Bounds for the normal approximation of the maximum likelihood estimator

Andreas Anastasiou

Supervisor: Professor Gesine Reinert

Department of Statistics

University of Oxford

This dissertation is submitted for the degree of

Doctor of Philosophy

Jesus College

Trinity Term, 2016

To my parents, Kyriacos and Maria, and my siblings, Anna and Tassos.

Acknowledgements

The first and most important person to thank is my supervisor, Professor Gesine Reinert. I am deeply grateful to her for the guidance, encouragement and support she offered me throughout my time as her student.

I would like to thank Professor Christophe Ley for a fruitful collaboration which lead to a paper, of which the results can be found in Chapter 5 of the thesis.

I would also like to thank Dr Robert Gaunt for various insightful comments and suggestions.

Last, but not of least importance, I would like to thank my parents who encouraged me to continue with postgraduate studies.

The research is conducted with the support of a Teaching Assistantship Bursary from the Department of Statistics, University of Oxford, and the Engineering and Physical Sciences Research Council (EPSRC) grant EP/K503113/1.

Abstract

The asymptotic normality of the maximum likelihood estimator (MLE) under regularity conditions is a long established and famous result. This is a qualitative result and the assessment of such a normal approximation is our main interest. For this task we partly use Stein's method, which is a probabilistic technique that can be used to explicitly measure the distributional distance between two distributions. Since its first appearance in 1972, the method has been developed for various distributions; here we use the results related to Stein's method for normal approximation.

In this thesis, we derive explicit upper bounds on the distributional distance between the distribution of the MLE and the normal distribution. First, the focus is on independent and identically distributed random variables from both discrete and continuous single-parameter distributions with particular attention to exponential families. For discrete distributions, the case where the MLE can be on the boundary of the parameter space is treated through a perturbation approach, which allows us to obtain bounds on the distributional distance of interest. The bounds are of order $\frac{1}{\sqrt{n}}$, where n is the number of observations. Simulation-based results are given to illustrate the power of the bound. Furthermore, often the MLE can not be obtained analytically and optimisation methods (such as the Newton-Raphson algorithm) are used. Even in such cases, order $\frac{1}{\sqrt{n}}$ bounds are given for the distributional distance related to the MLE.

The case of multi-parameter distributions follows smoothly after the detailed discussion

related to a scalar parameter. Apart from extending our approach to a multi-parameter setting, we also cover the case of independent but not necessarily identically distributed (i.n.i.d.) random vectors with specific focus on the widely applicable linear regression models.

Going back to the single-parameter setting a different approach to get an upper bound on the distributional distance between the distribution of the MLE and the normal distribution, based on the Delta method, is also developed. The MLE for a Generalised Gamma distribution gives an illustration of the results obtained through this Delta method approach.

Finally, we relax the independence assumption and results for the case of locally dependent random variables are obtained. An example of correlated sums of normally distributed random variables illustrates the bounds. Again, results that do not require an analytic expression of the MLE to be known are given. We end this thesis with ideas currently in progress and further open research questions.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Motivation for the thesis

Many statistical procedures rely on asymptotic results as exact finite-sample results are often difficult to derive. Yet a sample size is never infinite. In recent years, there has been a growing interest in assessing the quality of such approximations. The Central Limit Theorem (CLT), which is considered as one of the best known asymptotic developments in the area of probability and statistics, states that the sum of a large number of independent random variables, all of which have the same mean and variance, or in many alternative cases, random variables with specific types of dependence, will tend to be approximately normally distributed regardless of the random variables' exact distribution. The result can be extended to multidimensional random vectors.

Using the CLT, the Weak Law of Large Numbers and Taylor expansions, one is able to show that the asymptotic distribution of the maximum likelihood estimator (MLE) is normal under sufficient regularity conditions. This is an asymptotic result and the quality of the approximation may depend on the sample size as well as on the distribution of the random variables. The assessment of the quality of the normal approximation

related to the distribution of the MLE has been the motivation behind the research presented in this thesis.

The asymptotic normality of the MLE was first discussed in Fisher (1925). It is a fundamental qualitative result. Our aim is to find upper bounds on the distributional distance between the distribution of the MLE and the normal distribution under various general frameworks. We cover a variety of different scenarios related to the dimensionality of the MLE and also the dependence structure of possibly heterogeneous random variables (or vectors). Complementing the qualitative asymptotic result with a quantitative statement obtained from our bounds can be helpful to assess whether using the limiting distribution of the MLE is an acceptable approximation or not. From the opposite point of view, the results presented in this thesis can save both money and time by giving a good indication on whether a larger sample size is indeed necessary, for a good approximation to hold. These quantitative results for the MLE which are developed and explained in this thesis have already influenced researchers; see for example Pinelis and Molzon (2016).

The wide applicability of the maximum likelihood estimation method adds to the importance of our results. Among others, an MLE is used in ordinary and generalised linear models, time series analysis and a large number of other situations related to hypothesis testing and confidence intervals. They appear in a broad category of different fields, such as econometrics, computational biology and data modelling in physics and psychology. In addition to the usefulness of the MLE, in many cases its existence is easily ensured. Also, apart from the attractive asymptotic distribution property, the MLE satisfies many other exact sample size properties such as the functional invariance; if $\hat{\theta}$ is the MLE for θ, then the MLE for $g(\theta)$ is $g(\hat{\theta})$, with $g(\theta)$ being any transformation of the parameter. The numerous and useful properties of the MLE justify the fact that it is considered to be the most powerful and thus widely used estimation method.

1.2 Stein's method as a useful tool

For our results, we partly employ Stein's method, which was first introduced in 1972 by Charles Stein. It is a quite robust probabilistic technique that can measure the error in the approximation of one known distribution by another in a wide range of metrics. For the scope of this thesis, Stein's method for normal approximation is applied to a random variable

$$W = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i,$$

where $X_i, i = 1, 2, \ldots, n$ are mean zero random variables such that $\mathrm{Var}(W) = 1$. We will obtain upper bounds on the quantity

$$|\mathrm{E}[h(W)] - \mathrm{E}[h(Z)]|, \tag{1.1}$$

where Z follows the standard normal distribution and h belongs to a class of test functions, H. Depending on H, we get results in different metrics. Throughout this thesis, we mainly restrict ourselves to smooth test functions. Our approach is related to the work from Berry and Esséen for the known traditional convergence in distribution (Kolmogorov distance) result, which is of the form

$$\sup_{z \in \mathbb{R}} |\mathbb{P}(W \leq z) - \mathbb{P}(Z \leq z)| \leq C \frac{\mathrm{E}|X_1|^3}{\sqrt{n}}. \tag{1.2}$$

Esséen's original bound on C was 7.59, but has been decreased over the years and the latest improvement is $C \leq 0.4748$ as shown by Shevtsova (2014) using Fourier techniques. To see the result in (1.2) from Stein's method perspective, just take as test functions the class of indicator functions of half-lines $\left\{ \mathbb{1}_{[.\leq x]}, x \in \mathbb{R} \right\}$. By doing that, the expectations in (1.1) become cumulative distribution functions leading to the result in (1.2).

The Stein's method results are used here to give upper bounds on the distributional distance related to a sum of random variables. Since the MLE is not in general a sum, we find a way to represent the standardised MLE such that it contains a quantity which is a sum, plus a term which we can control. For this purpose we use second-order Taylor expansions of the first derivative of the log-likelihood function evaluated at the MLE, about the true value of the parameter. This approach gives as sum component the first derivative of the log-likelihood function, namely the score function, evaluated at the true value of the unknown parameter. We need to take the dependence structure of the random variables into account and use the appropriate results from Stein's method to bound the score function. The remainder terms get bounded using other techniques based mainly on Taylor expansions, conditional expectations and known probability inequalities, such as the Markov and Cauchy-Schwarz inequalities. This is the main approach followed in this thesis. However, another idea based on the widely known Delta method is also developed and used to give bounds on the required distributional distance between the exact distribution of the MLE and its limiting normal distribution in cases where the MLE can be expressed as a function of a sum of random variables. Stein's method and Taylor expansions are still our main tools.

It is reasonable to compare Stein's method over other known techniques that can give answers on distributional distance, such as moment generating or characteristic functions. Although more details are given later, perhaps the biggest advantage of Stein's method compared to these techniques is its appropriateness on various types of dependence structure. Another advantage of Stein's method is the flexibility to manipulate the distribution using the random quantities of which the quantity under interest is composed. In contrast, characteristic function methods are based on manipulating distributions through their transforms, leading to loss of probabilistic intuition. The immediate explicit upper bound on the distributional distance is another advantage (Reinert, 1998).

1.3 Outline of the thesis

In Chapter 2, some background information for the CLT, the MLE and Stein's method is provided. The notation used throughout the thesis is explained and assumptions under which existence and uniqueness of the MLE is secured are also given; these are split into two cases depending on whether the parameter space is open and connected, or compact. We then explain Fisher's original asymptotic result for the distribution of the MLE, under its most basic form of independent and identically distributed (i.i.d.) random variables. The sufficient regularity conditions under which this asymptotic result holds are also discussed and briefly explained.

The work presented in Chapter 3 is an already accepted paper (with Gesine Reinert) for *Bernoulli*. We begin the chapter presenting the bounded Wasserstein distance, which is the metric used for the single-parameter case. The relationship of this metric with the Kolmogorov distance, which leads to insights for confidence intervals, is given before proceeding to the main result of the chapter for the bounds on the distributional distance of interest. Special focus on the broad classification of exponential family distributions is given. Simulation-based results related to the exponential distribution illustrate the behaviour of the bounds through a graphical representation. We cover both the canonical and non-canonical parametrisation and compare these two situations. In some discrete distributions there is positive probability of the MLE being on the boundary of the parameter space, which leads to problems related to the applicability of Taylor expansions. We overcome this problem using a perturbation of both the parameter and the data, forcing them to be interior to the parameter space. The Poisson distribution with parameter space, Θ, the semi-closed interval $[0, \infty)$ serves as an illustration of these results. Under further assumptions, Chapter 3 gives a procedure to get an upper bound that does not contain any terms related to the MLE. The approach is based on finding upper bounds for the mean squared error. The new bound obtained is still applicable

in cases where an analytic expression of the MLE is not available. The results are applied to the Beta distribution with one of the two shape parameters being unknown. Simulation-based results are also given.

Chapter 4 treats the multi-parameter case for independent but not necessarily identically distributed (i.n.i.d.), high-dimensional random vectors. This is a single-author paper submitted to the *Journal of Multivariate Analysis*. We explain how an adaptation of the method developed in Chapter 3 for the case of a vector parameter is feasible. A multivariate Stein's method result by Reinert and Röllin (2009), multivariate Taylor expansions and conditional expectations lead to the final upper bound. We give a general upper bound but also a specific upper bound for linear regression models and we apply the specific bound to the simplest case of the straight-line regression with two unknown parameters (the intercept and the slope). The special case where the random vectors are identically distributed is also covered; simpler regularity conditions than those for the i.n.i.d. case are given in order for the normal approximation to hold. Following a similar approach as the one followed in the i.n.i.d. case we obtain the required, simpler upper bound. As an example, the normal distribution with both mean and variance unknown is treated. Furthermore, we give an upper bound when the vector MLE is not known explicitly. The approach is again based on firstly bounding the mean squared error of the vector MLE. The bound might seem slightly complicated and tedious to calculate. However, for the Beta distribution with both parameters being unknown, an expression for the bound is easily obtained.

Chapter 5 is based on joint work with Christophe Ley which has been submitted as a paper to the *Latin American Journal of Probability and Mathematical Statistics* (*ALEA*). This chapter uses the Delta method to obtain an expression of the MLE as a sum of random variables. More precisely, we use a one-to-one differentiable function in such a way that when applied to the MLE it gives a sum of random variables, allowing

the use of Stein's method for this quantity. There will also be error terms and those will be bounded through alternative techniques, based again on our main tool; Taylor expansions. We give the bound for the situation of a scalar parameter; special focus is again given to exponential families. The results are then applied to the generalised gamma distribution, of which the exponential distribution is a special case. This example allows a comparison of the results in this chapter with the results obtained in Chapter 3 for the exponential distribution.

In Chapter 6, the independence assumption of the previous chapters is relaxed to a local dependence structure between the random variables. We obtain bounds for the distributional distance of interest under this dependence structure. In an example related to normally distributed locally dependent random variables the bound behaves very well. In addition, as in the previous chapters, under further assumptions we give bounds when the MLE is not explicitly known.

Chapter 7 includes a summary of the results presented in this thesis and some ideas for future work. We cover applications as well as further theoretical extension of the results obtained in this thesis.

Chapter 2

Background

This chapter gives the background for the work presented in this thesis. In Section 2.1, a brief description of the Central Limit Theorem is given as it is the main tool used to prove the asymptotic normality of the MLE. In Section 2.2, we explain the main asymptotic result for the distribution of the MLE, which then incites us to find a way to assess the quality of such an approximation. For this purpose, Stein's method for normal approximation is employed. An introduction to the main results of this effective probabilistic method is fundamental for the understanding of the following chapters and is given in Section 2.3.

2.1 Central Limit theorem

The Central Limit Theorem was first discussed in its initial form related to Bernoulli random variables in a notable article by Abraham De Moivre in 1733 and since then it has become a cornerstone in probability and statistical theory. In general words, it states that the suitably scaled sample mean of a large number of random variables follows approximately the normal distribution regardless of the distribution of the random variables. As long as the assumptions are satisfied, the CLT has a number of

alternative versions depending on the dependence structure of the random variables, their dimensionality as well as whether we have identically distributed random variables (vectors) or not. From now on, $\mathrm{N}\left(\mu,\sigma^2\right)$ denotes the normal distribution with mean μ and variance σ^2. Below, we give the CLT in its classical form of i.i.d. random variables, known also as Lindeberg-Lévy CLT.

Theorem 2.1. *Let $X_1, X_2, \ldots$ be i.i.d. random variables with mean $\mu \in \mathbb{R}$ and variance $0 < \sigma^2 < \infty$. For $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $K \sim \mathrm{N}\left(0,\sigma^2\right)$,*

$$\sqrt{n}\left(\bar{X} - \mu\right) \xrightarrow[n\to\infty]{\mathrm{d}} K, \tag{2.1}$$

where $\xrightarrow{\mathrm{d}}$ denotes convergence in distribution.

The case $\sigma^2 = 0$ could be included by interpreting $\mathrm{N}(0,0)$ as a point mass at zero. A flavour of the result in (2.1) could be given through Abraham De Moivre's original article. We take a fair coin and we flip it many times. The distribution of each trial is Bernoulli with mean the probability to get 'heads', which is $\frac{1}{2}$. We add 1 (starting from zero) whenever the outcome of the coin is 'heads'. Then the distribution of the resulting sum over the total number of trials (the average) is well-approximated by the normal distribution with mean equal to $\frac{1}{2}$ and variance equal to $\frac{1}{4n}$. It is notable that through the CLT the exact distribution of the suitably scaled sum of discrete random variables can be approximated by the normal distribution, which is a continuous probability distribution.

The CLT is an asymptotic result and provides only an approximation for finite sample sizes. The quality of this approximation may depend on the magnitude of the sample size as well as the distribution of the random variables of which we take the sum. A quantification of this qualitative result is given through the celebrated Berry-Esséen theorem, which just requires the third central moment of $X_i, i = 1,2,\ldots,n$ to exist and to be finite. An upper bound of order $\frac{1}{\sqrt{n}}$ on the Kolmogorov distance between the exact distribution of the sample mean and the limiting normal distribution can be obtained; see

(1.2) for the bound. Generalising this result, we can obtain upper bounds on a variety of distributional distances using Stein's method; see Section 2.3 for a review of the method.

As our main interest is the MLE, which is not in general a sum of random variables, the CLT can not be directly applied to get the asymptotic normality of the MLE. The next section introduces the maximum likelihood estimation method and gives sufficient regularity conditions under which the asymptotic normality of the MLE holds.

2.2 Maximum likelihood estimation

In Statistics, maximum likelihood estimation is a widely used method of estimating d parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_d)$ found in a parametric statistical model. Having a fixed set of data and a defined statistical model, the maximum likelihood estimation gives a set of values $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})$ for the parameters of the model which maximises the likelihood function. Such a vector $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})$, whenever it exists, is called a maximum likelihood estimator (MLE).

To set the scene in the simplest scenario, suppose that $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ is a sample of n i.i.d. random variables with a joint density function $f(\boldsymbol{x}|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is an unknown parameter. For $X_i = x_i$, $i = 1, 2, \ldots, n$ being some observed values, we define the so-called likelihood function by $L(\boldsymbol{\theta}; \boldsymbol{x}) = f(\boldsymbol{x}|\boldsymbol{\theta})$, where $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$. If the distribution of the random variables is discrete (instead of a density, we have a probability mass function), then the likelihood function is the probability of observing the given data $\boldsymbol{x}$. Let $\boldsymbol{\theta_0} = (\theta_{0_1}, \theta_{0_2}, \ldots, \theta_{0_d})$, be the true (still unknown) value of the parameter vector $\boldsymbol{\theta}$. The aim is to find an estimate of $\boldsymbol{\theta_0}$ from the data $\boldsymbol{x}$ by maximising the likelihood. Rather than maximising the likelihood function, it is often more convenient to work with its logarithm, called the log-likelihood function and denoted by $l(\boldsymbol{\theta}; \boldsymbol{x})$. The logarithmic form makes calculations easier, particularly since $L(\boldsymbol{\theta}; \boldsymbol{x})$

will often (specifically in the case of independent data points) be a product of marginal probability density (mass) functions.

For many models the MLE exists and is unique; this is known as the 'regular' case. For a number of statistical models, however, uniqueness or even existence of the MLE is not secured. In the following proposition, we specify a number of sufficient conditions for both existence and uniqueness of the MLE, as found in Makelainen et al. (1981).

Proposition 2.1. *Let the likelihood function $L(\boldsymbol{\theta};\boldsymbol{x})$ be a twice continuously differentiable function with $\boldsymbol{\theta}$ varying in a connected open subset $\Theta \subset \mathbb{R}^d$, called the parameter space. Suppose that:*

(i) $\lim_{\boldsymbol{\theta}\to\partial\Theta} L(\boldsymbol{\theta};\boldsymbol{x}) = 0$, *where $\partial\Theta$ is the boundary of the parameter space Θ,*

(ii) With $\frac{\partial}{\partial\theta_i}$ *denoting now partial derivatives, the Hessian matrix*

$$\boldsymbol{H}(\boldsymbol{\theta};\boldsymbol{x}) = \left\{ \frac{\partial^2}{\partial\theta_i\partial\theta_j} L(\boldsymbol{\theta};\boldsymbol{x}) \right\}_{i,j=1,\cdots,d}$$

of second partial derivatives is negative definite at every point $\boldsymbol{\theta} \in \Theta$ for which the gradient vector

$$\nabla(L(\boldsymbol{\theta};\boldsymbol{x})) = \left\{ \tfrac{\partial}{\partial\theta_i} L(\boldsymbol{\theta};\boldsymbol{x}) \right\}_{i=1,\cdots,d}$$

vanishes. Then

1. *there is a unique maximum likelihood estimate $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{x}) \in \Theta$, and*
2. *the likelihood function attains*
   - *(a) no other maxima in Θ;*
   - *(b) no minima or other stationary points in Θ;*
   - *(c) its infimum value 0 on the boundary $\partial\Theta$ and nowhere else.*

If only $\nabla(L(\boldsymbol{\theta};\boldsymbol{x})) = \mathbf{0}$ is assumed, then there may be multiple solutions (stationary points) (Billingsley, 1961).

Remark 2.1. *Condition* (ii) *can be translated in terms of the log-likelihood function* $l(\boldsymbol{\theta};\boldsymbol{x})$. *Notice that*

$$\nabla(l(\boldsymbol{\theta};\boldsymbol{x})) = \frac{\nabla(L(\boldsymbol{\theta};\boldsymbol{x}))}{L(\boldsymbol{\theta};\boldsymbol{x})}.$$

Therefore, the vector of the first order partial derivatives of the likelihood function, vanishes whenever the respective vector related to the log-likelihood function vanishes, as long as $0 < L(\boldsymbol{\theta};\boldsymbol{x}) < \infty$. *In addition, the Hessian matrix for* $l(\boldsymbol{\theta};\boldsymbol{x})$ *is*

$$\boldsymbol{H_l}(\boldsymbol{\theta};\boldsymbol{x}) = \frac{\boldsymbol{H}(\theta;\boldsymbol{x})L(\boldsymbol{\theta};\boldsymbol{x}) - [\nabla(L(\boldsymbol{\theta};\boldsymbol{x}))]^2}{[L(\boldsymbol{\theta};\boldsymbol{x})]^2}.$$

For $\nabla(L(\boldsymbol{\theta};\boldsymbol{x})) = 0$, *notice that* $\boldsymbol{H_l}(\boldsymbol{\theta};\boldsymbol{x}) = \frac{\boldsymbol{H}(\theta;\boldsymbol{x})}{L(\boldsymbol{\theta};\boldsymbol{x})}$ *and thus* $\boldsymbol{H}(\boldsymbol{\theta};\boldsymbol{x})$ *is negative definite whenever* $\boldsymbol{H_l}(\boldsymbol{\theta};\boldsymbol{x})$ *is a negative definite matrix and* $L(\boldsymbol{\theta},\boldsymbol{x}) > 0$. *Therefore, for* $L(\boldsymbol{\theta},\boldsymbol{x}) > 0$ *we can easily express condition* (ii) *of Proposition 2.1 in terms of the log-likelihood function, which simplifies the calculations. It becomes*

(ii)' With $\frac{\partial}{\partial\theta_i}$ *denoting partial derivatives, the Hessian matrix of the log-likelihood function*

$$\boldsymbol{H_l}(\boldsymbol{\theta};\boldsymbol{x}) = \left\{\frac{\partial^2}{\partial\theta_i\partial\theta_j}l(\boldsymbol{\theta};\boldsymbol{x})\right\}_{i,j=1,\cdots,d}$$

of second partial derivatives is negative definite at every point $\boldsymbol{\theta} \in \Theta$ *for which the gradient vector*

$$\nabla(l(\boldsymbol{\theta};\boldsymbol{x})) = \left\{\frac{\partial}{\partial\theta_i}l(\boldsymbol{\theta};\boldsymbol{x})\right\}_{i=1,\cdots,d}$$

vanishes.

Proposition 2.1 provides sufficient conditions for the existence and uniqueness of the MLE when the parameter space, Θ, is an open and connected subset of $\mathbb{R}^d$. However, in some statistical models, the parameter space is compact. We use Kiefer (2008) for

the following proposition related to the existence and uniqueness of the MLE when it is obtained from parametric statistical models with compact parameter space.

Proposition 2.2. *If the parameter space Θ is compact and if the likelihood function is continuous on Θ then there exists a MLE. In addition, if the parameter space Θ is also convex and the likelihood function is strictly concave in θ, then the MLE is unique.*

The first result, related to the existence of the MLE, arises as a direct application of the Weierstrass Extreme Value Theorem, which states that if a real valued function f is continuous in a closed and bounded set, then it attains a maximum and a minimum, each at least once.

For this chapter and Chapters 3, 5 and 6, unless otherwise stated, the parameter θ is assumed to be scalar. The vector-parameter case is covered in Chapter 4. In the single-parameter setting, we denote the derivatives of the log-likelihood function with respect to θ, by $l'(\theta;\boldsymbol{x}), l''(\theta;\boldsymbol{x}), \cdots, l^{(j)}(\theta;\boldsymbol{x})$, for j any integer greater than 2. Unless otherwise stated, the parameter varies in an open interval (a,b), where $a,b \in \mathbb{R} \cup \{-\infty,\infty\}$ and $a < b$. Following (Casella and Berger, 2002, p.516), we also make the following assumptions:

(R1) the parameter is identifiable, meaning that if $\theta \neq \theta'$, then $\exists x : f(x|\theta) \neq f(x|\theta')$;

(R2) the density $f(x|\theta)$ is three times differentiable with respect to θ, the third derivative is continuous in θ and $\int f(x|\theta)\,\mathrm{d}x$ can be differentiated three times under the integral sign;

(R3) for any $\theta_0 \in \Theta$ and for S denoting the support of $f(x|\theta)$, there exists a positive number ε_0 and a function $M(x)$ (both of which may depend on θ_0) such that

$$\left|\frac{\mathrm{d}^3}{\mathrm{d}\theta^3}\log f(x|\theta)\right| \leq M(x) \quad \forall x \in S,\ \forall \theta \in \Theta \text{ such that } |\theta-\theta_0| < \varepsilon_0,$$

with $\mathrm{E}_{\theta_0}[M(X)] < \infty$;

(R4) $i(\theta_0) \neq 0$, where $i(\theta)$ is the expected Fisher Information number for one random variable.

The requirement (R2) that $\int f(x|\theta)\,\mathrm{d}x$ can be differentiated three times under the integral sign means that $\int_{\mathbb{R}} \frac{\mathrm{d}^j}{\mathrm{d}\theta^j} f(x|\theta)\,\mathrm{d}x = \frac{\mathrm{d}^j}{\mathrm{d}\theta^j} \int_{\mathbb{R}} f(x|\theta)\,\mathrm{d}x = 0, j \in \{1,2,3\}$. This condition ensures that if the expressions exist, then $\mathrm{E}_\theta[l'(\theta;\boldsymbol{X})] = 0$ and $\mathrm{Var}_\theta[l'(\theta;\boldsymbol{X})] = n\,i(\theta)$. The motivation of the work presented in this thesis are the results given in Theorem 2.2. The efficiency and asymptotic normality of the MLE have first been discussed in Fisher (1925). Here the i.i.d. case is presented; see Theorem 4.1 for the case of independent but not necessarily identically distributed random vectors. In addition, for dependent random variables the asymptotic result is given in Theorem 6.2. Here, $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ is the random vector and the MLE, $\hat{\theta}_n(\boldsymbol{X})$, is of course also random as a function of $\boldsymbol{X}$.

Theorem 2.2. *Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with probability density (or mass) function $f(x_i|\theta)$, where θ is the scalar parameter. Assume that the MLE exists, it is unique and (R1)-(R4) are satisfied. Then for $Z \sim \mathrm{N}(0,1)$*

$$(a)\ \frac{1}{\sqrt{n}} l'(\theta_0;\boldsymbol{X}) \xrightarrow[n\to\infty]{\mathrm{d}} \sqrt{i(\theta_0)}Z, \qquad (b)\ \sqrt{n\,i(\theta_0)}(\hat{\theta}_n(\boldsymbol{X}) - \theta_0) \xrightarrow[n\to\infty]{\mathrm{d}} Z. \tag{2.2}$$

Thus, in the case of i.i.d. random variables, under the regularity conditions (R1)-(R4), the asymptotic distribution of the suitably scaled MLE is the standard normal distribution. However, this is an asymptotic result and we will never have an infinite sample size. We will assess the quality of this approximation using partly Stein's method.

2.3 Stein's method

Stein's method was first introduced by Charles Stein in Stein (1972), while the monograph Stein (1986) explains in detail the method and it is in our opinion the most notable contribution. Stein's method is a powerful technique to assess whether a random variable, W, has a distribution close to a target distribution. The distance between the distribution of W and the target one is measured using a Zolotarev-type distance. If F, G are two random variables with values in $\mathbb{R}$ and H is a class of separating functions, then a Zolotarev-type distance between the laws of F and G, induced by H, is given by the quantity

$$d_H(F,G) = \{|\mathrm{E}[h(F)] - \mathrm{E}[h(G)]| : h \in H\}. \tag{2.3}$$

For example, taking $H = H_K = \left\{\mathbb{1}_{[.\leq x]}, x \in \mathbb{R}\right\}$ in (2.3), gives the Kolmogorov-Smirnov distance

$$d_K(F,G) = \sup_{z\in\mathbb{R}} |\mathbb{P}(F \leq z) - \mathbb{P}(G \leq z)|,$$

which is the maximum distance between two cumulative distribution functions. With

$$H_W = \{h : \mathbb{R} \to \mathbb{R} : |h(x) - h(y)| \leq |x - y|\}, \tag{2.4}$$

the set of Lipschitz functions with constant equal to one, (2.3) gives the Wasserstein distance,

$$d_W(F,G) = \sup\{|\mathrm{E}[h(F)] - \mathrm{E}[h(G)]| : h \in H_W\}. \tag{2.5}$$

Apart from just giving the general results, in the thesis we also focus on the Wasserstein as well as on the bounded Wasserstein (or Fortet-Mourier) distance, which is obtained for

$$H = H_{bW} = \left\{h : \mathbb{R} \to \mathbb{R} : \|h\|_{\mathrm{Lip}} + \|h\| \leq 1\right\}, \tag{2.6}$$

where from now on $\|.\|$ denotes the supremum norm ($\|.\|_\infty$) and $\|h\|_{\text{Lip}} = \sup_{\substack{x\neq y \\ x,y\in\mathbb{R}}} \frac{|h(x)-h(y)|}{|x-y|}$.
The bounded Wasserstein distance is denoted from now on by

$$d_{bW}(F,G) = \sup\{|\mathrm{E}[h(F)] - \mathrm{E}[h(G)]| : h \in H_{bW}\}, \tag{2.7}$$

with H_{bW} as in (2.6); see for example Nourdin and Peccati (2012). Using Rademacher's theorem, if $\|h\|_{\text{Lip}} \leq 1$, then h is differentiable almost everywhere, with h' denoting its derivative. Hence first-order Taylor expansion is possible.

An alternative to Stein's method are characteristic functions, which are quite powerful in examining if the distributions of two random variables are similar. Closeness of the characteristic functions implies closeness of the distributions. However, such methods lead to loss of probabilistic intuition as they rely on transformations of the quantity of interest. In addition, these methods are not easy to apply in the presence of a dependence structure; for limit theorems for sums of $m-$dependent random variables see Heinrich (1982). On the other hand, Stein's method is based on a characterisation of the target distribution and this makes it easier to manipulate the distribution through random quantities of which W is composed. Therefore, probabilistic intuition is not lost. In addition, Stein's method is quite powerful in the presence of dependence, providing explicit upper bounds through coupling constructions such as the exchangeable pair, the zero-bias and the size bias couplings.

Stein's method for normal approximation–the target distribution is the normal–is based on the well-known characterisation of the normal distribution that $K \sim \mathrm{N}(0,\sigma^2)$ if and only if

$$\mathrm{E}\left[Kt(K)\right] - \sigma^2 \mathrm{E}\left[t'(K)\right] = 0,$$

for all absolutely continuous and differentiable functions t for which the above expectations exist. If the distribution of the random variable, W, is close to the $\mathrm{N}(0,\sigma^2)$

distribution, then both $|\mathrm{E}[h(W)] - \mathrm{E}[h(K)]|$ and $\left|\mathrm{E}[Wt(W)] - \sigma^2 \mathrm{E}[t'(W)]\right|$ should be small, where h is a test function. These two expressions, more precisely the functions h and t, are now related through the so-called Stein equation for the $\mathrm{N}(0, \sigma^2)$ distribution

$$\sigma^2 t'(w) - wt(w) = h(w) - \mathrm{E}[h(K)]. \tag{2.8}$$

In order to obtain (2.3) for a given function h, it suffices to find the solution t_h of (2.8), evaluate $\sigma^2 t_h'(w) - wt_h(w)$ at W and take the absolute value of the expectation.

Lemma 2.1. *(Stein, 1972). For $Z \sim \mathrm{N}(0,1)$, the unique bounded solution of the Stein equation in* (2.8) *for $\sigma^2 = 1$ is given by*

$$f_h(w) = -\mathrm{e}^{\frac{w^2}{2}} \int_w^\infty \mathrm{e}^{-\frac{x^2}{2}} \left(h(x) - \mathrm{E}[h(Z)]\right) \mathrm{d}x. \tag{2.9}$$

Proof. See Chen et al. (2011). The idea is to solve a first order ordinary differential equation. Multiplying both sides of (2.8) for $\sigma^2 = 1$ by the integrating factor, which is $\mathrm{e}^{-\frac{w^2}{2}}$, yields

$$\frac{\mathrm{d}}{\mathrm{d}w}\left(f(w)\mathrm{e}^{-\frac{w^2}{2}}\right) = \mathrm{e}^{-\frac{w^2}{2}} \left(h(w) - \mathrm{E}[h(Z)]\right).$$

The next step is to integrate both sides to get, for $c \in \mathbb{R}$, that

$$\begin{aligned} f_h(w)\mathrm{e}^{-\frac{w^2}{2}} &= \int_{-\infty}^w \mathrm{e}^{-\frac{x^2}{2}} \left(h(x) - \mathrm{E}[h(Z)]\right) \mathrm{d}x + c \\ \Rightarrow f_h(w) &= \mathrm{e}^{\frac{w^2}{2}} \int_{-\infty}^w \mathrm{e}^{-\frac{x^2}{2}} \left(h(x) - \mathrm{E}[h(Z)]\right) \mathrm{d}x + c\mathrm{e}^{\frac{w^2}{2}}. \end{aligned}$$

Now, $|f_h(w)|$ is bounded if and only if $c = 0$. Thus, using also that

$$\int_{-\infty}^\infty \mathrm{e}^{-\frac{x^2}{2}} \left(h(x) - \mathrm{E}[h(Z)]\right) \mathrm{d}x = 0,$$

yields that the unique bounded solution is the one in (2.9). □

If $h : \mathbb{R} \to \mathbb{R}$, is absolutely continuous and bounded then the solution f_h in (2.9) satisfies

$$\|f_h''\| \leq 2\|h'\|, \tag{2.10}$$

see Chen et al. (2011), (2.13) in p.16. In the general case $\left(\sigma^2 \text{ not necessarily equal to } 1\right)$ the unique bounded solution, $t_h(w)$, of the Stein equation in (2.8) satisfies

$$t_h(w) = \frac{1}{\sigma} f_h\left(\frac{w}{\sigma}\right).$$

Thus, using (2.10),

$$\|t_h''\| \leq 2\frac{\|h'\|}{\sigma^3}. \tag{2.11}$$

A general result lies at the centre of our work and is given in Lemma 2.2.

Lemma 2.2. *Let $Y_1, Y_2, \ldots, Y_n$ be i.i.d. random variables with $\mathrm{E}(Y_i) = 0$, $\mathrm{Var}(Y_i) = \sigma^2 > 0$ and $\mathrm{E}|Y_i|^3 < \infty$. Let $W = \frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i$ and $K \sim \mathrm{N}(0, \sigma^2)$. Then for any function $h : \mathbb{R} \to \mathbb{R}$, such that h is absolutely continuous and bounded*

$$|\mathrm{E}[h(W)] - \mathrm{E}[h(K)]| \leq \frac{\|h'\|}{\sqrt{n}}\left(2 + \frac{1}{\sigma^3}\left[\mathrm{E}|Y_1|^3\right]\right). \tag{2.12}$$

Proof. The method of the proof is standard using the famous 'leave one out' approach and follows the steps in (Chen et al., 2011, p.5), where a sketch of the proof for $\sigma^2 = 1$ is given. We will show (2.12) by finding an upper bound on $|\sigma^2\mathrm{E}[t_h'(W)] - \mathrm{E}[Wt_h(W)]|$. Let $W_i = W - \frac{Y_i}{\sqrt{n}} = \frac{1}{\sqrt{n}}\sum_{\substack{j=1 \\ j\neq i}}^n Y_j$. A second-order Taylor expansion of $t_h(W)$ about W_i gives

$$\begin{aligned}
\mathrm{E}\left[Wt_h(W)\right] &= \frac{1}{\sqrt{n}}\mathrm{E}\left[\sum_{i=1}^n Y_i t_h(W)\right] \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathrm{E}\left[Y_i(t_h(W_i) + (W - W_i)t_h'(W_i) + \frac{1}{2}(W - W_i)^2 t_h''(W^*))\right],
\end{aligned}$$

where W^* is between W and W_i. Having $W - W_i = \frac{Y_i}{\sqrt{n}}$, with Y_i and W_i being independent yields

$$\begin{aligned}\mathrm{E}\left[Wt_h(W)\right] &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathrm{E}[Y_i]\mathrm{E}\left[t_h(W_i)\right] + \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[Y_i^2\right]\mathrm{E}\left[t_h'(W_i)\right] + \frac{1}{2n^{\frac{3}{2}}}\sum_{i=1}^{n}\mathrm{E}\left[Y_i^3 t_h''(W^*)\right] \\ &= \frac{\sigma^2}{n}\sum_{i=1}^{n}\mathrm{E}\left[t_h'(W_i)\right] + \frac{1}{2n^{\frac{3}{2}}}\sum_{i=1}^{n}\mathrm{E}\left[Y_i^3 t_h''(W^*)\right].\end{aligned} \tag{2.13}$$

Using (2.13),

$$\begin{aligned}&\left|\sigma^2\mathrm{E}\left[t_h'(W)\right] - \mathrm{E}[Wt_h(W)]\right| \\ &= \left|\sigma^2\mathrm{E}\left[t_h'(W)\right] - \frac{\sigma^2}{n}\sum_{i=1}^{n}\mathrm{E}\left[t_h'(W_i)\right] - \frac{1}{2n^{\frac{3}{2}}}\sum_{i=1}^{n}\mathrm{E}\left[Y_i^3 t_h''(W^*)\right]\right| \\ &= \left|\frac{\sigma^2}{n}\sum_{i=1}^{n}\mathrm{E}\left[t_h'(W) - t_h'(W_i)\right] - \frac{1}{2n^{\frac{3}{2}}}\sum_{i=1}^{n}\mathrm{E}\left[Y_i^3 t_h''(W^*)\right]\right| \\ &\leq \frac{\sigma^2}{n^{\frac{3}{2}}}\|t_h''\|\sum_{i=1}^{n}\mathrm{E}\left|Y_i\right| + \frac{1}{2n^{\frac{3}{2}}}\|t_h''\|\sum_{i=1}^{n}\mathrm{E}\left|Y_i^3\right|,\end{aligned}$$

by first-order Taylor expansion of $t_h'(W)$ about W_i. Applying the Cauchy-Schwarz inequality to the first term and using (2.11) yields

$$\begin{aligned}|\sigma^2\mathrm{E}[t_h'(W)] - \mathrm{E}[Wt_h(W)]| &\leq \frac{\sigma^3}{\sqrt{n}}\|t_h''\| + \frac{1}{2\sqrt{n}}\|t_h''\|\mathrm{E}\left|Y_1^3\right| \\ &\leq \frac{\|h'\|}{\sqrt{n}}\left(2 + \frac{1}{\sigma^3}\left[\mathrm{E}\left|Y_1\right|^3\right]\right).\end{aligned}$$

□

Remark 2.2. *Using* $Y_i = \frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_i|\theta_0)$*, notice that* (2.12) *can be used to assess the quality of* (a) *in* (2.2)*.*

Lemma 2.2 gives upper bounds in the case of independent random variables. However, sometimes independence is not secured and thus we need to use results related to dependent random variables. In Chapter 6 we give bounds when there is a local dependence structure between the random variables. Below, the notion of local dependence is

introduced before we give the Stein's method result that is used in the case of locally dependent random variables.

An m-dependent sequence of random variables $\{X_i, i \in \mathbb{N}\}$ is such that for each $i \in \mathbb{N}$ the sets of random variables $\{X_j, j \leq i\}$ and $\{X_j, j > i+m\}$ are independent. The Stein's method result for the case of locally dependent random variables is based on the local dependence condition (LD) which is explained below.

For a set of random variables $\{\xi_i, i = 1,2,\ldots,n\}$, for any $A \subset \{1,2,\ldots,n\}$ we define

$$A^c = \{i \in \{1,2,\ldots,n\} : i \notin A\}$$
$$\xi_A = \{\xi_i : i \in A\}$$

and we introduce the local dependence condition

(LD) For each $i \in \{1,2,\ldots,n\}$ there exist $A_i \subset B_i \subset \{1,2,\ldots,n\}$ such that ξ_i is independent of $\xi_{A_i^c}$ and ξ_{A_i} is independent of $\xi_{B_i^c}$.

Whenever this condition holds we set

$$\eta_i = \sum_{j \in A_i} \xi_j, \qquad \tau_i = \sum_{j \in B_i} \xi_j. \tag{2.14}$$

Lemma 2.3 below gives an upper bound for the Wasserstein distance between the distribution of a sum of m-dependent random variables satisfying (LD). The random variables are assumed to have mean zero with the variance of their sum being equal to one. The proof of the lemma is beyond the scope of the thesis and can be found in (Chen et al., 2011, p.134).

Lemma 2.3. *Let* $\{\xi_i, i = 1,2,\ldots,n\}$ *be a set of random variables with mean zero and* $\mathrm{Var}(W) = 1$*, where* $W = \sum_{i=1}^{n} \xi_i$*. If (LD) holds, then with* η_i *and* τ_i *as in* (2.14)*,*

$$d_W(W,Z) \leq 2\sum_{i=1}^{n} \left(\mathrm{E}|\xi_i\eta_i\tau_i| + |\mathrm{E}(\xi_i\eta_i)|\mathrm{E}|\tau_i|\right) + \sum_{i=1}^{n} \mathrm{E}\left|\xi_i\eta_i^2\right|. \tag{2.15}$$

In situations of dependent random variables, couplings turn out to be useful. Here the exchangeable pair coupling is briefly presented. For a given random variable of interest W, the exchangeable pair approach relies on a construction of a random variable W' such that (W,W') is an exchangeable pair; equivalently $(W,W') \overset{\mathrm{d}}{=\!=} (W',W)$, where $\overset{\mathrm{d}}{=\!=}$ denotes equality in distribution. An exchangeable pair is called a λ-Stein pair if it also satisfies

$$\mathrm{E}\left(W'|W\right) = (1-\lambda)W,$$

where $\lambda \in (0,1)$. An example of such an approach for independent random variables is given in (Chen et al., 2011, p.23), where $W = \sum_{i=1}^{n} \xi_i$, with ξ_i, $i \in \{1,2,\ldots,n\}$, are independent random variables with zero mean and such that $\mathrm{Var}(W) = 1$. Allowing $\{\xi_i', i = 1,2,\ldots,n\}$ to be an independent copy of $\{\xi_i, i = 1,2,\ldots,n\}$ and I to have the uniform distribution on $\{1,2,\ldots,n\}$ we define $W' = W - \xi_I + \xi_I'$. By construction (W,W') and (W',W) have the same distribution and simple calculations show that

$$\mathrm{E}\left(W'|W\right) = \left(1 - \frac{1}{n}\right)W.$$

Hence, (W,W') is a λ-Stein pair with $\lambda = \frac{1}{n}$. This approach can be extended to the multivariate setting, where W and W' are vectors. Such an approach will be used in Chapter 4 through a theorem by Reinert and Röllin (2009).

So far, this section considered univariate random variables. In this thesis Stein's method for multivariate normal approximation in $\mathbb{R}^d$ is also used. We start with the multivariate Stein equation. For $g : \mathbb{R}^d \to \mathbb{R}$ being twice differentiable and ∇g, $D^2 g$ being

the gradient vector and Hessian matrix of g, respectively, we use the Stein equation

$$\mathrm{Tr}\left(D^2 g(\boldsymbol{w})\right) - \boldsymbol{w}\nabla g(\boldsymbol{w}) = h(\boldsymbol{w}) - \mathrm{E}[h(\boldsymbol{Z})], \tag{2.16}$$

where $\boldsymbol{Z} \sim \mathrm{N}_d\left(\boldsymbol{0}, I_{d\times d}\right)$ and $\mathrm{Tr}(.)$ denotes the trace of a matrix. The function $h : \mathbb{R}^d \to \mathbb{R}$, still plays the role of the test function. Lemma 2.6 in (Chen et al., 2011, p.17) shows that if h has three bounded derivatives then the unique bounded solution of (2.16) is

$$g_h(\boldsymbol{w}) = -\int_0^\infty \left[\mathrm{E}\left[h\left(\boldsymbol{w}\mathrm{e}^{-u} + \sqrt{1-\mathrm{e}^{-2u}}\boldsymbol{Z}\right)\right] - \mathrm{E}[h(\boldsymbol{Z})]\right] \mathrm{d}u$$

and satisfies

$$\|g_h^{(\boldsymbol{k})}\| \leq \frac{1}{k}\|h^{(\boldsymbol{k})}\|.$$

Here for a vector $\boldsymbol{k} = (k_1, k_2, \ldots, k_d)$ with $k_j \in \mathbb{Z}_0^+$, we have $h^{(\boldsymbol{k})}(\boldsymbol{w}) = \frac{\partial^{|\boldsymbol{k}|}}{\prod_{j=1}^d \partial w_{k_j}} h(\boldsymbol{w})$ and $|\boldsymbol{k}| = \sum_{j=1}^d k_j$. The idea for the construction of exchangeable pair couplings is similar as in the univariate case.

Significant progress on Stein's method beyond normal approximation has been achieved. General distributions can be treated via the generator approach of Barbour (1990), the density approach of Stein et al. (2004), further developed in Ley and Swan (2013b) and Ley and Swan (2013a), and the parsimonious Stein operator approach of Ley et al. (2014). The most worth-mentioning extension is Stein's method for the Poisson distribution. This method is also known as Stein-Chen method and was first presented in Chen (1975). An illustration through various examples of the broad applicability and power of the Stein-Chen method for Poisson approximation can be found in Arratia et al. (1990). In addition, using Stein-Chen method, Barbour and Holst (1989) give upper bounds on the distributional distance between the distribution of a sum of Bernoulli random variables and the Poisson distribution, while Barbour et al. (1992) treats in detail the Stein-Chen method.

Further distributions in which Stein's method has been adapted include the uniform (Diaconis, 1989), point mass (Reinert, 1995), geometric (Peköz, 1996), Beta (Döbler, 2015), (Goldstein and Reinert, 2013), Laplace (Pike and Ren, 2014), binomial (Ehm, 1991), negative binomial (Barbour et al., 2015), exponential (Chatterjee et al., 2011), (Peköz and Röllin, 2011) and gamma distributions (Luk, 1994), (Nourdin and Peccati, 2009), (Pickett, 2004) as well as the class of variance-Gamma distributions (Gaunt, 2013). Related to the normal distribution, Barbour (1990) and Götze (1991) study similar results for more general Gaussian distributions. Nourdin and Peccati (2012) generalise the results to an infinite-dimensional setting. The survey article Ross (2011) gives the main concepts of Stein's method for distributional approximation by the normal, Poisson, exponential and geometric distributions.

In recent years there has been a growing interest in assessing the quality of asymptotic results in statistical theory using Stein's method. Recent examples include Ley et al. (2015) where the impact of the choice of the prior distribution in Bayesian statistics is assessed and Gaunt et al. (2015) for the chi-square approximation to the Pearson Statistic and of course the quantitative results of this thesis for assessing the normal approximation of the MLE. Our bounds have already influenced Pinelis and Molzon (2016).

Chapter 3

Bounds on the distance to normal for the MLE

In this chapter we assess the normal approximation of the MLE. Section 3.1 first briefly relates the Kolmogorov and the bounded Wasserstein distance. In addition, insights to conservative confidence intervals are given. Section 3.2 gives the main upper bound on the distributional distance between the distribution of the MLE and the normal distribution in terms of the bounded Wasserstein distance.

We specify the result for the distribution of the MLE for single-parameter exponential family distributions. Section 3.3 provides an upper bound for the distance between the distribution of the MLE for exponential family distributions and the normal distribution using the results of Section 3.2. We treat the example of the exponential distribution; simulation-based results illustrate the behaviour of the bound. In Section 3.4 we explain how, in discrete distributions, a perturbation method solves the issue of the MLE having positive probability of being on the boundary of the parameter space. The Poisson distribution with parameter space $[0,\infty)$ serves as an example to illustrate the results. Finally, in Section 3.5, bounds are provided for cases where a closed-form expression

for the MLE is not available. To achieve this task we first give bounds on the mean squared error of the MLE. The results of this chapter form a paper with Gesine Reinert, which is to appear in *Bernoulli*.

3.1 The bounded Wasserstein and the Kolmogorov distance

For $Z \sim \mathrm{N}(0,1)$, the aim is to bound the bounded Wasserstein distence

$$d_{bW}\left(\sqrt{n\,i(\theta_0)}(\hat{\theta}_n(\boldsymbol{X})-\theta_0), Z\right), \tag{3.1}$$

with $d_{bW}(.,.)$ as defined in (2.7). Recall that using $H_K = \left\{\mathbb{1}_{[.\leq x]}, x \in \mathbb{R}\right\}$ as the class of functions in (2.3) yields the Kolmogorov distance,

$$d_K\left(\sqrt{n\,i(\theta_0)}(\hat{\theta}_n(\boldsymbol{X})-\theta_0), Z\right).$$

The next proposition links our results to the Kolmogorov distance.

Proposition 3.1. *Let W be any real-valued random variable and $Z \sim \mathrm{N}(0,1)$. Assume that there exist $\delta_1 > 0$ and $\delta_2 \geq 0$, such that for any function h with $\|h'\|$ and $\|h\|$ bounded,*

$$|\mathrm{E}[h(W)] - \mathrm{E}[h(Z)]| \leq \delta_1 \|h'\| + \delta_2 \|h\|. \tag{3.2}$$

Then,

$$d_K(W,Z) \leq 2\sqrt{\delta_1} + \delta_2.$$

Proof. The proof of this proposition follows the approach explained in the proof of Theorem 3.3 of Chen et al. (2011), p.48, where a similar result is shown that links the

Wasserstein and the Kolmogorov distance. Let $z \in \mathbb{R}$ and for $\alpha = \sqrt{\delta_1}(2\pi)^{\frac{1}{4}}$, $z \in \mathbb{R}$, let

$$h_\alpha(w) = \begin{cases} 1, & \text{if } w \le z, \\ 1 + \frac{z-w}{\alpha}, & \text{if } z < w \le z + \alpha, \\ 0, & \text{if } w > z + \alpha, \end{cases}$$

so that h_α is bounded Lipschitz with $\|h_\alpha\| \le 1$ and $\|h'_\alpha\| \le \frac{1}{\alpha}$. The definition of $h_\alpha(.)$ gives that

$$\begin{aligned} \mathbb{P}(W \le z) - \mathbb{P}(Z \le z) &\le \mathrm{E}[h_\alpha(W)] - \mathrm{E}[h_\alpha(Z)] + \mathrm{E}[h_\alpha(Z)] - \mathbb{P}(Z \le z) \\ &\le |\mathrm{E}[h_\alpha(W)] - \mathrm{E}[h_\alpha(Z)]| + \mathrm{E}[h_\alpha(Z)] - \mathbb{P}(Z \le z) \\ &\le \delta_1 \|h'_\alpha\| + \delta_2 \|h_\alpha\| + \mathbb{P}(z < Z \le z + \alpha) \\ &\le \frac{\delta_1}{\alpha} + \delta_2 + \frac{\alpha}{\sqrt{2\pi}} \\ &\le 2\frac{\sqrt{\delta_1}}{(2\pi)^{\frac{1}{4}}} + \delta_2 \le 2\sqrt{\delta_1} + \delta_2. \end{aligned}$$

Similarly, using now

$$h^*_\alpha(w) = \begin{cases} 0, & \text{if } w \le z - \alpha, \\ \frac{w-(z-\alpha)}{\alpha}, & \text{if } z - \alpha < w \le z, \\ 1, & \text{if } w > z, \end{cases}$$

we can show that $\mathbb{P}(W \le z) - \mathbb{P}(Z \le z) \ge -\left(2\sqrt{\delta_1} + \delta_2\right)$, which completes the proof.

□

The Kolmogorov distance relates directly to exact conservative confidence intervals. Our bounds for the MLE (before specialising on d_{bW}) are of the form of the expression in (3.2), where in our case $W = \sqrt{ni(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right)$, δ_1 and δ_2 are explicitly given

with $\delta_1 = \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$ and $\delta_2 = \mathcal{O}\left(\frac{1}{n}\right)$. Using Proposition 3.1,

$$d_K\left(\sqrt{n\,i(\theta_0)}(\hat{\theta}_n(\boldsymbol{X}) - \theta_0), Z\right) \leq 2\sqrt{\delta_1} + \delta_2 =: B_K.$$

Therefore, for $y \in \mathbb{R}$:

$$\begin{aligned} &\left|\mathbb{P}\left(\sqrt{n\,i(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right) \leq y\right) - \mathbb{P}\left(Z \leq y\right)\right| \leq B_K \\ &\Leftrightarrow -B_K \leq \mathbb{P}\left(\sqrt{n\,i(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right) \leq y\right) - \mathbb{P}\left(Z \leq y\right) \leq B_K. \end{aligned} \tag{3.3}$$

For $\Phi^{-1}(.)$ the quantile function for the standard normal distribution, applying (3.3) to $y = \Phi^{-1}\left(\frac{\alpha}{2} - B_K\right)$ and to $y = \Phi^{-1}\left(1 - \frac{\alpha}{2} + B_K\right)$ yields

$$\mathbb{P}\left(\Phi^{-1}\left(\frac{\alpha}{2} - B_K\right) \leq \sqrt{n\,i(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right) \leq \Phi^{-1}\left(1 - \frac{\alpha}{2} + B_K\right)\right) \geq 1 - \alpha.$$

Hence, if the expected Fisher Information number for one random variable, $i(\theta_0)$, is known, then

$$\left(\hat{\theta}_n(\boldsymbol{X}) - \frac{\Phi^{-1}\left(1 - \frac{\alpha}{2} + B_K\right)}{\sqrt{n\,i(\theta_0)}}, \hat{\theta}_n(\boldsymbol{X}) - \frac{\Phi^{-1}\left(\frac{\alpha}{2} - B_K\right)}{\sqrt{n\,i(\theta_0)}}\right)$$

is a conservative $100(1-\alpha)\%$ confidence interval for θ_0. Note that B_K may depend on θ_0.

3.2 Bounds in terms of the bounded Wasserstein distance

The bounded Wasserstein distance links in well with Stein's method because the Lipschitz test functions are differentiable almost everywhere and so Taylor expansions can

be used. From now on, $\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta_0) := \frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta)\Big|_{\theta=\theta_0}$. The next two results provide a bound for (a) and (b) in (2.2), respectively.

Proposition 3.2. *Suppose $X_1, X_2, \ldots, X_n$ are i.i.d. random variables with probability density (or mass) function $f(x_i|\theta)$. Assume that the regularity conditions (R1)-(R4) are satisfied and $\mathrm{E}_{\theta_0}\left|\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta_0)\right|^3$ exists. Let $Z \sim N(0,1)$. Then for $h : \mathbb{R} \to \mathbb{R}$, such that h is absolutely continuous and bounded*

$$\left|\mathrm{E}_{\theta_0}\left[h\left(\frac{l'(\theta_0;\boldsymbol{X})}{\sqrt{n\,i(\theta_0)}}\right)\right] - \mathrm{E}[h(Z)]\right| \leq \frac{\|h'\|}{\sqrt{n}}\left(2 + \frac{1}{[i(\theta_0)]^{\frac{3}{2}}}\left[\mathrm{E}_{\theta_0}\left|\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta_0)\right|^3\right]\right). \tag{3.4}$$

In particular,

$$d_{bW}\left(\frac{l'(\theta_0;\boldsymbol{X})}{\sqrt{n\,i(\theta_0)}}, Z\right) \leq \frac{1}{\sqrt{n}}\left(2 + \frac{1}{[i(\theta_0)]^{\frac{3}{2}}}\left[\mathrm{E}_{\theta_0}\left|\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta_0)\right|^3\right]\right). \tag{3.5}$$

Proof. Let $Y_i = Y_i(X_i;\theta_0) = \frac{\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_i|\theta_0)}{\sqrt{i(\theta_0)}}, i = 1,2,\ldots,n$, which are i.i.d. random variables as $X_1, X_2, \ldots, X_n$ are i.i.d. The regularity conditions (R1)-(R4) ensure that $\mathrm{E}_{\theta_0}[Y_i] = 0$ and $\mathrm{Var}_{\theta_0}[Y_i] = 1$. Let $W = W(\boldsymbol{X};\theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i = \frac{l'(\theta_0;\boldsymbol{X})}{\sqrt{n\,i(\theta_0)}}$, so that $\mathrm{E}_{\theta_0}[W] = 0$ and $\mathrm{Var}_{\theta_0}[W] = 1$. Applying Lemma 2.2 to $K = Z \sim N(0,1)$ gives the result in (3.4). Using that $\|h'\| \leq 1$ for $h \in H_{bW}$, we obtain the bound in (3.5). □

The following theorem is the main result of this chapter; it is a bound on the distributional distance between the exact distribution of the MLE and its limiting normal distribution for the case of i.i.d. random variables from a single-parameter distribution. We specify the result for the bounded Wasserstein distance. For the sake of presentation we drop the subscript θ_0 from the expectation and variance.

Theorem 3.1. *Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with probability density (or mass) function $f(x_i|\theta)$ such that the MLE, $\hat{\theta}_n(\boldsymbol{X})$, exists and is unique and the regularity conditions (R1)-(R4) are satisfied. Assume that $\mathrm{E}\left|\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta_0)\right|^3 < \infty$ and*

that $\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^4\right]<\infty$. *Let* $0<\varepsilon=\varepsilon(\theta_0)=\varepsilon_0$ *be such that* $|\theta-\theta_0|<\varepsilon$ *with* ε_0 *as in (R3). Also, assume that* $\mathrm{E}\left(\left(\sum_{i=1}^n M(X_i)\right)^2\middle|\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right|<\varepsilon\right)<\infty$, *with* $M(X)$ *as in (R3). Let* $Z\sim \mathrm{N}(0,1)$. *Then for* $h:\mathbb{R}\to\mathbb{R}$, *such that* h *is absolutely continuous and bounded,*

$$
\begin{aligned}
&\left|\mathrm{E}\left[h\left(\sqrt{n\,i(\theta_0)}(\hat{\theta}_n(\boldsymbol{X})-\theta_0)\right)\right]-\mathrm{E}[h(Z)]\right| \\
&\leq \frac{\|h'\|}{\sqrt{n}}\left(2+\frac{1}{[i(\theta_0)]^{\frac{3}{2}}}\left[\mathrm{E}\left|\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta_0)\right|^3\right]\right) \\
&+\frac{\|h'\|}{\sqrt{i(\theta_0)}}\sqrt{\mathrm{Var}\left(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(X_1|\theta_0)\right)}\sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]} \\
&+\frac{2\|h\|}{\varepsilon^2}\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right] \\
&+\frac{\|h'\|}{2\sqrt{ni(\theta_0)}}\left[\mathrm{E}\left(\left(\sum_{i=1}^n M(X_i)\right)^2\middle|\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right|<\varepsilon\right)\right]^{\frac{1}{2}}\left[\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^4\right]\right]^{\frac{1}{2}}.
\end{aligned}
\tag{3.6}
$$

In particular, for $h\in H_{bW}$ *as in* (2.6), (3.6) *gives*

$$
\begin{aligned}
&d_{bW}\left(\sqrt{n\,i(\theta_0)}(\hat{\theta}_n(\boldsymbol{X})-\theta_0),Z\right)\leq\frac{1}{\sqrt{n}}\left(2+\frac{1}{[i(\theta_0)]^{\frac{3}{2}}}\left[\mathrm{E}\left|\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta_0)\right|^3\right]\right) \\
&+\frac{1}{\sqrt{i(\theta_0)}}\sqrt{\mathrm{Var}\left(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(X_1|\theta_0)\right)}\sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]} \\
&+\frac{2}{\varepsilon^2}\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right] \\
&+\frac{1}{2\sqrt{ni(\theta_0)}}\left[\mathrm{E}\left(\left(\sum_{i=1}^n M(X_i)\right)^2\middle|\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right|<\varepsilon\right)\right]^{\frac{1}{2}}\left[\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^4\right]\right]^{\frac{1}{2}}.
\end{aligned}
\tag{3.7}
$$

The following lemma, often known as Chebyshev's other inequality (see Fink and Jodeit (1984) for more information), is useful to obtain (3.7).

