The Annals of Applied Probability
2020, Vol. 30, No. 2, 608–643
https://doi.org/10.1214/19-AAP1510

BOUNDS FOR THE ASYMPTOTIC DISTRIBUTION OF THE LIKELIHOOD RATIO

BY ANDREAS ANASTASIOU[1] AND GESINE REINERT[2]

[1]*Department of Mathematics and Statistics, University of Cyprus, anastasiou.andreas@ucy.ac.cy*
[2]*Department of Statistics, University of Oxford, reinert@stats.ox.ac.uk*

In this paper, we give an explicit bound on the distance to chi-square for the likelihood ratio statistic when the data are realisations of independent and identically distributed random elements. To our knowledge, this is the first explicit bound which is available in the literature. The bound depends on the number of samples as well as on the dimension of the parameter space. We illustrate the bound with three examples: samples from an exponential distribution, samples from a normal distribution and logistic regression.

1. Introduction. One of the most celebrated theorems in theoretical statistics is Wilks' theorem, which states that under appropriate conditions, $2\times$ a log-likelihood ratio statistic is approximately chi-square distributed. This result is very useful when testing the null hypothesis $H_0 : \boldsymbol{\theta} \in \Theta_0$ against the alternative hypothesis $H_1 : \boldsymbol{\theta} \in \Theta$, where $\Theta_0 \subset \Theta$, using a generalised likelihood ratio test. Such tests arise, for example, in the area of model reduction, with the aim of finding a relatively simple model which explains the data reasonably well; see, for example, Chapter 6.5 in [6]. For this test, the number of degrees of freedom of the asymptotic chi-square distribution under the null hypothesis is $r = \dim(\Theta) - \dim(\Theta_0)$; see [25], as well as Chapter 12 in [16] and Chapter 16 in [23] for more details. This test is intriguing because of its generality; in [8] the term *Wilks' phenomenon* is coined for the fact that the asymptotic distribution of the likelihood ratio statistic does not depend on the "nuisance" parameters of the particular random mechanism which underlies the observations.

For any generalised likelihood ratio test, there are only finitely many observations available. The quality of the approximation will depend on the number of observations, and also on the distribution of the observations under the null hypothesis. As noted, for example, in [24], the quality of the chi-square approximation for a small sample size is unknown. To date, bounds on the distance to the chi-square distribution are only available in special cases. This paper addresses the problem through the use of Stein's method. The key ingredients are [10], where Stein's method for chi-square approximation is developed, and [1], where the distance to normality for maximum likelihood estimators is bounded using Stein's method for multivariate normal approximation. We shall apply these results in order to obtain our main theorem, Theorem 2.3. This theorem gives an explicit bound on the distance between the log-likelihood ratio statistic and the corresponding chi-square distribution.

Our results are the first ones which give an explicit bound to the chi-square distribution in Wilks' theorem under a general setting. These bounds are not optimised with respect to the constants. Their importance is of theoretical nature, but they can also be viewed as indicative of situations when the chi-square approximation does not hold. To illustrate this point, if d is the dimension of the parameter space Θ, r the number of degrees of freedom and n the number of observations, then our bounds tend to 0 as $n \to \infty$ when d is $o\left(n^{\frac{1}{18}}\right)$ when r is fixed, and when d is $o\left(n^{\frac{1}{23}}\right)$ when r is allowed to be of the same order as d. Hence the

Received August 2018; revised May 2019.
MSC2010 subject classifications. Primary 62H15, 62E17; secondary 60F05.
Key words and phrases. Log-likelihood ratio statistics, Wilks' theorem, Stein's method.

dimension of the parameter space is allowed to increase with n, but only very slowly. In particular, in the regime considered in [22], that d grows linearly with n, our bounds will tend to infinity with increasing sample size, as they should in this case. In the special case of logistic regression, [19] reported that the chi-square asymptotic is still valid when $d = o\left(n^{\frac{2}{3}}\right)$. While the criterion $d = o\left(n^{\frac{1}{23}}\right)$ is not as strong, our bound is explicit and is derived in a more general setting.

This paper opens up some avenues for further research. The first avenue relates to the order of the bound. In [10], for the Pearson chi-square statistic with fixed number of cells, a bound to the chi-square distribution of order n^{-1} is obtained, through making use of the quadratic form of the chi-square statistic. In contrast, Theorem 2.3 gives a bound of the order $n^{-1/2}$ for fixed d, with no clear possibility of improving the bounds. It has been suggested in the past that Pearson's chi-square statistic is closer to a chi-square distribution than the corresponding log-likelihood ratio statistic; see, for example, the chapter on historical perspective in [20]. It is probable that including a Bartlett correction in the log-likelihood ratio statistic as in [26] will improve its asymptotic performance; see Chapter 6.11 in [6]. In future work, it will be interesting to explore the discrepancy between the two tests. As a related question, in [22] it is shown that when the ratio $d(n)/n \to \kappa > 0$, a scaled chi-square statistic provides a good approximation for a class of logistic models; it would be interesting to explore this approximation further. Moreover, the bounds which are derived in this paper are not claimed to be tight—indeed the Cauchy–Schwarz inequality is repeatedly used. Obtaining tighter bounds remains an open problem.

A second avenue concerns the assumptions. In this paper, the observations $\mathbf{X}$ are assumed to be independent and identically distributed. As our proof is mainly based on Stein's method, generalisations to weakly dependent observations are straightforward in principle; see, for example, [4] and references therein. For simplicity of exposition, this paper concentrates on the classical i.i.d. case.

Moreover, for convenience in this paper we assume throughout that Θ is open, and Θ_0 is either open or a one-point set. In [25], it is assumed that Θ_0 is a hyperplane in the Euclidean space Θ. This assumption is weakened in [5], but an essential assumption for the chi-square asymptotics to hold is that the sets Θ_0 and Θ are (locally) equal to linear spaces (see [23], page 228); open sets, for example, satisfy this condition, but half-lines are not locally linear at their boundary points. A key assumption is that the true parameter does not lie on the boundary of the parameter space. This assumption is not easy to verify; when it is violated, then a rather different asymptotic behaviour can occur; see, for example, [5, 9, 21] and [11]. Allowing for more general parameter spaces is another item for further research.

The third avenue for further research concerns the method used for the proofs. This paper relies heavily on Stein's method. The conditions in our paper are such that the log-likelihood is locally linear, and hence resembles a quantitative approach to locally asymptotically normal models in the sense of Le Cam [14]. In contrast to Le Cam's general theory, instead of considering any small perturbation around the true parameter, we restrict attention to the maximum-likelihood estimator. This restriction allows to apply results from [2]. Expanding the results to provide a quantitative framework for Le Cam's theory will be part of future work.

The paper is structured as follows. Section 2 gives the general result. The proof is presented in modular form as a collection of lemmas, because the different steps in the approximation may be of independent interest. The proof also relies on Theorem 2.1 which is of interest in its own right as it gives an explicit bound on the distance to chi-square (in terms of smooth test functions) for a general standardised chi-square-type statistic based on score functions. Section 3 illustrates the result in three examples. First, we consider an example with a one-dimensional parameter, namely the exponential distribution. The second example is that of the normal distribution with two-dimensional parameter (μ, σ^2). The last example is logistic regression. Finally, Section 4 gives the proof of Theorem 2.1.

2. The general result. Before stating the general result, we introduce some notation. Let $C_b^n(\mathbb{R})$ denote the space of bounded functions on $\mathbb{R}$ with bounded kth order derivatives for $k \leq n$. For $h \in C_b^n$, we let $|||h|||_m = \max\{\|h^{(k)}\|, k = 0, 1, \dots, n\}$. Let $\nabla = (\frac{\partial}{\partial \theta_i}, i = 1, \dots, d)^\intercal$ denote the gradient operator, so that $\nabla\nabla^\intercal = (\frac{\partial^2}{\partial\theta_i\,\partial\theta_j})_{i,j=1,\dots,d}$ is the Hessian operator. The third-order tensor $\nabla^{(3)} := \nabla \otimes \nabla \otimes \nabla$ is so that $(\nabla^{(3)})_{i,j,k} = \frac{\partial^3}{\partial\theta_i\,\partial\theta_j\partial\theta_k}$ for $i, j, k = 1, \dots, d$ and for $d \times 1$ vectors $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$,

$$\nabla^{(3)} f[\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}] = \sum_{j=1}^{d}\sum_{k=1}^{d}\sum_{s=1}^{d} u_j v_k w_s \frac{\partial^3}{\partial\theta_j \partial\theta_k \partial\theta_s} f$$

is a scalar, while the vector $\nabla^{(3)} f[\boldsymbol{v}, \boldsymbol{w}]$ is given by

$$\nabla^{(3)} f[\boldsymbol{v}, \boldsymbol{w}] = \sum_{j=1}^{d}\sum_{k=1}^{d}\sum_{s=1}^{d} v_k w_s \frac{\partial^3}{\partial\theta_j\,\partial\theta_k \partial\theta_s} f.$$

Let $\boldsymbol{X} = (\boldsymbol{X_1}, \dots, \boldsymbol{X_n})$ be independent and identically distributed (i.i.d.) observations from a distribution with probability density function $f(\boldsymbol{x}|\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)^\intercal \in \Theta \subset \mathbb{R}^d$. The test problem is

$$H_0 : \theta_{0,j} = 0, \quad j = 1, \dots, r$$

against the general alternative $H_1 : \boldsymbol{\theta} \in \Theta$. Here, Θ is open, and Θ_0 is either open or a one-point set. Assume that $\dim(\Theta) = d$; then $\Theta_0 = \{\boldsymbol{\theta} \in \Theta : \theta_{0,j} = 0 \text{ for } j = 1, \dots, r\}$ has dimension $d - r$. Writing $\boldsymbol{\theta} = (\boldsymbol{\theta}_{[1:r]}, \boldsymbol{\theta}_{[r+1:d]})^\intercal$ where $\boldsymbol{\theta}_{[1:r]}$ is the vector of the first r components of $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_{[r+1:d]}$ is the vector of the remaining $d - r$ components of $\boldsymbol{\theta}$, the null hypothesis translates to $H_0 : \boldsymbol{\theta}_{0,[1:r]} = \boldsymbol{0}$.

Let $L(\boldsymbol{\theta}; \boldsymbol{x}) = \prod_{i=1}^n f(\boldsymbol{x_i}|\boldsymbol{\theta})$ denote the likelihood function. Set

$$\hat{\boldsymbol{\theta}}^{\text{res}}(\boldsymbol{x}) = \operatorname*{argmax}_{\boldsymbol{\theta}\in\Theta_0} L(\boldsymbol{\theta}; \boldsymbol{x}) = \big(\boldsymbol{0}_{[1:r]}, \hat{\boldsymbol{\theta}}^*_{[r+1:d]}(\boldsymbol{x})\big)^\intercal$$

$$\hat{\boldsymbol{\theta}}_n(\boldsymbol{x}) = \operatorname*{argmax}_{\boldsymbol{\theta}\in\Theta} L(\boldsymbol{\theta}; \boldsymbol{x});$$

under the conditions which will be specified, these quantities exist. The log-likelihood ratio statistic is

$$-2\log\Lambda = 2\log\left(\frac{T_1}{T_2}\right) \quad \text{with } T_1 = \frac{L(\hat{\boldsymbol{\theta}}_n(\boldsymbol{x}); \boldsymbol{x})}{L(\boldsymbol{\theta_0}; \boldsymbol{x})} \text{ and } T_2 = \frac{L(\hat{\boldsymbol{\theta}}^{\text{res}}(\boldsymbol{x}); \boldsymbol{x})}{L(\boldsymbol{\theta_0}; \boldsymbol{x})} \tag{2.1}$$

with $\boldsymbol{\theta_0}$ the unknown true parameter. Thus, T_1 is the likelihood ratio for testing the simple null hypothesis that $\boldsymbol{\theta} = \boldsymbol{\theta_0}$ against the alternative that $\boldsymbol{\theta} \in \Theta$, whereas T_2 is the likelihood ratio for testing the simple null hypothesis that $\boldsymbol{\theta} = \boldsymbol{\theta_0}$ against the alternative that $\boldsymbol{\theta} \in \Theta_0$.

The Fisher information matrix for one random vector is denoted by $I(\boldsymbol{\theta_0})$, which again is assumed to exist. We write $\ell(\boldsymbol{\theta}; \boldsymbol{x}) = \sum_{i=1}^n \ell_{\boldsymbol{x}_i}(\boldsymbol{\theta})$ with $\ell_{\boldsymbol{x}_i}(\boldsymbol{\theta}) = \log(f(\boldsymbol{x_i}|\boldsymbol{\theta}))$. The score function for $\boldsymbol{\theta_0}$ is

$$S(\boldsymbol{\theta_0}) = S(\boldsymbol{\theta_0}, \boldsymbol{x}) = \nabla \log L(\boldsymbol{\theta_0}; \boldsymbol{x}) = \sqrt{n}\begin{pmatrix}\boldsymbol{\xi}(\boldsymbol{\theta_0}, \boldsymbol{x})\\ \boldsymbol{\eta}(\boldsymbol{\theta_0}, \boldsymbol{x})\end{pmatrix} \tag{2.2}$$

with column vectors $\boldsymbol{\xi} = (\xi_1, \dots, \xi_r)^\intercal \in \mathbb{R}^r$ and $\boldsymbol{\eta} = (\eta_1, \dots, \eta_{d-r})^\intercal \in \mathbb{R}^{d-r}$. We omit the arguments $\boldsymbol{x}$ and $\boldsymbol{\theta}$ when they are obvious from the context. In the sequel, expectations are taken under the true parameter $\boldsymbol{\theta_0}$ unless otherwise indicated; $\mathbb{E}_{\boldsymbol{\theta}}$ signifies that the expectation is taken under $f(\boldsymbol{x}|\boldsymbol{\theta})$. We often abbreviate $\hat{\boldsymbol{\theta}}^*(\boldsymbol{x}) = \hat{\boldsymbol{\theta}}^*_{[r+1:d]}(\boldsymbol{x})$.

2.1. *The assumptions on the probability density function.* We write

$$I(\boldsymbol{\theta_0}) = \begin{pmatrix} A & B \\ B^\intercal & C \end{pmatrix}, \tag{2.3}$$

where for any $r \in \{1, 2, \dots, d\}$, $A = A^\intercal \in \mathbb{R}^{r\times r}$, $B \in \mathbb{R}^{r\times(d-r)}$, and $C = C^\intercal \in \mathbb{R}^{(d-r)\times(d-r)}$. We assume that the submatrices in (2.3) satisfy that C is invertible and that $A - BC^{-1}B^\intercal$ is positive definite. We make the following assumptions:

(C.1) Identifiability: the densities defined by any two different values of $\boldsymbol{\theta}$ are distinct;
(C.2) $\ell(\boldsymbol{\theta}; \boldsymbol{x})$ is three times differentiable with respect to the unknown vector parameter, $\boldsymbol{\theta}$, and the third partial derivatives are continuous in $\boldsymbol{\theta}$;
(C.3) for any $\boldsymbol{\theta_0} \in \Theta$ and for $\mathbb{X}$ denoting the support of the data,
 (a) there exists $\epsilon_1(\boldsymbol{\theta_0}) > 0$ and functions $M_{rst}(\boldsymbol{x})$ (they can depend on $\boldsymbol{\theta_0}$), such that for $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_d)$ and $r, s, t, j = 1, 2, \dots, d$,

$$\left|\frac{\partial^3}{\partial\theta_r\,\partial\theta_s\partial\theta_t}\ell(\boldsymbol{\theta}; \boldsymbol{x})\right| \le M_{rst}(\boldsymbol{x}) \quad \forall \boldsymbol{x} \in \mathbb{X}, |\theta_j - \theta_{0,j}| < \epsilon_1(\boldsymbol{\theta_0}),$$

 with $\mathbb{E}[M_{rst}(\boldsymbol{X})] < \infty$;
 (b) there exists $\epsilon_2(\boldsymbol{\theta_0}) > 0$ and functions $M^*_{k^*j^*l^*}(\boldsymbol{x})$, such that for all $k^*, j^*, l^*, j \in \{1, 2, \dots, d-r\}$ it holds that for all $\boldsymbol{x} \in \mathbb{X}$, if $|\theta^*_j - \theta_{*,j}| < \epsilon_2(\boldsymbol{\theta_0})$ then

$$\left|\frac{\partial^3}{\partial\theta_{k^*+r}\,\partial\theta_{j^*+r}\partial\theta_{l^*+r}}\ell(\boldsymbol{\theta^*}, \boldsymbol{x})\right| \le M^*_{k^*j^*l^*}(\boldsymbol{x}).$$

 In addition, $\mathbb{E}[M^*_{k^*j^*l^*}(\boldsymbol{X})] < \infty$;
(C.4) for all $\boldsymbol{\theta} \in \Theta$, $\mathbb{E}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{X}_i}(\boldsymbol{\theta})] = 0$;
(C.5) $I(\boldsymbol{\theta})$ is finite, symmetric and positive definite, and for $r, s = 1, 2, \dots, d$,

$$n\big[I(\boldsymbol{\theta})\big]_{rs} = \mathbb{E}_{\boldsymbol{\theta}}\left\{\frac{\partial}{\partial\theta_r}\ell(\boldsymbol{\theta}; \boldsymbol{X})\frac{\partial}{\partial\theta_s}\ell(\boldsymbol{\theta}; \boldsymbol{X})\right\} = -\mathbb{E}_{\boldsymbol{\theta}}\left\{\frac{\partial^2}{\partial\theta_r\,\partial\theta_s}\ell(\boldsymbol{\theta}; \boldsymbol{X})\right\}.$$

 This condition implies that $nI(\boldsymbol{\theta})$ is the covariance matrix of $\nabla(\ell(\boldsymbol{\theta}; \boldsymbol{x}))$;
(C.6) for $\kappa = 2, 4$,

$$\mathbb{E}\big((M_{kjv}(\boldsymbol{X}))^\kappa \,|\, |\hat{\theta}_n(\boldsymbol{x})_{(m)} - \theta_{0,(m)}| < \epsilon\big) < \infty,$$
$$\mathbb{E}\big((M^*_{k^*j^*v^*}(\boldsymbol{X}))^\kappa \,|\, |\hat{\theta}_*(\boldsymbol{x})_{(m^*)} - \theta_{*,(m^*)}| < \epsilon\big) < \infty,$$

 where $M_{kjv}(\boldsymbol{x})$ and $M^*_{k^*j^*v^*}(\boldsymbol{X})$ are as in (C.3);
(C.7) the random variables $Y_{i,j}(\boldsymbol{\theta}) = \frac{\partial}{\partial\theta_i}\log f(\boldsymbol{X_j}|\boldsymbol{\theta})$ have finite absolute moments up to 8th order.

Assumptions (C.1), (C.2), (C.3)(a), (C.4) and (C.5) are classical regularity conditions that were formulated mainly in the 1940s in order to make informal derivations mathematically thorough. Even though these conditions could be relaxed, mainly with respect to requirements related to the partial derivatives of the log-likelihood function, they are still of great interest because of their simple nature which can lead to simple and easy to understand proofs. Under these conditions, [7] proves that

$$\sqrt{n}(\boldsymbol{\hat{\theta}_n}(\boldsymbol{X}) - \boldsymbol{\theta_0}) \xrightarrow[n\to\infty]{\mathrm{d}} \big[I(\boldsymbol{\theta_0})\big]^{-\frac{1}{2}}\boldsymbol{Z},$$

where for d fixed, $I_{d\times d}$ is the $d \times d$ identity matrix, $\boldsymbol{Z} \sim \mathrm{N}_d(\boldsymbol{0}, I_{d\times d})$, and $\xrightarrow{\mathrm{d}}$ denotes convergence in distribution. Under the same conditions, [1] gives an explicit upper bound to the distance of the distribution of the MLE to the normal for i.i.d. random vectors.

