$

2.2 Location-Scale Family

For a linear regression model based on the location scale-family, the model is

$$y_i = x_i^\top \beta + \sigma \epsilon_i, \tag{2.7}$$

where the errors ϵ_i are assumed independent and identically distributed from a known distribution with continuous density $f(\epsilon)$. The model is parametrized by $\theta = (\beta, \sigma)$, we assume that the parameter of interest, ψ, is a component of β. For this model

$$\begin{aligned} q(\psi_0) &= s(\psi_0)/\rho(\psi_0), \text{ where} \\ s(\psi_0) &= \zeta_1(\psi_0)/j_{\mathrm{p}}^{1/2}(\hat{\psi}), \end{aligned} \tag{2.8}$$

s is the standard score test statistic and ρ is defined above. For this model

$$r_{np} = -\frac{1}{r}\log(\rho), \quad r_{inf} = \frac{1}{r}\log\left(\frac{s}{r}\right).$$

2.3 General Models

For general models the expression for q in (1.2) (see Reid (2003)) does not simplify to the Wald or score test, but depends on the derivatives of the log-likelihood in the sample space as well as in the parameter space.

3. Formal expansions of r_{inf} and r_{np}

In this section we obtain formal asymptotic expansions of r_{inf} and r_{np}, which detail the relationship between r and $r^{\star}$ in the linear exponential and location scale families, respectively. Analogous results to Theorem 1 and 2 were obtained by Cakmak et al. (1998) in the case of no nuisance parameters. The expansions for r_{inf} and r_{np} show that $r^{\star}$ is asymptotically equivalent to a location and scale adjustment to the likelihood root r.

We begin by establishing relationships between r, t and s.

Lemma 1. *Under Assumptions 1–3, for r, t and s defined in (1.1), (2.5) and (2.8):*

$$t = r\left\{1 + \frac{A_1}{n^{1/2}}r + \frac{B_1}{n}r^2 + O_p(n^{-3/2})\right\},$$
$$s = t\left\{1 + \frac{A_2}{n^{1/2}}t + \frac{B_2}{n}t^2 + O_p(n^{-3/2})\right\},$$

where

$$A_1 = -\frac{n^{1/2}}{6}\kappa_3(\hat{\psi}), \quad B_1 = \frac{n}{24}\kappa_4(\hat{\psi}) + \frac{5n}{72}\kappa_3^2(\hat{\psi}),$$
$$A_2 = \frac{n^{1/2}\kappa_3(\hat{\psi})}{2}, \qquad B_2 = -\frac{n\kappa_4(\hat{\psi})}{6}. \tag{3.9}$$

Proof. We begin by deriving the relationship between r and t, defined in (1.1) and (2.5) :

$$r^2 = 2\Big\{l_{\rm p}(\hat{\psi}) - l_{\rm p}(\psi_0)\Big\},$$

$$
\begin{aligned}
&= 2\Big\{(\hat{\psi}-\psi_0)\zeta_1(\hat{\psi}) - \frac{(\hat{\psi}-\psi_0)^2}{2}\zeta_2(\hat{\psi}) \\
&\quad + \frac{(\hat{\psi}-\psi_0)^3}{6}\zeta_3(\hat{\psi}) + \frac{(\hat{\psi}-\psi_0)^4}{24}\zeta_4(\hat{\psi}) + +O_p(n^{-3/2})\Big\} \\
&= t^2\Big\{1 + \frac{\kappa_3(\hat{\psi})}{3}t - \frac{\kappa_4(\hat{\psi})}{12}t^2 + O_p(n^{-3/2})\Big\}.
\end{aligned}
$$

The Taylor-series expansion for $(1+x)^{1/2}$ gives

$$
r = t\Big\{1 + \frac{\kappa_3(\hat{\psi})}{6}t - \frac{\kappa_4(\hat{\psi})}{24}t^2 - \frac{\kappa_3^2(\hat{\psi})}{72}t^2 + O_p(n^{-3/2})\Big\},
$$

which implies

$$
\begin{aligned}
t &= r\Big\{1 + \frac{\kappa_3(\hat{\psi})}{6}t - \frac{\kappa_4(\hat{\psi})}{24}t^2 - \frac{\kappa_3^2(\hat{\psi})}{72}t^2 + O_p(n^{-3/2})\Big\}^{-1} \\
&= r\Big\{1 - \frac{\kappa_3(\hat{\psi})}{6}t + \frac{\kappa_4(\hat{\psi})}{24}t^2 + \frac{\kappa_3^2(\hat{\psi})}{72}t^2 + \frac{\kappa_3^2(\hat{\psi})}{36}t^2 + O_p(n^{-3/2})\Big\} \\
&= r\Big\{1 - \frac{\kappa_3(\hat{\psi})}{6}t + \frac{\kappa_4(\hat{\psi})}{24}t^2 + \frac{\kappa_3^2(\hat{\psi})}{24}t^2 + O_p(n^{-3/2})\Big\}. \qquad (3.10)
\end{aligned}
$$

As t appears on both sides of the equation, we iteratively solve by substitution

$$
t = r\Big[1 - \frac{\kappa_3(\hat{\psi})}{6}\Big\{\Big(1 - \frac{\kappa_3(\hat{\psi})}{6}t\Big)r\Big\} + \frac{1}{24}\Big\{\kappa_3^2(\hat{\psi}) + \kappa_4(\hat{\psi})\Big\}r^2 + O_p(n^{-3/2})\Big] \qquad (3.11)
$$

$$
\begin{aligned}
&= r\Big[1 - \frac{\kappa_3(\hat{\psi})}{6}r + \Big\{\frac{\kappa_3^2(\hat{\psi})}{36}r\Big\}t + \frac{1}{24}\Big\{\kappa_3^2(\hat{\psi}) + \kappa_4(\hat{\psi})\Big\}r^2 + O_p(n^{-3/2})\Big] \\
&= r\Big\{1 - \frac{\kappa_3(\hat{\psi})}{6}r + \frac{5}{72}\kappa_3^2(\hat{\psi})r^2 + \frac{1}{24}\kappa_4(\hat{\psi})r^2 + O_p(n^{-3/2})\Big\} \qquad (3.12)
\end{aligned}
$$

For the expansion of s, we have

$$
s = \frac{\zeta_1(\psi_0)}{j_p^{1/2}(\hat{\psi})}
$$

$$
\begin{aligned}
&= \frac{1}{j_p^{1/2}(\hat{\psi})}\Big\{\zeta_1(\hat{\psi}) - \zeta_2(\hat{\psi})(\hat{\psi}-\psi_0) + \frac{\zeta_3(\hat{\psi})}{2}(\hat{\psi}-\psi_0)^2 - \frac{\zeta_4(\hat{\psi})}{6}(\hat{\psi}-\psi_0)^3 + O_p(n^{-3/2})\Big\} \\
&= t + \frac{\kappa_3(\hat{\psi})}{2}t^2 - \frac{\kappa_4(\hat{\psi})}{6}t^3 + O_p(n^{-3/2}).
\end{aligned}
$$

□

Remark 1. It is known that the Wald, score and the likelihood root statistics are first order equivalent; they converge to the same limit as $n \to \infty$ (Cox and Hinkley, 1974). The above expansions provide more precise statements.

Theorem 1. *Under Assumptions 1–3, for the linear exponential family,*

$$r_{inf} = -\frac{1}{6}\kappa_3(\hat{\psi}) + \Big\{\frac{1}{24}\kappa_4(\hat{\psi}) + \frac{4}{72}\kappa_3^2(\hat{\psi})\Big\}r + O_p(n^{-3/2}), \tag{3.13}$$

and for the location-scale family

$$r_{inf} = \frac{1}{3}\kappa_3(\hat{\psi}) - \Big\{\frac{3}{24}\kappa_4(\hat{\psi}) + \frac{11}{72}\kappa_3^2(\hat{\psi})\Big\}r + O_p(n^{-3/2}). \tag{3.14}$$