Lemma 3.1. *Let $M \geq 0$ be a random variable and $\varepsilon > 0$. For every continuous increasing function $f : [0,\infty) \to [0,\infty)$,*

$$\mathrm{E}[f(M)|M < \varepsilon] \leq \mathrm{E}[f(M)].$$

Proof of **Lemma 3.1**. Let $\varepsilon > 0$ and f a continuous increasing function with $f(m) \geq 0$ for $m > 0$. Then,

$$\begin{aligned}
\mathrm{E}[f(M)] &= \mathrm{E}[f(M)|M < \varepsilon]\,\mathbb{P}(M < \varepsilon) + \mathrm{E}[f(M)|M \geq \varepsilon]\,\mathbb{P}(M \geq \varepsilon) \\
&= \mathrm{E}[f(M)|M < \varepsilon]\,(1 - \mathbb{P}(M \geq \varepsilon)) + \mathrm{E}[f(M)|M \geq \varepsilon]\,\mathbb{P}(M \geq \varepsilon) \\
&= \mathrm{E}[f(M)|M < \varepsilon] + \mathbb{P}(M \geq \varepsilon)\,(\mathrm{E}[f(M)|M \geq \varepsilon] - \mathrm{E}[f(M)|M < \varepsilon]) \\
&\geq \mathrm{E}[f(M)|M < \varepsilon] \quad \text{as } f(m) \text{ is increasing.}
\end{aligned}$$

□

Proof of **Theorem 3.1**. The regularity conditions and the definition of the MLE ensure that $0 = l'(\hat{\theta}_n(\boldsymbol{x}); \boldsymbol{x})$. A second order Taylor expansion of $l'(\hat{\theta}_n(\boldsymbol{x}); \boldsymbol{x})$ about θ_0 gives

$$l''(\theta_0; \boldsymbol{x})\left(\hat{\theta}_n(\boldsymbol{x}) - \theta_0\right) = -l'(\theta_0; \boldsymbol{x}) - R_1(\theta_0; \boldsymbol{x}), \tag{3.8}$$

where

$$R_1(\theta_0; \boldsymbol{x}) = \frac{1}{2}\left(\hat{\theta}_n(\boldsymbol{x}) - \theta_0\right)^2 l^{(3)}(\theta^*; \boldsymbol{x}) \tag{3.9}$$

is the remainder term with θ^* lying between $\hat{\theta}_n(\boldsymbol{x})$ and θ_0. Rearranging (3.8) and adding zero gives

$$-n\,i(\theta_0)\left(\hat{\theta}_n(\boldsymbol{x}) - \theta_0\right) = -l'(\theta_0; \boldsymbol{x}) - R_1(\theta_0; \boldsymbol{x}) - \left(\hat{\theta}_n(\boldsymbol{x}) - \theta_0\right)\left[l''(\theta_0; \boldsymbol{x}) + n\,i(\theta_0)\right].$$

As $i(\theta_0) \neq 0$,

$$\sqrt{n\, i(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{x}) - \theta_0\right) = \frac{l'(\theta_0;\boldsymbol{x}) + R_1(\theta_0;\boldsymbol{x}) + R_2(\theta_0,\boldsymbol{x})}{\sqrt{n\, i(\theta_0)}},$$

where

$$R_2(\theta_0,\boldsymbol{x}) = (\hat{\theta}_n(\boldsymbol{x}) - \theta_0)\left(l''(\theta_0;\boldsymbol{x}) + n\, i(\theta_0)\right). \tag{3.10}$$

For $Z \sim N(0,1)$ and $h \in H$ given in (2.6),

$$\begin{aligned}
&\left|\mathrm{E}\left[h\left(\sqrt{n\, i(\theta_0)}(\hat{\theta}_n(\boldsymbol{X}) - \theta_0)\right)\right] - \mathrm{E}[h(Z)]\right| \leq \\
&\left|\mathrm{E}\left[h\left(\frac{l'(\theta_0;\boldsymbol{X}) + R_1(\theta_0;\boldsymbol{X}) + R_2(\theta_0;\boldsymbol{X})}{\sqrt{n\, i(\theta_0)}}\right) - h\left(\frac{l'(\theta_0;\boldsymbol{X})}{\sqrt{n\, i(\theta_0)}}\right)\right]\right| \qquad (3.11)\\
&+ \left|\mathrm{E}\left[h\left(\frac{l'(\theta_0;\boldsymbol{X})}{\sqrt{n\, i(\theta_0)}}\right)\right] - \mathrm{E}[h(Z)]\right|. \qquad (3.12)
\end{aligned}$$

The upper bound for (3.12) is given in Proposition 3.2. With regards to (3.11), let for ease of presentation

$$\begin{aligned}
C_1 = C_1(h,\theta_0;\boldsymbol{X}) &= h\left(\frac{l'(\theta_0;\boldsymbol{X}) + R_1(\theta_0;\boldsymbol{X}) + R_2(\theta_0;\boldsymbol{X})}{\sqrt{n\, i(\theta_0)}}\right) - h\left(\frac{l'(\theta_0;\boldsymbol{X})}{\sqrt{n\, i(\theta_0)}}\right)\\
C_2 = C_2(h,\theta_0;\boldsymbol{X}) &= h\left(\frac{l'(\theta_0;\boldsymbol{X}) + R_1(\theta_0;\boldsymbol{X})}{\sqrt{n\, i(\theta_0)}}\right) - h\left(\frac{l'(\theta_0;\boldsymbol{X})}{\sqrt{n\, i(\theta_0)}}\right). \qquad (3.13)
\end{aligned}$$

Using this notation, the triangle inequality yields

$$\begin{aligned}
&(3.11) = |\mathrm{E}\,[C_1]|\\
&= \left|\mathrm{E}\left[h\left(\frac{l'(\theta_0;\boldsymbol{X}) + R_1(\theta_0;\boldsymbol{X}) + R_2(\theta_0;\boldsymbol{X})}{\sqrt{n\, i(\theta_0)}}\right) - h\left(\frac{l'(\theta_0;\boldsymbol{X}) + R_1(\theta_0;\boldsymbol{X})}{\sqrt{n\, i(\theta_0)}}\right) + C_2\right]\right|\\
&\leq \frac{\|h'\|}{\sqrt{n\, i(\theta_0)}}\mathrm{E}\,|R_2(\theta_0;\boldsymbol{X})| + \mathrm{E}\,|C_2|\,, \qquad (3.14)
\end{aligned}$$

where the last inequality is due to a first order Taylor expansion of $h\left(\frac{l'(\theta_0;\boldsymbol{X}) + R_1(\theta_0;\boldsymbol{X}) + R_2(\theta_0;\boldsymbol{X})}{\sqrt{n\, i(\theta_0)}}\right)$ about $\frac{l'(\theta_0;\boldsymbol{X}) + R_1(\theta_0;\boldsymbol{X})}{\sqrt{n\, i(\theta_0)}}$. The definition of $R_2(\theta_0;\boldsymbol{x})$ in

(3.10) and the Cauchy - Schwarz inequality give that

$$\begin{aligned}\mathrm{E}\left|R_2\left(\theta_0;\boldsymbol{X}\right)\right| &= \mathrm{E}\left|\left(n\,i(\theta_0)+l''(\theta_0;\boldsymbol{X})\right)\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)\right| \\ &\leq \sqrt{\mathrm{E}\left[n\,i(\theta_0)+l''(\theta_0;\boldsymbol{X})\right]^2}\sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]} \\ &= \sqrt{n\mathrm{Var}\left(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(X_1|\theta_0)\right)}\sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]}. \qquad (3.15)\end{aligned}$$

To bound $\mathrm{E}\left|C_2\right|$ with C_2 as in (3.13), note that the term $R_1(\theta_0;\boldsymbol{X})$ is in general not uniformly bounded. For all x the rather crude bound $|C_2| \leq 2\|h\|$ is valid. If $\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right| < \varepsilon$ then a better bound is available. Hence, we condition on whether $\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right| \geq \varepsilon$ or $\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right| < \varepsilon$, with $\varepsilon = \varepsilon_0 > 0$ as in condition (R3). By Markov's inequality

$$\mathbb{P}\left(\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right| \geq \varepsilon\right) \leq \frac{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]}{\varepsilon^2}. \qquad (3.16)$$

Using the law of total expectation,

$$\begin{aligned}|\mathrm{E}[C_2]| \leq \mathrm{E}|C_2| =& \mathrm{E}\left(|C_2|\left|\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right| \geq \varepsilon\right)\mathbb{P}\left(\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right| \geq \varepsilon\right) \\ &+ \mathrm{E}\left(|C_2|\left|\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right| < \varepsilon\right)\mathbb{P}\left(\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right| < \varepsilon\right).\end{aligned}$$

Using (3.16) for the first term and a first order Taylor expansion of $h\left(\frac{l'(\theta_0;\boldsymbol{X})+R_1(\theta_0;\boldsymbol{X})}{\sqrt{n\,i(\theta_0)}}\right)$ about $\frac{l'(\theta_0;\boldsymbol{X})}{\sqrt{n\,i(\theta_0)}}$ for the second term gives

$$\begin{aligned}|\mathrm{E}[C_2]| &\leq \frac{2\|h\|}{\varepsilon^2}\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right] \\ &\quad + \left|\mathrm{E}\left(\frac{R_1(\theta_0,\boldsymbol{X})}{\sqrt{n\,i(\theta_0)}}h'(t(\boldsymbol{X}))\,\middle|\,\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right| < \varepsilon\right)\right|\mathbb{P}\left(\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right| < \varepsilon\right) \\ &\leq \frac{2\|h\|}{\varepsilon^2}\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right] \\ &\quad + \frac{\|h'\|}{2\sqrt{n\,i(\theta_0)}}\mathrm{E}\left(\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\left|l^{(3)}(\theta^*;\boldsymbol{X})\right|\left|\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right| < \varepsilon\right),\end{aligned}$$

where $t(\boldsymbol{X})$ lies between $\frac{l'(\theta_0;\boldsymbol{X})}{\sqrt{n\,i(\theta_0)}}$ and $\frac{l'(\theta_0;\boldsymbol{X})+R_1(\theta_0;\boldsymbol{X})}{\sqrt{n\,i(\theta_0)}}$. Using (3.14), (3.15) and the fact that for $\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right|<\varepsilon$,

$$\begin{aligned}|R_1(\theta_0;\boldsymbol{X})| &\leq \frac{1}{2}(\hat{\theta}_n(\boldsymbol{X})-\theta_0)^2 \sup_{\theta:|\theta-\theta_0|<\varepsilon}\left|l^{(3)}(\theta;\boldsymbol{X})\right| \\ &= \frac{1}{2}(\hat{\theta}_n(\boldsymbol{X})-\theta_0)^2 \sup_{\theta:|\theta-\theta_0|<\varepsilon}\left|\sum_{i=1}^{n}\frac{\mathrm{d}^3}{\mathrm{d}\theta^3}\log f(X_i|\theta)\right| \\ &\leq \frac{1}{2}(\hat{\theta}_n(\boldsymbol{X})-\theta_0)^2 \sum_{i=1}^{n}\left\{\sup_{\theta:|\theta-\theta_0|<\varepsilon}\left|\frac{\mathrm{d}^3}{\mathrm{d}\theta^3}\log f(X_i|\theta)\right|\right\} \\ &\leq \frac{1}{2}(\hat{\theta}_n(\boldsymbol{X})-\theta_0)^2 \sum_{i=1}^{n} M(X_i),\end{aligned} \tag{3.17}$$

then

$$\begin{aligned}|\mathrm{E}[C_1]| \leq & \frac{2\|h\|}{\varepsilon^2}\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right] \\ &+ \frac{\|h'\|}{\sqrt{i(\theta_0)}}\sqrt{\mathrm{Var}\left(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(X_1|\theta_0)\right)}\sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]} \\ &+ \frac{\|h'\|}{2\sqrt{n\,i(\theta_0)}}\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\sum_{i=1}^{n}M(X_i)\middle|\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right|<\varepsilon\right].\end{aligned}$$

The next step is based on the Cauchy - Schwarz inequality and the fact that

$$\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^4\middle|\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right|<\varepsilon\right] \leq \mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^4\right], \tag{3.18}$$

due to Lemma 3.1, giving

$$\begin{aligned}&|\mathrm{E}[C_1]| \leq \frac{2\|h\|}{\varepsilon^2}\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right] \\ &+ \frac{\|h'\|}{\sqrt{i(\theta_0)}}\sqrt{\mathrm{Var}\left(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(X_1|\theta_0)\right)}\sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]} \\ &+ \frac{\|h'\|}{2\sqrt{ni(\theta_0)}}\left[\mathrm{E}\left(\left(\sum_{i=1}^{n}M(X_i)\right)^2\middle|\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right|<\varepsilon\right)\right]^{\frac{1}{2}}\left[\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^4\right]\right]^{\frac{1}{2}}.\end{aligned} \tag{3.19}$$

The first result of the theorem in (3.6) is obtained using (3.4) and (3.19). Taking $h \in H_{bW}$ as in (2.6), then $\|h\| \leq 1$ and $\|h'\| \leq 1$ leading to the result in (3.7). □

Remark 3.1. ***(1)*** *Using* (3.10)*, if* $l''(\theta_0;\boldsymbol{x}) \equiv -n\,i(\theta_0)$ *then* $R_2(\theta_0;\boldsymbol{x}) \equiv 0$ *and the second term of the bounds given in Theorem 3.1 vanishes.*

(2) *For* $M_1 = M_1(\theta_0)$ *a constant that may depend on the unknown parameter* θ_0 *and for* $0 < \varepsilon = \varepsilon(\theta_0) = \varepsilon_0$ *as in (R3), if* $\left|\frac{\mathrm{d}^3}{\mathrm{d}\theta^3}\log f(x|\theta)\right| \leq M_1$ *then* $M(x) = M_1$ *and the bound is simpler as the conditional expectation in the last term of the bound vanishes. This is because step* (3.18) *in the proof is not any more necessary since*

$$\begin{aligned}
&\mathrm{E}\left[\sum_{i=1}^{n} M(X_i)\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2 \middle| \left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right| < \varepsilon\right] \\
&= nM_1\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2 \left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right| < \varepsilon\right] \leq nM_1\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right].
\end{aligned}$$

Thus, in this case the bound on the bounded Wasserstein distance becomes

$$\begin{aligned}
d_{bW}&\left(\sqrt{n\,i(\theta_0)}(\hat{\theta}_n(\boldsymbol{X})-\theta_0), Z\right) \leq \frac{1}{\sqrt{n}}\left(2+\frac{1}{[i(\theta_0)]^{\frac{3}{2}}}\left[\mathrm{E}\left|\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta_0)\right|^3\right]\right) \\
&+\frac{1}{\sqrt{i(\theta_0)}}\sqrt{\mathrm{Var}\left(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(X_1|\theta_0)\right)}\sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]} \\
&+\frac{2}{\varepsilon^2}\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right] \\
&+\frac{\sqrt{n}M_1}{2\sqrt{i(\theta_0)}}\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right].
\end{aligned}$$

Furthermore, if the third derivative of the log-likelihood function is uniformly bounded by a constant, $\left|\frac{\mathrm{d}^3}{\mathrm{d}\theta^3}\log f(x|\theta)\right| \leq \tilde{B}$*, then we do not need to use* ε *and with* C_1 *as in* (3.13)

$$\begin{aligned}
|\mathrm{E}[C_1]| \leq &\frac{\|h'\|}{\sqrt{i(\theta_0)}}\sqrt{\mathrm{Var}\left(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(X_1|\theta_0)\right)}\sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]} \\
&+\tilde{B}\frac{\|h'\|\sqrt{n}}{2\sqrt{i(\theta_0)}}\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]
\end{aligned}$$

leading to

$$d_{bW}\left(\sqrt{n\,i(\theta_0)}(\hat{\theta}_n(\boldsymbol{X})-\theta_0),Z\right) \leq \frac{1}{\sqrt{n}}\left(2+\frac{1}{[i(\theta_0)]^{\frac{3}{2}}}\left[\mathrm{E}\left|\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta_0)\right|^3\right]\right)$$
$$+\frac{1}{\sqrt{i(\theta_0)}}\sqrt{\mathrm{Var}\left(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(X_1|\theta_0)\right)}\sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]}$$
$$+\frac{\tilde{B}\sqrt{n}}{2\sqrt{i(\theta_0)}}\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right].$$

(3) *The rate of convergence of the mean squared error,* $\mathrm{E}\left[(\hat{\theta}_n(\boldsymbol{X})-\theta_0)^2\right]$*, is* $\mathcal{O}\left(\frac{1}{n}\right)$*. This result is obtained using that*

$$\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right] = \mathrm{Var}\left[\hat{\theta}_n(\boldsymbol{X})\right]+\mathrm{bias}^2\left[\hat{\theta}_n(\boldsymbol{X})\right]. \tag{3.20}$$

Under the standard asymptotics (from the regularity conditions (R1)-(R4)) the MLE is asymptotically efficient,

$$n\mathrm{Var}[\hat{\theta}_n(\boldsymbol{X})] \xrightarrow[n\to\infty]{} [i(\theta_0)]^{-1},$$

and hence the variance of the MLE is of order $\frac{1}{n}$*. In addition, from Theorem 2.2 the bias of the MLE is equal to zero up to the order* $\frac{1}{\sqrt{n}}$*; see also Cox and Snell (1968), where it is stated that the bias of the MLE is* $o\left(\frac{1}{n}\right)$*, but no explicit conditions are given. Combining these two results and using* (3.20) *shows that the mean squared error of the MLE is of order* $\frac{1}{n}$*. In the examples that follow, the remaining terms in the bound are of order at most* $\frac{1}{\sqrt{n}}$*.*

(4) *When the calculation of* $\mathrm{E}\left(\left|\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta_0)\right|^3\right)$ *is awkward, Hölder's inequality can be used, giving* $\mathrm{E}\left(\left|\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta_0)\right|^3\right) \leq \left[\mathrm{E}\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta_0)\right)^4\right]^{\frac{3}{4}}$.

3.3 Single-parameter exponential families

This section specifies Theorem 3.1 for the distribution of the MLE for single-parameter exponential family distributions. Many popular distributions which have the same underlying structure based on relatively simple properties are exponential families, such as the normal, Gamma and Laplace distributions. Generalisations of exponential families can be found in Lauritzen (1988) and Berk (1972).

This section focuses on the simplest case of a scalar parameter. The distribution of a random variable, X, is said to be a *single-parameter exponential family distribution* if the probability density (or mass) function is of the form

$$f(x|\theta) = \exp\{k(\theta)T(x) - A(\theta) + S(x)\}\, \mathbb{1}_{\{x\in B\}}, \tag{3.21}$$

where the set $B = \{x : f(x|\theta) > 0\}$ is the support of X and does not depend on θ; $k(\theta)$ and $A(\theta)$ are functions of the parameter; $T(x)$ and $S(x)$ are functions only of the data. The choice of the functions $k(\theta)$ and $T(X)$ is not unique. If $k(\theta) = \theta$ we have the so-called *canonical case*. In this case θ and $T(X)$ are called the *natural parameter* and *natural observation* (Casella and Berger, 2002). For $x_1, x_2, \ldots, x_n$ the data points, we make the following assumptions, where (Ass.Ex.1)-(Ass.Ex.3) are used to ensure the existence and uniqueness of the MLE and (A1)-(A4) follow from the regularity conditions in Section 2.2.

(Ass.Ex.1) $\Theta \subset \mathbb{R}$ is open and connected;

(Ass.Ex.2) $\lim_{\theta\to\partial\Theta} [k(\theta)\sum_{i=1}^n T(x_i) - nA(\theta) + \sum_{i=1}^n S(x_i)] = -\infty$;

(Ass.Ex.3) We have $k''(\theta)\sum_{i=1}^n T(x_i) - n\,A''(\theta) < 0$ at every point θ, for which $k'(\theta)\sum_{i=1}^n T(x_i) - nA'(\theta) = 0$;

(A1) $k'(\theta) \neq 0, \forall\theta \in \Theta$ and $D(\theta) = \frac{A'(\theta)}{k'(\theta)}$ is invertible;

(A2) $l(\theta;x)$ is three times continuously differentiable with respect to θ, which is equivalent to both $k^{(3)}(\theta)$ and $A^{(3)}(\theta)$ existing and being continuous. In addition, integration of the density function over x and differentiation with respect to θ are three times interchangeable;

(A3) for any $\theta_0 \in \Theta$, there exists a positive number ε_0 and a function $M(x)$ (both of which may depend on θ_0) such that

$$\left|k^{(3)}(\theta)T(x) - A^{(3)}(\theta)\right| \leq M(x), \ \ \forall x \in B, \ \forall \theta \in \Theta : |\theta - \theta_0| < \varepsilon_0,$$

with $\mathrm{E}[M(X)] < \infty$;

(A4) $\mathrm{Var}[T(X)] > 0$;

(A5) $\mathrm{E}\,|T(X) - D(\theta_0)|^3$ exists. This assumption is required for meaningful bounds.

Proposition 3.3. *In the case of the single-parameter exponential family, the assumptions (A1)-(A4) imply the regularity conditions (R1)-(R4) as expressed in Chapter 3.*

Proof. First, according to (Davison, 2008, p.172), an exponential family of order p is written in its minimal representation if the set $\{1, k_1(\theta), k_2(\theta), \cdots, k_p(\theta)\}$ is linearly independent. From (A1), because $k'(\theta) \neq 0$, then $k(\theta)$ is not constant in θ. Therefore, a single-parameter exponential family is always in minimal representation and thus identifiable (Geyer, 2013). Continuing, (A2) and (A3) are just the versions of (R2) and (R3) in the case that the random variables follow a single-parameter exponential family distribution. Finally, using (3.21), simple steps yield

$$i(\theta_0) = \mathrm{Var}\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X|\theta_0)\right) = \mathrm{Var}(k'(\theta_0)T(X) - A'(\theta_0)) = \left[k'(\theta_0)\right]^2 \mathrm{Var(T(X))},$$

which indicates that since, from (A1), $k'(\theta) \neq 0$, $\forall \theta \in \Theta$, then

$$i(\theta_0) > 0 \Leftrightarrow \text{Var}(T(X)) > 0.$$

Thus, (A4) also implies (R4). □

Since exponential family distributions have the same structure for the probability density (or mass) function, then it would make sense for the MLE to have a specific general form for these distributions. This is shown in the proposition that follows.

Proposition 3.4. *Let $X_1, X_2, \cdots, X_n$ be i.i.d. random variables from a single-parameter exponential family distribution. The probability density (or mass) function is given in* (3.21)*. Assume that (A1), (A2) and (Ass.Ex.1)-(Ass.Ex.3) hold. The MLE in this case is unique and equal to*

$$\hat{\theta}_n(\boldsymbol{X}) = D^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} T(X_i)\right). \tag{3.22}$$

Proof. From (A2) and (3.21),

$$l^{(j)}(\theta; \boldsymbol{x}) = k^{(j)}(\theta)\sum_{i=1}^{n} T(x_i) - nA^{(j)}(\theta), \text{ for } j = 1,2,3$$

and straightforward calculations show that $l'\left(\hat{\theta}_n(\boldsymbol{X}); \boldsymbol{X}\right) = 0$ has the unique solution $\hat{\theta}_n(\boldsymbol{X}) = D^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} T(X_i)\right)$, where $D(\theta) = \frac{A'(\theta)}{k'(\theta)}$ is invertible by (A1). Using now (Ass.Ex.1) and (Ass.Ex.3) gives that $\hat{\theta}_n(\boldsymbol{X}) = D^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} T(X_i)\right)$ is the unique MLE in the case of single-parameter exponential family distributions. □

The following corollary of Theorem 3.1 gives an upper bound on the bounded Wasserstein distance between the distribution of the MLE as in (3.22) and the normal distribution.

Corollary 3.1. *Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with the probability density (or mass) function of a single-parameter exponential family distribution. Assume that*

(A1)-(A5) are satisfied and that (Ass.Ex.1)-(Ass.Ex.3) hold. With $Z \sim \mathrm{N}(0,1)$, $h \in H_{bW}$ *as in* (2.6), $0 < \varepsilon = \varepsilon(\theta_0) = \varepsilon_0$ *such that* $\theta \in (\theta_0 - \varepsilon, \theta_0 + \varepsilon) \cap \Theta$ *where* ε_0 *is as in (A3) and for* $\mathrm{E}\left((\sum_{i=1}^n M(X_i))^2 \middle| \left|\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right| < \varepsilon \right) < \infty$, *with* $M(X)$ *as in (A3), it holds that*

$$
\begin{aligned}
&d_{bW}\left(\sqrt{n\, i(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right), Z\right) \leq \frac{1}{\sqrt{n}}\left(2 + \frac{\mathrm{E}|T(X_1) - D(\theta_0)|^3}{[\mathrm{Var}[T(X_1)]]^{\frac{3}{2}}}\right) \\
&+ \frac{|k''(\theta_0)|}{\sqrt{i(\theta_0)}}\sqrt{\mathrm{Var}\left(T(X_1)\right)}\sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right)^2\right]} \\
&+ \frac{2}{\varepsilon^2}\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right)^2\right] \\
&+ \frac{1}{2\sqrt{ni(\theta_0)}}\left[\mathrm{E}\left(\left(\sum_{i=1}^n M(X_i)\right)^2 \middle| \left|\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right| < \varepsilon\right)\right]^{\frac{1}{2}}\left[\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right)^4\right]\right]^{\frac{1}{2}}.
\end{aligned}
\tag{3.23}
$$

Proof. For the first term of the bound let $Y_i = Y_i(X_i;\theta_0) = \frac{\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_i|\theta_0)}{\sqrt{i(\theta_0)}}, i = 1,2,\ldots,n$. Proposition 3.2 requires the calculation of $\mathrm{E}\,|Y_1|^3$ and

$$
\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_i|\theta)\Big|_{\theta=\theta_0} = k'(\theta_0)T(X_i) - A'(\theta_0)
$$

yields

$$
\begin{aligned}
\mathrm{E}\left|\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_i|\theta_0)\right|^3 &= \mathrm{E}\left|k'(\theta_0)T(X_i) - A'(\theta_0)\right|^3 = \mathrm{E}\left|k'(\theta_0)(T(X_i) - D(\theta_0))\right|^3 \\
&= |k'(\theta_0)|^3\mathrm{E}|T(X_i) - D(\theta_0)|^3, \quad \forall i \in \{1,2,\ldots,n\}.
\end{aligned}
$$

In addition, $i(\theta_0) = \mathrm{Var}\left[\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_i|\theta_0)\right] = [k'(\theta_0)]^2\mathrm{Var}[T(X_i)] > 0$ from (A1) and (A4). These quantities can now be applied to get the first term of the bound in (3.23). For the remaining terms, using (3.7),

$$
\mathrm{Var}\left(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(X_1|\theta_0)\right) = \mathrm{Var}\left(k''(\theta_0)T(X_1) - A''(\theta)\right) = \left[k''(\theta_0)\right]^2\mathrm{Var}\left(T(X_1)\right),
$$

which yields the assertion of the corollary. □

Remark 3.2. *In the canonical case, $k''(\theta_0) \equiv 0$ and the second term of the bound vanishes. Also, $\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(x|\theta) = -A''(\theta)$ and $i(\theta_0) = A''(\theta_0)$. In addition, $\frac{\mathrm{d}^3}{\mathrm{d}\theta^3}\log f(x|\theta) = -A^{(3)}(\theta)$ is independent of the random variables. Thus, using* ***(2)*** *of Remark 3.1 we get*

$$d_{bW}\left(\sqrt{n\,i(\theta_0)}(\hat{\theta}_n(\boldsymbol{X})-\theta_0), Z\right) \leq \frac{1}{\sqrt{n}}\left(2+\frac{\mathrm{E}|T(X_1)-D(\theta_0)|^3}{[\mathrm{Var}[T(X_1)]]^{\frac{3}{2}}}\right)$$
$$+\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]\left(\frac{2}{\varepsilon^2}+\frac{\sqrt{n}}{2\sqrt{A''(\theta_0)}}\sup_{\theta:|\theta-\theta_0|<\varepsilon}\left|A^{(3)}(\theta)\right|\right). \tag{3.24}$$

3.3.1 The exponential distribution

Here, we consider two examples using the exponential distribution, firstly, in its canonical form, and then under a change of parametrisation. In each case, we check if all necessary assumptions are satisfied in order to obtain the bound in (3.23).

The canonical case

In the case of $X_1, X_2, \ldots, X_n$ exponentially distributed, $\mathrm{Exp}(\theta)$, i.i.d. random variables where $\theta > 0$, the probability density function is

$$f(x|\theta) = \theta\exp\{-\theta x\} = \exp\{\log\theta - \theta x\} = \exp\{k(\theta)T(x) - A(\theta) + S(x)\}\,\mathbb{1}_{\{x\in B\}}, \tag{3.25}$$

where $B = (0,\infty)$, $\theta \in \Theta = (0,\infty)$, $T(x) = -x$, $k(\theta) = \theta$, $A(\theta) = -\log\theta$ and $S(x) = 0$. Hence $\mathrm{Exp}(\theta)$ is a single-parameter canonical exponential family distribution.

Corollary 3.2. *Let $X_1, X_2, \cdots, X_n$ be i.i.d. random variables that follow the Exp(θ_0) distribution. The MLE exists, it is unique, equal to $\hat{\theta}_n(\boldsymbol{X}) = \frac{1}{\bar{X}}$ and (A1)-(A5) are satisfied. For $h \in H_{bW}$ as in* (2.6)*,*

$$d_{bW}\left(\sqrt{n\,i(\theta_0)}(\hat{\theta}_n(\boldsymbol{X})-\theta_0), Z\right) < \frac{4.41456}{\sqrt{n}} + \frac{8(n+2)}{(n-1)(n-2)} + \frac{8\sqrt{n}(n+2)}{(n-1)(n-2)}. \tag{3.26}$$

Proof. We check the existence and uniqueness of the MLE through the assumptions (Ass.Ex.1)-(Ass.Ex.3). (Ass.Ex.1) holds since the parameter space Θ is open and connected. For (Ass.Ex.2),

$$k(\theta)\sum_{i=1}^{n}T(x_i)-nA(\theta)+\sum_{i=1}^{n}S(x_i)=-\theta\sum_{i=1}^{n}x_i+n\log\theta\xrightarrow[\theta\to\infty]{\theta\to 0}-\infty.$$

For (Ass.Ex.3), simple steps give

$$l'(\theta;\boldsymbol{x})=\frac{n}{\theta}-\sum_{i=1}^{n}x_i,\quad l''(\theta;\boldsymbol{x})=-\frac{n}{\theta^2}. \tag{3.27}$$

Thus, $l'(\theta;\boldsymbol{X})=0$ if and only if $\theta=\hat{\theta}_n(\boldsymbol{X})=\frac{n}{\sum_{i=1}^{n}X_i}$. In addition, $l''(\hat{\theta}_n(\boldsymbol{X});\boldsymbol{X})<0$ and thus (Ass.Ex.3) holds. The MLE is $\hat{\theta}_n(\boldsymbol{X})=\frac{n}{\sum_{i=1}^{n}X_i}$ and it is unique. To verify the assumptions (A1)-(A5), for (A1), $k(\theta)=\theta$ and thus $k'(\theta)=1\neq 0$, $\forall\theta\in\Theta$ and $D(\theta)=A'(\theta)=-\frac{1}{\theta}$ with $D^{-1}(\theta)=D(\theta)=-\frac{1}{\theta}$. For (A2), the first two derivatives for the sample are shown in (3.27). Also, $l^{(3)}(\theta;\boldsymbol{x})=\frac{2n}{\theta^3}$. Using (3.25) gives

$$\begin{aligned}
\int_0^\infty \frac{\mathrm{d}}{\mathrm{d}\theta}f(x|\theta)\,\mathrm{d}x&=\int_0^\infty \exp\{-\theta x\}-\theta x\exp\{-\theta x\}\,\mathrm{d}x=\frac{1}{\theta}-\mathrm{E}(X)=0\\
\int_0^\infty \frac{\mathrm{d}^2}{\mathrm{d}\theta^2}f(x|\theta)\,\mathrm{d}x&=\int_0^\infty -2x\exp\{-\theta x\}+\theta x^2\exp\{-\theta x\}\,\mathrm{d}x\\
&=-\frac{2}{\theta}\mathrm{E}(X)+\mathrm{E}\left(X^2\right)=0\\
\int_0^\infty \frac{\mathrm{d}^3}{\mathrm{d}\theta^3}f(x|\theta)\,\mathrm{d}x&=\int_0^\infty 3x^2\exp\{-\theta x\}-\theta x^3\exp\{-\theta x\}\,\mathrm{d}x\\
&=\frac{3}{\theta}\mathrm{E}\left(X^2\right)-\mathrm{E}\left(X^3\right)=0.
\end{aligned}$$

Now,

$$\left|k^{(3)}(\theta)T(x)-A^{(3)}(\theta)\right|=\left|-A^{(3)}(\theta)\right|=\left|\frac{2}{\theta^3}\right|, \tag{3.28}$$

which needs to get bounded in a neighbourhood of θ_0 in order for (A3) to hold. Let $\varepsilon = \varepsilon(\theta_0) = \varepsilon_0$ be a positive constant such that $\theta \in (\theta_0 - \varepsilon, \theta_0 + \varepsilon) \cap \Theta$. From (3.28), if $\varepsilon > \theta_0$ then $(\theta_0 - \varepsilon, \theta_0 + \varepsilon) \cap \Theta = (0, \theta_0 + \varepsilon)$ and the result in (3.28) can not be upper bounded in $(0, \theta_0 + \varepsilon)$. Therefore, $0 < \varepsilon < \theta_0$ with $(\theta_0 - \varepsilon, \theta_0 + \varepsilon) \cap \Theta = (\theta_0 - \varepsilon, \theta_0 + \varepsilon)$ and

$$\left| k^{(3)}(\theta) T(x) - A^{(3)}(\theta) \right| = \left| -A^{(3)}(\theta) \right| = \left| \frac{2}{\theta^3} \right| < \frac{2}{(\theta_0 - \varepsilon)^3} = M(x),$$

with $\mathrm{E}[M(X)] = \frac{2}{(\theta_0 - \varepsilon)^3} < \infty$, and hence (A3) holds. We choose ε to be in the middle of the interval $(0, \theta_0)$ as there is a trade-off on its choice for the second and third term of the bound in (3.24). The second term is a product of the mean squared error with $\frac{2}{\varepsilon^2}$ indicating that we should choose ε away from zero. However, the third term of the bound includes $\sup_{\theta : |\theta - \theta_0| < \varepsilon} \left| A^{(3)}(\theta) \right| = \frac{2}{(\theta_0 - \varepsilon)^3}$ and a wise choice for ε would be such that it is not close to θ_0. An optimisation process with respect to ε becomes quite tedious and therefore we choose ε to be the midpoint of $(0, \theta_0)$, which is away from both zero and θ_0 and also behaves very well. Now, since $T(x) = x$, we have that $\mathrm{Var}(T(X)) = \mathrm{Var}(X) = \frac{1}{\theta_0^2} > 0$ and (A4) is also satisfied. For (A5), basic calculations of integrals show that $\mathrm{E}|T(X) - D(\theta_0)|^3 = \mathrm{E}\left| \frac{1}{\theta_0} - X \right|^3 < \frac{2.41456}{\theta_0^3}$ and (A5) is satisfied. We have already showed that $\sup_{\theta : |\theta - \theta_0| < \varepsilon} \left| A^{(3)}(\theta) \right| = \frac{2}{(\theta_0 - \varepsilon)^3}$. For $\varepsilon = \frac{\theta_0}{2}$, $\sup_{\theta : |\theta - \theta_0| < \varepsilon} \left| A^{(3)}(\theta) \right| = \frac{16}{\theta_0^3}$. In addition, since $X_i \sim \mathrm{Exp}(\theta), \forall i \in \{1, 2, \cdots, n\}$ we have that $\bar{X} \sim \mathrm{G}(n, n\theta)$, with $G(\alpha, \beta)$ being the Gamma distribution with shape parameter α and rate parameter β. Using now the results in pp.70-73 of Kendall and Stuart (1969) gives that

$$\begin{aligned} \mathrm{E}\left[(\hat{\theta}_n(\boldsymbol{X}) - \theta_0)^2 \right] &= \mathrm{E}\left[(\hat{\theta}_n(\boldsymbol{X})^2 \right] - 2\theta_0 \mathrm{E}\left[\hat{\theta}_n(\boldsymbol{X}) \right] + \theta_0^2 \\ &= \frac{(n\theta_0)^2}{(n-1)(n-2)} - \frac{2n\theta_0^2}{n-1} + \theta_0^2 = \frac{(n+2)\theta_0^2}{(n-1)(n-2)}. \end{aligned}$$

Using (3.24) we obtain the result in (3.26). □

Remark 3.3. ***(1)*** *The rate of convergence of the bound is* $\frac{1}{\sqrt{n}}$*. Note that the bound does not depend on the value of* θ_0*.*

(2) *The calculation of* $\mathrm{E}\left|\frac{1}{\theta_0}-X\right|^3$ *requires a significant amount of steps. Instead, one could use Hölder's inequality with* $\mathrm{E}\left|\frac{1}{\theta_0}-X\right|^3 \leq \left[\mathrm{E}\left(\frac{1}{\theta_0}-X\right)^4\right]^{\frac{3}{4}} = \frac{9^{\frac{3}{4}}}{\theta_0^3}$ *using the results in pp.70-73 of Kendall and Stuart (1969).*

The non-canonical case

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables from $\mathrm{Exp}\left(\frac{1}{\theta}\right)$, with p.d.f.

$$\begin{aligned} f(x|\theta) &= \frac{1}{\theta}\exp\left\{-\frac{1}{\theta}x\right\} = \exp\left\{-\log\theta - \frac{1}{\theta}x\right\} \\ &= \exp\left\{k(\theta)T(x) - A(\theta) + S(x)\right\}\mathbb{1}_{\{x\in B\}}, \end{aligned} \tag{3.29}$$

where $B=(0,\infty)$, $\theta\in\Theta=(0,\infty)$, $T(x)=-x$, $k(\theta)=\frac{1}{\theta}$, $A(\theta)=\log\theta$ and $S(x)=0$. Thus, $\mathrm{Exp}\left(\frac{1}{\theta}\right)$ is a non-canonical exponential family distribution.

Corollary 3.3. *Let* $X_1, X_2, \cdots, X_n$ *be i.i.d. random variables that follow the* $Exp\left(\frac{1}{\theta_0}\right)$ *distribution. The MLE exists, it is unique, equal to* $\hat{\theta}_n(\boldsymbol{X}) = \bar{X}$ *and (A1)-(A5) are satisfied. For* $h \in H_{bW}$ *as in* (2.6)*,*

$$\begin{aligned} d_{bW}\left(\sqrt{n\,i(\theta_0)}(\hat{\theta}_n(\boldsymbol{X})-\theta_0), Z\right) <& \frac{4.41456}{\sqrt{n}} + \frac{8}{n} + \frac{2}{\sqrt{n}} \\ &+ \frac{1}{\sqrt{n}}\left(80\left[3\left(\frac{2}{n}+1\right)\right]^{\frac{1}{2}}\right). \end{aligned} \tag{3.30}$$

Proof. Corollary 3.2 and the invariance property of the MLE give that $\hat{\theta}_n(\boldsymbol{X}) = \bar{X}$. To check (A1)-(A5), condition (A1) holds since $k'(\theta) = -\frac{1}{\theta^2} \neq 0$, $\forall\theta\in\Theta$ and $D^{-1}(\theta) = D(\theta) = -\theta$. Equation (3.29) and simple calculations show that the log-likelihood is three times differentiable with respect to the parameter. For one observation, x, we have

that

$$\begin{aligned}
\int_0^\infty \frac{\mathrm{d}}{\mathrm{d}\theta} f(x|\theta)\,\mathrm{d}x &= \int_0^\infty -\frac{1}{\theta^2}\exp\left\{-\frac{1}{\theta}x\right\} + \frac{x}{\theta^3}\exp\left\{-\frac{1}{\theta}x\right\}\mathrm{d}x \\
&= -\frac{1}{\theta} + \frac{1}{\theta^2}\mathrm{E}(X) = 0 \\
\int_0^\infty \frac{\mathrm{d}^2}{\mathrm{d}\theta^2} f(x|\theta)\,\mathrm{d}x &= \int_0^\infty \exp\left\{-\frac{1}{\theta}x\right\}\left(\frac{2}{\theta^3} - \frac{4x}{\theta^4} + \frac{x^2}{\theta^5}\right)\mathrm{d}x \\
&= \frac{2}{\theta^2} - \frac{4}{\theta^3}\mathrm{E}(X) + \frac{1}{\theta^4}\mathrm{E}\left(X^2\right) = 0 \\
\int_0^\infty \frac{\mathrm{d}^3}{\mathrm{d}\theta^3} f(x|\theta)\,\mathrm{d}x &= \int_0^\infty \exp\left\{-\frac{1}{\theta}x\right\}\left(-\frac{6}{\theta^4} + \frac{18x}{\theta^5} - \frac{9x^2}{\theta^6} + \frac{x^3}{\theta^7}\right)\mathrm{d}x \\
&= -\frac{6}{\theta^3} + \frac{18}{\theta^4}\mathrm{E}(X) - \frac{9}{\theta^5}\mathrm{E}\left(X^2\right) + \frac{1}{\theta^6}\mathrm{E}\left(X^3\right) = 0
\end{aligned}$$

and thus (A2) is satisfied. Now, using the triangle inequality,

$$\left|k^{(3)}(\theta)T(x) - A^{(3)}(\theta)\right| = \left|\frac{6x}{\theta^4} - \frac{2}{\theta^3}\right| \leq \frac{6x}{\theta^4} + \frac{2}{\theta^3}. \tag{3.31}$$

To upper bound (3.31) in a neighbourhood of θ so that (A3) holds, we take as in the canonical case $0 < \varepsilon < \theta_0$, since if $\varepsilon > \theta_0$ then $(\theta_0 - \varepsilon, \theta_0 + \varepsilon) \cap \Theta = (0, \theta_0 + \varepsilon)$ and (3.31) cannot get bounded. Thus, $\varepsilon < \theta_0$ and $(\theta_0 - \varepsilon, \theta_0 + \varepsilon) \cap \Theta = (\theta_0 - \varepsilon, \theta_0 + \varepsilon)$ which leads to

$$\left|k^{(3)}(\theta)T(x) - A^{(3)}(\theta)\right| = \left|\frac{6x}{\theta^4} - \frac{2}{\theta^3}\right| \leq \frac{6x}{\theta^4} + \frac{2}{\theta^3} \leq \frac{6x}{(\theta_0 - \varepsilon)^4} + \frac{2}{(\theta_0 - \varepsilon)^3} = M(x), \tag{3.32}$$

with $\mathrm{E}[M(X)] = \frac{6\theta_0}{(\theta_0-\varepsilon)^4} + \frac{2}{(\theta_0-\varepsilon)^3} < \infty$ showing that (A3) is satisfied. Since $T(x) = x$, then $\mathrm{Var}(T(X)) = \mathrm{Var}(X) = \theta_0^2$ which is positive and (A4) holds. For (A5),

$$\mathrm{E}|T(X) - D(\theta_0)|^3 = \mathrm{E}\left[\theta_0 - X\right]^3 < 2.41456\theta_0^3.$$

The mean squared error is $\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]=\mathrm{E}\left[(\bar{X}-\theta_0)^2\right]=\frac{\theta_0^2}{n}$. Using (3.32) we have that

$$\sum_{i=1}^{n} M(X_i)=\frac{6}{(\theta_0-\varepsilon)^4}\sum_{i=1}^{n} X_i+\frac{2n}{(\theta_0-\varepsilon)^3}=\frac{2n}{(\theta_0-\varepsilon)^4}\left(3\hat{\theta}_n(\boldsymbol{X})+\theta_0-\varepsilon\right).$$

Therefore,

$$\begin{aligned}
&\left[\mathrm{E}\left(\left(\sum_{i=1}^{n} M(X_i)\right)^2\middle|\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right|<\varepsilon\right)\right]^{\frac{1}{2}}\\
&=\left[\mathrm{E}\left(\left(\frac{2n}{(\theta_0-\varepsilon)^4}(3\hat{\theta}_n(\boldsymbol{X})+\theta_0-\varepsilon)\right)^2\middle|\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right|<\varepsilon\right)\right]^{\frac{1}{2}}\\
&\leq\frac{2n}{(\theta_0-\varepsilon)^4}\left[\mathrm{E}\left(\left(3\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right|+4\theta_0-\varepsilon\right)^2\middle|\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right|<\varepsilon\right)\right]^{\frac{1}{2}}\\
&<\frac{2n}{(\theta_0-\varepsilon)^4}\left[(2\varepsilon+4\theta_0)^2\right]^{\frac{1}{2}},\qquad \text{as } 4\theta_0-\varepsilon>0\\
&=\frac{4n(2\theta_0+\varepsilon)}{(\theta_0-\varepsilon)^4}.
\end{aligned}$$

The quantity $\left[\mathrm{E}\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^4\right]^{\frac{1}{2}}$ is calculated using the results in p.73 and the equations (3.38), p.70 of Kendall and Stuart (1969) along with the fact that $\hat{\theta}_n(\boldsymbol{X})=\bar{X}\sim G\left(n,\frac{n}{\theta_0}\right)$. We obtain that $\mathrm{E}(\hat{\theta}_n(\boldsymbol{X})-\theta_0)^4=\frac{3\theta_0^4}{n^2}\left(\frac{2}{n}+1\right)$. Therefore,

$$\begin{aligned}
&\left[\mathrm{E}\left(\left(\sum_{i=1}^{n} M(X_i)\right)^2\middle|\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right|<\varepsilon\right)\right]^{\frac{1}{2}}\left[\mathrm{E}\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^4\right]^{\frac{1}{2}}\\
&<\frac{4n(2\theta_0+\varepsilon)}{(\theta_0-\varepsilon)^4}\left[\frac{3\theta_0^4}{n^2}\left(\frac{2}{n}+1\right)\right]^{\frac{1}{2}}=\frac{4(2\theta_0+\varepsilon)}{(\theta_0-\varepsilon)^4}\left[3\theta_0^4\left(\frac{2}{n}+1\right)\right]^{\frac{1}{2}}. \qquad (3.33)
\end{aligned}$$

We are now in a position to apply the general result in (3.1) to the specific case of i.i.d. random variables, which follow the exponential distribution with rate parameter $\frac{1}{\theta_0}$. There is, as in the canonical case, a trade-off on the choice of ε. On the one hand, the bound related to the third term in (3.23) requires ε to be large. On the other hand,

for the upper bound of $\left[\mathrm{E}\left((\sum_{i=1}^{n} M(X_i))^2 \middle| |\hat{\theta}_n(\boldsymbol{X}) - \theta_0| < \varepsilon\right)\right]^{\frac{1}{2}}$ as shown in (3.33) it would be better to have ε away from θ_0. It has already been explained that $\varepsilon \in (0, \theta_0)$ and we choose again $\varepsilon = \frac{\theta_0}{2}$, which yields the result in (3.30). □

Remark 3.4. ***(1)*** *The rate of convergence related to the sample size of the above upper bound is* $\frac{1}{\sqrt{n}}$ *and the bound does not depend on* θ_0.

(2) *The first term of the upper bound in* (3.30) *is the same as the one in* (3.26). *However, the rest of the bound is larger in* (3.30) *than in* (3.26) $\forall n \in \mathbb{N}$.

(3) *In the specific setting of independent, exponentially distributed random variables with rate parameter* $\frac{1}{\theta_0}$, *the MLE exists, it is unique and equal to* $\hat{\theta}_n(\boldsymbol{X}) = \bar{X}$. *Define* $W = \frac{\sqrt{n}(\bar{X}-\theta_0)}{\theta_0} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i$, *where* $Y_i = \frac{X_i-\theta_0}{\theta_0}$ *are independent, zero mean and unit variance random variables. In addition,* $\mathrm{E}(W) = 0$ *and* $\mathrm{Var}(W) = \frac{n}{\theta_0^2}\mathrm{Var}(\bar{X}) = \frac{1}{n\theta_0^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = 1$. *Therefore, the result in* (2.12) *gives*

$$d_{bW}\left(\sqrt{n\,i(\theta_0)}(\hat{\theta}_n(\boldsymbol{X}) - \theta_0), Z\right) \leq \frac{1}{\sqrt{n}}\left(2 + \frac{1}{\theta_0^3}\mathrm{E}|X_1 - \theta_0|^3\right) < \frac{4.41456}{\sqrt{n}}. \quad (3.34)$$

The upper bound given in (3.34) *as a result of the direct use of Stein's method is smaller than the upper bound given in* (3.30) *using the general method from Section 3.2. However, in order to apply Stein's method directly, we require the quantity* $\sqrt{n\,i(\theta_0)}(\hat{\theta}_n(\boldsymbol{x}) - \theta_0)$ *to be a sum of independent random variables. The general method, on the other hand, gives an upper bound for* (3.1), *whatever the MLE is, as long as the assumptions (A1)-(A5) hold.*

In the next subsection results based on simulations are presented for both the canonical and the non-canonical parametrisation of the exponential distribution and for various sample sizes.

3.3.2 Empirical results

Here, we study the accuracy of our bounds by simulations. We start by generating 10,000 trials of n random independent observations, x, from the exponential distribution. The means for the canonical and the non-canonical case are equal to 1 and 2, respectively. The MLE, $\hat{\theta}_n(\boldsymbol{X})$, is evaluated in each trial, which in turn gives a vector of 10,000 values. After these values are standardised, we apply to them the function $h(x) = \frac{1}{x^2+2}$ and we calculate their sample mean, denoted by $\hat{\mathrm{E}}\left[h\left(\sqrt{n\,i(\theta_0)}\,(\hat{\theta}_n(\boldsymbol{X}) - \theta_0)\right)\right]$. The function h is a member of the class H_{bW} as in (2.6) and $\|h\| = 0.5$, $\|h'\| = \frac{3\sqrt{1.5}}{16}$. We use these values to calculate the bound in (3.6). We define

$$Q_h(\theta_0) := \left|\hat{\mathrm{E}}\left[h\left(\sqrt{n\,i(\theta_0)}\,(\hat{\theta}_n(\boldsymbol{X}) - \theta_0)\right)\right] - \tilde{\mathrm{E}}[h(Z)]\right|,$$

where $\tilde{\mathrm{E}}[h(Z)] = 0.379$ is the approximation of $\mathrm{E}[h(Z)]$ up to three decimal places. We compare $Q_h(\theta_0)$ with the bound in (3.6), using the difference between their values as a measure of the error. The results presented in the following tables are based on this particular function h while the bounded Wasserstein metric is a supremum over a broader class of test functions, given in (2.6). The results from the simulations are shown in Tables 3.1 and 3.2.

Table 3.1: Simulation results from the Exp(1) distribution

n	$Q_h(\theta_0)$	Upper bound	Error
10	0.008	1.955	1.947
100	0.002	0.336	0.334
1000	0.001	0.094	0.093
10,000	0.0002	0.029	0.0288
100,000	0.0001	0.009	0.0089

Table 3.2: Simulation results from the Exp(0.5) distribution treated as a non-canonical exponential family

n	$Q_h(\theta_0)$	Upper bound	Error	Bound from Lemma 2.2
10	0.004	11.888	11.884	0.321
100	0.002	3.401	3.399	0.101
1000	0.002	1.058	1.056	0.032
10,000	0.001	0.333	0.332	0.010
100,000	0.0005	0.105	0.1045	0.003

The tables indicate that $Q_h(\theta_0)$, the bound and the error, decrease as the sample size gets larger. All the values in Table 3.1 are smaller than the respective ones in Table 3.2, as expected from Remark 3.4. The bounds are not very good for n up to 100. For the non-canonical case the bounds using directly Lemma 2.2 are, as expected, much better than those from the general approach. The bounds are conceptual and better constants may be possible.

As a graphical tool, both the density plots and the normal Quantile-Quantile (Q-Q) plot (for $n = 1000$) for the two examples are used, in order to observe the behaviour of the standardised values of the MLE as the sample size varies. In addition, the bar plots give the magnitude of the error against the sample size in order to capture the performance of Stein's method for each example. Figures 3.1 and 3.2 show the density plots, while the Q-Q plots with the bar plots for the canonical and the non-canonical case are presented in Figures 3.3 and 3.4, respectively.

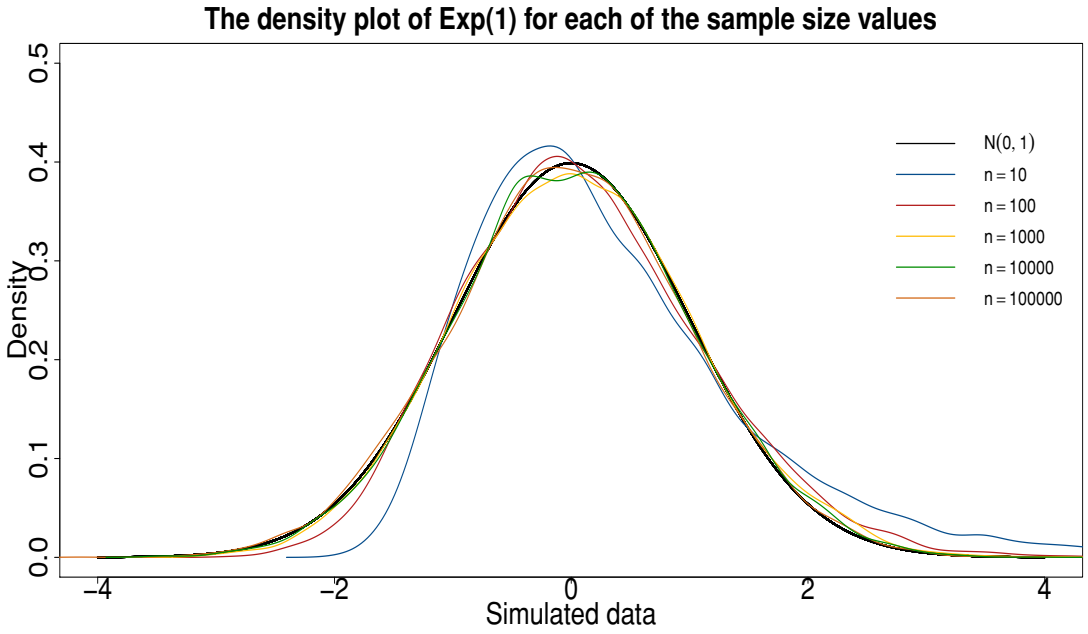

Fig. 3.1: The density function of $\sqrt{n\,i(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right)$ for various sample sizes. In this case $X_1, X_2, \cdots, X_n$ are i.i.d. random variables which follow the Exp(1) distribution.

The density plots verify what has already been observed in Tables 3.1 and 3.2. The distribution of the standardised MLE gets closer to the standard normal as the sample size increases. Specifically, for both examples, the curves produced when $n = 10$ do not approximate the simulated standard normal distribution well. There is a significant but not adequate improvement when $n = 100$, since the density plots generated are closer to the density of the N(0,1) distribution. Therefore, if the sample size gets greater than 100, then a significant drop in the difference between the distributional distance and the upper bound provided by Stein's method is expected. The 'Error' columns of the tables verify this conjecture.

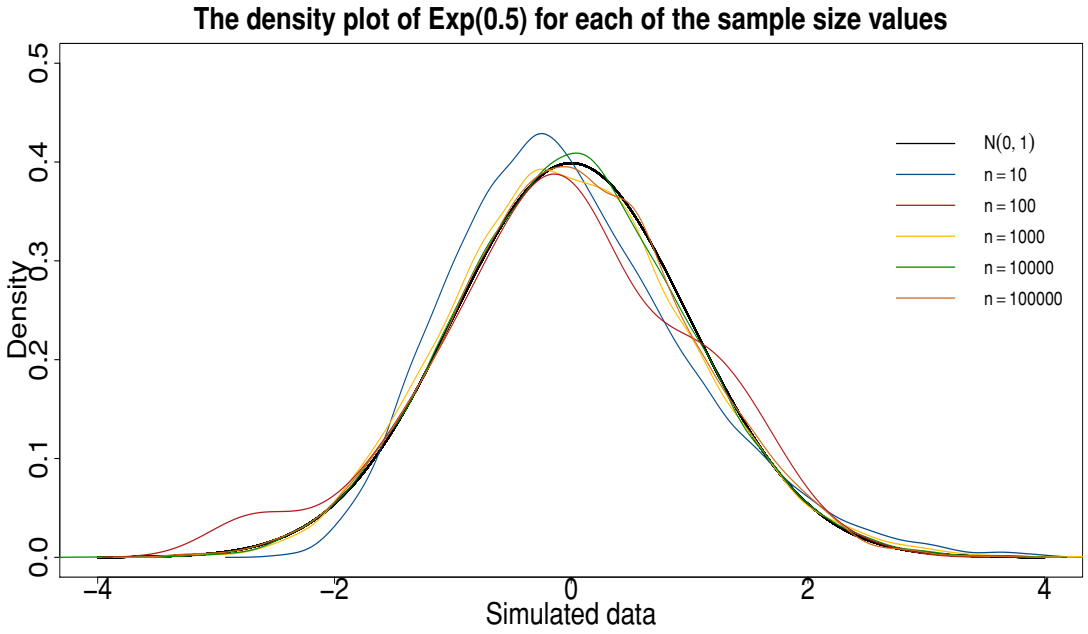

Fig. 3.2: The density function of $\sqrt{n\,i(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right)$ for various sample sizes. In this case $X_1, X_2, \cdots, X_n$ are i.i.d. random variables which follow the $\text{Exp}(0.5)$ distribution.