There are different sets of assumptions available which ensure existence and asymptotic normality of the MLE. In Theorem 5.39 of [23], the asymptotic normality of the MLE is proven under weaker assumptions; mainly a Lipschitz condition on the Hellinger distance, but without an explicit bound on the rate of convergence. Furthermore, [23] proves similar results for the asymptotic normality of the wider class of M-estimators; the explicit bounds in [1] have not yet been extended to such conditions.

Let the subscript $(m) \in \{1, 2, \dots, d\}$ denote an index for which the deviation $|\hat{\theta}_n(\boldsymbol{x})_{(m)} - \theta_{0,(m)}|$ is the largest among the d components;

$$(m) \in \{1, \dots, d\} \quad \text{such that } \left|\hat{\theta}_n(\boldsymbol{x})_{(m)} - \theta_{0,(m)}\right| = \max_{j=1,\dots,d} \left|\hat{\theta}_n(\boldsymbol{x})_j - \theta_{0,j}\right|$$

and similarly $(m^*) \in \{1, 2, \dots, d-r\}$ is defined with $\boldsymbol{\theta_0}$ replaced by $\boldsymbol{\theta^*} = (\theta_{*,1}, \dots, \theta_{*,d-r})$.

2.2. *A chi-square approximation.* First, we give a general chi-square approximation result which may be of independent interest and which is crucial for our overall bound; the proof is found in Section 4.

THEOREM 2.1. *Let $Z_{i,j}, i = 1, \dots, r, j = 1, \dots, n$ be mean zero random variables such that $Z_{i,j}$ is independent of $\{Z_{k,\ell}, k = 1, \dots, r, \ell \neq j\}$ and $Z_{i,j}$ has the same distribution as $Z_{i,\ell}$, for $i = 1, \dots, r$. Moreover, assume that $Z_{i,j}$ has moments up to order 8 so that*

$$\beta(I) = \mathbb{E}\Big(\prod_{i \in I} Z_{i,1}\Big), \tag{2.4}$$

for I a multiset of indices in $\{1, \dots, r\}$, exists for $|I| \le 8$. Let

$$Z_i = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} Z_{i,j}; \quad i = 1, \dots, r, \quad \text{and} \quad \mathbf{Z} = (Z_1, \dots, Z_r)^{\intercal}.$$

Let $\tau_{i,k} = \operatorname{Cov}(Z_{i,1}, Z_{k,1})$ and assume that the $r \times r$ matrix $\tau = (\tau_{i,k})_{i,k=1,\dots,r}$ is invertible. Let $U = \tau^{-1}$ and let $T = \mathbf{Z}^{\intercal} U \mathbf{Z}$. Then for all functions $g \in C_b^3(\mathbb{R})$,

$$\left|\mathbb{E}[g(T))] - \mathbb{E}[g(\chi_r)]\right| \le \frac{16 |||g|||_3}{r\sqrt{n}} R(r)$$

with

$$\begin{aligned} R(r) &= \frac{4}{\sqrt{n}} + \frac{1}{2\sqrt{n}} \mathbb{E}(W^2) \\ &\quad + \sqrt{B_1(r,n)} \\ &\quad \times \left(\frac{r}{2} + \frac{3}{2} \sqrt{\sum_{i,k,a,b,,e,f=1}^{r} U_{i,k} U_{a,b} U_{e,f} \beta(i,a,b) \beta(k,e,f) + \frac{r^{\frac{3}{2}}}{n}}\right) \\ &\quad + \frac{\sqrt{B_2(r,n)}}{\sqrt{n}} \left(\frac{1}{4}\left(\sqrt{\mathbb{E}(W^2)} + \sqrt{r}\right) + \sqrt{\mathbb{E}(W^2) n^{-\frac{1}{2}} + r^2 + 2r}\right), \end{aligned} \tag{2.5}$$

where $W = \sum_{i,k=1}^{r} U_{i,k} Z_{i,1} Z_{k,1}$,

$$B_1(r,n) = \frac{4rn}{n-1} + \frac{1}{n} \mathbb{E}(W^2)$$

and

$$B_2(r,n) = 96\mathbb{E}(W^2) + \frac{1}{n}\Bigg(\frac{1}{n}\mathbb{E}(W^4) + 24\mathbb{E}(W^3) + 32\sum_{a,b,i,k,c,d,e,f}^{r} U_{a,b}U_{i,k}U_{c,d}U_{e,f}\beta(i,a,c,e,b)\beta(c,k,f) + 16\sum_{a,b,i,k,c,d,e,f}^{r} U_{a,b}U_{i,k}U_{c,d}U_{e,f}\beta(i,a,c,e)\beta(b,d,k,f)\Bigg).$$

REMARK 2.2.

1. If $\mathbb{E}(W^k)$ is of order r^k for $k=2,3,4$, which is the case when U is diagonal with entries of order 1, and $Z'_{i,j}$s being of order 1, then $B_1 = O\left(r + \frac{r^2}{n}\right)$, $B_2 = O\left(r^2 + \frac{r^8}{n}\right)$ and the overall bound tends to 0 as $n \to \infty$ if $r = o\left(n^{\frac{3}{8}}\right)$. As the proof of Theorem 2.1 involves repeated applications of the Cauchy–Schwarz inequality, we do not expect this bound to be tight.
2. Theorem 2.1 can be applied to the score-test like statistic

$$\begin{pmatrix}\boldsymbol{\xi}\\ \boldsymbol{\eta}\end{pmatrix}^{\intercal}[I(\boldsymbol{\theta_0})]^{-1}\begin{pmatrix}\boldsymbol{\xi}\\ \boldsymbol{\eta}\end{pmatrix},$$

which is closely related to the classical score test statistic in which $I(\boldsymbol{\theta_0})$ is replaced by $I(\hat{\boldsymbol{\theta}})$. Using Taylor expansion to assess $[I(\boldsymbol{\theta_0})]^{-1} - [I(\hat{\boldsymbol{\theta}})]^{-1}$, it is straightforward to obtain a bound on the distance to the appropriate chi-square distribution for the score test. Due to space issues, we do not pursue this application here.

2.3. *A bound on the distance to chi-square for Wilks' statistic*. The main result of this paper is as follows.

THEOREM 2.3. *Let $\boldsymbol{X_1}, \boldsymbol{X_2}, \dots, \boldsymbol{X_n}$ be i.i.d. $\mathbb{R}^t$-valued, $t \in \mathbb{Z}^+$, random vectors with probability density* (*or mass*) *function $f(\boldsymbol{x_1}|\boldsymbol{\theta})$, for which the parameter space Θ is an open subset of $\mathbb{R}^d$. Assume that the MLE exists and is unique and that* (*C*.1)–(*C*.7) *are satisfied. Then for $-2\log\Lambda$ as in* (2.1), *$h \in \mathbb{C}_b^3(\mathbb{R})$ and $K \sim \chi_r^2$, it holds that*

$$\begin{aligned}&\big|\mathbb{E}[h(-2\log\Lambda)] - \mathbb{E}[h(K)]\big|\\ &\quad\le \frac{16\|h\|_3}{r\sqrt{n}}R(r)\\ &\qquad+\frac{1}{\sqrt{n}}\big(\|h'\|\big(K_1(\boldsymbol{\theta_0}) + K_1^*(\boldsymbol{\theta_0})\big) + K_2(\boldsymbol{\theta_0}) + K_2^*(\boldsymbol{\theta_0})\big),\end{aligned} \tag{2.6}$$

where $R(r)$ is given in (2.5) *in Theorem* 2.1 *with*

$$Z_{i,j} = \frac{\partial}{\partial\theta_i}\log\big(f(\boldsymbol{X_j}|\boldsymbol{\theta_0})\big) - \sum_{k=1}^{d-r}(BC^{-1})_{i,k}\frac{\partial}{\partial\theta_{k+d}}\log\big(f(\boldsymbol{X_j}|\boldsymbol{\theta_0})\big). \tag{2.7}$$

In addition, for $0 < \epsilon \le \epsilon(\boldsymbol{\theta_0})$,

$$K_1(\boldsymbol{\theta_0}) = 3n\sum_{j=1}^{d}\sum_{k=1}^{d}\big[\mathbb{E}(Q_j^2Q_k^2)\big]^{\frac{1}{2}}\bigg[\mathrm{Var}\bigg(\frac{\partial^2}{\partial\theta_j\,\partial\theta_k}\log f(\boldsymbol{X_1}|\boldsymbol{\theta_0})\bigg)\bigg]^{\frac{1}{2}}$$

$$
\begin{aligned}
&+\sum_{l=1}^{d}\sum_{m=1}^{d}[[I(\boldsymbol{\theta_0})]^{-1}]_{lm} \\
&\times\sum_{j=1}^{d}\sum_{k=1}^{d}\sqrt{\operatorname{Var}\left(\frac{\partial^2}{\partial\theta_l\,\partial\theta_j}\log f(X_1|\boldsymbol{\theta_0})\right)}[\mathbb{E}(Q_j^6)\mathbb{E}(Q_k^6)\mathbb{E}(T_{mk}^6)]^{\frac{1}{6}}
\end{aligned} \tag{2.8}
$$

and

$$
\begin{aligned}
&K_2(\boldsymbol{\theta_0}) \\
&\quad=\frac{2\|h\|\sqrt{n}}{\epsilon^2}\mathbb{E}\left(\sum_{j=1}^{d}Q_j^2\right) \\
&\qquad+\|h'\|\sqrt{n}\frac{7}{3}\sum_{j=1}^{d}\sum_{k=1}^{d}\sum_{l=1}^{d}[\mathbb{E}(Q_j^2Q_k^2Q_l^2)]^{\frac{1}{2}}[\mathbb{E}[(M_{jkl}(\boldsymbol{X}))^2||Q_{(m)}|<\epsilon]]^{\frac{1}{2}} \\
&\qquad+\frac{\|h'\|}{\sqrt{n}}\sum_{q=1}^{d}\sum_{k=1}^{d}|[[I(\boldsymbol{\theta_0})]^{-1}]_{kq}| \\
&\qquad\times\sum_{j=1}^{d}\sum_{l=1}^{d}\sum_{s=1}^{d}\sqrt{\mathbb{E}(Q_j^2Q_l^2Q_s^2)}[\mathbb{E}(T_{kj}^4||Q_{(m)}|<\epsilon)]^{\frac{1}{4}} \\
&\qquad\times[\mathbb{E}((M_{qsl}(\boldsymbol{X}))^4||Q_{(m)}|<\epsilon)]^{\frac{1}{4}} \\
&\qquad+\frac{\|h'\|}{4\sqrt{n}}\sum_{b=1}^{d}\sum_{k=1}^{d}\sum_{s=1}^{d}\sum_{q=1}^{d}\sum_{l=1}^{d}\sum_{j=1}^{d}|[[I(\boldsymbol{\theta_0})]^{-1}]_{qb}|\sqrt{\mathbb{E}(Q_k^2Q_s^2Q_j^2Q_l^2)} \\
&\qquad\times[\mathbb{E}((M_{bsk}(\boldsymbol{X}))^4||Q_{(m)}|<\epsilon)]^{\frac{1}{4}}[\mathbb{E}((M_{qjl}(\boldsymbol{X}))^4||Q_{(m)}|<\epsilon)]^{\frac{1}{4}}
\end{aligned} \tag{2.9}
$$

and $K_1^*(\boldsymbol{\theta_0})$, $K_2^*(\boldsymbol{\theta_0})$ *are the versions of* $K_1(\boldsymbol{\theta_0})$ *and* $K_2(\boldsymbol{\theta_0})$, *respectively*, *under the null hypothesis*. *Here*,

$$
\mathbf{Q}=\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(X)-\boldsymbol{\theta_0};\qquad \mathbf{Q}^*=(Q_1^*,\dots,Q_{d-r}^*)^{\mathsf{T}};
$$
$$
T=(T_{i,j})_{i,j=1,\dots,d};\qquad T^*=(T_{i,j}^*)_{i,j=1,\dots,d-r}
$$

with

$$
\begin{aligned}
&Q_j^*=Q_j^*(\boldsymbol{X},\boldsymbol{\theta_*}):=\hat{\theta}^*(\boldsymbol{X})_j-\theta_j^*\quad\forall j=1,2,\dots,d-r, \\
&T_{lj}=T_{lj}(\boldsymbol{\theta_0},\boldsymbol{X})=\frac{\partial^2}{\partial\theta_l\,\partial\theta_j}\ell(\boldsymbol{\theta_0};\boldsymbol{X})+n\big[I(\boldsymbol{\theta_0})\big]_{lj},\quad j,l\in\{1,2,\dots,d\}, \\
&T_{lj}^*=T_{lj}^*(\boldsymbol{\theta_0},\boldsymbol{X})=\frac{\partial^2}{\partial\theta_{l+r}\,\partial\theta_{j+r}}\ell(\boldsymbol{\theta_0};\boldsymbol{X})+nC_{lj},\quad j,l\in\{1,2,\dots,d-r\}.
\end{aligned} \tag{2.10}
$$

REMARK 2.4. The differentiability assumptions for Theorem 2.3 are made for convenience rather than for mathematical requirement. Indeed in [15] the classical assumptions for asymptotic normality of maximum likelihood estimators are weakened by replacing differentiability requirements with Hellinger-differentiability conditions; see also [18]. This avenue of research will be part of future investigation. Further discussion about alternative assumptions for Wilks' theorem can be found in Chapter 12 of [16].

REMARK 2.5. (1) At first glance, the bound seems complicated. However, the examples that follow show that the terms are easily calculated.

(2) For Q_j as in (2.10), to see that $\mathbb{E}(Q_j^2) = \mathcal{O}(\frac{1}{n})$, from the asymptotic normality of the MLE, it follows that $\sqrt{n}\mathbb{E}(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) - \boldsymbol{\theta_0}) \xrightarrow[n\to\infty]{} \mathbf{0}$ and $\mathrm{Cov}(\sqrt{n}[I(\boldsymbol{\theta_0})]^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) - \boldsymbol{\theta_0})) \xrightarrow[n\to\infty]{} I_{d\times d}$. Therefore, for $j = 1, \dots, d$, $\mathbb{E}(Q_j) = o\left(\frac{1}{\sqrt{n}}\right)$, and

$$n[I(\boldsymbol{\theta_0})]^{\frac{1}{2}}\mathrm{Cov}(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}))[I(\boldsymbol{\theta_0})]^{\frac{1}{2}} \xrightarrow[n\to\infty]{} I_{d\times d}.$$

Hence

$$\mathrm{Var}(\hat{\theta}_n(\boldsymbol{X})_j) = \mathcal{O}\left(\frac{1}{n}\right), \quad \forall j \in \{1, 2, \dots, d\}$$

and

$$\mathbb{E}(Q_j^2) = \mathrm{Var}(\hat{\theta}_n(\boldsymbol{X})_j) + [\mathrm{E}(Q_j)]^2 = \mathcal{O}\left(\frac{1}{n}\right). \tag{2.11}$$

(3) With T_{lj} as in (2.10), using (C.5) and the fact that $\boldsymbol{X_1}, \boldsymbol{X_2}, \dots, \boldsymbol{X_n}$ are i.i.d. yields

$$\begin{aligned}\mathbb{E}(T_{lj}^2) &= \mathbb{E}\left(\frac{\partial^2}{\partial\theta_l\,\partial\theta_j}\ell(\boldsymbol{\theta_0}; \boldsymbol{X}) + n[I(\boldsymbol{\theta_0})]_{lj}\right)^2 \\ &= \mathrm{Var}\left(\frac{\partial^2}{\partial\theta_l\,\partial\theta_j}\ell(\boldsymbol{\theta_0}; \boldsymbol{X})\right) \qquad (2.12) \\ &= n\,\mathrm{Var}\left(\frac{\partial^2}{\partial\theta_l\,\partial\theta_j}\log(f(\boldsymbol{X_1}|\boldsymbol{\theta_0}))\right),\end{aligned}$$

showing that $\mathbb{E}(T_{lj}^2)$ is $\mathcal{O}(n)$.

(4) For fixed d, the upper bound we give in (2.6) is $\mathcal{O}\left(n^{-1/2}\right)$. The expression for $R(r)$ given in Theorem 2.1, is $\mathcal{O}(1)$. In addition, using (2.11) and (2.12) it can be deduced that

$$K_1(\boldsymbol{\theta_0}) = \mathcal{O}(1), \qquad K_1^*(\boldsymbol{\theta_0}) = \mathcal{O}(1), \qquad K_2(\boldsymbol{\theta_0}) = \mathcal{O}(1), \qquad K_2^*(\boldsymbol{\theta_0}) = \mathcal{O}(1).$$

Hence, the upper bound in Theorem 2.3 is $\mathcal{O}\left(n^{-1/2}\right)$.

(5) If the dimensionality of the parameter is not fixed but if the entries of $I(\boldsymbol{\theta_0})$ are of order 1, the entries of the matrix U are of order $\mathcal{O}\left((d-r)^2\right)$. In addition, $Z_{i,j}$ as in (2.7) are also of order $\mathcal{O}((d-r)^2)$, and hence W in (2.5) is of order $\mathcal{O}(r^2(d-r)^6)$. Therefore, the first term of the bound is $\mathcal{O}(\sqrt{r^5(d-r)^{18}n^{-1}})$ which is of maximal order $\mathcal{O}(\sqrt{d^{23}n^{-1}})$. This term is small when $d = o(n^{\frac{1}{23}})$. For fixed r, the term is small when $d = o(n^{\frac{1}{18}})$. Using (2.11) and (2.12), the second and third terms (related to $K_1(\boldsymbol{\theta_0})$ and $K_1^*(\boldsymbol{\theta_0})$) of the bound are of order $d^2n^{-1/2}$, while the fourth and fifth terms (related to $K_2(\boldsymbol{\theta_0})$ and $K_2^*(\boldsymbol{\theta_0})$) are both $\mathcal{O}(d^3n^{-1/2})$. Hence, the overall order of the bound in the chi-square approximation for the likelihood ratio test is at most of order $d^{23/2}n^{-1/2}$ when both r and d are not fixed and the chi-square approximation is justified when $d = o(n^{1/23})$. The proof of Theorem 2.3 involves repeated applications of the Cauchy–Schwarz inequality, and hence we do not expect this bound to be tight.

(6) Due to the smoothness assumptions in this paper, the bound in Theorem 2.3 is not given in a standard probability distance. Instead it could be re-phrased in terms of the integral probability metric

$$d(\mu, \nu) = \sup_{h\in\mathbb{C}_b^2(\mathbb{R}):|||h|||_3\le 1} |\mathbb{E}h(X) - \mathbb{E}h(Y)|$$

where $X \sim \mu$ and $Y \sim \nu$. For more details on such metrics see for example [27] and [12].

2.4. *Investigation of the rate of d with respect to n.* Remark 2.5 indicates that our bound still goes to zero even when both the number d of the explanatory parameters as well as the number r of the restricted parameters under the null hypothesis are allowed to grow at a quite low rate with the sample size. In order to investigate further how large d (and also r) can be for the χ^2 approximation to be valid, we run some simple simulations from the specific example of the multivariate Student's t-distribution with v degrees of freedom; from now on, this is denoted by t_v. As in [17], n independent draws from the multivariate t_v distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ are realised as the product $Y_i r_i$, $i = 1, \dots, n$, with

$$Y_i | \boldsymbol{\mu}, \Sigma, r \overset{\text{ind}}{\sim} \mathrm{N}(\boldsymbol{\mu}, \Sigma), \quad \text{for } i = 1, 2, \dots, n,$$
$$r_i | v \overset{\text{i.i.d}}{\sim} \mathrm{Gamma}\left(\frac{v}{2}, \frac{v}{2}\right), \quad \text{for } i = 1, 2, \dots, n.$$

When both $\boldsymbol{Y} = (Y_1, Y_2, \dots, Y_n)$ and $\boldsymbol{r} = (r_1, r_2, \dots, r_n)$ are considered observed, then $(Y_1, Y_2, \dots, Y_n, r_1, r_2, \dots, r_n)$ comprise the complete data. Therefore, the complete-data likelihood function can be factored into the product of two distinct functions: the likelihood of $(\boldsymbol{\mu}, \Sigma)$ corresponding to the conditional distribution of $\boldsymbol{Y}$ given $\boldsymbol{r}$, and the likelihood function of v corresponding to the marginal distribution of $\boldsymbol{r}$. In this case, the maximum likelihood estimators for $\boldsymbol{\mu}$ and Σ are

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \frac{\sum_{i=1}^n r_i Y_i}{\sum_{i=1}^n r_i}, \\ \hat{\Sigma} &= \frac{1}{n} \sum_{i=1}^n r_i (Y_i - \hat{\boldsymbol{\mu}})(Y_i - \hat{\boldsymbol{\mu}})^\intercal. \end{aligned} \tag{2.13}$$

Consider the test problem $H_0 = \boldsymbol{\mu} = \mathbf{0}$ against the general alternative where both $\boldsymbol{\mu}$ and Σ are unrestricted. Using (2.13), the MLE under the alternative is $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) = (\hat{\boldsymbol{\mu}}, \hat{\Sigma})^\intercal$, while under the null, $\hat{\Sigma}_* = \frac{1}{n} \sum_{i=1}^n r_i Y_i Y_i^\intercal$. The log-likelihood ratio statistic is

$$-2 \log \Lambda = 2\big(\ell(\hat{\boldsymbol{\mu}}, \hat{\Sigma} | \boldsymbol{Y}, \boldsymbol{r}) - \ell(\hat{\Sigma}_* | \boldsymbol{Y}, \boldsymbol{r})\big). \tag{2.14}$$

Below we investigate the behaviour of the log-likelihood ratio statistic through some simulations.