Theorem 2. *Under Assumptions 1–3, for the linear exponential family*

$$r_{np} = \frac{1}{2}\frac{\gamma_1(\hat{\psi})}{j_{\mathrm{p}}(\hat{\psi})^{1/2}} - \Big\{\frac{1}{12}\frac{\kappa_3(\hat{\psi})\gamma_1(\hat{\psi})}{j_{\mathrm{p}}(\hat{\psi})^{1/2}} - \frac{1}{4}\frac{\gamma_2(\hat{\psi})}{j_{\mathrm{p}}(\hat{\psi})}\Big\}r + O_p(n^{-3/2}), \tag{3.15}$$

and for the location-scale family

$$r_{np} = -\frac{1}{2}\frac{\gamma_1(\hat{\psi})}{j_{\mathrm{p}}(\hat{\psi})^{1/2}} + \Big\{\frac{1}{12}\frac{\kappa_3(\hat{\psi})\gamma_1(\hat{\psi})}{j_{\mathrm{p}}(\hat{\psi})^{1/2}} - \frac{1}{4}\frac{\gamma_2(\hat{\psi})}{j_{\mathrm{p}}(\hat{\psi})}\Big\}r + O_p(n^{-3/2}). \tag{3.16}$$

Proof of Theorems 1 and 2. **Linear Exponential Family:** Using (2.6) and Lemma 1 for the linear exponential family we have

$$\begin{aligned} r_{inf} &= \frac{1}{r}\log\left(\frac{t}{r}\right) \\ &= \frac{A_1}{n^{1/2}} + \left(\frac{B_1}{n} - \frac{A_1^2}{2n}\right) r + O_p(n^{-3/2}) \end{aligned}$$

where substitution by 3.9 give 3.13 A similar expansion can be developed for r_{np}:

$$\begin{aligned} r_{np} &= \frac{1}{2r}\log\left\{\frac{|j_{\lambda\lambda}(\hat{\psi},\hat{\lambda})|}{|j_{\lambda\lambda}(\psi_0,\hat{\lambda}_{\psi_0})|}\right\} \\ &= \frac{1}{2r}\left[\frac{\gamma_1(\hat{\psi})}{\{-\zeta_2(\hat{\psi})\}^{1/2}}t + \frac{\gamma_2(\hat{\psi})}{2\zeta_2(\hat{\psi})}t^2 + O_p\Big(n^{-3/2}\Big)\right] && (3.17) \\ &= \frac{1}{2}\left[\left(1 + \frac{A_1}{n^{1/2}}r\right)\frac{\gamma_1(\hat{\psi})}{\{-\zeta_2(\hat{\psi})\}^{1/2}} + \frac{\gamma_2(\hat{\psi})}{2\zeta_2(\hat{\psi})}r + O_p\Big(n^{-3/2}\Big)\right] && (3.18) \end{aligned}$$

using Lemma 1; substitution (3.15) gives the result.

Location-Scale Family: The expansion for the nuisance parameter adjustment r_{np} is the same as in the exponential family, except for a change in sign. From Lemma 1,

$$\begin{aligned} r_{inf} &= \frac{1}{r}\log\left[\frac{1}{r}\left\{t + \frac{\kappa_3(\hat{\psi})}{2}t^2 - \frac{\kappa_4(\hat{\psi})}{6}t^3 + O_p(n^{-3/2})\right\}\right] \\ &= \frac{1}{r}\log\left[1 + \frac{A_2}{n^{1/2}}r + \frac{B_2}{n}r^2 + \frac{\kappa_3(\hat{\psi})}{2}\left\{1 + \frac{A_2}{n^{1/2}}r\right\}^2 r - \frac{\kappa_4(\hat{\psi})}{6}r^2 + O_p(n^{-3/2})\right] \\ &= \frac{1}{3}\kappa_3(\hat{\psi}) - \left\{\frac{3}{24}\kappa_4(\hat{\psi}) + \frac{11}{72}\kappa_3^2(\hat{\psi})\right\}r + O_p(n^{-3/2}), \end{aligned}$$

substitution (3.15) gives the result. □

Remark 2. From Theorems 2 and 1, for the linear exponential family,

$$r^{\star} = -\frac{1}{6}\kappa_3(\hat{\psi}) + \frac{1}{2}\frac{\gamma_1(\hat{\psi})}{\{-\zeta_2(\hat{\psi})\}^{1/2}} + \left[1 + \frac{1}{24}\kappa_4(\hat{\psi}) + \frac{4}{72}\kappa_3^2(\hat{\psi}) - \frac{1}{12}\frac{\kappa_3(\hat{\psi})\gamma_1(\hat{\psi})}{\{-\zeta_2(\hat{\psi})\}^{1/2}} - \frac{1}{4}\frac{\gamma_2(\hat{\psi})}{\zeta_2(\hat{\psi})}\right] r + O_p(n^{-3/2}),$$

and for the location-scale family,

$$r^{\star} = \frac{1}{3}\kappa_3(\hat{\psi}) - \frac{1}{2}\frac{\gamma_1(\hat{\psi})}{\{-\zeta_2(\hat{\psi})\}^{1/2}} + \left[1 - \frac{3}{24}\kappa_4(\hat{\psi}) - \frac{11}{72}\kappa_3^2(\hat{\psi}) + \frac{1}{12}\frac{\kappa_3(\hat{\psi})\gamma_1(\hat{\psi})}{\{-\zeta_2(\hat{\psi})\}^{1/2}} + \frac{1}{4}\frac{\gamma_2(\hat{\psi})}{\zeta_2(\hat{\psi})}\right] r + O_p(n^{-3/2}).$$

From these, we obtain for both families:

$$r^{\star} = \frac{r - \tilde{A}/n^{1/2}}{\{1 + \tilde{B}/n\}} + O_p(n^{-3/2}),$$

which shows that $r^{\star}$ is a location and scaling correction of r up to the third order, where $\tilde{A}$ and $\tilde{B}$ are $O_p(1)$.

Corollary 1. *Under Assumptions 1–3, the expansions of r_{inf} and r_{np} can be extended to arbitrary order:*

$$r_{inf} = \sum_{k=1}^{m} \frac{A_k}{n^{k/2}} r^{k-1} + O_p(n^{-(m+1)/2}), \quad r_{np} = \sum_{k=1}^{m} \frac{B_k}{n^{k/2}} r^{k-1} + O_p(n^{-(m+1)/2}), \tag{3.19}$$

for arbitrary $m \in \mathbb{N}$ where the coefficients $A_k = O_p(1)$ and $B_k = O_p(1)$.

From the substitution argument employed in (3.11) to (3.12), we can deduce that for the expansion of r_{inf} the coefficient of r^{k-1} is $O_p(1)$ and

$$A_k = n^{k/2} \sum_{j=1}^{k} \sum_{\{i_1,\cdots,i_j\}\in S_j} Z_j(\{i_1,\cdots,i_j\}) \prod_{l=1}^{j} \hat{\kappa}_{i_l},$$

where indices $\{i_1, \cdots, i_j\}$ take values in $\{3, \cdots, k+2\}$, S_j is the set of all indices such that $\sum_{l=1}^{j}(i_l - 2) = k$, and $Z_j(\cdot)$ is a combinatorial constant.

From equation (3.18), when grouping terms in powers of r, we obtain the following expressions for the coefficients in the expansion of r_{np}:

$$B_k = \sum_{j=0}^{k-1} n^{(k-j)/2} C_j \frac{\gamma_{k-j}(\hat{\psi})}{\{-\zeta_2(\hat{\psi})\}^{(k-j)/2}} \quad \text{and} \quad C_k = \sum_{m=1}^{k} \sum_{\{i_1,\cdots,i_m\}\in D_{m,k}} \prod_{l=1}^{m} A_{i_l},$$

where $C_0 = 1$, the indices $\{i_1, \cdots, i_m\}$ range from 1 to k and the set $D_{m,k}$ is the set of all indices $\{i_1, \cdots, i_m\}$ such that $\sum_{l=1}^{m} i_l = k$. Some examples of terms which appear in A_k, B_k and C_k are given in Figure 1.