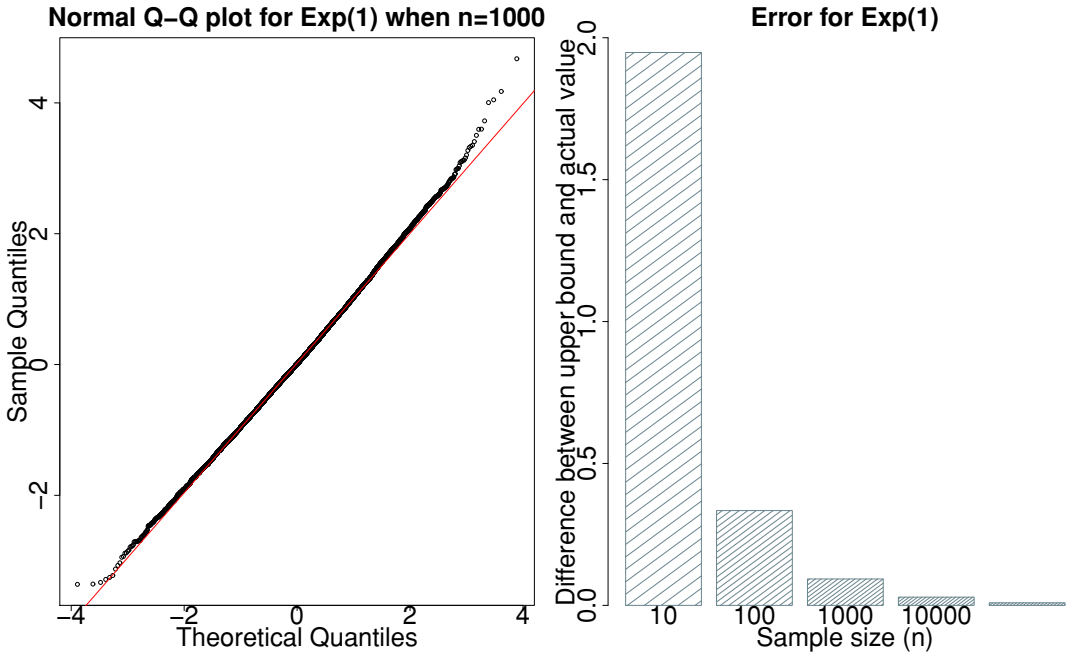

Fig. 3.3: The Q-Q plot, when $n = 1000$, and the bar plot of the error for different sample sizes for the exponential distribution example, $\text{Exp}(1)$.

Figures 3.3 and 3.4 present two symmetric normal Q-Q plots, with the majority of the observations lying on the straight line that passes through the first and third quartiles.

There are a few extreme values in the tails of the plots, but the number of those cases is relatively small, compared to the total of 10,000 values for the standardised MLE. The bar plots presented in the figures show graphically what we concluded from the density plots. When the sample size, n, is greater than 100, the error drops very quickly.

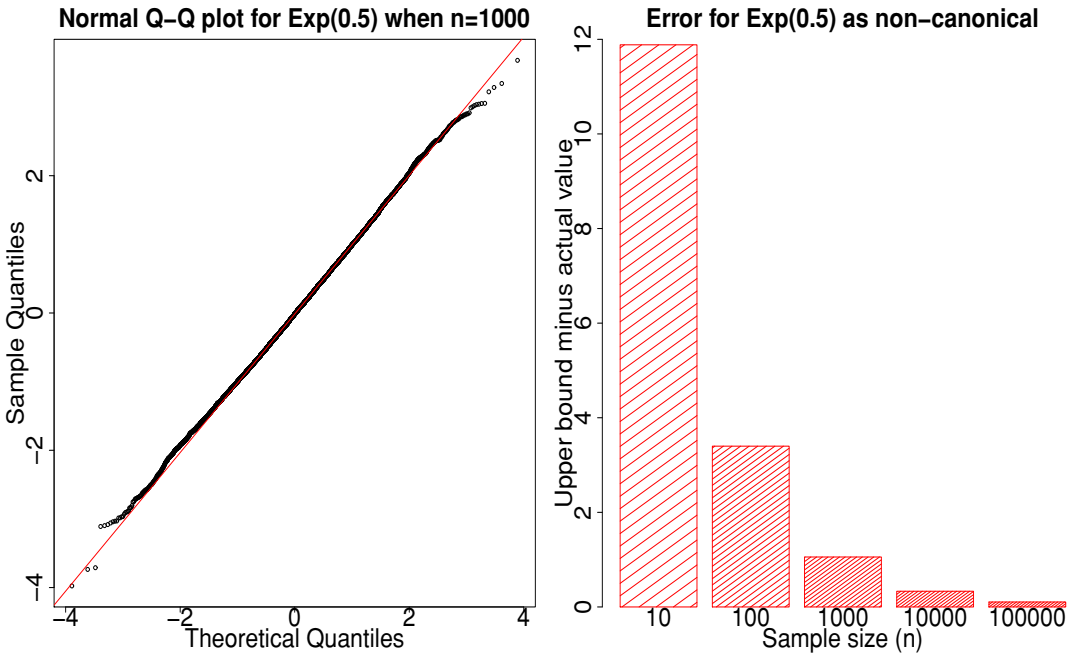

Fig. 3.4: The Q-Q plot, when $n = 1000$, and the bar plot of the error for different sample sizes for the exponential distribution example, $\text{Exp}(0.5)$.

3.4 Discrete distributions: The boundary issue

In this section, we consider discrete distributions that face the problem of the MLE having positive probability of being on the boundary of the parameter space. This issue causes complications for results which are based on asymptotic properties that require both the MLE and the parameter to be interior to the parameter space, such as Taylor expansions. The idea we follow is based on a perturbation method applied both on the parameter and the sample. The new, perturbed values of the parameter and the MLE will be interior to Θ, allowing us to apply part of the general results as explained

in Section 3.2. The above method is assessed through the Poisson distribution with parameter θ and parameter space $\Theta = [0, \infty)$.

3.4.1 The perturbation approach

A perturbation method based on a perturbation function, should be such that first of all, the function should perturb the quantity of interest in a way that ensures it will be interior to its domain. The second requirement is that the perturbed quantity should be as close as possible to the initial quantity. Let X be a random variable with support B, which is a connected, closed (or semi-closed) interval $[a,b]$ ($(a,b]$ or $[a,b)$), where $-\infty < a < b < \infty$. For $0 < \delta < \frac{b-a}{2}$, we are looking for a perturbation function, $q : B \to \overset{\circ}{B}$ (where in this case, $\overset{\circ}{B}$ denotes the interior of the set B) with $q(x) = kx + d$, such that:

(1) $q(a) = a + \delta$ and $q(b) = b - \delta$.

(2) $\sup_{x \in B} |q(x) - x|$ is minimal.

Solving this problem for k and d, gives $k = 1 - \frac{2\delta}{b-a}$ and $d = \delta + \frac{2a}{b-a}\delta$. There is only one solution, which minimises $\sup_{x \in B} |q(x) - x|$. Thus, the second requirement is also satisfied. We choose $\delta = \delta(n) = \frac{c}{n}$ with $0 < c < \frac{n(b-a)}{2}$. Finally, the perturbation function is

$$q_c(x) = x + \frac{c}{n} - \frac{2c}{n}\left(\frac{x-a}{b-a}\right), \; x \in B. \tag{3.35}$$

In the case where $B = (-\infty, b]$ or $B = [a, \infty)$, then $q_c(x) = x - \frac{c}{n}$ or $q_c(x) = x + \frac{c}{n}$, respectively.

Assuming existence and uniqueness of the MLE, $\hat{\theta}_n(\boldsymbol{X})$, for the parameter θ_0, of a discrete distribution with parameter space as in the previous paragraph, the aim is to find an upper bound on

$$d_{bW}\left(\sqrt{n}(\hat{\theta}_n(\boldsymbol{X}) - \theta_0), K\right),$$

where $K \sim \mathrm{N}\left(0, \frac{1}{i(\theta_0)}\right)$. The quantity to bound is not exactly the one shown in (3.1), because the Expected Fisher Information number might not exist or not be finite when θ_0 lies on the boundary of the parameter space. For this purpose, the perturbation function in (3.35) is used for both the parameter and the data.

First we introduce some notation. For S being the discrete sample space, let $a := \inf\Theta$, $b := \sup\Theta$, $\underline{S} := \inf S$, $\overline{S} := \sup S$ and $0 < c_1 < \frac{n(b-a)}{2}$, $0 < c_2 < \frac{n(\overline{S}-\underline{S})}{2}$. In addition,

$$q_{c_1}(\theta_0) = \theta_0 + \frac{c_1}{n} - \frac{2c_1}{n}\left(\frac{\theta_0 - a}{b-a}\right) \tag{3.36}$$

is the perturbed parameter and

$$q_{c_2}(x_i) = x_i + \frac{c_2}{n} - \frac{2c_2}{n}\left(\frac{x_i - \underline{S}}{\overline{S}-\underline{S}}\right)$$

is the perturbed value for x_i. The perturbed data is

$$q_{c_2}(\boldsymbol{x}) = (q_{c_2}(x_1), q_{c_2}(x_2), \ldots, q_{c_2}(x_n)). \tag{3.37}$$

We now assume that the domain of the MLE, $\hat{\theta}_n(\boldsymbol{x})$, can be extended so that $\hat{\theta}_n(q_{c_2}(\boldsymbol{x})) := \hat{\theta}_n(\boldsymbol{y})|_{\boldsymbol{y}=q_{c_2}(\boldsymbol{x})}$ is defined and gives values that belong to the parameter space Θ. The perturbed MLE is denoted by

$$\hat{\theta}_n^*(\boldsymbol{x}) := \hat{\theta}_n(q_{c_2}(\boldsymbol{x})) := \hat{\theta}_n(\boldsymbol{y})|_{\boldsymbol{y}=q_{c_2}(\boldsymbol{x})}. \tag{3.38}$$

Also,

$$l'(q_{c_1}(\theta_0); q_{c_2}(\boldsymbol{x})) := l'(\theta; \boldsymbol{y})\Big|_{\substack{\theta=q_{c_1}(\theta_0),\\ \boldsymbol{y}=q_{c_2}(\boldsymbol{x})}} \qquad l'\left(\hat{\theta}_n^*(\boldsymbol{x}); q_{c_2}(\boldsymbol{x})\right) := l'(\theta; \boldsymbol{y})\Big|_{\substack{\theta=\hat{\theta}_n^*(\boldsymbol{x}),\\ \boldsymbol{y}=q_{c_2}(\boldsymbol{x})}}$$

$$l''(q_{c_1}(\theta_0); q_{c_2}(\boldsymbol{x})) := l''(\theta; \boldsymbol{y})\Big|_{\substack{\theta=q_{c_1}(\theta_0),\\ \boldsymbol{y}=q_{c_2}(\boldsymbol{x})}} \qquad l^{(3)}(\theta; q_{c_2}(\boldsymbol{x})) = l^{(3)}(\theta; \boldsymbol{y})\Big|_{\boldsymbol{y}=q_{c_2}(\boldsymbol{x})}.$$

For ease of presentation, abbreviate

$$Y_i = \frac{1}{\sqrt{n} i\left(q_{c_1}(\theta_0)\right)} \left[\frac{\mathrm{d}}{\mathrm{d}\theta} \log f\left(U_i|\theta\right) \Big|_{\substack{\theta = q_{c_1}(\theta_0) \\ U_i = q_{c_2}(X_i)}} \right], \; i \in \{1, 2, \ldots, n\} \tag{3.39}$$

while

$$w_1 := w_1\left(n, c_1, c_2, \theta_0\right) \;\text{ and }\; w_2 := w_2\left(n, c_1, c_2, \theta_0\right)$$

are its expectation and variance, respectively.

Theorem 3.2. *Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables which follow a single-parameter discrete distribution with parameter space the connected, closed or semi-closed interval $\Theta \subset \mathbb{R}$ and discrete sample space S. Assume that $i(\theta_0) > 0$ and let $\frac{1}{i(\theta_0)} = 0$ be the continuous extension of $\frac{1}{i(\theta)}$ to θ_0 when θ_0 is such that $i(\theta)$ approaches negative or positive infinity as θ approaches θ_0. Let $h \in H_{bW}$ and $0 < \varepsilon = \varepsilon(q_{c_1}(\theta_0)) = \varepsilon_0$ such that as in (R3), $\left| \frac{\mathrm{d}^3}{\mathrm{d}\theta^3} \log f(q_{c_2}(x)|\theta) \right| \leq M(q_{c_2}(x))$, $\forall x \in S$, $\forall \theta \in \overset{\circ}{\Theta}$ such that $|\theta - q_{c_1}(\theta_0)| < \varepsilon_0$, where $\overset{\circ}{\Theta}$ denotes the interior of the parameter space Θ. Also, assume that $\mathrm{E}\left(\left(\sum_{i=1}^n M\left(q_{c_2}(X_i)\right)\right)^2 \middle| \left|\hat{\theta}_n^*(\boldsymbol{X}) - q_{c_1}(\theta_0)\right| < \varepsilon \right) < \infty$. Let $K \sim \mathrm{N}\left(0, \frac{1}{i(\theta_0)}\right)$. Then for $0 < c_1 < \frac{n(b-a)}{2}$, $0 < c_2 < \frac{n(\overline{S}-\underline{S})}{2}$, $q_{c_1}(\theta_0)$ as in (3.36) and $q_{c_2}(\boldsymbol{x})$ as in (3.37),*

$$\begin{aligned}
d_{bW}\left(\sqrt{n}\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right), K\right) &\leq \frac{c_1}{\sqrt{n}} \left| 1 - 2\left(\frac{\theta_0 - a}{b - a}\right) \right| + \sqrt{n}\mathrm{E}\left|\hat{\theta}_n(\boldsymbol{X}) - \hat{\theta}_n^*(\boldsymbol{X})\right| \\
&+ \left[\left| 1 - \frac{1}{\sqrt{w_2 n\, i(\theta_0)}} \right| \sqrt{n\, w_2 + (n\, w_1)^2} + \frac{\sqrt{n}|w_1|}{\sqrt{w_2\, i(\theta_0)}} \right] \\
&+ \frac{1}{\sqrt{n}} \left(2 + \frac{1}{(w_2)^{\frac{3}{2}}} \mathrm{E}|Y_1 - w_1|^3 \right) \mathbb{1}\left\{ \frac{1}{i(\theta_0)} > 0 \right\} + \frac{2}{\varepsilon^2} \mathrm{E}\left[\left(\hat{\theta}_n^*(\boldsymbol{X}) - q_{c_1}(\theta_0)\right)^2 \right] \\
&+ \frac{1}{\sqrt{n} i\left(q_{c_1}(\theta_0)\right)} \left\{ \sqrt{\mathrm{E}\left[l''\left(q_{c_1}(\theta_0); q_{c_2}(\boldsymbol{X})\right) + n i\left(q_{c_1}(\theta_0)\right) \right]^2} \sqrt{\mathrm{E}\left[\left(\hat{\theta}_n^*(\boldsymbol{X}) - q_{c_1}(\theta_0)\right)^2 \right]} \right. \\
&\quad + \frac{1}{2} \left[\mathrm{E}\left(\left(\sum_{i=1}^n M\left(q_{c_2}(X_i)\right) \right)^2 \middle| \left|\hat{\theta}_n^*(\boldsymbol{X}) - q_{c_1}(\theta_0)\right| < \varepsilon \right) \right]^{\frac{1}{2}} \\
&\quad \left. \times \left[\mathrm{E}\left[\left(\hat{\theta}_n^*(\boldsymbol{X}) - q_{c_1}(\theta_0)\right)^4 \right] \right]^{\frac{1}{2}} \right\}.
\end{aligned} \tag{3.40}$$

Proof. For ease of presentation, the proof of the theorem is split into three steps.

Step 1: Perturbation of θ_0. Using the triangle inequality and then a first order Taylor expansion of $h(\sqrt{n}(\hat{\theta}_n(\boldsymbol{X})-\theta_0))$ about $\sqrt{n}\left(\hat{\theta}_n(\boldsymbol{X})-q_{c_1}(\theta_0)\right)$ gives

$$\begin{aligned}
&\left|\mathrm{E}\left[h\left(\sqrt{n}\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)\right)\right]-\mathrm{E}[h(K)]\right| \leq \left|\mathrm{E}\left[h\left(\sqrt{n}\left(\hat{\theta}_n(\boldsymbol{X})-q_{c_1}(\theta_0)\right)\right)\right]-\mathrm{E}[h(K)]\right| \\
&+\left|\mathrm{E}\left[h\left(\sqrt{n}\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)\right)-h\left(\sqrt{n}\left(\hat{\theta}_n(\boldsymbol{X})-q_{c_1}(\theta_0)\right)\right)\right]\right| \\
&\leq \left|\mathrm{E}\left[h\left(\sqrt{n}\left(\hat{\theta}_n(\boldsymbol{X})-q_{c_1}(\theta_0)\right)\right)\right]-\mathrm{E}[h(K)]\right|+\sqrt{n}\|h'\|\mathrm{E}\left|q_{c_1}(\theta_0)-\theta_0\right| \\
&= \left|\mathrm{E}\left[h\left(\sqrt{n}\left(\hat{\theta}_n(\boldsymbol{X})-q_{c_1}(\theta_0)\right)\right)\right]-\mathrm{E}[h(K)]\right|+\frac{\|h'\|c_1}{\sqrt{n}}\left|1-2\left(\frac{\theta_0-a}{b-a}\right)\right|. \qquad (3.41)
\end{aligned}$$

Step 2: Perturbation of the MLE. To perturb the MLE, we perturb the data. This construction ensures that the MLE evaluated at $q_{c_2}(\boldsymbol{x})$ (as defined in (3.37)) is not on the boundary of the parameter space. Following the same process as in (3.41), using the triangle inequality and a first order Taylor expansion of $h\left(\sqrt{n}\left(\hat{\theta}_n(\boldsymbol{X})-q_{c_1}(\theta_0)\right)\right)$ about $\sqrt{n}\left(\hat{\theta}_n^*(\boldsymbol{X})-q_{c_1}(\theta_0)\right)$, where $\hat{\theta}_n^*(\boldsymbol{X})$ is as in (3.38), gives

$$\begin{aligned}
&\left|\mathrm{E}\left[h\left(\sqrt{n}\left(\hat{\theta}_n(\boldsymbol{X})-q_{c_1}(\theta_0)\right)\right)\right]-\mathrm{E}[h(K)]\right| \\
&\leq \left|\mathrm{E}\left[h\left(\sqrt{n}\left(\hat{\theta}_n^*(\boldsymbol{X})-q_{c_1}(\theta_0)\right)\right)\right]-\mathrm{E}[h(K)]\right| \\
&\quad+\left|\mathrm{E}\left[h\left(\sqrt{n}\left(\hat{\theta}_n(\boldsymbol{X})-q_{c_1}(\theta_0)\right)\right)-h\left(\sqrt{n}\left(\hat{\theta}_n^*(\boldsymbol{X})-q_{c_1}(\theta_0)\right)\right)\right]\right| \\
&\leq \left|\mathrm{E}\left[h\left(\sqrt{n}\left(\hat{\theta}_n^*(\boldsymbol{X})-q_{c_1}(\theta_0)\right)\right)\right]-\mathrm{E}[h(K)]\right|+\sqrt{n}\|h'\|\mathrm{E}\left|\hat{\theta}_n(\boldsymbol{X})-\hat{\theta}_n^*(\boldsymbol{X})\right|. \qquad (3.42)
\end{aligned}$$

Step 3: The final bound. It remains to bound

$$\left|\mathrm{E}\left[h\left(\sqrt{n}\left(\hat{\theta}_n^*(\boldsymbol{X})-q_{c_1}(\theta_0)\right)\right)\right]-\mathrm{E}[h(K)]\right|.$$

Since both $q_{c_1}(\theta_0)$ and $\hat{\theta}_n^*(\boldsymbol{x})$ are interior to Θ, a second-order Taylor expansion of $l'\left(\hat{\theta}_n^*(\boldsymbol{x}); q_{c_2}(\boldsymbol{x})\right) = l'\left(\hat{\theta}_n\left(q_{c_2}(\boldsymbol{x})\right); q_{c_2}(\boldsymbol{x})\right) = 0$ about $q_{c_1}(\theta_0)$ yields

$$0 = l'\left(q_{c_1}(\theta_0); q_{c_2}(\boldsymbol{x})\right) + \left(\hat{\theta}_n^*(\boldsymbol{x}) - q_{c_1}(\theta_0)\right) l''\left(q_{c_1}(\theta_0); q_{c_2}(\boldsymbol{x})\right) + R_1\left(q_{c_1}(\theta_0); q_{c_2}(\boldsymbol{x})\right), \tag{3.43}$$

where, similarly as in Section 3.2,

$$R_1\left(q_{c_1}(\theta_0); q_{c_2}(\boldsymbol{x})\right) = \frac{1}{2}\left(\hat{\theta}_n^*(\boldsymbol{x}) - q_{c_1}(\theta_0)\right)^2 l^{(3)}\left(\tilde{\theta}; q_{c_2}(\boldsymbol{x})\right)$$

with

$$l^{(3)}\left(\tilde{\theta}; q_{c_2}(\boldsymbol{x})\right) = l^{(3)}(\theta; \boldsymbol{y})\Big|_{\substack{\theta=\tilde{\theta} \\ \boldsymbol{y}=q_{c_2}(\boldsymbol{x})}}$$

for $\tilde{\theta}$ between $\hat{\theta}_n^*(\boldsymbol{x})$ and $q_{c_1}(\theta_0)$. A simple rearrangement of the terms in (3.43), leads to

$$\hat{\theta}_n^*(\boldsymbol{x}) - q_{c_1}(\theta_0) = \frac{-l'\left(q_{c_1}(\theta_0); q_{c_2}(\boldsymbol{x})\right) - R_1\left(q_{c_1}(\theta_0); q_{c_2}(\boldsymbol{x})\right)}{l''\left(q_{c_1}(\theta_0); q_{c_2}(\boldsymbol{x})\right)}.$$

In general $l''\left(q_{c_1}(\theta_0); q_{c_2}(\boldsymbol{x})\right) \neq -n\, i\left(q_{c_1}(\theta_0)\right)$ and similar steps to those used in the proof of Theorem 3.1 yield

$$\hat{\theta}_n^*(\boldsymbol{x}) - q_{c_1}(\theta_0) = \frac{l'\left(q_{c_1}(\theta_0); q_{c_2}(\boldsymbol{x})\right) + R_1\left(q_{c_1}(\theta_0); q_{c_2}(\boldsymbol{x})\right) + R_2\left(q_{c_1}(\theta_0); q_{c_2}(\boldsymbol{x})\right)}{n\, i\left(q_{c_1}(\theta_0)\right)},$$

where

$$R_2\left(q_{c_1}(\theta_0); q_{c_2}(\boldsymbol{x})\right) = \left(\hat{\theta}_n^*(\boldsymbol{x}) - q_{c_1}(\theta_0)\right)\left[l''\left(q_{c_1}(\theta_0); q_{c_2}(\boldsymbol{x})\right) + n\, i\left(q_{c_1}(\theta_0)\right)\right].$$

Using that $q_{c_2}(\boldsymbol{X}) = (q_{c_2}(X_1), q_{c_2}(X_2), \ldots, q_{c_2}(X_n))$, the triangle inequality gives

$$
\begin{aligned}
&\left|\mathrm{E}\left[h\left(\sqrt{n}\left(\hat{\theta}_n^*(\boldsymbol{X})-q_{c_1}(\theta_0)\right)\right)\right]-\mathrm{E}[h(K)]\right| \\
&\leq \left|\mathrm{E}\left[h\left(\frac{l'\left(q_{c_1}(\theta_0);q_{c_2}(\boldsymbol{X})\right)}{\sqrt{n}\,i\left(q_{c_1}(\theta_0)\right)}\right)\right]-\mathrm{E}[h(K)]\right| \qquad (3.44) \\
&\quad+\left|\mathrm{E}\left[h\left(\frac{l'\left(q_{c_1}(\theta_0);q_{c_2}(\boldsymbol{X})\right)+R_1\left(q_{c_1}(\theta_0);q_{c_2}(\boldsymbol{X})\right)+R_2\left(q_{c_1}(\theta_0);q_{c_2}(\boldsymbol{X})\right)}{\sqrt{n}\,i\left(q_{c_1}(\theta_0)\right)}\right)\right. \\
&\qquad\left.-h\left(\frac{l'\left(q_{c_1}(\theta_0);q_{c_2}(\boldsymbol{X})\right)}{\sqrt{n}\,i\left(q_{c_1}(\theta_0)\right)}\right)\right]\Bigg|. \qquad (3.45)
\end{aligned}
$$

(A) To find an upper bound on (3.44) using Lemma 2.2, note that

$$
\frac{l'\left(q_{c_1}(\theta_0);q_{c_2}(\boldsymbol{X})\right)}{\sqrt{n}\,i\left(q_{c_1}(\theta_0)\right)}=\sum_{i=1}^{n}Y_i,
$$

where Y_i as in (3.39) and $w_1 := w_1(n,c_1,c_2,\theta_0)$, $w_2 := w_2(n,c_1,c_2,\theta_0)$ are its expectation and variance, respectively. These quantities depend on the sample size and on the perturbed values ($q_{c_1}(\theta_0)$ and $q_{c_2}(x_i)$). Define $\tilde{Y}_i=\frac{Y_i-w_1}{\sqrt{w_2\,i(\theta_0)}},\forall i\in\{1,2,\ldots,n\}$ with $\mathrm{E}(\tilde{Y}_i)=0$ and $\mathrm{Var}(\tilde{Y}_i)=\frac{1}{i(\theta_0)}$. As a consequence of $X_1,X_2,\ldots,X_n$ being i.i.d. random variables, $\tilde{Y}_1,\tilde{Y}_2,\ldots,\tilde{Y}_n$ are i.i.d. random variables too. Using the triangle inequality and that

$$
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{Y}_i=\frac{1}{\sqrt{w_2 n\,i(\theta_0)}}\left(\frac{l'\left(q_{c_1}(\theta_0);q_{c_2}(\boldsymbol{X})\right)}{\sqrt{n}i\left(q_{c_1}(\theta_0)\right)}-n\,w_1\right) \qquad (3.46)
$$

gives

$$
\begin{aligned}
&\left|\mathrm{E}\left[h\left(\frac{l'\left(q_{c_1}(\theta_0);q_{c_2}(\boldsymbol{X})\right)}{\sqrt{n}\,i\left(q_{c_1}(\theta_0)\right)}\right)\right]-\mathrm{E}[h(K)]\right|\leq \\
&\left|\mathrm{E}\left[h\left(\frac{1}{\sqrt{w_2 n\,i(\theta_0)}}\left(\frac{l'\left(q_{c_1}(\theta_0);q_{c_2}(\boldsymbol{X})\right)}{\sqrt{n}\,i\left(q_{c_1}(\theta_0)\right)}-n\,w_1\right)\right)\right]-\mathrm{E}[h(K)]\right| \\
&+\left|\mathrm{E}\left[h\left(\frac{l'\left(q_{c_1}(\theta_0);q_{c_2}(\boldsymbol{X})\right)}{\sqrt{n}\,i\left(q_{c_1}(\theta_0)\right)}\right)-h\left(\frac{1}{\sqrt{w_2 n\,i(\theta_0)}}\left(\frac{l'\left(q_{c_1}(\theta_0);q_{c_2}(\boldsymbol{X})\right)}{\sqrt{n}\,i\left(q_{c_1}(\theta_0)\right)}-n\,w_1\right)\right)\right]\right|. \qquad (3.47)
\end{aligned}
$$

The first term of the bound in (3.47) will be bounded using Lemma 2.2 with $W = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{Y}_i$ as in (3.46). Thus,

$$\begin{aligned}
&\left|\mathrm{E}\left[h\left(\frac{1}{\sqrt{w_2 n\, i(\theta_0)}}\left(\frac{l'(q_{c_1}(\theta_0);q_{c_2}(\boldsymbol{X}))}{\sqrt{n}\, i(q_{c_1}(\theta_0))} - n\, w_1\right)\right)\right] - \mathrm{E}[h(K)]\right| \\
&\leq \frac{\|h'\|}{\sqrt{n}}\left(2 + [i(\theta_0)]^{\frac{3}{2}}\mathrm{E}\left|\tilde{Y}_1\right|^3\right) = \frac{\|h'\|}{\sqrt{n}}\left(2 + \frac{1}{(w_2)^{\frac{3}{2}}}\mathrm{E}|Y_1 - w_1|^3\right). \qquad (3.48)
\end{aligned}$$

When $\frac{1}{i(\theta_0)} = 0$ then $\tilde{Y}_i = 0, \forall i \in \{1,2,\ldots,n\}$ and the first term of the bound in (3.47) is equal to zero. For the second term of the upper bound in (3.47) a first-order Taylor expansion and the Cauchy-Schwarz inequality yield

$$\begin{aligned}
&\left|\mathrm{E}\left[h\left(\frac{l'(q_{c_1}(\theta_0);q_{c_2}(\boldsymbol{X}))}{\sqrt{n}\, i(q_{c_1}(\theta_0))}\right) - h\left(\frac{1}{\sqrt{w_2 n\, i(\theta_0)}}\left(\frac{l'(q_{c_1}(\theta_0);q_{c_2}(\boldsymbol{X}))}{\sqrt{n}\, i(q_{c_1}(\theta_0))} - n\, w_1\right)\right)\right]\right| \\
&\leq \|h'\|\left|1 - \frac{1}{\sqrt{w_2 n\, i(\theta_0)}}\right|\mathrm{E}\left|\frac{l'(q_{c_1}(\theta_0);q_{c_2}(\boldsymbol{X}))}{\sqrt{n}\, i(q_{c_1}(\theta_0))}\right| + \frac{\|h'\|\sqrt{n}|w_1|}{\sqrt{w_2\, i(\theta_0)}} \\
&\leq \|h'\|\left|1 - \frac{1}{\sqrt{w_2 n\, i(\theta_0)}}\right|\left(\mathrm{Var}\left(\frac{l'(q_{c_1}(\theta_0);q_{c_2}(\boldsymbol{X}))}{\sqrt{n}\, i(q_{c_1}(\theta_0))}\right) + \frac{[\mathrm{E}(l'(q_{c_1}(\theta_0);q_{c_2}(\boldsymbol{X})))]^2}{n\,[i(q_{c_1}(\theta_0))]^2}\right)^{\frac{1}{2}} \\
&\quad + \frac{\|h'\|\sqrt{n}|w_1|}{\sqrt{w_2\, i(\theta_0)}} \\
&= \|h'\|\left[\left|1 - \frac{1}{\sqrt{w_2 n\, i(\theta_0)}}\right|\sqrt{n\, w_2 + (n\, w_1)^2} + \frac{\sqrt{n}|w_1|}{\sqrt{w_2\, i(\theta_0)}}\right]. \qquad (3.49)
\end{aligned}$$

(B) To complete the proof, it remains to find an upper bound for (3.45). The idea is the same as the one used for (3.19). We condition on whether $\left|\hat{\theta}_n^*(\boldsymbol{X}) - q_{c_1}(\theta_0)\right| \geq \varepsilon$ or $\left|\hat{\theta}_n^*(\boldsymbol{X}) - q_{c_1}(\theta_0)\right| < \varepsilon$, where $0 < \varepsilon = \varepsilon(q_{c_1}(\theta_0))$ such that $\left|\frac{\mathrm{d}^3}{\mathrm{d}\theta^3}\log f(u|\theta)\Big|_{u=q_{c_2}(x)}\right| \leq M(q_{c_2}(x))$, $\forall x \in S$, $\forall \theta \in \overset{\circ}{\Theta}$ with $|\theta - q_{c_1}(\theta_0)| < \varepsilon$. Following the same process as in Section 3.2 yields

$$
\begin{aligned}
&\left| \mathrm{E}\left[h\left(\frac{l'\left(q_{c_1}(\theta_0); q_{c_2}(\boldsymbol{X})\right) + R_1\left(q_{c_1}(\theta_0); q_{c_2}(\boldsymbol{X})\right) + R_2\left(q_{c_1}(\theta_0); q_{c_2}(\boldsymbol{X})\right)}{\sqrt{n}\, i\left(q_{c_1}(\theta_0)\right)} \right) \right.\right. \\
&\qquad \left.\left. - h\left(\frac{l'\left(q_{c_1}(\theta_0); q_{c_2}(\boldsymbol{X})\right)}{\sqrt{n}\, i\left(q_{c_1}(\theta_0)\right)} \right) \right] \right| \\
&\leq \frac{2\|h\|}{\varepsilon^2} \mathrm{E}\left[\left(\hat{\theta}_n^*(\boldsymbol{X}) - q_{c_1}(\theta_0) \right)^2 \right] \\
&+ \frac{\|h'\|}{\sqrt{n} i\left(q_{c_1}(\theta_0)\right)} \left\{ \sqrt{\mathrm{E}\left[l''\left(q_{c_1}(\theta_0); q_{c_2}(\boldsymbol{X})\right) + n i\left(q_{c_1}(\theta_0)\right) \right]^2} \sqrt{\mathrm{E}\left[\left(\hat{\theta}_n^*(\boldsymbol{X}) - q_{c_1}(\theta_0) \right)^2 \right]} \right. \\
&\qquad + \frac{1}{2} \left[\mathrm{E}\left(\left(\sum_{i=1}^{n} M\left(q_{c_2}(X_i)\right) \right)^2 \middle| \left| \hat{\theta}_n^*(\boldsymbol{X}) - q_{c_1}(\theta_0) \right| < \varepsilon \right) \right]^{\frac{1}{2}} \\
&\qquad\qquad \left. \times \left[\mathrm{E}\left[\left(\hat{\theta}_n^*(\boldsymbol{X}) - q_{c_1}(\theta_0) \right)^4 \right] \right]^{\frac{1}{2}} \right\}. \qquad (3.50)
\end{aligned}
$$

Combining (3.41), (3.42), (3.48), (3.49) and (3.50) and the fact that $\|h\| \leq 1$, $\|h'\| \leq 1$ gives the result in (3.40). □

Remark 3.5. ***(1)*** *In order for the above bound to approach zero as the sample size, n, increases we require that* $\mathrm{E}\left| \hat{\theta}_n(\boldsymbol{X}) - \hat{\theta}_n^*(\boldsymbol{X}) \right| = o\left(\frac{1}{\sqrt{n}} \right)$.

(2) *When both endpoints of the parameter space are not finite, then parameter perturbation is not necessary. In the case where one of the two endpoints of the now semi-closed parameter space is infinite, then it suffices to change the form of the perturbed parameter, which now becomes*

$$
\begin{aligned}
q_{c_1}(\theta_0) &= \theta_0 - \frac{c_1}{n}, \quad \text{if the left endpoint is equal to } -\infty \\
q_{c_1}(\theta_0) &= \theta_0 + \frac{c_1}{n}, \quad \text{if the right endpoint is equal to } \infty.
\end{aligned}
$$

The same remark holds regarding the sample space and the relevant perturbation of the data.

3.4.2 Example: The Poisson distribution

In this subsection the Poisson distribution with parameter $\theta \in \Theta = [0,\infty)$ is treated. The value $\theta = 0$ must be in the parameter space in order for the MLE, $\hat{\theta}_n(\boldsymbol{X}) = \bar{X}$, to exist and to be unique. The Poisson(θ) distribution with the aforementioned parameter space is not a single-parameter exponential family. When $\theta = 0$ is included in the parameter space the requirements of an exponential family are not satisfied as the set of values x for which the relevant probability mass function

$$f(x|\theta) = \frac{\mathrm{e}^{-\theta}\theta^x}{x!},\ \theta \in [0,\infty),\ x \in \mathbb{Z}_0^+$$

is positive, is different for $\theta = 0$ than for any other value of the parameter θ; the support of the distribution depends on the parameter. As the parameter space is neither an open and connected nor a compact subset of $\mathbb{R}$, we can not use Proposition 2.1 or Proposition 2.2, respectively. Therefore, the steps that prove existence and uniqueness of the MLE are presented below.

Let $X_1, X_2, \cdots, X_n$ be i.i.d. random variables from Poisson(θ), $\theta \in [0,\infty)$. The likelihood function is

$$L(\theta;\boldsymbol{x}) = \prod_{i=1}^{n} f(x_i|\theta) = \frac{\exp\{-n\theta\}\,\theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}.$$

To maximise the likelihood function we account for whether $\sum_{i=1}^n x_i = 0$ or not. Each case is examined separately.

Case 1: If $\sum_{i=1}^n x_i = 0$ (equivalently $x_1 = x_2 = \cdots = x_n = 0$), then

$$L(\theta;\mathbf{0}) = \exp\{-n\theta\}$$

is a decreasing function of θ. It takes its maximum value on $[0,\infty)$ when θ attains its minimum; hence $\hat{\theta}_n(\boldsymbol{X}) = 0 = \bar{X}$.

Case 2: If $\sum_{i=1}^{n} x_i \neq 0$, we follow the usual concept on finding the MLE. The first derivative of the log-likelihood function with respect to θ is

$$l'(\theta;\boldsymbol{x}) = -n + \frac{\sum_{i=1}^{n} x_i}{\theta}. \tag{3.51}$$

We can see that $l'(\theta;\boldsymbol{x}) = 0 \Leftrightarrow \frac{\sum_{i=1}^{n} x_i}{\theta} = n \Leftrightarrow \theta = \hat{\theta}_n(\boldsymbol{X}) = \bar{X}$. In order to verify that $\hat{\theta}_n(\boldsymbol{X}) = \bar{X}$ is the MLE, $l''(\hat{\theta}_n(\boldsymbol{x});\boldsymbol{x}) < 0$ is required. Simple calculations yield

$$l''(\theta;\boldsymbol{x}) = -\frac{\sum_{i=1}^{n} x_i}{\theta^2} \Rightarrow l''(\hat{\theta}_n(\boldsymbol{x});\boldsymbol{x}) = -\frac{n\bar{x}}{\bar{x}^2} = -\frac{n}{\bar{x}} < 0.$$

In conclusion, using the results from Cases 1 and 2 above, we deduce that for $X_i, X_2, \cdots, X_n$ being i.i.d. random variables that follow the Poisson$(\theta), \theta \in [0,\infty)$, the MLE exists, it is unique and equal to $\hat{\theta}_n(\boldsymbol{X}) = \bar{X}$.

Following the steps of the proof of Theorem 3.2, using also Hölder's inequality for the third absolute moment in the fourth term of the bound in (3.40) gives the following result.

Corollary 3.4. *Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables which follow the Poisson(θ_0) distribution, with $\theta_0 \in [0,\infty)$. For $K \sim \mathrm{N}(0,\theta_0)$, $h \in H_{bW}$ as in* (2.6) *and $0 < c_1 \leq c_2$ positive constants and $C = c_2 - c_1 \geq 0$,*

1) if $\theta_0 > 0$ then

$$\begin{aligned} d_{bW}\left(\sqrt{n}(\hat{\theta}_n(\boldsymbol{X}) - \theta_0), K\right) &< \frac{c_1}{\sqrt{n}} + \frac{c_2}{\sqrt{n}} + \frac{1}{\sqrt{n}}\left[2 + \frac{(3\theta_0+1)^{\frac{3}{4}}}{\theta_0^{\frac{3}{4}}}\right] \\ &+ \frac{C}{\sqrt{n}} + \frac{8}{n\left(\theta_0 + \frac{c_1}{n}\right)^2}\left(\theta_0 + \frac{C^2}{n}\right) + \frac{1}{\sqrt{n}\left(\theta_0 + \frac{c_1}{n}\right)}\left(\theta_0 + \frac{C^2}{n}\right) \\ &+ \frac{12}{\sqrt{n}(\theta_0 + \frac{c_1}{n})}\left[\frac{\theta_0}{n}(1+3n\theta_0) + \frac{4C\theta_0}{n} + \frac{6C^2\theta_0}{n} + \frac{C^4}{n^2}\right]^{\frac{1}{2}}. \end{aligned} \tag{3.52}$$

2) if $\theta_0 = 0$ *then*

$$d_{bW}\left(\sqrt{n}\hat{\theta}_n(\boldsymbol{X}), K\right) = 0.$$

Proof. We have already proved that the unique MLE in this case is $\hat{\theta}_n(\boldsymbol{X}) = \bar{X}$. Using the second result of Remark 3.5, the perturbed parameter is $q_{c_1}(\theta_0) = \theta_0 + \frac{c_1}{n}$, while the perturbed data are $q_{c_2}(x_i) = x_i + \frac{c_2}{n}$ so that the perturbed MLE is

$$\hat{\theta}_n^*(\boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^{n}\left(X_i + \frac{c_2}{n}\right) = \hat{\theta}_n(\boldsymbol{X}) + \frac{c_2}{n},$$

where c_1 and c_2 are chosen such that $0 < c_1 \leq c_2$. It is easy to deduce that, in this specific example, the first term of the general bound in (3.40) becomes $\frac{c_1}{\sqrt{n}}$. For the second term,

$$\sqrt{n}\mathrm{E}\left|\hat{\theta}_n(\boldsymbol{X}) - \hat{\theta}_n^*(\boldsymbol{X})\right| = \sqrt{n}\left|-\frac{c_2}{n}\right| = \frac{c_2}{\sqrt{n}}.$$

The expected Fisher information number is

$$i(\theta_0) = \mathrm{Var}\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X|\theta_0)\right) = \frac{1}{\theta_0^2}\mathrm{Var}(X) = \frac{1}{\theta_0}.$$

To find $w_1 := w_1(n)$ and $w_2 := w_2(n)$ in the general bound, using (3.51)

$$Y_i = \frac{q_{c_1}(\theta_0)}{\sqrt{n}}\left(-1 + \frac{q_{c_2}(X_i)}{q_{c_1}(\theta_0)}\right) = \frac{q_{c_2}(X_i) - q_{c_1}(\theta_0)}{\sqrt{n}} = \frac{1}{\sqrt{n}}\left(X_i + \frac{c_2}{n} - \theta_0 - \frac{c_1}{n}\right),$$

results in $w_1 = \frac{c_2 - c_1}{n^{\frac{3}{2}}}$ and $w_2 = \frac{\theta_0}{n}$. Thus, for the third term of the general upper bound in (3.40),

$$\begin{aligned}
&\left|1 - \frac{1}{\sqrt{w_2 n\, i(\theta_0)}}\right|\sqrt{n\, w_2 + (n\, w_1)^2} + \frac{\sqrt{n}|w_1|}{\sqrt{w_2\, i(\theta_0)}} \\
&= \left|1 - \frac{\sqrt{\theta_0}}{\sqrt{\theta_0}}\right|\sqrt{\theta_0 + \frac{c_2 - c_1}{n^2}} + \frac{\sqrt{n}|c_2 - c_1|}{n} = \frac{|c_2 - c_1|}{\sqrt{n}}.
\end{aligned}$$

For the fourth term in (3.40), the equations (3.38), p.70 of Kendall and Stuart (1969), and the results in p.73 of the book show that $\mathrm{E}(X_i-\theta_0)^4=\theta_0(1+3\theta_0)$. This result and Hölder's inequality yield

$$\frac{1}{\sqrt{n}}\left(2+\frac{1}{(w_2)^{\frac{3}{2}}}\mathrm{E}|Y_1-w_1|^3\right)\leq\frac{1}{\sqrt{n}}\left(2+\frac{1}{(w_2)^{\frac{3}{2}}}\left[\mathrm{E}(Y_1-w_1)^4\right]^{\frac{3}{4}}\right)$$
$$=\frac{1}{\sqrt{n}}\left(2+\left(\frac{n}{\theta_0}\right)^{\frac{3}{2}}\left[\mathrm{E}\left(\frac{X_1-\theta_0}{\sqrt{n}}\right)^4\right]^{\frac{3}{4}}\right)=\frac{1}{\sqrt{n}}\left(2+\left(\frac{3\theta_0+1}{\theta_0}\right)^{\frac{3}{4}}\right).$$

Denoting by $C=c_2-c_1\geq 0$, then $\mathrm{E}\left[\left(\hat{\theta}_n^*(\boldsymbol{X})-q_{c_1}(\theta_0)\right)^2\right]=\frac{\theta_0}{n}+\frac{C^2}{n^2}$. Also,

$$\mathrm{E}\left[\left(\hat{\theta}_n^*(\boldsymbol{X})-q_{c_1}(\theta_0)\right)^4\right]=\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0+\frac{C}{n}\right)^4\right]$$
$$=\mathrm{E}\left[\left(\frac{\sum_{i=1}^n X_i-n\theta_0+C}{n}\right)^4\right]$$
$$=\frac{1}{n^4}\mathrm{E}\left[\left(\sum_{i=1}^n X_i-n\theta_0\right)^4\right]+\frac{4C}{n^4}\mathrm{E}\left[\left(\sum_{i=1}^n X_i-n\theta_0\right)^3\right]+\frac{6C^2}{n^4}\mathrm{E}\left[\left(\sum_{i=1}^n X_i-n\theta_0\right)^2\right]$$
$$+\frac{4C^3}{n^4}\mathrm{E}\left(\sum_{i=1}^n X_i-n\theta_0\right)+\frac{C^4}{n^4}.$$

Using again the results in Kendall and Stuart (1969) and the fact that $\sum_{i=1}^n X_i\sim\text{Poisson}(n\theta_0)$ yields

$$\mathrm{E}\left[\left(\hat{\theta}_n^*(\boldsymbol{X})-q_{c_1}(\theta_0)\right)^4\right]=\frac{\theta_0}{n^3}(1+3n\theta_0)+\frac{4C\theta_0}{n^3}+\frac{6C^2\theta_0}{n^3}+\frac{C^4}{n^4}.$$

We have that $\left|\frac{\mathrm{d}^3}{\mathrm{d}\theta^3}\log f(U|\theta)\Big|_{U=q_{c_2}(X)}\right|=\left|\frac{2\left(X+\frac{c_2}{n}\right)}{\theta^3}\right|$. To find the supremum in a neighbourhood of θ, we take $0<\varepsilon=\varepsilon(q_{c_1}(\theta_0))$ such that $\varepsilon<q_{c_1}(\theta_0)$, since if $\varepsilon>q_{c_1}(\theta_0)$ then $(q_{c_1}(\theta_0)-\varepsilon,q_{c_1}(\theta_0)+\varepsilon)\cap\overset{\circ}{\Theta}=(0,q_{c_1}(\theta_0)+\varepsilon)$ and then $\sup_{\theta:\left|\theta-q_{c_1}(\theta_0)\right|<\varepsilon}\left|\frac{2\left(X_i+\frac{c_2}{n}\right)}{\theta^3}\right|$ is not finite. Thus, for $0<\varepsilon<q_{c_1}(\theta_0)$,

$$\sup_{\theta:\left|\theta-q_{c_1}(\theta_0)\right|<\varepsilon}\left|\frac{2\left(X_i+\frac{c_2}{n}\right)}{\theta^3}\right|=\frac{2\left(X_i+\frac{c_2}{n}\right)}{\left(q_{c_1}(\theta_0)-\varepsilon\right)^3}=M\left(q_{c_2}(X_i)\right)$$

and

$$\begin{aligned}
&\left[\mathrm{E}\left(\left(\sum_{i=1}^{n} M\left(q_{c_2}(X_i)\right)\right)^2 \middle| \left|\hat{\theta}_n^*(\boldsymbol{X}) - q_{c_1}(\theta_0)\right| < \varepsilon\right)\right]^{\frac{1}{2}} \\
&= \left[\mathrm{E}\left(\left(\frac{2n\hat{\theta}_n^*(\boldsymbol{X})}{\left(q_{c_1}(\theta_0)-\varepsilon\right)^3}\right)^2 \middle| \left|\hat{\theta}_n^*(\boldsymbol{X}) - q_{c_1}(\theta_0)\right| < \varepsilon\right)\right]^{\frac{1}{2}} \\
&= \frac{2n}{\left(q_{c_1}(\theta_0)-\varepsilon\right)^3}\left[\mathrm{E}\left(\left(\hat{\theta}_n^*(\boldsymbol{X})\right)^2 \middle| \left|\hat{\theta}_n^*(\boldsymbol{X}) - q_{c_1}(\theta_0)\right| < \varepsilon\right)\right]^{\frac{1}{2}} \\
&= \frac{2n}{\left(q_{c_1}(\theta_0)-\varepsilon\right)^3}\left[\mathrm{E}\left(\left(\left|\hat{\theta}_n^*(\boldsymbol{X}) - q_{c_1}(\theta_0) + q_{c_1}(\theta_0)\right|\right)^2 \middle| \left|\hat{\theta}_n^*(\boldsymbol{X}) - q_{c_1}(\theta_0)\right| < \varepsilon\right)\right]^{\frac{1}{2}} \\
&\leq \frac{2n}{\left(q_{c_1}(\theta_0)-\varepsilon\right)^3}\left[\mathrm{E}\left(\left(\left|\hat{\theta}_n^*(\boldsymbol{X}) - q_{c_1}(\theta_0)\right| + q_{c_1}(\theta_0)\right)^2 \middle| \left|\hat{\theta}_n^*(\boldsymbol{X}) - q_{c_1}(\theta_0)\right| < \varepsilon\right)\right]^{\frac{1}{2}} \\
&< \frac{2n}{\left(q_{c_1}(\theta_0)-\varepsilon\right)^3}\left[\left(\varepsilon + q_{c_1}(\theta_0)\right)^2\right]^{\frac{1}{2}} = \frac{2n}{\left(q_{c_1}(\theta_0)-\varepsilon\right)^3}\left(\varepsilon + q_{c_1}(\theta_0)\right).
\end{aligned}$$

There is one last term to bound in order to apply the results of Theorem 3.2 to this specific example of the Poisson distribution. Simple steps yield

$$\begin{aligned}
&\sqrt{\mathrm{E}\left[\left(l''\left(q_{c_1}(\theta_0); q_{c_2}(\boldsymbol{X})\right) + ni\left(q_{c_1}(\theta_0)\right)\right)^2\right]} \\
&= \sqrt{\mathrm{E}\left[\left(-\frac{\sum_{i=1}^{n} q_{c_2}(X_i)}{\left(q_{c_1}(\theta_0)\right)^2} + \frac{n}{q_{c_1}(\theta_0)}\right)^2\right]} \\
&= \frac{n}{\left(q_{c_1}(\theta_0)\right)^2}\sqrt{\mathrm{E}\left[\left(\hat{\theta}_n^*(\boldsymbol{X}) - q_{c_1}(\theta_0)\right)^2\right]} = \frac{n}{\left(\theta_0 + \frac{c_1}{n}\right)^2}\sqrt{\left(\frac{\theta_0}{n} + \frac{C^2}{n^2}\right)}.
\end{aligned}$$

Using all these results to (3.40), gives

$$\begin{aligned}
&d_{bW}\left(\sqrt{n}(\hat{\theta}_n(\boldsymbol{X}) - \theta_0), K\right) < \frac{c_1}{\sqrt{n}} + \frac{c_2}{\sqrt{n}} + \frac{1}{\sqrt{n}}\left[2 + \frac{(3\theta_0+1)^{\frac{3}{4}}}{\theta_0^{\frac{3}{4}}}\right]\mathbb{1}\{\theta_0 > 0\} \\
&+ \frac{C}{\sqrt{n}} + \frac{2}{\varepsilon^2}\left(\frac{\theta_0}{n} + \frac{C^2}{n^2}\right) + \frac{\sqrt{n}}{\left(\theta_0 + \frac{c_1}{n}\right)}\left(\frac{\theta_0}{n} + \frac{C^2}{n^2}\right) \\
&+ \frac{\theta_0 + \frac{c_1}{n}}{2\sqrt{n}}\left(\frac{2n\left(\varepsilon + q_{c_1}(\theta_0)\right)}{\left(q_{c_1}(\theta_0)-\varepsilon\right)^3}\right)\left[\frac{\theta_0}{n^3}(1+3n\theta_0) + \frac{4C\theta_0}{n^3} + \frac{6C^2\theta_0}{n^3} + \frac{C^4}{n^4}\right]^{\frac{1}{2}}.
\end{aligned}$$

We have already explained that $\varepsilon \in (0, q_{c_1}(\theta_0))$. There is a trade-off on the choice of ε for the fifth and the seventh term of the general bound in (3.40). The fifth term indicates that as the value of ε gets nearer zero the quality of the bound is worse, while the seventh term requires ε to be as small as possible since $M\left(q_{c_2}(X_i)\right) = \frac{2\left(X_i+\frac{c_2}{n}\right)}{\left(q_{c_1}(\theta_0)-\varepsilon\right)^3}$. An optimisation process is quite tedious and thus we choose $\varepsilon = \frac{q_{c_1}(\theta_0)}{2} = \frac{\theta_0+\frac{c_1}{n}}{2}$. Giving this value to ε yields the result of the Corollary. □

Remark 3.6. ***(1)*** *The upper bound in* (3.52) *for the distributional distance between the actual distribution of the MLE and the normal distribution in the case of i.i.d. random variables following the Poisson(θ) distribution, with $\theta \in [0,\infty)$ is of order $\frac{1}{\sqrt{n}}$.*
(2) *We notice that the bound in* (3.52) *achieves its minimum as a function of C when $C = 0$ (equivalently $c_1 = c_2 = c > 0$). This yields*

$$d_{bW}\left(\sqrt{n}(\hat{\theta}_n(\boldsymbol{X}) - \theta_0), K\right) < \frac{2c}{\sqrt{n}} + \frac{1}{\sqrt{n}}\left[2 + \frac{(3\theta_0+1)^{\frac{3}{4}}}{\theta_0^{\frac{3}{4}}}\right] + \frac{8\theta_0}{n\left(\theta_0+\frac{c}{n}\right)^2} + \frac{\theta_0}{\sqrt{n}\left(\theta_0+\frac{c}{n}\right)} + \frac{12}{\sqrt{n}(\theta_0+\frac{c}{n})}\left[\frac{\theta_0}{n} + 3\theta_0^2\right]^{\frac{1}{2}}.$$

(3) *Since the MLE is unique and equal to $\hat{\theta}_n(\boldsymbol{X}) = \bar{X}$, Lemma 2.2 could be used directly for $\bar{X}$. Define $W = \sqrt{n}(\bar{X} - \theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i$, where $Y_i = X_i - \theta_0$ are independent, zero mean random variables. Also,* $\mathrm{E}(W) = 0$ *and* $\mathrm{Var}(W) = n\mathrm{Var}(\bar{X}) = \frac{1}{n}\sum_{i=1}^n \mathrm{Var}(X_i) = \theta_0$. *Therefore,* (2.12) *for $K \sim \mathrm{N}(0,\theta_0)$ and Hölder's inequality give for $\theta_0 > 0$*

$$d_{bW}\left(\sqrt{n}(\hat{\theta}_n(\boldsymbol{X}) - \theta_0), K\right) \leq \frac{1}{\sqrt{n}}\left(2 + \frac{1}{\theta_0^{\frac{3}{2}}}\left[\mathrm{E}(Y_1)^4\right]^{\frac{3}{4}}\right) = \frac{1}{\sqrt{n}}\left(2 + \frac{(3\theta_0+1)^{\frac{3}{4}}}{\theta_0^{\frac{3}{4}}}\right).$$

This bound, obtained by the direct application of Stein's method, is smaller than the bound given in Corollary 3.4. However the interest in the example treated in this section, where $\Theta = [0,\infty)$, is in adapting the approach to such cases where the MLE could be on the boundary of the parameter space and also it is not assumed that the MLE is a sum of random variables.

Empirical results

As in the exponential distribution example, the accuracy of the distributional upper bound given in (3.52) is studied using some simulated data. Similar steps as in Section 3.3 are followed. We produce 10,000 trials of n random independent observations x from the Poisson distribution with mean equal to 2. We then take the mean (which is the MLE) of the observations for each trial. However, these 10,000 values are not standardised now. We just subtract the value of the parameter and multiply by $\sqrt{n}$. Then, the function $h(x) = \frac{1}{x^2+2}, \forall x \in \mathbb{Z}_0^+$ is applied to these values and we take the sample mean, namely $\hat{\mathrm{E}}\left[h\left(\sqrt{n}\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right)\right)\right]$. Since $K \sim \mathrm{N}(0,2)$, basic calculations give $\tilde{\mathrm{E}}[h(K)] = 0.328$ which is the approximation of $\mathrm{E}[h(K)]$ up to three decimal places. With

$$Q_2(h,\theta_0) := \left|\hat{\mathrm{E}}\left[h\left(\sqrt{n}\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right)\right)\right] - \tilde{\mathrm{E}}[h(K)]\right|,$$

we compare $Q_2(h,\theta_0)$ with the bound in (3.52). Table 3.3 summarises the results for each sample size, n. In the table we denote by B_c the bound for $c_1 = c_2 = c$, such that $C = 0$ in (3.52). We present the results for $c \in \{0.1, 1, 2\}$. The error column in the table is the difference between $Q_2(h,\theta_0)$ and the value of the bound for $c = 1$, denoted by B_1.

Table 3.3: Simulation results from the Poisson(2) distribution

n	$Q_2(h,\theta_0)$	B_1	Error	$B_{0.1}$	B_2	Bound from Lemma 2.2
10	0.0008	1.2626	1.2618	1.1756	1.3642	0.4762
100	0.0006	0.3699	0.3693	0.3297	0.4147	0.1506
1000	0.0005	0.1130	0.1125	0.9999	0.1275	0.0476
10000	0.0004	0.0353	0.0349	0.0311	0.0399	0.0151
100000	0.0002	0.0111	0.0109	0.0098	0.0125	0.0048

Table 3.3 indicates that the quantity $Q_2(h,\theta_0)$, the upper bounds and the relevant error decrease as the sample size increases. In addition, the smaller the constant c is, the

smaller the bound. The error (for $c = 1$) is adequately small when the sample size is equal to 100. The upper bound obtained by applying directly Lemma 2.2 is, as expected, smaller than the one we get from the general perturbation method.