Results from simulations. First, we generate 100 trials of n random independent observations, x, from the standard multivariate t_{10} distribution with dimension d. We take $n = 50, 100, 500, 1000$ and in all cases $d = 0.1n, 0.2n, \dots, 0.8n$, resulting in 4 vectors of length 8 each, and consider the test for the null hypothesis $\boldsymbol{\mu} = \mathbf{0}$. In general, there are $d^2 + d$ (d^2 for $\hat{\Sigma}$ and d for $\hat{\boldsymbol{\mu}}$) quantities to be estimated and under the null we restrict d parameters. At each trial, we evaluate $-2 \log \Lambda$ as in (2.14), which in turn for each combination of n and d, as above, gives a vector of 100 values. In two different experiments, we apply to these values the functions $h_1(x) = x$ and $h_2(x) = (x^2 + 2)^{-1}$. Then, for $j = 1, 2$, we calculate the sample means $\hat{\mathbb{E}}[h_j(-2 \log \Lambda)]$ and the relative differences

$$\mathrm{RD}_{h_j} := \frac{|\hat{\mathbb{E}}[h_j(-2 \log \Lambda)] - \tilde{\mathbb{E}}[h_j(K)]|}{\tilde{\mathbb{E}}[h_j(K)]}, \quad j = 1, 2,$$

where $K \sim \chi^2_d$ and $\tilde{\mathbb{E}}[h_j(K)]$ is the approximation of $\mathbb{E}[h_j(K)]$ up to three decimal places. To gauge the quality of the approximation, for each of the 4 vectors we find $m_j^{c^*}(n)$, which for each n denotes the smallest value of d such that $\mathrm{RD}_{h_j} > c^*$, where $j = 1, 2$ and c^* takes values in the set $\{0.1, 0.3, 0.5\}$. The behaviour for these values of d was monotone in our simulations. For $\boldsymbol{m_j^k} = (m_j^k(50), m_j^k(100), m_j^k(500), m_j^k(1000))$ the fitted slope against $\log(n)$ for the results related to $\boldsymbol{m_1^{0.3}}$, $\boldsymbol{m_1^{0.5}}$ and $\boldsymbol{m_2^{0.5}}$ is equal to 1, while the slopes related to $\boldsymbol{m_1^{0.1}}$, $\boldsymbol{m_2^{0.1}}$ and $\boldsymbol{m_2^{0.3}}$ are, up to 3 decimal places, equal to 0.956, 0.986, 0.943, respectively. Additionally, for h_1 the median of the observed absolute deviations of (estimated) expectations was 22.18. For each of $n = 50, 100, 500, 1000$, we found the first d such that the absolute deviation exceeded this median; in the coarse grained setup, we obtained, in order of n, $d = 35, 60, 200, 300$. A log-log fit resulted in a slope of 0.8554.

This set of simulations assumes independent observations. As an example for dependent observations, we use the t-distribution but now with covariance matrix $\Sigma = A^T A$, where A has i.i.d. uniform[0, 1] entries. We use the same test function h_1 and 4 replicas each for $n = 50, 100, 250, 500, 750$ with $d = 0.1n, 0.2n, \dots, 0.8n$; giving 160 observations. To gauge the distance, for each of the 4 observation vectors for each n we found the first $d = 0.1 * n$ such that the difference between the estimated expectation and the corresponding chi-square expectation exceeds the median (16.41) of these differences. We then fitted a slope in the log-log plot of $\log d$ against $\log n$. (We also considered h_2 in this simulation but the absolute difference decreased with increasing n, and for the relative difference there was no variation in the value of d, and hence we did not fit a slope.) The fitted slope for exceedances of the median was 0.837. This result indicates that in the presence of dependence, d should not grow as fast as n. The estimated slopes indicate an exponent n^γ of $\gamma > \frac{2}{3}$, where $\frac{2}{3}$ is the value obtained for the logistic regression case. However, the function considered is very special, and hence we cannot draw a general conclusion. Keeping in mind that the simulations are only indicative, they would however suggest that d could grow much faster with n than the theoretical results ensure.

2.5. *Proof of Theorem* 2.3. The log-likelihood ratio statistic can be expressed as in (2.1). The expected Fisher information matrix is given in (2.3); with C^{-1} assumed to exist. From now on, we will use the notation introduced in (2.10). The different steps are disentangled into results which hold for every realisation $\boldsymbol{x}$, and results which hold when taking expectations over test functions.

2.5.1. *Approximation for* $2\log T_1$ *and for* $2\log T_2$. The first step in the proof is to derive an approximation for $2\log T_1$.

LEMMA 2.6. *Under the assumptions of Theorem* 2.3,

$$2\log T_1 = n\big(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) - \boldsymbol{\theta_0}\big)^\intercal I(\boldsymbol{\theta_0})\big(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) - \boldsymbol{\theta_0}\big) + R_1 + R_2,$$

where using the notation in (2.10),

$$R_1 = R_1(\boldsymbol{X}, \boldsymbol{\theta_0}) = -\boldsymbol{Q}^\intercal T \boldsymbol{Q}$$

and

$$R_2 = R_2(\boldsymbol{X}, \boldsymbol{\theta_0}) = \nabla^{(3)}\left\{\frac{1}{3}\ell(\tilde{\boldsymbol{\theta}}; \boldsymbol{X}) - \ell(\tilde{\tilde{\boldsymbol{\theta}}}; \boldsymbol{X})\right\}[\boldsymbol{Q}, \boldsymbol{Q}, \boldsymbol{Q}]$$

for some $\tilde{\boldsymbol{\theta}}, \tilde{\tilde{\boldsymbol{\theta}}}$ *between* $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})$ *and* $\boldsymbol{\theta_0}$.

PROOF. The regularity condition (C.2) and a third-order Taylor expansion of $\ell(\boldsymbol{\theta_0}; \boldsymbol{x})$ about $\hat{\boldsymbol{\theta}}_n(\boldsymbol{x})$ yield

$$\ell(\boldsymbol{\theta_0}; \boldsymbol{x}) = \ell\big(\hat{\boldsymbol{\theta}}_n(\boldsymbol{x}); \boldsymbol{x}\big) - \boldsymbol{Q}^\intercal \nabla \ell\big(\hat{\boldsymbol{\theta}}_n(\boldsymbol{x}); \boldsymbol{x}\big) + \frac{1}{2}\boldsymbol{Q}^\intercal \nabla\nabla^\intercal\big(\ell\big(\hat{\boldsymbol{\theta}}_n(\boldsymbol{x}); \boldsymbol{x}\big)\boldsymbol{Q} - \frac{1}{6}\nabla^{(3)}\ell(\tilde{\boldsymbol{\theta}}; \boldsymbol{x})[\boldsymbol{Q}, \boldsymbol{Q}, \boldsymbol{Q}],$$

where $\tilde{\boldsymbol{\theta}}$ is between $\boldsymbol{\theta_0}$ and $\hat{\boldsymbol{\theta}}_n(\boldsymbol{x})$. As $\nabla\ell(\hat{\boldsymbol{\theta}}_n(\boldsymbol{x});\boldsymbol{x}) = \mathbf{0}$,

$$\begin{aligned}2\log T_1 &= -\boldsymbol{Q}^\intercal \nabla\nabla^\intercal\big(\ell(\hat{\boldsymbol{\theta}}_n(\boldsymbol{x});\boldsymbol{x}\big)\boldsymbol{Q} + \frac{1}{3}\nabla^{(3)}\ell(\tilde{\boldsymbol{\theta}};\boldsymbol{X})[\boldsymbol{Q},\boldsymbol{Q},\boldsymbol{Q}]\\ &= n\boldsymbol{Q}^\intercal I(\boldsymbol{\theta_0})\boldsymbol{Q} - \boldsymbol{Q}^\intercal\big\{\nabla\nabla^\intercal\big(\ell(\hat{\boldsymbol{\theta}}_n(\boldsymbol{x});\boldsymbol{X}\big) - nI(\boldsymbol{\theta_0})\big\}\boldsymbol{Q}\\ &\quad + \frac{1}{3}\nabla^{(3)}\ell(\tilde{\boldsymbol{\theta}};\boldsymbol{X})[\boldsymbol{Q},\boldsymbol{Q},\boldsymbol{Q}].\end{aligned}$$

Then, a first-order Taylor expansion of $\nabla\nabla^\intercal\ell(\hat{\boldsymbol{\theta}}_n(\boldsymbol{x});\boldsymbol{X})$ about $\boldsymbol{\theta_0}$ gives

$$\begin{aligned}2\log T_1 &= n\boldsymbol{Q}^\intercal I(\boldsymbol{\theta_0})\boldsymbol{Q} + R_1 + R_2\\ &= n\big(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) - \boldsymbol{\theta_0}\big)^\intercal I(\boldsymbol{\theta_0})\big(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) - \boldsymbol{\theta_0}\big) + R_1 + R_2.\end{aligned}$$

This completes the proof. □

From Lemma 2.6, the next approximation of the log-likelihood ratio is almost immediate. Using (2.1), Lemma 2.6 and its analogous expression for $2\log T_2$ with $\hat{\boldsymbol{\theta}}_n(\boldsymbol{x})$ replaced by $\hat{\boldsymbol{\theta}}^{\text{res}}(\boldsymbol{x})$,

$$\begin{aligned}-2\log\Lambda &= n\big(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) - \boldsymbol{\theta_0}\big)^\intercal I(\boldsymbol{\theta_0})\big(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) - \boldsymbol{\theta_0}\big) + R_1 + R_2\\ &\quad - n\big(\hat{\boldsymbol{\theta}}^{\mathbf{res}}(\boldsymbol{X}) - \boldsymbol{\theta_0}\big)^\intercal C\big(\hat{\boldsymbol{\theta}}^{\mathbf{res}}(\boldsymbol{X}) - \boldsymbol{\theta_0}\big) - R_1^* - R_2^*,\end{aligned}$$

where R_1 and R_2 are as in Lemma 2.6 and R_1^* and R_2^* are the corresponding expressions with $\hat{\boldsymbol{\theta}}_n(\boldsymbol{x})$ replaced by $\hat{\boldsymbol{\theta}}^{\text{res}}(\boldsymbol{x})$.

2.5.2. *Approximation for the score function.* From now on, let

$$G = G(\mathbf{X}) = \begin{pmatrix}\boldsymbol{\xi}\\ \boldsymbol{\eta}\end{pmatrix}^\intercal \big[I(\boldsymbol{\theta_0})\big]^{-1}\begin{pmatrix}\boldsymbol{\xi}\\ \boldsymbol{\eta}\end{pmatrix} - \boldsymbol{\eta}^\intercal C^{-1}\boldsymbol{\eta}.$$

It is straightforward to simplify this expression to give

$$G = (\boldsymbol{\xi} - BC^{-1}\boldsymbol{\eta})^\intercal(A - BC^{-1}B^\intercal)^{-1}(\boldsymbol{\xi} - BC^{-1}\boldsymbol{\eta}).$$

LEMMA 2.7. *Under the assumptions of Theorem* 2.3,

$$\begin{aligned}&n\big(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) - \boldsymbol{\theta_0}\big)^\intercal I(\boldsymbol{\theta_0})\big(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) - \boldsymbol{\theta_0}\big)\\ &\quad = \begin{pmatrix}\boldsymbol{\xi}\\ \boldsymbol{\eta}\end{pmatrix}^\intercal \big[I(\boldsymbol{\theta_0})\big]^{-1}\begin{pmatrix}\boldsymbol{\xi}\\ \boldsymbol{\eta}\end{pmatrix}\\ &\qquad - (\boldsymbol{R_3} + \boldsymbol{R_4})^\intercal(\boldsymbol{R_3} + \boldsymbol{R_4}) + 2\sqrt{n}\big(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) - \boldsymbol{\theta_0}\big)^\intercal\big[I(\boldsymbol{\theta_0})\big]^{\frac{1}{2}}(\boldsymbol{R_3} + \boldsymbol{R_4}),\end{aligned}$$

with

$$\begin{aligned}\boldsymbol{R_3} = \boldsymbol{R_3}(\boldsymbol{X},\boldsymbol{\theta_0}) &= \frac{1}{\sqrt{n}}\big[I(\boldsymbol{\theta_0})\big]^{-\frac{1}{2}}\boldsymbol{Q}^\intercal\big(\nabla\nabla^\intercal\big(\ell(\boldsymbol{\theta_0};\boldsymbol{X})\big) + n\big[I(\boldsymbol{\theta_0})\big]\big),\\ \boldsymbol{R_4} = \boldsymbol{R_4}(\boldsymbol{X},\boldsymbol{\theta_0}) &= \frac{1}{2\sqrt{n}}\big[I(\boldsymbol{\theta_0})\big]^{-\frac{1}{2}}\big(\nabla^{(3)}\ell(\boldsymbol{\theta};\boldsymbol{X})|_{\boldsymbol{\theta}=\boldsymbol{\theta_0^*}}\big)[\boldsymbol{Q},\boldsymbol{Q}]\end{aligned}\tag{2.15}$$

for some $\boldsymbol{\theta_0^*}$ *between* $\boldsymbol{\theta_0}$ *and* $\hat{\boldsymbol{\theta}}_n(\boldsymbol{x})$.

PROOF. For $\boldsymbol{\theta}_0^*$ between $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}_n(\boldsymbol{x})$, similarly as in [1], the regularity condition (C.2) allows to expand the vector

$$nI(\boldsymbol{\theta}_0)\boldsymbol{Q} = \nabla\ell(\boldsymbol{\theta}_0;\boldsymbol{x}) + T\boldsymbol{Q} + \frac{1}{2}(\nabla^{(3)}\ell(\boldsymbol{\theta};\boldsymbol{x})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0^*})[\boldsymbol{Q},\boldsymbol{Q}]$$

and, therefore,

$$\begin{aligned}&\sqrt{n}[I(\boldsymbol{\theta}_0)]^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_n(\boldsymbol{x}) - \boldsymbol{\theta}_0)\\ &\quad= \frac{1}{\sqrt{n}}[I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}}\{\nabla(\ell(\boldsymbol{\theta}_0;\boldsymbol{x})) + \boldsymbol{Q}^\intercal(\nabla\nabla^\intercal(\ell(\boldsymbol{\theta}_0;\boldsymbol{x})) + nI(\boldsymbol{\theta}_0)\}\\ &\qquad+ \frac{1}{2\sqrt{n}}[I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}}\{(\nabla^{(3)}\ell(\boldsymbol{\theta};\boldsymbol{x})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0^*})[\boldsymbol{Q},\boldsymbol{Q}]\}.\end{aligned}$$

Using the score vector notation (2.2),

$$\sqrt{n}[I(\boldsymbol{\theta}_0)]^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) - \boldsymbol{\theta}_0) = [I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}}\begin{pmatrix}\boldsymbol{\xi}\\ \boldsymbol{\eta}\end{pmatrix} + \boldsymbol{R}_3 + \boldsymbol{R}_4, \tag{2.16}$$

where $\boldsymbol{R}_3$ and $\boldsymbol{R}_4$ are as in (2.15). Using (2.16) and that $I(\boldsymbol{\theta}_0)$ is a symmetric matrix leads to

$$\begin{aligned}&n(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) - \boldsymbol{\theta}_0)^\intercal I(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) - \boldsymbol{\theta}_0)\\ &\quad= \begin{pmatrix}\boldsymbol{\xi}\\ \boldsymbol{\eta}\end{pmatrix}^\intercal [I(\boldsymbol{\theta}_0)]^{-1}\begin{pmatrix}\boldsymbol{\xi}\\ \boldsymbol{\eta}\end{pmatrix} + (\boldsymbol{R}_3 + \boldsymbol{R}_4)^\intercal(\boldsymbol{R}_3 + \boldsymbol{R}_4)\\ &\qquad+ 2\begin{pmatrix}\boldsymbol{\xi}\\ \boldsymbol{\eta}\end{pmatrix}^\intercal [I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}}(\boldsymbol{R}_3 + \boldsymbol{R}_4).\end{aligned} \tag{2.17}$$

However, from (2.16),

$$\begin{pmatrix}\boldsymbol{\xi}\\ \boldsymbol{\eta}\end{pmatrix}^\intercal [I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}} = \sqrt{n}(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) - \boldsymbol{\theta}_0)^\intercal[I(\boldsymbol{\theta}_0)]^{\frac{1}{2}} - (\boldsymbol{R}_3 + \boldsymbol{R}_4)^\intercal,$$

so that

$$\begin{aligned}\begin{pmatrix}\boldsymbol{\xi}\\ \boldsymbol{\eta}\end{pmatrix}^\intercal [I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}}(\boldsymbol{R}_3 + \boldsymbol{R}_4) &= \sqrt{n}(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) - \boldsymbol{\theta}_0)^\intercal[I(\boldsymbol{\theta}_0)]^{\frac{1}{2}}(\boldsymbol{R}_3 + \boldsymbol{R}_4)\\ &\quad- (\boldsymbol{R}_3 + \boldsymbol{R}_4)^\intercal(\boldsymbol{R}_3 + \boldsymbol{R}_4).\end{aligned}$$

Using this in (2.17) yields the assertion. □

A similar result holds for T_2. Following exactly the same steps as for Lemma (2.7), but now with $\boldsymbol{\theta}^*$ instead of $\boldsymbol{\theta}_0$,

$$\begin{aligned}2\log T_2 &= \boldsymbol{\eta}^\intercal C^{-1}\boldsymbol{\eta} - R_1^* + R_2^* - (\boldsymbol{R}_3^* + \boldsymbol{R}_4^*)^\intercal(\boldsymbol{R}_3^* + \boldsymbol{R}_4^*)\\ &\quad+ 2\sqrt{n}(\hat{\boldsymbol{\theta}}_*(\boldsymbol{X}) - \boldsymbol{\theta}_*)^\intercal C^{\frac{1}{2}}(\boldsymbol{R}_3^* + \boldsymbol{R}_4^*).\end{aligned} \tag{2.18}$$

Combining (2.1), Lemma 2.6 and (2.18) with the notation (2.5.2) gives

$$-2\log\Lambda = G + R_{A_1} + R_{A_2} + R_{B_1} + R_{B_2}, \tag{2.19}$$

where

$$\begin{aligned}R_{A_1} &= R_1 - \boldsymbol{R}_3{}^\intercal\boldsymbol{R}_3 + 2\sqrt{n}(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) - \boldsymbol{\theta}_0)^\intercal[I(\boldsymbol{\theta}_0)]^{\frac{1}{2}}\boldsymbol{R}_3,\\ R_{A_2} &= -R_1{}^* + (\boldsymbol{R}_3^*)^\intercal\boldsymbol{R}_3^* - 2\sqrt{n}(\hat{\boldsymbol{\theta}}_*(\boldsymbol{X}) - \boldsymbol{\theta}_*)^\intercal C^{\frac{1}{2}}\boldsymbol{R}_3^*,\end{aligned}$$

$$
\begin{aligned}
R_{B_1} &= R_2 - \boldsymbol{R_4}^\intercal(\boldsymbol{R_3} + \boldsymbol{R_4}) - \boldsymbol{R_3}^\intercal \boldsymbol{R_4} \\
&\quad + 2\sqrt{n}\big(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) - \boldsymbol{\theta_0}\big)^\intercal \big[I(\boldsymbol{\theta_0})\big]^{\frac{1}{2}} \boldsymbol{R_4}, \\
R_{B_2} &= -R_2^* + \big(\boldsymbol{R_4^*}\big)^\intercal\big(\boldsymbol{R_3^*} + \boldsymbol{R_4}^*\big) + \big(\boldsymbol{R_3^*}\big)^\intercal \boldsymbol{R_4^*} \\
&\quad - 2\sqrt{n}\big(\hat{\boldsymbol{\theta}}_{\boldsymbol{*}}(\boldsymbol{X}) - \boldsymbol{\theta_*}\big)^\intercal C^{\frac{1}{2}} \boldsymbol{R_4^*}.
\end{aligned} \tag{2.20}
$$

Here, R_1, R_2 are as in Lemma 2.6 and R_1^*, R_2^* are their respective versions under the null hypothesis. Furthermore, $\boldsymbol{R_3}$ and $\boldsymbol{R_4}$ are as in Lemma 2.7, and $\boldsymbol{R_3}^*$ and $\boldsymbol{R_4}^*$ are the corresponding remainder terms from Lemma 2.7 with $\hat{\boldsymbol{\theta}}_n(\boldsymbol{x})$ replaced by $\hat{\boldsymbol{\theta}}^{\text{res}}(\boldsymbol{x})$. Note that R_{A_1} and R_{B_1} contain the terms that are obtained through $2\log T_1$, whereas R_{A_2} and R_{B_2} contain the quantities that are due to $2\log T_2$.