3.1 r as a series in q

In this section, we obtain an expansion of r as a polynomial in q, which is key to the proof of the accuracy of the normal approximation to the distribution of $r^{\star}$, see Brazzale et al. (2007, §8) and Davison and Reid (2021, §4.4). To our knowledge this has not been established for models with nuisance parameters, and is a direct consequence of Lemma 1. We do not give the exact forms of the coefficients which appear in the polynomial

k	A_k	B_k	C_k
1	$n^{1/2}\kappa_3(\hat{\psi})$	$\frac{n^{1/2}\gamma_1(\hat{\psi})}{\{-\zeta_2(\hat{\psi})\}^{1/2}}$	A_1
2	$n\kappa_3^2(\hat{\psi})$, $n\kappa_4(\hat{\psi})$	$\frac{n^{1/2}C_1\gamma_1(\hat{\psi})}{\{-\zeta_2(\hat{\psi})\}^{1/2}}$, $\frac{n\gamma_2(\hat{\psi})}{-\zeta_2(\hat{\psi})}$	A_2, A_1^2
3	$n^{3/2}\kappa_3^3(\hat{\psi})$, $n^{3/2}\kappa_3(\hat{\psi})\kappa_4(\hat{\psi})$, $n^{3/2}\kappa_5(\hat{\psi})$	$\frac{C_2n^{1/2}\gamma_1(\hat{\psi})}{\{-\zeta_2(\hat{\psi})\}^{1/2}}$, $\frac{C_1n\gamma_2(\hat{\psi})}{-\zeta_2(\hat{\psi})}$, $\frac{n^{3/2}\gamma_3(\hat{\psi})}{\{-\zeta_2(\hat{\psi})\}^{3/2}}$	A_3,A_1A_2, A_1^3
4	$n^2\kappa_3^4(\hat{\psi}), n^2\kappa_3^2(\hat{\psi})\kappa_4(\hat{\psi})$, $n^2\kappa_4^2(\hat{\psi}), n^2\kappa_5(\hat{\psi})\kappa_3(\hat{\psi})$, $n^2\kappa_6(\hat{\psi})$	$\frac{C_3n^{1/2}\gamma_1(\hat{\psi})}{\{-\zeta_2(\hat{\psi})\}^{1/2}}$, $\frac{C_2n\gamma_2(\hat{\psi})}{\{-\zeta_2(\hat{\psi})\}}$, $\frac{C_1n^{3/2}\gamma_3(\hat{\psi})}{\{-\zeta_2(\hat{\psi})\}^{3/2}}$, $\frac{n^2\gamma_4(\hat{\psi})}{\{-\zeta_2(\hat{\psi})\}^2}$	A_4, A_3A_1, A_2^2, A_1^2, A_2, A_1^4

Figure 1: The first four terms of A_k, B_k and C_k. The order of the κ_j terms are given in Lemma 2 in §6.1.

expansion, as they are not as simple as those in §3.

Theorem 3. *Under Assumptions 1–3, for the linear exponential and location-scale families,*

$$r = q + \frac{A}{n^{1/2}}q^2 + \frac{B}{n}q^3 + O_p\left(n^{-3/2}\right),$$

where A and B are $O_p(1)$.

We defer the proof to §6.1.

4. $r^\star$ in high dimensions

We discuss the behaviour of $r^\star$ in the setting where p increases with n. This asymptotic paradigm is increasingly relevant as datasets observed increase in size not only in the number of samples but also in the number of covariates. It is known that first-order inferential procedures, such as the likelihood ratio test, have poor finite sample performance in this setting (Sur and Candès, 2019). We quantify the order of the adjustment factors r_{inf} and r_{np} in this asymptotic regime, and show that the order is potentially much larger than in the p-fixed asymptotic regime. In fact, in the p-fixed asymptotic regime, the order of both adjustments are $O_p(n^{-1/2})$, this is no longer the case if p is allowed to increase. The results agree with those in Tang and Reid (2020), which uses a different approach that relies on a direct expansion of the adjustment factor.

For this regime we assume a weaker set of assumptions, Assumptions 4 – 8, these new Assumptions are given and discussed in §6.4 and are easier to satisfy in the high-dimensional setting. In addition we assume that $j_{\psi\lambda}(\theta) = O_p(n^{1/2})$ for all $\theta \in \{\theta : \|\theta - \theta_0\|_2 < \delta\}$ for some $\delta > 0$. This follows if the parametrization (ψ, λ) is orthogonal in the Cox and Reid (1987) sense, and an even stronger version of this is possible in the linear exponential family. We restrict the size of the eigenvalues of the third likelihood derivative matrices and we require that $p = o(n^{1/2}/\log(n))$. A discussion of the technical aspects of these results is provided in §6.4 and §6.5.

4.1 Linear Exponential Family

In the linear exponential family, there exists an orthogonal parametrization (ψ, λ) for which $\hat{\lambda}_\psi = \hat{\lambda}$, so that $d^k\hat{\lambda}_\psi/d\psi^k = 0$ for all integer-valued k. For this sub-section assume that the maximum singular value of $j_{\psi\lambda\lambda}(\hat{\theta})$ is $O_p(n)$. We consider the leading terms which appear in r_{np} and r_{inf} in (3.15) and (3.13), as the subsequent terms are of smaller order as shown in Theorems 4 and 5 in §6.4. We find that $\kappa_3(\hat{\psi})$ and $\kappa_4(\hat{\psi})$, defined in (2.3), satisfy

$$\kappa_3(\hat{\psi}) = O_p(n^{-1/2}), \quad \kappa_4(\hat{\psi}) = O_p(n^{-1}),$$

which implies that

$$r_{inf} = O_p(n^{-1/2});$$

the order is independent of the dimension of the parameter, as long as $p = o(n^{1/2}/\log(n))$.

For r_{np}, we have

$$\left|\gamma_1(\hat{\psi})\right| = \left|\mathrm{tr}\left[j_{\lambda\lambda}^{-1}(\hat{\theta})j_{\psi\lambda\lambda}(\hat{\theta})\right]\right| \leq \sigma_{\max}\{j_{\psi\lambda\lambda}(\hat{\theta})\}\,\mathrm{tr}[j_{\lambda\lambda}^{-1}(\hat{\theta})] = O_p(p),$$

using von Neumann's trace inequality and

$$\begin{aligned}
\left|\gamma_2(\hat{\psi})\right| &= \left|\mathrm{tr}\left[j_{\lambda\lambda}^{-1}(\hat{\theta}) j_{\psi\lambda\lambda}(\hat{\theta}) j_{\lambda\lambda}^{-1}(\hat{\theta}) j_{\psi\lambda\lambda}(\hat{\theta})\right] + \mathrm{tr}\left[j_{\lambda\lambda}^{-1}(\hat{\theta}) j_{\psi\psi\lambda\lambda}(\hat{\theta})\right]\right| \\
&\leq \sigma_{\max}^2\{j_{\psi\lambda\lambda}(\hat{\theta})\}\,\mathrm{tr}[\{j_{\lambda\lambda}^{-1}(\hat{\theta})\}^2] + \left\|j_{\lambda\lambda}^{-1}(\hat{\theta})\right\|_F \left\|j_{\psi\psi\lambda\lambda}(\hat{\theta})\right\|_F \\
&\leq \sigma_{\max}^2\{j_{\psi\lambda\lambda}(\hat{\theta})\} p \sigma_{\max}[\{j_{\lambda\lambda}^{-1}(\hat{\theta})\}^2] + p^{1/2}\sigma_{\max}\{j_{\lambda\lambda}^{-1}(\hat{\theta})\} \left\|j_{\psi\psi\lambda\lambda}(\hat{\theta})\right\|_F \\
&= O_p(p) + O_p(p^{3/2}).
\end{aligned}$$

The largest singular value $\sigma_{\max}$ is bounded by the Frobenius norm. This with Theorem 2 implies that

$$r_{np} = O_p\left(pn^{-1/2}\right).$$

These results imply that r coincides with $r^\star$ asymptotically in distribution if $p = o(n^{1/2}/\log(n))$, and agree with results in Tang and Reid (2020).

4.2 Location-Scale Family

We again consider the leading terms which appear in r_{inf} and r_{np} in (3.14) and (3.16), as the subsequent terms are of smaller order as shown in Theorems 4 and 5 in §6.4. Under the orthogonal parametrization (Tang and Reid, 2020, Lemma 1):

$$\left\|\frac{d\hat{\lambda}_\psi}{d\psi}|_{\psi=\hat{\psi}}\right\|_2 = O_p(p^{1/2}n^{-1/2}),$$

showing that $\kappa_3(\hat{\psi}) = O_p(p/n^{1/2})$, and $\kappa_4(\hat{\psi}) = O_p(p^2/n)$, which further implies that

$$r_{inf} = O_p\{(\max(p/n^{1/2}, p^2/n)\},$$

showing a dependence in p not present in the linear exponential family. Next we examine the size of r_{np}. As the derivatives of the constrained maximum likelihood estimate are not 0, we make the assumption that $\max_{i=1,\cdots,p} \sigma_{\max}\{j_{\theta_i\lambda\lambda}(\hat{\theta})\} = O_p(n)$. Then