The accuracy of Stein's method in the case of i.i.d. random variables that follow the Poisson distribution is verified using some graphical tools. The behaviour of the MLE is shown in the density and the Q-Q plots presented in Figure 3.5 and 3.6, respectively. Both of these plots show that the distribution of the MLE is very close to the normal distribution. The bar plot in Figure 3.6 gives the magnitude of the error and its considerable decrease when the sample size gets greater than or equal to 100.

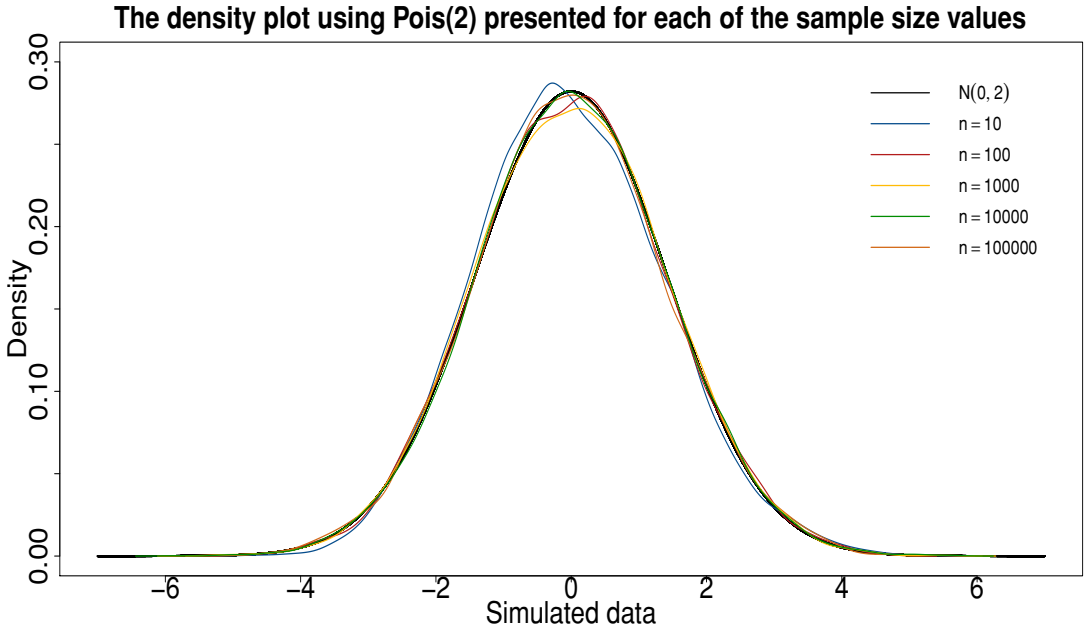

Fig. 3.5: The density function of $\sqrt{n}(\hat{\theta}_n(\boldsymbol{X}) - \theta_0)$ for various sample sizes. In this case $X_1, X_2, \cdots, X_n$ are i.i.d. random variables which follow the Poisson(2) distribution.

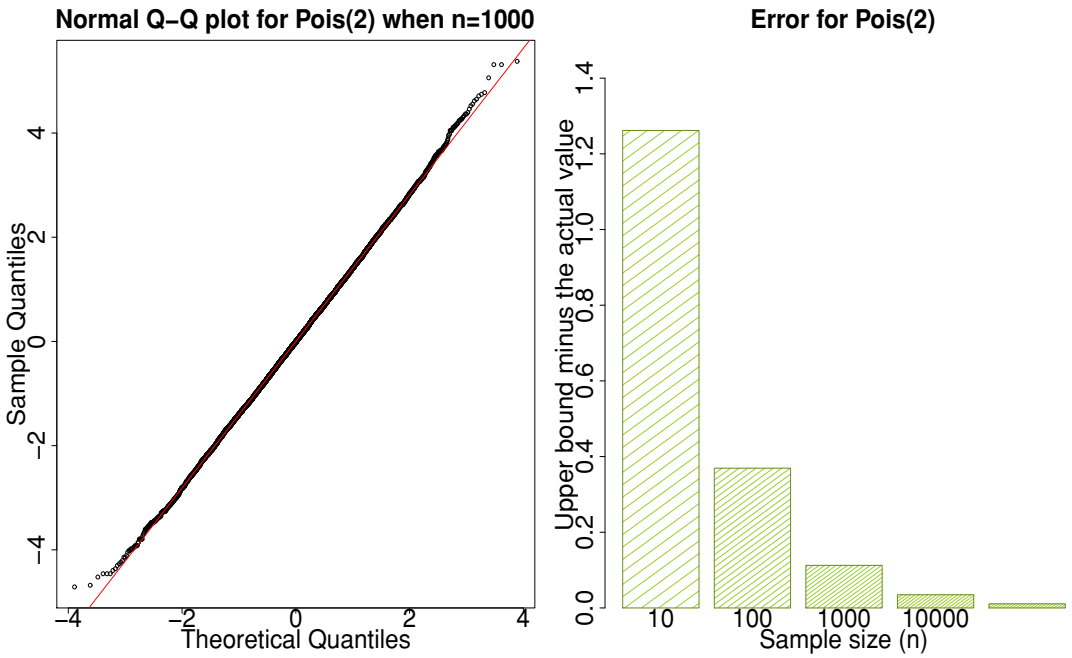

Fig. 3.6: The Q-Q plot, when $n = 1000$, and the bar plot of the error for different sample sizes for the Poisson distribution with mean equal to 2.

3.5 Bounds on the mean squared error

The previous sections considered models where the MLE exists, it is unique and can be found analytically as an explicit function of the random variables $X_1, X_2, \ldots, X_n$. The general upper bound in Section 3.2 has been proven without assuming or using an explicit form of the MLE. The bound in (3.7) includes terms, such as the mean squared error, $\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right)^2\right]$, of which the calculation requires to know the expression of the MLE. This fact generates problems in models where no closed-form solution to the maximization problem is known or available; in these cases, a numerical method, such as Newton-Raphson algorithm, can often be used to approximate the MLE. Hence a normal approximation is of interest in these cases too.

This section focuses on finding upper bounds for situations where an analytic form for the MLE is not available. In the proof for the final upper bound in Theorem 3.1, an

explicit form of the MLE was not used. However, if the MLE is not known, then the mean squared error (MSE) of the MLE, appearing in the bound for (3.1) should be bounded by a quantity which is independent of $\hat{\theta}_n(\boldsymbol{X})$.

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables. Apart from the regularity conditions, first defined in Section 2.2, we require the following further assumptions that make the steps and the calculations easier and ensure a meaningful upper bound:

(Fur.1) The support, S, is bounded;

(Fur.2) For all $\theta_0 \in \Theta$ there exists $\varepsilon_0 = \varepsilon_0(\theta_0) > 0$ such that $\left|\frac{\mathrm{d}^3}{\mathrm{d}\theta^3} \log f(x_1|\theta)\right| \leq M$ for all $\theta \in \Theta$ such that $|\theta - \theta_0| < \varepsilon_0$, where $M = M(\theta_0)$ is a constant that may depend on the unknown parameter θ_0;

(Fur.3) The sample size, n, satisfies

$$n > \frac{\|x\|^2 \left[M\varepsilon_0 + \sqrt{(M\varepsilon_0)^2 + 8[i(\theta_0)]^2}\right]^2}{4[i(\theta_0)]^3 \varepsilon_0^2}. \tag{3.53}$$

This ensures that $1 - 2\frac{\|x^2\|}{n\, i(\theta_0)\varepsilon_0^2} - \frac{\|x\|M}{\sqrt{n}[i(\theta_0)]^{\frac{3}{2}}} > 0$ for ε_0 as in (Fur.2).

Towards the aim of bounding the MSE by a quantity which is independent of the MLE, we will use the smooth test function $h(x) = x^2$, which under (Fur.1) is bounded, in (3.6) of Theorem 3.1 and rearrange the bound. For ease of presentation, let

$$D_1 = D_1(\theta_0, x, n) = 1 - 2\frac{\|x^2\|}{n\, i(\theta_0)\varepsilon_0^2} - \frac{\|x\|M}{\sqrt{n}[i(\theta_0)]^{\frac{3}{2}}}.$$

Theorem 3.3. *Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with probability density (or mass) function $f(x_i|\theta)$ with bounded support. Assume that the regularity conditions (R1) - (R4), as well as the assumptions (Fur.1) - (Fur.3) are satisfied. Also assume that the MLE exists and that it is unique. Then $A_1 = A_1(\theta_0, n)$ is an upper bound for*

$\sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]}$, *where*

$$A_1 = \frac{1}{D_1\sqrt{ni(\theta_0)}}\left\{\frac{\|x\|\sqrt{\mathrm{Var}\left[\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(X_1|\theta_0)\right]}}{\sqrt{n}i(\theta_0)}\right. \tag{3.54}$$
$$\left. + \left[\frac{\|x\|^2\mathrm{Var}\left[\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(X_1|\theta_0)\right]}{n[i(\theta_0)]^2} + D_1\left[1+\frac{2\|x\|}{\sqrt{n}}\left(2+\frac{\mathrm{E}\left|\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta_0)\right|^3}{[i(\theta_0)]^{\frac{3}{2}}}\right)\right]\right]^{\frac{1}{2}}\right\}.$$

Proof. Using the notation for the remainder terms, the triangle inequality, conditional expectations, Markov's inequality and Stein's method, as in Section 3.2, gives for $\varepsilon = \varepsilon(\theta_0) = \varepsilon_0$

$$\begin{aligned}
&\left|\mathrm{E}\left[h\left(\sqrt{n\,i(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)\right)\right]-\mathrm{E}[h(Z)]\right| \leq \frac{\|h'\|}{\sqrt{n}}\left(2+\frac{\left[\mathrm{E}\left|\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta_0)\right|^3\right]}{[i(\theta_0)]^{\frac{3}{2}}}\right)\\
&+\frac{\|h'\|}{\sqrt{i(\theta_0)}}\sqrt{\mathrm{Var}\left(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(X_1|\theta_0)\right)}\sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]}\\
&+\frac{2\|h\|}{\varepsilon^2}\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]\\
&+\frac{\|h'\|}{2\sqrt{ni(\theta_0)}}\mathrm{E}\left(\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\sup_{\theta:|\theta-\theta_0|<\varepsilon}\left|l^{(3)}(\theta;\boldsymbol{X})\right|\middle||\hat{\theta}_n(\boldsymbol{X})-\theta_0|<\varepsilon\right).
\end{aligned}$$

Using now (Fur.2)

$$\begin{aligned}
&\left|\mathrm{E}\left[h\left(\sqrt{n\,i(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)\right)\right]-\mathrm{E}[h(Z)]\right| \leq \frac{\|h'\|}{\sqrt{n}}\left(2+\frac{\mathrm{E}\left|\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta_0)\right|^3}{[i(\theta_0)]^{\frac{3}{2}}}\right)\\
&+\frac{\|h'\|}{\sqrt{i(\theta_0)}}\sqrt{\mathrm{Var}\left(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(X_1|\theta_0)\right)}\sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]}\\
&+\frac{\|h'\|\sqrt{n}M}{2\sqrt{i(\theta_0)}}\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]+\frac{2\|h\|}{\varepsilon^2}\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right].
\end{aligned} \tag{3.55}$$

Straightforward calculations and denoting by B_{x^2} the upper bound for (3.1) when $h(x) = x^2$, lead to

$$\begin{aligned} \mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right)^2\right] &= \frac{1}{n\,i(\theta_0)}\left|\mathrm{E}\left[\sqrt{n\,i(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right)\right]^2 - \mathrm{E}\left(Z^2\right) + \mathrm{E}\left(Z^2\right)\right| \\ &\leq \frac{1}{n\,i(\theta_0)}(B_{x^2} + 1), \end{aligned} \tag{3.56}$$

where

$$\begin{aligned} B_{x^2} \leq & 2\frac{\|x\|}{\sqrt{n}}\left(2 + \frac{\mathrm{E}\left|\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta_0)\right|^3}{[i(\theta_0)]^{\frac{3}{2}}}\right) + \left[\frac{2\|x^2\|}{\varepsilon^2} + \frac{\|x\|\sqrt{n}M}{\sqrt{i(\theta_0)}}\right]\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right)^2\right] \\ & + 2\frac{\|x\|}{\sqrt{i(\theta_0)}}\sqrt{\mathrm{Var}\left(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(X_1|\theta_0)\right)}\sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right)^2\right]}. \end{aligned} \tag{3.57}$$

Combining the results in (3.56) and (3.57) yields

$$\begin{aligned} &\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right)^2\right] \leq 2\frac{\|x\|}{n^{\frac{3}{2}}i(\theta_0)}\left(2 + \frac{\mathrm{E}\left|\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta_0)\right|^3}{[i(\theta_0)]^{\frac{3}{2}}}\right) \\ &+ \frac{1}{n\,i(\theta_0)}\left[\frac{2\|x^2\|}{\varepsilon^2} + \frac{\|x\|\sqrt{n}M}{\sqrt{i(\theta_0)}}\right]\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right)^2\right] \\ &+ 2\frac{\|x\|}{n[i(\theta_0)]^{\frac{3}{2}}}\sqrt{\mathrm{Var}\left(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(X_1|\theta_0)\right)}\sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right)^2\right]} + \frac{1}{n\,i(\theta_0)}. \end{aligned} \tag{3.58}$$

The next step is to solve the simple quadratic inequality in (3.58), with unknown $A = A(\theta_0) = \sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right)^2\right]}$. We get

$$\begin{aligned} A^2 \leq \frac{1}{n\,i(\theta_0)}\Bigg(1 + \frac{2\|x\|}{\sqrt{n}}\left(2 + \frac{\mathrm{E}\left|\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta_0)\right|^3}{[i(\theta_0)]^{\frac{3}{2}}}\right) + \left[\frac{2\|x^2\|}{\varepsilon^2} + \frac{\|x\|\sqrt{n}M}{\sqrt{i(\theta_0)}}\right]A^2 \\ + \frac{2\|x\|\sqrt{\mathrm{Var}\left(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(X_1|\theta_0)\right)}}{\sqrt{i(\theta_0)}}A\Bigg), \end{aligned}$$

which yields

$$D_1 A^2 - 2\frac{\|x\|\sqrt{\mathrm{Var}\left(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(X_1|\theta_0)\right)}}{n[i(\theta_0)]^{\frac{3}{2}}}A - \frac{1}{n\,i(\theta_0)}\left[1+2\frac{\|x\|}{\sqrt{n}}\left(2+\frac{\mathrm{E}\left|\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta_0)\right|^3}{[i(\theta_0)]^{\frac{3}{2}}}\right)\right] \leq 0. \tag{3.59}$$

Thus, the two solutions of the above quadratic inequality are given by

$$A_{1,2} = \frac{1}{D_1\sqrt{ni(\theta_0)}}\left\{\sqrt{\mathrm{Var}\left(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(X_1|\theta_0)\right)}\frac{\|x\|}{\sqrt{n}i(\theta_0)} \pm \left[\frac{\|x\|^2\mathrm{Var}\left(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(X_1|\theta_0)\right)}{n[i(\theta_0)]^2} + D_1\left(1+\frac{2\|x\|}{\sqrt{n}}\left(2+\frac{\mathrm{E}\left|\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta_0)\right|^3}{[i(\theta_0)]^{\frac{3}{2}}}\right)\right)\right]^{\frac{1}{2}}\right\}. \tag{3.60}$$

Since $D_1 > 0$, the root with the plus sign in (3.60) is positive, while the other is negative. Therefore the inequality in (3.59) is satisfied for $A = \sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]} \leq A_1$, with A_1 as in (3.54). □

Remark 3.7. ***(1)*** *Using Theorem 3.3, the final upper bound for* (3.1) *which is useful when no analytic expression of the MLE is available, becomes*

$$d_{bW}\left(\sqrt{n\,i(\theta_0)}(\hat{\theta}_n(\boldsymbol{X})-\theta_0), Z\right) \leq \frac{1}{\sqrt{n}}\left(2+\frac{\mathrm{E}\left|\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta_0)\right|^3}{[i(\theta_0)]^{\frac{3}{2}}}\right) + \left[\frac{2}{\varepsilon^2}+\frac{\sqrt{n}M}{2\sqrt{i(\theta_0)}}\right](A_1)^2 + \frac{\sqrt{\mathrm{Var}\left(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(X_1|\theta_0)\right)}}{\sqrt{i(\theta_0)}}A_1. \tag{3.61}$$

(2) *The order of A_1 in terms of the sample size is $\frac{1}{\sqrt{n}}$. Hence, the order of the final upper bound in* (3.61) *is also $\frac{1}{\sqrt{n}}$.*

(3) *The result of the Theorem holds when (Fur.3) is satisfied, meaning that we have a lower bound on the sample size n, as can be seen from* (3.53)*. This lower bound can be*

quite large, and thus Theorem 3.3 is mainly of theoretical interest.

(4) *The result holds only when the support of the distribution is bounded; see (Fur.1). An approach that holds when the support of the distribution is not necessarily bounded, but assumes boundedness of the parameter space, is explained in Theorem 6.5, where the random variables are not necessarily independent.*

Example: The Beta distribution

Consider the example of i.i.d random variables from the Beta distribution with one of the two shape parameters being unknown. In this case, the MLE can only be expressed in terms of the inverse of the digamma function, $\Psi(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta}\log\Gamma(\theta)$. The general result in Theorem 3.3 is used, in order to obtain an upper bound for the MSE and use it to get an upper bound for (3.1). The following corollary gives the result.

Corollary 3.5. *Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables from the Beta(θ_0, β) distribution, where β is known and θ_0 is the unknown parameter. Let*

$$B_1 = B_1(\theta_0) = 8\left(\Psi_3(\theta_0) + \Psi_3(\theta_0+\beta) + 3[\Psi_1(\theta_0)]^2 + 3[\Psi_1(\theta_0+\beta)]^2\right),$$

where $\Psi_j(\theta), j \in \mathbb{N}$ is the j^{th} derivative of the digamma function, $\Psi(\theta)$. Also, let

$$B_2 = B_2(\theta_0) = \frac{96\beta}{\theta_0^4} + \frac{\beta\pi^4}{15}, \qquad D_{\Psi 1} = D_{\Psi 1}(\theta_0, \beta) = \Psi_1(\theta_0) - \Psi_1(\theta_0+\beta)$$

and

$$B_3 = B_3(\theta_0, n) = \frac{\left[\left(1 + \frac{2}{\sqrt{n}}\left(2 + \frac{(B_1)^{\frac{3}{4}}}{(D_{\Psi 1})^{\frac{3}{2}}}\right)\right)\left(1 - \frac{8}{n\theta_0^2 D_{\Psi 1}} - \frac{B_2}{\sqrt{n}(D_{\Psi 1})^{\frac{3}{2}}}\right)\right]^{\frac{1}{2}}}{\sqrt{D_{\Psi 1}} - \frac{8}{n\theta_0^2\sqrt{D_{\Psi 1}}} - \frac{B_2}{\sqrt{n}D_{\Psi 1}}}. \quad (3.62)$$

Let

$$n > \frac{\left[B_2\frac{\theta_0}{2} + \sqrt{\frac{(B_2\theta_0)^2}{4} + 8[\Psi_1(\theta_0) - \Psi_1(\theta_0+\beta)]^2}\right]^2}{[\Psi_1(\theta_0) - \Psi_1(\theta_0+\beta)]^3\theta_0^2}.$$

Then for $Z \sim N(0,1)$

$$d_{bW}\left(\sqrt{n\,i(\theta_0)}(\hat{\theta}_n(\boldsymbol{X})-\theta_0), Z\right) \leq \frac{1}{\sqrt{n}}\left(2+\frac{(B_1)^{\frac{3}{4}}}{[\Psi_1(\theta_0)-\Psi_1(\theta_0+\beta)]^{\frac{3}{2}}}\right) + \frac{8}{n\theta_0^2}(B_3)^2 + \frac{B_2(B_3)^2}{2\sqrt{n}[\Psi_1(\theta_0)-\Psi_1(\theta_0+\beta)]^{\frac{1}{2}}}. \tag{3.63}$$

Proof. The probability density function is

$$f(x|\theta) = \frac{\Gamma(\theta+\beta)}{\Gamma(\theta)\Gamma(\beta)}x^{\theta-1}(1-x)^{\beta-1}, \tag{3.64}$$

with $\theta > 0$ and $x \in [0,1]$. Hence

$$\begin{aligned} l(\theta;\boldsymbol{x}) =& n[\log(\Gamma(\theta+\beta)) - \log(\Gamma(\theta)) - \log(\Gamma(\beta))] \\ &+ (\theta-1)\sum_{i=1}^{n}\log x_i + (\beta-1)\sum_{i=1}^{n}\log(1-x_i) \end{aligned} \tag{3.65}$$

and

$$l'(\theta;\boldsymbol{x}) = n[\Psi(\theta+\beta)-\Psi(\theta)] + \sum_{i=1}^{n}\log x_i$$

$$l^{(j)}(\theta;\boldsymbol{x}) = n(\Psi_{j-1}(\theta+\beta)-\Psi_{j-1}(\theta)),\ j \in \mathbb{N}\setminus\{1\}.$$

Now we show that the conditions (R1)-(R4) and the assumptions (Fur.1)-(Fur.3) are satisfied. For (R1) it is obvious. As for (R2), the three times differentiability of the density function can be verified from (3.65). In addition, using (3.64) and the expressions for the logarithmic expectations of a Beta distributed random variable, it is easy to verify $\int_0^1 \frac{\mathrm{d}^j}{\mathrm{d}\theta^j} f(x|\theta)\,\mathrm{d}x = \frac{\mathrm{d}^j}{\mathrm{d}\theta^j}\int_0^1 f(x|\theta)\,\mathrm{d}x = 0, j \in \{1,2,3\}$ for (R2). Now for (R3), a first order Taylor expansion yields

$$\left|\frac{\mathrm{d}^3}{\mathrm{d}\theta^3}\log f(x|\theta)\right| = |\Psi_2(\theta+\beta)-\Psi_2(\theta)| = \beta\,|\Psi_3(\theta^*)|\,, \tag{3.66}$$

for θ^* between θ and $\theta+\beta$. Using that

$$\Psi_m(z) = (-1)^{m+1} m! \sum_{k=0}^{\infty} \frac{1}{(z+k)^{m+1}}, \text{ for } z \in \mathbb{C} \setminus \{\mathbb{Z}_0^-\} \text{ and m} > 0, \tag{3.67}$$

gives

$$\Psi_3(z) = 6 \sum_{k=0}^{\infty} \frac{1}{(z+k)^4}, \tag{3.68}$$

with $\Psi_3(z)$ being a decreasing function of z. We need to upper bound (3.66) in a neighbourhood of θ in order for (R3) to hold. Let $\varepsilon = \varepsilon(\theta_0) = \varepsilon_0$ be a positive constant such that $\theta \in (\theta_0 - \varepsilon, \theta_0 + \varepsilon) \cap \Theta$ as in (R3). If $\varepsilon > \theta_0$ then $(\theta_0 - \varepsilon, \theta_0 + \varepsilon) \cap \Theta = (0, \theta_0 + \varepsilon)$ and the result in (3.66) can not be upper bounded in $(0, \theta_0 + \varepsilon)$ as $\Psi_3(\theta)$ is not bounded for $\theta \in (0, \theta_0 + \varepsilon)$. Therefore, $0 < \varepsilon < \theta_0$ and using that $\sum_{i=1}^{\infty} \frac{1}{i^4} = \frac{\pi^4}{90}$ gives that for θ^* between θ and $\theta+\beta$

$$\begin{aligned}
&\sup_{\theta:|\theta-\theta_0|<\varepsilon} \left| \frac{\mathrm{d}^3}{\mathrm{d}\theta^3} \log f(x|\theta) \right| = \sup_{\theta:|\theta-\theta_0|<\varepsilon} |\Psi_2(\theta+\beta) - \Psi_2(\theta)| \\
&= \beta \sup_{\theta:|\theta-\theta_0|<\varepsilon} |\Psi_3(\theta^*)| \leq \beta \sup_{\theta:|\theta-\theta_0|<\varepsilon} |\Psi_3(\theta)|, \text{ as } \Psi_3(.) \text{ is a decreasing function} \\
&\leq \beta |\Psi_3(\theta_0 - \varepsilon)| = 6\beta \sum_{k=0}^{\infty} \frac{1}{(\theta_0 - \varepsilon + k)^4} \\
&\leq 6\beta \left[\frac{1}{(\theta_0 - \varepsilon)^4} + \sum_{k=1}^{\infty} \frac{1}{k^4} \right] \\
&= \frac{6\beta}{(\theta_0 - \varepsilon)^4} + \frac{\beta \pi^4}{15} = M =: M(\varepsilon),
\end{aligned} \tag{3.69}$$

with $M < \infty$. Hence, (R3) holds as well. Also, $i(\theta_0) = \Psi_1(\theta_0) - \Psi_1(\theta_0 + \beta)$ which is positive since it is obvious from (3.67) that $\Psi_1(z)$ is a decreasing function. The assumption (Fur.1) holds with $\|x\| \leq 1$. The assumption (Fur.2) is satisfied from (3.69). Now, since $i(\theta_0) = \Psi_1(\theta_0) - \Psi_1(\theta_0 + \beta)$ take

$$n > \frac{\left[M\varepsilon + \sqrt{(M\varepsilon)^2 + 8[\Psi_1(\theta_0) - \Psi_1(\theta_0 + \beta)]^2} \right]^2}{4\varepsilon^2 [\Psi_1(\theta_0) - \Psi_1(\theta_0 + \beta)]^3}$$

in order for (Fur.3) to be satisfied. To find A_1 as in (3.54), firstly, as $\mathrm{E}\left|\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta_0)\right|^3$ is not straightforward to evaluate due to the absolute value in the expectation, it is easily seen that using Hölder's inequality $\mathrm{E}\left|\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta_0)\right|^3 \leq \left[\mathrm{E}\left[\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta_0)\right)^4\right]\right]^{\frac{3}{4}}$ we find an upper bound for

$$\begin{aligned}\mathrm{E}\left[\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta_0)\right)^4\right] &= \mathrm{E}\left[(\log X_1 + \Psi(\theta_0+\beta) - \Psi(\theta_0))^4\right] \\ &= \mathrm{E}\left[(\log X_1 - \mathrm{E}(\log X_1))^4\right].\end{aligned}$$

If $G_1 \sim \Gamma(\theta_0,\lambda)$ and $G_2 \sim \Gamma(\beta,\lambda)$ are independent, then $\frac{G_1}{G_1+G_2} \sim \mathrm{Beta}(\theta_0,\beta)$. Thus, with $X_1 = \frac{G_1}{G_1+G_2}$

$$\begin{aligned}&\mathrm{E}\left[\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta_0)\right)^4\right] \\ &= \mathrm{E}\left[((\log G_1 - \mathrm{E}[\log G_1]) + (\mathrm{E}(\log(G_1+G_2)) - \log(G_1+G_2)))^4\right] \\ &\leq 8\left\{\mathrm{E}\left[(\log G_1 - \mathrm{E}(\log G_1))^4\right] + \mathrm{E}\left[(\log(G_1+G_2) - \mathrm{E}(\log(G_1+G_2)))^4\right]\right\}.\end{aligned} \tag{3.70}$$

Now the fourth central moment of the logarithm of a Gamma distributed random variable is calculated using that $\int_0^\infty \frac{z^{\alpha-1}\mathrm{e}^{-z}(\log z)^k}{\Gamma(\alpha)}\,\mathrm{d}z = \frac{\Gamma^{(k)}(\alpha)}{\Gamma(\alpha)}$, for any $\alpha > 0$ and $k \in \mathbb{N}$. For $Y \sim \Gamma(\alpha,\lambda)$ and letting $z = \lambda y$

$$\mathrm{E}(\log Y) = \Psi(\alpha) - \log\lambda.$$

Using again $z = \lambda y$,

$$\begin{aligned}\mathrm{E}\left[(\log Y - \mathrm{E}(\log Y))^4\right] &= \int_0^\infty \frac{z^{\alpha-1}\mathrm{e}^{-z}}{\Gamma(\alpha)}\left(\log\left(\frac{z}{\lambda}\right) - \mathrm{E}\left(\log\left(\frac{Z}{\lambda}\right)\right)\right)^4 \mathrm{d}z \\ &= \int_0^\infty \frac{z^{\alpha-1}\mathrm{e}^{-z}}{\Gamma(\alpha)}\left(\log z - \mathrm{E}\left(\log Z\right)\right)^4 \mathrm{d}z\end{aligned}$$

$$= \frac{1}{\Gamma(\alpha)} \sum_{k=0}^{4} \binom{4}{k} (-1)^k [\Psi(\alpha)]^{4-k} \int_0^\infty z^{\alpha-1} \mathrm{e}^{-z} (\log z)^k \, \mathrm{d}z$$

$$= -3[\Psi(\alpha)]^4 + 6[\Psi(\alpha)]^2 \frac{\Gamma''(\alpha)}{\Gamma(\alpha)} - 4\Psi(\alpha) \frac{\Gamma^{(3)}(\alpha)}{\Gamma(\alpha)} + \frac{\Gamma^{(4)}(\alpha)}{\Gamma(\alpha)}.$$

At this point, the digamma function can be used in order to simplify the expression above. Following simple steps it can be easily verified that

$$\frac{\Gamma''(\alpha)}{\Gamma(\alpha)} = \Psi_1(\alpha) + [\Psi(\alpha)]^2, \quad \frac{\Gamma^{(3)}(\alpha)}{\Gamma(\alpha)} = \Psi_2(\alpha) + 3\Psi(\alpha)\Psi_1(\alpha) + [\Psi(\alpha)]^3,$$
$$\frac{\Gamma^{(4)}(\alpha)}{\Gamma(\alpha)} = \Psi_3(\alpha) + 4\Psi_2(\alpha)\Psi(\alpha) + 6\Psi_1(\alpha)[\Psi(\alpha)]^2 + 3[\Psi_1(\alpha)]^2 + [\Psi(\alpha)]^4.$$

Hence for $Y \sim \Gamma(\alpha, \lambda)$

$$\mathrm{E}\left[(\log Y - \mathrm{E}(\log Y))^4 \right] = \Psi_3(\alpha) + 3[\Psi_1(\alpha)]^2 \tag{3.71}$$

and therefore, from (3.70),

$$\mathrm{E}\left[\left(\frac{\mathrm{d}}{\mathrm{d}\theta} \log f(X_1|\theta_0) \right)^4 \right]$$
$$\leq 8 \left(\Psi_3(\theta_0) + \Psi_3(\theta_0 + \beta) + 3\left[\Psi_1(\theta_0)\right]^2 + 3\left[\Psi_1(\theta_0 + \beta)\right]^2 \right) = B_1.$$

With $M = M(\varepsilon)$ as in (3.69), taking $\varepsilon = \frac{\theta_0}{2}$, we conclude that

$$\sup_{\theta : |\theta - \theta_0| < \varepsilon} \left| \frac{\mathrm{d}^3}{\mathrm{d}\theta^3} \log f(x_1|\theta) \right| \leq \frac{96\beta}{\theta_0^4} + \frac{\beta\pi^4}{15} = B_2.$$

Using (3.65), gives

$$\mathrm{Var}\left(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} \log f(X_1|\theta_0) \right) = \mathrm{Var}(\Psi_1(\theta_0 + \beta) - \Psi_1(\theta_0)) = 0.$$

Having found all the necessary quantities, the upper bound in (3.54) is calculated and multiplied by $\sqrt{n}$. This is equal to B_3 shown in (3.62), which is an upper bound for

$\sqrt{n\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]}$ in the specific case of i.i.d. random variables from the Beta distribution. Using this bound in (3.55) gives the result in (3.63). □

Now, we study the accuracy of our bound for the MSE of the MLE by simulations. For the simulations, $\theta_0 = 1.5$, $\beta = 1$ and in this case of β being equal to 1, the MLE is analytically known with $\hat{\theta}_n(\boldsymbol{X}) = -\frac{n}{\sum_{i=1}^{n}\log X_i}$. We find that $n \geq 7398$, in order for (Fur.3) to be satisfied. The process to simulate is quite simple. Let $n \in \{7398, 7399, \ldots, 8397\}$ and for each n, start by generating 10,000 trials of n random independent observations, x, from the Beta distribution with parameter values as above. The MLE, $\hat{\theta}_n(\boldsymbol{X})$, is evaluated in each trial, which in turn gives a vector of 10,000 values. Thus, for each n from 7398 to 8397, we evaluate the sample MSE, $\hat{\mathrm{E}}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right] = \frac{1}{10{,}000}\sum_{i=1}^{10{,}000}\left(\hat{\theta}_n(\boldsymbol{x})_i-\theta_0\right)^2$ and compare it with its upper bound, $\left(\frac{B_3}{\sqrt{n}}\right)^2$, where B_3 is given in (3.62). The difference between their values measures the error of our bound on the MSE. Some simulation results are shown in Table 3.4.

Table 3.4: Simulation results from the Beta$(1.5, 1)$ distribution

n	$\hat{\mathrm{E}}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]$	Upper bound	Error
7500	0.0002	0.0999	0.0997
7700	0.0002	0.0333	0.0331
7900	0.0002	0.0197	0.0195
8100	0.0002	0.0139	0.0137
8300	0.0002	0.0107	0.0105

The table indicates that the bound and the error decrease as the sample size increases, as expected, since the order of the upper bound for the MSE is $\frac{1}{n}$. In addition, the smaller the sample size, the larger the bound is. The bounds are considerably larger than the estimated MSE and they are not numerically sharp. In addition, because of the relatively strong requirement that $n \geq 7398$, these bounds on the MSE are mainly of theoretical interest.

Chapter 4

Multi-parameter distributions

This chapter extends the work presented in Chapter 3 by focusing on the multi-parameter case in the presence of independent but not necessarily identically distributed (i.n.i.d.) and possibly high-dimensional random vectors. The interest is on assessing the asymptotic normality of the vector MLE. The approach we follow is based on Stein's method under a multivariate setting. Let

$$H = \left\{ h : \mathbb{R}^d \to \mathbb{R} : h \text{ is three times differentiable with bounded derivatives} \right\} \quad (4.1)$$

be the class of test functions we use in this chapter. We abbreviate $\|h\|_1 := \sup_i \|\frac{\partial}{\partial x_i} h\|$, $\|h\|_2 := \sup_{i,j} \|\frac{\partial^2}{\partial x_i \partial x_j} h\|$ and $\|h\|_3 := \sup_{i,j,k} \|\frac{\partial^3}{\partial x_i \partial x_j \partial x_k} h\|$. For upper bounds for the distributional distance of interest we employ Lemma 4.1 below, which is based on the notion of exchangeability. For two random vectors $\boldsymbol{W}$ and $\boldsymbol{W}'$ in $\mathbb{R}^d$, $d \in \mathbb{Z}^+$, we say that $(\boldsymbol{W}, \boldsymbol{W}')$ is an exchangeable pair if and only if $(\boldsymbol{W}, \boldsymbol{W}') \overset{\mathrm{d}}{=\!=} (\boldsymbol{W}', \boldsymbol{W})$, where $\overset{\mathrm{d}}{=\!=}$ denotes equality in distribution.

Lemma 4.1. *(Reinert and Röllin, 2009). Assume that $(\boldsymbol{W}, \boldsymbol{W}')$ is an exchangeable pair of $\mathbb{R}^d-$valued random vectors such that $\mathrm{E}(\boldsymbol{W}) = \boldsymbol{0}$, $\mathrm{E}(\boldsymbol{W}\boldsymbol{W}^\intercal) = \Sigma$, with $\Sigma \in \mathbb{R}^{d \times d}$*

symmetric and positive definite. Suppose further that

$$\mathrm{E}\left(\boldsymbol{W}' - \boldsymbol{W} | \boldsymbol{W}\right) = -\Lambda \boldsymbol{W} + \boldsymbol{R} \tag{4.2}$$

for an invertible $d \times d$ matrix Λ and a $\sigma(\boldsymbol{W})$-measurable random vector $\boldsymbol{R}$. Then if $\boldsymbol{Z}$ has the d-dimensional standard normal distribution, we have for every $h \in H$ as in (4.1)*,*

$$\left|\mathrm{E}[h(\boldsymbol{W})] - \mathrm{E}\left[h\left(\Sigma^{\frac{1}{2}}\boldsymbol{Z}\right)\right]\right| \leq \frac{\|h\|_2}{4}A + \frac{\|h\|_3}{12}B + \left(\|h\|_1 + \frac{1}{2}d\|\Sigma\|^{\frac{1}{2}}\|h\|_2\right)C,$$

where, with $\lambda^{(i)} := \sum_{m=1}^{d}\left|\left(\Lambda^{-1}\right)_{m,i}\right|$,

$$\begin{aligned}
A &= \sum_{i=1}^{d}\sum_{j=1}^{d}\lambda^{(i)}\sqrt{\mathrm{Var}\left(\mathrm{E}\left(\left(W_i' - W_i\right)\left(W_j' - W_j\right)\middle|\boldsymbol{W}\right)\right)}, \\
B &= \sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{k=1}^{d}\lambda^{(i)}\mathrm{E}\left|\left(W_i' - W_i\right)\left(W_j' - W_j\right)\left(W_k' - W_k\right)\right|, \\
C &= \sum_{i=1}^{d}\lambda^{(i)}\sqrt{\mathrm{Var}(R_i)}.
\end{aligned}$$

Reinert and Röllin (2009) were not the first to use Stein's original approach of exchangeable pairs for multivariate normal approximation. This goes back to Chatterjee and Meckes (2008) where three abstract normal approximation theorems using exchangeable pairs are presented that frequently produce total variation distance bounds of the optimal order in the univariate case. Chatterjee and Meckes (2008) also explain that because the differential equation (2.8) in the univariate case is first order rather than second, it allows for reducing the degree of smoothness of the test function needed by one, over what is required in the multivariate case. This lead to obtaining bounds on the Kolmogorov distance, as explained in Section 3.1. The multivariate Stein equation in (2.16) is a second order differential equation and an improvement on bounds for the total variation distance in the multivariate case is not available. Although Chatterjee

and Meckes (2008) give in general sharper bounds, the results of Reinert and Röllin (2009) are more widely applicable; see Meckes (2009) for a comparison of these two different approaches. In addition, bounds for the so-called Fisher information distance for sums of i.i.d. random vectors are given in Nourdin et al. (2014).

The chapter is organised as follows. Section 4.1 treats the case of i.n.i.d. random vectors. After presenting the main result for the asymptotic normality of the MLE under sufficient regularity conditions, we give an upper bound on the distributional distance between the distribution of the vector MLE and the multivariate normal distribution. Special attention is given to linear regression models with an application to the simplest case of the straight-line model. In Section 4.2, under weaker regularity conditions, we present the upper bound for the case of i.i.d. random vectors. Specific results for independent random variables that follow the normal distribution with unknown mean and variance are also given. Subsection 4.2.2 contains an upper bound on the distributional distance of interest, which holds even in cases where no analytic expression of the vector MLE is available. We illustrate the results through the Beta distribution with both shape parameters unknown. The results of this chapter have been submitted to the *Journal of Multivariate Analysis* as a single-author paper.

4.1 Non-identically distributed random vectors

In this section we examine the case of i.n.i.d. t-dimensional random vectors, for $t \in \mathbb{N}$. Apart from the assumptions set in Propositions 2.1 and 2.2 for the existence and uniqueness of the MLE, we use some regularity conditions, first stated in Hoadley (1971), in order to establish the asymptotic normality of the MLE.

4.1.1 A general bound

For $\boldsymbol{X_1}, \boldsymbol{X_2}, \ldots, \boldsymbol{X_n}$ being i.n.i.d. random vectors, we denote by $f_i(\boldsymbol{x}, \boldsymbol{\theta})$ the probability density (or mass) function for $\boldsymbol{X_i}$. Thus, the likelihood function is $L(\boldsymbol{\theta}, \boldsymbol{x}) = \prod_{i=1}^{n} f_i(\boldsymbol{x_i}|\boldsymbol{\theta})$, with its logarithm still being denoted by $l(\boldsymbol{\theta}; \boldsymbol{x})$. Assuming that the parameter space is an open subset of $\mathbb{R}^d$, Proposition 2.1 is sufficient for the existence and uniqueness of $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})$. In addition, we work under the following regularity conditions for the asymptotic normality of the MLE to hold, (Hoadley, 1971):

(N1) $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) \xrightarrow{\text{p}} \boldsymbol{\theta_0}$, as $n \to \infty$, where $\boldsymbol{\theta_0}$ is the true parameter value;

(N2) the gradient vector $\nabla(\log(f_k(\boldsymbol{X_k}|\boldsymbol{\theta}))) \in \mathbb{R}^{d\times 1}$ and the Hessian matrix $J_k(\boldsymbol{X_k}, \boldsymbol{\theta}) = \left\{ \frac{\partial^2}{\partial\theta_i \partial\theta_j} \log(f_k(\boldsymbol{X_k}|\boldsymbol{\theta})) \right\}_{i,j=1,2,\ldots,d} \in \mathbb{R}^{d\times d}$ exist almost surely $\forall k \in \{1, 2, \ldots, n\}$ with respect to the probability measure $\mathbb{P}$, where $\mathbb{P}[.]$ denotes probability with respect to $\boldsymbol{\theta_0}$;

(N3) $J_k(\boldsymbol{X_k}, \boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$, uniformly in k, almost surely with respect to $\mathbb{P}$ and is a measurable function of $\boldsymbol{X_k}$;

(N4) $\mathrm{E}[\nabla(\log(f_k(\boldsymbol{X_k}|\boldsymbol{\theta})))] = \boldsymbol{0}$, $k = 1, 2, \ldots, n$;

(N5) with $[\nabla(\log(f_k(\boldsymbol{X_k}|\boldsymbol{\theta})))]^{\intercal} \in \mathbb{R}^{1\times d}$ denoting the transpose of $\nabla(\log(f_k(\boldsymbol{X_k}|\boldsymbol{\theta})))$,

$$I_k(\boldsymbol{\theta}) = \mathrm{E}\left[[\nabla(\log(f_k(\boldsymbol{X_k}|\boldsymbol{\theta})))]\,[\nabla(\log(f_k(\boldsymbol{X_k}|\boldsymbol{\theta})))]^{\intercal}\right] = -\mathrm{E}\left[J_k(\boldsymbol{X_k}, \boldsymbol{\theta})\right];$$

(N6) for

$$\bar{I}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{j=1}^{n} I_j(\boldsymbol{\theta}), \tag{4.3}$$

there exists a matrix $\bar{I}(\boldsymbol{\theta}) \in \mathbb{R}^{d\times d}$ such that $\bar{I}_n(\boldsymbol{\theta}) \xrightarrow[n\to\infty]{} \bar{I}(\boldsymbol{\theta})$. In addition, $\bar{I}_n(\boldsymbol{\theta}), \bar{I}(\boldsymbol{\theta})$ are symmetric matrices for all $\boldsymbol{\theta}$ and $\bar{I}(\boldsymbol{\theta})$ is positive definite;

(N7) for some $\delta > 0$, $\frac{\sum_k \mathrm{E}|\boldsymbol{\lambda}^{\intercal}\nabla(\log(f_k(\boldsymbol{X_k}|\boldsymbol{\theta_0})))|^{2+\delta}}{n^{\frac{2+\delta}{2}}} \xrightarrow[n\to\infty]{} 0$ for all $\boldsymbol{\lambda} \in \mathbb{R}^d$;

(N8) with $\|.\|$ the ordinary Euclidean norm on $\mathbb{R}^d$, then for $k,i,j \in \{1,2,\ldots,d\}$ there exist $\varepsilon > 0$, $K > 0$, $\delta > 0$ and random variables $B_{k,ij}(\boldsymbol{X_k})$ such that

(i) $\sup\left\{\left|\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log(f_k(\boldsymbol{X_k}|\boldsymbol{t}))\right| : \|\boldsymbol{t}-\boldsymbol{\theta_0}\| \leq \varepsilon\right\} \leq B_{k,ij}(\boldsymbol{X_k})$;

(ii) $\mathrm{E}\left|B_{k,ij}(\boldsymbol{X_k})\right|^{1+\delta} \leq K$.

Assuming that $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})$ exists and is unique, the following theorem gives the result for the asymptotic normality of the MLE in the i.n.i.d. case, in a slightly different way than the result in Hoadley (1971).

Theorem 4.1. *Let $\boldsymbol{X_1}, \boldsymbol{X_2}, \ldots, \boldsymbol{X_n}$ be independent random vectors with probability density (or mass) functions $f_i(\boldsymbol{x_i}|\boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$. Assume that the MLE exists and it is unique and that the regularity conditions (N1)-(N8) hold. Also let $\boldsymbol{Z} \sim \mathrm{N}_d(\boldsymbol{0}, I_{d\times d})$, where $\boldsymbol{0}$ is the $d \times 1$ zero vector and $I_{d\times d}$ is the $d \times d$ identity matrix. Then,*

$$\left[\sum_{k=1}^{n} I_k(\boldsymbol{\theta_0})\right]^{\frac{1}{2}} (\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) - \boldsymbol{\theta_0}) \xrightarrow[n\to\infty]{\mathrm{d}} \boldsymbol{Z}. \tag{4.4}$$

Proof. Hoadley (1971) proves in Theorem 2, p.1983 that under the regularity conditions (N1)-(N8)

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) - \boldsymbol{\theta_0}\right) \xrightarrow[n\to\infty]{\mathrm{d}} [\bar{I}(\boldsymbol{\theta_0})]^{-\frac{1}{2}} \boldsymbol{Z}.$$

Using this result and (N6) we obtain that

$$\left[\frac{1}{n}\sum_{k=1}^{n} I_k(\boldsymbol{\theta_0})\right]^{\frac{1}{2}} \sqrt{n}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) - \boldsymbol{\theta_0}\right) \xrightarrow[n\to\infty]{\mathrm{d}} [\bar{I}(\boldsymbol{\theta_0})]^{\frac{1}{2}} [\bar{I}(\boldsymbol{\theta_0})]^{-\frac{1}{2}} \boldsymbol{Z} = \boldsymbol{Z},$$

which is the result of the theorem. □

Our motivation remains to assess the quality of the normal approximation in (4.4) through explicit upper bounds on the distributional distance of interest. From now on, unless otherwise stated, $\bar{I}_n(\boldsymbol{\theta})$ is as in (4.3). Let the subscript (m) denote an index for

which $\left|\hat{\theta}_n(\boldsymbol{x})_{(m)} - \theta_{0_{(m)}}\right|$ is the largest among the d components;

$$(m) \in \{1,2,\ldots,d\} \text{ is such that}$$
$$\left|\hat{\theta}_n(\boldsymbol{x})_{(m)} - \theta_{0_{(m)}}\right| \geq \left|\hat{\theta}_n(\boldsymbol{x})_j - \theta_{0_j}\right|, \ \forall j \in \{1,2,\ldots,d\}$$

and also, for ease of presentation, let

$$C_{(m)} = C_{(m)}(\boldsymbol{X}, \boldsymbol{\theta_0}) := \hat{\theta}_n(\boldsymbol{X})_{(m)} - \theta_{0_{(m)}}. \tag{4.5}$$

Our main result is as follows.

Theorem 4.2. *Let $\boldsymbol{X_1}, \boldsymbol{X_2}, \ldots, \boldsymbol{X_n}$ be i.n.i.d. $\mathbb{R}^t$-valued, $t \in \mathbb{Z}^+$, random vectors with probability density (or mass) function $f_i(\boldsymbol{x_i}|\boldsymbol{\theta})$, for which the parameter space Θ is an open subset of $\mathbb{R}^d$. Asssume that (N1)-(N8) are satisfied and that the MLE exists and it is unique. In addition, assume that for any $\boldsymbol{\theta_0} \in \Theta$ there exists $0 < \varepsilon = \varepsilon(\boldsymbol{\theta_0})$ and functions $M_{kjl}(\boldsymbol{x})$, $\forall k,j,l \in \{1,2,\ldots,d\}$ such that $\left|\frac{\partial^3}{\partial\theta_k\partial\theta_j\partial\theta_l} l(\boldsymbol{\theta},\boldsymbol{x})\right| \leq M_{kjl}(\boldsymbol{x})$ for all $\boldsymbol{\theta} \in \Theta$ such that $|\theta_j - \theta_{0_j}| < \varepsilon \ \forall j \in \{1,2,\ldots,d\}$. Also, for $C_{(m)}$ as in* (4.5)*, assume that $\mathrm{E}\left(\left(M_{kjv}(\boldsymbol{X})\right)^2 \middle| \left|C_{(m)}\right| < \varepsilon\right) < \infty$. Let $\left\{\boldsymbol{X}'_i, i = 1,2,\ldots,n\right\}$ be an independent copy of $\{\boldsymbol{X_i}, i = 1,2,\ldots,n\}$. For $h \in H$, with H as in* (4.1) *and $\boldsymbol{Z} \sim \mathrm{N}_d(\boldsymbol{0}, I_{d\times d})$ it holds that*

$$\begin{aligned}
&\left|\mathrm{E}\left[h\left(\sqrt{n}[\bar{I}_n(\boldsymbol{\theta_0})]^{\frac{1}{2}}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) - \boldsymbol{\theta_0}\right)\right)\right] - \mathrm{E}[h(\boldsymbol{Z})]\right| \\
&\leq \frac{\|h\|_1}{\sqrt{n}} K_1(\boldsymbol{\theta_0}) + \frac{\|h\|_2}{\sqrt{n}} K_2(\boldsymbol{\theta_0}) + \frac{\|h\|_3}{\sqrt{n}} K_3(\boldsymbol{\theta_0}) + \frac{2\|h\|}{\varepsilon^2}\mathrm{E}\left(\sum_{j=1}^{d}\left(\hat{\theta}_n(\boldsymbol{X})_j - \theta_{0_j}\right)^2\right),
\end{aligned} \tag{4.6}$$

where

$$\begin{aligned}
K_1(\boldsymbol{\theta_0}) = \sum_{k=1}^{d}\sum_{l=1}^{d}&\left|\left[[\bar{I}_n(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{lk}\right| \\
&\times \sum_{j=1}^{d}\left[\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})_j - \theta_{0_j}\right)^2\right]\mathrm{E}\left[\left(\frac{\partial^2}{\partial\theta_j\partial\theta_k} l(\boldsymbol{\theta_0};\boldsymbol{X}) + n[\bar{I}_n(\boldsymbol{\theta_0})]_{kj}\right)^2\right]\right]^{\frac{1}{2}}
\end{aligned}$$

$$+\frac{1}{2}\left\{\sum_{k=1}^{d}\sum_{l=1}^{d}\left|\left[\left[\bar{I}_n(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}}\right]_{lk}\right|\sum_{j=1}^{d}\sum_{v=1}^{d}\left[\mathrm{E}\left(\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\left(\hat{\theta}_n(\boldsymbol{X})_v-\theta_{0_v}\right)^2\right)\right]^{\frac{1}{2}}\right.$$

$$\left.\times\left[\mathrm{E}\left(\left(M_{kjv}(\boldsymbol{X})\right)^2\middle|\left|C_{(m)}\right|<\varepsilon\right)\right]^{\frac{1}{2}}\right\}, \tag{4.7}$$

$$K_2(\boldsymbol{\theta_0})=\frac{1}{4\sqrt{n}}\sum_{j=1}^{d}\left[\sum_{i=1}^{n}\mathrm{Var}\left(\left(\sum_{k=1}^{d}\left[\bar{I}_n(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}}_{jk}\frac{\partial}{\partial\theta_k}\log(f_i(\boldsymbol{X_i}|\boldsymbol{\theta_0}))\right)^2\right)\right]^{\frac{1}{2}}$$

$$+\frac{1}{2\sqrt{n}}\sum_{k=1}^{d-1}\sum_{j>k}^{d}\left[\sum_{i=1}^{n}\mathrm{Var}\left(\sum_{q=1}^{d}\sum_{v=1}^{d}\left[\bar{I}_n(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}}_{jq}\left[\bar{I}_n(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}}_{kv}\right.\right.$$

$$\left.\left.\times\frac{\partial}{\partial\theta_q}\log(f_i(\boldsymbol{X_i}|\boldsymbol{\theta_0}))\frac{\partial}{\partial\theta_v}\log(f_i(\boldsymbol{X_i}|\boldsymbol{\theta_0}))\right)\right]^{\frac{1}{2}} \tag{4.8}$$

and

$$K_3(\boldsymbol{\theta_0})=\frac{1}{12n}\sum_{i=1}^{n}\mathrm{E}\left(\sum_{m=1}^{d}\left|\sum_{l=1}^{d}\left[\bar{I}_n(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}}_{ml}\right.\right.$$

$$\left.\left.\times\left(\frac{\partial}{\partial\theta_l}\left\{\log(f_i(\boldsymbol{X_i'}|\boldsymbol{\theta_0}))-\log(f_i(\boldsymbol{X_i}|\boldsymbol{\theta_0}))\right\}\right)\right|\right)^3. \tag{4.9}$$

The following lemma can be seen as a multivariate extension of Chebyshev's other inequality in Lemma 3.1 and it is useful for bounding conditional expectations, which sometimes can be difficult to derive.