2.5.3. *Bounding the remainder terms*. In this section, we shall bound expectations of the log-likelihood ratio statistics under smooth test functions. Let $h \in \mathbb{C}_b^3(\mathbb{R})$ and $K \sim \chi_r^2$. Using the triangle inequality and (2.19),

$$
\begin{aligned}
&\big|\mathbb{E}\big[h(-2\log\Lambda)\big] - \mathbb{E}\big[h(K)\big]\big| \\
&\quad = \big|\mathbb{E}\big[h(G + R_{A_1} + R_{A_2} + R_{B_1} + R_{B_2})\big] - \mathbb{E}\big[h(K)\big]\big| \\
&\quad \le \big|\mathbb{E}\big[h(G + R_{A_1} + R_{A_2} + R_{B_1} + R_{B_2}) - h(G + R_{B_1} + R_{B_2})\big]\big| \\
&\qquad + \big|\mathbb{E}\big[h(G + R_{B_1} + R_{B_2}) - h(G)\big]\big| + \big|\mathbb{E}\big[h(G)\big] - \mathbb{E}\big[h(K)\big]\big|.
\end{aligned}
$$

The terms to bound are hence

$$
\big|\mathbb{E}\big[h(G + R_{A_1} + R_{A_2} + R_{B_1} + R_{B_2}) - h(G + R_{B_1} + R_{B_2})\big]\big| \tag{2.21}
$$

and

$$
\big|\mathbb{E}\big[h(G + R_{B_1} + R_{B_2}) - h(G)\big]\big| \tag{2.22}
$$

as well as

$$
\big|\mathbb{E}\big[h(G)\big] - \mathbb{E}\big[h(K)\big]\big|. \tag{2.23}
$$

The bound for $|\mathbb{E}[h(-2\log\Lambda)] - \mathbb{E}[h(K)]|$ is split into the above three terms in order to help in the understanding of this proof. The quantity in (2.23) will be bounded using Theorem 2.1, while we distinguish between (2.21) and (2.22) because the terms R_{A_1} and R_{A_2} are uniformly bounded, whereas R_{B_1} and R_{B_2} are not (some conditioning on the distance between the MLE and the value of the parameter will be needed in order to treat R_{B_1} and R_{B_2}). We now proceed to bound these three terms in (2.21), (2.22) and (2.23) in order to complete the proof of Theorem 2.3.

1. *Bounding term* (2.21). For some $t(\boldsymbol{X})$ between $G + R_{A_1} + R_{A_2} + R_{B_1} + R_{B_2}$ and $G + R_{B_1} + R_{B_2}$, a first- order Taylor expansion yields

$$
(2.21) = \big|\mathbb{E}\big[h'\big(t(\boldsymbol{X})\big)(R_{A_1} + R_{A_2})\big]\big| \le \|h'\|\mathbb{E}\big[|R_{A_1}| + |R_{A_2}|\big].
$$

We start by bounding $\mathbb{E}|R_{A_1}|$, where

$$
\mathbb{E}|R_{A_1}| \le \mathbb{E}|R_1| + \mathbb{E}\big|\boldsymbol{R_3^\intercal R_3}\big| + 2\sqrt{n}\mathbb{E}\big|\big(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) - \boldsymbol{\theta_0}\big)^\intercal\big[I(\boldsymbol{\theta_0})\big]^{\frac{1}{2}}\boldsymbol{R_3}\big|. \tag{2.24}
$$

With the notation in (2.10), as $R_1 = -\boldsymbol{Q}^\intercal T \boldsymbol{Q}$,

$$
\mathbb{E}|R_1| \le \sum_{j=1}^{d}\sum_{k=1}^{d} \mathbb{E}|Q_j Q_k T_{kj}|
$$

$$\leq \sum_{j=1}^{d}\sum_{k=1}^{d}[\mathbb{E}(Q_j^2Q_k^2)]^{\frac{1}{2}}[\mathbb{E}(T_{kj}^2)]^{\frac{1}{2}} \tag{2.25}$$

$$= \sqrt{n}\sum_{j=1}^{d}\sum_{k=1}^{d}[\mathbb{E}(Q_j^2Q_k^2)]^{\frac{1}{2}}\left[\mathrm{Var}\left(\frac{\partial^2}{\partial\theta_j\,\partial\theta_k}\log f(\boldsymbol{X_1}|\boldsymbol{\theta_0})\right)\right]^{\frac{1}{2}}.$$

Using Hölder's inequality twice,

$$\begin{aligned}
&\mathbb{E}|\boldsymbol{R_3}^\intercal\boldsymbol{R_3}| \\
&\leq \frac{1}{n}\sum_{l=1}^{d}\sum_{m=1}^{d}[[I(\boldsymbol{\theta_0})]^{-1}]_{lm}\sum_{j=1}^{d}\sum_{k=1}^{d}\mathbb{E}|Q_jQ_kT_{lj}T_{mk}| \\
&\leq \frac{1}{\sqrt{n}}\sum_{l=1}^{d}\sum_{m=1}^{d}[[I(\boldsymbol{\theta_0})]^{-1}]_{lm} \\
&\quad\times\sum_{j=1}^{d}\sum_{k=1}^{d}\sqrt{\mathrm{Var}\left(\frac{\partial^2}{\partial\theta_l\,\partial\theta_j}\log f(X_1|\boldsymbol{\theta_0})\right)}[\mathbb{E}(Q_j^6)\mathbb{E}(Q_k^6)\mathbb{E}(T_{mk}^6)]^{\frac{1}{6}}.
\end{aligned} \tag{2.26}$$

Moreover,

$$\begin{aligned}
&2\sqrt{n}\mathbb{E}|(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})-\boldsymbol{\theta_0})^\intercal[I(\boldsymbol{\theta_0})]^{\frac{1}{2}}\boldsymbol{R_3}| \\
&\leq 2\sum_{l=1}^{d}\sum_{j=1}^{d}\mathbb{E}|Q_lQ_jT_{lj}| \\
&\leq 2\sqrt{n}\sum_{l=1}^{d}\sum_{j=1}^{d}\sqrt{\mathbb{E}(Q_l^2Q_j^2)}\sqrt{\mathrm{Var}\left(\frac{\partial^2}{\partial\theta_l\,\partial\theta_j}\log f(X_1|\boldsymbol{\theta_0})\right)}.
\end{aligned} \tag{2.27}$$

Combining the results in (2.24), (2.25), (2.26) and (2.27) yields

$$\mathbb{E}|R_{A_1}| \leq \frac{1}{\sqrt{n}}K_1(\boldsymbol{\theta_0}),$$

with $K_1(\boldsymbol{\theta_0})$ as in (2.8). In order to bound $\mathbb{E}|R_{A_2}|$, we follow exactly the same process that was followed to bound $\mathbb{E}|R_{A_1}|$, but now under the null hypothesis, to conclude that $\mathbb{E}|R_{A_2}| \leq \frac{1}{\sqrt{n}}K_1^*(\boldsymbol{\theta_0})$ and, therefore,

$$(2.21) \leq \frac{\|h'\|}{\sqrt{n}}\big(K_1(\boldsymbol{\theta_0}) + K_1^*(\boldsymbol{\theta_0})\big). \tag{2.28}$$

2. *Bounding term* (2.22). The terms in R_{B_1} and R_{B_2} of (2.20) may not be uniformly bounded in $\boldsymbol{\theta}$. The triangle inequality leads to

$$(2.22) \leq \big|\mathbb{E}\big[h(G+R_{B_1}+R_{B_2}) - h(G+R_{B_2})\big]\big| \tag{2.29}$$

$$+\big|\mathbb{E}\big[h(G+R_{B_2}) - h(G)\big]\big|. \tag{2.30}$$

Bound for (2.29). Let $0 < \epsilon \leq \epsilon(\boldsymbol{\theta_0})$. With $Q_{(m)}$ as in (2.10), the law of total expectation, the Cauchy–Schwarz inequality and Markov's inequality yield

$$\begin{aligned}
(2.29) &\leq \mathbb{E}\big|h(G+R_{B_1}+R_{B_2}) - h(G+R_{B_2})\big| \\
&\leq 2\|h\|\mathbb{P}\big(|Q_m| \geq \epsilon\big)
\end{aligned}$$

$$+\mathbb{E}\big[\big|h(G+R_{B_1}+R_{B_2})-h(G+R_{B_2})\big|\big||Q_{(m)}|<\epsilon\big]\mathbb{P}\big(|Q_{(m)}|<\epsilon\big)$$
$$\le 2\frac{\|h\|}{\epsilon^2}\mathbb{E}\left(\sum_{j=1}^{d}Q_j^2\right)+\mathbb{E}\big[\big|h(G+R_{B_1}+R_{B_2})-h(G+R_{B_2})\big|\big||Q_{(m)}|<\epsilon\big].$$

A first-order Taylor expansion gives

$$\begin{aligned}&\mathbb{E}\big[\big|h(G+R_{B_1}+R_{B_2})-h(G+R_{B_2})\big|\big||Q_{(m)}|<\epsilon\big]\\&\quad\le \|h'\|\mathbb{E}\big[|R_{B_1}|\big||Q_{(m)}|<\epsilon\big]\\&\quad\le \|h'\|\mathbb{E}\big[\big(|R_2|+\big|\boldsymbol{R_4}^\intercal(\boldsymbol{R_3}+\boldsymbol{R_4})\big|+\big|\boldsymbol{R_3}^\intercal\boldsymbol{R_4}\big|\\&\qquad+2\sqrt{n}\big|(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})-\boldsymbol{\theta_0})^\intercal[I(\boldsymbol{\theta_0})]^{\frac{1}{2}}\boldsymbol{R_4}\big|\big)\big||Q_{(m)}|<\epsilon\big].\end{aligned}\tag{2.31}$$

From now on, we denote

$$\Delta_{qsl}:=\Delta_{qsl}(\boldsymbol{X},\boldsymbol{\theta_0})=\frac{\partial^3}{\partial\theta_q\,\partial\theta_s\,\partial\theta_l}\ell(\boldsymbol{\theta_0^*};\boldsymbol{X})\tag{2.32}$$

and we bound the terms in (2.31) in turns.

Bound for $\mathbb{E}|R_2|$. With R_2 as in Lemma 2.6, it is straightforward that for Δ_{qml} as in (2.32),

$$\begin{aligned}\mathbb{E}\big(|R_2|\big||Q_{(m)}|<\epsilon\big)&\le\frac{4}{3}\sum_{j=1}^{d}\sum_{k=1}^{d}\sum_{s=1}^{d}\mathbb{E}\big(|Q_jQ_kQ_s\Delta_{jks}|\big||Q_{(m)}|<\epsilon\big)\\&\le\frac{4}{3}\sum_{j=1}^{d}\sum_{k=1}^{d}\sum_{s=1}^{d}\sqrt{\mathbb{E}\big(Q_j^2Q_k^2Q_s^2\big)}\big[\mathbb{E}\big((M_{jks}(\boldsymbol{X}))^2\big||Q_{(m)}|<\epsilon\big)\big]^{\frac{1}{2}}.\end{aligned}$$

Bound for $\mathbb{E}(|\boldsymbol{R_4}^\intercal\boldsymbol{R_3}|\,||Q_{(m)}|<\epsilon)$ *and* $\mathbb{E}(|\boldsymbol{R_3}^\intercal\boldsymbol{R_4}|\,||Q_{(m)}|<\epsilon)$. With Δ_{qsl} as in (2.32) and T_{kj} as in (2.10), using Hölder's inequality and [1], Lemma 4.1, we obtain that

$$\begin{aligned}&\mathbb{E}\big(\big|\boldsymbol{R_4}^\intercal\boldsymbol{R_3}\big|\big||Q_{(m)}|<\epsilon\big)\\&\quad\le\frac{1}{2n}\sum_{q=1}^{d}\sum_{k=1}^{d}\big|\big[[I(\boldsymbol{\theta_0})]^{-1}\big]_{kq}\big|\\&\qquad\times\sum_{j=1}^{d}\sum_{l=1}^{d}\sum_{s=1}^{d}\mathbb{E}\big[|Q_jQ_lQ_sT_{kj}\Delta_{qsl}|\big||Q_{(m)}|<\epsilon\big]\\&\quad\le\frac{1}{2n}\sum_{q=1}^{d}\sum_{k=1}^{d}\big|\big[[I(\boldsymbol{\theta_0})]^{-1}\big]_{kq}\big|\\&\qquad\times\sum_{j=1}^{d}\sum_{l=1}^{d}\sum_{s=1}^{d}\sqrt{\mathbb{E}\big(Q_j^2Q_l^2Q_s^2\big)}\big[\mathbb{E}\big(T_{kj}^4\big||Q_{(m)}|<\epsilon\big)\big]^{\frac{1}{4}}\\&\qquad\times\big[\mathbb{E}\big((M_{qml}(\boldsymbol{X}))^4\big||Q_{(m)}|<\epsilon\big)\big]^{\frac{1}{4}}.\end{aligned}\tag{2.33}$$

Since $\boldsymbol{R_3}^\intercal\boldsymbol{R_4}=\boldsymbol{R_4}^\intercal\boldsymbol{R_3}$, $\mathbb{E}(|\boldsymbol{R_3}^\intercal\boldsymbol{R_4}|\,||Q_{(m)}|<\epsilon)$ can also be bounded by (2.33).

Bound for $\mathbb{E}(|\boldsymbol{R_4}^\intercal\boldsymbol{R_4}|\,||Q_{(m)}|<\epsilon)$. Again with [1], Lemma 4.1, and the Cauchy–Schwarz inequality, and with Δ_{qml} as in (2.32),

$$\begin{aligned}&\mathbb{E}\big(\big|\boldsymbol{R_4}^\intercal\boldsymbol{R_4}\big|\big||Q_{(m)}|<\epsilon\big)\\&\quad\le\frac{1}{4n}\sum_{b=1}^{d}\sum_{k=1}^{d}\sum_{s=1}^{d}\sum_{q=1}^{d}\sum_{l=1}^{d}\sum_{j=1}^{d}\big|\big[[I(\boldsymbol{\theta_0})]^{-1}\big]_{qb}\big|\end{aligned}$$

$$
\begin{aligned}
&\times \mathbb{E}\big[|Q_k Q_s Q_j Q_l \Delta_{bsk} \Delta_{qjl}|\big||Q_{(m)}| < \epsilon\big] \\
&\le \frac{1}{4n} \sum_{b=1}^{d} \sum_{k=1}^{d} \sum_{s=1}^{d} \sum_{q=1}^{d} \sum_{l=1}^{d} \sum_{j=1}^{d} \big|\big[[I(\boldsymbol{\theta_0})]^{-1}\big]_{qb}\big| \\
&\times \sqrt{\mathbb{E}\big(Q_k^2 Q_s^2 Q_j^2 Q_l^2\big)} \\
&\times \big[\mathbb{E}\big(\big(M_{bsk}(\boldsymbol{X})\big)^4 \big| |Q_{(m)}| < \epsilon\big)\big]^{\frac{1}{4}} \big[\mathbb{E}\big(\big(M_{qjl}(\boldsymbol{X})\big)^4 \big| |Q_{(m)}| < \epsilon\big)\big]^{\frac{1}{4}}.
\end{aligned} \tag{2.34}
$$

Bound for $\mathbb{E}(2\sqrt{n}|(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) - \boldsymbol{\theta_0})^\intercal [I(\boldsymbol{\theta_0})]^{\frac{1}{2}} \boldsymbol{R_4}|\,||Q_{(m)}| < \epsilon)$. A similar process as the one to obtain the bounds in (2.33) and (2.34) yields

$$
\begin{aligned}
&\mathbb{E}\big(2\sqrt{n}\big|(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) - \boldsymbol{\theta_0})^\intercal [I(\boldsymbol{\theta_0})]^{\frac{1}{2}} \boldsymbol{R_4}\big|\big||Q_{(m)}| < \epsilon\big) \\
&\le \sum_{l=1}^{d} \sum_{j=1}^{d} \sum_{q=1}^{d} \mathbb{E}\big(|Q_l Q_j Q_q \Delta_{ljq}|\big||Q_{(m)}| < \epsilon\big) \\
&\le \sum_{l=1}^{d} \sum_{j=1}^{d} \sum_{q=1}^{d} \sqrt{\mathbb{E}\big(Q_l^2 Q_j^2 Q_q^2\big)} \big[\mathbb{E}\big(\big(M_{ljq}(\boldsymbol{X})\big)^2 \big| |Q_{(m)}| < \epsilon\big)\big]^{\frac{1}{2}}.
\end{aligned} \tag{2.35}
$$

Combining (2.29), (2.31), (2.33), (2.34) and (2.35),

$$
(2.29) \le \frac{1}{\sqrt{n}} K_2(\boldsymbol{\theta_0}), \tag{2.36}
$$

with $K_2(\boldsymbol{\theta_0})$ as in (2.9).

Bound for (2.30). With $Q^*_{(m)}$ as in (2.10), the law of total expectation, the Cauchy–Schwarz inequality and Markov's inequality yield

$$
\begin{aligned}
(2.30) &\le \mathbb{E}\big|h(G + R_{B_2}) - h(G)\big| \\
&\le 2\|h\| \mathbb{P}\big(\big|Q^*_m\big| \ge \epsilon\big) \\
&\quad + \mathbb{E}\big[\big|h(G + R_{B_2}) - h(G)\big|\big|\big|Q^*_{(m)}\big| < \epsilon\big] \mathbb{P}\big(\big|Q^*_{(m)}\big| < \epsilon\big).
\end{aligned}
$$

Finding an upper bound for this expression follows the same arguments as the one to bound (2.29) and, therefore, it will not be repeated; the result is

$$
(2.30) \le \frac{1}{\sqrt{n}} K_2^*(\boldsymbol{\theta_0}), \tag{2.37}
$$

where $K_2^*(\boldsymbol{\theta_0})$ is the version of $K_2(\boldsymbol{\theta_0})$ under the null hypothesis.