$$\begin{aligned}
|\gamma_1(\hat{\psi})| &= \left|\mathrm{tr}\left[j_{\lambda\lambda}^{-1}(\hat{\theta})\frac{d}{d\psi}j_{\lambda\lambda}(\hat{\theta}_\psi)|_{\psi=\hat{\psi}}\right]\right| \leq \sigma_{\max}\left\{\frac{d}{d\psi}j_{\lambda\lambda}(\hat{\theta}_\psi)|_{\psi=\hat{\psi}}\right\}\mathrm{tr}[j_{\lambda\lambda}^{-1}(\hat{\theta})] \\
&= \sigma_{\max}\left\{j_{\psi\lambda\lambda}(\hat{\theta}) + \sum_{i=1}^{p-1}\frac{\partial\hat{\lambda}_{\psi,i}}{\partial\psi}|_{\psi=\hat{\psi}}\ j_{\lambda_i\lambda\lambda}(\hat{\theta})\right\}\mathrm{tr}[j_{\lambda\lambda}^{-1}(\hat{\theta})] \\
&\leq \left[\sigma_{\max}\left\{j_{\psi\lambda\lambda}(\hat{\theta})\right\} + \left\|\frac{\partial\hat{\lambda}_{\psi,i}}{\partial\psi}|_{\psi=\hat{\psi}}\right\|_1 \sigma_{\max}\left\{j_{\lambda_i\lambda\lambda}(\hat{\theta})\right\}\right]\mathrm{tr}[j_{\lambda\lambda}^{-1}(\hat{\theta})] \\
&\leq \left[\sigma_{\max}\left\{j_{\psi\lambda\lambda}(\hat{\theta})\right\} + p^{1/2}\left\|\frac{\partial\hat{\lambda}_{\psi,i}}{\partial\psi}|_{\psi=\hat{\psi}}\right\|_2 \sigma_{\max}\left\{j_{\lambda_i\lambda\lambda}(\hat{\theta})\right\}\right]\mathrm{tr}[j_{\lambda\lambda}^{-1}(\hat{\theta})] \\
&= \{O_p(n) + O_p(pn^{1/2})\}O_p(p/n) = O_p\left\{\max(p, p^2/n^{1/2})\right\}.
\end{aligned}$$

Also,

$$\begin{aligned}
\left|\gamma_2(\hat{\psi})\right| &= \left|\mathrm{tr}\left[j_{\lambda\lambda}^{-1}(\hat{\theta})\frac{d}{d\psi}j_{\lambda\lambda}(\hat{\theta}_\psi)|_{\psi=\hat{\psi}}\ j_{\lambda\lambda}^{-1}(\hat{\theta})\frac{d}{d\psi}j_{\lambda\lambda}(\hat{\theta}_\psi)|_{\psi=\hat{\psi}}\right] + \mathrm{tr}\left[j_{\lambda\lambda}^{-1}(\hat{\theta})\frac{d^2}{d\psi^2}j_{\lambda\lambda}(\hat{\theta}_\psi)|_{\psi=\hat{\psi}}\right]\right| \\
&\leq \sigma_{\max}^2\left\{\frac{d}{d\psi}j_{\lambda\lambda}(\hat{\theta}_\psi)|_{\psi=\hat{\psi}}\right\}\mathrm{tr}[\{j_{\lambda\lambda}^{-1}(\hat{\theta})\}^2] + \mathrm{tr}[j_{\lambda\lambda}^{-1}(\hat{\theta})]\sigma_{\max}\left\{\frac{d^2}{d\psi^2}j_{\lambda\lambda}(\hat{\theta}_\psi)|_{\psi=\hat{\psi}}\right\} \\
&= O_p\{\max(p^2, p^4/n)\} + O_p\{\max(pn, p^2n^{1/2}, p^3)\}.
\end{aligned}$$

Detailed calculation of the maximum singular value of $d^2 j_{\lambda\lambda}(\hat{\theta}_\psi)/d\psi^2|_{\psi=\hat{\psi}}$ is deferred to §6.3. If $p = o(n^{1/2}/\log(n))$, then

$$r_{np} = O_p(p/n^{1/2}),$$

which agrees with the result in Tang and Reid (2020, §5.2).

5. Simulations

We present simulations to illustrate the results in §3; for the linear exponential and location scale family

$$r^\star = A^\star/n^{1/2} + (1 + B^\star/n)r + O_p(n^{-3/2}), \tag{5.20}$$

by Remark 2. To provide numerical evidence that (5.20) holds, note that,

$$r^\star - A^\star/n^{1/2} - (1 + B^\star/n)r = O_p(n^{-3/2}), \tag{5.21}$$

and a sufficient condition for a random variable to be $O_p(n^{-3/2})$ is for both its mean and standard deviation to be $O(n^{-3/2})$. We illustrate the relationship in (5.21) graphically by plotting the value of simulation the mean and standard deviation of

$$\tilde{r}^\star = r^\star - A^\star - (1 + B^\star)r \tag{5.22}$$

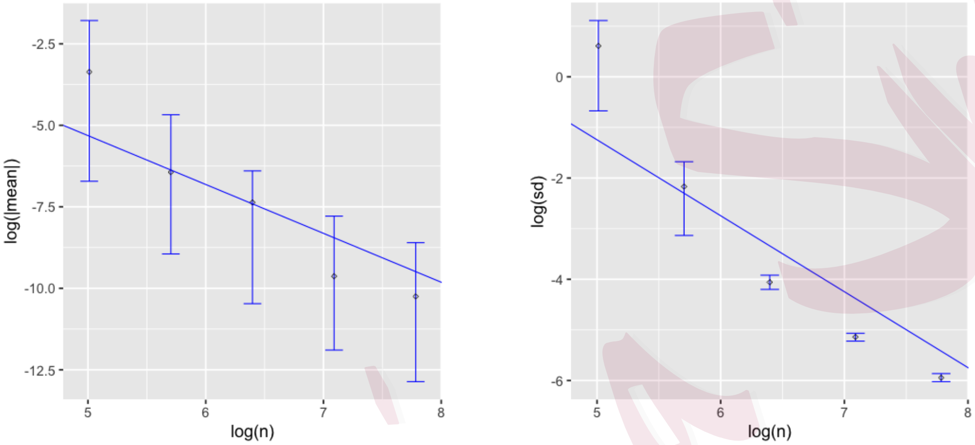
as a function of n; we expect this to roughly follow a linear trend when plotted against $\log(n)$, with slope at most $-3/2$.

5.1 Logistic Regression

We first consider an example based on logistic regression in which there are 25 covariates associated with each y_i, taken to be independent and identically distributed standard normals. The true regression coefficients are $\beta = (1, 0, 1, 1, 1, 0, \ldots, 0)$, and the intercept is $\beta_0 = 1$, and we test $H_0 : \beta_1 = 0$. We obtain estimates of $A^\star$ and $B^\star$ in (5.21) by numerical differentiation of the profile log-likelihood and the log-determinant of the information matrix. For each $n = 150, 300, 600, 1200, 2400$, we simulate 2000 values of $r^\star - A^\star - (1 + B^\star)r$ and plot the 95% bootstrap confidence intervals of the empirical mean and standard deviation from 1000 bootstrap simulations. In Figure 2, these are compared to a line with slope $-3/2$.

5.2 Linear regression with student t Errors

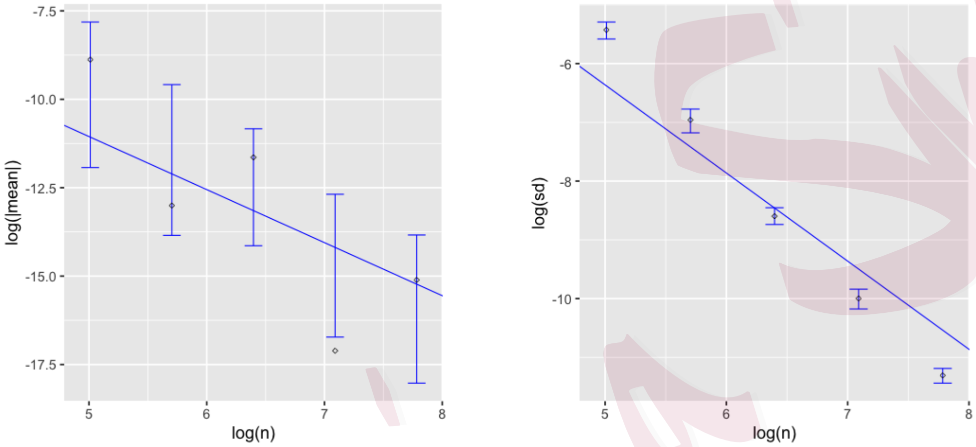
We consider an example based on a location-scale regression model where the error follows a student t distribution with 5 degrees of freedom. The values of the regression coefficients and the distribution of the covariates are the same as §5.1. The results of the simulations are given in Figure 3.