Lemma 4.2. *Let $\boldsymbol{M}\in\mathbb{R}^d$ be a random vector with $M_i>0\ \forall i=1,2,\ldots,d$ and $\varepsilon>0$. For every continuous function $f:\mathbb{R}^d\to\mathbb{R}$ such that $f(\boldsymbol{m})$ is increasing and $f(\boldsymbol{m})\geq 0$, for $m_i>0\ \forall i\in\{1,2,\ldots,d\}$, where $\boldsymbol{m}=(m_1,m_2,\ldots,m_d)$,*

$$\mathrm{E}[f(\boldsymbol{M})|M_i<\varepsilon\ \forall i=1,2,\ldots,d]\leq\mathrm{E}[f(\boldsymbol{M})].$$

Proof of **Lemma 4.2**. Let $k \in \{1,2,\ldots,d\}$. We set $M_{d+1} = 0$. It will be shown that for $k = 1,2,\ldots,d$ we have that

$$\mathrm{E}[f(\boldsymbol{M})|M_i < \varepsilon \ , \ i = k,\ldots,d] \leq \mathrm{E}[f(\boldsymbol{M})|M_i < \varepsilon \ , \ i = k+1,\ldots,d].$$

From the law of total expectation,

$$\begin{aligned}
&\mathrm{E}[f(\boldsymbol{M})|M_i < \varepsilon \ , \ i = k+1,\ldots,d] \\
&= \mathrm{E}[f(\boldsymbol{M})|M_i < \varepsilon \ , \ i = k,\ldots,d]\mathbb{P}\left[M_k < \varepsilon | M_i < \varepsilon \ , \ i = k+1,\ldots,d\right] \\
&+ \mathrm{E}[f(\boldsymbol{M})|M_i < \varepsilon \ , \ i = k+1,\ldots,d, \ M_k \geq \varepsilon]\mathbb{P}\left[M_k \geq \varepsilon | M_i < \varepsilon \ , \ i = k+1,\ldots,d\right].
\end{aligned}$$

Using that

$$\mathbb{P}\left[M_k < \varepsilon | M_i < \varepsilon \ , \ i = k+1,\ldots,d\right] = 1 - \mathbb{P}\left[M_k \geq \varepsilon | M_i < \varepsilon \ , \ i = k+1,\ldots,d\right]$$

yields

$$\begin{aligned}
&\mathrm{E}[f(\boldsymbol{M})|M_i < \varepsilon \ , \ i = k+1,\ldots,d] \\
&= \mathrm{E}[f(\boldsymbol{M})|M_i < \varepsilon \ , \ i = k,\ldots,d] \\
&+ \mathbb{P}\left[M_k \geq \varepsilon | M_i < \varepsilon \ , \ i = k+1,\ldots,d\right] \Bigg\{ \mathrm{E}[f(\boldsymbol{M})|M_i < \varepsilon \ , \ i = k+1,\ldots,d, \ M_k \geq \varepsilon] \\
&\qquad\qquad - \mathrm{E}[f(\boldsymbol{M})|M_i < \varepsilon \ , \ i = k,\ldots,d] \Bigg\}.
\end{aligned} \tag{4.10}$$

Since $f(\boldsymbol{m})$ is an increasing function,

$$\mathrm{E}[f(\boldsymbol{M})|M_i < \varepsilon \ , \ i = k+1,\ldots,d, \ M_k \geq \varepsilon] - \mathrm{E}[f(\boldsymbol{M})|M_i < \varepsilon \ , \ i = k,\ldots,d] \geq 0. \tag{4.11}$$

Applying this to (4.10) gives that

$$\mathrm{E}[f(\boldsymbol{M})|M_i < \varepsilon\ ,\ i = k,\ldots,d] \leq \mathrm{E}[f(\boldsymbol{M})|M_i < \varepsilon\ ,\ i = k+1,\ldots,d].$$

A simple iteration over k gives that

$$\mathrm{E}[f(\boldsymbol{M})|M_i < \varepsilon\ \forall i = 1,2,\ldots,d] \leq \mathrm{E}[f(\boldsymbol{M})], \tag{4.12}$$

which is the result of the lemma. □

Proof of **Theorem 4.2**. The regularity conditions and the definition of the MLE give that

$$\frac{\partial}{\partial\theta_k} l\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{x});\boldsymbol{x}\right) = 0\ \forall k \in \{1,2,\ldots,d\}.$$

Using a second-order Taylor expansion of $\frac{\partial}{\partial\theta_k} l(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{x});\boldsymbol{x})$ about $\boldsymbol{\theta_0}$,

$$\begin{aligned} 0 =& \frac{\partial}{\partial\theta_k} l(\boldsymbol{\theta_0};\boldsymbol{x}) + \sum_{j=1}^{d} (\hat{\theta}_n(\boldsymbol{x})_j - \theta_{0_j}) \left(\frac{\partial^2}{\partial\theta_k\partial\theta_j} l(\boldsymbol{\theta_0};\boldsymbol{x})\right) \\ &+ \frac{1}{2}\sum_{j=1}^{d}\sum_{q=1}^{d} (\hat{\theta}_n(\boldsymbol{x})_j - \theta_{0_j})(\hat{\theta}_n(\boldsymbol{x})_q - \theta_{0_q}) \left(\frac{\partial^3}{\partial\theta_k\partial\theta_j\partial\theta_q} l(\boldsymbol{\theta};\boldsymbol{x})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0^*}\right), \end{aligned}$$

where $\boldsymbol{\theta}_0^*$ is between $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{x})$ and $\boldsymbol{\theta_0}$. Rearranging,

$$\begin{aligned} &\sum_{j=1}^{d} (\hat{\theta}_n(\boldsymbol{x})_j - \theta_{0_j}) \left(\frac{\partial^2}{\partial\theta_k\partial\theta_j} l(\boldsymbol{\theta_0};\boldsymbol{x})\right) \\ &= -\frac{\partial}{\partial\boldsymbol{\theta}_\kappa} l(\boldsymbol{\theta_0};\boldsymbol{x}) - \frac{1}{2}\sum_{j=1}^{d}\sum_{q=1}^{d} (\hat{\theta}_n(\boldsymbol{x})_j - \theta_{0_j})(\hat{\theta}_n(\boldsymbol{x})_q - \theta_{0_q}) \left(\frac{\partial^3}{\partial\theta_k\partial\theta_j\partial\theta_q} l(\boldsymbol{\theta};\boldsymbol{x})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0^*}\right), \end{aligned}$$

which gives that for all $k = 1,2,\ldots,d$,

$$\begin{aligned} &\sum_{j=1}^{d} (\hat{\theta}_n(\boldsymbol{x})_j - \theta_{0_j}) \left(-n[\bar{I}_n(\boldsymbol{\theta_0})]_{kj} + \frac{\partial^2}{\partial\theta_k\partial\theta_j} l(\boldsymbol{\theta_0};\boldsymbol{x}) + n[\bar{I}_n(\boldsymbol{\theta_0})]_{kj}\right) \\ &= -\frac{\partial}{\partial\theta_k} l(\boldsymbol{\theta_0};\boldsymbol{x}) - \frac{1}{2}\sum_{j=1}^{d}\sum_{q=1}^{d} (\hat{\theta}_n(\boldsymbol{x})_j - \theta_{0_j})(\hat{\theta}_n(\boldsymbol{x})_q - \theta_{0_q}) \left(\frac{\partial^3}{\partial\theta_k\partial\theta_j\partial\theta_q} l(\boldsymbol{\theta};\boldsymbol{x})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0^*}\right) \end{aligned}$$

so that

$$\sum_{j=1}^{d} n[\bar{I}_n(\boldsymbol{\theta_0})]_{kj}(\hat{\theta}_n(\boldsymbol{x})_j - \theta_{0_j}) = \frac{\partial}{\partial\theta_k} l(\boldsymbol{\theta_0};\boldsymbol{x})$$
$$+ \sum_{j=1}^{d} (\hat{\theta}_n(\boldsymbol{x})_j - \theta_{0_j}) \left(\frac{\partial^2}{\partial\theta_k \partial\theta_j} l(\boldsymbol{\theta_0};\boldsymbol{x}) + n[\bar{I}_n(\boldsymbol{\theta_0})]_{kj} \right)$$
$$+ \frac{1}{2} \sum_{j=1}^{d} \sum_{q=1}^{d} (\hat{\theta}_n(\boldsymbol{x})_j - \theta_{0_j})(\hat{\theta}_n(\boldsymbol{x})_q - \theta_{0_q}) \left(\frac{\partial^3}{\partial\theta_k \partial\theta_j \partial\theta_q} l(\boldsymbol{\theta};\boldsymbol{x})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0^*} \right). \tag{4.13}$$

Using (4.13), which holds $\forall k \in \{1,2,\ldots,d\}$,

$$\sqrt{n}[\bar{I}_n(\boldsymbol{\theta_0})]^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{x}) - \boldsymbol{\theta_0})$$
$$= \frac{1}{\sqrt{n}}[\bar{I}_n(\boldsymbol{\theta_0})]^{-\frac{1}{2}} \left\{ \nabla(l(\boldsymbol{\theta_0};\boldsymbol{x})) \right.$$
$$\left. + \sum_{j=1}^{d} (\hat{\theta}_n(\boldsymbol{x})_j - \theta_{0_j}) \left(\nabla \left(\frac{\partial}{\partial\theta_j} l(\boldsymbol{\theta_0};\boldsymbol{x}) \right) + n[\bar{I}_n(\boldsymbol{\theta_0})]_{[j]} \right) \right\}$$
$$+ \frac{1}{2\sqrt{n}}[\bar{I}_n(\boldsymbol{\theta_0})]^{-\frac{1}{2}} \left\{ \sum_{j=1}^{d} \sum_{q=1}^{d} (\hat{\theta}_n(\boldsymbol{x})_j - \theta_{0_j})(\hat{\theta}_n(\boldsymbol{x})_q - \theta_{0_q}) \right.$$
$$\left. \times \left(\nabla \left(\frac{\partial^2}{\partial\theta_j \partial\theta_q} l(\boldsymbol{\theta};\boldsymbol{x})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0^*} \right) \right) \right\},$$

where $[\bar{I}_n(\boldsymbol{\theta_0})]_{[j]}$ is the j^{th} column of the matrix $\bar{I}_n(\boldsymbol{\theta_0})$. The triangle inequality gives

$$\left| \mathrm{E}\left[h\left(\sqrt{n}\,[\bar{I}_n(\boldsymbol{\theta_0})]^{\frac{1}{2}} \left(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) - \boldsymbol{\theta_0} \right) \right) \right] - \mathrm{E}[h(\boldsymbol{Z})] \right|$$
$$\leq \left| \mathrm{E}\left[h\left(\frac{1}{\sqrt{n}}\,[\bar{I}_n(\boldsymbol{\theta_0})]^{-\frac{1}{2}} \nabla(l(\boldsymbol{\theta_0};\boldsymbol{X})) \right) \right] - \mathrm{E}[h(\boldsymbol{Z})] \right| \tag{4.14}$$
$$+ \left| \mathrm{E}\left[h\left(\sqrt{n}\,[\bar{I}_n(\boldsymbol{\theta_0})]^{\frac{1}{2}} \left(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) - \boldsymbol{\theta_0} \right) \right) - h\left(\frac{1}{\sqrt{n}}\,[\bar{I}_n(\boldsymbol{\theta_0})]^{-\frac{1}{2}} \nabla(l(\boldsymbol{\theta_0};\boldsymbol{X})) \right) \right] \right|. \tag{4.15}$$

Step 1: Upper bound for (4.14). First of all, $\nabla(l(\boldsymbol{\theta_0};\boldsymbol{x})) = \sum_{i=1}^{n} \nabla\left(\log(f_i(\boldsymbol{x_i}|\boldsymbol{\theta_0}))\right)$ due to independence. The results of Lemma 4.1 will be used for

$$\boldsymbol{W} = \frac{1}{\sqrt{n}} \left[\bar{I}_n(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}} \sum_{i=1}^{n} \nabla(\log(f_i(\boldsymbol{X_i}|\boldsymbol{\theta_0}))) = (W_1, W_2, \ldots, W_d)^{\intercal} \in \mathbb{R}^{d\times 1}. \quad (4.16)$$

From (4.16) we have that for all $k \in \{1,2,\ldots,d\}$, $W_k = \sum_{i=1}^{n} \xi_{ik}$, with

$$\xi_{ik} = \frac{1}{\sqrt{n}} \sum_{j=1}^{d} \left[\left[\bar{I}_n(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}}\right]_{kj} \frac{\partial}{\partial \theta_j} \log(f_i(\boldsymbol{X_i}|\boldsymbol{\theta_0})). \quad (4.17)$$

From the regularity conditions, $\mathrm{E}\left(\nabla(l(\boldsymbol{\theta_0};\boldsymbol{X}))\right) = \boldsymbol{0}$ and thus $\mathrm{E}(\boldsymbol{W}) = \boldsymbol{0}$. Also, $\bar{I}_n(\boldsymbol{\theta_0})$ is symmetric. Therefore, $\left[\bar{I}_n(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}}$ is also symmetric. From the regularity conditions $\sum_{i=1}^{n} \mathrm{Cov}\left[\nabla\left(\log\left(f_i(\boldsymbol{X_i}|\boldsymbol{\theta_0})\right)\right)\right] = n\bar{I}_n(\boldsymbol{\theta_0})$ and basic calculations show that $\mathrm{Cov}(\boldsymbol{W}) = I_{d\times d}$. Since $\mathrm{E}(\boldsymbol{W}) = \boldsymbol{0}$ and $\mathrm{E}(\boldsymbol{W}\boldsymbol{W}^{\intercal}) = I_{d\times d}$, the first assumption of Lemma 4.1 is satisfied.

The next step is to define $\boldsymbol{W}'$ such that $(\boldsymbol{W}, \boldsymbol{W}')$ is an exchangeable pair satisfying (4.2). Let $\left\{\boldsymbol{X_i'}, i = 1,2,\ldots,n\right\}$ be an independent copy of $\{\boldsymbol{X_i}, i = 1,2,\ldots,n\}$ and let the index $I \in \{1,2,\ldots,n\}$ follow the uniform distribution on $\{1,2,\ldots,n\}$, independently of $\left\{\boldsymbol{X_i}, \boldsymbol{X_i'}, i = 1,2,\ldots,n\right\}$. Let

$$\xi_{ik}' = \frac{1}{\sqrt{n}} \sum_{j=1}^{d} \left[\left[\bar{I}_n(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}}\right]_{kj} \frac{\partial}{\partial \theta_j} \log(f_i(\boldsymbol{X_i'}|\boldsymbol{\theta_0}))$$

and

$$W_k' = W_k - \xi_{Ik} + \xi_{Ik}', \ \forall k \in \{1,2,\ldots,d\},$$

with $\mathrm{E}\left(W_k' - W_k|\boldsymbol{W}\right) = \mathrm{E}\left(\xi_{Ik}' - \xi_{Ik}|\boldsymbol{W}\right) = -\mathrm{E}\left(\xi_{Ik}|\boldsymbol{W}\right) = -\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left(\xi_{ik}|\boldsymbol{W}\right) = -\frac{W_k}{n}$. Hence (4.2) is satisfied with $\Lambda = \frac{1}{n} I_{d\times d}$ and $\boldsymbol{R} = \boldsymbol{0}$. We conclude that the assumptions

of Lemma 4.1 are satisfied and Lemma 4.1 gives that

$$|\mathrm{E}[h(\boldsymbol{W})]-\mathrm{E}[h(\boldsymbol{Z})]| \leq n\left(\frac{\|h\|_2}{4}\sum_{i=1}^{d}\sum_{j=1}^{d}\left[\mathrm{Var}\left(\mathrm{E}\left[\left(W_i'-W_i\right)\left(W_j'-W_j\right)|\boldsymbol{W}\right]\right)\right]^{\frac{1}{2}}\right) \tag{4.18}$$

$$+n\left(\frac{\|h\|_3}{12}\sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{k=1}^{d}\mathrm{E}\left|\left(W_i'-W_i\right)\left(W_j'-W_j\right)\left(W_k'-W_k\right)\right|\right). \tag{4.19}$$

To bound the variance of the conditional expectations in (4.18), let $\mathscr{A} = \sigma(\boldsymbol{X_1}, \boldsymbol{X_2}, \ldots, \boldsymbol{X_n})$. Since $\sigma(\boldsymbol{W}) \subset \mathscr{A}$, for any random variable Y, $\mathrm{Var}(\mathrm{E}[Y|\boldsymbol{W}]) \leq \mathrm{Var}(\mathrm{E}[Y|\mathscr{A}])$. Simple steps yield

$$(4.18) \leq n\frac{\|h\|_2}{4}\left\{\sum_{j=1}^{d}\left[\mathrm{Var}\left(\mathrm{E}\left[(\xi_{Ij}'-\xi_{Ij})^2|\mathscr{A}\right]\right)\right]^{\frac{1}{2}}\right.$$

$$\left.+2\sum_{j>k}^{d}\sum_{k=1}^{d-1}\left[\mathrm{Var}\left(\mathrm{E}\left[\left(\xi_{Ik}'-\xi_{Ik}\right)\left(\xi_{Ij}'-\xi_{Ij}\right)|\mathscr{A}\right]\right)\right]^{\frac{1}{2}}\right\}. \tag{4.20}$$

After straightforward calculations and using that $\left\{\boldsymbol{X_i'}, i=1,2,\ldots,n\right\}$ is an independent copy of $\{\boldsymbol{X_i}, i=1,2,\ldots,n\}$ and ξ_{ik}' is independent of $\mathscr{A}$ we get that

$$(4.20) = n\frac{\|h\|_2}{4}\left\{\sum_{j=1}^{d}\left[\mathrm{Var}\left(\mathrm{E}\left[(\xi_{Ij}')^2\right]-2\mathrm{E}[\xi_{Ij}']\mathrm{E}[\xi_{Ij}|\mathscr{A}]+\mathrm{E}\left[\xi_{Ij}^2|\mathscr{A}\right]\right)\right]^{\frac{1}{2}}\right.$$

$$\left.+2\sum_{k=1}^{d-1}\sum_{j>k}^{d}\left[\mathrm{Var}\left(\mathrm{E}\left[\xi_{Ik}'\xi_{Ij}'\right]-\mathrm{E}\left[\xi_{Ij}'\right]\mathrm{E}\left[\xi_{Ik}|\mathscr{A}\right]-\mathrm{E}\left[\xi_{Ik}'\right]\mathrm{E}\left[\xi_{Ij}|\mathscr{A}\right]+\mathrm{E}\left[\xi_{Ik}\xi_{Ij}|\mathscr{A}\right]\right)\right]^{\frac{1}{2}}\right\}. \tag{4.21}$$

Using that $\mathrm{E}\left[\xi_{ik}'\right]=0$,

$$
\begin{aligned}
(4.21) &= n\frac{\|h\|_2}{4}\left\{\sum_{j=1}^{d}\left[\mathrm{Var}\left(\mathrm{E}\left[(\xi'_{Ij})^2\right]+\mathrm{E}\left[\xi^2_{Ij}|\mathscr{A}\right]\right)\right]^{\frac{1}{2}}\right. \\
&\qquad \left. +2\sum_{k=1}^{d-1}\sum_{j>k}^{d}\left[\mathrm{Var}\left(\mathrm{E}\left[\xi'_{Ik}\xi'_{Ij}\right]+\mathrm{E}\left[\xi_{Ik}\xi_{Ij}|\mathscr{A}\right]\right)\right]^{\frac{1}{2}}\right\} \\
&= n\frac{\|h\|_2}{4}\left\{\sum_{j=1}^{d}\left[\frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n}\mathrm{E}\left[\xi^2_{ij}|\mathscr{A}\right]\right)\right]^{\frac{1}{2}}+2\sum_{j>k}^{d}\sum_{k=1}^{d-1}\left[\frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n}\mathrm{E}\left[\xi_{ik}\xi_{ij}|\mathscr{A}\right]\right)\right]^{\frac{1}{2}}\right\} \\
&= \frac{\|h\|_2}{4}\left\{\sum_{j=1}^{d}\left[\mathrm{Var}\left(\sum_{i=1}^{n}\xi^2_{ij}\right)\right]^{\frac{1}{2}}+2\sum_{j>k}^{d}\sum_{k=1}^{d-1}\left[\mathrm{Var}\left(\sum_{i=1}^{n}\xi_{ik}\xi_{ij}\right)\right]^{\frac{1}{2}}\right\}=\frac{\|h\|_2}{\sqrt{n}}K_2(\boldsymbol{\theta_0}),
\end{aligned}
$$

with $K_2(\boldsymbol{\theta_0})$ in (4.8). For (4.19), using (4.17), after basic calculations we obtain that

$$
(4.19) \leq \frac{\|h\|_3}{\sqrt{n}}K_3(\boldsymbol{\theta_0}),
$$

with $K_3(\boldsymbol{\theta_0})$ as in (4.9). Thus,

$$
(4.14) \leq \frac{\|h\|_2}{\sqrt{n}}K_2(\boldsymbol{\theta_0})+\frac{\|h\|_3}{\sqrt{n}}K_3(\boldsymbol{\theta_0}). \tag{4.22}
$$

Step 2: Upper bound for (4.15). We will use multivariate Taylor expansions, conditional expectations and Markov's inequality. The approach is similar to the one in the single-parameter case of Chapter 3. For ease of presentation, let

$$
\begin{aligned}
\boldsymbol{R_1}(\boldsymbol{\theta_0};\boldsymbol{x}) &= \frac{1}{2\sqrt{n}}[\bar{I}_n(\boldsymbol{\theta_0})]^{-\frac{1}{2}} \\
&\qquad \times\sum_{j=1}^{d}\sum_{q=1}^{d}(\hat{\theta}_n(\boldsymbol{x})_j-\theta_{0_j})(\hat{\theta}_n(\boldsymbol{x})_q-\theta_{0_q})\left(\nabla\left(\frac{\partial^2}{\partial\theta_j\partial\theta_q}l(\boldsymbol{\theta};\boldsymbol{x})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*_{\mathbf{0}}}\right)\right) \\
T_1 = T_1(\boldsymbol{\theta_0};\boldsymbol{X},h) &:= h\left(\sqrt{n}\,[\bar{I}_n(\boldsymbol{\theta_0})]^{\frac{1}{2}}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})-\boldsymbol{\theta_0}\right)\right) \\
&\qquad -h\left(\frac{1}{\sqrt{n}}\left[\bar{I}_n(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}}\left(\nabla\left(l(\boldsymbol{\theta_0};\boldsymbol{X})\right)\right)\right) \\
T_2 = T_2(\boldsymbol{\theta_0};\boldsymbol{X},h) &:= h\left(\frac{1}{\sqrt{n}}[\bar{I}_n(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\left(\nabla(l(\boldsymbol{\theta_0};\boldsymbol{x}))\right)+\boldsymbol{R_1}(\boldsymbol{\theta_0};\boldsymbol{x})\right) \\
&\qquad -h\left(\frac{1}{\sqrt{n}}\left[\bar{I}_n(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}}\left(\nabla\left(l(\boldsymbol{\theta_0};\boldsymbol{X})\right)\right)\right).
\end{aligned} \tag{4.23}
$$

Using the above notation,

$$
\begin{aligned}
(4.15) &= |\mathrm{E}[T_1]| \\
&= \left|\mathrm{E}\left[h\left(\sqrt{n}\,[\bar{I}_n(\boldsymbol{\theta_0})]^{\frac{1}{2}}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})-\boldsymbol{\theta_0}\right)\right)\right. \\
&\qquad \left. -h\left(\frac{1}{\sqrt{n}}[\bar{I}_n(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\left(\nabla(l(\boldsymbol{\theta_0};\boldsymbol{x}))\right)+\boldsymbol{R_1}(\boldsymbol{\theta_0};\boldsymbol{X})\right)+T_2\right]\right|.
\end{aligned}
$$

Now, the triangle inequality gives that

$$
\begin{aligned}
(4.15) \leq \mathrm{E}&\left|h\left(\sqrt{n}\,[\bar{I}_n(\boldsymbol{\theta_0})]^{\frac{1}{2}}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})-\boldsymbol{\theta_0}\right)\right)\right. \\
&\left. -h\left(\frac{1}{\sqrt{n}}[\bar{I}_n(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\left(\nabla(l(\boldsymbol{\theta_0};\boldsymbol{x}))\right)+\boldsymbol{R_1}(\boldsymbol{\theta_0};\boldsymbol{X})\right)\right| \qquad (4.24) \\
&+\mathrm{E}|T_2|. \qquad (4.25)
\end{aligned}
$$

The next step is based on a first order multivariate Taylor expansion. With $A_{[j]}$ denoting the j^{th} row of a matrix A, we get that

$$
\begin{aligned}
h\left(\sqrt{n}\,[\bar{I}_n(\boldsymbol{\theta_0})]^{\frac{1}{2}}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})-\boldsymbol{\theta_0}\right)\right) &= h\left(\frac{1}{\sqrt{n}}[\bar{I}_n(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\left(\nabla(l(\boldsymbol{\theta_0};\boldsymbol{x}))\right)+\boldsymbol{R_1}(\boldsymbol{\theta_0};\boldsymbol{X})\right) \\
+\sum_{j=1}^{d}\Bigg(&\sqrt{n}\left[[\bar{I}_n(\boldsymbol{\theta_0})]^{\frac{1}{2}}\right]_{[j]}(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})-\boldsymbol{\theta_0})-\frac{1}{\sqrt{n}}\left[[\bar{I}_n(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{[j]}\left(\nabla\left(l(\boldsymbol{\theta_0};\boldsymbol{X})\right)\right) \\
&-\frac{1}{2\sqrt{n}}\left[[\bar{I}_n(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{[j]}\Bigg\{\sum_{k=1}^{d}\sum_{q=1}^{d}(\hat{\theta}_n(\boldsymbol{x})_k-\theta_{0_k})(\hat{\theta}_n(\boldsymbol{x})_q-\theta_{0_q}) \\
&\qquad\times\left(\nabla\left(\frac{\partial^2}{\partial\theta_k\partial\theta_q}l(\boldsymbol{\theta};\boldsymbol{x})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0^*}\right)\right)\Bigg\}\Bigg)\frac{\partial}{\partial x_j}h(\boldsymbol{t}(\boldsymbol{X})),
\end{aligned}
$$

where $\boldsymbol{t}(\boldsymbol{X})$ is between $\frac{1}{\sqrt{n}}[\bar{I}_n(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\left(\nabla\left(l(\boldsymbol{\theta_0};\boldsymbol{X})\right)\right)+\boldsymbol{R_1}(\boldsymbol{\theta_0};\boldsymbol{X})$ and $\sqrt{n}\,[\bar{I}_n(\boldsymbol{\theta_0})]^{\frac{1}{2}}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})-\boldsymbol{\theta_0}\right)$. Thus,

$$\left| h\left(\sqrt{n}[\bar{I}_n(\boldsymbol{\theta_0})]^{\frac{1}{2}}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})-\boldsymbol{\theta_0}\right)\right) - h\left(\frac{1}{\sqrt{n}}[\bar{I}_n(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\left(\nabla(l(\boldsymbol{\theta_0};\boldsymbol{x}))\right)+\boldsymbol{R_1}(\boldsymbol{\theta_0};\boldsymbol{X})\right)\right|$$

$$\leq \|h\|_1 \left| \sum_{j=1}^{d} \left(\sqrt{n}\left[[\bar{I}_n(\boldsymbol{\theta_0})]^{\frac{1}{2}}\right]_{[j]}(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})-\boldsymbol{\theta_0}) - \frac{1}{\sqrt{n}}\left[[\bar{I}_n(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{[j]}\nabla\left(l(\boldsymbol{\theta_0};\boldsymbol{X})\right) \right.\right.$$

$$\left. -\frac{1}{2\sqrt{n}}\left[[\bar{I}_n(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{[j]} \left\{ \sum_{k=1}^{d}\sum_{q=1}^{d}(\hat{\theta}_n(\boldsymbol{x})_k-\theta_{0_k})(\hat{\theta}_n(\boldsymbol{x})_q-\theta_{0_q}) \right.\right.$$

$$\left.\left.\left. \times\left(\nabla\left(\frac{\partial^2}{\partial\theta_k\partial\theta_q}l(\boldsymbol{\theta};\boldsymbol{x})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0^*}\right)\right)\right\}\right)\right|.$$

Using (4.13) component-wise, we conclude that

$$(4.24) \leq \frac{\|h\|_1}{\sqrt{n}}\sum_{k=1}^{d}\sum_{l=1}^{d}\left|\left[[\bar{I}_n(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{lk}\right|\sum_{j=1}^{d}\mathrm{E}\left|\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)\right.$$

$$\left.\times\left(\frac{\partial^2}{\partial\theta_j\partial\theta_k}l(\boldsymbol{\theta_0};\boldsymbol{X})+n[\bar{I}_n(\boldsymbol{\theta_0})]_{kj}\right)\right|$$

$$\leq \frac{\|h\|_1}{\sqrt{n}}\sum_{k=1}^{d}\sum_{l=1}^{d}\left|\left[[\bar{I}_n(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{lk}\right|\sum_{j=1}^{d}\left[\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\right]\right.$$

$$\left.\times\,\mathrm{E}\left[\left(\frac{\partial^2}{\partial\theta_j\partial\theta_k}l(\boldsymbol{\theta_0};\boldsymbol{X})+n[\bar{I}_n(\boldsymbol{\theta_0})]_{kj}\right)^2\right]\right]^{\frac{1}{2}}, \tag{4.26}$$

where for the second inequality we used Cauchy-Schwarz inequality. To bound now $\mathrm{E}\,|T_2|$, with T_2 as in (4.23), we need to take into account that $\frac{\partial^3}{\partial\theta_k\partial\theta_q\partial\theta_j}l(\boldsymbol{\theta};\boldsymbol{x})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0^*}$ is not uniformly bounded. Therefore, with $\varepsilon>0$, the law of total expectation yields

$$\mathrm{E}\,|T_2| \leq 2\|h\|\mathbb{P}\left(\left|\hat{\theta}_n(\boldsymbol{X})_{(m)}-\theta_{0_{(m)}}\right|\geq\varepsilon\right)+\mathrm{E}\left(|T_2|\left|\left|\hat{\theta}_n(\boldsymbol{X})_{(m)}-\theta_{0_{(m)}}\right|<\varepsilon\right.\right).$$

Using Chebyshev's inequality,

$$\mathbb{P}\left(\left|\hat{\theta}_n(\boldsymbol{X})_{(m)}-\theta_{0_{(m)}}\right|\geq\varepsilon\right)\leq\frac{1}{\varepsilon^2}\mathrm{E}\left(\sum_{j=1}^{d}\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\right). \tag{4.27}$$

To bound $\mathrm{E}\left(|T_2|\middle|\left|\hat{\theta}_n(\boldsymbol{X})_{(m)}-\theta_{0_{(m)}}\right|<\varepsilon\right)$, we use a first-order Taylor expansion and

$$h\left(\frac{1}{\sqrt{n}}[\bar{I}_n(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\left(\nabla(l(\boldsymbol{\theta_0};\boldsymbol{x}))\right)+\boldsymbol{R_1}(\boldsymbol{\theta_0};\boldsymbol{x})\right)=h\left(\frac{1}{\sqrt{n}}[\bar{I}_n(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\left(\nabla\left(l(\boldsymbol{\theta_0};\boldsymbol{X})\right)\right)\right)$$
$$+\sum_{j=1}^{d}\left(\frac{1}{2\sqrt{n}}\left[[\bar{I}_n(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{[j]}\left\{\sum_{k=1}^{d}\sum_{q=1}^{d}(\hat{\theta}_n(\boldsymbol{x})_k-\theta_{0_k})(\hat{\theta}_n(\boldsymbol{x})_q-\theta_{0_q})\right.\right.$$
$$\left.\left.\times\left(\nabla\left(\frac{\partial^2}{\partial\theta_j\partial\theta_q}l(\boldsymbol{\theta};\boldsymbol{x})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0^*}\right)\right)\right\}\right)\frac{\partial}{\partial x_j}h(\tilde{\boldsymbol{t}}(\boldsymbol{X})),$$

where now $\tilde{\boldsymbol{t}}(\boldsymbol{X})$ is between $\frac{1}{\sqrt{n}}[\bar{I}_n(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\left(\nabla(l(\boldsymbol{\theta_0};\boldsymbol{x}))\right)+\boldsymbol{R_1}(\boldsymbol{\theta_0};\boldsymbol{x})$ and $\frac{1}{\sqrt{n}}[\bar{I}_n(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\left(\nabla(l(\boldsymbol{\theta_0};\boldsymbol{x}))\right)$. Using again (4.13),

$$|T_2|\leq\frac{\|h\|_1}{2\sqrt{n}}\sum_{k=1}^{d}\sum_{l=1}^{d}\left|\left[[\bar{I}_n(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{lk}\right|$$
$$\times\left\{\sum_{j=1}^{d}\sum_{v=1}^{d}\left|\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)\left(\hat{\theta}_n(\boldsymbol{X})_v-\theta_{0_v}\right)\frac{\partial^3}{\partial\theta_k\partial\theta_j\partial\theta_v}l(\boldsymbol{\theta};\boldsymbol{X})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0^*}\right|\right\}. \tag{4.28}$$

Therefore, from (4.27) and (4.28) we have for $C_{(m)}$ as in (4.5) that

$$\mathrm{E}|T_2|\leq\frac{2\|h\|}{\varepsilon^2}\mathrm{E}\left(\sum_{j=1}^{d}\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\right)$$
$$+\frac{\|h\|_1}{2\sqrt{n}}\sum_{k=1}^{d}\sum_{l=1}^{d}\left|\left[[\bar{I}_n(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{lk}\right|\mathrm{E}\left(\sum_{j=1}^{d}\sum_{v=1}^{d}\left|\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)\left(\hat{\theta}_n(\boldsymbol{X})_v-\theta_{0_v}\right)\right.\right.$$
$$\left.\left.\times\frac{\partial^3}{\partial\theta_k\partial\theta_j\partial\theta_v}l(\boldsymbol{\theta};\boldsymbol{X})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0^*}\right|\middle|\left|C_{(m)}\right|<\varepsilon\right).$$

The Cauchy-Schwarz inequality and Lemma 4.2 yield

$$
\begin{aligned}
\mathrm{E}|T_2| &\leq \frac{2\|h\|}{\varepsilon^2}\mathrm{E}\left(\sum_{j=1}^{d}\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\right)\\
&+\frac{\|h\|_1}{2\sqrt{n}}\left\{\sum_{k=1}^{d}\sum_{l=1}^{d}\left|\left[\bar{I}_n(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}}\right]_{lk}\Big|\sum_{j=1}^{d}\sum_{v=1}^{d}\left[\mathrm{E}\left(\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\left(\hat{\theta}_n(\boldsymbol{X})_v-\theta_{0_v}\right)^2\right)\right]^{\frac{1}{2}}\right.\\
&\qquad\left.\times\left[\mathrm{E}\left(\left(\sup_{\substack{\boldsymbol{\theta}:\left|\theta_m-\theta_{0_m}\right|<\varepsilon\\ \forall m\in\{1,2,\ldots,d\}}}\left|\frac{\partial^3}{\partial\theta_k\partial\theta_j\partial\theta_v}l(\boldsymbol{\theta};\boldsymbol{X})\right|\right)^2\Bigg|\left|C_{(m)}\right|<\varepsilon\right)\right]^{\frac{1}{2}}\right\}\\
&\leq \frac{2\|h\|}{\varepsilon^2}\mathrm{E}\left(\sum_{j=1}^{d}\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\right)\\
&+\frac{\|h\|_1}{2\sqrt{n}}\left\{\sum_{k=1}^{d}\sum_{l=1}^{d}\left|\left[\bar{I}_n(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}}\right]_{lk}\Big|\sum_{j=1}^{d}\sum_{v=1}^{d}\left[\mathrm{E}\left(\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\left(\hat{\theta}_n(\boldsymbol{X})_v-\theta_{0_v}\right)^2\right)\right]^{\frac{1}{2}}\right.\\
&\qquad\left.\times\left[\mathrm{E}\left(\left(M_{kjv}(\boldsymbol{X})\right)^2\Big|\left|C_{(m)}\right|<\varepsilon\right)\right]^{\frac{1}{2}}\right\}. \qquad (4.29)
\end{aligned}
$$

Therefore, from (4.26) and (4.29) we obtain that

$$
(4.15)\leq\frac{2\|h\|}{\varepsilon^2}\mathrm{E}\left(\sum_{j=1}^{d}\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\right)+\frac{\|h\|_1}{\sqrt{n}}K_1(\boldsymbol{\theta_0}), \qquad (4.30)
$$

where $K_1(\boldsymbol{\theta_0})$ is as in (4.7). Using now (4.22) and (4.30) gives the assertion. □

Remark 4.1. ***(1)*** *Assuming that* $\bar{I}_n(\boldsymbol{\theta_0})=\mathcal{O}(1)$ *in* (4.3) *and using Theorem 4.1 yields, for fixed d,* $\mathrm{E}\left(\sum_{j=1}^{d}\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\right)=\mathcal{O}\left(\frac{1}{n}\right)$. *To see this, use that from the asymptotic normality of the MLE as expressed in Theorem 4.1,* $\sqrt{n}\mathrm{E}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})-\boldsymbol{\theta_0}\right)\xrightarrow[n\to\infty]{}\mathbf{0}$ *and thus*

$$
\mathrm{E}\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)=o\left(\frac{1}{\sqrt{n}}\right),\ \forall j\in\{1,2,\ldots,d\}. \qquad (4.31)
$$

Theorem 4.1 shows that $\text{Cov}\left(\sqrt{n}\,[\bar{I}_n(\boldsymbol{\theta_0})]^{\frac{1}{2}}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})-\boldsymbol{\theta_0}\right)\right) \xrightarrow[n\to\infty]{} I_{d\times d}$. *Therefore,*

$$n\,[\bar{I}_n(\boldsymbol{\theta_0})]^{\frac{1}{2}}\,\text{Cov}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})\right)[\bar{I}_n(\boldsymbol{\theta_0})]^{\frac{1}{2}} \xrightarrow[n\to\infty]{} I_{d\times d}. \tag{4.32}$$

Assuming that the matrix $\bar{I}_n(\boldsymbol{\theta_0})$ *as defined in* (4.3) *is* $\mathcal{O}(1)$*, it follows from* (4.32) *that*

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

Combining these results,

$$\text{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\right] = \text{Var}\left(\hat{\theta}_n(\boldsymbol{X})_j\right)+\left[\text{E}\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)\right]^2 = \mathcal{O}\left(\frac{1}{n}\right). \tag{4.33}$$

Furthermore, using (4.31), (4.33) *then if* $\bar{I}_n(\boldsymbol{\theta_0}) = \mathcal{O}(1)$ *it can be deduced that*

$$\begin{aligned} K_1(\boldsymbol{\theta_0}) &= \mathcal{O}(1)\\ K_2(\boldsymbol{\theta_0}) &= \mathcal{O}(1)\\ K_3(\boldsymbol{\theta_0}) &= \mathcal{O}(1), \end{aligned} \tag{4.34}$$

where $K_1(\boldsymbol{\theta_0}), K_2(\boldsymbol{\theta_0}), K_3(\boldsymbol{\theta_0})$ *are as in* (4.7), (4.8), (4.9)*, respectively. Hence, using* (4.33) *and* (4.34)*, if* $\bar{I}_n(\boldsymbol{\theta_0}) = \mathcal{O}(1)$ *then the upper bound in Theorem 4.2 is* $\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$.

(2) *Often Hölder's inequality will be used to bound the third term as the calculation of absolute third moments can be quite complicated, even for simple multi-parameter distributions.*

(3) *In terms of the dimensionality d of the parameter,* $K_1(\boldsymbol{\theta_0}) = \mathcal{O}\left(d^4\right), K_2(\boldsymbol{\theta_0}) = \mathcal{O}\left(d^4\right)$ *and* $K_3(\boldsymbol{\theta_0}) = \mathcal{O}\left(d^8\right)$ *as can be deduced from* (4.7), (4.8) *and* (4.9)*, respectively. The last term of the bound in* (4.6) *is of order d in terms of the dimensionality of the parameter. Thus, for* $d \gg n$ *the bound does not behave well, but d could grow moderately with n. For example* $d = o\left(n^{\alpha}\right),\ 0 < \alpha < \frac{1}{16}$ *would still yield a bound which goes to zero as n goes to infinity.*

4.1.2 Linear regression

This subsection calculates the bound in (4.6) for linear regression models. The asymptotic normality of the MLE in linear regression models has been proven in Fahrmeir and Kaufmann (1985). We give the example of a straight-line regression and the bound turns out to be of order $\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$, where n is the sample size. The following notation is used throughout this subsection. The vector $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)^\intercal \in \mathbb{R}^{n\times 1}$ denotes the response variable for the linear regression, while $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_d)^\intercal \in \mathbb{R}^{d\times 1}$ is the vector of the d parameters and $\boldsymbol{\epsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^\intercal \in \mathbb{R}^{n\times 1}$ is the vector of the error terms, which are i.i.d. random variables with $\varepsilon_i \sim \mathrm{N}(0, \sigma^2) \forall i \in \{1, 2, \ldots n\}$. The true value of the unknown parameter $\boldsymbol{\beta}$ is denoted by $\boldsymbol{\beta_0} = (\beta_{0_1}, \beta_{0_2}, \ldots, \beta_{0_d})^\intercal \in \mathbb{R}^{d\times 1}$. The design matrix is

$$X = \begin{pmatrix} 1 & x_{1,2} & \ldots & x_{1,d} \\ 1 & x_{2,2} & \ldots & x_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,2} & \ldots & x_{n,d} \end{pmatrix}.$$

For the model

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

the aim is to find bounds on the distributional distance between the distribution of the MLE, $\hat{\boldsymbol{\beta}}$, and the normal distribution. In the case where $\beta_j, j = 1, 2, \ldots, d$ are independent random variables, then we have linear random-effects models and bounds related to the normal approximation of variance component maximum likelihood estimators are given in Dicker and Erdogdu (2016+).

The probability density function for Y_i is

$$f(y_i|\boldsymbol{\beta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\left(y_i - X_{[i]}\boldsymbol{\beta}\right)^2\right\}, \tag{4.35}$$

where $X_{[i]}$ denotes the i^{th} row of the design matrix. The parameter space $\Theta = \mathbb{R}^d$ is open and if $X^\intercal X$ is of full rank, the matrix $X^\intercal X$ is invertible and the vector MLE is

$$\hat{\beta} = (X^\intercal X)^{-1} X^\intercal \boldsymbol{Y}. \tag{4.36}$$

We now bound the corresponding distributional distance.

Corollary 4.1. *Let $Y_i, i \in \{1,2,\ldots,n\}$ be independent normal random variables with*

$$Y_i \sim \mathrm{N}\left(X_{[i]}\beta_0, \sigma^2\right),$$

where $X_{[i]}$ denotes the i^{th} row of the design matrix X and σ^2 is known. Assume that the $d \times d$ matrix $X^\intercal X$ is of full rank. Let $\{Y_i', i = 1,2,\ldots,n\}$ be an independent copy of $\{Y_i, i = 1,2,\ldots,n\}$ and $\boldsymbol{Z} \sim \mathrm{N}_d(\mathbf{0}, I_{d\times d})$ and $\bar{I}_n(\beta) = \frac{1}{n}\sum_{j=1}^n I_j(\beta)$. Then for $h : \mathbb{R}^d \to \mathbb{R}$, a three times differentiable function with bounded derivatives,

$$\begin{aligned}
&\left|\mathrm{E}\left[h\left(\sqrt{n}[\bar{I}_n(\beta)]^{\frac{1}{2}}\left(\hat{\beta} - \beta_0\right)\right)\right] - \mathrm{E}[h(\boldsymbol{Z})]\right| \\
&\leq \frac{\|h\|_2}{4}\sum_{j=1}^d\left[\sum_{i=1}^n \mathrm{Var}\left(\left(\sum_{k=1}^d \frac{X_{ik}}{\sigma}\left[[X^\intercal X]^{-\frac{1}{2}}\right]_{jk}\left(Y_i - \sum_{m=1}^d X_{im}\beta_{0_m}\right)\right)^2\right)\right]^{\frac{1}{2}} \\
&+ \frac{\|h\|_2}{2}\sum_{j>k}^d\sum_{k=1}^{d-1}\left[\sum_{i=1}^n \mathrm{Var}\left(\sum_{q=1}^d\sum_{v=1}^d \frac{X_{iq}X_{iv}}{\sigma^2}\left[[X^\intercal X]^{-\frac{1}{2}}\right]_{jq}\left[[X^\intercal X]^{-\frac{1}{2}}\right]_{kv}\right.\right. \\
&\qquad\qquad \left.\left.\times\left(Y_i - \sum_{m=1}^d X_{im}\beta_{0_m}\right)^2\right)\right]^{\frac{1}{2}} \\
&+ \frac{\|h\|_3}{12}\sum_{i=1}^n \mathrm{E}\left(\sum_{m=1}^d\left|\sum_{l=1}^d \frac{X_{il}}{\sigma}\left[[X^\intercal X]^{-\frac{1}{2}}\right]_{ml}\left(Y_i - Y_i'\right)\right|\right)^3.
\end{aligned} \tag{4.37}$$

Proof. Using (4.35) and a vector representation, we get the following expressions for the log-likelihood and its first derivative with respect to the parameter vector:

$$
\begin{aligned}
l(\boldsymbol{\beta};\boldsymbol{y}) &= -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}(\boldsymbol{y}-X\boldsymbol{\beta})^\intercal(\boldsymbol{y}-X\boldsymbol{\beta}) \\
\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\beta}}l(\boldsymbol{\beta};\boldsymbol{y}) &= -\frac{1}{2\sigma^2}\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\beta}}\left\{\left(\boldsymbol{y}^\intercal-\boldsymbol{\beta}^\intercal X^\intercal\right)(\boldsymbol{y}-X\boldsymbol{\beta})\right\} \\
&= -\frac{1}{2\sigma^2}\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\beta}}\left\{\boldsymbol{y}^\intercal\boldsymbol{y}-\boldsymbol{\beta}^\intercal X^\intercal\boldsymbol{y}-\boldsymbol{y}^\intercal X\boldsymbol{\beta}+\boldsymbol{\beta}^\intercal X^\intercal X\boldsymbol{\beta}\right\} \\
&= \frac{1}{\sigma^2}\left(X^\intercal\boldsymbol{y}-X^\intercal X\boldsymbol{\beta}\right).
\end{aligned}
$$

The Hessian matrix for the log-likelihood function,

$$
\boldsymbol{H}(\boldsymbol{\beta};\boldsymbol{y}) = \left\{\frac{\partial^2}{\partial\beta_i\partial\beta_j}l(\boldsymbol{\beta};\boldsymbol{y})\right\}_{i,j=1,\cdots,d} = -\frac{1}{\sigma^2}X^\intercal X,
$$

does not depend on $\boldsymbol{y}$. Thus, $\bar{I}_n(\boldsymbol{\beta_0}) = \frac{1}{n\sigma^2}X^\intercal X$ and so $\left[\bar{I}_n(\boldsymbol{\beta})\right]^{-\frac{1}{2}} = \sigma\sqrt{n}\left[X^\intercal X\right]^{-\frac{1}{2}}$. The result in (4.36) gives that

$$
\begin{aligned}
\sqrt{n}\left[\bar{I}_n(\boldsymbol{\beta_0})\right]^{\frac{1}{2}}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta_0}\right) &= \frac{1}{\sigma}\left\{\left[X^\intercal X\right]^{-\frac{1}{2}}X^\intercal\boldsymbol{Y}-\left[X^\intercal X\right]^{\frac{1}{2}}\boldsymbol{\beta_0}\right\} \\
&= \frac{1}{\sqrt{n}}\left[\sigma\sqrt{n}\left[X^\intercal X\right]^{-\frac{1}{2}}\right]\frac{1}{\sigma^2}\left(X^\intercal\boldsymbol{Y}-X^\intercal X\boldsymbol{\beta_0}\right) \\
&= \frac{1}{\sqrt{n}}\left[I_n(\boldsymbol{\beta_0})\right]^{-\frac{1}{2}}\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\beta}}l(\boldsymbol{\beta};\boldsymbol{y})\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta_0}}. \qquad (4.38)
\end{aligned}
$$

The expression in (4.38) is the same as W in (4.16) and therefore the quantity of interest $\left|\mathrm{E}\left[h\left(\sqrt{n}[\bar{I}_n(\boldsymbol{\beta_0})]^{\frac{1}{2}}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta_0}\right)\right)\right]-\mathrm{E}[h(\boldsymbol{Z})]\right|$ is equal to (4.14) with (4.15) being equal to zero for this specific case of the linear regression model. Therefore, using (4.22) and

$$
\frac{\partial}{\partial\beta_k}\log(f_i(Y_i|\boldsymbol{\beta_0})) = \frac{X_{ik}}{\sigma^2}\left(Y_i-\sum_{m=1}^{d}X_{im}\beta_{0_m}\right)
$$

in Theorem 4.2 yields the result of the Corollary. □

Example: The simple linear model (d=2)

Here, we apply the results of (4.37) to the case of a straight-line regression with two unknown parameters. The model is

$$Y_i = \beta_1 + \beta_2(x_i - \bar{x}) + \varepsilon_i, \quad \forall i \in \{1,2,\ldots,n\}.$$

The unknown parameters β_1 and β_2 are the intercept and slope of the regression, respectively. The i.i.d. random variables $\varepsilon_i \sim \mathrm{N}(0,\sigma^2), \forall i \in \{1,2,\ldots,n\}$ are as before. It is well known that the MLE exists, it is unique and equal to $\hat{\beta} = \left(\bar{Y}, \frac{\sum_{i=1}^n (x_i-\bar{x})Y_i}{\sum_{i=1}^n (x_i-\bar{x})^2}\right)^\intercal$. The following corollary of Theorem 4.2 gives, for this specific example, an upper bound on the distributional distance between the exact distribution of $\hat{\beta}$ and the normal distribution.

Corollary 4.2. *Let $Y_1, Y_2, \ldots, Y_n$ be independent random variables with $Y_i \sim \mathrm{N}(\beta_1 + \beta_2(x_i - \bar{x}), \sigma^2)$, with the case of $x_i = x_j$ for $i \neq j$, $i, j \in \{1,2,\ldots,n\}$ being excluded. For $\boldsymbol{Z} \sim \mathrm{N}_2(\mathbf{0}, I_{2\times 2})$ and $h \in H$ as in* (4.1)*,*

$$\left|\mathrm{E}\left[h\left(\sqrt{n}[\bar{I}_n(\beta)]^{\frac{1}{2}}\left(\hat{\beta} - \beta_{\mathbf{0}}\right)\right)\right] - \mathrm{E}[h(\boldsymbol{Z})]\right| \leq \frac{\|h\|_2}{4}\left(\sqrt{\frac{2}{n}} + \frac{\sqrt{2\sum_{i=1}^n (x_i-\bar{x})^4}}{\sum_{i=1}^n (x_i-\bar{x})^2}\right) + \frac{\|h\|_2}{\sqrt{2n}} + \frac{8\|h\|_3}{3\sqrt{\pi}}\left(\frac{1}{\sqrt{n}} + \frac{\sum_{i=1}^n |x_i-\bar{x}|^3}{[\sum_{i=1}^n (x_i-\bar{x})^2]^{\frac{3}{2}}}\right).$$

Proof. We have that

$$X = \begin{pmatrix} 1 & x_1 - \bar{x} \\ 1 & x_2 - \bar{x} \\ \vdots & \vdots \\ 1 & x_n - \bar{x} \end{pmatrix}, \; X^\intercal X = \begin{pmatrix} n & 0 \\ 0 & \sum_{i=1}^n (x_i-\bar{x})^2 \end{pmatrix}, \; [X^\intercal X]^{-\frac{1}{2}} = \begin{pmatrix} \frac{1}{\sqrt{n}} & 0 \\ 0 & \frac{1}{\sqrt{\sum_{i=1}^n (x_i-\bar{x})^2}} \end{pmatrix}, \tag{4.39}$$

which shows that $X^\intercal X$ is invertible if and only if $\sum_{i=1}^n (x_i - \bar{x})^2 \neq 0$, which holds if x_i's are not all identical. The quantities of the bound in (4.37) are calculated for this specific

case. We use that $Y_i - \beta_1 - (x_i - \bar{x})\beta_2 \overset{\mathrm{d}}{=\!=} \sigma Z_i$, where $Z_i \sim \mathrm{N}(0,1)$. For the first term in (4.37) we obtain that

$$\begin{aligned}
&\sum_{j=1}^{2}\left[\sum_{i=1}^{n}\mathrm{Var}\left(\left(\sum_{k=1}^{2}\frac{X_{ik}}{\sigma}\left[[X^{\intercal}X]^{-\frac{1}{2}}\right]_{jk}\left(Y_i-\sum_{m=1}^{2}X_{im}\beta_m\right)\right)^2\right)\right]^{\frac{1}{2}}\\
&=\sum_{j=1}^{2}\left[\sum_{i=1}^{n}\mathrm{Var}\left(\left(\left(\frac{X_{i1}}{\sigma}\left[[X^{\intercal}X]^{-\frac{1}{2}}\right]_{j1}+\frac{X_{i2}}{\sigma}\left[[X^{\intercal}X]^{-\frac{1}{2}}\right]_{j2}\right)(\sigma Z_i)\right)^2\right)\right]^{\frac{1}{2}}\\
&=\left[\sum_{i=1}^{n}\mathrm{Var}\left(\left(\frac{1}{\sqrt{n}}Z_i\right)^2\right)\right]^{\frac{1}{2}}+\left[\sum_{i=1}^{n}\mathrm{Var}\left(\left(\frac{x_i-\bar{x}}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}Z_i\right)^2\right)\right]^{\frac{1}{2}}\\
&=\frac{1}{n}\left[\sum_{i=1}^{n}\mathrm{Var}\left(Z_i^2\right)\right]^{\frac{1}{2}}+\frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\left[\sum_{i=1}^{n}(x_i-\bar{x})^4\mathrm{Var}\left(Z_i^2\right)\right]^{\frac{1}{2}}\\
&=\sqrt{\frac{2}{n}}+\frac{\sqrt{2\sum_{i=1}^{n}(x_i-\bar{x})^4}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}.
\end{aligned}\tag{4.40}$$

For an upper bound for the second term of (4.37), since $d=2$ we have that $k=1$, $j=2$ leading to

$$\begin{aligned}
&\left[\sum_{i=1}^{n}\mathrm{Var}\left(\sum_{q=1}^{2}\sum_{v=1}^{2}\frac{X_{iq}X_{iv}}{\sigma^2}\left[[X^{\intercal}X]^{-\frac{1}{2}}\right]_{2q}\left[[X^{\intercal}X]^{-\frac{1}{2}}\right]_{1v}\left(Y_i-\sum_{m=1}^{2}X_{im}\beta_m\right)^2\right)\right]^{\frac{1}{2}}\\
&=\left[\sum_{i=1}^{n}\mathrm{Var}\left(\frac{X_{i2}X_{i1}}{\sigma^2}\left[[X^{\intercal}X]^{-\frac{1}{2}}\right]_{22}\left[[X^{\intercal}X]^{-\frac{1}{2}}\right]_{11}(\sigma Z_i)^2\right)\right]^{\frac{1}{2}}\\
&=\left[\sum_{i=1}^{n}\mathrm{Var}\left(\frac{x_i-\bar{x}}{\sqrt{n\sum_{i=1}^{n}(x_i-\bar{x})^2}}Z_i^2\right)\right]^{\frac{1}{2}}=\frac{1}{\sqrt{n\sum_{i=1}^{n}(x_i-\bar{x})^2}}\left[\sum_{i=1}^{n}\mathrm{Var}\left(Z_i^2\right)(x_i-\bar{x})^2\right]^{\frac{1}{2}}\\
&=\sqrt{\frac{2}{n}}.
\end{aligned}\tag{4.41}$$

For the final term of (4.37), because Y_i' is an independent copy of Y_i, then $Y_i' - Y_i \sim \mathrm{N}(0, 2\sigma^2)$, with $\mathrm{E}\,|Y_i' - Y_i|^3 = 8\frac{\sigma^3}{\sqrt{\pi}}$. Using that

$$(|a|+|b|)^3 \le 4\left(|a|^3+|b|^3\right),\ \ a,b\in\mathbb{R} \tag{4.42}$$

yields

$$\begin{aligned}
&\sum_{i=1}^{n} \mathrm{E}\left(\sum_{m=1}^{2}\left|\sum_{l=1}^{2}\frac{X_{il}}{\sigma}\left[[X^{\intercal}X]^{-\frac{1}{2}}\right]_{ml}\left(Y_i-Y_i'\right)\right|\right)^3 \\
&=\sum_{i=1}^{n}\mathrm{E}\left(\left|\left(\frac{X_{i1}}{\sigma}\left[[X^{\intercal}X]^{-\frac{1}{2}}\right]_{11}+\frac{X_{i2}}{\sigma}\left[[X^{\intercal}X]^{-\frac{1}{2}}\right]_{22}\right)\left(Y_i-Y_i'\right)\right|\right)^3 \\
&\leq\sum_{i=1}^{n}\mathrm{E}\left(\left(\frac{1}{\sigma\sqrt{n}}+\frac{|x_i-\bar{x}|}{\sigma\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}\right)\left|Y_i'-Y_i\right|\right)^3 \\
&\leq 4\sum_{i=1}^{n}\left(\frac{8}{n^{\frac{3}{2}}\sqrt{\pi}}+\frac{8|x_i-\bar{x}|^3}{[\sum_{i=1}^{n}(x_i-\bar{x})^2]^{\frac{3}{2}}\sqrt{\pi}}\right)=\frac{32}{\sqrt{\pi}}\left(\frac{1}{\sqrt{n}}+\frac{\sum_{i=1}^{n}|x_i-\bar{x}|^3}{[\sum_{i=1}^{n}(x_i-\bar{x})^2]^{\frac{3}{2}}}\right). \quad (4.43)
\end{aligned}$$

Summarizing, in the case of $Y_1, Y_2, \ldots, Y_n$ being independent random variables with $Y_i \sim \mathrm{N}\left(\beta_1+\beta_2(x_i-\bar{x}), \sigma^2\right)$, we apply to (4.37) the results of (4.40), (4.41) and (4.43) to obtain the assertion of the corollary. □

Remark 4.2. *The bound is* $\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$.

4.2 Identically distributed random vectors

In this section we use weaker regularity conditions than (N1)-(N8) in order to find an upper bound in the case of independent and identically distributed random vectors. Following Davison (2008), we make the following assumptions:

(R.C.1) The densities defined by any two different values of $\boldsymbol{\theta}$ are distinct;

(R.C.2) the log-likelihood function is three times differentiable with respect to the unknown vector parameter, $\boldsymbol{\theta}$, and the third partial derivatives are continuous in $\boldsymbol{\theta}$;

(R.C.3) for any $\boldsymbol{\theta_0} \in \boldsymbol{\Theta}$ and for $\mathbb{X}$ denoting the support of the data $\boldsymbol{x} = (\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n})$, there exists $\varepsilon_0 > 0$ and functions $M_{rst}(\boldsymbol{x})$ (they can depend on $\boldsymbol{\theta_0}$), such that

for $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d)$ and $r, s, t, j = 1, \ldots, d$,

$$\frac{1}{n}\left|\frac{\partial^3}{\partial\theta_r \partial\theta_s \partial\theta_t} l(\boldsymbol{\theta}; \boldsymbol{x})\right| \leq M_{rst}(\boldsymbol{x}),\ \forall \boldsymbol{x} \in \mathbb{X},\ \left|\theta_j - \theta_{0_j}\right| < \varepsilon_0,$$

with $\mathrm{E}(M_{rst}(\boldsymbol{X})) < \infty$;

(R.C.4) the expected Fisher information matrix $I(\boldsymbol{\theta})$ is finite, symmetric and positive definite. For $r, s = 1, 2, \ldots d$, its elements satisfy

$$[I(\boldsymbol{\theta})]_{rs} = \mathrm{E}\left\{\frac{\partial}{\partial\theta_r} l(\boldsymbol{\theta}; \boldsymbol{X}) \frac{\partial}{\partial\theta_s} l(\boldsymbol{\theta}; \boldsymbol{X})\right\} = \mathrm{E}\left\{-\frac{\partial^2}{\partial\theta_r \partial\theta_s} l(\boldsymbol{\theta}; \boldsymbol{X})\right\}.$$

This condition implies that $I(\boldsymbol{\theta})$ is the covariance matrix of the score vector.

These regularity conditions in the multi-parameter case resemble those in Section 2.2 where it is assumed that the parameter is scalar. From now on, unless otherwise stated, the notation $I(\boldsymbol{\theta})$ stands for the expected Fisher information matrix for one random vector. Under (R.C.1)-(R.C.4), (Davison, 2008, p.118) shows that

$$\sqrt{n}[I(\boldsymbol{\theta_0})]^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) - \boldsymbol{\theta_0}) \xrightarrow[n\to\infty]{\mathrm{d}} \mathrm{N}_d(\mathbf{0}, I_{d\times d})).$$

The upper bound on the distributional distance between the distribution of a vector MLE and the multivariate normal in the case of i.i.d. random vectors is the same as the bound in Theorem 4.2 and thus it is not given again. The bound can be simplified due to the fact that in the i.i.d. case $\bar{I}_n(\boldsymbol{\theta_0}) = I(\boldsymbol{\theta_0})$ and $f_i(\boldsymbol{x_i}) = f(\boldsymbol{x_i})$, $\forall i \in \{1, 2, \ldots, n\}$. In the next example of independent random variables from the normal distribution with both mean and variance unknown the bound can be easily calculated and it is, as expected, of the order $\frac{1}{\sqrt{n}}$.

4.2.1 The normal distribution

Here, we apply Theorem 4.2 in the case of $X_1, X_2, \ldots, X_n$ independent and identically distributed random variables from $\mathrm{N}(\mu, \sigma^2)$ with $\boldsymbol{\theta_0} = (\mu, \sigma^2)$. It is well-known that the MLE exists, it is unique and equal to $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) = \left(\hat{\mu}, \hat{\sigma^2}\right)^{\mathsf{T}} = \left(\bar{X}, \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^{\mathsf{T}}$; see for example (Davison, 2008, p.116). In addition, the regularity conditions (R.C.1)-(R.C.4) are satisfied.