3. *Bounding term* (2.23). With $Z_{i,j}$ as in (2.7), and the notation of Theorems 2.1 and 2.2,

$$
\begin{aligned}
\mathbb{E}[\mathbf{Z}^\intercal \mathbf{Z}] &= \mathbb{E}\big[(\boldsymbol{\xi} - BC^{-1}\boldsymbol{\eta})^\intercal (\boldsymbol{\xi} - BC^{-1}\boldsymbol{\eta})\big] \\
&= A - BC^{-1}B^\intercal.
\end{aligned}
$$

Thus $Z_{i,j}$, $i = 1, \dots, r$, $j = 1, \dots, n$ satisfy the assumptions of Theorem 2.1 with

$$
\tau = A - BC^{-1}B^\intercal, \tag{2.38}
$$

which is assumed to be positive definite; hence its inverse U exists. Applying Theorem 2.1 gives the bound

$$
\big|\mathbb{E}\big[h(G)\big] - \mathbb{E}\big[h(K)\big]\big| \le \frac{16 |||h|||_3}{r\sqrt{n}} R(r) \tag{2.39}
$$

with $R(r)$ given in (2.5). The results in (2.39), (2.28), (2.36) and (2.37) conclude the proof of Theorem 2.3. □

The next section gives three examples to illustrate the approach. First, we consider an example with a one-dimensional parameter, namely the exponential distribution. The second example is that of the normal distribution with two-dimensional parameter (μ, σ^2). The last example is logistic regression.

3. Examples.

3.1. *Single-parameter-case example*: *The exponential distribution.* Here, we apply Theorem 2.3 in an example from a single-parameter distribution. We highlight that in the single-parameter case the interest is on assessing the asymptotic χ_1^2 distribution of $2(l(\hat{\theta}_n(\boldsymbol{X}); \boldsymbol{X}) - l(\theta_0; \boldsymbol{X}))$, where θ_0 is the true value of the unknown parameter θ. The log-likelihood ratio in (2.1) reduces to $-2\log\Lambda = 2\log T_1$, so that there is no need to introduce T_2 as defined in (2.1) and the terms $K_1^*(\theta_0)$ and $K_2^*(\theta_0)$ in the expression of (2.6) vanish.

To illustrate the single-parameter case, we consider an example from the exponential distribution with mean θ_0. For $X \sim \mathrm{Exp}(\frac{1}{\theta})$, $\theta > 0$ the p.d.f. is $f(x|\theta) = \frac{1}{\theta}\exp\{-\frac{x}{\theta}\}$, for $x > 0$.

COROLLARY 3.1. *Let $X_1, X_2, \dots, X_n$ be i.i.d. random variables that follow the* $\mathrm{Exp}(\frac{1}{\theta_0})$ *distribution. The MLE exists, it is unique, equal to $\hat{\theta}_n(\boldsymbol{X}) = \bar{X}$ and the regularity conditions* (*C*.1)–(*C*.7) *are satisfied. For $h \in \mathbb{C}_b^3(\mathbb{R})$ and $K \sim \chi_1^2$,*

$$
\begin{aligned}
&\big|\mathbb{E}\big[h\big(2\big(\ell\big(\hat{\theta}_n(\boldsymbol{X}); \boldsymbol{X}\big) - \ell(\theta_0; \boldsymbol{X})\big)\big)\big] - \mathbb{E}\big[h(K)\big]\big| \qquad (3.1)\\
&\quad \le 8\frac{\|h\|}{n} + \frac{16\|h\|_3}{\sqrt{n}}\left(\frac{17}{2\sqrt{n}} + \sqrt{\frac{4n}{n-1} + \frac{9}{n}}\left(\frac{7}{2} + \frac{1}{n}\right)\right.\\
&\qquad \left. + \frac{\sqrt{864n^2 + 10{,}472n + 14{,}833}}{n^{3/2}}\left(1 + \sqrt{9n^{-\frac{1}{2}} + 3}\right)\right)\\
&\qquad + \frac{\|h'\|}{\sqrt{n}}\left\{6\sqrt{3 + \frac{6}{n}} + \sqrt{15 + \frac{130}{n} + \frac{120}{n^2}}\left(\frac{1120}{3} + \frac{320(3 + \frac{6}{n})^{\frac{1}{4}} + 4}{\sqrt{n}}\right)\right. \qquad (3.2)\\
&\qquad \left. + \frac{6400}{\sqrt{n}}\sqrt{105 + \frac{2380}{n} + \frac{7308}{n^2} + \frac{5040}{n^3}}\right\}.
\end{aligned}
$$

REMARK 3.2. (1) The upper bound in (3.1) is $\mathcal{O}(\frac{1}{\sqrt{n}})$.

(2) The bound does not depend on the parameter θ_0.

PROOF. It is easy to check that the assumptions of Theorem 2.3 hold. Here, we choose $\epsilon(\theta_0) = \frac{1}{2}\theta_0$ and with $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$,

$$
M_{i,j,k}(\boldsymbol{x}) = M_{1,1,1}(\boldsymbol{x}) = \frac{96}{\theta_0^4}\sum_{i=1}^n x_i + \frac{16}{\theta_0^3} = \frac{96n}{\theta^4}\left(3\bar{x} + \frac{1}{2}\theta_0\right).
$$

In addition, straightforward calculations lead to $\hat{\theta}_n(\boldsymbol{X}) = \bar{X}$. The expected Fisher information number for one random variable is $I(\theta_0) = \frac{1}{\theta_0^2}$. We start with the calculation of the first term of the bound in (2.6). With $d = 1$ and $r = 1$, the $\boldsymbol{W} = (\boldsymbol{\xi}, \boldsymbol{\eta})$ reduces to $\xi = \frac{1}{\sqrt{n}}\sum_{j=1}^n Y_j$, where $Y_j(\theta_0) = \frac{\mathrm{d}}{\mathrm{d}\theta} \log f(X_j|\theta_0) = \frac{X_j - \theta_0}{\theta_0^2}$. In addition, since $I(\theta_0)$ is now a scalar and equal to $\frac{1}{\theta_0^2}$, the matrix τ as in (2.38) is also equal to $\frac{1}{\theta_0^2}$ which means that $U = \tau^{-1} = \theta_0^2$. In addition, $D = BC^{-1} = 0$ in this example and $Z_{i,j} = Z_j = Y_j(\theta_0)$. Therefore, in our example $W = UZ_1^2 = \theta_0^2 Z_1^2$ and the aim is to bound $R(r)$ as in (2.5). With

$$\mathbb{E}(W^2) = \theta_0^4 \mathbb{E}(Z_1^4) = \frac{1}{\theta_0^4}\mathbb{E}(X_1 - \theta_0)^4 = 9,$$

$$\mathbb{E}(W^3) = \theta_0^6 \mathbb{E}(Z_1^6) = \frac{1}{\theta_0^6}\mathbb{E}(X_1 - \theta_0)^6 = 265,$$

$$\mathbb{E}(W^4) = \theta_0^8 \mathbb{E}(Z_1^8) = \frac{1}{\theta_0^8}\mathbb{E}(X_1 - \theta_0)^8 = 14{,}833,$$

then in this example

$$\begin{aligned} B_1(r,n) &= \frac{4n}{n-1} + \frac{9}{n}, \\ B_2(r,n) &= 864 + \frac{1}{n}\left(\frac{14{,}833}{n} + 6360 + \theta_0^8\big(32\mathbb{E}(Z_1^5)\mathbb{E}(Z_1^3) + 16\mathbb{E}^2(Z_1^4)\big)\right) \\ &= 864 + \frac{1}{n}\left(\frac{14{,}833}{n} + 6360 + 2816 + 1296\right) \\ &= 864 + \frac{1}{n}\left(\frac{14{,}833}{n} + 10{,}472\right). \end{aligned}$$

In addition, simple steps lead to

$$\sqrt{\sum_{i,k,a,b,,e,f=1}^{r} U_{i,k}U_{a,b}U_{e,f}\beta(i,a,b)\beta(k,e,f)} = 2.$$

Therefore,

$$\begin{aligned} R(r) = {} & \frac{4}{\sqrt{n}} + \frac{9}{2\sqrt{n}} + \sqrt{\frac{4n}{n-1} + \frac{9}{n}}\left(\frac{7}{2} + \frac{1}{n}\right) \\ & + \frac{\sqrt{864n^2 + 10{,}472n + 14{,}833}}{n^{3/2}}\big(1 + \sqrt{9n^{-\frac{1}{2}} + 3}\big). \end{aligned} \tag{3.3}$$

The next task is to bound $K_1(\theta_0)$ as in (2.8), for $d = 1$. Using the definition of Q_j in (2.10), $Q_1 = \bar{X} - \theta_0$. The moments of Q_1 are calculated using standard results from [13] along with the fact that $\bar{X} \sim G(n, \frac{n}{\theta_0})$, giving

$$3n\sqrt{\mathbb{E}(Q_1)^4}\left[\mathrm{Var}\left(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(X_1|\theta_0)\right)\right]^{\frac{1}{2}} = 6\sqrt{3 + \frac{6}{n}}. \tag{3.4}$$

For the second quantity in (2.8), with the definition of T_{11} in (2.10),

$$\begin{aligned} & \frac{1}{i(\theta_0)}\left[\mathrm{Var}\left(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(X_1|\theta_0)\right)\right]^{\frac{1}{2}}[\mathbb{E}(Q^6)]^{\frac{1}{3}}[\mathbb{E}(T_{11}^6)]^{\frac{1}{6}} \\ & = 2[\mathbb{E}(\bar{X} - \theta_0)^6]^{\frac{1}{3}}\left[\mathbb{E}\left(-\frac{2n\bar{X}}{\theta_0^3} + \frac{2n}{\theta_0^2}\right)^6\right]^{\frac{1}{6}} \\ & = \frac{4n}{\theta_0^3}\sqrt{\mathbb{E}(\bar{X} - \theta_0)^6} = \frac{4}{\sqrt{n}}\sqrt{15 + \frac{130}{n} + \frac{120}{n^2}}. \end{aligned} \tag{3.5}$$

Combining (3.4) and (3.5),

$$K_1(\theta_0) = 6\sqrt{3+\frac{6}{n}} + \frac{4}{\sqrt{n}}\sqrt{15+\frac{130}{n}+\frac{120}{n^2}}. \tag{3.6}$$

We proceed to find a bound for $K_2(\boldsymbol{\theta_0})$, as defined in (2.9). The calculation of the first term is straightforward:

$$2\sqrt{n}\frac{\|h\|}{\epsilon^2}\mathbb{E}(\bar{X}-\theta_0)^2 = \frac{2\|h\|\theta_0^2}{\sqrt{n}\epsilon^2}.$$

The second term of (2.9) requires the calculation of conditional expectations related to $M_{111}(\boldsymbol{X})$. For $\epsilon = \frac{1}{2}\theta_0$,

$$\begin{aligned}
&\sqrt{n}\|h'\|\frac{7}{3}\sqrt{\mathbb{E}(Q_1^6)}\big[\mathbb{E}\big[(M_{111}(\boldsymbol{X}))^2 \big| |Q_1| < \epsilon\big]\big]^{\frac{1}{2}}\\
&\quad = \sqrt{n}\|h'\|\frac{7}{3}\sqrt{\mathbb{E}(\bar{X}-\theta_0)^6}\bigg[\mathbb{E}\bigg[\frac{96^2n^2}{\theta^8}\bigg(3\bar{X}+\frac{1}{2}\theta_0\bigg)^2\bigg| |\bar{X}-\theta_0| < \frac{1}{2}\theta_0\bigg]\bigg]^{\frac{1}{2}}\\
&\quad \le \frac{448\|h'\|\theta_0^3}{3\theta_0^4}\sqrt{15+\frac{130}{n}+\frac{120}{n^2}}\bigg(2\theta_0+\frac{1}{2}\theta_0\bigg)\\
&\quad = \frac{1120}{3}\|h'\|\sqrt{15+\frac{130}{n}+\frac{120}{n^2}}.
\end{aligned}$$

Bounding the third term of (2.9) requires the calculation of conditional expectations related to T_{11} of (2.10) and $M_{111}(\boldsymbol{X})$. It is easy to see that T_{11} can be written as a continuous, increasing function of Q_1. Therefore, employing Lemma 2.1 of [2], leads to

$$\begin{aligned}
&\frac{\|h'\|}{\sqrt{n}}\frac{1}{i(\theta_0)}\sqrt{\mathbb{E}(Q_1^6)}\bigg[\mathbb{E}\bigg(T_{11}^4 \big| |Q_1| < \frac{1}{2}\theta_0\bigg)\bigg]^{\frac{1}{4}}\bigg[\mathbb{E}\bigg((M_{111}(\boldsymbol{X}))^4 \big| |Q_1| < \frac{1}{2}\theta_0\bigg)\bigg]^{\frac{1}{4}}\\
&\quad \le \frac{128n^{\frac{3}{2}}\|h'\|}{\theta_0^5}\sqrt{\frac{\theta_0^6}{n^3}\bigg(15+\frac{130}{n}+\frac{120}{n^2}\bigg)}\big[\mathbb{E}(\bar{X}-\theta_0)^4\big]^{\frac{1}{4}}\frac{5}{2}\theta_0.
\end{aligned}$$

An upper bound for the fourth term of (2.9) is found in a similar way. Collecting these bounds give

$$\begin{aligned}
K_2(\theta_0) \le{}& \frac{8\|h\|}{\sqrt{n}} + \|h'\|\sqrt{15+\frac{130}{n}+\frac{120}{n^2}}\bigg(\frac{1120}{3}+\frac{320(3)^{\frac{1}{4}}}{\sqrt{n}}\bigg(\frac{2}{n}+1\bigg)^{\frac{1}{4}}\bigg)\\
&+\frac{6400\|h'\|}{\sqrt{n}}\sqrt{105+\frac{2380}{n}+\frac{7308}{n^2}+\frac{5040}{n^3}}.
\end{aligned} \tag{3.7}$$

Combining the results in (3.3), (3.6) and (3.7) yields the assertion. □

REMARK 3.3. We chose $\epsilon(\theta_0)$ to be the mid-point of the interval $(0, \theta_0)$ as there is a trade off on its choice for $K_2(\boldsymbol{\theta_0})$. A more systematic choice of $\epsilon(\theta_0)$ based on numerical solutions of inequalities could be of interest in principle. As our bounds are not optimised with respect to the constants, for space reasons this systematic choice is not carried out.

Empirical results. Here, we study the accuracy of our bounds by simulations. We start by generating 100 trials of n random independent observations, x, from $\mathrm{Exp}(\frac{1}{\theta})$, where $n = 1000 \times j$, $j = 1, 2, \dots, 1000$ and $\theta = 3$. We evaluate the MLE, $\hat{\theta}_n(\boldsymbol{X})$ of the parameter in

TABLE 1
Simulation results for the exponential distribution example

n	$Q_{h_t}(\theta_0)$	Upper bound	Error
10^4	0.009	5.934	5.925
10^5	0.007	1.516	1.509
10^6	0.001	0.443	0.442

each trial and then the log-likelihood ratio statistic, $2(l(\hat{\theta}_n(\boldsymbol{X}); \boldsymbol{X}) - l(\theta_0; \boldsymbol{X}))$, which in turn gives a vector of 100 values. It is easy to show that for this example

$$2\big(l\big(\hat{\theta}_n(\boldsymbol{X}); \boldsymbol{X}\big) - l(\theta_0; \boldsymbol{X})\big) = 2n\bigg(\log\bigg(\frac{\theta_0}{\bar{X}}\bigg) + \frac{\bar{X}}{\theta_0} - 1\bigg).$$

We apply to these values the function $h_t(x) = (x^2+2)^{-1}$ and we calculate their sample mean, denoted by $\hat{\mathbb{E}}[h_t(2(l(\hat{\theta}_n(\boldsymbol{X}); \boldsymbol{X}) - l(\theta_0; \boldsymbol{X})))]$. The function h_t is a member of the class $C_b^3(\mathbb{R})$ with

$$\|h_t\| = 0.5, \qquad \|h_t'\| = \frac{3\sqrt{1.5}}{16},$$
$$\|h_t''\| = 0.5, \qquad \|h_t\|_3 = \frac{15}{128}\sqrt{25 + 11\sqrt{5}}.$$

We use these values to calculate the bound in (3.1). We define

$$Q_{h_t}(\theta_0) := \big|\hat{\mathbb{E}}\big[h_t\big(2\big(l\big(\hat{\theta}_n(\boldsymbol{X}); \boldsymbol{X}\big) - l(\theta_0; \boldsymbol{X})\big)\big)\big] - \tilde{\mathbb{E}}\big[h_t(K)\big]\big|, \tag{3.8}$$

where $\tilde{\mathbb{E}}[h_t(K)] = 0.373$ is the approximation of $\mathbb{E}[h_t(K)]$ up to three decimal places, where $K \sim \chi_1^2$. We compare $Q_{h_t}(\theta_0)$ with the bound in (3.1), using the difference between their values as a measure of the error. The results for $n = 10^j$, $j = 4, 5, 6$ are presented in Table 1 and are based on this particular function h_t, while our theoretical bounds hold for any test function that belongs in the class $C_b^3(\mathbb{R})$.

The table indicates that both the bound and the error decrease as the sample size gets larger. When at each step we increase the sample size by a factor of ten, then the value of the upper bound drops by approximately a $\sqrt{10}$ factor, which is expected as the expression in (3.1) is $\mathcal{O}(n^{-1/2})$. It is instructive to examine the contributions of each term to the bound. For example, when $n = 10^5$, where the bound is equal to 1.516,

$$\frac{16\|h_t\|_3}{r\sqrt{n}}R(r) = 0.304, \qquad \frac{\|h_t'\|}{\sqrt{n}}K_1(\theta_0) = 0.008, \qquad \frac{1}{\sqrt{n}}K_2(\theta_0) = 1.204.$$

We see that the bound is mostly dependent on the quantity related to $K_2(\theta_0)$ due to the large and nonoptimised constants in (3.7).

3.2. *Example*: *The normal distribution*. Here, we apply Theorem 2.3 in the case of $X_1, X_2, \dots, X_n$ i.i.d. random variables from $\mathrm{N}(\mu, \sigma^2)$ with $\boldsymbol{\theta} = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+$. We consider the test problem $H_0 : \mu = 0$ against the general alternative. It is well known that under the alternative, the MLE is equal to $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) = (\hat{\mu}, \hat{\sigma}^2)^\intercal = (\bar{X}, \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2)^\intercal$; see, for example, [7], page 116. Under the null, simple calculations show that the MLE for σ^2 is $\hat{\theta}_*(\boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^n X_i^2$. In addition, the regularity conditions are satisfied.

COROLLARY 3.4. *Let $X_1, X_2, \dots, X_n$ be i.i.d. random variables that follow the $N(\mu, \sigma^2)$ distribution. For the likelihood ratio test $H_0 : \mu = 0$ against the general alternative, for $h \in \mathbb{C}_b^3(\mathbb{R})$ and $K \sim \chi_1^2$, it holds that*

$$
\begin{aligned}
&\left|\mathbb{E}\left[h(-2\log\Lambda)\right] - \mathbb{E}\left[h(K)\right]\right| \\
&\quad \le \frac{\|h\|}{n}\left(\frac{8}{\sigma^2} + 32\right) \\
&\qquad + \frac{16\|h\|_3}{\sqrt{n}}\left(\frac{11}{2\sqrt{n}} + \sqrt{\frac{4n}{n-1} + \frac{3}{n}}\left(\frac{1}{2} + \frac{1}{n}\right)\right. \\
&\qquad \left. + \frac{\sqrt{288n^2 + 504n + 105}}{n^{3/2}}\left(\frac{1}{4}(1+\sqrt{3}) + \sqrt{3\left(n^{-\frac{1}{2}} + 1\right)}\right)\right) \\
&\qquad + \frac{\|h'\|}{\sqrt{n}}\left(16{,}086 + \frac{13{,}527{,}826}{\sqrt{n}} + 448\sqrt{\frac{2}{n} + \frac{\sigma^2}{2}}\right. \\
&\qquad + 860\sqrt{1 + 648\left(\left(\frac{3}{2} + \frac{\sigma^2}{4}\right)^2 + \frac{3}{n^2}\right)} + \frac{7416}{\sqrt{n}}\left(\frac{3}{n^2} + \frac{\sigma^4}{16}\right)^{\frac{1}{4}} \\
&\qquad + \frac{1}{\sqrt{n}}\left(1 + 839{,}808\left(\left(\frac{3}{2} + \frac{\sigma^2}{4}\right)^4 + \frac{105}{n^4}\right)\right)^{\frac{1}{4}} \\
&\qquad \times \left(21{,}984 + 23{,}616\left(\frac{3}{n^2} + \frac{\sigma^4}{16}\right)^{\frac{1}{4}}\right) \\
&\qquad + \frac{87{,}104}{\sqrt{n}}\sqrt{\frac{3}{n^2} + \frac{\sigma^4}{16}} + \frac{38{,}512}{\sqrt{n}} \\
&\qquad \left. \times \sqrt{1 + 839{,}808\left(\left(\frac{3}{2} + \frac{\sigma^2}{4}\right)^4 + \frac{105}{n^4}\right)}\right).
\end{aligned} \tag{3.9}
$$

REMARK 3.5. (1) For fixed σ^2, the upper bound in Corollary 3.4 is of order $\frac{1}{\sqrt{n}}$. There is no claim that the constants are optimal.