(a) Empirical mean of $\tilde{r}^{\star}$.

(b) Empirical standard deviation of $\tilde{r}^{\star}$

Figure 2: Plots for logistic regression illustrating order of $\tilde{r}^{\star}$ (5.22). The mean, and the standard deviation functions of n, are plotted against $\log(n)$ along with bootstrap 95% confidence intervals. The solid line has slope $-3/2$ and the intercept is fitted using the least squares estimate.



(a) Empirical mean of $\tilde{r}^\star$.

(b) Empirical standard deviation of $\tilde{r}^\star$

Figure 3: Plots for t_5 based regression illustrating the order of $\tilde{r}^\star$ (5.22). The mean, and the standard deviation functions of n, are plotted against $\log(n)$ along with bootstrap 95% confidence intervals. The solid line has slope $-3/2$ and the intercept is fitted using the least squares estimate.

6. Additional Proof Details

6.1 Proof of Theorem 3

Proof. Note that $\rho^{-1} = O_p(1)$ by Assumption 3. We prove the result in the case for the linear exponential family; the proof for the location-scale family is similar. We use capital letters to denote terms of order $O_p(1)$. From (3.10),

$$\begin{aligned} t &= r\Big\{1 - \frac{\kappa_3(\hat\psi)}{6}t + \frac{\kappa_4(\hat\psi)}{24}t^2 + \frac{\kappa_3^2(\hat\psi)}{24}t^2 + O_p(n^{-3/2})\Big\} \\ &= r\left[1 - \frac{\kappa_3(\hat\psi)}{6\rho}q + \left\{\frac{1}{24\rho^2}\kappa_4(\hat\psi) + \frac{1}{24\rho^2}\kappa_3^2(\hat\psi)\right\}q^2 + O_p(n^{-3/2})\right] \\ &= r\left\{1 + \frac{C}{n^{1/2}}q + \frac{D}{n}q^2 + O_p(n^{-3/2})\right\}. \qquad (6.23) \end{aligned}$$

We expand $|j_{\lambda\lambda}(\hat\theta_{\psi_0})|$,

$$\begin{aligned} |j_{\lambda\lambda}(\hat\theta_{\psi_0})| &= |j_{\lambda\lambda}(\hat\theta)| + (\psi_0 - \hat\psi)\frac{d|j_{\lambda\lambda}(\hat\theta_{\psi_0})|}{d\psi}|_{\psi=\hat\psi} + \frac{1}{2}(\psi_0 - \hat\psi)^2\frac{d|j_{\lambda\lambda}(\hat\theta_{\psi_0})|}{d\psi^2}|_{\psi=\hat\psi} + \cdots \\ &= |j_{\lambda\lambda}(\hat\theta)|\left\{1 + (\hat\psi - \psi_0)\gamma_1(\hat\psi) + (\hat\psi - \psi_0)^2\gamma_2(\hat\psi) + O_p(n^{-3/2})\right\} \\ &= |j_{\lambda\lambda}(\hat\theta)|\left[1 + \frac{\gamma_1(\hat\psi)}{\{\rho j_{\mathrm{p}}^{1/2}(\hat\psi)\}}q + \frac{\gamma_2(\hat\psi)}{\rho^2 j_{\mathrm{p}}(\hat\psi)}q^2 + O_p(n^{-3/2})\right] \\ &= |j_{\lambda\lambda}(\hat\theta)|\left\{1 + \frac{E}{n^{1/2}}q + \frac{F}{n}q^2 + O_p(n^{-3/2})\right\}. \end{aligned}$$

Therefore,

$$\rho = \left\{\frac{|j_{\lambda\lambda}(\hat\theta)|}{|j_{\lambda\lambda}(\hat\theta_{\psi_0})|}\right\}^{1/2} = \left[\frac{1}{1 + Cq/n^{1/2} + Dq^2/n + O_p(n^{-3/2})}\right]^{1/2}$$

$$= 1 + \frac{G}{n^{1/2}} q + \frac{H}{n} q^2 + O_p(n^{-3/2}). \tag{6.24}$$

Note that $\gamma_1(\psi) = O_p(1)$ and $\gamma_2(\psi) = O_p(1)$ by Lemma 2 in §6.2. Combining (6.23) and (6.24),

$$q = t\rho = r\left\{1 + \frac{C}{n^{1/2}} q + \frac{D}{n} q^2 + O_p(n^{-3/2})\right\}\left\{1 + \frac{G}{n^{1/2}} q + \frac{H}{n} q^2 + O_p(n^{-3/2})\right\}.$$

$$\begin{aligned} r &= q\left\{1 + \frac{qC}{n^{1/2}} + \frac{q^2 D}{n} + O_p(n^{-3/2})\right\}^{-1}\left\{1 + \frac{G}{n^{1/2}} q + \frac{H}{n} q^2 + O_p(n^{-3/2})\right\}^{-1} \\ &= q\left\{1 + \frac{A}{n^{1/2}} q + \frac{B}{n} q^2 + O_p(n^{-3/2})\right\} = q + \frac{A}{n^{1/2}} q^2 + \frac{B}{n} q^3 + O_p(n^{-3/2}), \end{aligned}$$

which shows the desired result. For the location scale family, we use the same arguments and apply them to s instead of t. □

6.2 Order of γ_k and κ_k

We first establish the order of $\gamma_k(\hat{\psi})$ and $\zeta_k(\hat{\psi})$ for arbitrary integer k. All results are proved for ψ and λ scalar parameters; generalization to vector λ is straightforward but notationally tedious.

Lemma 2. *For all integers k,*

$$\frac{\partial^k \hat{\lambda}_\psi}{\partial \psi^k}|_{\psi=\hat{\psi}} = O_p(1), \quad \frac{d^k}{d\psi^k}\tilde{\jmath}_{\lambda\psi}|_{\psi=\hat{\psi}} = O_p(n), \quad \frac{d^k}{d\psi^k}\tilde{\jmath}_{\lambda\lambda}|_{\psi=\hat{\psi}} = O_p(n),$$

$$\kappa_k(\hat{\psi}) = O_p\{n^{-(k-2)/2}\}, \quad \gamma_k(\hat{\psi}) = O_p(1).$$

Proof. Differentiating the expression

$$0 = \tilde{l}_{\lambda},$$

we obtain,

$$\tilde{j}_{\lambda\lambda} \frac{\partial \hat{\lambda}_{\psi}}{\partial \psi} = -\tilde{j}_{\psi\lambda}, \tag{6.25}$$

thus,

$$\frac{\partial \hat{\lambda}_{\psi}}{\partial \psi}|_{\psi=\hat{\psi}} = -\hat{j}_{\lambda\lambda}^{-1}\hat{j}_{\psi\lambda} = O_p(1), \tag{6.26}$$

by Assumptions 1 and 3, we use the above to show

$$\frac{d}{d\psi}\tilde{j}_{\lambda\lambda}|_{\psi=\hat{\psi}} = \hat{j}_{\psi\psi} + \frac{\partial \hat{\lambda}}{\partial \psi}|_{\psi=\hat{\psi}}\ \hat{j}_{\psi\lambda} = O_p(n),$$

and by the same argument show that $d\tilde{j}_{\lambda\psi}/d\psi|_{\psi=\hat{\psi}} = O_p(n)$. This proves the first three claims hold for $k = 1$.