Corollary 4.3. *Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables that follow the $N(\mu, \sigma^2)$ distribution. For $\boldsymbol{Z} \sim \mathrm{N}_d(\mathbf{0}, I_{d\times d})$ and $h \in H$ as defined in* (4.1)*,*

$$\begin{aligned}\left|\mathrm{E}\left[h\left(\sqrt{n}[I(\boldsymbol{\theta_0})]^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) - \boldsymbol{\theta_0})\right)\right] - \mathrm{E}[h(\boldsymbol{Z})]\right| &\leq \frac{5}{2}\frac{\|h\|_2}{\sqrt{n}} + 19\frac{\|h\|_3}{\sqrt{n}} \\ &+ \frac{8\|h\|}{n\sigma^2}\left(1 + 2\sigma^2\right) + 4\sqrt{2}\frac{\|h\|_1}{\sqrt{n}} + \frac{4\|h\|_1}{\sqrt{n}}\left[3 + 16\sqrt{2}\sqrt{\frac{1}{n} + \frac{\sigma^2}{4}}\right] \\ &+ \frac{32\|h\|_1}{\sqrt{n}}\left[1 + 648\left[\left(\frac{3}{2} + \frac{\sigma^2}{4}\right)^2 + \frac{3}{n^2}\right]\right]^{\frac{1}{2}}. \end{aligned} \tag{4.44}$$

Proof. The first and second-order partial derivatives are

$$\begin{aligned}&\frac{\partial}{\partial\mu}l(\boldsymbol{\theta_0};\boldsymbol{x}) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu), && \frac{\partial}{\partial\sigma^2}l(\boldsymbol{\theta_0};\boldsymbol{x}) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2, \\ &\frac{\partial^2}{\partial\mu^2}l(\boldsymbol{\theta_0};\boldsymbol{x}) = -\frac{n}{\sigma^2}, && \frac{\partial^2}{\partial\sigma^4}l(\boldsymbol{\theta_0};\boldsymbol{x}) = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{n}(x_i - \mu)^2, \\ &\frac{\partial^2}{\partial\mu\partial\sigma^2}l(\boldsymbol{\theta_0};\boldsymbol{x}) = \frac{\partial^2}{\partial\sigma^2\partial\mu}l(\boldsymbol{\theta_0};\boldsymbol{x}) = -\frac{n}{\sigma^4}(\bar{X} - \mu). \end{aligned} \tag{4.45}$$

Hence the expected Fisher Information matrix for one random variable is

$$I(\boldsymbol{\theta_0}) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}, \text{ so that } [I(\boldsymbol{\theta_0})]^{-\frac{1}{2}} = \begin{pmatrix} \sigma & 0 \\ 0 & \sqrt{2}\sigma^2 \end{pmatrix}. \tag{4.46}$$

Using (4.16), let

$$\boldsymbol{W} = \frac{1}{\sqrt{n}}\left(\frac{1}{\sigma}\sum_{i=1}^{n}(X_i-\mu), -\frac{n}{\sqrt{2}} + \frac{1}{\sqrt{2}\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2\right)^{\mathsf{T}} = (W_1, W_2)^{\mathsf{T}},$$

where $W_i = \sum_{j=1}^{n}\xi_{ji}$ with $\xi_{i1} = \frac{X_i-\mu}{\sqrt{n}\sigma}$ and $\xi_{i2} = \frac{(X_i-\mu)^2-\sigma^2}{\sigma^2\sqrt{2n}}, i = 1,2,\ldots,n$. Now we bound the terms in Theorem 4.2 in order of appearance. We start with the term $\frac{\|h\|_1}{\sqrt{n}}K_1(\boldsymbol{\theta_0})$, where $K_1(\boldsymbol{\theta_0})$ is given in (4.7). For the first quantity, using that $\mathrm{E}\left[(\bar{X}-\mu)^2\right] = \frac{\sigma^2}{n}$ and $\mathrm{E}\left[\left(\hat{\sigma}^2-\sigma^2\right)^2\right] = \frac{\sigma^4}{n}\left(2-\frac{1}{n}\right)$ yields

$$\begin{aligned}
&\frac{\|h\|_1}{\sqrt{n}}\sum_{k=1}^{2}\sum_{l=1}^{2}\left|\left[\left[\bar{I}_n(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}}\right]_{lk}\right| \\
&\qquad\times\sum_{j=1}^{2}\left[\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\right]\mathrm{E}\left[\left(\frac{\partial^2}{\partial\theta_j\partial\theta_k}l(\boldsymbol{\theta_0};\boldsymbol{X})+n[\bar{I}_n(\boldsymbol{\theta_0})]_{kj}\right)^2\right]\right]^{\frac{1}{2}} \\
&= \frac{\sigma\|h\|_1}{\sqrt{n}}\left[\mathrm{E}\left[\left(\hat{\sigma}^2-\sigma^2\right)^2\right]\mathrm{E}\left[\left(\frac{-n(\bar{X}-\mu)}{\sigma^4}\right)^2\right]\right]^{\frac{1}{2}} \\
&\quad+\frac{\sqrt{2}\sigma^2\|h\|_1}{\sqrt{n}}\left\{\left[\mathrm{E}\left[(\bar{X}-\mu)^2\right]\mathrm{E}\left[\left(\frac{-n(\bar{X}-\mu)}{\sigma^4}\right)^2\right]\right]^{\frac{1}{2}}\right. \\
&\qquad\left.+\left[\mathrm{E}\left[\left(\hat{\sigma}^2-\sigma^2\right)^2\right]\mathrm{E}\left[\left(\frac{n}{\sigma^4}-\frac{1}{\sigma^6}\sum_{i=1}^{n}(X_i-\mu)^2\right)^2\right]\right]^{\frac{1}{2}}\right\} \\
&\le \frac{\sqrt{n}\|h\|_1}{\sigma^3}\sqrt{\mathrm{E}\left(\hat{\sigma}^2-\sigma^2\right)^2}\sqrt{\mathrm{E}(\bar{X}-\mu)^2}+\frac{\sqrt{2n}\|h\|_1}{\sigma^2}\mathrm{E}\left[(\bar{X}-\mu)^2\right] \\
&\quad+\frac{\sqrt{2}\sigma^2\|h\|_1}{\sqrt{n}}\sqrt{\mathrm{E}\left[\left(\hat{\sigma}^2-\sigma^2\right)^2\right]\mathrm{E}\left[\left(\frac{n}{\sigma^4}-\frac{1}{\sigma^6}\sum_{i=1}^{n}(X_i-\mu)^2\right)^2\right]} \\
&\le \frac{\|h\|_1}{\sqrt{n}}\sqrt{2-\frac{1}{n}}+\frac{\sqrt{2}\|h\|_1}{\sqrt{n}} \\
&\quad+\frac{\sqrt{2}\sigma^4\|h\|_1}{n}\sqrt{\left(2-\frac{1}{n}\right)\left(\frac{n^2}{\sigma^8}-\frac{2n}{\sigma^{10}}\mathrm{E}\left[\sum_{i=1}^{n}(X_i-\mu)^2\right]+\frac{1}{\sigma^{12}}\left[\mathrm{E}\left[\sum_{i=1}^{n}(X_i-\mu)^2\right]\right]^2\right)} \\
&= \frac{\|h\|_1}{\sqrt{n}}\sqrt{2-\frac{1}{n}}+\frac{\sqrt{2}\|h\|_1}{\sqrt{n}}+\frac{\sqrt{2}\sigma^4\|h\|_1}{n}\sqrt{\left(2-\frac{1}{n}\right)\left(\frac{n^2}{\sigma^8}-\frac{2n^2}{\sigma^8}+\frac{n}{\sigma^8}(2+n)\right)}
\end{aligned}$$

$$= \frac{\|h\|_1}{\sqrt{n}}\left(\sqrt{2-\frac{1}{n}}+\sqrt{2}\right)+\frac{\sqrt{2}\|h\|_1}{\sqrt{n}}\sqrt{2\left(2-\frac{1}{n}\right)} \leq 4\sqrt{2}\frac{\|h\|_1}{\sqrt{n}}. \tag{4.47}$$

For the second quantity in (4.7), we find an upper bound for

$$\sup_{\boldsymbol{\theta}:\left|\theta_m-\theta_{0_m}\right|<\varepsilon}\left|\frac{\partial^3}{\partial\theta_k\partial\theta_j\partial\theta_i}l(\boldsymbol{\theta};\boldsymbol{X})\right| \leq nM_{kji}(\boldsymbol{X}),$$

where $m,k,j,i \in \{1,2\}$ and $M_{kji}(\boldsymbol{x})$ is as in (R.C.3). Below, $\boldsymbol{\theta}=(\theta_1,\theta_2)$ is the vector parameter and $\boldsymbol{\theta_0}=(\theta_{0_1},\theta_{0_2})=(\mu,\sigma^2)$ is the true, unknown value. We have

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

as well as

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}:\left|\theta_m-\theta_{0_m}\right|<\varepsilon}\left|\frac{\partial^3}{\partial\theta_2^3}l(\boldsymbol{\theta};\boldsymbol{X})\right| = \sup_{\boldsymbol{\theta}:\left|\theta_m-\theta_{0_m}\right|<\varepsilon}\left|-\frac{n}{\theta_2^3}+\frac{3}{\theta_2^4}\sum_{i=1}^n(X_i-\theta_1)^2\right| \\
&< \frac{n}{(\sigma^2-\varepsilon)^3}+\frac{3}{(\sigma^2-\varepsilon)^4}\sup_{\boldsymbol{\theta}:\left|\theta_m-\theta_{0_m}\right|<\varepsilon}\left|\sum_{i=1}^n(X_i-\bar{X}+\bar{X}-\mu+\mu-\theta_1)^2\right| \\
&\leq \frac{n}{(\sigma^2-\varepsilon)^3}+\frac{3}{(\sigma^2-\varepsilon)^4}\sup_{\boldsymbol{\theta}:\left|\theta_m-\theta_{0_m}\right|<\varepsilon}\left|3\sum_{i=1}^n\left[(X_i-\bar{X})^2+(\bar{X}-\mu)^2+(\mu-\theta_1)^2\right]\right| \\
&< \frac{n}{(\sigma^2-\varepsilon)^3}+\frac{9n}{(\sigma^2-\varepsilon)^4}\left(\hat{\sigma}^2+(\bar{X}-\mu)^2+\varepsilon^2\right) =: nM_{222}(\boldsymbol{X}).
\end{aligned} \tag{4.48}$$

Moreover,

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}:\left|\theta_m-\theta_{0_m}\right|<\varepsilon}\left|\frac{\partial^3}{\partial\theta_1\partial\theta_2^2}l(\boldsymbol{\theta};\boldsymbol{X})\right| = \sup_{\boldsymbol{\theta}:\left|\theta_m-\theta_{0_m}\right|<\varepsilon}\left|\frac{\partial^3}{\partial\theta_2^2\partial\theta_1}l(\boldsymbol{\theta};\boldsymbol{X})\right| \\
&= \sup_{\boldsymbol{\theta}:\left|\theta_m-\theta_{0_m}\right|<\varepsilon}\left|\frac{2n}{\theta_2^3}(\bar{X}-\theta_1)\right| \\
&= \sup_{\boldsymbol{\theta}:\left|\theta_m-\theta_{0_m}\right|<\varepsilon}\left|\frac{2n}{\theta_2^3}(\bar{X}-\mu+\mu-\theta_1)\right| \\
&< \frac{2n}{(\sigma^2-\varepsilon)^3}(|\bar{X}-\mu|+\varepsilon) =: nM_{122}(\boldsymbol{X})
\end{aligned} \tag{4.49}$$

and

$$\sup_{\theta:\left|\theta_m-\theta_{0_m}\right|<\varepsilon}\left|\frac{\partial^3}{\partial\theta_1^2\partial\theta_2}l(\boldsymbol{\theta};\boldsymbol{X})\right| = \sup_{\theta:\left|\theta_m-\theta_{0_m}\right|<\varepsilon}\left|\frac{\partial^3}{\partial\theta_2\partial\theta_1^2}l(\boldsymbol{\theta};\boldsymbol{X})\right|$$
$$= \sup_{\theta:\left|\theta_m-\theta_{0_m}\right|<\varepsilon}\left|\frac{n}{\theta_2^2}\right| < \frac{n}{(\sigma^2-\varepsilon)^2} =: nM_{112}(\boldsymbol{X}). \tag{4.50}$$

In addition, we calculate $\mathrm{E}\left(\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\left(\hat{\theta}_n(\boldsymbol{X})_i-\theta_{0_i}\right)^2\right), \forall i,j\in\{1,2\}$. For this purpose we use that in the case of i.i.d. random variables from the Normal distribution,

$$\mathrm{Cov}\left(\bar{X}-\mu, \frac{1}{n}\sum_{i=1}^n (X_i-\bar{X})^2-\sigma^2\right)=0 \qquad \text{(Casella and Berger, 2002, p.218).} \tag{4.51}$$

We have that

$$\mathrm{E}\,(\bar{X}-\mu)^4 = \frac{3\sigma^4}{n^2}$$

and

$$\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^n (X_i-\bar{X})^2-\sigma^2\right)^4 = \frac{\sigma^8}{n^2}\left(12+\frac{4}{n}-\frac{15}{n^2}\right) < 16\frac{\sigma^8}{n^2}, \text{ for all } n\in\mathbb{N}$$

and with $G\sim\chi^2_{n-1}$, so that $\mathrm{E}(G)=n-1$ and $\mathrm{Var}(G)=2(n-1)$,

$$\begin{aligned}
&\mathrm{E}\left[(\bar{X}-\mu)^2\left(\frac{1}{n}\sum_{i=1}^n (X_i-\bar{X})^2-\sigma^2\right)^2\right]\\
&= \mathrm{E}\left[(\bar{X}-\mu)^2\right]\mathrm{E}\left[\left(\frac{1}{n}\sum_{i=1}^n (X_i-\bar{X})^2-\sigma^2\right)^2\right] \quad (\text{ using (4.51)})\\
&= \mathrm{E}\left[(\bar{X}-\mu)^2\right]\frac{\sigma^4}{n^2}\mathrm{E}\left[\left(\sum_{i=1}^n\left(\frac{X_i-\bar{X}}{\sigma}\right)^2-n\right)^2\right]\\
&= \frac{\sigma^6}{n^3}\mathrm{E}\left[(G-(n-1)-1)^2\right]\\
&= \frac{\sigma^6}{n^3}\left(\mathrm{E}\left[(G-(n-1))^2\right]-2\mathrm{E}\left[G-(n-1)\right]+1\right) = \frac{\sigma^6}{n^3}\left(\mathrm{Var}\left[G\right]+1\right)\\
&= \frac{\sigma^6}{n^3}\left(2(n-1)+1\right) = \frac{\sigma^6}{n^2}\left(2-\frac{1}{n}\right) < 2\frac{\sigma^6}{n^2}.
\end{aligned}$$

For $C_{(m)}$ as in (4.5) and $[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}$ in (4.46), the second quantity in $K_1(\boldsymbol{\theta_0})$ becomes

$$
\begin{aligned}
&\frac{\|h\|_1}{2\sqrt{n}}\left\{\sum_{k=1}^{2}\sum_{l=1}^{2}\left|\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{lk}\right|\sum_{j=1}^{2}\sum_{i=1}^{2}\left[\mathrm{E}\left(\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\left(\hat{\theta}_n(\boldsymbol{X})_i-\theta_{0_i}\right)^2\right)\right]^{\frac{1}{2}}\right.\\
&\qquad\left.\times\left[\mathrm{E}\left(\left(nM_{kji}(\boldsymbol{X})\right)^2\middle||C_{(m)}|<\varepsilon\right)\right]^{\frac{1}{2}}\right\}\\
&=\frac{\|h\|_1}{2\sqrt{n}}\left\{\left|\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{11}\right|\sum_{j=1}^{2}\sum_{i=1}^{2}\left[\mathrm{E}\left(\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\left(\hat{\theta}_n(\boldsymbol{X})_i-\theta_{0_i}\right)^2\right)\right]^{\frac{1}{2}}\right.\\
&\qquad\times\left[\mathrm{E}\left(\left(nM_{1ji}(\boldsymbol{X})\right)^2\middle||C_{(m)}|<\varepsilon\right)\right]^{\frac{1}{2}}\\
&\qquad+\left|\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{22}\right|\sum_{j=1}^{2}\sum_{i=1}^{2}\left[\mathrm{E}\left(\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\left(\hat{\theta}_n(\boldsymbol{X})_i-\theta_{0_i}\right)^2\right)\right]^{\frac{1}{2}}\\
&\qquad\left.\times\left[\mathrm{E}\left(\left(nM_{2ji}(\boldsymbol{X})\right)^2\middle||C_{(m)}|<\varepsilon\right)\right]^{\frac{1}{2}}\right\}\\
&=\frac{\|h\|_1\sigma}{2\sqrt{n}}\left\{\frac{2n}{(\sigma^2-\varepsilon)^2}\left[\mathrm{E}\left((\bar{X}-\mu)^2\left(\hat{\sigma^2}-\sigma^2\right)^2\right)\right]^{\frac{1}{2}}\right.\\
&\qquad\left.+\left[\mathrm{E}\left[\left(\hat{\sigma^2}-\sigma^2\right)^4\right]\right]^{\frac{1}{2}}\frac{2n}{(\sigma^2-\varepsilon)^3}\left[\mathrm{E}\left((|\bar{X}-\mu|+\varepsilon)^2\middle||C_{(m)}|<\varepsilon\right)\right]^{\frac{1}{2}}\right\}\\
&\quad+\frac{\|h\|_1\sqrt{2}\sigma^2}{2\sqrt{n}}\left\{\frac{n}{(\sigma^2-\varepsilon)^2}\left[\mathrm{E}(\bar{X}-\mu)^4\right]^{\frac{1}{2}}\right.\\
&\qquad+\frac{4n}{(\sigma^2-\varepsilon)^3}\left[\mathrm{E}\left((\bar{X}-\mu)^2\left(\hat{\sigma^2}-\sigma^2\right)^2\right)\right]^{\frac{1}{2}}\left[\mathrm{E}\left((|\bar{X}-\mu|+\varepsilon)^2\middle||C_{(m)}|<\varepsilon\right)\right]^{\frac{1}{2}}\\
&\qquad\left.+n\left[\mathrm{E}\left[\left(\hat{\sigma^2}-\sigma^2\right)^4\right]\right]^{\frac{1}{2}}\left[\mathrm{E}\left(\left(M_{222}(\boldsymbol{X})\right)^2\middle||C_{(m)}|<\varepsilon\right)\right]^{\frac{1}{2}}\right\}\\
&\leq\frac{\|h\|_1\sigma}{2\sqrt{n}}\left\{\frac{2\sqrt{2}\sigma^3}{(\sigma^2-\varepsilon)^2}+\frac{8\sigma^4}{(\sigma^2-\varepsilon)^3}\left[\mathrm{E}\left((|\bar{X}-\mu|+\varepsilon)^2\middle||C_{(m)}|<\varepsilon\right)\right]^{\frac{1}{2}}\right\}\\
&\quad+\frac{\|h\|_1\sigma^2}{\sqrt{2n}}\left\{\frac{\sqrt{3}\sigma^2}{(\sigma^2-\varepsilon)^2}+4\sqrt{2}\frac{\sigma^3}{(\sigma^2-\varepsilon)^3}\left[\mathrm{E}\left((|\bar{X}-\mu|+\varepsilon)^2\middle||C_{(m)}|<\varepsilon\right)\right]^{\frac{1}{2}}\right.\\
&\qquad\left.+4\sigma^4\left[\mathrm{E}\left(\left(M_{222}(\boldsymbol{X})\right)^2\middle||C_{(m)}|<\varepsilon\right)\right]^{\frac{1}{2}}\right\}. \qquad (4.52)
\end{aligned}
$$

We next bound $\left[\mathrm{E}\left((M_{222}(\boldsymbol{X}))^2\middle||C_{(m)}|<\varepsilon\right)\right]^{\frac{1}{2}}$ and $\left[\mathrm{E}\left((|\bar{X}-\mu|+\varepsilon)^2\middle||C_{(m)}|<\varepsilon\right)\right]^{\frac{1}{2}}$. For $\left[\mathrm{E}\left((|\bar{X}-\mu|+\varepsilon)^2\middle||C_{(m)}|<\varepsilon\right)\right]^{\frac{1}{2}}$, we have

$$\begin{aligned}\left[\mathrm{E}\left((|\bar{X}-\mu|+\varepsilon)^2\middle||C_{(m)}|<\varepsilon\right)\right]^{\frac{1}{2}} &\leq \left[2\mathrm{E}\left((\bar{X}-\mu)^2\middle||C_{(m)}|<\varepsilon\right)+2\varepsilon^2\right]^{\frac{1}{2}} \\ &\leq \sqrt{2}\left[\mathrm{E}\left[(\bar{X}-\mu)^2\right]+\varepsilon^2\right]^{\frac{1}{2}} \quad \text{using Lemma 4.2} \\ &= \sqrt{2}\sqrt{\frac{\sigma^2}{n}+\varepsilon^2}. \end{aligned} \tag{4.53}$$

In addition, simple steps yield

$$\begin{aligned} M_{222}(\boldsymbol{X}) &= \frac{1}{(\sigma^2-\varepsilon)^3}+\frac{9}{(\sigma^2-\varepsilon)^4}\left(\hat{\sigma}^2+(\bar{X}-\mu)^2+\varepsilon^2\right) \\ &= \frac{1}{(\sigma^2-\varepsilon)^3}+\frac{9}{(\sigma^2-\varepsilon)^4}\left(\hat{\sigma}^2-\sigma^2+(\bar{X}-\mu)^2+\varepsilon^2+\sigma^2\right) \\ &\leq \frac{1}{(\sigma^2-\varepsilon)^3}+\frac{9}{(\sigma^2-\varepsilon)^4}\left(\left|\hat{\sigma}^2-\sigma^2\right|+(\bar{X}-\mu)^2+\varepsilon^2+\sigma^2\right), \end{aligned}$$

which leads to

$$\begin{aligned}[M_{222}(\boldsymbol{X})]^2 &\leq 2\left[\frac{1}{(\sigma^2-\varepsilon)^6}+\frac{81}{(\sigma^2-\varepsilon)^8}\left(\left|\hat{\sigma}^2-\sigma^2\right|+(\bar{X}-\mu)^2+\varepsilon^2+\sigma^2\right)^2\right] \\ &\leq 2\left[\frac{1}{(\sigma^2-\varepsilon)^6}+\frac{162}{(\sigma^2-\varepsilon)^8}\left[\left(\left|\hat{\sigma}^2-\sigma^2\right|+\varepsilon^2+\sigma^2\right)^2+(\bar{X}-\mu)^4\right]\right]. \end{aligned} \tag{4.54}$$

Using the result in (4.54) and Lemma 4.2 yields

$$\begin{aligned}&\left[\mathrm{E}\left((M_{222}(\boldsymbol{X}))^2\middle||C_{(m)}|<\varepsilon\right)\right]^{\frac{1}{2}} \\ &\leq \sqrt{2}\left[\frac{1}{(\sigma^2-\varepsilon)^6}+\frac{162}{(\sigma^2-\varepsilon)^8}\left[\left(\varepsilon+\varepsilon^2+\sigma^2\right)^2+3\frac{\sigma^4}{n^2}\right]\right]^{\frac{1}{2}}. \end{aligned} \tag{4.55}$$

Using (4.52),(4.53) and (4.55), the second term in (4.7) multiplied by $\frac{\|h\|_1}{\sqrt{n}}$ becomes

$$\begin{aligned}
&\frac{\|h\|_1}{2\sqrt{n}}\left\{\sum_{k=1}^{d}\sum_{l=1}^{d}\left|\left[I(\boldsymbol{\theta_0})^{-\frac{1}{2}}\right]_{lk}\right|\sum_{j=1}^{d}\sum_{i=1}^{d}\left[\mathrm{E}\left(\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\left(\hat{\theta}_n(\boldsymbol{X})_i-\theta_{0_i}\right)^2\right)\right]^{\frac{1}{2}}\right.\\
&\qquad\left.\times\left[\mathrm{E}\left(\left(nM_{kji}(\boldsymbol{X})\right)^2\middle||C_{(m)}|<\varepsilon\right)\right]^{\frac{1}{2}}\right\}\\
&\leq\frac{\|h\|_1\sigma^4}{\sqrt{n}\left(\sigma^2-\varepsilon\right)^2}\left[\sqrt{2}+\frac{4\sqrt{2}\sigma}{\sigma^2-\varepsilon}\sqrt{\frac{\sigma^2}{n}+\varepsilon^2}\right]\\
&\quad+\frac{\|h\|_1\sigma^4}{\sqrt{2n}(\sigma^2-\varepsilon)^2}\left[\sqrt{3}+\frac{8\sigma}{\sigma^2-\varepsilon}\sqrt{\frac{\sigma^2}{n}+\varepsilon^2}\right]\\
&\quad+\frac{4\|h\|_1\sigma^6}{\sqrt{n}\left(\sigma^2-\varepsilon\right)^3}\left[1+\frac{162}{\left(\sigma^2-\varepsilon\right)^2}\left[\left(\varepsilon+\varepsilon^2+\sigma^2\right)^2+3\frac{\sigma^4}{n^2}\right]\right]^{\frac{1}{2}}\\
&=\frac{\|h\|_1\sigma^4}{\sqrt{n}\left(\sigma^2-\varepsilon\right)^2}\left[\sqrt{2}+\sqrt{\frac{3}{2}}+\frac{4\sigma}{\sigma^2-\varepsilon}\sqrt{\frac{\sigma^2}{n}+\varepsilon^2}\left(\sqrt{2}+\frac{2}{\sqrt{2}}\right)\right]\\
&\quad+\frac{4\|h\|_1\sigma^6}{\sqrt{n}\left(\sigma^2-\varepsilon\right)^3}\left[1+\frac{162}{\left(\sigma^2-\varepsilon\right)^2}\left[\left(\varepsilon+\varepsilon^2+\sigma^2\right)^2+3\frac{\sigma^4}{n^2}\right]\right]^{\frac{1}{2}}\\
&\leq\frac{\|h\|_1\sigma^4}{\sqrt{n}\left(\sigma^2-\varepsilon\right)^2}\left[3+\frac{8\sqrt{2}\sigma}{\sigma^2-\varepsilon}\sqrt{\frac{\sigma^2}{n}+\varepsilon^2}\right]\\
&\quad+\frac{4\|h\|_1\sigma^6}{\sqrt{n}\left(\sigma^2-\varepsilon\right)^3}\left[1+\frac{162}{\left(\sigma^2-\varepsilon\right)^2}\left[\left(\varepsilon+\varepsilon^2+\sigma^2\right)^2+3\frac{\sigma^4}{n^2}\right]\right]^{\frac{1}{2}}. \qquad (4.56)
\end{aligned}$$

Taking $\varepsilon=\frac{\sigma^2}{2}$ yields

$$(4.56)=\frac{4\|h\|_1}{\sqrt{n}}\left[3+16\sqrt{2}\sqrt{\frac{1}{n}+\frac{\sigma^2}{4}}\right]+\frac{32\|h\|_1}{\sqrt{n}}\left[1+648\left[\left(\frac{3}{2}+\frac{\sigma^2}{4}\right)^2+\frac{3}{n^2}\right]\right]^{\frac{1}{2}}. \qquad (4.57)$$

Combining the results of (4.47) and (4.57) yields to a bound on the first term, $\frac{\|h\|_1}{\sqrt{n}}K_1(\boldsymbol{\theta_0})$, of the general upper bound in (4.6). The second term, $\frac{\|h\|_2}{\sqrt{n}}K_2(\boldsymbol{\theta_0})$, is a sum of two quantities as (4.8) shows. For the first quantity,

$$\begin{aligned}
&\frac{\|h\|_2}{4\sqrt{n}}\sum_{j=1}^{d}\left[\mathrm{Var}\left(\left(\sum_{k=1}^{d}\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{jk}\frac{\partial}{\partial\theta_k}\log f(X_1|\boldsymbol{\theta_0})\right)^2\right)\right]^{\frac{1}{2}}\\
&=\frac{\|h\|_2}{4\sqrt{n}}\left\{\left[\mathrm{Var}\left(\frac{(X_1-\mu)^2}{\sigma^2}\right)\right]^{\frac{1}{2}}+\left[\mathrm{Var}\left(\frac{(X_1-\mu)^4}{2\sigma^4}-\frac{(X_1-\mu)^2}{\sigma^2}\right)\right]^{\frac{1}{2}}\right\}\\
&=\frac{\|h\|_2}{4\sqrt{n}}\left\{\sqrt{2}+\left[\frac{1}{4}\mathrm{Var}\left[\left(\frac{X_1-\mu}{\sigma}\right)^4\right]+\mathrm{Var}\left[\left(\frac{X_1-\mu}{\sigma}\right)^2\right]\right.\right.\\
&\qquad\left.\left.-\mathrm{Cov}\left(\left(\frac{X_1-\mu}{\sigma}\right)^4,\left(\frac{X_1-\mu}{\sigma}\right)^2\right)\right]^{\frac{1}{2}}\right\}. \qquad (4.58)
\end{aligned}$$

Since $\frac{X_i-\mu}{\sigma}\sim \mathrm{N}(0,1)$, simple steps yield

$$\begin{aligned}
\mathrm{Var}\left[\left(\frac{X_1-\mu}{\sigma}\right)^4\right]&=\mathrm{E}\left[\left(\frac{X_1-\mu}{\sigma}\right)^8\right]-\left[\mathrm{E}\left[\left(\frac{X_1-\mu}{\sigma}\right)^4\right]\right]^2=96\\
\mathrm{Var}\left[\left(\frac{X_1-\mu}{\sigma}\right)^2\right]&=\mathrm{E}\left[\left(\frac{X_1-\mu}{\sigma}\right)^4\right]-\left[\mathrm{E}\left[\left(\frac{X_1-\mu}{\sigma}\right)^2\right]\right]^2=2\\
\mathrm{Cov}\left(\left(\frac{X_1-\mu}{\sigma}\right)^4,\left(\frac{X_1-\mu}{\sigma}\right)^2\right)&=\mathrm{E}\left[\left(\frac{X_1-\mu}{\sigma}\right)^6\right]\\
&\quad-\mathrm{E}\left[\left(\frac{X_1-\mu}{\sigma}\right)^4\right]\mathrm{E}\left[\left(\frac{X_1-\mu}{\sigma}\right)^2\right]=12.
\end{aligned}$$

Applying the above results to (4.58),

$$\frac{\|h\|_2}{4\sqrt{n}}\sum_{j=1}^{d}\left[\mathrm{Var}\left(\left(\sum_{k=1}^{d}\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{jk}\frac{\partial}{\partial\theta_k}\log f(X_1|\boldsymbol{\theta_0})\right)^2\right)\right]^{\frac{1}{2}}=\frac{\|h\|_2}{4\sqrt{n}}\left(\sqrt{2}+\sqrt{14}\right). \qquad (4.59)$$

The second quantity in $\frac{\|h\|_2}{\sqrt{n}}K_2(\boldsymbol{\theta_0})$ is also easily calculated. Since $[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}$ is diagonal, we obtain that

$$\frac{\|h\|_2}{2\sqrt{n}} \sum_{j>k}^{d} \sum_{k=1}^{d-1} \left[\mathrm{Var} \left(\sum_{q=1}^{d} \sum_{v=1}^{d} \left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}} \right]_{jq} \frac{\partial}{\partial \theta_q} \log f(X_1|\boldsymbol{\theta_0}) \right.\right.$$
$$\left.\left. \times \left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}} \right]_{kv} \frac{\partial}{\partial \theta_v} \log f(X_1|\boldsymbol{\theta_0}) \right) \right]^{\frac{1}{2}}$$
$$= \frac{\|h\|_2}{2\sqrt{n}} \left[\mathrm{Var} \left(\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}} \right]_{11} \frac{\partial}{\partial \mu} \log f(X_1|\boldsymbol{\theta_0}) \left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}} \right]_{22} \frac{\partial}{\partial \sigma^2} \log f(X_1|\boldsymbol{\theta_0}) \right) \right]^{\frac{1}{2}}. \tag{4.60}$$

Using (4.45) for one random variable and (4.46) yields

$$(4.60) = \frac{\|h\|_2}{2\sqrt{n}} \left[\mathrm{Var} \left(\left(\frac{X_1-\mu}{\sigma} \right) \left(\frac{(X_1-\mu)^2}{\sqrt{2}\sigma^2} - \frac{1}{\sqrt{2}} \right) \right) \right]^{\frac{1}{2}}$$
$$= \frac{\|h\|_2}{2\sqrt{2n}} \left[\mathrm{Var} \left(\left(\frac{X_1-\mu}{\sigma} \right)^3 - \left(\frac{X_1-\mu}{\sigma} \right) \right) \right]^{\frac{1}{2}}$$
$$= \frac{\|h\|_2}{2\sqrt{2n}} \left[\mathrm{Var} \left(\left(\frac{X_1-\mu}{\sigma} \right)^3 \right) + \mathrm{Var} \left(\frac{X_1-\mu}{\sigma} \right) - 2\mathrm{Cov} \left(\left(\frac{X_1-\mu}{\sigma} \right)^3, \frac{X_1-\mu}{\sigma} \right) \right]^{\frac{1}{2}}$$
$$= \frac{\|h\|_2}{2\sqrt{2n}} \left[\mathrm{E} \left[\left(\frac{X_1-\mu}{\sigma} \right)^6 \right] + 1 - 2\mathrm{E} \left[\left(\frac{X_1-\mu}{\sigma} \right)^4 \right] \right]^{\frac{1}{2}}$$
$$= \frac{\|h\|_2}{2\sqrt{2n}} \sqrt{10} = \frac{\sqrt{5}\|h\|_2}{2\sqrt{n}}. \tag{4.61}$$

For an upper bound for the third term $\frac{\|h\|_3}{\sqrt{n}} K_3(\boldsymbol{\theta_0})$ in (4.6), with $K_3(\boldsymbol{\theta_0})$ as in (4.9), we use that X_1' is an independent copy of X_1. The triangle inequality and (4.42) give that

$$\frac{\|h\|_3}{12\sqrt{n}} \mathrm{E} \left(\sum_{i=1}^{d} \left| \sum_{l=1}^{d} \left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}} \right]_{il} \left(\frac{\partial}{\partial \theta_l} \log f(X_1'|\boldsymbol{\theta_0}) - \frac{\partial}{\partial \theta_l} \log f(X_1|\boldsymbol{\theta_0}) \right) \right| \right)^3$$
$$= \frac{\|h\|_3}{12\sqrt{n}} \mathrm{E} \left(\left| \frac{X_1'-\mu}{\sigma} - \frac{X_1-\mu}{\sigma} \right| + \frac{1}{\sqrt{2}\sigma^2} \left| (X_1'-\mu)^2 - (X_1-\mu)^2 \right| \right)^3$$
$$\leq \frac{\|h\|_3}{3\sqrt{n}} \left[\mathrm{E} \left| \frac{X_1'-X_1}{\sigma} \right|^3 + \mathrm{E} \left| \frac{(X_1'-\mu)^2-(X_1-\mu)^2}{\sqrt{2}\sigma^2} \right|^3 \right]$$
$$\leq \frac{8\|h\|_3}{3\sqrt{n}} \left[\frac{\mathrm{E}|X_1|^3}{\sigma^3} + \frac{\mathrm{E}(X_1-\mu)^6}{2\sqrt{2}\sigma^6} \right] = \frac{8\|h\|_3}{3\sqrt{n}} \left[\frac{2\sqrt{2}}{\sqrt{\pi}} + \frac{15}{2\sqrt{2}} \right] \leq \frac{19}{\sqrt{n}} \|h\|_3. \tag{4.62}$$

For the last term of (4.6), $\mathrm{E}\left(\bar{X}-\mu\right)^2=\frac{\sigma^2}{n}$ and $\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2-\sigma^2\right)^2=\frac{\sigma^4}{n}\left(2-\frac{1}{n}\right)$. For the choice of $\varepsilon=\varepsilon_0$ as in (R.C.3), (4.48), (4.49) and (4.50) require that $0<\varepsilon<\sigma^2$. Again, due to a trade-off between the first and the last term of the bound in (4.6), we choose $\varepsilon=\frac{\sigma^2}{2}$ which gives that

$$\begin{aligned}\frac{2\|h\|}{\varepsilon^2}\left(\sum_{j=1}^{d}\mathrm{E}\left[\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})_j-\boldsymbol{\theta}_{0_j}\right)^2\right]\right)&=\frac{8\|h\|}{\sigma^4}\left(\frac{\sigma^2}{n}+\frac{\sigma^4}{n}\left(2-\frac{1}{n}\right)\right)\\&=\frac{8\|h\|}{n\sigma^2}\left(1+\sigma^2\left(2-\frac{1}{n}\right)\right)\\&\leq\frac{8\|h\|}{n\sigma^2}\left(1+2\sigma^2\right).\end{aligned}\tag{4.63}$$

Using the results in (4.47), (4.57), (4.59), (4.61), (4.62) and (4.63) we get the result of the Corollary. □

Remark 4.3. ***(1)*** *The rate of convergence of the upper bound in* (4.44) *is* $\frac{1}{\sqrt{n}}$.

(2) *There might be cases where the parameters depend on the sample size, so that* $\mu=\mu(n)$ *and* $\sigma^2=\sigma^2(n)$. *The bound in* (4.44) *does not depend on* $\mu(n)$ *and goes to zero as long as*

(i) $\frac{1}{n\sigma^2(n)}\xrightarrow[n\to\infty]{}0$,

(ii) $\frac{\sigma^2(n)}{\sqrt{n}}\xrightarrow[n\to\infty]{}0$,

are both satisfied. From (i), the order of $\sigma^2(n)$ *should not be less than or equal to* $\frac{1}{n}$, *while from (ii) we see that* $\sigma^2(n)$ *should be of order smaller than* $\sqrt{n}$. *For instance,* $\sigma^2(n)=cn^{\frac{1}{4}}$, *where* $c\in\mathbb{R}$ *is a constant, satisfies the above limits. The bound in* (4.44) *is then of order* $\frac{1}{n^{\frac{1}{4}}}$.

4.2.2 Bounds when the MLE is not known explicitly

In the single-parameter case, it was possible to find an upper bound for the mean squared error and then use it to get upper bounds on the distributional distance of interest which can be applied when the MLE is not expressed in a closed-form. In this subsection, we give similar bounds for the multi-parameter case with multivariate (dimensionality $t \in \mathbb{N}$) i.i.d. random vectors $\boldsymbol{X_1}, \boldsymbol{X_2}, \ldots, \boldsymbol{X_n}$, with $\boldsymbol{X_i} = (X_{i1}, X_{i2}, \ldots, X_{it})^\intercal$. We make some extra assumptions,

(Con.1) $\forall j \in \{1,2,\ldots,t\}$, the support S_j of X_{ij} is a bounded interval in $\mathbb{R}$; let $s_j := \sup_{x_{ij} \in S_j} \{|x_{ij}|\}$ and $s := \max\{s_1, s_2, \ldots, s_t\}$;

(Con.2) for all $\boldsymbol{\theta_0} \in \Theta$, where Θ is the open parameter space, there exists an $\varepsilon_0 = \varepsilon_0(\boldsymbol{\theta_0}) > 0$ such that for all $\boldsymbol{\theta} \in \Theta$ with $\left|\theta_j - \theta_{0_j}\right| < \varepsilon_0, \forall j = 1,2,\ldots,d$

$$\sup_{\substack{\boldsymbol{\theta}: \left|\theta_q - \theta_{0_q}\right| < \varepsilon_0 \\ \forall q \in \{1,2,\ldots,d\}}} \left| \frac{\partial^3}{\partial\theta_k \partial\theta_j \partial\theta_i} \log f(\boldsymbol{x_1}|\boldsymbol{\theta}) \right| \leq M_{kji},$$

where $M_{kji} = M_{kji}(\boldsymbol{\theta_0})$ is a constant that may depend only on the unknown parameter $\boldsymbol{\theta_0}$;

(Con.3) for $M = \sup_{i,j} \left\{ \left| \left[[I(\boldsymbol{\theta_0})]^{-1} \right]_{ij} \right| \right\}$, the sample size satisfies

$$n > \frac{s^2 d^2}{4\varepsilon_0^2} \left(M\varepsilon_0 \sum_{l=1}^{d} \sum_{k=1}^{d} \left| \left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}} \right]_{lk} \right| \sum_{m=1}^{d} \sum_{i=1}^{d} M_{kim} + \left[M^2 \varepsilon_0^2 \left(\sum_{l=1}^{d} \sum_{k=1}^{d} \left| \left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}} \right]_{lk} \right| \sum_{m=1}^{d} \sum_{i=1}^{d} M_{kim} \right)^2 + 8M \right]^{\frac{1}{2}} \right)^2.$$

If (Con.1)-(Con.3) hold, then

$$2d^2 s^2 M + dsM\sqrt{n}\varepsilon_0^2 \sum_{l=1}^{d} \sum_{k=1}^{d} \left| \left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}} \right]_{lk} \right| \sum_{m=1}^{d} \sum_{i=1}^{d} M_{kim} - n\varepsilon_0^2 < 0$$

holds with ε_0 as in (Con.2). The upper bound in (4.6) can be split into terms coming from Lemma 4.1, and terms due to Taylor expansions and conditional expectations. For ease of presentation, since the terms coming from Lemma 4.1 are not related to $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})$, from now on, unless otherwise stated, we abbreviate

$$
\begin{aligned}
D := D(\boldsymbol{\theta_0}, h, \boldsymbol{X}) := {} & \frac{\|h\|_2}{4\sqrt{n}} \sum_{j=1}^{d} \left[\mathrm{Var} \left(\left(\sum_{k=1}^{d} \left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}} \right]_{jk} \frac{\partial}{\partial \theta_k} \log f(\boldsymbol{X_1}|\boldsymbol{\theta_0}) \right)^2 \right) \right]^{\frac{1}{2}} \\
& + \frac{\|h\|_2}{2\sqrt{n}} \sum_{k=1}^{d-1} \sum_{j>k}^{d} \left[\mathrm{Var} \left(\sum_{q=1}^{d} \sum_{v=1}^{d} \left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}} \right]_{jq} \frac{\partial}{\partial \theta_q} \log f(\boldsymbol{X_1}|\boldsymbol{\theta_0}) \right. \right. \\
& \qquad \left. \left. \times \left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}} \right]_{kv} \frac{\partial}{\partial \theta_v} \log f(\boldsymbol{X_1}|\boldsymbol{\theta_0}) \right) \right]^{\frac{1}{2}} \\
& + \frac{\|h\|_3}{12\sqrt{n}} \mathrm{E} \left(\sum_{i=1}^{d} \left| \sum_{l=1}^{d} \left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}} \right]_{il} \left(\frac{\partial}{\partial \theta_l} \log f(\boldsymbol{X_1'}|\boldsymbol{\theta_0}) - \frac{\partial}{\partial \theta_l} \log f(\boldsymbol{X_1}|\boldsymbol{\theta_0}) \right) \right| \right)^3 .
\end{aligned}
\tag{4.64}
$$

In order to give an upper bound when $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})$ is not known explicitly, we bound $\sqrt{\mathrm{E}\left[\sum_{j=1}^{d} \left(\hat{\theta}_n(\boldsymbol{X})_j - \theta_{0_j}\right)^2\right]}$ by a quantity which does not require knowledge of the MLE. The result is given in the following theorem.

Theorem 4.3. *Let $\boldsymbol{X_1}, \boldsymbol{X_2}, \ldots, \boldsymbol{X_n}$ be i.i.d. $\mathbb{R}^t$-valued random elements, for $t \in \mathbb{N}$, with probability density (or mass) function $f(\boldsymbol{x_i}|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the d-valued vector parameter. Assume that (R.C.1)-(R.C.4) and also (Con.1)-(Con.3) are satisfied. Also $\boldsymbol{Z} \sim \mathrm{N}_d(\mathbf{0}, I_{d\times d})$ and we assume existence and uniqueness of the MLE, $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})$. For $\varepsilon = \varepsilon_0$ as in (Con.2) and using the notation*

$$\gamma = \sum_{j=1}^{d} \left| \left[[I(\boldsymbol{\theta_0})]^{-1} \right]_{jj} \right| + \frac{M}{2\sqrt{n}} \sum_{j=1}^{d} \sqrt{\operatorname{Var}\left(\left(\sum_{k=1}^{d} \left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}} \right]_{jk} \frac{\partial}{\partial \theta_k} \log f(\boldsymbol{X_1}|\boldsymbol{\theta_0}) \right)^2 \right)}$$

$$+ \frac{M}{\sqrt{n}} \sum_{k=1}^{d-1} \sum_{j>k}^{d} \left[\operatorname{Var}\left(\sum_{q=1}^{d} \sum_{v=1}^{d} \left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}} \right]_{jq} \left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}} \right]_{kv} \right.\right.$$

$$\left.\left. \times \frac{\partial}{\partial \theta_q} \log f(\boldsymbol{X_1}|\boldsymbol{\theta_0}) \frac{\partial}{\partial \theta_v} \log f(\boldsymbol{X_1}|\boldsymbol{\theta_0}) \right) \right]^{\frac{1}{2}}$$

$$\omega = 1 - 2\frac{d^2 s^2 M}{n\varepsilon^2} - \frac{dsM}{\sqrt{n}} \sum_{l=1}^{d} \sum_{k=1}^{d} \left| \left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}} \right]_{lk} \right| \sum_{m=1}^{d} \sum_{i=1}^{d} M_{kmi}$$

$$v = 2d^{\frac{3}{2}} sM \sum_{l=1}^{d} \sum_{k=1}^{d} \left| \left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}} \right]_{lk} \right| \sqrt{\sum_{i=1}^{d} \operatorname{Var}\left(\frac{\partial^2}{\partial \theta_k \partial \theta_i} \log f(\boldsymbol{X_1}|\boldsymbol{\theta_0}) \right)} \tag{4.65}$$

it holds that

$$\sqrt{\mathrm{E}\left[\sum_{j=1}^{d} \left(\hat{\theta}_n(\boldsymbol{X})_j - \theta_{0_j} \right)^2 \right]} \leq \frac{1}{\sqrt{n}} \left(\frac{\frac{v}{\sqrt{n}} + \sqrt{\frac{v^2}{n} + 4\omega\gamma}}{2\omega} \right). \tag{4.66}$$

Remark 4.4. *The quantities γ, ω and v are $\mathcal{O}(1)$.*

Proof. Using (4.6) and (4.64) gives that for $C_{(m)}$ as in (4.5),

$$\left| \mathrm{E}\left[h\left(\sqrt{n} [I(\boldsymbol{\theta_0})]^{\frac{1}{2}} (\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) - \boldsymbol{\theta_0}) \right) \right] - \mathrm{E}[h(\boldsymbol{Z})] \right| \leq D + \frac{2\|h\|}{\varepsilon^2} \mathrm{E}\left[\sum_{j=1}^{d} \left(\hat{\theta}_n(\boldsymbol{X})_j - \theta_{0_j} \right)^2 \right]$$

$$+ \frac{\|h\|_1}{\sqrt{n}} \sum_{k=1}^{d} \sum_{l=1}^{d} \left| \left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}} \right]_{lk} \right|$$

$$\times \sum_{j=1}^{d} \left[\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})_j - \theta_{0_j} \right)^2 \right] \mathrm{E}\left[\left(\frac{\partial^2}{\partial \theta_j \partial \theta_k} l(\boldsymbol{\theta_0}; \boldsymbol{X}) + n[I(\boldsymbol{\theta_0})]_{kj} \right)^2 \right] \right]^{\frac{1}{2}} \tag{4.67}$$

$$+ \frac{\|h\|_1}{2\sqrt{n}} \left\{ \sum_{k=1}^{d} \sum_{l=1}^{d} \left| \left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}} \right]_{lk} \right| \right.$$

$$\left. \times \mathrm{E}\left(\left| \sum_{j=1}^{d} \sum_{i=1}^{d} \left(\hat{\theta}_n(\boldsymbol{X})_j - \theta_{0_j} \right) \left(\hat{\theta}_n(X)_i - \theta_{0_i} \right) \frac{\partial^3}{\partial \theta_k \partial \theta_j \partial \theta_i} l(\boldsymbol{\theta_0^*}; \boldsymbol{X}) \right| \middle| |C_{(m)}| < \varepsilon \right) \right\}. \tag{4.68}$$

Step 1: Upper bound for (4.67). Using that $\mathrm{E}\left(\frac{\partial^2}{\partial\theta_j\partial\theta_k}l(\boldsymbol{\theta_0};\boldsymbol{X})+n[I(\boldsymbol{\theta_0})]_{kj}\right)=0$, $\forall j,k\in\{1,2,\ldots,d\}$,

$$\begin{aligned}(4.67) &= \|h\|_1\sum_{k=1}^{d}\sum_{l=1}^{d}\left|\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{lk}\right| \\ &\quad\times\sum_{j=1}^{d}\sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\right]}\sqrt{\mathrm{Var}\left(\frac{\partial^2}{\partial\theta_k\partial\theta_j}\log f(\boldsymbol{X_1}|\boldsymbol{\theta_0})\right)} \\ &\leq \|h\|_1\sum_{k=1}^{d}\sum_{l=1}^{d}\left|\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{lk}\right| \\ &\quad\times\sum_{j=1}^{d}\sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\right]}\sqrt{\sum_{i=1}^{d}\mathrm{Var}\left(\frac{\partial^2}{\partial\theta_k\partial\theta_i}\log f(\boldsymbol{X_1}|\boldsymbol{\theta_0})\right)}, \qquad (4.69)\end{aligned}$$

where for the last inequality the trivial bound

$$\mathrm{Var}\left(\frac{\partial^2}{\partial\theta_k\partial\theta_j}\log f(\boldsymbol{X_1}|\boldsymbol{\theta_0})\right)\leq\sum_{i=1}^{d}\mathrm{Var}\left(\frac{\partial^2}{\partial\theta_k\partial\theta_i}\log f(\boldsymbol{X_1}|\boldsymbol{\theta_0})\right)$$

was used, since the variance of a random variable is always non-negative. Now, using that $\left(\sum_j^d\alpha_j\right)^2\leq d\left(\sum_{j=1}^d\alpha_j^2\right)$ for $\alpha_j\in\mathbb{R}$, yields

$$\left(\sum_{j=1}^{d}\sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\right]}\right)^2\leq d\sum_{j=1}^{d}\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\right].$$

Taking square roots in both sides of the above inequality and applying this result to (4.69) gives

$$\begin{aligned}(4.67) &\leq \|h\|_1\sqrt{d}\sum_{k=1}^{d}\sum_{l=1}^{d}\left|\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{lk}\right| \\ &\quad\times\sqrt{\sum_{i=1}^{d}\mathrm{Var}\left(\frac{\partial^2}{\partial\theta_k\partial\theta_i}\log f(\boldsymbol{X_1}|\boldsymbol{\theta_0})\right)}\sqrt{\mathrm{E}\left[\sum_{j=1}^{d}\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\right]}. \qquad (4.70)\end{aligned}$$

Step 2: Upper bound for (4.68). Notice that from (Con.2), if $\left|\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right|<\varepsilon$, $\forall j\in\{1,2,\ldots,d\}$, then $\left|\frac{\partial^3}{\partial\theta_k\partial\theta_j\partial\theta_i}l(\boldsymbol{\theta_0^*};\boldsymbol{x})\right|=\left|\sum_{l=1}^n\frac{\partial^3}{\partial\theta_k\partial\theta_j\partial\theta_i}\log f(\boldsymbol{x_l}|\boldsymbol{\theta_0^*})\right|\leq nM_{kji}$. In

addition,

$$\sum_{j=1}^{d}\sum_{i=1}^{d}\left|\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)\left(\hat{\theta}_n(\boldsymbol{X})_i-\theta_{0_i}\right)\right|M_{kji}$$
$$=\sum_{j=1}^{d}\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2M_{kjj}+2\sum_{j>i}^{d}\sum_{i=1}^{d-1}\left|\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right|\left|\hat{\theta}_n(\boldsymbol{X})_i-\theta_{0_i}\right|M_{kij}.$$

Using now that $2\alpha\beta\leq\alpha^2+\beta^2,\forall\alpha,\beta\in\mathbb{R}$,

$$\begin{aligned}
&\sum_{j=1}^{d}\sum_{i=1}^{d}\left|\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)\left(\hat{\theta}_n(\boldsymbol{X})_i-\theta_{0_i}\right)\right|M_{kji}\\
&\leq\sum_{j=1}^{d}\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2M_{kjj}+\sum_{j>i}^{d}\sum_{i=1}^{d-1}\left[\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2+\left(\hat{\theta}_n(\boldsymbol{X})_i-\theta_{0_i}\right)^2\right]M_{kji}\\
&=\sum_{j=1}^{d}\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\sum_{i=1}^{d}M_{kji}\\
&\leq\sum_{j=1}^{d}\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\sum_{m=1}^{d}\sum_{i=1}^{d}M_{kmi}. \qquad (4.71)
\end{aligned}$$

Using (4.71) yields

$$\begin{aligned}
(4.68)&\leq\frac{\|h\|_1}{2\sqrt{n}}\sum_{k=1}^{d}\sum_{l=1}^{d}\left|\left[I(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}}\right]_{lk}\Big|\\
&\quad\times\mathrm{E}\left(n\sum_{j=1}^{d}\sum_{i=1}^{d}M_{kij}\left|\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right|\left|\hat{\theta}_n(\boldsymbol{X})_i-\theta_{0_i}\right|\middle|\left|C_{(m)}\right|<\varepsilon\right)\\
&\leq\frac{\|h\|_1\sqrt{n}}{2}\sum_{k=1}^{d}\sum_{l=1}^{d}\left|\left[I(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}}\right]_{lk}\Big|\sum_{m=1}^{d}\sum_{i=1}^{d}M_{kmi}\mathrm{E}\left[\sum_{j=1}^{d}\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\right]. \qquad (4.72)
\end{aligned}$$

Hence, from (4.70) and (4.72),

$$\left|\mathrm{E}\left[h\left(\sqrt{n}[I(\boldsymbol{\theta_0})]^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})-\boldsymbol{\theta_0})\right)\right]-\mathrm{E}[h(\boldsymbol{Z})]\right| \leq D+\frac{2\|h\|}{\varepsilon^2}\mathrm{E}\left[\sum_{j=1}^{d}\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\right]$$
$$+\|h\|_1\sqrt{d}\sum_{k=1}^{d}\sum_{l=1}^{d}\left|\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{lk}\right|$$
$$\times\sqrt{\sum_{i=1}^{d}\mathrm{Var}\left(\frac{\partial^2}{\partial\theta_k\partial\theta_i}\log f(\boldsymbol{X_1}|\boldsymbol{\theta_0})\right)}\sqrt{\mathrm{E}\left[\sum_{j=1}^{d}\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\right]}$$
$$+\frac{\|h\|_1\sqrt{n}}{2}\sum_{k=1}^{d}\sum_{l=1}^{d}\left|\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{lk}\right|\sum_{m=1}^{d}\sum_{i=1}^{d}M_{kmi}\mathrm{E}\left[\sum_{j=1}^{d}\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\right]. \tag{4.73}$$

Since D as defined in (4.64), is not related to the MLE, the upper bound in (4.73) depends on $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})$ only through $\mathrm{E}\left[\sum_{j=1}^{d}\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\right]$. Our purpose is to find a bound for $\mathrm{E}\left[\sum_{j=1}^{d}\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\right]$ that does not depend on $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})$.