(2) The normal bound is only small when σ^2 is neither too large nor too small, so that $n^{-1} \ll \sigma^2 \ll n^{1/2}$.

PROOF. We will use the result of Theorem 2.3. In this case $d = 2$ and $r = 1$. The expected Fisher information matrix for one random variable is

$$
I(\boldsymbol{\theta_0}) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}, \quad \text{so that} \quad \left[I(\boldsymbol{\theta_0})\right]^{-1} = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix}. \tag{3.10}
$$

The assumptions (C.1)–(C.7) are verified for $\epsilon(\boldsymbol{\theta_0}) < \infty$ and

$$
\sup_{\boldsymbol{\theta}:|\theta_m - \theta_{0,m}|<\epsilon}\left|\frac{\partial^3}{\partial\theta_1^3}\ell(\boldsymbol{\theta}; \boldsymbol{X})\right| = 0 =: M_{111}(\boldsymbol{X})
$$

as well as

$$
\begin{aligned}
&\sup_{\boldsymbol{\theta}:|\theta_m-\theta_{0,m}|<\epsilon}\left|\frac{\partial^3}{\partial\theta_2^3}\ell(\boldsymbol{\theta};\boldsymbol{X})\right| \\
&\quad=\sup_{\boldsymbol{\theta}:|\theta_m-\theta_{0,m}|<\epsilon}\left|-\frac{n}{\theta_2^3}+\frac{3}{\theta_2^4}\sum_{i=1}^{n}(X_i-\theta_1)^2\right| \\
&\quad\le\frac{n}{(\sigma^2-\epsilon)^3}+\frac{9n}{(\sigma^2-\epsilon)^4}\big(\hat{\sigma^2}+(\bar{X}-\mu)^2+\epsilon^2\big)=:M_{222}(\boldsymbol{X}).
\end{aligned}
\tag{3.11}
$$

Moreover,

$$
\begin{aligned}
&\sup_{\boldsymbol{\theta}:|\theta_m-\theta_{0,m}|<\epsilon}\left|\frac{\partial^3}{\partial\theta_1\,\partial\theta_2^2}\ell(\boldsymbol{\theta};\boldsymbol{X})\right| \\
&\quad=\sup_{\boldsymbol{\theta}:|\theta_m-\theta_{0,m}|<\epsilon}\left|\frac{\partial^3}{\partial\theta_2^2\,\partial\theta_1}\ell(\boldsymbol{\theta};\boldsymbol{X})\right| \\
&\quad\le\frac{2n}{(\sigma^2-\epsilon)^3}\big(|\bar{X}-\mu|+\epsilon\big)=:M_{122}(\boldsymbol{X})
\end{aligned}
\tag{3.12}
$$

and

$$
\begin{aligned}
&\sup_{\boldsymbol{\theta}:|\theta_m-\theta_{0,m}|\le\epsilon}\left|\frac{\partial^3}{\partial\theta_1^2\,\partial\theta_2}\ell(\boldsymbol{\theta};\boldsymbol{X})\right| \\
&\quad=\sup_{\boldsymbol{\theta}:|\theta_m-\theta_{0,m}|<\epsilon}\left|\frac{\partial^3}{\partial\theta_2\,\partial\theta_1^2}\ell(\boldsymbol{\theta};\boldsymbol{X})\right| \\
&\quad=\sup_{\boldsymbol{\theta}:|\theta_m-\theta_{0,m}|<\epsilon}\left|\frac{n}{\theta_2^2}\right|\le\frac{n}{(\sigma^2-\epsilon)^2}=:M_{112}(\boldsymbol{X}).
\end{aligned}
\tag{3.13}
$$

We start with the calculation of the first term of the bound in (2.6). From (3.10), we have that $A=\frac{1}{\sigma^2}$, $B=0$, $C=\frac{1}{2\sigma^4}$, so that $\tau=\frac{1}{\sigma^2}$ and, therefore, $U=\sigma^2$ and $D=BC^{-1}=0$. Therefore, in our example, for $j=1,2,\dots,n$ we have that $Z_{1,j}=Z_j=Y_{1,j}(\boldsymbol{\theta_0})=\frac{\partial}{\partial\theta_1}\log f(X_j|\boldsymbol{\theta_0})=\frac{X_j-\mu}{\sigma^2}$. Now, $W=UZ_1^2=\frac{(X_1-\mu)^2}{\sigma^2}$ and

$$
\begin{aligned}
\mathbb{E}(W^2)&=\frac{1}{\sigma^4}\mathbb{E}\big((X_1-\mu)^4\big)=3,\\
\mathbb{E}(W^3)&=\frac{1}{\sigma^6}\mathbb{E}\big((X_1-\mu)^6\big)=15,\\
\mathbb{E}(W^4)&=\frac{1}{\sigma^8}\mathbb{E}\big((X_1-\mu)^8\big)=105.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
B_1(r,n)&=\frac{4n}{n-1}+\frac{3}{n},\\
B_2(r,n)&=288+\frac{1}{n}\left(\frac{105}{n}+360+32\sigma^8\mathbb{E}(Z_1^5)\mathbb{E}(Z_1^3)+16\sigma^8\mathbb{E}^2(Z_1^4)\right)\\
&=288+\frac{1}{n}\left(\frac{105}{n}+504\right).
\end{aligned}
$$

In addition, because $\mathbb{E}(Z_1^3) = \frac{1}{\sigma^6}\mathbb{E}[(X_1-\mu)^3] = 0$,

$$\sqrt{\sum_{i,k,a,b,,e,f=1}^{r} U_{i,k}U_{a,b}U_{e,f}\beta(i,a,b)\beta(k,e,f)} = 0$$

and, therefore,

$$\begin{aligned} R(r) &= \frac{4}{\sqrt{n}} + \frac{3}{2\sqrt{n}} + \sqrt{\frac{4n}{n-1} + \frac{3}{n}}\left(\frac{1}{2} + \frac{1}{n}\right) \\ &\quad + \frac{\sqrt{288n^2 + 504n + 105}}{n^{3/2}}\left(\frac{1}{4}(1+\sqrt{3}) + \sqrt{3(n^{-\frac{1}{2}} + 1)}\right), \end{aligned} \tag{3.14}$$

which is then used to obtain an upper bound for the first term in the general expression of (2.6). We now proceed to bound $K_1(\boldsymbol{\theta_0})$ in (2.8). In regards to the first quantity, the expressions for the partial derivatives of the log-likelihood and the fact that in the case of i.i.d. random variables from the normal distribution, $\bar{X}$ and $\hat{\sigma}^2$ are independent random variables ([3], page 218) lead to

$$\begin{aligned} &3n\sum_{j=1}^{2}\sum_{k=1}^{2}[\mathbb{E}(Q_j^2 Q_k^2)]^{\frac{1}{2}}\left[\mathrm{Var}\left(\frac{\partial^2}{\partial\theta_j\,\partial\theta_k}\log f(X_1|\boldsymbol{\theta_0})\right)\right]^{\frac{1}{2}} \\ &\quad = 3n\sqrt{\mathbb{E}(\hat{\sigma}^2 - \sigma^2)^4}\sqrt{\mathrm{Var}\left(\frac{1}{\sigma^6}(X_1-\mu)^2\right)} \\ &\qquad + 6n\sqrt{\mathbb{E}(\bar{X}-\mu)^2\mathbb{E}(\hat{\sigma}^2-\sigma^2)^2}\sqrt{\mathrm{Var}\left(\frac{1}{\sigma^4}(X_1-\mu)\right)} \\ &\quad \le 3n\sqrt{\frac{16\sigma^8}{n^2}}\sqrt{\frac{2}{\sigma^8}} + 6n\sqrt{2\frac{\sigma^6}{n^2}}\sqrt{\frac{1}{\sigma^6}} = 18\sqrt{2}. \end{aligned} \tag{3.15}$$

For the second quantity in (2.8), let $G_\kappa \sim \chi^2_\kappa$; then

$$\begin{aligned} \mathbb{E}(Q_1^6) &= \mathbb{E}(\bar{X}-\mu)^6 = 15\frac{\sigma^6}{n^3}, \\ \mathbb{E}(Q_2^6) &= \mathbb{E}(\hat{\sigma}^2-\sigma^2)^6 = \frac{\sigma^{12}}{n^6}\mathbb{E}(G_{n-1}-n)^6 \\ &= \frac{\sigma^{12}}{n^3}\left(120 + \frac{940}{n} - \frac{114}{n^2} - \frac{945}{n^3}\right) \le \frac{1060}{n^3}\sigma^{12}. \end{aligned} \tag{3.16}$$

Using now (3.10),

$$\mathbb{E}(T_{11}^6) = 0, \quad \mathbb{E}(T_{12}^6) = \mathbb{E}(T_{21}^6) = \frac{1}{\sigma^{18}}\mathbb{E}\left(\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)\right)^6 = \frac{15n^3}{\sigma^{18}},$$

$$\mathbb{E}(T_{22}^6) = \frac{1}{\sigma^{24}}\mathbb{E}(G_n - n)^6 = \frac{40n^3}{\sigma^{24}}\left(3 + \frac{52}{n} + \frac{96}{n^2}\right) \le \frac{6040n^3}{\sigma^{24}}.$$

With inequalities (3.10) and (3.16), this yields

$$\sum_{l=1}^{2}\sum_{m=1}^{2}[[I(\boldsymbol{\theta_0})]^{-1}]_{lm}\sum_{j=1}^{2}\sum_{k=1}^{2}\sqrt{\mathrm{Var}\left(\frac{\partial^2}{\partial\theta_l\,\partial\theta_j}\log f(X_1|\boldsymbol{\theta_0})\right)} \times [\mathbb{E}(Q_j^6)\mathbb{E}(Q_k^6)\mathbb{E}(T_{mk}^6)]^{\frac{1}{6}} \leq \frac{212}{\sqrt{n}}. \tag{3.17}$$

Combining the results in (3.15) and (3.17),

$$K_1(\boldsymbol{\theta_0}) \leq 26 + \frac{212}{\sqrt{n}}. \tag{3.18}$$

Following the same steps as in (3.15) and (3.17) under the null hypothesis $\mu = 0$, with $\hat{\theta}_*(\boldsymbol{X})_1 = \frac{1}{n}\sum_{i=1}^n X_i^2$,

$$K_1^*(\boldsymbol{\theta_0}) \leq 33 + \frac{220}{\sqrt{n}}. \tag{3.19}$$

We proceed to find a bound for $K_2(\boldsymbol{\theta_0})$, as defined in (2.9). The calculation of the first term is straightforward and

$$2\sqrt{n}\frac{\|h\|}{\epsilon^2}\sum_{j=1}^{2}\mathbb{E}(Q_j^2) = 2\sqrt{n}\frac{\|h\|}{\epsilon^2}\left(\frac{\sigma^2}{n}+\frac{\sigma^4}{n}\left(2-\frac{1}{n}\right)\right) \leq \frac{2\|h\|\sigma^2}{\sqrt{n}\epsilon^2}(1+2\sigma^2). \tag{3.20}$$

Using (3.11), (3.12) and (3.13), we are able to find an upper bound for the second term in (2.9). Simple calculations yield

$$\begin{aligned}\sqrt{n}\|h'\|\frac{7}{3}\sum_{j=1}^{2}\sum_{k=1}^{2}\sum_{m=1}^{2}&[\mathbb{E}(Q_j^2Q_k^2Q_m^2)]^{\frac{1}{2}}[\mathbb{E}[(M_{jkm}(\boldsymbol{X}))^2||Q_{(m)}|<\epsilon]]^{\frac{1}{2}}\\ &= \|h'\|\left\{\frac{7\sqrt{6}\sigma^4}{(\sigma^2-\epsilon)^2}+\frac{56\sqrt{2}\sigma^5}{(\sigma^2-\epsilon)^3}\sqrt{\frac{\sigma^2}{n}+\epsilon^2}\right.\\ &\quad\left.+\frac{7\sqrt{2120}\sigma^6}{3(\sigma^2-\epsilon)^3}\sqrt{1+\frac{162}{(\sigma^2-\epsilon)^2}\left((\epsilon+\epsilon^2+\sigma^2)^2+\frac{3\sigma^4}{n^2}\right)}\right\}.\end{aligned} \tag{3.21}$$

The third term of (2.9) requires the calculation of conditional expectations related to T_{kj} of (2.10) and $M_{qml}(\boldsymbol{X})$, where $k, j, q, m, l \in \{1, 2\}$. It is easy to see that both T_{12} and T_{22} can be written as continuous, increasing functions of Q_1 and Q_2. Therefore, with $Q_{(m)}$ as in (2.10) and for $G_n \sim \chi_n^2$, employing Lemma 4.1 of [1], leads to

$$\mathbb{E}(T_{11}^4||Q_{(m)}|<\epsilon) = 0, \tag{3.22}$$

$$\mathbb{E}(T_{12}^4||Q_{(m)}|<\epsilon) = \mathbb{E}\left(\frac{1}{\sigma^{16}}\left(\sum_{i=1}^{n}(X_i-\mu)\right)^4\Big||Q_{(m)}|<\epsilon\right) \leq \frac{n^4}{\sigma^{16}}\mathbb{E}(\bar{X}-\mu)^4 = \frac{3n^2}{\sigma^{12}},$$

$$\begin{aligned}\mathbb{E}(T_{22}^4 \| Q_{(m)}| < \epsilon) &= \mathbb{E}\left(\left(\frac{n}{\sigma^4} - \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2\right)^4 \middle| |Q_{(m)}| < \epsilon\right) \\ &\leq \mathbb{E}\left(\frac{n}{\sigma^4} - \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2\right)^4 \\ &= \frac{1}{\sigma^{16}}\mathbb{E}(G_n - n)^4 = \frac{12n^2}{\sigma^{16}}\left(1 + \frac{4}{n}\right) \leq \frac{60n^2}{\sigma^{16}}.\end{aligned}$$

Using the results of (3.11), (3.12) and (3.22), after simple calculations, we get for the third term of (2.9) that

$$\begin{aligned}&\frac{\|h'\|}{\sqrt{n}}\sum_{q=1}^{2}\sum_{k=1}^{2}|[[I(\boldsymbol{\theta_0})]^{-1}]_{kq}| \\ &\quad\times \sum_{j=1}^{2}\sum_{l=1}^{2}\sum_{m=1}^{2}[\mathbb{E}(Q_j^2 Q_l^2 Q_m^2)]^{\frac{1}{2}}[\mathbb{E}(T_{kj}^4 \| Q_{(m)}| < \epsilon)]^{\frac{1}{4}} \\ &\quad\times [\mathbb{E}(M_{qml}^4(\boldsymbol{X}) \| Q_{(m)}| < \epsilon)]^{\frac{1}{4}} \qquad (3.23)\\ &\leq \frac{\|h'\|}{\sqrt{n}}\left(\frac{35\sigma^4}{(\sigma^2-\epsilon)^2} + \frac{339\sigma^6}{(\sigma^2-\epsilon)^3}\left(\frac{3}{n^2} + \left(\frac{\epsilon}{\sigma}\right)^4\right)^{\frac{1}{4}}\right) \\ &\quad+ \frac{2700\sigma^6\|h'\|}{\sqrt{n}(\sigma^2-\epsilon)^3}\left(1 + \frac{52{,}488}{(\sigma^2-\epsilon)^4}\left((\epsilon+\epsilon^2+\sigma^2)^4 + \frac{105\sigma^8}{n^4}\right)\right)^{\frac{1}{4}}.\end{aligned}$$

To find an upper bound for the fourth term of (2.9), using (3.10), (3.11), (3.12),

$$\begin{aligned}&\frac{\|h'\|}{4\sqrt{n}}\sum_{b=1}^{2}\sum_{k=1}^{2}\sum_{m=1}^{2}\sum_{q=1}^{2}\sum_{l=1}^{2}\sum_{j=1}^{2}|[[I(\boldsymbol{\theta_0})]^{-1}]_{qb}|\sqrt{\mathbb{E}(Q_k^2 Q_m^2 Q_j^2 Q_l^2)} \\ &\quad\times [\mathbb{E}((M_{bmk}(\boldsymbol{X}))^4 \| Q_{(m)}| < \epsilon)]^{\frac{1}{4}}[\mathbb{E}((M_{qjl}(\boldsymbol{X}))^4 \| Q_{(m)}| < \epsilon)]^{\frac{1}{4}} \\ &\leq \frac{\|h'\|}{\sqrt{n}}\left\{\frac{13\sigma^8}{(\sigma^2-\epsilon)^4} + \frac{147\sigma^{10}}{(\sigma^2-\epsilon)^5}\left(\frac{3}{n^2} + \left(\frac{\epsilon}{\sigma}\right)^4\right)^{\frac{1}{4}}\right. \\ &\quad+ \frac{1361\sigma^{12}}{(\sigma^2-\epsilon)^6}\sqrt{\frac{3}{n^2} + \left(\frac{\epsilon}{\sigma}\right)^4} \qquad (3.24)\\ &\quad+ \left(1 + \frac{52{,}488}{(\sigma^2-\epsilon)^4}\left((\epsilon+\epsilon^2+\sigma^2)^4 + \frac{105\sigma^8}{n^4}\right)\right)^{\frac{1}{4}} \\ &\quad\times\left(\frac{12\sigma^{10}}{(\sigma^2-\epsilon)^5} + \frac{369\sigma^{12}}{(\sigma^2-\epsilon)^6}\left(\frac{3}{n^2} + \left(\frac{\epsilon}{\sigma}\right)^4\right)^{\frac{1}{4}}\right) \\ &\quad\left.+ \frac{\sqrt{362{,}096}\sigma^{12}}{(\sigma^2-\epsilon)^6}\sqrt{1 + \frac{52{,}488}{(\sigma^2-\epsilon)^4}\left((\epsilon+\epsilon^2+\sigma^2)^4 + \frac{105\sigma^8}{n^4}\right)}\right\}.\end{aligned}$$

The bounds in (3.20), (3.21), (3.23) and (3.24) depend on the constant ϵ as defined in the statement of Theorem 2.3. For the choice of ϵ, (3.11), (3.12) and (3.13) require that $0 < \epsilon < \sigma^2$. There is trade off related to the choice of ϵ between the expressions (3.20) and (3.24).