Induction Step: We assume that for all $k_1 \leq k$ the result holds, we need to show that for $k+1$

$$\frac{\partial^{k+1}\hat{\lambda}_{\psi}}{\partial \psi^{k+1}}|_{\psi=\hat{\psi}} = O_p(1), \quad \frac{d^{k+1}}{d\psi^{k+1}}\tilde{j}_{\lambda\lambda}|_{\psi=\hat{\psi}} = O_p(n), \quad \frac{d^{k+1}}{d\psi^{k+1}}\tilde{j}_{\lambda\psi}|_{\psi=\hat{\psi}} = O_p(n).$$

First, we differentiate (6.26) k times to obtain

$$\frac{\partial^{k+1}\hat{\lambda}_{\psi}}{\partial \psi^{k+1}} = -\sum_{i=1}^{k}\binom{k}{i}\frac{d^{i}(\tilde{j}_{\lambda\lambda})^{-1}}{d\psi^{i}}\frac{d^{k-i}\tilde{j}_{\psi\lambda}}{d\psi^{k-i}}.$$

Using Faa di Bruno's formula for the differentiation of a composition of functions we obtain

$$\frac{d^i}{d\psi^i}(\tilde{\jmath}_{\lambda\lambda})^{-1} = \sum_{k=1}^{i}(-1)^k k!\{\tilde{\jmath}_{\lambda\lambda}\}^{(-k-1)} B_{i,k}\left(\frac{d}{d\psi}\tilde{\jmath}_{\lambda\lambda}, \cdots, \frac{d}{d\psi^{i-k+1}}\tilde{\jmath}_{\lambda\lambda}\right), \tag{6.27}$$

where

$$B_{i,k}(x_1, \cdots, x_{i-k+1}) = \sum \frac{i!}{j_1!j_2!\cdots j_{i-k+1}!}\left(\frac{x_1}{1!}\right)^{j_1}\left(\frac{x_2}{2!}\right)^{j_2}\cdots\left(\frac{x_{i-k+1}}{(i-k+1)!}\right)^{j_{i-k+1}},$$

and the summation in the above expression is taken over all sets of $j_1, \ldots, j_{i-k+1}$ such that,

$$j_1 + j_2 + \cdot + j_{i-k+1} = k, \quad j_1 + 2j_2 + \cdots + (i-k+1)j_{i-k+1} = i.$$

The polynomials $B_{i,k}$ are the partial Bell polynomials. From (6.27) we deduce that $d(\tilde{\jmath}_{\lambda\lambda})^{-1}/d\psi|_{\psi=\hat{\psi}} = O_p(n^{-1})$, since the constraint $j_1 + j_2 + \cdot + j_{n-l+1} = k$ implies that

$$B_{i,k}(d\tilde{\jmath}_{\lambda\lambda}/d\psi|_{\psi=\hat{\psi}}, \cdots, d^{i-k+1}\tilde{\jmath}_{\lambda\lambda}/d\psi^{i-k+1}|_{\psi=\hat{\psi}}) = O_p(n^k),$$

and $(\hat{\jmath}_{\lambda\lambda})^{-k-1} = O_p(n^{-k-1})$, which implies that every term in the summation is $O_p(n^{-1})$.

Thus, by the induction assumption, we have $d^{k-i}\tilde{\jmath}_{\psi\lambda}/d\psi^{k-i}|_{\psi=\hat{\psi}} = O_p(n)$ for $i = 1, \cdots, k-1$. Therefore we have the desired result for the constrained derivative of the maximum likelihood estimate. Next we show that

$$\frac{d^{k+1}}{d\psi^{k+1}}\tilde{\jmath}_{\lambda\lambda}|_{\psi=\hat{\psi}} = O_p(n), \quad \frac{d^{k+1}}{d\psi^{k+1}}\tilde{\jmath}_{\lambda\psi}|_{\psi=\hat{\psi}} = O_p(n).$$

For this,

$$\frac{d^{k+1}}{d\psi^{k+1}}\tilde{\jmath}_{\lambda\psi} = \sum_{i,j,l,m=1}^{k+1} a_{i,j,l,m}\frac{\partial^{i+j}\tilde{\jmath}_{\psi\lambda}(\psi,\hat{\lambda}_\psi)}{\partial\psi^i\partial\lambda^j}\left(\frac{\partial^l\hat{\lambda}_\psi}{\partial\psi^l}\right)^m = O_p(n), \qquad (6.28)$$

which can be obtained through successive applications of the chain rule, some of the coefficients $a_{i,j,l,m}$ may be 0. The result follows from the fact that all derivatives of the constrained maximum likelihood estimate are $O_p(1)$ up to the $(k+1)$ order when evaluated at $\hat{\psi}$ and log-likelihood derivatives are assumed to be $O_p(n)$ when evaluated at $\hat{\theta}$. A similar argument can be made for the derivatives of $\tilde{\jmath}_{\lambda\lambda}$.

Order of $\kappa_k(\psi)$

The total derivative of the profile log-likelihood function is a summation of partial derivatives multiplied by the derivative of the constrained maximum likelihood estimate, so the result is obtained from arguments used in (6.28).

Order of $\gamma_k(\psi)$

We have

$$\gamma_k(\psi) = \sum_{i+j=k,\ j\geq 1} \mathrm{tr}\left[\frac{d^i}{d\psi^i}(\tilde{\jmath}_{\lambda\lambda})^{-1}\frac{d^j}{d\psi^j}\tilde{\jmath}_{\lambda\lambda}\right]. \qquad (6.29)$$

Using $d^i(\tilde{\jmath}_{\lambda\lambda})^{-1}/d\psi^i|_{\psi=\hat{\psi}} = O_p(n^{-i})$ from (6.27) and $d^j\tilde{\jmath}_{\lambda\lambda}/d\psi_i|_{\psi=\hat{\psi}} = O_p(n)$ from (6.28) we conclude $\gamma_k(\hat{\psi}) = O_p(1)$. □

6.3 Order of Maximum Singular Value in §4

We obtain the order of the maximum singular value of the second derivative of the information matrix for the location-scale model in the high-dimensional setting. We have

$$\begin{aligned} \frac{d^2}{d\psi^2}\tilde{\jmath}_{\lambda\lambda}|_{\psi=\hat{\psi}} = \hat{\jmath}_{\psi\psi\lambda\lambda} &+ 2\sum_{i=1}^{p-1}\frac{\partial\hat{\lambda}_{\psi,i}}{\partial\psi}|_{\psi=\hat{\psi}}\ \hat{\jmath}_{\psi\lambda_i\lambda\lambda} + \sum_{i=1}^{p-1}\frac{\partial^2\hat{\lambda}_{\psi,i}}{\partial\psi^2}|_{\psi=\hat{\psi}}\ \hat{\jmath}_{\lambda_i\lambda\lambda} \\ &+ \sum_{i=1}^{p-1}\sum_{j=1}^{p-1}\frac{\partial\hat{\lambda}_{\psi,i}}{\partial\psi}|_{\psi=\hat{\psi}}\frac{\partial\hat{\lambda}_{\psi,j}}{\partial\psi}|_{\psi=\hat{\psi}}\ \hat{\jmath}_{\lambda_i\lambda_j\lambda\lambda}. \end{aligned}$$

Now the maximal singular values of the matrices of interest are:

$$\begin{aligned} &\sigma_{\max}\{\hat{\jmath}_{\psi\psi\lambda\lambda}\} \le \|\hat{\jmath}_{\psi\psi\lambda\lambda}\|_F = O_p(pn). \\ &\sigma_{\max}\left\{\sum_{i=1}^{p-1}\frac{\partial\hat{\lambda}_{\psi,i}}{\partial\psi}|_{\psi=\hat{\psi}}\ \hat{\jmath}_{\psi\lambda_i\lambda\lambda}\right\} \\ &\le \sum_{i=1}^{p-1}\left|\frac{\partial\hat{\lambda}_{\psi,i}}{\partial\psi}|_{\psi=\hat{\psi}}\right|\sigma_{\max}\left\{\hat{\jmath}_{\psi\lambda_i\lambda\lambda}\right\} \\ &\le \sum_{i=1}^{p-1}\left|\frac{\partial\hat{\lambda}_{\psi,i}}{\partial\psi}|_{\psi=\hat{\psi}}\right|\|\hat{\jmath}_{\psi\lambda_i\lambda\lambda}\|_F \\ &\le p^{1/2}\left\|\frac{\partial\hat{\lambda}_{\psi}}{\partial\psi}|_{\psi=\hat{\psi}}\right\|_2\max_{i=1,\dots,p}\|\hat{\jmath}_{\psi\lambda_i\lambda\lambda}\|_F = O_p(p^2n^{1/2}). \\ &\sigma_{\max}\left\{\sum_{i=1}^{p-1}\frac{\partial^2\hat{\lambda}_{\psi,i}}{\partial\psi^2}|_{\psi=\hat{\psi}}\ \hat{\jmath}_{\lambda_1\lambda\lambda}\right\} \end{aligned}$$

$$\leq \sum_{i=1}^{p-1} \left| \frac{\partial^2 \hat{\lambda}_{\psi,i}}{\partial \psi^2}|_{\psi=\hat{\psi}} \right| \ \sigma_{\max}\left\{\hat{\jmath}_{\lambda_i\lambda\lambda}\right\} = O_p(pn)$$