Step 3: The MSE test function. To this purpose define the test function

$$h=h_{\theta_0}:\mathbb{R}^d\to\mathbb{R}:h(\boldsymbol{x})=\boldsymbol{x}^\intercal[I(\boldsymbol{\theta_0})]^{-1}\boldsymbol{x}. \tag{4.74}$$

Then,

$$\begin{aligned}
&h\left(\sqrt{n}[I(\boldsymbol{\theta_0})]^{\frac{1}{2}}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{x})-\boldsymbol{\theta_0}\right)\right)\\
&=\left[\sqrt{n}[I(\boldsymbol{\theta_0})]^{\frac{1}{2}}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{x})-\boldsymbol{\theta_0}\right)\right]^\intercal[I(\boldsymbol{\theta_0})]^{-1}\sqrt{n}[I(\boldsymbol{\theta_0})]^{\frac{1}{2}}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{x})-\boldsymbol{\theta_0}\right)\\
&=n\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})-\boldsymbol{\theta_0}\right)^\intercal\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})-\boldsymbol{\theta_0}\right),
\end{aligned}$$

since $\left([I(\boldsymbol{\theta_0})]^{\frac{1}{2}}\right)^\intercal=[I(\boldsymbol{\theta_0})]^{\frac{1}{2}}$ as $[I(\boldsymbol{\theta_0})]^{\frac{1}{2}}$ is symmetric. Thus,

$$h\left(\sqrt{n}[I(\boldsymbol{\theta_0})]^{\frac{1}{2}}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{x})-\boldsymbol{\theta_0}\right)\right)=n\sum_{j=1}^{d}\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2.$$

Hence,

$$\begin{aligned}\mathrm{E}[h(\boldsymbol{Z})] &= \sum_{j=1}^{d}\left[[I(\boldsymbol{\theta_0})]^{-1}\right]_{jj}\mathrm{E}\left[Z_j^2\right] + 2\sum_{\substack{k=1\\k<i}}^{d}\sum_{i=1}^{d}\left[[I(\boldsymbol{\theta_0})]^{-1}\right]_{ki}\mathrm{E}[Z_k]\mathrm{E}[Z_i] \\ &= \sum_{j=1}^{d}\left[[I(\boldsymbol{\theta_0})]^{-1}\right]_{jj}.\end{aligned}$$

Denoting with B_h the bound in (4.73) for the test function h as in (4.74),

$$\begin{aligned}&\mathrm{E}\left[\sum_{j=1}^{d}\left(\hat{\theta}_n(\boldsymbol{X})_j - \theta_{0_j}\right)^2\right] \\ &= \frac{1}{n}\left|\mathrm{E}\left[h\left(\sqrt{n}[I(\boldsymbol{\theta_0})]^{\frac{1}{2}}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X}) - \boldsymbol{\theta_0}\right)\right)\right] - \mathrm{E}[h(\boldsymbol{Z})] + \mathrm{E}[h(\boldsymbol{Z})]\right| \\ &\leq \frac{1}{n}\left(B_h + \sum_{j=1}^{d}\left|\left[[I(\boldsymbol{\theta_0})]^{-1}\right]_{jj}\right|\right). \qquad (4.75)\end{aligned}$$

For the calculation of B_h, note that with s as in (Con.1) and M as in (Con.3),

$$\begin{aligned}&\|h\| = d^2s^2M, && \|h\|_1 = \sup_j\left|\frac{\partial}{\partial x_j}h(\boldsymbol{x})\right| = 2dMs, \\ &\|h\|_2 = \sup_{i,j}\left|\frac{\partial^2}{\partial x_i\partial x_j}h(\boldsymbol{x})\right| = 2M, && \|h\|_3 = \sup_{i,j,k}\left|\frac{\partial^3}{\partial x_i\partial x_j\partial x_k}h(\boldsymbol{x})\right| = 0.\end{aligned}$$

Using (4.64) and (4.73) yields

$$\begin{aligned}B_h &= \frac{M}{2\sqrt{n}}\sum_{j=1}^{d}\left[\mathrm{Var}\left(\left(\sum_{k=1}^{d}\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{jk}\frac{\partial}{\partial\theta_k}\log f(\boldsymbol{X_1}|\boldsymbol{\theta_0})\right)^2\right)\right]^{\frac{1}{2}} \\ &+ \frac{M}{\sqrt{n}}\sum_{k=1}^{d-1}\sum_{j>k}^{d}\left[\mathrm{Var}\left(\sum_{q=1}^{d}\sum_{v=1}^{d}\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{jq}\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{kv}\right.\right. \\ &\qquad\left.\left.\times\frac{\partial}{\partial\theta_q}\log f(\boldsymbol{X_1}|\boldsymbol{\theta_0})\frac{\partial}{\partial\theta_v}\log f(\boldsymbol{X_1}|\boldsymbol{\theta_0})\right)\right]^{\frac{1}{2}} \\ &+ 2\frac{d^2s^2M}{\varepsilon^2}\mathrm{E}\left(\sum_{j=1}^{d}\left(\hat{\theta}_n(\boldsymbol{X})_j - \theta_{0_j}\right)^2\right) \\ &+ \sqrt{n}dsM\sum_{l=1}^{d}\sum_{k=1}^{d}\left|\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{lk}\right|\sum_{m=1}^{d}\sum_{i=1}^{d}M_{kmi}\mathrm{E}\left(\sum_{j=1}^{d}\left(\hat{\theta}_n(\boldsymbol{X})_j - \theta_{0_j}\right)^2\right)\end{aligned}$$

$$+2d^{\frac{3}{2}}sM\sum_{l=1}^{d}\sum_{k=1}^{d}\left|\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{lk}\right| \times\sqrt{\sum_{i=1}^{d}\mathrm{Var}\left(\frac{\partial^2}{\partial\theta_k\partial\theta_i}\log f(\boldsymbol{X_1}|\boldsymbol{\theta_0})\right)}\sqrt{\mathrm{E}\left(\sum_{j=1}^{d}\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\right)}. \tag{4.76}$$

Let $U:=\sqrt{\mathrm{E}\left[\sum_{j=1}^{d}\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\right]}$. The results in (4.75) and (4.76) give

$$0\leq U^2\left(2\frac{d^2s^2M}{n\varepsilon^2}+\frac{dsM}{\sqrt{n}}\sum_{l=1}^{d}\sum_{k=1}^{d}\left|\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{lk}\right|\sum_{m=1}^{d}\sum_{i=1}^{d}M_{kmi}-1\right) +U\left(\frac{2}{n}d^{\frac{3}{2}}sM\sum_{l=1}^{d}\sum_{k=1}^{d}\left|\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{lk}\right|\sqrt{\sum_{i=1}^{d}\mathrm{Var}\left(\frac{\partial^2}{\partial\theta_k\partial\theta_i}\log f(\boldsymbol{X_1}|\boldsymbol{\theta_0})\right)}\right)+\frac{\gamma}{n}, \tag{4.77}$$

with γ as in (4.65). Solving the quadratic inequality in (4.77) (with unknown U) and using (Con.3) related to the sample size, n, we obtain for v and ω as in (4.65) that

$$U\leq U_1=\frac{1}{\sqrt{n}}\left(\frac{\frac{v}{\sqrt{n}}+\sqrt{\frac{v^2}{n}+4\omega\gamma}}{2\omega}\right), \tag{4.78}$$

proving the result of the theorem. □

Remark 4.5. *As the bound* (4.66) *does not include* $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})$*, in cases where a closed-form expression for the vector MLE is not available, we can still get an upper bound on the distributional distance between the distribution of the MLE and the* d*-variate standard normal,*

$$
\begin{aligned}
&\left|\mathrm{E}\left[h\left(\sqrt{n}[I(\boldsymbol{\theta_0})]^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})-\boldsymbol{\theta_0})\right)\right]-\mathrm{E}[h(\boldsymbol{Z})]\right| \\
&\leq \frac{\|h\|_2}{4\sqrt{n}}\sum_{j=1}^{d}\left[\mathrm{Var}\left(\left(\sum_{k=1}^{d}\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{jk}\frac{\partial}{\partial\theta_k}\log f(\boldsymbol{X_1}|\boldsymbol{\theta_0})\right)^2\right)\right]^{\frac{1}{2}} \\
&\quad+\frac{\|h\|_2}{2\sqrt{n}}\sum_{k<j}^{d}\sum_{j=1}^{d}\left[\mathrm{Var}\left(\sum_{q=1}^{d}\sum_{v=1}^{d}\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{jq}\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{kv}\right.\right. \\
&\qquad\qquad\left.\left.\times\frac{\partial}{\partial\theta_q}\log f(\boldsymbol{X_1}|\boldsymbol{\theta_0})\frac{\partial}{\partial\theta_v}\log f(\boldsymbol{X_1}|\boldsymbol{\theta_0})\right)\right]^{\frac{1}{2}} \\
&\quad+\frac{\|h\|_3}{12\sqrt{n}}\mathrm{E}\left(\sum_{i=1}^{d}\left|\sum_{l=1}^{d}\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{il}\left(\frac{\partial}{\partial\theta_l}\log f(\boldsymbol{X_1'}|\boldsymbol{\theta_0})-\frac{\partial}{\partial\theta_l}\log f(\boldsymbol{X_1}|\boldsymbol{\theta_0})\right)\right|\right)^3 \\
&\quad+2\frac{\|h\|}{\varepsilon^2}U_1^2 \\
&\quad+\|h\|_1\sqrt{d}U_1\sum_{k=1}^{d}\sum_{l=1}^{d}\left|\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{lk}\right|\sqrt{\sum_{i=1}^{d}\mathrm{Var}\left(\frac{\partial^2}{\partial\theta_k\partial\theta_i}\log f(\boldsymbol{X_1}|\boldsymbol{\theta_0})\right)} \\
&\quad+\frac{\|h\|_1\sqrt{n}}{2}U_1^2\sum_{k=1}^{d}\sum_{l=1}^{d}\left|\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{lk}\right|\sum_{m=1}^{d}\sum_{i=1}^{d}M_{kmi}. \qquad (4.79)
\end{aligned}
$$

Example: The Beta distribution

Here, we find an upper bound for the specific example of i.i.d. random variables from the Beta distribution with both shape parameters being unknown. An analytic expression for the MLE is not available. Applying the result of (4.78) to bound $\mathrm{E}\left(\sum_{j=1}^{2}\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\right)$, gives an upper bound for the distributional distance of interest. Some useful notations are now presented. Firstly, as in Section 3.5, $\Psi_j(.)$ is the j^{th} derivative of the digamma function Ψ, with $\Psi(z)=\frac{\Gamma'(z)}{\Gamma(z)}, z>0$. From (3.67), $\Psi_j(z)$ can be defined through a sum. For $\alpha,\beta,x,y>0$ and $0<\varepsilon<\min\{x,y\}$, let

$$
\begin{aligned}
&\delta_I := \Psi_1(\alpha)\Psi_1(\beta) - \Psi_1(\alpha+\beta)(\Psi_1(\alpha)+\Psi_1(\beta)),\\
&C_1(x,y) := \Psi_3(x) + \Psi_3(x+y) + 3[\Psi_1(x)]^2 + 3[\Psi_1(x+y)]^2,\\
&C_2(x,y) := \Psi_1(x) - \Psi_1(x+y) + \sqrt{\delta_I},\\
&C_3(x,y) := C_1(x,y)[C_2(y,x)]^2,\\
&C_4(x,y) := \frac{6x}{(y-\varepsilon)^4} + \frac{x\pi^4}{15} + \frac{6}{(x+y-\varepsilon)^3} + 7.26,\\
&M_B := \frac{1}{\delta_I}\sup\{\Psi_1(\alpha+\beta), \Psi_1(\min\{\alpha,\beta\}) - \Psi_1(\alpha+\beta)\},\\
&\gamma_B := \frac{4M_B}{\sqrt{n}\delta_I[C_2(\alpha,\beta)+C_2(\beta,\alpha)]}\left\{\left[(C_2(\beta,\alpha))^4 C_1(\alpha,\beta) + [\Psi_1(\alpha+\beta)]^4 C_1(\beta,\alpha)\right]^{\frac{1}{2}}\right.\\
&\qquad\left. + \left[(C_2(\alpha,\beta))^4 C_1(\beta,\alpha) + [\Psi_1(\alpha+\beta)]^4 C_1(\alpha,\beta)\right]^{\frac{1}{2}}\right\}\\
&\quad + \frac{M_B\sqrt{24}}{\sqrt{n}\delta_I[C_2(\alpha,\beta)+C_2(\beta,\alpha)]}\left\{[\Psi_1(\alpha+\beta)]^2(C_3(\alpha,\beta)+C_3(\beta,\alpha))\right.\\
&\qquad\left. + 2\sqrt{C_1(\alpha,\beta)C_1(\beta,\alpha)}\left[[\Psi_1(\alpha+\beta)]^4 + [C_2(\alpha,\beta)]^2[C_2(\beta,\alpha)]^2\right]\right\}^{\frac{1}{2}}\\
&\quad + \frac{\Psi_1(\beta)+\Psi_1(\alpha)-2\Psi_1(\alpha+\beta)}{\delta_I},\\
&\omega_B := 1 - 8\frac{M_B}{n\varepsilon^2} - \frac{2M_B\left\{\left(\Psi_1(\beta)+\sqrt{\delta_I}\right)C_4(\beta,\alpha) + \left(\Psi_1(\alpha)+\sqrt{\delta_I}\right)C_4(\alpha,\beta)\right\}}{\sqrt{n\delta_I(C_2(\alpha,\beta)+C_2(\beta,\alpha))}}.
\end{aligned}
\tag{4.80}
$$

Corollary 4.4. *Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables from the Beta(α,β) distribution with $\boldsymbol{\theta_0} = (\alpha,\beta)$. Let $m = \min\{\alpha,\beta\}$, $\varepsilon = \frac{m}{2} > 0$ and*

$$
\begin{aligned}
n \geq \frac{4}{m^2}\left(\right.&\frac{mM_B\left(\Psi_1(\beta)+\sqrt{\delta_I}\right)C_4(\beta,\alpha) + \left(\Psi_1(\alpha)+\sqrt{\delta_I}\right)C_4(\alpha,\beta)}{2\sqrt{\delta_I(C_2(\alpha,\beta)+C_2(\beta,\alpha))}}\\
&+\left.\left[\left(\frac{mM_B\left(\Psi_1(\beta)+\sqrt{\delta_I}\right)C_4(\beta,\alpha) + \left(\Psi_1(\alpha)+\sqrt{\delta_I}\right)C_4(\alpha,\beta)}{2\sqrt{\delta_I(C_2(\alpha,\beta)+C_2(\beta,\alpha))}}\right)^2 + 8M_B\right]^{\frac{1}{2}}\right)^2.
\end{aligned}
\tag{4.81}
$$

Then,

a) *When n satisfies* (4.81), $\mathrm{E}\left[\sum_{j=1}^{2}\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\right]\leq\sqrt{\frac{\gamma_B}{n\omega_B}}$.

b) *For* $\boldsymbol{Z}\sim \mathrm{N}_2(\mathbf{0},I_{2\times 2})$, *we obtain*

$$\begin{aligned}
&\left|\mathrm{E}\left[h\left(\sqrt{n}[I(\boldsymbol{\theta_0})]^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})-\boldsymbol{\theta_0})\right)\right]-\mathrm{E}[h(\boldsymbol{Z})]\right|\leq \\
&\frac{2\|h\|_2}{\sqrt{n}\delta_I\left[C_2(\alpha,\beta)+C_2(\beta,\alpha)\right]}\left\{\left[(C_2(\beta,\alpha))^4C_1(\alpha,\beta)+[\Psi_1(\alpha+\beta)]^4C_1(\beta,\alpha)\right]^{\frac{1}{2}}\right. \\
&\qquad\left.+\left[(C_2(\alpha,\beta))^4C_1(\beta,\alpha)+[\Psi_1(\alpha+\beta)]^4C_1(\alpha,\beta)\right]^{\frac{1}{2}}\right\} \\
&+\frac{\|h\|_2\sqrt{6}}{\sqrt{n}\delta_I\left[C_2(\alpha,\beta)+C_2(\beta,\alpha)\right]}\left\{[\Psi_1(\alpha+\beta)]^2\left(C_3(\alpha,\beta)+C_3(\beta,\alpha)\right)\right. \\
&\qquad\left.+2\sqrt{C_1(\alpha,\beta)C_1(\beta,\alpha)}\left[[\Psi_1(\alpha+\beta)]^4+[C_2(\alpha,\beta)]^2[C_2(\beta,\alpha)]^2\right]\right\}^{\frac{1}{2}} \\
&+\frac{32\|h\|_3 8^{\frac{3}{4}}}{3\sqrt{n}\left[\delta_I\left[C_2(\alpha,\beta)+C_2(\beta,\alpha)\right]\right]^{\frac{3}{2}}}\left\{\left[[C_2(\beta,\alpha)]^3+[\Psi_1(\alpha+\beta)]^3\right][C_1(\alpha,\beta)]^{\frac{3}{4}}\right. \\
&\qquad\left.+\left[[C_2(\alpha,\beta)]^3+[\Psi_1(\alpha+\beta)]^3\right][C_1(\beta,\alpha)]^{\frac{3}{4}}\right\} \\
&+\frac{8\|h\|\gamma_B}{nm^2\omega_B}+\frac{\|h\|_1\gamma_B\left\{\left(\Psi_1(\beta)+\sqrt{\delta_I}\right)C_4(\beta,\alpha)+\left(\Psi_1(\alpha)+\sqrt{\delta_I}\right)C_4(\alpha,\beta)\right\}}{2\sqrt{n}\omega_B\sqrt{\delta_I\left(C_2(\alpha,\beta)+C_2(\beta,\alpha)\right)}}.
\end{aligned}\tag{4.82}$$

The following lemma gives a useful approach to finding a square root of a matrix. The proof of the lemma is given in Somayya (1997).

Lemma 4.3. *Let* $M=\begin{pmatrix}A & B\\ C & D\end{pmatrix}$, *where* A,B,C,D *may be real or complex numbers. In addition, let* τ *and* δ *be the trace and determinant of M, respectively. Let s and t, such*

that $s^2 = \delta$ *and* $t^2 = \tau + 2s$*. Then, if* $t \neq 0$*, a square root of* M *is*

$$R = \frac{1}{t}\begin{pmatrix} A+s & B \\ C & D+s \end{pmatrix}.$$

Remark 4.6. *The lemma holds for* $s = \pm\sqrt{\delta}$ *and* $t = \pm\sqrt{\tau+2s}$*. Using different signs for those quantities gives different roots of the matrix* M*. From now on, unless otherwise stated, we take the positive roots for* s *and* t*.*

Proof of **Corollary 4.4**.

Part **a)**. The probability density function is

$$f(x|\boldsymbol{\theta}) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1},$$

with $\alpha, \beta > 0$ and $x \in [0,1]$. Hence, for $j,k \in \mathbb{Z}^+$

$$\begin{aligned}
&\frac{\partial}{\partial\alpha}\log f(x|\boldsymbol{\theta}) = \Psi(\alpha+\beta) - \Psi(\alpha) + \log(x) \\
&\frac{\partial}{\partial\beta}\log f(x|\boldsymbol{\theta}) = \Psi(\alpha+\beta) - \Psi(\beta) + \log(1-x) \\
&\frac{\partial^{j+1}}{\partial\alpha^{j+1}}\log f(x|\boldsymbol{\theta}) = \Psi_j(\alpha+\beta) - \Psi_j(\alpha) \\
&\frac{\partial^{j+1}}{\partial\beta^{j+1}}\log f(x|\boldsymbol{\theta}) = \Psi_j(\alpha+\beta) - \Psi_j(\beta) \\
&\frac{\partial^{k+j}}{\partial\alpha^k\partial\beta^j}\log f(x|\boldsymbol{\theta}) = \Psi_{k+j-1}(\alpha+\beta). \qquad (4.83)
\end{aligned}$$

The assumptions (Con.1)-(Con.3) are satisfied. For (Con.1) notice that $x \in [0,1]$. For (Con.2), let $\varepsilon_0 = \varepsilon = \varepsilon(\boldsymbol{\theta_0}) : 0 < \varepsilon < \min\{\alpha,\beta\}$. In addition, $\sum_{i=1}^{\infty}\frac{1}{i^4} = \frac{\pi^4}{90}$ and $\sum_{i=1}^{\infty}\frac{1}{i^3} = \zeta(3)$, where $\zeta(.)$ is the Riemann zeta function and $\zeta(3)$ is known as Apery's constant, which is an irrational number (≈ 1.202). For the calculations that follow,

$\zeta(3) < 1.21$ is used. With $\boldsymbol{\theta} = (\theta_1, \theta_2)$, using (3.67), (3.68) and (3.69) yields

$$\begin{aligned}
&\sup_{\substack{\boldsymbol{\theta}:|\theta_1-\alpha|<\varepsilon \\ |\theta_2-\beta|<\varepsilon}} \left| \frac{\partial^3}{\partial \theta_1^3} \log f(x|\boldsymbol{\theta}) \right| \leq \frac{6\beta}{(\alpha-\varepsilon)^4} + \frac{\beta\pi^4}{15} := M_{111} \\
&\sup_{\substack{\boldsymbol{\theta}:|\theta_1-\alpha|<\varepsilon \\ |\theta_2-\beta|<\varepsilon}} \left| \frac{\partial^3}{\partial \theta_2^3} \log f(x|\boldsymbol{\theta}) \right| \leq \frac{6\alpha}{(\beta-\varepsilon)^4} + \frac{\alpha\pi^4}{15} := M_{222} \\
&\sup_{\substack{\boldsymbol{\theta}:|\theta_1-\alpha|<\varepsilon \\ |\theta_2-\beta|<\varepsilon}} \left| \frac{\partial^3}{\partial \theta_1^2 \partial \theta_2} \log f(x|\boldsymbol{\theta}) \right| = \sup_{\substack{\boldsymbol{\theta}:|\theta_1-\alpha|<\varepsilon \\ |\theta_2-\beta|<\varepsilon}} \left| \frac{\partial^3}{\partial \theta_1 \partial \theta_2^2} \log f(x|\boldsymbol{\theta}) \right| \\
&\qquad\qquad \leq \frac{2}{(\alpha+\beta-2\varepsilon)^3} + 2\zeta(3) < \frac{2}{(\alpha+\beta-2\varepsilon)^3} + 2.42 \\
&\qquad\qquad := M_{112} = M_{121} = M_{211} = M_{122} = M_{212} = M_{221}. \quad (4.84)
\end{aligned}$$

Therefore, (Con.2) is also satisfied. Using (4.83), it is straightforward that the expected Fisher Information matrix in this case becomes

$$I(\boldsymbol{\theta_0}) = \begin{pmatrix} \Psi_1(\alpha) - \Psi_1(\alpha+\beta) & -\Psi_1(\alpha+\beta) \\ -\Psi_1(\alpha+\beta) & \Psi_1(\beta) - \Psi_1(\alpha+\beta) \end{pmatrix}.$$

The inverse of this 2×2 matrix is

$$[I(\boldsymbol{\theta_0})]^{-1} = \frac{1}{\delta_I} \begin{pmatrix} \Psi_1(\beta) - \Psi_1(\alpha+\beta) & \Psi_1(\alpha+\beta) \\ \Psi_1(\alpha+\beta) & \Psi_1(\alpha) - \Psi_1(\alpha+\beta) \end{pmatrix}$$

and using Lemma 4.3

$$[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}} = \frac{1}{[\delta_I (C_2(\alpha,\beta) + C_2(\beta,\alpha))]^{\frac{1}{2}}} \begin{pmatrix} C_2(\beta,\alpha) & \Psi_1(\alpha+\beta) \\ \Psi_1(\alpha+\beta) & C_2(\alpha,\beta) \end{pmatrix}. \quad (4.85)$$

For $M = M_B$, as defined in (Con.3), we have that

$$M_B = \frac{1}{\delta_I} \sup\{\Psi_1(\alpha+\beta), \Psi_1(\alpha) - \Psi_1(\alpha+\beta), \Psi_1(\beta) - \Psi_1(\alpha+\beta)\}$$

$$= \frac{1}{\delta_I} \sup\{\Psi_1(\alpha+\beta), \Psi_1(\min\{\alpha,\beta\}) - \Psi_1(\alpha+\beta)\},$$

as $\Psi_1(.)$ is a decreasing function. Having found M_B and $[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}$ and using (4.84) with the notation (4.80), (Con.3) is satisfied for

$$n \geq \frac{4}{m^2}\left(\frac{mM_B\left(\Psi_1(\beta)+\sqrt{\delta_I}\right)C_4(\beta,\alpha)+\left(\Psi_1(\alpha)+\sqrt{\delta_I}\right)C_4(\alpha,\beta)}{2\sqrt{\delta_I\left(C_2(\alpha,\beta)+C_2(\beta,\alpha)\right)}}\right.$$
$$\left.+\left[\left(\frac{mM_B\left(\Psi_1(\beta)+\sqrt{\delta_I}\right)C_4(\beta,\alpha)+\left(\Psi_1(\alpha)+\sqrt{\delta_I}\right)C_4(\alpha,\beta)}{2\sqrt{\delta_I\left(C_2(\alpha,\beta)+C_2(\beta,\alpha)\right)}}\right)^2+8M_B\right]^{\frac{1}{2}}\right)^2,$$

with $C_4(x,y)$ as in (4.80). We proceed with the bound for $\mathrm{E}\left[\sum_{j=1}^{2}\left(\hat{\theta}_n(\boldsymbol{X})_j-\theta_{0_j}\right)^2\right]$. Firstly, as the second-order partial derivatives of the log-likelihood function for the Beta distribution are not random,

$$\mathrm{Var}\left(\frac{\partial^2}{\partial\alpha^2}\log f(X|\boldsymbol{\theta_0})\right)=\mathrm{Var}\left(\frac{\partial^2}{\partial\beta^2}\log f(X|\boldsymbol{\theta_0})\right)=\mathrm{Var}\left(\frac{\partial^2}{\partial\alpha\partial\beta}\log f(X|\boldsymbol{\theta_0})\right)=0$$

and hence $v=0$. Therefore, for the Beta distribution

$$\mathrm{E}\left(\sum_{j=1}^{2}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})_j-\boldsymbol{\theta}_{\mathbf{0}_j}\right)^2\right)\leq U_1=\frac{\sqrt{4\omega\gamma}}{2\sqrt{n}\omega}=\sqrt{\frac{\gamma}{n\omega}}.$$

Our focus now turns to find an upper bound for γ. The first quantity of the result for γ in (4.65) is

$$\sum_{j=1}^{d}\left|\left[[I(\boldsymbol{\theta_0})]^{-1}\right]_{jj}\right|=\frac{1}{\delta_I}[\Psi_1(\beta)+\Psi_1(\alpha)-2\Psi_1(\alpha+\beta)]. \tag{4.86}$$

For the second term, using equations (3.70) and (3.71),

$$\begin{aligned}\mathrm{E}\left[\left(\frac{\partial}{\partial\alpha}\log f(X_1|\boldsymbol{\theta_0})\right)^4\right]&\leq 8C_1(\alpha,\beta)\\ \mathrm{E}\left[\left(\frac{\partial}{\partial\beta}\log f(X_1|\boldsymbol{\theta_0})\right)^4\right]&\leq 8C_1(\beta,\alpha),\end{aligned} \tag{4.87}$$

with $C_1(x,y)$ as in (4.80). Using (4.87) and (4.85) gives

$$
\begin{aligned}
&\left[\mathrm{Var}\left(\left(\sum_{k=1}^{2}\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{1k}\frac{\partial}{\partial\theta_k}\log f(X_1|\boldsymbol{\theta_0})\right)^2\right)\right]^{\frac{1}{2}} \\
&+\left[\mathrm{Var}\left(\left(\sum_{k=1}^{2}\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{2k}\frac{\partial}{\partial\theta_k}\log f(X_1|\boldsymbol{\theta_0})\right)^2\right)\right]^{\frac{1}{2}} \\
&\leq\left[\mathrm{E}\left(\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{11}\frac{\partial}{\partial\alpha}\log f(X_1|\boldsymbol{\theta_0})+\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{12}\frac{\partial}{\partial\beta}\log f(X_1|\boldsymbol{\theta_0})\right)^4\right]^{\frac{1}{2}} \\
&+\left[\mathrm{E}\left(\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{12}\frac{\partial}{\partial\alpha}\log f(X_1|\boldsymbol{\theta_0})+\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{22}\frac{\partial}{\partial\beta}\log f(X_1|\boldsymbol{\theta_0})\right)^4\right]^{\frac{1}{2}} \\
&\leq\sqrt{8}\left[\left(\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{11}\right)^4\mathrm{E}\left(\frac{\partial}{\partial\alpha}\log f(X_1|\boldsymbol{\theta_0})\right)^4\right. \\
&\qquad\left.+\left(\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{12}\right)^4\mathrm{E}\left(\frac{\partial}{\partial\beta}\log f(X_1|\boldsymbol{\theta_0})\right)^4\right]^{\frac{1}{2}} \\
&+\sqrt{8}\left[\left(\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{12}\right)^4\mathrm{E}\left(\frac{\partial}{\partial\alpha}\log f(X_1|\boldsymbol{\theta_0})\right)^4\right. \\
&\qquad\left.+\left(\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{22}\right)^4\mathrm{E}\left(\frac{\partial}{\partial\beta}\log f(X_1|\boldsymbol{\theta_0})\right)^4\right]^{\frac{1}{2}} \\
&\leq\frac{8}{\delta_I[C_2(\alpha,\beta)+C_2(\beta,\alpha)]}\left\{\left[(C_2(\beta,\alpha))^4C_1(\alpha,\beta)+(\Psi_1(\alpha+\beta))^4C_1(\beta,\alpha)\right]^{\frac{1}{2}}\right. \\
&\qquad\left.+\left[(\Psi_1(\alpha+\beta))^4C_1(\alpha,\beta)+(C_2(\alpha,\beta))^4C_1(\beta,\alpha)\right]^{\frac{1}{2}}\right\}.
\end{aligned}
\tag{4.88}
$$

For the third term of the expression for γ as in (4.65) since $d=2$,

$$\left[\mathrm{Var}\left(\sum_{q=1}^{2}\sum_{v=1}^{2}\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{2q}\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{1v}\frac{\partial}{\partial\theta_q}\log f(X_1|\boldsymbol{\theta_0})\frac{\partial}{\partial\theta_v}\log f(X_1|\boldsymbol{\theta_0})\right)\right]^{\frac{1}{2}}$$

$$\leq\left[\mathrm{E}\left(\sum_{q=1}^{2}\sum_{v=1}^{2}\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{2q}\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{1v}\frac{\partial}{\partial\theta_q}\log f(X_1|\boldsymbol{\theta_0})\frac{\partial}{\partial\theta_v}\log f(X_1|\boldsymbol{\theta_0})\right)^2\right]^{\frac{1}{2}}$$

$$=\Bigg[\mathrm{E}\Bigg(\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{12}\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{11}\left(\frac{\partial}{\partial\alpha}\log f(X_1|\boldsymbol{\theta_0})\right)^2$$
$$+\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{12}\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{22}\left(\frac{\partial}{\partial\beta}\log f(X_1|\boldsymbol{\theta_0})\right)^2$$
$$+\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{22}\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{11}\frac{\partial}{\partial\alpha}\log f(X_1|\boldsymbol{\theta_0})\frac{\partial}{\partial\beta}\log f(X_1|\boldsymbol{\theta_0})$$
$$+\left(\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{12}\right)^2\frac{\partial}{\partial\alpha}\log f(X_1|\boldsymbol{\theta_0})\frac{\partial}{\partial\beta}\log f(X_1|\boldsymbol{\theta_0})\Bigg)^2\Bigg]^{\frac{1}{2}}. \tag{4.89}$$

Using now the known inequality, $(\sum_{i=1}^{n}a_i)^2\leq n\sum_{i=1}^{n}a_i^2$ for $a_i\in R$, and also the Cauchy-Schwarz inequality yields

$$(4.89)\leq\Bigg[3\Bigg(\left(\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{12}\right)^2\left(\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{11}\right)^2\mathrm{E}\left(\frac{\partial}{\partial\alpha}\log f(X_1|\boldsymbol{\theta_0})\right)^4$$
$$+\left(\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{22}\right)^2\left(\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{12}\right)^2\mathrm{E}\left(\frac{\partial}{\partial\beta}\log f(X_1|\boldsymbol{\theta_0})\right)^4$$
$$+2\left(\left(\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{22}\right)^2\left(\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{11}\right)^2+\left(\left[[I(\boldsymbol{\theta_0})]^{-\frac{1}{2}}\right]_{12}\right)^4\right)$$
$$\times\sqrt{\mathrm{E}\left(\frac{\partial}{\partial\alpha}\log f(X_1|\boldsymbol{\theta_0})\right)^4\mathrm{E}\left(\frac{\partial}{\partial\beta}\log f(X_1|\boldsymbol{\theta_0})\right)^4}\Bigg)\Bigg]^{\frac{1}{2}}$$
$$\leq\frac{\sqrt{24}}{\delta_I[C_2(\alpha,\beta)+C_2(\beta,\alpha)]}\Bigg\{(\Psi_1(\alpha+\beta))^2\left[C_3(\alpha,\beta)+C_3(\beta,\alpha)\right]$$
$$+2\left[(C_2(\alpha,\beta))^2(C_2(\beta,\alpha))^2+(\Psi_1(\alpha+\beta))^4\right]\sqrt{C_1(\alpha,\beta)C_1(\beta,\alpha)}\Bigg\}^{\frac{1}{2}}. \tag{4.90}$$

The inequalities in (4.86), (4.88) and (4.90) show that $\gamma \leq \gamma_B$ as in (4.80). To calculate ω,

$$\begin{aligned}
&\sum_{l=1}^{2}\sum_{k=1}^{2}\left|\left[[I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}}\right]_{lk}\right|\sum_{m=1}^{2}\sum_{i=1}^{2}M_{kim}\\
&=\sum_{k=1}^{2}\left[\left|\left[[I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}}\right]_{1k}\right|+\left|\left[[I(\boldsymbol{\theta}_0)]^{-\frac{1}{2}}\right]_{2k}\right|\right]\sum_{m=1}^{2}\sum_{i=1}^{2}M_{kim}\\
&=\frac{\Psi_1(\beta)+\sqrt{\delta_I}}{\sqrt{\delta_I\left(C_2(\alpha,\beta)+C_2(\beta,\alpha)\right)}}\sum_{m=1}^{2}\sum_{i=1}^{2}M_{1im}+\frac{\Psi_1(\alpha)+\sqrt{\delta_I}}{\sqrt{\delta_I\left(C_2(\alpha,\beta)+C_2(\beta,\alpha)\right)}}\sum_{m=1}^{2}\sum_{i=1}^{2}M_{2im}\\
&=\frac{\Psi_1(\beta)+\sqrt{\delta_I}}{\sqrt{\delta_I\left(C_2(\alpha,\beta)+C_2(\beta,\alpha)\right)}}\left(\frac{6\beta}{(\alpha-\varepsilon)^4}+\frac{\beta\pi^4}{15}+\frac{6}{(\alpha+\beta-\varepsilon)^3}+7.26\right)\\
&\quad+\frac{\Psi_1(\alpha)+\sqrt{\delta_I}}{\sqrt{\delta_I\left(C_2(\alpha,\beta)+C_2(\beta,\alpha)\right)}}\left(\frac{6\alpha}{(\beta-\varepsilon)^4}+\frac{\alpha\pi^4}{15}+\frac{6}{(\alpha+\beta-\varepsilon)^3}+7.26\right)\\
&=\frac{\left(\Psi_1(\beta)+\sqrt{\delta_I}\right)C_4(\beta,\alpha)+\left(\Psi_1(\alpha)+\sqrt{\delta_I}\right)C_4(\alpha,\beta)}{\sqrt{\delta_I\left(C_2(\alpha,\beta)+C_2(\beta,\alpha)\right)}}
\end{aligned}\tag{4.91}$$

shows that $\omega=\omega_B$ as defined in (4.80). Since $\omega_B>0$ for n satisfying (Con.3), we conclude that

$$\mathrm{E}\left(\sum_{j=1}^{2}\left(\hat{\theta}_n(\boldsymbol{X})_j-\boldsymbol{\theta}_{0_j}\right)^2\right)\leq\frac{\sqrt{4\omega_B\gamma_B}}{2\sqrt{n}\omega_B}=\frac{\sqrt{\omega_B\gamma_B}}{\sqrt{n}\omega_B}=\sqrt{\frac{\gamma_B}{n\omega_B}}.\tag{4.92}$$

Part **b)**. For (4.82) we use the general expression of the bound in (4.79). Upper bounds for the first two terms of (4.79) have already been obtained in (4.88) and (4.90). Now, for ease of presentation let

$$\begin{aligned}
K_1(X_1,X_1')&:=\frac{\partial}{\partial\alpha}\log f(X_1'|\boldsymbol{\theta}_0)-\frac{\partial}{\partial\alpha}\log f(X_1|\boldsymbol{\theta}_0)\\
K_2(X_1,X_1')&:=\frac{\partial}{\partial\beta}\log f(X_1'|\boldsymbol{\theta}_0)-\frac{\partial}{\partial\beta}\log f(X_1|\boldsymbol{\theta}_0).
\end{aligned}\tag{4.93}$$

Using (4.93) and the inequality in (4.42), we have for the third term in (4.79),

$$
\begin{aligned}
&\mathrm{E}\left(\sum_{i=1}^{2}\left|\sum_{l=1}^{2}\left[I(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}}_{il}\left(\frac{\partial}{\partial\boldsymbol{\theta}_l}\log f(X_1'|\boldsymbol{\theta_0})-\frac{\partial}{\partial\boldsymbol{\theta}_l}\log f(X_1|\boldsymbol{\theta_0})\right)\right|\right)^3\\
&=\mathrm{E}\left(\left|\sum_{l=1}^{2}\left[I(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}}_{1l}\left(\frac{\partial}{\partial\boldsymbol{\theta}_l}\log f(X_1'|\boldsymbol{\theta_0})-\frac{\partial}{\partial\boldsymbol{\theta}_l}\log f(X_1|\boldsymbol{\theta_0})\right)\right|\right.\\
&\qquad\left.+\left|\sum_{l=1}^{2}\left[I(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}}_{2l}\left(\frac{\partial}{\partial\boldsymbol{\theta}_l}\log f(X_1'|\boldsymbol{\theta_0})-\frac{\partial}{\partial\boldsymbol{\theta}_l}\log f(X_1|\boldsymbol{\theta_0})\right)\right|\right)^3\\
&\le 4\mathrm{E}\left(\left|\sum_{l=1}^{2}\left[I(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}}_{1l}\left(\frac{\partial}{\partial\boldsymbol{\theta}_l}\log f(X_1'|\boldsymbol{\theta_0})-\frac{\partial}{\partial\boldsymbol{\theta}_l}\log f(X_1|\boldsymbol{\theta_0})\right)\right|^3\right.\\
&\qquad\left.+\left|\sum_{l=1}^{2}\left[I(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}}_{2l}\left(\frac{\partial}{\partial\boldsymbol{\theta}_l}\log f(X_1'|\boldsymbol{\theta_0})-\frac{\partial}{\partial\boldsymbol{\theta}_l}\log f(X_1|\boldsymbol{\theta_0})\right)\right|^3\right)\\
&=4\mathrm{E}\left(\left|\left[I(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}}_{11}\left(K_1(X_1,X_1')\right)+\left[I(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}}_{12}\left(K_2(X_1,X_1')\right)\right|^3\right.\\
&\qquad\left.+\left|\left[I(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}}_{12}\left(K_1(X_1,X_1')\right)+\left[I(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}}_{22}\left(K_2(X_1,X_1')\right)\right|^3\right)\\
&\le 16\left(\left[\left|\left[I(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}}_{11}\right|^3+\left|\left[I(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}}_{12}\right|^3\right]\mathrm{E}\left(\left|K_1(X_1,X_1')\right|^3\right)\right.\\
&\qquad\left.+\left[\left|\left[I(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}}_{12}\right|^3+\left|\left[I(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}}_{22}\right|^3\right]\mathrm{E}\left(\left|K_2(X_1,X_1')\right|^3\right)\right)\\
&\le 128\left(\left[\left|\left[I(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}}_{11}\right|^3+\left|\left[I(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}}_{12}\right|^3\right]8^{\frac{3}{4}}[C_1(\alpha,\beta)]^{\frac{3}{4}}\right.\\
&\qquad\left.+\left[\left|\left[I(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}}_{12}\right|^3+\left|\left[I(\boldsymbol{\theta_0})\right]^{-\frac{1}{2}}_{22}\right|^3\right]8^{\frac{3}{4}}[C_1(\beta,\alpha)]^{\frac{3}{4}}\right)\\
&=\frac{128}{\left[\delta_I\left(C_2(\alpha,\beta)+C_2(\beta,\alpha)\right)\right]^{\frac{3}{2}}}\left\{8^{\frac{3}{4}}[C_1(\alpha,\beta)]^{\frac{3}{4}}\left([C_2(\beta,\alpha)]^3+[\Psi_1(\alpha+\beta)]^3\right)\right.\\
&\qquad\left.+8^{\frac{3}{4}}[C_1(\beta,\alpha)]^{\frac{3}{4}}\left([C_2(\alpha,\beta)]^3+[\Psi_1(\alpha+\beta)]^3\right)\right\}.
\end{aligned}
\tag{4.94}
$$

The above inequalities are a result of (4.42). To be more specific, for the last inequality in (4.94) we use that

$$\begin{aligned}
&\left|\frac{\partial}{\partial\theta_j}\log f(X_1'|\boldsymbol{\theta_0}) - \frac{\partial}{\partial\theta_j}\log f(X_1|\boldsymbol{\theta_0})\right|^3 \\
&\leq \left(\left|\frac{\partial}{\partial\theta_j}\log f(X_1'|\boldsymbol{\theta_0})\right| + \left|\frac{\partial}{\partial\theta_j}\log f(X_1|\boldsymbol{\theta_0})\right|\right)^3 \\
&\leq 4\left(\left|\frac{\partial}{\partial\theta_j}\log f(X_1'|\boldsymbol{\theta_0})\right|^3 + \left|\frac{\partial}{\partial\theta_j}\log f(X_1|\boldsymbol{\theta_0})\right|^3\right)
\end{aligned}$$

and therefore,

$$\begin{aligned}
&\mathrm{E}\left(\left|\frac{\partial}{\partial\theta_j}\log f(X_1'|\boldsymbol{\theta_0}) - \frac{\partial}{\partial\theta_j}\log f(X_1|\boldsymbol{\theta_0})\right|^3\right) \\
&\leq 4\left(\mathrm{E}\left(\left|\frac{\partial}{\partial\theta_j}\log f(X_1'|\boldsymbol{\theta_0})\right|^3\right) + \mathrm{E}\left(\left|\frac{\partial}{\partial\theta_j}\log f(X_1|\boldsymbol{\theta_0})\right|^3\right)\right) \\
&= 8\,\mathrm{E}\left(\left|\frac{\partial}{\partial\theta_j}\log f(X_1|\boldsymbol{\theta_0})\right|^3\right) \leq 8\left[\mathrm{E}\left(\frac{\partial}{\partial\theta_j}\log f(X_1|\boldsymbol{\theta_0})\right)^4\right]^{\frac{3}{4}}.
\end{aligned}$$

Thus, applying the results of (4.88), (4.90), (4.91), (4.92) and (4.94) to (4.79) Corollary 4.4 follows. □

Remark 4.7. *Using the notation* (4.80)*, it is straightforward that the first three terms of the bound are* $\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$*. In addition, since* γ_B *and* ω_B *are* $\mathcal{O}(1)$*, the fourth and the fifth term of the bound are of order* $\frac{1}{n}$ *and* $\frac{1}{\sqrt{n}}$*, respectively. Combining these results for the order of each of the terms, the order of the bound* (4.82) *is* $\frac{1}{\sqrt{n}}$*.*

Chapter 5

An approach using the Delta method

The work presented in this chapter is based on a paper with Christophe Ley that has been submitted to the *Latin American Journal of Probability and Mathematical Statistics* (*ALEA*). We return to the single-parameter setting and our aim is still to provide sharp explicit upper bounds on Zolotarev-type distances between the distribution of the MLE and the normal distribution. Our approach is based on a combination of the Delta method, Stein's method, Taylor expansions and conditional expectations. The bounds presented in this chapter are slightly restrictive in the sense that they are applicable only in situations where a function of the MLE can be expressed as a sum of independent and identically distributed terms which are functions of the observations. In these cases, such bounds are at least as sharp as those presented in Chapter 3.

The chapter is organised as follows. In Section 5.1 the new bound in the case where a function of the MLE is a sum of i.i.d. terms is described, proved and compared to the bound given in Theorem 3.1. In Section 5.2, the results are applied to the class of single-parameter exponential family distributions and some examples are treated in detail.

5.1 New bounds on the distance to normal for the MLE

The aim of the section is to find an upper bound for $d_{bW}\left(\sqrt{n\,i(\theta_0)}(\hat{\theta}_n(\boldsymbol{X})-\theta_0), Z\right)$. The tools used to reach this result are now the Delta method, Stein's method (through Lemma 2.2) and, as in Chapter 3, Taylor expansions and conditional expectations.

Consider a sample of i.i.d. random variables $X_1, X_2, \ldots, X_n$ from a single-parameter distribution and, as in Chapter 3, $\hat{\theta}_n(\boldsymbol{X})$ denotes the MLE for the parameter $\theta \in \Theta$, which is as usual to exist. We are interested in settings where there exists a one-to-one twice differentiable function $q : \Theta \to \mathbb{R}$ such that

$$q\left(\hat{\theta}_n(\boldsymbol{X})\right) = \frac{1}{n}\sum_{i=1}^{n} g(X_i) \tag{5.1}$$

for some $g : \mathbb{R} \to \mathbb{R}$. Some classical examples include

1) The normal distribution with density $f\left(x|\mu,\sigma^2\right) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$, $x \in \mathbb{R}$, for which $\mu \in \mathbb{R}$ is an unknown parameter, whereas $\sigma > 0$ is considered to be known. The MLE for μ is

$$\hat{\theta}_n(\boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i;$$

in this case $q(\theta) = \theta$ and $g(x) = x$.

2) The normal distribution, where now the mean μ is known and σ^2 represents the unknown parameter, with

$$\hat{\theta}_n(\boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^{n} (X_i-\mu)^2;$$

in this case $q(\theta) = \theta$ and $g(x) = (x-\mu)^2$.

3) The Weibull distribution with density $f(x|\alpha,\sigma) = \frac{\alpha}{\sigma}\left(\frac{x}{\sigma}\right)^{\alpha-1}\exp\left\{-\left(\frac{x}{\sigma}\right)^{\alpha}\right\}$, $x \geq 0$, where σ is the unknown scale parameter and $\alpha > 0$ is fixed. The MLE for σ is

$$\hat{\theta}_n(\boldsymbol{X}) = \left(\frac{1}{n}\sum_{i=1}^{n} X_i^{\alpha}\right)^{\frac{1}{\alpha}};$$

in this case, it is easily deduced that $q(\theta) = \theta^{\alpha}$ and $g(x) = x^{\alpha}$.

4) The Laplace scale model with density $f(x|\sigma) = \frac{1}{2\sigma}\exp\left\{-\frac{|x|}{\sigma}\right\}, \sigma > 0$, over $\mathbb{R}$, for which

$$\hat{\theta}_n(\boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^{n} |X_i|;$$

in this case $q(\theta) = \theta$ and $g(x) = |x|$.

Moreover, the broad single-parameter exponential families, discussed in Section 3.3 satisfy condition (5.1); see Proposition 5.1 for details. Therefore, the results presented in this section apply to many of the well-known distributions.

Our strategy consists in benefiting from the special form of $q(\hat{\theta}_n(\boldsymbol{X}))$, which is a sum of random variables, and thus allows us to use the sharp bound of Lemma 2.2. It is precisely at this point that the Delta method comes into play: instead of comparing $\hat{\theta}_n(\boldsymbol{X})$ to $Z \sim \mathrm{N}(0,1)$ we rather compare $q(\hat{\theta}_n(\boldsymbol{X}))$ to Z, and then bound the distance between $\hat{\theta}_n(\boldsymbol{X})$ and $q(\hat{\theta}_n(\boldsymbol{X}))$. The univariate Delta method is given in the following theorem.

Theorem 5.1. *Let θ be the parameter with parameter space $\Theta \subset \mathbb{R}$. Also, let Y_n be a sequence of random variables which satisfies*

$$\sqrt{n}(Y_n - \theta) \xrightarrow[n\to\infty]{\mathrm{d}} \sigma Z.$$

Then, for any function $g : \mathbb{R} \to \mathbb{R}$ such that $g'(\theta)$ exists and is not equal to zero we obtain

$$\sqrt{n}(g(Y_n) - g(\theta)) \xrightarrow[n\to\infty]{\mathrm{d}} g'(\theta)\sigma Z. \tag{5.2}$$

For a proof of the theorem see (Casella and Berger, 2002, p.243). The following lemma is useful for the proof of the main result of this section.

Lemma 5.1. *Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables following a distribution with unknown parameter θ. Its true value is denoted by θ_0. Assume that (R1)-(R4) of Section 2.2 are satisfied and that the MLE $\hat{\theta}_n(\boldsymbol{X})$ exists and is unique. Let $q : \Theta \to \mathbb{R}$ be a differentiable function such that $q(\hat{\theta}_n(\boldsymbol{X})) = \frac{1}{n}\sum_{i=1}^n g(X_i)$, for some function $g : \mathbb{R} \to \mathbb{R}$. With* $\mathrm{E}[.]$ *and* $\mathrm{Var}[.]$ *the expectation and variance under θ_0, then if* $\mathrm{E}[g(X_1)]$ *and* $\mathrm{Var}[g(X_1)]$ *exist, we have that*

$$\mathrm{E}\left[g(X_1)\right] = q(\theta_0) \qquad \textit{and} \qquad \mathrm{Var}\left[g(X_1)\right] = \frac{[q'(\theta_0)]^2}{i(\theta_0)}.$$

Proof. Since $q\left(\hat{\theta}_n(\boldsymbol{X})\right) = \frac{1}{n}\sum_{i=1}^n g(X_i)$, we have

$$\begin{aligned} \mathrm{E}\left[q(\hat{\theta}_n(\boldsymbol{X})\right] &= \frac{1}{n}\sum_{i=1}^n \mathrm{E}[g(X_i)] = \mathrm{E}[g(X_1)] \\ \mathrm{Var}\left[\sqrt{n}q\left(\hat{\theta}_n(\boldsymbol{X})\right)\right] &= \frac{1}{n}\sum_{i=1}^n \mathrm{Var}[g(X_i)] = \mathrm{Var}[g(X_1)]. \end{aligned} \tag{5.3}$$

In addition, since the regularity conditions (R1)-(R4) are satisfied then the result (b) in (2.2) holds and the Delta method result in (5.2) yields

$$\sqrt{n}(q(\hat{\theta}_n(\boldsymbol{X})) - q(\theta_0)) \xrightarrow[n\to\infty]{\mathrm{d}} \frac{q'(\theta_0)}{\sqrt{i(\theta_0)}}Z,$$

which indicates that

$$\sqrt{n}\left(\mathrm{E}\left(q\left(\hat{\theta}_n(\boldsymbol{X})\right)\right) - q(\theta_0)\right) \xrightarrow[n\to\infty]{} 0 \text{ and } \mathrm{Var}\left(\sqrt{n}q\left(\hat{\theta}_n(\boldsymbol{X})\right)\right) \xrightarrow[n\to\infty]{} \frac{[q'(\theta_0)]^2}{i(\theta_0)}. \tag{5.4}$$

Combining the results in (5.3) and (5.4) we summarise that $\mathrm{E}\left[q\left(\hat{\theta}_n(\boldsymbol{X})\right)\right] = \mathrm{E}[g(X_1)]$ and as n goes to infinity the limit of $\mathrm{E}(q(\hat{\theta}_n(\boldsymbol{X})))$ is $q(\theta_0)$, meaning that $\mathrm{E}[g(X_1)] = q(\theta_0)$. Applying the same idea for $\mathrm{Var}\left[\sqrt{n}q(\hat{\theta}_n(\boldsymbol{X}))\right]$ which is equal to $\mathrm{Var}[g(X_1)]$ and converges to $\frac{[q'(\theta_0)]^2}{i(\theta_0)}$, yields the result of the Lemma. □

We are now ready to give an upper bound on the distributional distance of interest using the Delta method from Theorem 5.1.

Theorem 5.2. *Let $X_1,\ldots,X_n$ be i.i.d. random variables with probability density (or mass) function $f(x_i|\theta)$ and let $Z \sim \mathrm{N}(0,1)$. Assume that the regularity conditions (R1)-(R4) are satisfied and that the MLE, $\hat{\theta}_n(\boldsymbol{X})$, exists and is unique. Furthermore let $q : \Theta \to \mathbb{R}$ be a one-to-one twice differentiable function with $q'(\theta) \neq 0$ $\forall\theta \in \Theta$ such that $q\left(\hat{\theta}_n(\boldsymbol{X})\right) = \frac{1}{n}\sum_{i=1}^n g(X_i)$, where the mapping $g : \mathbb{R} \to \mathbb{R}$ is such that $\mathrm{E}\left[|g(X_1) - q(\theta_0)|^3\right] < \infty$. Then,*

(1) if $q''(.)$ is not uniformly bounded but for any $\theta_0 \in \Theta$ there exists $0 < \varepsilon = \varepsilon(\theta_0)$ such that $\sup_{\theta:|\theta-\theta_0|<\varepsilon} |q''(\theta)| < \infty$, then

$$d_{bW}\left(\sqrt{n\,i(\theta_0)}(\hat{\theta}_n(\boldsymbol{X}) - \theta_0), Z\right) \leq \frac{1}{\sqrt{n}}\left(2 + \frac{[i(\theta_0)]^{\frac{3}{2}}}{|q'(\theta_0)|^3}\mathrm{E}\left[|g(X_1) - q(\theta_0)|^3\right]\right)$$
$$+ \mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right)^2\right]\mathbb{1}_{\{q(\theta)\neq\theta\}}\left(\frac{2}{\varepsilon^2} + \frac{\sqrt{n\,i(\theta_0)}}{2\,|q'(\theta_0)|}\sup_{\theta:|\theta-\theta_0|<\varepsilon}\left|q''(\theta)\right|\right). \quad (5.5)$$

(2) If $q''(.)$ is uniformly bounded, with $|q''(\theta)| \leq B^*, \forall\theta \in \Theta$, then

$$d_{bW}\left(\sqrt{n\,i(\theta_0)}(\hat{\theta}_n(\boldsymbol{X}) - \theta_0), Z\right) \leq \frac{1}{\sqrt{n}}\left(2 + \frac{[i(\theta_0)]^{\frac{3}{2}}}{|q'(\theta_0)|^3}\mathrm{E}\left[|g(X_1) - q(\theta_0)|^3\right]\right)$$
$$+ B^*\frac{\sqrt{n\,i(\theta_0)}}{2\,|q'(\theta_0)|}\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right)^2\right]\mathbb{1}_{\{q(\theta)\neq\theta\}}. \quad (5.6)$$

Proof. The asymptotic normality of the MLE is explicitly stated in Theorem 2.2. Applying (5.2) to this result in combination with the requirement $q'(\theta_0) \neq 0$ yields

$$\frac{\sqrt{n\, i(\theta_0)}}{q'(\theta_0)} \left(q\left(\hat{\theta}_n(\boldsymbol{X})\right) - q(\theta_0)\right) \xrightarrow[n\to\infty]{\mathrm{d}} Z,$$

with $q\left(\hat{\theta}_n(\boldsymbol{X})\right) = \frac{1}{n}\sum_{i=1}^{n} g(X_i)$. Using the triangle inequality,

$$\begin{aligned}
&\left|\mathrm{E}\left[h\left(\sqrt{n\, i(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right)\right)\right] - \mathrm{E}[h(Z)]\right| \\
&\leq \left|\mathrm{E}\left[h\left(\frac{\sqrt{n\, i(\theta_0)}}{q'(\theta_0)}\left(q\left(\hat{\theta}_n(\boldsymbol{X})\right) - q(\theta_0)\right)\right)\right] - \mathrm{E}[h(Z)]\right| \qquad (5.7) \\
&+ \left|\mathrm{E}\left[h\left(\sqrt{n\, i(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right)\right) - h\left(\frac{\sqrt{n\, i(\theta_0)}}{q'(\theta_0)}\left(q\left(\hat{\theta}_n(\boldsymbol{X})\right) - q(\theta_0)\right)\right)\right]\right|. \qquad (5.8)
\end{aligned}$$

To find an upper bound for (5.7), some simple rewriting yields

$$\begin{aligned}
\frac{\sqrt{n\, i(\theta_0)}}{q'(\theta_0)}\left(q\left(\hat{\theta}_n(\boldsymbol{X})\right) - q(\theta_0)\right) &= \frac{\sqrt{n\, i(\theta_0)}}{q'(\theta_0)}\left(\frac{1}{n}\sum_{i=1}^{n} g(X_i) - q(\theta_0)\right) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{\sqrt{i(\theta_0)}}{q'(\theta_0)}\left(g(X_i) - q(\theta_0)\right)\right\} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i,
\end{aligned}$$

where $Y_i = \frac{\sqrt{i(\theta_0)}}{q'(\theta_0)}\left(g(X_i) - q(\theta_0)\right), i = 1, \ldots, n$ which are i.i.d. random variables. Using the result of Lemma 5.1 we deduce that $\mathrm{E}[Y_i] = 0$ and $\mathrm{Var}[Y_i] = 1$, as Lemma 2.2 requires, which for $\sigma^2 = 1$ gives

$$d_{bW}\left(\frac{\sqrt{n\, i(\theta_0)}}{q'(\theta_0)}(q(\hat{\theta}_n(\boldsymbol{X})) - q(\theta_0)), Z\right) \leq \frac{1}{\sqrt{n}}\left(2 + \frac{[i(\theta_0)]^{\frac{3}{2}}}{|q'(\theta_0)|^3}\mathrm{E}\left[|g(X_1) - q(\theta_0)|^3\right]\right). \qquad (5.9)$$

For an upper bound on (5.8), since the case $q(\theta) = \theta$ is obvious, assume from now on that $q(\theta) \neq \theta$. We denote by

$$A := A(h,q,\theta_0,\boldsymbol{X}) := h\left(\sqrt{n\,i(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)\right) - h\left(\frac{\sqrt{n\,i(\theta_0)}}{q'(\theta_0)}\left(q\left(\hat{\theta}_n(\boldsymbol{X})\right)-q(\theta_0)\right)\right).$$

To find an upper bound for $|\mathrm{E}(A)|$, the approach is similar to the one developed in the proof of Theorem 3.1.

Case 1: $\forall\theta\in\Theta$ there exists $0<\varepsilon=\varepsilon(\theta_0)$ such that $\sup_{\theta:|\theta-\theta_0|<\varepsilon}|q''(\theta)|<\infty$. Then, using the law of total expectation related to conditioning on $\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right|\geq\varepsilon$ or $\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right|<\varepsilon$ yields

$$\begin{aligned}|\mathrm{E}[A]|\leq\mathrm{E}[|A|] &= \mathrm{E}\left[|A|\;\middle|\;\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right|\geq\varepsilon\right]\mathbb{P}\left(\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right|\geq\varepsilon\right)\\ &\quad+\mathrm{E}\left[|A|\;\middle|\;\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right|<\varepsilon\right]\mathbb{P}\left(\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right|<\varepsilon\right).\end{aligned}$$

Markov's inequality and the elementary results of $\mathbb{P}\left(\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right|<\varepsilon\right)\leq 1$ and $|A|\leq 2\|h\|$ further yield

$$|\mathrm{E}(A)|\leq 2\|h\|\frac{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]}{\varepsilon^2}+\mathrm{E}\left[|A|\;\middle|\;\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right|<\varepsilon\right]. \tag{5.10}$$

We now focus on the conditional expectation on the right-hand side of (5.10). A second-order Taylor expansion of $q\left(\hat{\theta}_n(\boldsymbol{x})\right)$ about θ_0 gives

$$q\left(\hat{\theta}_n(\boldsymbol{x})\right)=q(\theta_0)+\left(\hat{\theta}_n(\boldsymbol{x})-\theta_0\right)q'(\theta_0)+\frac{1}{2}\left(\hat{\theta}_n(\boldsymbol{x})-\theta_0\right)^2 q''(\theta^*), \tag{5.11}$$

for θ^* between $\hat{\theta}_n(\boldsymbol{x})$ and θ_0. Since $q'(\theta)\neq 0\;\forall\theta\in\Theta$ is assumed, we can multiply both sides in (5.11) with $\frac{\sqrt{n\,i(\theta_0)}}{q'(\theta_0)}$. Rearranging the terms,

$$\frac{\sqrt{n\,i(\theta_0)}\left(q\left(\hat{\theta}_n(\boldsymbol{x})\right)-q(\theta_0)\right)}{q'(\theta_0)}=\sqrt{n\,i(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{x})-\theta_0\right)+\frac{\sqrt{n\,i(\theta_0)}}{2q'(\theta_0)}q''\left(\theta^*\right)\left(\hat{\theta}_n(\boldsymbol{x})-\theta_0\right)^2.$$

The above result along with another first-order Taylor expansion gives

$$h\left(\sqrt{n\,i(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{x})-\theta_0\right)\right)-h\left(\frac{\sqrt{n\,i(\theta_0)}}{q'(\theta_0)}\left(q\left(\hat{\theta}_n(\boldsymbol{x})\right)-q(\theta_0)\right)\right)$$
$$=-\frac{\sqrt{n\,i(\theta_0)}}{2q'(\theta_0)}q''(\theta^*)\left(\hat{\theta}_n(\boldsymbol{x})-\theta_0\right)^2 h'(t(\boldsymbol{x})), \tag{5.12}$$

where $t(\boldsymbol{x})$ is between $\sqrt{n\,i(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{x})-\theta_0\right)$ and $\frac{\sqrt{n\,i(\theta_0)}}{q'(\theta_0)}\left(q\left(\hat{\theta}_n(\boldsymbol{x})\right)-q(\theta_0)\right)$. Equality (5.12) combined with Lemma 3.1 related to conditional expectations yield

$$\begin{aligned}
&\mathrm{E}\left[|A|\;\middle|\;\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right|<\varepsilon\right]\\
&=\mathrm{E}\left[\left|-\frac{\sqrt{n\,i(\theta_0)}}{2q'(\theta_0)}q''(\theta^*)\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2 h'(t(\boldsymbol{X}))\right|\;\middle|\;\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right|<\varepsilon\right]\\
&\leq\frac{\|h'\|\sqrt{n\,i(\theta_0)}}{2\,|q'(\theta_0)|}\mathrm{E}\left[\left|q''(\theta^*)\right|\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\;\middle|\;\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right|<\varepsilon\right]\\
&\leq\frac{\|h'\|\sqrt{n\,i(\theta_0)}}{2\,|q'(\theta_0)|}\sup_{\theta:|\theta-\theta_0|<\varepsilon}\left|q''(\theta)\right|\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\;\middle|\;\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right|<\varepsilon\right]\\
&\leq\frac{\|h'\|\sqrt{n\,i(\theta_0)}}{2\,|q'(\theta_0)|}\sup_{\theta:|\theta-\theta_0|<\varepsilon}\left|q''(\theta)\right|\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right].
\end{aligned}\tag{5.13}$$

Combining the bounds in (5.9), (5.10) and (5.13) we get, for $h\in H_{bW}$ as in (2.6), the result in (5.5).