We choose $\epsilon = \frac{\sigma^2}{2}$. Using this value in (3.20), (3.21), (3.23) and (3.24), leads to

$$\begin{aligned}
K_2(\boldsymbol{\theta_0}) \leq{}& \frac{8\|h\|}{\sqrt{n}\sigma^2}(1+2\sigma^2) + \|h'\|\biggl\{28\sqrt{6} + 448\sqrt{\frac{2}{n} + \frac{\sigma^2}{2}} + \frac{348}{\sqrt{n}} \\
&+ 860\sqrt{1 + 648\biggl(\biggl(\frac{3}{2} + \frac{\sigma^2}{4}\biggr)^2 + \frac{3}{n^2}\biggr)} + \frac{7416}{\sqrt{n}}\biggl(\frac{3}{n^2} + \frac{\sigma^4}{16}\biggr)^{\frac{1}{4}} \\
&+ \frac{1}{\sqrt{n}}\biggl(1 + 839{,}808\biggl(\biggl(\frac{3}{2} + \frac{\sigma^2}{4}\biggr)^4 + \frac{105}{n^4}\biggr)\biggr)^{\frac{1}{4}} \qquad (3.25)\\
&\times \biggl(21{,}984 + 23{,}616\biggl(\frac{3}{n^2} + \frac{\sigma^4}{16}\biggr)^{\frac{1}{4}}\biggr) \\
&+ \frac{87{,}104}{\sqrt{n}}\sqrt{\frac{3}{n^2} + \frac{\sigma^4}{16}} + \frac{38{,}512}{\sqrt{n}}\sqrt{1 + 839{,}808\biggl(\biggl(\frac{3}{2} + \frac{\sigma^2}{4}\biggr)^4 + \frac{105}{n^4}\biggr)}\biggr\}.
\end{aligned}$$

It remains to find an upper bound for $K_2^*(\boldsymbol{\theta_0})$, which is the version of $K_2(\boldsymbol{\theta_0})$ under the null hypothesis of $\mu = 0$. This requires the calculation of conditional expectations related to T_{11}^* of (2.10) and $M_{111}^*(\boldsymbol{X})$ as defined in (C.3)(b). Simple calculations yield

$$M_{111}^*(\boldsymbol{X}) = \frac{n}{(\sigma^2-\epsilon)^3} + \frac{3}{(\sigma^2-\epsilon)^4}\sum_{i=1}^n X_i^2, \qquad T_{11}^* = \frac{n}{\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^n X_i^2.$$

Using the above results and Lemma 4.1 from [1], we obtain that for $\epsilon = \frac{\sigma^2}{2}$,

$$\begin{aligned}
K_2^*(\boldsymbol{\theta_0}) ={}& 2\sqrt{n}\frac{\|h\|}{\epsilon^2}\mathbb{E}(Q_1^*)^2 \\
&+ \sqrt{n}\|h'\|\frac{7}{3}\sqrt{\mathbb{E}(Q_1^*)^6}\sqrt{\mathbb{E}[(M_{111}^*(\boldsymbol{X}))^2 | |Q_{(m)}^*| < \epsilon]} \\
&+ \frac{2\|h'\|\sigma^4}{\sqrt{n}}\sqrt{\mathbb{E}(Q_j^*)^6}[\mathbb{E}((T_{11}^*)^4 | |Q_{(m)}^*| < \epsilon)]^{\frac{1}{4}} \qquad (3.26)\\
&\times [\mathbb{E}((M_{111}^*(\boldsymbol{X}))^4 | |Q_{(m)}^*| < \epsilon)]^{\frac{1}{4}} \\
&+ \frac{\sigma^4\|h'\|}{2\sqrt{n}}\sqrt{\mathbb{E}(Q_1^*)^8}\sqrt{\mathbb{E}((M_{111}^*(\boldsymbol{X}))^4 | |Q_{(m)}^*| < \epsilon)} \\
\leq{}& \frac{16\|h\|}{\sqrt{n}} + 15{,}958\|h'\| + \frac{13{,}527{,}046}{\sqrt{n}}\|h'\|.
\end{aligned}$$

Applying the results of (3.14), (3.18), (3.19), (3.25) and (3.26), to the expression of the general upper bound in (2.6) yields (3.9). □

Empirical results. As in the exponential distribution example, we assess the accuracy of our bounds by simulations. We start by generating 100 trials of n random independent observations, x, from $\mathrm{N}(\mu, \sigma^2)$, where $n = 1000 \times j$, $j = 1, 2, \ldots, 1000$ and $\mu = 0, \sigma = 1$. We evaluate the log-likelihood ratio statistic, $-2\log\Lambda$, in each trial, which in turn gives a vector of 100 values. For this example, simple steps yield

$$-2\log\Lambda = n\biggl(\log\biggl(\frac{\sum_{i=1}^n X_i^2}{\sum_{i=1}^n (X_i - \bar{X})^2}\biggr)\biggr).$$

TABLE 2
Simulation results for the normal distribution example

n	$Q_{h_t}^{\mathrm{N}}$	Upper bound	Error
10^4	0.023	2960.752	2960.729
10^5	0.020	323.226	323.206
10^6	0.016	40.924	40.908

We apply to these values the function $h_t(x) = (x^2+2)^{-1}$ again and we calculate their sample mean, denoted by $\hat{\mathbb{E}}[h_t(-2\log\Lambda)]$. We define

$$Q_{h_t}^{\mathrm{N}} := \left|\hat{\mathbb{E}}\big[h_t(-2\log\Lambda)\big] - \tilde{\mathbb{E}}\big[h_t(K)\big]\right|, \tag{3.27}$$

where $\tilde{\mathbb{E}}[h_t(K)] = 0.373$. We compare $Q_{h_t}^{\mathrm{N}}$ with the bound in (3.9), using the difference between their values as a measure of the error. The results for $n = 10^j$, $j = 4, 5, 6$ are presented in Table 2 below. The table indicates that when at each step the sample size is increased, then the value of the upper bound decreases. The values of the bound are quite large, but this is due to the large constants in (3.9). Smaller constants are obtainable, but this, in our understanding, will make the bound in (3.9) more tedious and difficult to present. Our main purpose has been to show the order of the bound with respect to the sample size, n.

3.3. *Example*: *Logistic regression*. In binomial regression, the data are i.i.d. observations $(\boldsymbol{X_i}, Y_i)$, $i = 1, \dots, n$, where $\boldsymbol{X_i} \in \mathbb{R}^d$ and $Y_i \in \{0, 1\}$; see, for example, [23], page 66. The binary regression model is that

$$\mathbb{P}_{\boldsymbol{\theta}}(Y_i = 1 | \boldsymbol{X_i} = \boldsymbol{x}) = \psi(\boldsymbol{\theta}^\top \boldsymbol{x})$$

for $\boldsymbol{\theta} \in \mathbb{R}^d$ and $\psi : \mathbb{R} \to [0, 1]$ continuously differentiable, monotone, with derivatives bounded away from 0 and ∞. For logistic regression,

$$\psi(\theta) = \frac{1}{1 + e^{-\theta}}$$

and here we restrict ourselves to this case, although generalisations are straightforward. We assume that the distribution of $\boldsymbol{X}$ is such that $\boldsymbol{X}$ has finite moments up to order 6. To ensure that the MLE, $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})$, for $\boldsymbol{\theta}$ exists and is unique, we assume that the $\boldsymbol{X_i}$'s do not concentrate on a $(d-1)$-dimensional affine subspace of $\mathbb{R}^d$.

Consider as in [22] to test the simple hypothesis that $\boldsymbol{\theta_0} = \boldsymbol{0}$ against the general alternative. The likelihood in this case is

$$L(\boldsymbol{\theta}; (\boldsymbol{x_i}, y_i), i = 1, \dots, n) = \prod_{i=1}^{n} \left\{\psi(\boldsymbol{\theta}^\top \boldsymbol{x_i})^{y_i} \left(1 - \psi(\boldsymbol{\theta}^\top \boldsymbol{x_i})\right)^{1-y_i}\right\}$$

so that the score function is

$$S(\boldsymbol{\theta}) = \frac{y - \psi(\boldsymbol{\theta}^\top \boldsymbol{x})}{\psi(\boldsymbol{\theta}^\top \boldsymbol{x})(1 - \psi(\boldsymbol{\theta}^\top \boldsymbol{x}))} \psi'(\boldsymbol{\theta}^\top \boldsymbol{x}) \boldsymbol{x}$$

while the Fisher information matrix is $I(\boldsymbol{\theta}) = \mathbb{E}[\psi'(\boldsymbol{\theta}^\top \boldsymbol{X}))\boldsymbol{X}\boldsymbol{X}^\top]$. For testing $H_0 : \theta_1 = 0$, we have for $\boldsymbol{x} = (x_1, \dots, x_d)$,

$$\xi(\boldsymbol{x}) = \frac{1}{\sqrt{n}} (y_1 - \psi(\boldsymbol{\theta}^\top \boldsymbol{x})) x_1$$

and

$$\boldsymbol{\eta}(\boldsymbol{x}) = \frac{1}{\sqrt{n}}\big((y_2 - \psi(\boldsymbol{\theta}^\top \boldsymbol{x}))x_2, \dots, (y_p - \psi(\boldsymbol{\theta}^\top \boldsymbol{x}))x_p\big)^\top.$$

To check the assumptions on the third derivative of the log-likelihood, we calculate

$$\frac{\partial^3}{\partial\theta_k\,\partial\theta_j\partial\theta_i}\ell(\boldsymbol{\theta};(\boldsymbol{x},y)) = \frac{e^{\boldsymbol{\theta}^\top \boldsymbol{x}}(1+e^{2\boldsymbol{\theta}^\top \boldsymbol{x}} - 4e^{\boldsymbol{\theta}^\top \boldsymbol{x}})}{(1+e^{\boldsymbol{\theta}^\top \boldsymbol{x}})^4} x_i x_j x_k$$

so that

$$\left|\frac{\partial^3}{\partial\theta_k\,\partial\theta_j\partial\theta_i}\ell(\boldsymbol{\theta};(\boldsymbol{x},y))\right| \le |x_i x_j x_k| =: M_{k,j,i}(\boldsymbol{x}).$$

In particular, $\mathbb{E}(M_{k,j,i}(\mathbf{X})^2) = \mathbb{E}(X_i^2 X_j^2 X_k^2)$ and $\mathbb{E}(M_{k,j,i}(\mathbf{X})^4) = \mathbb{E}(X_i^4 X_j^4 X_k^4)$.

In order to apply Theorem 2.1, we use the variables

$$Z_{i,j} = \frac{\partial}{\partial\theta_j}(y_i \log\psi(\boldsymbol{\theta}^\top \boldsymbol{x_i}) + (1-y_i)\log(1-\psi(\boldsymbol{\theta}^\top \boldsymbol{x_i})) = (y_i - \psi(\boldsymbol{\theta}^\top \boldsymbol{x_i}))x_{i,j}$$

for $i = 1,\dots,n$, $j = 1,\dots,d$. Due to the binomial structure and the fact that $|y_i - \psi(\boldsymbol{\theta}^\top \boldsymbol{X_i})| \le 1$, it holds that $\mathbb{E}|Z_{i,j}| \le \mathbb{E}|X_{i,j}|$.

Note that for logistic regression the MLE in general does not have a closed form. Hence we cannot evaluate (2.10) explicitly, although with given data sets a numerical evaluation is possible. For our purposes, it suffices to illustrate the applicability of the bound as well as its behaviour in terms of d and n.

4. Proof of Theorem 2.1.

PROOF. Here is the proof of Theorem 2.1. The proof is based on Stein's method, as follows. From [10], a random variable X has the chi-square distribution with r degrees of freedom if and only if

$$\mathbb{E}\left[Xf''(X) + \frac{1}{2}(r-X)f'(X)\right] = 0 \tag{4.1}$$

for all twice differentiable functions $f : \mathbb{R}^+ \to \mathbb{R}$ such that the expectations in (4.1) exist. Moreover, if g is bounded and has three bounded derivatives, then the Stein equation

$$xf''(x) + \frac{1}{2}(r-x)f'(x) = g(x) - \mathbb{E}[g(\chi_r)] \tag{4.2}$$

has solution f which satisfies

$$\|f^{(k)}\| \le \frac{4}{r+k-1}(3\|g^{(k-1}\| + \|g^{(k-2)}\|) \tag{4.3}$$

for $k = 2, 3, 4$. Using (4.2) with solution $f = f_g$, for any random variable T,

$$\mathbb{E}[g(T)] - \mathbb{E}[g(\chi_r)] = \mathbb{E}\left[Tf''(T) + \frac{1}{2}(r-T)f'(T)\right]. \tag{4.4}$$

For $T = \mathbf{Z}^\top U\mathbf{Z}$, we expand

$$T = \frac{1}{n}\sum_{i=1}^{r}\sum_{k=1}^{r} U_{i,k}\sum_{j=1}^{n}\sum_{\ell=1}^{n} Z_{i,j}Z_{k,\ell}.$$

Now, for any $s, k = 1, \dots, r$, from the definition of the inverse,

$$\sum_{i=1}^{r} U_{i,k}\tau_{i,s} = \sum_{i=1}^{r} U_{k,i}\tau_{i,s} = \mathbb{1}(s = k) \tag{4.5}$$

and as T is a quadratic form, it has mean

$$\mathbb{E}[T] = \mathrm{trace}(U\tau) = r. \tag{4.6}$$

As $Z_{i,j}$ is independent of $Z_{k,\ell}$ for $\ell \neq j$, $k = 1, \dots, r$, T is a sum of locally dependent summands. For $j, \ell = 1, \dots, n$, set

$$T^j = \frac{1}{n}\sum_{i=1}^{r}\sum_{k=1}^{r}\sum_{s\neq j}\sum_{t\neq j} Z_{i,s}U_{i,k}Z_{k,t}.$$

Then T^j is independent of $Z_{a,j}$ for all $a = 1, \dots, r$. Moreover, using the symmetry of U,

$$T - T^j = \frac{1}{n}\sum_{i=1}^{r}\sum_{k=1}^{r} U_{i,k}Z_{i,j}\biggl(Z_{k,j} + 2\sum_{s\neq j} Z_{k,s}\biggr). \tag{4.7}$$

Next, Taylor expansion gives

$$\mathbb{E}[Tf'(T)] = \frac{1}{n}\sum_{i=1}^{r}\sum_{k=1}^{r} U_{i,k}\sum_{j=1}^{n}\sum_{\ell=1}^{n}\mathbb{E}[Z_{i,j}Z_{k,\ell}f'(T^j)] \tag{4.8}$$

$$+ \frac{1}{n}\sum_{i=1}^{r}\sum_{k=1}^{r} U_{i,k}\sum_{j=1}^{n}\sum_{\ell=1}^{n}\mathbb{E}[Z_{i,j}Z_{k,\ell}(T - T^j)f''(T^j)] \tag{4.9}$$

$$+ 2R_1,$$

with

$$R_1 = \frac{1}{4n}\sum_{i,k=1}^{r} U_{i,k}\sum_{j,\ell=1}^{n}\mathbb{E}[Z_{i,j}Z_{k,\ell}(T - T^j)^2 f^{(3)}(T^j + \rho T)] \tag{4.10}$$

for some $0 < \rho < 1$. We shall return to this remainder term later. First, from the independence, (4.8) yields

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{r}\sum_{k=1}^{r} U_{i,k}\sum_{j=1}^{n}\sum_{\ell=1}^{n}\mathbb{E}[Z_{i,j}Z_{k,\ell}f'(T^j)] \\
&\quad = \sum_{i=1}^{r}\sum_{k=1}^{r} U_{i,k}\tau_{i,k}\mathbb{E}[f'(T)] + 2R_2 \\
&\quad = r\mathbb{E}[f'(T)] + 2R_2,
\end{aligned}$$

where we used (4.6) in the last step, and

$$R_2 = \frac{1}{2}\sum_{i=1}^{r}\sum_{k=1}^{r} U_{i,k}\tau_{i,k}\biggl(\frac{1}{n}\sum_{j=1}^{n}\mathbb{E}[f'(T^j)] - \mathbb{E}[f'(T)]\biggr). \tag{4.11}$$

For (4.9), with (4.7),

$$\frac{1}{n}\sum_{i=1}^{r}\sum_{k=1}^{r} U_{i,k}\sum_{j=1}^{n}\sum_{\ell=1}^{n}\mathbb{E}[Z_{i,j}Z_{k,\ell}(T - T^j)f''(T^j)]$$

$$
\begin{aligned}
&= \frac{1}{n^2} \sum_{i,k,a,b=1}^{r} U_{i,k} U_{a,b} \beta(i,k,a,b) \sum_{j=1}^{n} \mathbb{E}[f''(T^j)] \\
&\quad + \frac{1}{n^2} \sum_{i,k,a,b=1}^{r} U_{i,k} U_{a,b} \beta(i,a,b) \Biggl(\sum_{j=1}^{n} \sum_{\ell \neq j} \mathbb{E}[Z_{k,\ell} f''(T^j)] \\
&\quad + 2 \sum_{j=1}^{n} \sum_{s \neq j} \mathbb{E}[Z_{g,s} f''(T^j)] \Biggr) \\
&\quad + \frac{2}{n^2} \sum_{i,k,a,b=1}^{r} U_{i,k} U_{a,b} \tau_{i,a} \sum_{j=1}^{n} \sum_{s \neq j} \sum_{\ell \neq j} \mathbb{E}[Z_{k,\ell} Z_{g,s} f''(T^j)] \\
&= 2R_3 + 2R_4 + \frac{2}{n^2} \sum_{k,b=1}^{r} U_{k,b} \sum_{j=1}^{n} \sum_{s \neq j} \sum_{\ell \neq j} \mathbb{E} Z_{k,\ell} Z_{g,s} f''(T^j),
\end{aligned}
$$

where we used (4.5) in the last step and put

$$
R_3 = \frac{1}{2n^2} \sum_{i,k,a,b=1}^{r} U_{i,k} U_{a,b} \beta(i,k,a,b) \sum_{j=1}^{n} \mathbb{E}[f''(T^j)] \tag{4.12}
$$

and

$$
R_4 = \frac{3}{2n^2} \sum_{i,k,a,b=1}^{r} U_{i,k} U_{a,b} \beta(i,a,b) \sum_{j=1}^{n} \sum_{\ell \neq j} \mathbb{E}[Z_{k,\ell} f''(T^j)]. \tag{4.13}
$$

Thus for (4.4),

$$
\begin{aligned}
&\mathbb{E}[g(T)] - \mathbb{E}[g(\chi_r)] \\
&\quad = \mathbb{E}\left[T f''(T) + \frac{1}{2}(r - T) f'(T) \right] \\
&\quad = \frac{1}{n} \sum_{k,b=1}^{r} U_{k,b} \sum_{s=1}^{n} \sum_{\ell=1}^{n} \mathbb{E}\left[Z_{k,\ell} Z_{g,s} \left(f''(T) - \frac{1}{n} \sum_{j \neq \ell, s} f''(T^j) \right) \right] - \sum_{i=1}^{4} R_i \\
&\quad = -\sum_{i=1}^{4} R_i + R_5 + R_6
\end{aligned}
$$

with

$$
R_5 = \frac{1}{n} \mathbb{E}[T f''(T)] \tag{4.14}
$$

and

$$
R_6 = \frac{1}{n^2} \sum_{k,b=1}^{r} U_{k,b} \sum_{s=1}^{n} \sum_{\ell=1}^{n} \mathbb{E}\left[Z_{k,\ell} Z_{g,s} \sum_{j \neq \ell, s} \bigl(f''(T) - f''(T^j) \bigr) \right]. \tag{4.15}
$$

It remains to bound the remainder terms. To this purpose, we carry out an ancillary calculation. With (4.7) and using the independence,

$$
\begin{aligned}
&\mathbb{E}[(T - T^j)^2] \\
&\quad = \frac{1}{n^2} \sum_{a,b,i,k} U_{a,b} U_{i,k} \mathbb{E}\left[Z_{i,j} Z_{a,j} \left(Z_{b,j} + 2 \sum_{\ell \neq j} Z_{b,\ell} \right) \left(Z_{k,j} + 2 \sum_{s \neq j} Z_{k,s} \right) \right]
\end{aligned}
$$

$$= \frac{1}{n^2} \sum_{a,b,i,k} U_{a,b} U_{i,k} \big(\beta(a,b,i,k) + 4(n-1)\tau_{i,a}\tau_{b,k}\big)$$
$$= \frac{1}{n^2} \bigg(\sum_{a,b,i,k} U_{a,b} U_{i,k} \beta(a,b,i,k) + 4(n-1)r \bigg)$$
$$= \frac{1}{n} B_1(r,n),$$

where we employed (4.5) and (4.6) for the last step and

$$B_1(r,n) = \frac{4rn}{n-1} + \frac{1}{n}\mathbb{E}(W^2)$$

with $W = \sum_{i,k=1}^{r} U_{i,k} Z_{i,1} Z_{k,1}$. Similarly,

$$\mathbb{E}\big[(T-T^j)^4\big]$$
$$= \frac{1}{n^4} \sum_{a,b,i,k,c,d,e,f} U_{a,b} U_{i,k} U_{c,d} U_{e,f} \mathbb{E}\bigg[Z_{i,j} Z_{a,j} Z_{c,j} Z_{e,j}$$
$$\times \bigg(Z_{b,j} + 2\sum_{\ell \neq j} Z_{b,\ell} \bigg) \bigg(Z_{k,j} + 2\sum_{s \neq j} Z_{k,s} \bigg)$$
$$\times \bigg(Z_{d,j} + 2\sum_{t \neq j} Z_{d,t} \bigg) \bigg(Z_{f,j} + 2\sum_{u \neq j} Z_{f,u} \bigg) \bigg]$$
$$= \frac{1}{n^4} \sum_{a,b,i,k,c,d,e,f} U_{a,b} U_{i,k} U_{c,d} U_{e,f} \big(\beta(a,b,c,d,e,f,i,k)$$
$$+ 24\beta(i,a,c,e,b,d)\tau(k,f)(n-1) + 32\beta(i,a,c,e,b)\beta(c,k,f)(n-1)$$
$$+ 16\beta(i,a,c,e)\big(\beta(b,d,k,f)(n-1) + 6n(n-1)\tau_{b,d}\tau_{k,f}\big)$$
$$\leq \frac{1}{n^2} B_2(r,n)$$

with

$$B_2(r,n) = 96\mathbb{E}(W^2)$$
$$+ \frac{1}{n} \bigg(\frac{1}{n}\mathbb{E}(W^4) + 24\mathbb{E}(W^3)$$
$$+ 32 \sum_{a,b,i,k,c,d,e,f} U_{a,b} U_{i,k} U_{c,d} U_{e,f} \beta(i,a,c,e,b)\beta(c,k,f)$$
$$+ 16 \sum_{a,b,i,k,c,d,e,f} U_{a,b} U_{i,k} U_{c,d} U_{e,f} \beta(i,a,c,e)\beta(b,d,k,f) \bigg).$$

Now the ingredients are in place to bound $|R_i|, i = 1, \dots, 6$.