$$\sigma_{\max}\left\{ \sum_{i=1}^{p-1}\sum_{j=1}^{p-1} \frac{\partial \hat{\lambda}_{\psi,i}}{\partial \psi}|_{\psi=\hat{\psi}} \frac{\partial \hat{\lambda}_{\psi,j}}{\partial \psi}|_{\psi=\hat{\psi}} \ \hat{\jmath}_{\lambda_i\lambda_j\lambda\lambda} \right\}$$

$$\leq \sum_{i=1}^{p-1}\sum_{j=1}^{p-1} \left| \frac{\partial \hat{\lambda}_{\psi,i}}{\partial \psi}|_{\psi=\hat{\psi}} \right| \left| \frac{\partial \hat{\lambda}_{\psi,j}}{\partial \psi}|_{\psi=\hat{\psi}} \right| \ \sigma_{\max}\left\{\hat{\jmath}_{\lambda_i\lambda_j\lambda\lambda}\right\}$$

$$\leq \sum_{i=1}^{p-1}\sum_{j=1}^{p-1} \left| \frac{\partial \hat{\lambda}_{\psi,i}}{\partial \psi}|_{\psi=\hat{\psi}} \right| \left| \frac{\partial \hat{\lambda}_{\psi,j}}{\partial \psi}|_{\psi=\hat{\psi}} \right| \left\| \hat{\jmath}_{\lambda_i\lambda_j\lambda\lambda} \right\|_F$$

$$= p \left\| \frac{\partial \hat{\lambda}_{\psi}}{\partial \psi}|_{\psi=\hat{\psi}} \right\|_2^2 \max_{i,j=1,\ldots,p} \left\| \hat{\jmath}_{\lambda_i\lambda_j\lambda\lambda} \right\|_F = O_p(p^3).$$

Using the triangle inequality results in the rates obtained for $\gamma_1(\hat{\psi})$ and $\gamma_2(\hat{\psi})$ given in §5.2.

6.4 Assumptions in high dimensions

To derive the relationships between r, s and t in high dimensions, we require a different set of Assumptions:

Assumption 4. $|\hat{\psi} - \psi_0| = O_p(n^{-\delta})$ for some $\delta > 0$ as p and n tend to ∞.

Assumption 5. $\kappa_3(\hat{\psi}), \kappa_4(\hat{\psi})$ and $\kappa_5(\tilde{\psi})$ are $O_p(n^{-\epsilon})$, $O_p(n^{-2\epsilon})$ and $O_p(n^{-3\epsilon})$ respectively, for $\tilde{\psi}$ in a $O(n^{-\delta'})$ neighborhood of ψ_0 for some $\delta' < \delta$ and some $\epsilon > 0$.

Assumption 6. $\gamma_1(\hat{\psi}), \gamma_2(\hat{\psi})$ and $\gamma_3(\tilde{\psi})$ are $O_p(1)$ for $\tilde{\psi}$ in a $O(n^{-\delta'})$ neighborhood of ψ_0 for some $\delta' < \delta$.

Assumption 7. Either r, s or t is $O_p(1)$ as p and n increase to ∞.

Assumption 8. The log-likelihood is six times differentiable in a L^2 ball of radius δ around θ_0.

Lemma 3. *Under Assumptions 4–8, for r, t and s defined in (1.1), (2.5) and (2.8):*

$$t = r\left\{1 + A_1 r + B_1 r^2 + O_p(n^{-3\epsilon})\right\},$$

$$s = t\left\{1 + A_2 t + B_2 t^2 + O_p(n^{-3\epsilon})\right\},$$

where,

$$A_1 = -\frac{1}{6}\kappa_3(\hat{\psi}), \quad B_1 = \frac{1}{24}\kappa_4(\hat{\psi}) + \frac{5}{72}\kappa_3^2(\hat{\psi}),$$

$$A_2 = \frac{\kappa_3(\hat{\psi})}{2}, \qquad B_2 = -\frac{\kappa_4(\hat{\psi})}{6}.$$

Proof. The proof structure is similar to the proof of Lemma 1, therefore some details are omitted. We derive the relationship between r and t:

$$\begin{aligned}
r^2 &= 2\Big\{l_{\mathrm{p}}(\hat{\psi}) - l_{\mathrm{p}}(\psi_0)\Big\}, \\
&= 2\Big\{(\hat{\psi}-\psi_0)\zeta_1(\hat{\psi}) - \frac{(\hat{\psi}-\psi_0)^2}{2}\zeta_2(\hat{\psi}) \\
&\quad + \frac{(\hat{\psi}-\psi_0)^3}{6}\zeta_3(\hat{\psi}) + \frac{(\hat{\psi}-\psi_0)^4}{24}\zeta_4(\hat{\psi}) + \frac{(\hat{\psi}-\psi_0)^5}{120}\zeta_5(\tilde{\psi})\Big\} \\
&= t^2\Big\{1 + \frac{\kappa_3(\hat{\psi})}{3}t - \frac{\kappa_4(\hat{\psi})}{12}t^2 + O_p(n^{-3\epsilon})\Big\},
\end{aligned}$$

for some $\tilde{\psi}$ between ψ_0 and $\hat{\psi}$, by Assumptions 4, and 5 and Taylor's theorem (justified by Assumption 8). The other steps are the same as in the proof of Lemma 1. Keeping track of the remainder term, we obtain:

$$t = r\Big\{1 - \frac{\kappa_3(\hat{\psi})}{6}r + \frac{5}{72}\kappa_3^2(\hat{\psi})r^2 + \frac{1}{24}\kappa_4(\hat{\psi})r^2 + O_p(n^{-3\epsilon})\Big\}.$$

The Taylor series expansion of $(1+x)^{-1}$ is valid as the argument is $o_p(1)$, thus will eventually be smaller than 1 with probability 1. For the expansion of s:

$$\begin{aligned} s &= \frac{\zeta_1(\psi_0)}{j_p^{1/2}(\hat{\psi})} \\ &= \frac{1}{j_p^{1/2}(\hat{\psi})}\Big\{\zeta_1(\hat{\psi}) - \zeta_2(\hat{\psi})(\hat{\psi}-\psi_0) + \frac{\zeta_3(\hat{\psi})}{2}(\hat{\psi}-\psi_0)^2 - \frac{\zeta_4(\hat{\psi})}{6}(\hat{\psi}-\psi_0)^3 + \frac{\zeta_5(\tilde{\psi})}{24}(\hat{\psi}-\psi_0)^4\Big\} \\ &= t + \frac{\kappa_3(\hat{\psi})}{2}t^2 - \frac{\kappa_4(\hat{\psi})}{6}t^3 + O_p(n^{-3\epsilon}), \end{aligned}$$

for some $\tilde{\psi}$ between ψ_0 and $\hat{\psi}$, the final line follows by the same arguments as above. □

There are also high-dimensional versions of Theorems 1 and 2 which can be established using the same assumptions as above.

Theorem 4. *Under Assumptions 4–8, for the linear exponential family,*

$$r_{inf} = -\frac{1}{6}\kappa_3(\hat{\psi}) + \Big\{\frac{1}{24}\kappa_4(\hat{\psi}) + \frac{4}{72}\kappa_3^2(\hat{\psi})\Big\}r + O_p(n^{-3\epsilon}), \qquad (6.30)$$

and for the location-scale family

$$r_{inf} = \frac{1}{3}\kappa_3(\hat{\psi}) - \Big\{\frac{3}{24}\kappa_4(\hat{\psi}) + \frac{11}{72}\kappa_3^2(\hat{\psi})\Big\}r + O_p(n^{-3\epsilon}). \qquad (6.31)$$

Theorem 5. *Under Assumptions 4–8, for the linear exponential family*

$$r_{np} = \frac{1}{2}\frac{\gamma_1(\hat{\psi})}{j_{\mathrm{p}}(\hat{\psi})^{1/2}} - \Big\{\frac{1}{12}\frac{\kappa_3(\hat{\psi})\gamma_1(\hat{\psi})}{j_{\mathrm{p}}(\hat{\psi})^{1/2}} - \frac{1}{4}\frac{\gamma_2(\hat{\psi})}{j_{\mathrm{p}}(\hat{\psi})}\Big\}r + O_p(n^{-3\epsilon}), \qquad (6.32)$$

and for the location-scale family

$$r_{np} = -\frac{1}{2}\frac{\gamma_1(\hat{\psi})}{j_{\mathrm{p}}(\hat{\psi})^{1/2}} + \Big\{\frac{1}{12}\frac{\kappa_3(\hat{\psi})\gamma_1(\hat{\psi})}{j_{\mathrm{p}}(\hat{\psi})^{1/2}} - \frac{1}{4}\frac{\gamma_2(\hat{\psi})}{j_{\mathrm{p}}(\hat{\psi})}\Big\}r + O_p(n^{-3\epsilon}). \qquad (6.33)$$