Bounding R_1. For R_1 from (4.10) and the Cauchy–Schwarz inequality, using the notation (2.4),

$$R_1 = \frac{1}{4n} \sum_{i,k=1}^{r} U_{i,k} \sum_{j,\ell=1}^{n} \mathbb{E}\big[Z_{i,j} Z_{k,\ell} (T-T^j)^2 f^{(3)}(T^j + \rho T)\big]$$

(4.16)

$$= R_{1,1} + R_{1,2}$$

with

$$|R_{1,1}| \leq \frac{1}{4n}\|f^{(3)}\| \sum_{j=1}^{n} \mathbb{E}\big[|W|\big(T - T^j\big)^2\big]$$
$$\leq \frac{1}{4n}\|f^{(3)}\|\sqrt{B_2(r,n)}\sqrt{\mathbb{E}[W^2]}$$

and

$$|R_{1,2}| \leq \frac{1}{4n}\|f^{(3)}\| \sum_{j=1}^{n} \mathbb{E}\left[\left|\sum_{i,k=1}^{r} U_{i,k} Z_{i,j} \sum_{\ell \neq j} Z_{k,\ell}\right| \big(T - T^j\big)^2\right]$$
$$\leq \frac{1}{4n^2}\|f^{(3)}\|\sqrt{B_2(r,n)} \sum_{j=1}^{n} \sqrt{\mathbb{E}\left[\left(\sum_{i,k=1}^{r} U_{i,k} Z_{i,j} \sum_{\ell \neq j} Z_{k,\ell}\right)^2\right]}.$$

We calculate

$$\mathbb{E}\left[\left(\sum_{i,k=1}^{r} U_{i,k} Z_{i,j} \sum_{\ell \neq j} Z_{k,\ell}\right)^2\right] = \sum_{i,k=1}^{r} U_{i,k} \sum_{a,b=1}^{r} U_{a,b} \tau_{i,a} \tau_{b,k}$$
$$(4.17) \qquad = r$$

and thus

$$|R_{1,2}| \leq \frac{\sqrt{r}}{4n}\|f^{(3)}\|\sqrt{B_2(r,n)}.$$

Hence, for (4.16),

$$|R_1| \leq \frac{1}{4n}\|f^{(3)}\|\sqrt{B_2(r,n)}\big(\sqrt{\mathbb{E}(W^2)} + \sqrt{r}\big).$$

Bounding R_2. For R_2 from (4.11), by Taylor expansion and the Cauchy–Schwarz inequality,

$$|R_2| \leq \frac{1}{2n}\|f''\| \sum_{j=1}^{n} \left|\sum_{i=1}^{r} \sum_{k=1}^{r} U_{i,k} \tau_{i,k}\right| \mathbb{E}\big[\big|T - T^j\big|\big]$$
$$\leq \frac{r}{2\sqrt{n}}\|f''\|\sqrt{B_1(r,n)}.$$

Bounding R_3. For R_3 from (4.12), it is straightforward to bound

$$|R_3| \leq \frac{1}{2n}\|f''\| \left|\sum_{i,k,a,b=1}^{r} U_{i,k} U_{a,b} \beta(i,k,a,b)\right| = \frac{1}{2n}\|f''\| \mathbb{E}\big(W^2\big).$$

Bounding R_4. For R_4 from (4.13), for $\ell \neq j$ we introduce

$$T^{j,\ell} = \frac{1}{n} \sum_{i=1}^{r} \sum_{k=1}^{r} \sum_{s \neq j} \sum_{t \neq j,\ell} Z_{i,s} U_{i,k} Z_{k,t}.$$

Then $T^{j,\ell}$ is independent of $Z_{a,\ell}$ for all $a = 1, \dots, r$. Moreover, as for (4.7),

$$T^j - T^{j,\ell} = \frac{1}{n} \sum_{i=1}^{r} \sum_{k=1}^{r} U_{i,k} Z_{i,\ell} \left(Z_{k,\ell} + 2 \sum_{s \neq j,\ell} Z_{k,s}\right).$$

Thus, for $\ell \neq j$,

$$\mathbb{E}\big[Z_{k,\ell} f''\big(T^j\big)\big] = \mathbb{E}\big[Z_{k,\ell}\big(T^j - T^{j,\ell}\big) f^{(3)}\big(T^j + \rho\big(T^{j,\ell} - T^j\big)\big)\big]$$

for some $0 < \rho < 1$. Hence using the Cauchy–Schwarz inequality,

$$|R_4| \leq \frac{3}{2}\|f^{(3)}\|\frac{1}{\sqrt{n}}\sqrt{B_1(r,n)}\sqrt{\mathbb{E}\left[\left(\sum_{i,k,a,b=1}^{r} U_{i,k}U_{a,b}\beta(i,a,b)Z_{k,\ell}\right)^2\right]}$$

$$\leq \frac{3}{2}\|f^{(3)}\|\frac{1}{\sqrt{n}}\sqrt{B_1(r,n)}\sqrt{\sum_{i,k,a,b,e,f=1}^{r} U_{i,k}U_{a,b}U_{e,f}\beta(i,a,b)\beta(k,e,f)},$$

where we used (4.5).

Bounding R_5. Bounding R_5 from (4.14) is straightforward using Lemma 2.1 from [10] which states that for g bounded, the solution f of the Stein equation satisfies $\|xf''(x)\| \leq 4\|g\|$,

$$|R_5| = \frac{1}{n}|\mathbb{E}[Tf''(T)]| \leq \frac{4}{n}\|g\|.$$

Bounding R_6. For R_6 from (4.15),

$$R_6 = \frac{1}{n^2}\sum_{k,b=1}^{r} U_{k,b}\sum_{s=1}^{n}\sum_{\ell=1}^{n}\sum_{j\neq\ell,s}\mathbb{E}[Z_{k,\ell}Z_{b,s}(T-T^j)f^{(3)}(T^j)] \tag{4.18}$$

$$+\frac{1}{2n^2}\sum_{k,b=1}^{r} U_{k,b} \times\sum_{s=1}^{n}\sum_{\ell=1}^{n}\sum_{j\neq\ell,s}\mathbb{E}[Z_{k,\ell}Z_{b,s}(T-T^j)^2 f^{(4)}(T^j+\rho(T-T^j))] \tag{4.19}$$

for some $0 < \rho < 1$. Now, for $j \neq \ell, s$, with (4.7) and using the independence,

$$\sum_{s\neq\ell}\sum_{k,b=1}^{r} U_{k,b}\mathbb{E}[Z_{k,\ell}Z_{b,s}(T-T^j)f^{(3)}(T^j)]$$

$$= \frac{1}{n}\sum_{s\neq\ell}\sum_{k,b=1}^{r} U_{k,b}\sum_{c,d=1}^{r} U_{c,d}\tau(c,d)\mathbb{E}[Z_{k,\ell}Z_{b,s}f^{(3)}(T^j)]$$

$$= \frac{r}{n}\sum_{s\neq\ell}\sum_{k,b=1}^{r} U_{k,b}\mathbb{E}[Z_{k,\ell}Z_{b,s}f^{(3)}(T^j)]$$

$$= \frac{r}{n}\sum_{s\neq\ell}\mathbb{E}\left[\sum_{k,b=1}^{r} U_{k,b}Z_{k,\ell}Z_{b,s}(f^{(3)}(T^j)-f^{(3)}(T^{j,\ell}))\right]$$

so that

$$\sum_{s\neq\ell}\sum_{k,b=1}^{r} U_{k,b}\mathbb{E}[Z_{k,\ell}Z_{b,s}(T-T^j)f^{(3)}(T^j)]$$

$$\leq \frac{r}{n^{\frac{3}{2}}}\|f^{(4)}\|\sqrt{B_1(r,n)}\sqrt{\mathbb{E}\left[\left(\sum_{k,b=1}^{r} U_{k,b}Z_{k,\ell}\sum_{s\neq\ell}Z_{b,s}\right)^2\right]}$$

$$= \|f^{(4)}\|\left(\frac{r}{n}\right)^{\frac{3}{2}}\sqrt{B_1(r,n)},$$

where we used (4.18). Hence for (4.18), with the Cauchy–Schwarz inequality,

$$\frac{1}{n^2}\left|\sum_{k,b=1}^{r} U_{k,b}\sum_{s=1}^{n}\sum_{\ell=1}^{n}\sum_{j\neq\ell,s}\mathbb{E}\big[Z_{k,\ell}Z_{g,s}\big(T-T^j\big)f^{(3)}\big(T^j\big)\big]\right|$$
$$\leq\left(\frac{r}{n}\right)^{\frac{3}{2}}\|f^{(4)}\|\sqrt{B_1(r,n)}.$$

For (4.20), again with the Cauchy–Schwarz inequality,

$$\frac{1}{2n^2}\left|\sum_{k,b=1}^{r} U_{k,b}\sum_{s=1}^{n}\sum_{\ell=1}^{n}\sum_{j\neq\ell,s}\mathbb{E}\big[Z_{k,\ell}Z_{b,s}\big(T-T^j\big)^2 f^{(4)}\big(T^j+\rho\big(T-T^j\big)\big)\big]\right|$$
$$\leq\|f^{(4)}\|\frac{1}{2n^2}\sum_{j=1}^{n}\sqrt{\mathbb{E}\left[\left(\sum_{k,b=1}^{r}U_{k,b}\sum_{s\neq j}\sum_{\ell\neq j}Z_{k,\ell}Z_{b,s}\right)^2\right]}\sqrt{\mathbb{E}\big[\big(T-T^j\big)^4\big]}.$$

We calculate

$$\mathbb{E}\left[\left(\sum_{k,b=1}^{r}U_{k,b}\sum_{s\neq j}\sum_{\ell\neq j}Z_{k,\ell}Z_{b,s}\right)^2\right]$$
$$=(n-1)\mathbb{E}\big(W^2\big)+(n-1)^2\big(r^2+2r\big).$$

Hence,

$$\frac{1}{2n^2}\left|\sum_{k,b=1}^{r} U_{k,b}\sum_{s=1}^{n}\sum_{\ell=1}^{n}\sum_{j\neq\ell,s}\mathbb{E}\big[Z_{k,\ell}Z_{b,s}\big(T-T^j\big)^2\right.$$
$$\left.\times f^{(4)}\big(T^j+\rho\big(T-T^j\big)\big)\big]\right|$$
$$\leq\frac{1}{2n}\|f^{(4)}\|\sqrt{B_2(r,n)}\sqrt{\mathbb{E}\big(W^2\big)n^{-\frac{1}{2}}+r^2+2r}.$$

Thus,

$$|R_6|\leq\frac{1}{n}\|f^{(4)}\|\left(\frac{r^{\frac{3}{2}}}{\sqrt{n}}\sqrt{B_1(r,n)}+\sqrt{B_2(r,n)}\sqrt{\mathbb{E}\big(W^2\big)n^{-\frac{1}{2}}+r^2+2r}\right).$$

Collecting the bounds gives that

$$\sum_{i=1}^{6}|R_i|$$
$$\leq\frac{4}{n}\|g\|$$
$$+\|f''\|\left(\frac{r}{2\sqrt{n}}\sqrt{B_1(r,n)}+\frac{1}{n}\mathbb{E}\big(W^2\big)\right)$$
$$+\|f^{(3)}\|\left(\frac{3}{2\sqrt{n}}\sqrt{B_1(r,n)}\sqrt{\sum_{i,k,a,b,e,f=1}^{r}U_{i,k}U_{a,b}U_{e,f}\beta(i,a,b)\beta(k,e,f)}\right.$$

$$+ \frac{1}{4n}\sqrt{B_2(r,n)}\big(\sqrt{\mathbb{E}(W^2)}+\sqrt{r}\big)\bigg)$$

$$+ \frac{1}{2n}\|f^{(4)}\|\bigg(\frac{r^{\frac{3}{2}}}{\sqrt{n}}\sqrt{B_1(r,n)}+\sqrt{B_2(r,n)}\sqrt{\mathbb{E}(W^2)n^{-\frac{1}{2}}+r^2+2r}\bigg).$$

Employing the bounds $\|f^{(k)}\| \leq \frac{16}{r}\|\|g\|\|_3$ from (4.3) for the solution of the Stein equation gives the assertion. □

Acknowledgements. The authors would like to thank the anonymous referees for their helpful comments which have lead to an improvement of the paper.

The second author was supported in part by the Alan Turing Institute.

REFERENCES

[1] ANASTASIOU, A. (2018). Assessing the multivariate normal approximation of the maximum likelihood estimator from high-dimensional, heterogeneous data. *Electron. J. Stat.* **12** 3794–3828. MR3881763 https://doi.org/10.1214/18-EJS1492

[2] ANASTASIOU, A. and REINERT, G. (2017). Bounds for the normal approximation of the maximum likelihood estimator. *Bernoulli* **23** 191–218. MR3556771 https://doi.org/10.3150/15-BEJ741

[3] CASELLA, G. and BERGER, R. L. (2002). *Statistical Inference*, 2nd ed. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA. MR1051420

[4] CHEN, L. H. Y., GOLDSTEIN, L. and SHAO, Q.-M. (2011). *Normal Approximation by Stein's Method. Probability and Its Applications* (*New York*). Springer, Heidelberg. MR2732624 https://doi.org/10.1007/978-3-642-15007-4

[5] CHERNOFF, H. (1954). On the distribution of the likelihood ratio. *Ann. Math. Stat.* **25** 573–578. MR0065087 https://doi.org/10.1214/aoms/1177728725

[6] COX, D. R. (2006). *Principles of Statistical Inference*. Cambridge Univ. Press, Cambridge. MR2278763 https://doi.org/10.1017/CBO9780511813559

[7] DAVISON, A. C. (2003). *Statistical Models. Cambridge Series in Statistical and Probabilistic Mathematics* **11**. Cambridge Univ. Press, Cambridge. MR1998913 https://doi.org/10.1017/CBO9780511815850

[8] FAN, J., ZHANG, C. and ZHANG, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Statist.* **29** 153–193. MR1833962 https://doi.org/10.1214/aos/996986505

[9] FEDER, P. I. (1968). On the distribution of the log likelihood ratio test statistic when the true parameter is "near" the boundaries of the hypothesis regions. *Ann. Math. Stat.* **39** 2044–2055. MR0234553 https://doi.org/10.1214/aoms/1177698032

[10] GAUNT, R. E., PICKETT, A. M. and REINERT, G. (2017). Chi-square approximation by Stein's method with application to Pearson's statistic. *Ann. Appl. Probab.* **27** 720–756. MR3655852 https://doi.org/10.1214/16-AAP1213

[11] GEYER, C. J. (1994). On the asymptotics of constrained M-estimation. *Ann. Statist.* **22** 1993–2010. MR1329179 https://doi.org/10.1214/aos/1176325768

[12] GIBBS, A. L. and SU, F. E. (2002). On choosing and bounding probability metrics. *Int. Stat. Rev.* **70** 419–435.

[13] KENDALL, M. G. and STUART, A. (1969). *The Advanced Theory of Statistics*, 3rd ed. *Distribution Theory* **1**. Charles Griffin and Company Limited, London.

[14] LE CAM, L. (1960). Locally asymptotically normal families of distributions. Certain approximations to families of distributions and their use in the theory of estimation and testing hypotheses. *Univ. California Publ. Statist.* **3** 37–98. MR0126903

[15] LE CAM, L. (1970). On the assumptions used to prove asymptotic normality of maximum likelihood estimates. *Ann. Math. Stat.* **41** 802–828. MR0267676 https://doi.org/10.1214/aoms/1177696960

[16] LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing Statistical Hypotheses*, 3rd ed. *Springer Texts in Statistics*. Springer, New York. MR2135927

[17] LIU, C. and RUBIN, D. B. (1995). ML estimation of the t distribution using EM and its extensions, ECM and ECME. *Statist. Sinica* **5** 19–39. MR1329287

[18] POLLARD, D. (2013). A note on insufficiency and the preservation of Fisher information. In *From Probability to Statistics and Back*: *High-Dimensional Models and Processes*. *Inst. Math. Stat.* (*IMS*) *Collect.* **9** 266–275. IMS, Beachwood, OH. MR3202639 https://doi.org/10.1214/12-IMSCOLL919

[19] PORTNOY, S. (1988). Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *Ann. Statist.* **16** 356–366. MR0924876 https://doi.org/10.1214/aos/1176350710

[20] READ, T. R. C. and CRESSIE, N. A. C. (2012). *Goodness-of-Fit Statistics for Discrete Multivariate Data*. *Springer Series in Statistics*. Springer, New York. MR0955054 https://doi.org/10.1007/978-1-4612-4578-0

[21] SELF, S. G. and LIANG, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Amer. Statist. Assoc.* **82** 605–610. MR0898365

[22] SUR, P., CHEN, Y. and CANDÈS, E. J. (2020). The likelihood ratio test in high-dimensional logistic regression is asymptotically a rescaled chi-square. *Probab. Theory Related Fields* **175** 487–558. MR4009715

[23] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256

[24] VISSCHER, P. M. (2006). A note on the asymptotic distribution of likelihood ratio tests to test variance components. *Twin Research and Human Genetics* **9** 490–495.

[25] WILKS, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Stat.* **9** 60–62.

[26] WIT, E. C. and BAKEWELL, D. J. (2012). Borrowing strength: A likelihood ratio test for related sparse signals. *Bioinformatics* **28** 1980–1989.

[27] ZOLOTAREV, V. M. (1983). Probability metrics. *Teor. Veroyatn. Primen.* **28** 264–287. MR0700210