Proof. **Linear Exponential Family:** Using (2.6) and Lemma 1 for a linear exponential family we have

$$\begin{aligned} r_{inf} &= \frac{1}{r}\log\left(\frac{t}{r}\right) \\ &= A_1 + \left(B_1 - A_1^2\right) r + O_p(n^{-3\epsilon}) \\ &= -\frac{1}{6}\kappa_3(\hat{\psi}) + \Big\{\frac{1}{24}\kappa_4(\hat{\psi}) + \frac{4}{72}\kappa_3^2(\hat{\psi})\Big\}r + O_p(n^{-3\epsilon}). \qquad (6.34) \end{aligned}$$

A similar expansion can be developed for r_{np}:

$$\begin{aligned} r_{np} &= \frac{1}{2r}\log\left\{\frac{|j_{\lambda\lambda}(\hat{\psi},\hat{\lambda})|}{|j_{\lambda\lambda}(\psi_0,\hat{\lambda}_{\psi_0})|}\right\} \\ &= \frac{1}{2r}\left[\frac{\gamma_1(\hat{\psi})}{\{-\zeta_2(\hat{\psi})\}^{1/2}}t + \frac{\gamma_2(\hat{\psi})}{2\zeta_2(\hat{\psi})}t^2 + \frac{\gamma_3(\tilde{\psi})}{6(-\zeta_2(\hat{\psi}))^{3/2}}t^3\right] \qquad (6.35) \\ &= \frac{1}{2}\left[\left(1 + \frac{A_1}{n^{1/2}}r\right)\frac{\gamma_1(\hat{\psi})}{\{-\zeta_2(\hat{\psi})\}^{1/2}} + \frac{\gamma_2(\hat{\psi})}{2\zeta_2(\hat{\psi})}r + O_p\left(n^{-3\epsilon}\right)\right] \end{aligned}$$

$$= \frac{1}{2} \frac{\gamma_1(\hat{\psi})}{\{-\zeta_2(\hat{\psi})\}^{1/2}} + \left[\frac{1}{2} \frac{A_1 \gamma_1(\hat{\psi})}{\{-\zeta_2(\hat{\psi}) n\}^{1/2}} + \frac{\gamma_2(\hat{\psi})}{4\zeta_2(\hat{\psi})} \right] r + O_p\Big(n^{-3\epsilon}\Big), \quad (6.36)$$

where the third equality uses Lemma 3.

The proof for r_{inf} in the location-scale family is the same as in the proof of the exponential family but with differing coefficients. □

6.5 Checking the Assumptions for GLMs

We show Assumptions 4–8 are satisfied for a non-trivial model in high dimensions. We consider generalized linear models with $p = o(n^{1/2}/\log(n))$ and smooth likelihoods for example logistic regression, gamma regression or Poisson regression, and show that these models satisfy our Assumptions. In what follows we use β to denote the regression coefficients, β_0 the data generating regression coefficients, and X the design matrix.

Under conditions 1 and 2 of Fan et al. (2019), Assumption 4 holds with a rate of $O(n^{-1/2}\log(n))$ (by their Equation (7) and with $\gamma = 1/2$). Assumption 7 for t holds by Theorem 1 in Fan et al. (2019). Assumption 8 holds for the class of smooth likelihoods we are considering.

For Assumptions 5 – 6, these are more readily checked under an orthogonal parameterization under which $\hat{\lambda}_\psi = \hat{\lambda}$, as in Tang and Reid (2020), as the derivatives of the profile likelihood simplifies dramatically. This is justified as r_{inf} and r_{np} are invariant to re-parametrization. To check these assumptions we may take $\delta = 1/2 + 2\log\log(n)$. Before we show this, we require:

Lemma 4. *Under Conditions 1 and 2 of Fan et al. (2019), the likelihood derivatives evaluated in an $O(\log(n)/n^{1/2})$ L^2 neighborhood of the MLE and the constrained MLE are $O_p(n)$ uniformly if $|X\beta_0|$ is a uniformly bounded vector with probability 1 as both n and p tend to ∞.*

The requirement that $|X\beta_0|$ is bounded uniformly makes sure that the true generative model for each observation remains bounded in probability.

Proof. It is shown in Fan et al. (2019) that with probability tending to 1, $\max_{i=1,\ldots,n}|X_i(\hat\beta - \beta_0)| = O((p/n)^{1/2}) \to 0$, under the null and if $p = o(n)$, showing that $\max_{i=1,\ldots,n}|X_i\hat\beta|$ is bounded as well. Using the fact that the observations are assumed to be independent and the likelihoods are analytic functions, this then implies that all likelihood derivatives are of order $O_p(n)$ when evaluated at the MLE or the constrained MLE under the null hypothesis.

Similarly, in a $O(\log(n)/n^{1/2})$ neighborhood of the MLE or constrained MLE, $X_i(\hat\beta + e_n)$, for $\|e_n\|_2 = O(\log(n)/n^{1/2})$, also tend to the value of $X_i\beta_0$ uniformly, by Condition 1 in Fan et al. (2019) on the norm of X_i, we have that $\max_{i=1,\ldots,n}|X_ie_n|$ tends to 0 with probability 1. □

This, with Condition 2 in Fan et al. (2019) (which guarantees that $j_p^{-1}(\tilde{\psi}) = O(n^{-1})$ for $\tilde{\psi} - \psi_0 = O_p(n^{-\delta})$) along with our condition on $\sigma_{\max}\{j_{\psi\lambda\lambda}(\hat{\theta})\} = O(n)$, we see that γ_k and κ_k are decreasing in asymptotic order. As the derivatives of the likelihood are all of order $O_p(n)$, and under the orthogonal parameterization $\hat{\lambda}_\psi = \hat{\lambda}$, then by the same calculations as in §4.1 it can be seen that $\kappa_k = O(n^{k-2})$ for $i = 3, 4, 5$ and $\gamma_k = O_p((p/n^{1/2})^k)$ $k = 1, 2, 3$, showing that Assumptions 5 and 6 hold thus verifying all the Assumptions.

Acknowledgements

This research was partially completed while Y.T. was affiliated with the University of Toronto. We thank Nicola Sartori, Michele Lambardi di San Miniato, Ioannis Kosmidis, Heather Battey and Michaël Lalancette and anonymous reviewers for helpful comments. This research was partially supported by the Natural Sciences and Engineering Research Council of Canada and the Vector Institute.

References

Barndorff-Nielsen, O. E. and D. R. Cox (1994). *Inference and Asymptotics.* Chapman & Hall, New York.

Brazzale, A., A. Davison, and N. Reid (2007). *Applied Asymptotics: Case Studies in Small-Sample Statistics.* Cambridge University Press, Cambridge.

Cakmak, S., D. Fraser, P. McDunnough, N. Reid, and X. Yuan (1998). Likelihood centered asymptotic model exponential and location model versions. *J. Statist. Plann. and Inf.* **66**, 211–222.

Cox, D. R. and D. V. Hinkley (1974). *Theoretical Statistics.* Chapman & Hall, New York.

Cox, D. R. and N. Reid (1987). Parameter orthogonality and approximate conditional inference (with discussion). *J. R. Statist. Soc. B* **49**, 1–39.

Davison, A. C. and N. Reid (2021). The tangent exponential model. *arXiv:2106.10496*.

Fan, Y., E. Demirkaya, and J. Lv (2019). Nonuniformity of p-values can occur early in diverging dimensions. *J. Mach. Learn. Res.* **20**, 1–33.

Mirsky, L. (1975). A trace inequality of John von Neumann. *Monatshefte für Mathematik* **79**, 303–306.

Pierce, D. A. and D. Peters (1992). Practical use of higher order asymptotics for multiparameter exponential families (with discussion). *J. R. Statist. Soc. B* **54**, 701–737.

Reid, N. (2003). Asymptotics and the theory of inference. *Ann. Statist.* **31**, 1695–1731.

Sur, P. and E. J. Candès (2019). A modern maximum likelihood theory for high-dimensional logistic regression. *Proc. Nat. Acad. Sci.* **116**, 14516–14525.

Tang, Y. and N. Reid (2020). Modified likelihood root in high dimensions. *J. R. Statist. Soc.*

B **82**, 1349–1369.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

Imperial College London

E-mail: yanbo.tang@imperial.ac.uk

University of Toronto

E-mail: nancym.reid@utoronto.ca