Case 2: $q''(.)$ is uniformly bounded with $|q''(\theta)|\leq B^*$, $\forall\theta\in\Theta$. In this case, we do not need to take conditional expectations. The result in (5.12) gives

$$\begin{aligned}
\mathrm{E}\left[|A|\right]&\leq\frac{\|h'\|\sqrt{n\,i(\theta_0)}}{2\,|q'(\theta_0)|}\mathrm{E}\left[\left|q''(\theta^*)\right|\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]\\
&\leq B^*\frac{\|h'\|\sqrt{n\,i(\theta_0)}}{2\,|q'(\theta_0)|}\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right].
\end{aligned}$$

The above result and the bound in (5.9) give for $h\in H_{bW}$ as in (2.6) the bound in (5.6) for the case where $q''(.)$ is uniformly bounded in θ. □

Remark 5.1. ***(1)*** *The convergence of the second and third terms in* (5.5) *is governed by the asymptotic behaviour of* $\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]$*, whose rate of convergence is* $\mathcal{O}\left(\frac{1}{n}\right)$*; see Remark 3.1.*

(2) *Note that if* $q''(.)$ *is uniformly bounded, then* $0<\varepsilon=\varepsilon(\theta_0)$ *is not needed leading to a simpler bound; see* (5.6)*.*

(3) *In the simplest possible situation where* $\hat{\theta}_n(\boldsymbol{X})$ *is already a sum of i.i.d. terms,* $q(x)=x$ *and hence the upper bounds in* (5.5) *and* (5.6) *simplify to*

$$d_{bW}\left(\sqrt{n\,i(\theta_0)}(\hat{\theta}_n(\boldsymbol{X})-\theta_0),Z\right)\leq\frac{1}{\sqrt{n}}\left(2+[i(\theta_0)]^{\frac{3}{2}}\mathrm{E}\left[|g(X_1)-\theta_0|^3\right]\right),$$

where Lemma 2.2 has been directly used.

We now compare this bound to (3.7). The bound using Delta method is simpler. A drawback is that in order to apply the Delta method approach an analytic expression for the MLE is required which in addition needs to be written as a function of a sum of independent terms. On the other hand, the approach through the score function, as explained in Proposition 3.2 and in the proof of Theorem 3.1 holds whatever the form of the MLE is, or even when an explicit analytic expression of the MLE is not known; we showed that in Section 3.5. Let us now comment on the two bounds in (3.7) and (5.5) term by term.

- For the first term of the bounds, the different positioning of the expected Fisher information number is explained by the fact that we apply Lemma 2.2 to the standardised version of $g(X_1),g(X_2),\ldots,g(X_n)$, which have variance $\frac{[q'(\theta_0)]^2}{i(\theta_0)}$ (note that $i(\theta_0)$ is in the denominator), while in Theorem 3.1 the result was obtained by applying the lemma after standardising the first derivative of the log-likelihood function for each random variable. The variance in that case is equal to $i(\theta_0)$;
- the second and the third terms of the bound in (5.5) vanish when $q(\theta)=\theta$. For $q(\theta)\neq\theta$ the second term is the same in both bounds, whereas the third term in

(5.5) reads

$$\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]\frac{\sqrt{n\,i(\theta_0)}}{2\,|q'(\theta_0)|}\sup_{\theta:|\theta-\theta_0|<\varepsilon}\left|q''(\theta)\right| \tag{5.14}$$

and is to be compared to

$$\begin{aligned}&\frac{1}{\sqrt{i(\theta_0)}}\sqrt{\mathrm{Var}\left(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(X_1|\theta_0)\right)}\sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]}\\&+\frac{1}{2\sqrt{ni(\theta_0)}}\left[\mathrm{E}\left(\left(\sum_{i=1}^{n}M(X_i)\right)^2\middle|\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right|<\varepsilon\right)\right]^{\frac{1}{2}}\\&\qquad\times\left[\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^4\right]\right]^{\frac{1}{2}}.\end{aligned} \tag{5.15}$$

The second derivative, $q''(\theta)$ in (5.14) plays the role of $l^{(3)}(\theta;\boldsymbol{X})$, up to a difference: in a neighbourhood of θ_0, $l^{(3)}(\theta;\boldsymbol{X})$ is bounded by $\sum_{i=1}^{n}M(X_i)$. Consequently, the second term of (5.15) has $\sqrt{n}$ in its numerator, exactly as in (5.14). The distinct positioning of the information number $i(\theta_0)$ has the same reason as for the first term. The bound in (5.14) is clearly simpler (has one less term) and sharper than (5.15) at the level of moments of $\hat{\theta}_n(\boldsymbol{X})-\theta_0$ since

$$\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]\leq\left[\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^4\right]\right]^{\frac{1}{2}}$$

by the Cauchy-Schwarz inequality.

Therefore, in the simple, but still widely applicable setting where $\hat{\theta}_n(\boldsymbol{X})$ is a function of a sum of i.i.d. terms, the bound using the Delta method is simpler than the one in Theorem 3.1 and also of the same order. As already mentioned, an obvious advantage of the approach followed in Chapter 3 is its wider applicability as it works for all MLE settings and when no closed-form expression of the MLE is available.

5.2 Comparison in exponential families

In this section, we give the bound using the Delta method in the case of single-parameter exponential family distributions. The Generalised-Gamma distribution is also discussed and an upper bound is given. To allow for a further comparison between the two bounds, the exponential distribution both in the canonical and the non-canonical form as in Section 3.3 is used.

5.2.1 Bounds for single-parameter exponential families

For the single-parameter exponential families framework, where the probability density (or mass) function of a random variable is as in (3.21), it is straightforward to see that the function $q(.)$ is defined in general.

Proposition 5.1. *Suppose $X_1, \ldots, X_n$ are i.i.d. random variables with probability density (or mass) function that can be expressed in the form of* (3.21)*. Assume that (Ass.Ex.1)-(Ass.Ex.3) and (A1)-(A5) (Section 3.3) are satisfied. Then, with $D(\theta) = \frac{A'(\theta)}{k'(\theta)}$ as in (A1), we obtain that $q(\theta_0) = D(\theta_0)$ for $q : \Theta \to \mathbb{R}$ as in Theorem 5.2.*

Proof. Using (3.21), we have that

$$l'(\theta; \boldsymbol{x}) = k'(\theta) \sum_{i=1}^{n} T(x_i) - nA'(\theta) = 0 \Longleftrightarrow D(\theta) = \frac{1}{n} \sum_{i=1}^{n} T(x_i).$$

Since $D(.)$ is invertible (see condition (A1)), $\hat{\theta}_n(\boldsymbol{X}) = D^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} T(X_i)\right)$. The claim readily follows. □

The broad single-parameter exponential families satisfy (5.1). Consequently, Theorem 5.2 can be applied to (3.21), resulting in

Corollary 5.1. *Let $X_1,\dots,X_n$ be i.i.d. random variables with the probability density (or mass) function of a single-parameter exponential family. Assume that (Ass.Ex.1)-(Ass.Ex.3) and (A1)-(A5) are satisfied. With $Z\sim \mathrm{N}(0,1)$, we obtain*

(1) if $D''(.)$ is not uniformly bounded but for any $\theta_0\in\Theta$ there exists $0<\varepsilon=\varepsilon(\theta_0)$ such that $\sup_{\theta:|\theta-\theta_0|<\varepsilon}|D''(\theta)|<\infty$, then

$$d_{bW}\left(\sqrt{n\,i(\theta_0)}(\hat\theta_n(\boldsymbol{X})-\theta_0),Z\right)\le\frac{1}{\sqrt{n}}\left(2+\frac{|k'(\theta_0)|^3\mathrm{E}\left[|T(X_1)-D(\theta_0)|^3\right]}{|A''(\theta_0)-k''(\theta_0)D(\theta_0)|^{\frac{3}{2}}}\right)$$
$$+\mathrm{E}\left[\left(\hat\theta_n(\boldsymbol{X})-\theta_0\right)^2\right]\mathbb{1}_{\{D(\theta)\neq\theta\}}\left(\frac{2}{\varepsilon^2}\right.$$
$$\left.+\frac{\sqrt{n}|k'(\theta_0)|}{2\sqrt{|A''(\theta_0)-k''(\theta_0)D(\theta_0)|}}\sup_{\theta:|\theta-\theta_0|<\varepsilon}\left|D''(\theta)\right|\right). \tag{5.16}$$

(2) If $D''(.)$ is uniformly bounded, with $|D''(\theta)|\le B^*,\forall\theta\in\Theta$, then

$$d_{bW}\left(\sqrt{n\,i(\theta_0)}(\hat\theta_n(\boldsymbol{X})-\theta_0),Z\right)\le\frac{1}{\sqrt{n}}\left(2+\frac{|k'(\theta_0)|^3\mathrm{E}\left[|T(X_1)-D(\theta_0)|^3\right]}{|A''(\theta_0)-k''(\theta_0)D(\theta_0)|^{\frac{3}{2}}}\right)$$
$$+B^*\frac{\sqrt{n}|k'(\theta_0)|}{2\sqrt{|A''(\theta_0)-k''(\theta_0)D(\theta_0)|}}\mathrm{E}\left[\left(\hat\theta_n(\boldsymbol{X})-\theta_0\right)^2\right]\mathbb{1}_{\{D(\theta)\neq\theta\}}. \tag{5.17}$$

Proof. We readily have

$$i(\theta_0)=\mathrm{E}\left[-\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(X_1|\theta_0)\right]=A''(\theta_0)-k''(\theta_0)\mathrm{E}[T(X_1)]$$
$$=\frac{A''(\theta_0)k'(\theta_0)-k''(\theta_0)A'(\theta_0)}{k'(\theta_0)}$$
$$q'(\theta_0)=\frac{A''(\theta_0)k'(\theta_0)-k''(\theta_0)A'(\theta_0)}{[k'(\theta_0)]^2}.$$

Therefore,

$$\frac{\sqrt{i(\theta_0)}}{|q'(\theta_0)|} = \frac{|k'(\theta_0)|^{\frac{3}{2}}}{\sqrt{|A''(\theta_0)k'(\theta_0) - k''(\theta_0)A'(\theta_0)|}} = \frac{|k'(\theta_0)|}{\sqrt{|A''(\theta_0) - k''(\theta_0)D(\theta_0)|}}.$$

This result, along with the fact that $g(x) = T(x)$ and $q(\theta) = D(\theta)$ by definition, allows to deduce the upper bounds from Theorem 5.2. □

Remark 5.2. *It is particularly interesting to spell out this bound in the canonical case ($k(\theta) = \theta$). Since $k'(\theta) = 1$, then $D(\theta) = A'(\theta)$ and thus*

$$d_{bW}\left(\sqrt{n\,i(\theta_0)}(\hat{\theta}_n(\boldsymbol{X}) - \theta_0), Z\right) \leq \frac{1}{\sqrt{n}}\left(2 + \frac{\mathrm{E}\left[|T(X_1) - A'(\theta_0)|^3\right]}{|A''(\theta_0)|^{\frac{3}{2}}}\right)$$
$$+ \mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right)^2\right] \mathbb{1}_{\{A'(\theta)\neq\theta\}}\left(\frac{2}{\varepsilon^2} + \frac{\sqrt{n}}{2\sqrt{|A''(\theta_0)|}} \sup_{\theta:|\theta-\theta_0|<\varepsilon}\left|A^{(3)}(\theta)\right|\right),$$

which is exactly the same as the bound in (3.24)*. If $\left|A^{(3)}(\theta)\right|$ is uniformly bounded by a constant B^* then*

$$d_{bW}\left(\sqrt{n\,i(\theta_0)}(\hat{\theta}_n(\boldsymbol{X}) - \theta_0), Z\right) \leq \frac{1}{\sqrt{n}}\left(2 + \frac{\mathrm{E}\left[|T(X_1) - A'(\theta_0)|^3\right]}{|A''(\theta_0)|^{\frac{3}{2}}}\right)$$
$$+ B^* \frac{\sqrt{n}}{2\sqrt{|A''(\theta_0)|}} \mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right)^2\right] \mathbb{1}_{\{A'(\theta)\neq\theta\}}.$$

We now proceed to give the upper bound for the generalised Gamma distribution, which is a generalisation of the Gamma distribution with three parameters $\theta, p, d > 0$ and from now is denoted by $GG(\theta, d, p)$. With $\Gamma(.)$ denoting as usual the gamma function, the probability density function for $x > 0$ is

$$f(x|\theta) = \frac{p x^{d-1}}{\theta^d} \frac{\exp\left\{-\left(\frac{x}{\theta}\right)^p\right\}}{\Gamma\left(\frac{d}{p}\right)}. \tag{5.18}$$

The generalised gamma distribution includes many other known distributions as special cases. If $d = p$ then the generalized gamma distribution becomes the Weibull distribu-

tion. Furthermore, if $p = 1$, we get the gamma distribution (if in addition $d = 1$, then the exponential distribution with mean θ is obtained). When d and p are known and the scale θ_0 is the unknown parameter of interest, the MLE for θ_0 exists, it is unique and $\hat{\theta}_n(\boldsymbol{X}) = \left(\frac{p}{nd}\sum_{i=1}^n x_i^p\right)^{\frac{1}{p}}$.

Corollary 5.2. *Let $X_1, \ldots, X_n$ be i.i.d. random variables from the generalized gamma GG(θ, d, p) distribution, where the shape parameters $d, p > 0$ are known and the scale parameter $\theta_0 > 0$ is the unknown parameter of interest. Then,*

$$\begin{aligned} d_{bW}\left(\sqrt{n\,i(\theta_0)}(\hat{\theta}_n(\boldsymbol{X}) - \theta_0), Z\right) &\leq \frac{1}{\sqrt{n}}\left(2 + \left(3 + 6\frac{p}{d}\right)^{\frac{3}{4}}\right) \\ &+ \left(1 - 2\left(\frac{p}{nd}\right)^{\frac{1}{p}}\frac{\Gamma\left(\frac{nd+1}{p}\right)}{\Gamma\left(\frac{nd}{p}\right)} + \left(\frac{p}{nd}\right)^{\frac{2}{p}}\frac{\Gamma\left(\frac{nd+2}{p}\right)}{\Gamma\left(\frac{nd}{p}\right)}\right)\mathbb{1}_{\{\{d\neq 1\}\cup\{p\neq 1\}\}} \\ &\times \left[8 + \frac{\sqrt{nd\,p}|p-1|}{2}\left(\frac{1}{2^{p-2}}\mathbb{1}_{\{p<2\}} + \left(\frac{3}{2}\right)^{p-2}\mathbb{1}_{\{p\geq 2\}}\right)\right]. \end{aligned} \tag{5.19}$$

Proof. Rearranging the terms in (5.18) yields

$$f(x|\theta) = \exp\left\{-\left(\frac{x}{\theta}\right)^p + \log p - d\log\theta + (d-1)\log x - \log\left(\Gamma\left(\frac{d}{p}\right)\right)\right\}$$

and in the terminology of single-parameter exponential families, $B = (0,\infty)$, $\Theta = (0,\infty)$, $T(x) = x^p$, $k(\theta) = -\frac{1}{\theta^p}$, $A(\theta) = d\log\theta$ and $S(x) = \log p + (d-1)\log x - \log\left(\Gamma\left(\frac{d}{p}\right)\right)$. Simple steps yield

$$l(\theta_0;\boldsymbol{x}) = -\frac{1}{\theta_0^p}\sum_{i=1}^n x_i^p + n\log p - nd\log\theta_0 + (d-1)\log\left(\prod_{i=1}^n x_i\right) - n\log\left(\Gamma\left(\frac{d}{p}\right)\right)$$

and

$$l'(\theta_0;\boldsymbol{x}) = \frac{p}{\theta_0^{p+1}}\sum_{i=1}^n x_i^p - n\frac{d}{\theta_0}.$$

Thus, $l'(\theta;\boldsymbol{x}) = 0$ if and only if $\theta = \hat{\theta}_n(\boldsymbol{x}) = \left(\frac{p}{nd}\sum_{i=1}^n x_i^p\right)^{\frac{1}{p}}$. Also, $l''(\hat{\theta}_n(\boldsymbol{x});\boldsymbol{x}) = -n\frac{pd}{[\hat{\theta}_n(\boldsymbol{x})]^2} < 0$, which shows that the MLE exists and is unique. To obtain the bound,

the results of Corollary 5.1 are employed. For the generalised gamma distribution, $D(\theta_0) = q(\theta_0) = \frac{d}{p}\theta_0^p$ and thus

$$\mathrm{E}\left[|T(X) - D(\theta_0)|^3\right] = \mathrm{E}\left[\left|X^p - \frac{d}{p}\theta_0^p\right|^3\right].$$

This third absolute moment is complicated to calculate. Using the change-of-variable technique, it can be easily shown that if $X \sim \mathrm{GG}(\theta_0, d, p)$, then $X^p \sim \text{Gamma}\left(\frac{d}{p}, \frac{1}{\theta_0^p}\right)$. Therefore, Hölder's inequality gives

$$\begin{aligned}
&\mathrm{E}\left[\left|X^p - \frac{d}{p}\theta_0^p\right|^3\right] \leq \left[\mathrm{E}\left[\left(X^p - \frac{d}{p}\theta_0^p\right)^4\right]\right]^{\frac{3}{4}} \\
&= \left[\mathrm{E}\left(X^{4p}\right) + \left(\frac{d\theta_0^p}{p}\right)^4 + 6\left(\frac{d\theta_0^p}{p}\right)^2 \mathrm{E}\left(X^{2p}\right) - 4\left(\frac{d\theta_0^p}{p}\right)^3 \mathrm{E}\left(X^p\right) - 4\frac{d}{p}\theta_0^p \mathrm{E}\left(X^{3p}\right)\right]^{\frac{3}{4}} \\
&= \left[\theta_0^{4p}\frac{d}{p}\left(6 + 3\frac{d}{p}\right)\right]^{\frac{3}{4}} = \theta_0^{3p}\left(\frac{d}{p}\right)^{\frac{3}{4}}\left(6 + 3\frac{d}{p}\right)^{\frac{3}{4}}. \qquad (5.20)
\end{aligned}$$

In addition, since $k(\theta) = -\frac{1}{\theta^p}, A(\theta) = d\log\theta$ and $D(\theta) = \frac{d}{p}\theta^p$

$$\begin{aligned}
\frac{|k'(\theta_0)|}{\sqrt{|A''(\theta_0) - k''(\theta_0)D(\theta_0)|}} &= \frac{p\theta_0^{-(p+1)}}{\sqrt{\left|-\frac{d}{\theta_0^2} + p(p+1)\theta_0^{-(p+2)}\frac{d}{p}\theta_0^p\right|}} \\
&= \frac{p\theta_0^{-(p+1)}}{\sqrt{\left|-\frac{d}{\theta_0^2} + \frac{d(p+1)}{\theta_0^2}\right|}} = \frac{p\theta_0^{-(p+1)}}{\sqrt{\frac{dp}{\theta_0^2}}} = \frac{\sqrt{p}}{\sqrt{d}\theta_0^p}. \qquad (5.21)
\end{aligned}$$

Using that $X_i^p \sim \text{Gamma}\left(\frac{d}{p}, \frac{1}{\theta_0^p}\right) \Rightarrow \sum_{i=1}^n X_i^p \sim \text{Gamma}\left(n\frac{d}{p}, \frac{1}{\theta_0^p}\right)$, we get

$$\begin{aligned}
&\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right)^2\right] = \left(\frac{p}{nd}\right)^{\frac{2}{p}} \mathrm{E}\left[\left(\sum_{i=1}^n X_i^p\right)^{\frac{2}{p}}\right] + \theta_0^2 - 2\left(\frac{p}{nd}\right)^{\frac{1}{p}}\theta_0 \mathrm{E}\left[\left(\sum_{i=1}^n X_i^p\right)^{\frac{1}{p}}\right] \\
&= \theta_0^2\left(\frac{p}{nd}\right)^{\frac{2}{p}} \frac{\Gamma\left(\frac{nd+2}{p}\right)}{\Gamma\left(\frac{nd}{p}\right)} + \theta_0^2 - 2\theta_0^2\left(\frac{p}{nd}\right)^{\frac{1}{p}} \frac{\Gamma\left(\frac{nd+1}{p}\right)}{\Gamma\left(\frac{nd}{p}\right)}
\end{aligned}$$

$$= \theta_0^2 \left(1 - 2 \left(\frac{p}{nd} \right)^{\frac{1}{p}} \frac{\Gamma\left(\frac{nd+1}{p} \right)}{\Gamma\left(\frac{nd}{p} \right)} + \left(\frac{p}{nd} \right)^{\frac{2}{p}} \frac{\Gamma\left(\frac{nd+2}{p} \right)}{\Gamma\left(\frac{nd}{p} \right)} \right). \tag{5.22}$$

We now check whether $D''(\theta)$ is uniformly bounded. Simple steps yield

$$\left| D''(\theta) \right| = \left| d(p-1)\theta^{p-2} \right|,$$

which is not bounded for $\theta \in (0, \infty)$. However for any $\theta_0 \in \Theta$ there exists $0 < \varepsilon = \varepsilon(\theta_0)$ such that $\sup_{\theta : |\theta - \theta_0| < \varepsilon} |D''(\theta)| < \infty$. Therefore, this example falls under the first case of Corollary 5.1. Below, we see that the supremum depends on the value of p:

$$\begin{aligned} \sup_{\theta : |\theta - \theta_0| < \varepsilon} \left| d(p-1)\theta^{p-2} \right| &= d|p-1| \sup_{\theta : |\theta - \theta_0| < \varepsilon} \left| \theta^{p-2} \right| \\ &= d|p-1| \begin{cases} (\theta_0 - \varepsilon)^{p-2}, & \text{if } 0 < p < 2 \\ \\ (\theta_0 + \varepsilon)^{p-2}, & \text{if } p \geq 2. \end{cases} \end{aligned} \tag{5.23}$$

We now explain how $0 < \varepsilon$ has been chosen such that $\left| \theta^{p-2} \right|$ is bounded for all $\theta \in (\theta_0 - \varepsilon, \theta_0 + \varepsilon) \cap \Theta$. We have that $\left| \theta^{p-2} \right|$ in (5.23) is bounded for $p < 2$ when $0 < \varepsilon < \theta_0$. In both cases of (5.23), $\sup_{\theta : |\theta - \theta_0| < \varepsilon} \left| d(p-1)\theta^{p-2} \right|$ gets smaller as ε approaches zero. On the other hand, the second term of the bound in (5.16) becomes smaller as ε gets larger. An optimisation process with respect to ε is quite complicated and we choose $\varepsilon = \frac{\theta_0}{2}$, which is far away from both zero and θ_0. Using this value of ε and applying the results of (5.20), (5.21), (5.22) and (5.23) to (5.16) , we obtain the result in (5.19). □

Remark 5.3. ***(1)*** *The bound in* (5.19) *is* $\mathcal{O}\left(\frac{1}{\sqrt{n}} \right)$. *This is not obvious due to the term* $\left(1 - 2 \left(\frac{p}{nd} \right)^{\frac{1}{p}} \frac{\Gamma\left(\frac{nd+1}{p} \right)}{\Gamma\left(\frac{nd}{p} \right)} + \left(\frac{p}{nd} \right)^{\frac{2}{p}} \frac{\Gamma\left(\frac{nd+2}{p} \right)}{\Gamma\left(\frac{nd}{p} \right)} \right)$. *Using the following Taylor expansion for a*

ratio of Gamma functions (see Tricomi and Erdélyi (1951))

$$\frac{\Gamma(z+a)}{\Gamma(z+b)} = z^{a-b}\left(1+\frac{(a-b)(a+b-1)}{2z}+\mathcal{O}\left(|z|^{-2}\right)\right)$$

for large z $\left(\text{here, } \frac{nd}{p}\right)$ *and bounded a and b, we can see that this term is of order* $\frac{1}{n}$*, leading to the overall order of* $\frac{1}{\sqrt{n}}$.

(2) *The indicator function in* (5.19) *comes from the fact that* $q(\theta)=\theta \Leftrightarrow d, p=1$.

5.2.2 Bounds for the exponential distribution

In this subsection, we consider as an example one special case of the generalised gamma distribution; the exponential distribution. This choice of distribution allows us to compare the results of this section with those in Section 3.3. First, the canonical form of the distribution is treated and then we will change the parameterisation.

The canonical case: Exp(θ)

We start with $X_1,\ldots,X_n$ exponentially distributed i.i.d. random variables with scale parameter $\theta>0$. The probability density function is given in (3.25). With this in hand, it is easily noticed that $D(\theta_0)=q(\theta_0)=\frac{A'(\theta_0)}{k'(\theta_0)}=-\frac{1}{\theta_0}$. Equation (5.21) shows that for $p=d=1$

$$\frac{|k'(\theta_0)|}{\sqrt{|A''(\theta_0)-k''(\theta_0)D(\theta_0)|}}=\theta_0.$$

The rest of the quantities in the bound of Corollary 5.1 have already been calculated in Section 3.3 and for $\varepsilon=\frac{\theta_0}{2}$, we obtain

$$\begin{aligned} d_{bW}\left(\sqrt{n\,i(\theta_0)}(\hat{\theta}_n(\boldsymbol{X})-\theta_0),Z\right) <& \frac{4.41456}{\sqrt{n}}+8\frac{(n+2)}{(n-1)(n-2)} \\ &+8\frac{\sqrt{n}(n+2)}{(n-1)(n-2)}=\mathcal{O}\left(\frac{1}{\sqrt{n}}\right). \end{aligned} \tag{5.24}$$

Remark 5.4. ***(1)*** *The bound is* $\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$.

(2) *The result in* (5.24) *is the same as in* (3.26)*, which is not a surprise in view of Remark 5.2.*

The non-canonical case: Exp$\left(\frac{1}{\theta}\right)$

We now examine again the case where $X_1,\ldots,X_n$ are i.i.d. random variables from Exp$\left(\frac{1}{\theta}\right)$. The probability density function is found in (3.29). It is straightforward that $D(\theta)=q(\theta)=\theta$ and $|D''(\theta)|=0$, thus uniformly bounded, meaning that we are under the second case of Corollary 5.1 and we will use the bound in (5.17) which does not require to define $\varepsilon=\varepsilon(\theta_0)>0$. The fact that $D(\theta)=\theta$ makes the last term of the bound in (5.17) vanish. Having also that $\mathrm{E}\,|T(X)-D(\theta_0)|^3=\mathrm{E}\,|X-\theta_0|^3<2.41456\theta_0^3$ and

$$\frac{|k'(\theta_0)|}{\sqrt{|A''(\theta_0)-k''(\theta_0)D(\theta_0)|}}=\frac{1}{\theta_0}$$

yields

$$d_{bW}\left(\sqrt{n\,i(\theta_0)}(\hat{\theta}_n(\boldsymbol{X})-\theta_0),Z\right)<\frac{4.41456}{\sqrt{n}}. \tag{5.25}$$

Remark 5.5. ***(1)*** *The order of the bound in terms of the sample size is, as expected,* $\frac{1}{\sqrt{n}}$.

(2) *For the same distribution, Corollary 3.3 gives*

$$\frac{4.41456}{\sqrt{n}}+\frac{8}{n}+\frac{2}{\sqrt{n}}+\frac{1}{\sqrt{n}}\left(80\left[3\left(\frac{2}{n}+1\right)\right]^{\frac{1}{2}}\right),$$

showing that the bound using the Delta method is an improvement of this bound. In Remark 3.4, it is also made clear that because in this specific example the MLE is a sum of random variables then someone can use directly Lemma 2.2 and bound the distributional distance of interest, which also gives the bound in (5.25).

Chapter 6

Locally dependent random variables

In this chapter, we assess the normal approximation of the MLE under the presence of a local dependence structure between the random variables. Stein's method for normal approximation is employed, but this time we use the results of Lemma 2.3 since they are appropriate due to the structure of the random variables. We work under assumptions that allow us to express the bound in terms of the Wasserstein distance which is defined in (2.5). A general approach is first developed to get upper bounds on the Wasserstein distance between the distribution of the suitably scaled MLE and the standard normal distribution. An example of normally distributed locally dependent random variables serves as an illustration of these results. After that, the asymptotic normality result is given under rather restrictive but sufficient regularity conditions and a new bound is obtained. Bounds are also derived when the MLE can not be expressed in a closed form.

The notion of an m-dependent sequence of random variables $\{X_i, i \in \mathbb{N}\}$ has already been introduced in Section 2.3. The sets $\{X_j, j \leq i\}$ and $\{X_j, j > i+m\}$ are independent. The joint density function of $X_1, X_2, \ldots, X_n$ is

$$f(\boldsymbol{x}|\theta) = L(\theta;\boldsymbol{x}) = f(x_1;\theta)f(x_2|x_1;\theta)\dots f(x_n|x_{n-1},\dots,x_{n-m};\theta)$$
$$= f(x_1;\theta)\prod_{i=2}^{n} f(x_i|x_{i-1},\dots,x_{m_i^*};\theta), \tag{6.1}$$

where $m_i^* = \max\{i-m, 1\}$.

In Section 6.1 we explain the process of finding an upper bound on the Wasserstein distance between the distribution of the suitably scaled MLE and the standard normal distribution, that holds in a general framework. The quantity we are interested in is split into two terms with the one being bounded using Stein's method and the other using alternative techniques based mainly on Taylor expansions. These results are applied in Section 6.2 to the example of locally dependent normally distributed random variables. In Section 6.3, we discuss on a set of sufficient regularity conditions for the asymptotic normality of the MLE to hold. Under these, rather strict, conditions a different bound is obtained for the Wasserstein distance between the distribution of the MLE and its limiting normal distribution. Under further assumptions, Section 6.4 gives upper bounds for situations where the MLE can not be defined analytically.

6.1 A general approach

Under the assumption of the parameter space being an open and connected interval, Proposition 2.1 is sufficient for the existence and uniqueness of $\hat{\theta}_n(\boldsymbol{X})$. Our purpose is to obtain an upper bound on the Wasserstein distance between the distribution of an appropriately scaled MLE and the standard normal distribution. The results of Lemma 2.3 will be applied to a sequence $\{\xi_i, i = 1,2,\dots,n\}$ of $2m$-dependent random variables. We denote by

$$M_{1j} := \max\{1, j-2m\} \qquad M_{2j} := \min\{n, j+2m\}$$
$$K_{1j} := \max\{1, j-4m\} \qquad K_{2j} := \min\{n, j+4m\}.$$

In addition, the dependency neighbourhoods, A_j and B_j, as defined in (LD) on p.20 are

$$\begin{aligned} A_j &= \{M_{1j}, M_{1j}+1, \ldots, M_{2j}-1, M_{2j}\} \\ B_j &= \{K_{1j}, K_{1j}+1, \ldots, K_{2j}-1, K_{2j}\}. \end{aligned} \tag{6.2}$$

Having that $\forall i \in \{1,2,\ldots,n\}$, $|A_i|$ and $|B_i|$ denote the number of elements in the sets A_i and B_i, respectively, then

$$|A_i| \leq 4m+1, \qquad |B_i| \leq 8m+1.$$

Figure 6.1 below is a graphical representation of our notation related to the $2m$-dependence structure.

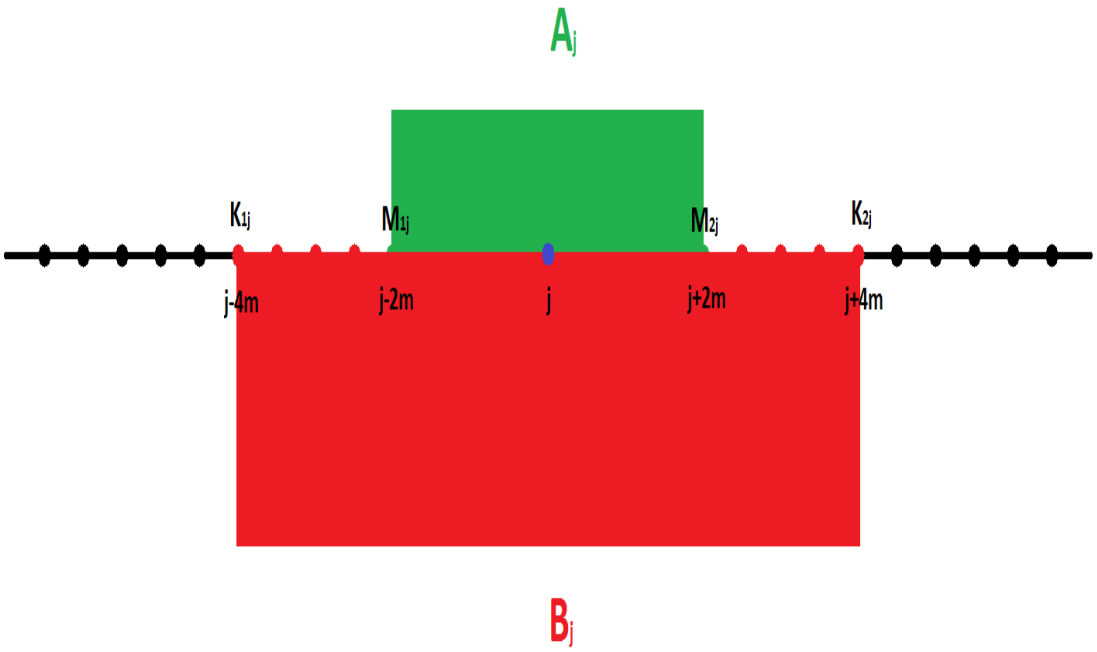

Fig. 6.1: Structure of a $2m$-dependent sequence.

In this section, for $\{X_i, i = 1,2,\ldots,n\}$ being an m-dependent sequence, we work under the following assumptions:

(A.D.1) The log-likelihood function is three times differentiable with uniformly bounded third derivative in $\theta \in \Theta$, $(x_1, x_2, \ldots, x_n) \in S$. We denote the supremum by

$$S_d(n) := \sup_{\substack{\theta \in \Theta \\ \boldsymbol{x} \in S}} \left| l^{(3)}(\theta; \boldsymbol{x}) \right| < \infty. \tag{6.3}$$

(A.D.2) $\mathrm{E}\left[\frac{\mathrm{d}}{\mathrm{d}\theta} \log f(X_1|\theta)\right] = \mathrm{E}\left[\frac{\mathrm{d}}{\mathrm{d}\theta} \log f(X_i|X_{i-1}, \ldots, X_{i-m}; \theta)\right] = 0$, for $i = 2, 3, \ldots, n$.

(A.D.3) With θ_0, as usual, denoting the true value of the unknown parameter,

$$\sqrt{n}\mathrm{E}\left[\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right] \xrightarrow[n\to\infty]{} 0. \tag{6.4}$$

(A.D.4) The limit of the reciprocal of $n\mathrm{Var}\left(\hat{\theta}_n(\boldsymbol{X})\right)$ exists and from now on, unless otherwise stated,

$$0 < i_2(\theta_0) = \lim_{n\to\infty} \frac{1}{n\mathrm{Var}(\hat{\theta}_n(\boldsymbol{X}))}. \tag{6.5}$$

The assumption (A.D.1) is stronger than (R3) of Section 2.2 which was used in the case of independent random variables. Using uniform boundedness of the third derivative of the log-likelihood function in (A.D.1) allows us to get bounds on the Wasserstein distance related to the MLE. The following theorem gives the bound.

Theorem 6.1. *Let $\{X_i, i = 1, 2, \ldots, n\}$ be an m-dependent sequence of identically distributed random variables with probability density (or mass) function $f(x_i|x_{i-1}, \ldots, x_{i-m}; \theta)$, where $\theta \in \Theta$ and $(x_1, x_2, \ldots, x_n) \in S$, the support of the joint probability density (or mass) function. Assume that $\hat{\theta}_n(\boldsymbol{X})$ exists and is unique. In addition, assume that (A.D.1)-(A.D.4) hold and that $\mathrm{Var}[l'(\theta_0; \boldsymbol{X})] > 0$. Let*

$$\alpha := \alpha(\theta_0, n) := \sqrt{\frac{\mathrm{Var}\left(l'(\theta_0; \boldsymbol{X})\right)}{\mathrm{Var}\left(\hat{\theta}_n(\boldsymbol{X})\right)}}, \tag{6.6}$$

which is assumed to be finite and not equal to zero. In addition, we denote by

$$\xi_1 = \frac{\mathrm{d}}{\mathrm{d}\theta} \log f(X_1|\theta)\Big|_{\theta=\theta_0} \sqrt{\frac{n}{\mathrm{Var}(l'(\theta_0;\boldsymbol{X}))}}$$

and for $i = 2,3,\ldots,n$,

$$\xi_i = \frac{\mathrm{d}}{\mathrm{d}\theta} \log f(X_i|X_{i-1},\ldots,X_{i-m};\theta)\Big|_{\theta=\theta_0} \sqrt{\frac{n}{\mathrm{Var}(l'(\theta_0;\boldsymbol{X}))}}.$$

For $Z \sim \mathrm{N}(0,1)$,

$$\begin{aligned}
&d_W\left(\sqrt{n\,i_2(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right), Z\right) \\
&\leq \frac{2}{n^{\frac{3}{2}}} \sum_{i=1}^{n} \sum_{j\in A_i} \sum_{k\in B_i} \left[\mathrm{E}\left((\xi_i)^4\right) \mathrm{E}\left((\xi_j)^4\right) \mathrm{E}\left((\xi_k)^4\right)\right]^{\frac{1}{4}} \\
&+ \frac{2}{n^{\frac{3}{2}}} \sum_{i=1}^{n} \sum_{j\in A_i} \sum_{k\in B_i} \left[\mathrm{E}\left((\xi_i)^2\right) \mathrm{E}\left((\xi_j)^2\right) \mathrm{E}\left((\xi_k)^2\right)\right]^{\frac{1}{2}} \\
&+ \frac{1}{n^{\frac{3}{2}}} \sum_{i=1}^{n} |A_i| \sum_{j\in A_i} \left[\mathrm{E}\left((\xi_i)^2\right) \mathrm{E}\left((\xi_j)^4\right)\right]^{\frac{1}{2}} \\
&+ \left| \frac{\sqrt{n\,i_2(\theta_0)\mathrm{Var}[l'(\theta_0;\boldsymbol{X})]}}{\alpha} - 1 \right| \\
&+ \frac{S_d(n)\sqrt{n\,i_2(\theta_0)}}{2\alpha} \mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right] \\
&+ \frac{\sqrt{n\,i_2(\theta_0)}}{\alpha} \sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]} \sqrt{\mathrm{E}\left[\left(l''(\theta_0;\boldsymbol{X})+\alpha\right)^2\right]}. \qquad (6.7)
\end{aligned}$$

Proof. By the definition of the MLE and (A.D.1), $l'\left(\hat{\theta}_n(\boldsymbol{x});\boldsymbol{x}\right) = 0$. With $R_1(\theta_0;\boldsymbol{X})$ as in (3.9), a second order Taylor expansion gives that

$$\begin{aligned}
&\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right) l''(\theta_0;\boldsymbol{X}) = -l'(\theta_0;\boldsymbol{X}) - R_1(\theta_0;\boldsymbol{X}) \\
&\Rightarrow \left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)\left(\alpha + l''(\theta_0;\boldsymbol{X}) - \alpha\right) = -l'(\theta_0;\boldsymbol{X}) - R_1(\theta_0;\boldsymbol{X}) \\
&\Rightarrow -\alpha\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right) = -l'(\theta_0;\boldsymbol{X}) - R_1(\theta_0;\boldsymbol{X}) - \left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)\left(l''(\theta_0;\boldsymbol{X})+\alpha\right)
\end{aligned}$$

Multiplying both sides by $-\frac{\sqrt{n\,i_2(\theta_0)}}{\alpha}$,

$$\sqrt{n\,i_2(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)=\frac{\sqrt{n\,i_2(\theta_0)}}{\alpha}\Bigg[l'(\theta_0;\boldsymbol{X})+R_1(\theta_0;\boldsymbol{X}) + \left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)\left(l''(\theta_0;\boldsymbol{X})+\alpha\right)\Bigg]. \tag{6.8}$$

Applying the triangle inequality,

$$\begin{aligned}&\left|\mathrm{E}\left[h\left(\sqrt{n\,i_2(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)\right)\right]-\mathrm{E}[h(Z)]\right| \\ &\leq \left|\mathrm{E}\left[h\left(\frac{\sqrt{n\,i_2(\theta_0)}l'(\theta_0;\boldsymbol{X})}{\alpha}\right)\right]-\mathrm{E}[h(Z)]\right| \qquad (6.9)\\ &+\left|\mathrm{E}\left[h\left(\sqrt{n\,i_2(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)\right)-h\left(\frac{\sqrt{n\,i_2(\theta_0)}l'(\theta_0;\boldsymbol{X})}{\alpha}\right)\right]\right|. \qquad (6.10)\end{aligned}$$

Step 1: Bound for (6.9). Let, for ease of presentation $l'(\theta_0;\boldsymbol{X})=\sum_{i=1}^n\tilde{\xi}_i$, where

$$\begin{aligned}\tilde{\xi}_1&=\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta)\Big|_{\theta=\theta_0},\\ \tilde{\xi}_i&=\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_i|X_{i-1},\ldots,X_{i-m};\theta)\Big|_{\theta=\theta_0}\ \text{ for }\ i=2,3,\ldots,n.\end{aligned}$$

Assumption (A.D.2) ensures that $\tilde{\xi}_i, i=1,2,\ldots,n$ have mean zero. Furthermore, for some function $g:\mathbb{R}^{m+1}\to\mathbb{R}$, it holds that $\tilde{\xi}_i=g(X_i,X_{i-1},\ldots,X_{i-m})$ and taking into account that $\{X_i,i=1,2,\ldots,n\}$ is an m-dependent sequence, we conclude that $\left\{\tilde{\xi}_i,i=1,2,\ldots,n\right\}$ forms a $2m$-dependent sequence. Define now

$$W:=\frac{l'(\theta_0;\boldsymbol{X})}{\sqrt{\mathrm{Var}\left[l'(\theta_0;\boldsymbol{X})\right]}}=\sum_{i=1}^n\left(\frac{\xi_i}{\sqrt{n}}\right), \tag{6.11}$$

with

$$\xi_i=\tilde{\xi}_i\sqrt{\frac{n}{\mathrm{Var}\left[l'(\theta_0;\boldsymbol{X})\right]}},\ \forall i\in\{1,2,\ldots,n\}.$$

It follows that $\left\{\frac{\xi_i}{\sqrt{n}}, i = 1,2,\ldots,n\right\}$ is a random $2m$-dependent sequence with mean zero and also $\mathrm{Var}(W) = 1$. In addition, (LD) is satisfied with A_j and B_j as in (6.2). A simple triangle inequality gives that

$$(6.9) \leq |\mathrm{E}[h(W)] - \mathrm{E}[h(Z)]| \tag{6.12}$$

$$+ \left|\mathrm{E}\left[h\left(\frac{\sqrt{n\,i_2(\theta_0)}\,l'(\theta_0;\boldsymbol{X})}{\alpha}\right) - h(W)\right]\right|. \tag{6.13}$$

Since the assumptions of Lemma 2.3 are satisfied for W as in (6.11), one can directly use (2.15) in order to find an upper bound for (6.12). For (6.13), a first order Taylor expansion of $h\left(\frac{\sqrt{n\,i_2(\theta_0)}\,l'(\theta_0;\boldsymbol{X})}{\alpha}\right)$ about W yields

$$\begin{aligned} &h\left(\frac{\sqrt{n\,i_2(\theta_0)}\,l'(\theta_0;\boldsymbol{X})}{\alpha}\right) - h\left(\frac{l'(\theta_0;\boldsymbol{X})}{\sqrt{\mathrm{Var}(l'(\theta_0;\boldsymbol{X}))}}\right) \\ &= \left(\frac{\sqrt{n\,i_2(\theta_0)}\,l'(\theta_0;\boldsymbol{X})}{\alpha} - \frac{l'(\theta_0;\boldsymbol{X})}{\sqrt{\mathrm{Var}(l'(\theta_0;\boldsymbol{X}))}}\right) h'(t_1(\boldsymbol{X})), \end{aligned}$$

where $t_1(\boldsymbol{X})$ is between $\frac{\sqrt{n\,i_2(\theta_0)}\,l'(\theta_0;\boldsymbol{X})}{\alpha}$ and $\frac{l'(\theta_0;\boldsymbol{X})}{\sqrt{\mathrm{Var}(l'(\theta_0;\boldsymbol{X}))}}$. Therefore,

$$\begin{aligned} (6.13) &\leq \|h'\| \left|\frac{\sqrt{n\,i_2(\theta_0)}}{\alpha} - \frac{1}{\sqrt{\mathrm{Var}(l'(\theta_0;\boldsymbol{X}))}}\right| \mathrm{E}\left|l'(\theta_0;\boldsymbol{X})\right| \\ &\leq \|h'\| \left|\frac{\sqrt{n\,i_2(\theta_0)\mathrm{Var}(l'(\theta_0;\boldsymbol{X}))}}{\alpha} - 1\right|. \end{aligned} \tag{6.14}$$

Thus, for $h \in H_W$ as in (2.4), we conclude that

$$\begin{aligned} (6.9) \leq \frac{2}{n^{\frac{3}{2}}}\left[\sum_{i=1}^{n}(\mathrm{E}|\xi_i\eta_i\tau_i|) + \sum_{i=1}^{n}(|\mathrm{E}(\xi_i\eta_i)|\,\mathrm{E}|\tau_i|)\right] + \frac{1}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\mathrm{E}\left|\xi_i\eta_i^2\right| \\ + \left|\frac{\sqrt{n\,i_2(\theta_0)\mathrm{Var}(l'(\theta_0;\boldsymbol{X}))}}{\alpha} - 1\right|, \end{aligned} \tag{6.15}$$

with η_i and τ_i as in (2.14). The absolute expectations in (6.15) can be difficult to bound and we express the first three quantities of the above bound in terms of more easily

calculable terms. For the first term in (6.15), using Hölder's inequality

$$\begin{aligned}
\mathrm{E}|\xi_i\eta_i\tau_i| &= \mathrm{E}\left|\xi_i \sum_{j\in A_i}\xi_j \sum_{k\in B_i}\xi_k\right| \\
&\leq \sum_{j\in A_i}\sum_{k\in B_i}\mathrm{E}\left|\xi_i\xi_j\xi_k\right| \\
&\leq \sum_{j\in A_i}\sum_{k\in B_i}\left[\mathrm{E}\left(|\xi_i|^3\right)\mathrm{E}\left(|\xi_j|^3\right)\mathrm{E}\left(|\xi_k|^3\right)\right]^{\frac{1}{3}} \\
&\leq \sum_{j\in A_i}\sum_{k\in B_i}\left[\mathrm{E}\left((\xi_i)^4\right)\mathrm{E}\left((\xi_j)^4\right)\mathrm{E}\left((\xi_k)^4\right)\right]^{\frac{1}{4}}. \qquad (6.16)
\end{aligned}$$

For the second term of the bound in (6.15), the Cauchy-Schwarz inequality yields

$$\begin{aligned}
|\mathrm{E}(\xi_i\eta_i)|\,\mathrm{E}|\tau_i| &= \left|\mathrm{E}\left(\xi_i\sum_{j\in A_i}\xi_j\right)\right|\mathrm{E}\left|\sum_{k\in B_i}\xi_k\right| \\
&\leq \sum_{j\in A_i}\mathrm{E}\left|\xi_i\xi_j\right|\sum_{k\in B_i}\mathrm{E}|\xi_k| \\
&\leq \sum_{j\in A_i}\sum_{k\in B_i}\left[\mathrm{E}\left((\xi_i)^2\right)\mathrm{E}\left((\xi_j)^2\right)\mathrm{E}\left((\xi_k)^2\right)\right]^{\frac{1}{2}}. \qquad (6.17)
\end{aligned}$$

For the third term, we also employ Jensen's inequality to get that

$$\left(\sum_{i\in J}|a_i|\right)^z \leq J^{z-1}\sum_{i\in J}|a_i|^z,\ \forall a_i\in\mathbb{R}\ \text{ and }\ z\in\mathbb{N}$$

and therefore

$$\begin{aligned}
\mathrm{E}\left|\xi_i\eta_i^2\right| &= \mathrm{E}\left|\xi_i\left(\sum_{j\in A_i}\xi_j\right)^2\right| \\
&\leq |A_i|\,\mathrm{E}\left|\xi_i\sum_{j\in A_i}\xi_j^2\right| \\
&\leq |A_i|\sum_{j\in A_i}\mathrm{E}\left|\xi_i\xi_j^2\right| \\
&\leq |A_i|\sum_{j\in A_i}\left[\mathrm{E}\left((\xi_i)^2\right)\mathrm{E}\left((\xi_j)^4\right)\right]^{\frac{1}{2}}. \qquad (6.18)
\end{aligned}$$

The results in (6.16), (6.17) and (6.18) yield

$$
\begin{aligned}
(6.12) &\leq \frac{2}{n^{\frac{3}{2}}}\left[\sum_{i=1}^{n}\left(\mathrm{E}|\xi_i\eta_i\tau_i|\right)+\sum_{i=1}^{n}\left(|\mathrm{E}(\xi_i\eta_i)|\,\mathrm{E}|\tau_i|\right)\right]+\frac{1}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\mathrm{E}\left|\xi_i\eta_i^2\right| \\
&\leq \frac{2}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j\in A_i}\sum_{k\in B_i}\left[\mathrm{E}\left((\xi_i)^4\right)\mathrm{E}\left(\left(\xi_j^4\right)\right)\mathrm{E}\left((\xi_k)^4\right)\right]^{\frac{1}{4}} \\
&\quad+\frac{2}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j\in A_i}\sum_{k\in B_i}\left[\mathrm{E}\left((\xi_i)^2\right)\mathrm{E}\left((\xi_j)^2\right)\mathrm{E}\left((\xi_k)^2\right)\right]^{\frac{1}{2}} \\
&\quad+\frac{1}{n^{\frac{3}{2}}}\sum_{i=1}^{n}|A_i|\sum_{j\in A_i}\left[\mathrm{E}\left((\xi_i)^2\right)\mathrm{E}\left((\xi_j)^4\right)\right]^{\frac{1}{2}}. \qquad (6.19)
\end{aligned}
$$

The bound for (6.12) is now obviously a function only of $\mathrm{E}\left(\xi_i^2\right)$ and $\mathrm{E}\left(\xi_i^4\right)$.

Step 2: Bound for (6.10). As in Chapter 3 where the case of independent random variables was treated, the main tool used here is again Taylor expansions. However, comparing (A.D.1) with (R3) from Chapter 2, we do not need to use conditional expectations in this case as the third derivative of the log-likelihood function is assumed to be uniformly bounded in the whole parameter space and not just in a neighbourhood of it. This also means that no positive constant ε is used. For ease of presentation, let

$$
\begin{aligned}
\tilde{C}(\theta_0) &= \tilde{C}(h,\theta_0;\boldsymbol{X}) := h\left(\sqrt{n\,i_2(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)\right)-h\left(\frac{\sqrt{n\,i_2(\theta_0)}\,l'(\theta_0;\boldsymbol{X})}{\alpha}\right) \\
&= h\left(\frac{\sqrt{n\,i_2(\theta_0)}\left[l'(\theta_0;\boldsymbol{X})+R_1(\theta_0;\boldsymbol{X})+\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)\left(l''(\theta_0;\boldsymbol{X})+\alpha\right)\right]}{\alpha}\right) \\
&\quad-h\left(\frac{\sqrt{n\,i_2(\theta_0)}\,l'(\theta_0;\boldsymbol{X})}{\alpha}\right),
\end{aligned}
$$

from the result in (6.8). Using now a first order Taylor expansion of $h\left(\frac{\sqrt{n\,i_2(\theta_0)}\left[l'(\theta_0;\boldsymbol{X})+R_1(\theta_0;\boldsymbol{X})+\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)\left(l''(\theta_0;\boldsymbol{X})+\alpha\right)\right]}{\alpha}\right)$ about $\frac{\sqrt{n\,i_2(\theta_0)}\,l'(\theta_0;\boldsymbol{X})}{\alpha}$ yields

$$
\begin{aligned}
(6.10) = \left|\mathrm{E}\left[\tilde{C}(\theta_0)\right]\right| \leq \frac{\sqrt{n\,i_2(\theta_0)}}{\alpha}\|h'\|\Bigg(&\mathrm{E}\left[\frac{1}{2}\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\left|l^{(3)}(\theta^*;\boldsymbol{X})\right|\right] \\
&+\mathrm{E}\left|\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)\left(l''(\theta_0;\boldsymbol{X})+\alpha\right)\right|\Bigg)
\end{aligned}
$$

$$\leq \frac{\sqrt{n\,i_2(\theta_0)}}{\alpha}\|h'\| \left(\frac{S_d(n)}{2} \mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0 \right)^2 \right] \right.$$
$$\left. + \sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0 \right)^2 \right]} \sqrt{\mathrm{E}\left[\left(l''(\theta_0; \boldsymbol{X}) + \alpha \right)^2 \right]} \right), \quad (6.20)$$

where for the last step we used Cauchy-Schwarz inequality and $S_d(n)$ is as in (6.3). We conclude that (6.14), (6.19) and (6.20) yield, for $h \in H_W$, the assertion of the theorem as expressed in (6.7). □

6.2 Locally dependent normal random variables

To illustrate the general results, as an example assume that we have a sequence $\{S_1, S_2, \ldots, S_n\}$ of random variables where for $k \in \mathbb{Z}^+$ and $\forall j \in \{1, 2, \ldots, n\}$,

$$S_j = \sum_{i=(j-1)k}^{jk} X_i,$$

for $X_i, i = 0, 1, 2, \ldots, nk$ i.i.d. random variables from the $\mathrm{N}(\mu, \sigma^2)$ distribution with $\mu = \theta \in \mathbb{R}$ being the unknown parameter and σ^2 is known. It is easy to see that $\{S_i\}_{i=1,2,\ldots,n}$ is a 1-dependent sequence of random variables since for $\delta \in \mathbb{Z} \setminus \{0\}$

$$\mathrm{Cov}(S_i, S_{i+\delta}) = \begin{cases} \mathrm{Var}(X_1) = \sigma^2, & \text{if } |\delta| = 1 \\ 0, & \text{if } |\delta| > 1. \end{cases}$$

Furthermore,

$$S_i \sim \mathrm{N}((k+1)\theta, (k+1)\sigma^2) \quad (6.21)$$

as it is a sum of $k+1$ independent normally distributed random variables with mean θ and variance σ^2.

It is standard, see (Casella and Berger, 2002, p.177), that if Y_1, Y_2 are two normally distributed random variables with $Y_1 \sim \mathrm{N}(\mu_1, \sigma_1^2)$ and $Y_2 \sim \mathrm{N}(\mu_2, \sigma_2^2)$, then

$$(Y_2|Y_1 = y_1) \sim \mathrm{N}\left(\mu_2 + \frac{\sigma_2}{\sigma_1}\rho(y_1 - \mu_1),\, (1-\rho^2)\sigma_2^2\right),$$

where $\rho = \frac{\mathrm{Cov}(Y_1,Y_2)}{\sqrt{\mathrm{Var}(Y_1)\mathrm{Var}(Y_2)}}$ denotes the correlation between Y_1 and Y_2. In our example,

$$\rho = \frac{\mathrm{Cov}(S_{i-1}, S_i)}{\sqrt{\mathrm{Var}(S_{i-1})\mathrm{Var}(S_i)}} = \frac{\sigma^2}{(k+1)\sigma^2} = \frac{1}{k+1},\ \forall i \in \{2,3,\ldots,n\}.$$

Therefore, for $i = 2,3,\ldots,n$

$$(S_i|S_{i-1} = s_{i-1}) \sim \mathrm{N}\left((k+1)\theta + \frac{1}{k+1}\left(s_{i-1} - (k+1)\theta\right), \frac{k(k+2)}{k+1}\sigma^2\right). \tag{6.22}$$

The likelihood function for the parameter θ under $\boldsymbol{S} = (S_1, S_2, \ldots, S_n)$ is

$$\begin{aligned}
L(\theta; \boldsymbol{S}) &= f(S_1|\theta)\prod_{i=2}^{n} f(S_i|S_{i-1};\theta) \\
&= \frac{1}{\sqrt{2\pi(k+1)\sigma^2}}\exp\left\{-\frac{(S_1-(k+1)\theta)^2}{2(k+1)\sigma^2}\right\} \\
&\quad \times \prod_{i=2}^{n}\frac{\sqrt{k+1}}{\sqrt{2\pi k(k+2)\sigma^2}}\exp\left\{-\frac{(k+1)\left(S_i - \left((k+1)\,\theta + \frac{1}{k+1}(S_{i-1}-(k+1)\theta)\right)\right)^2}{2k(k+2)\sigma^2}\right\} \\
&= \frac{(k+1)^{\frac{n-1}{2}}}{\sqrt{2\pi(k+1)\sigma^2}(2\pi k(k+2)\sigma^2)^{\frac{n-1}{2}}}\exp\left\{-\frac{(S_1-(k+1)\theta)^2}{2(k+1)\sigma^2}\right. \\
&\qquad \left. -\frac{k+1}{2k(k+2)\sigma^2}\sum_{i=2}^{n}\left(S_i - \left((k+1)\,\theta + \frac{1}{k+1}(S_{i-1}-(k+1)\theta)\right)\right)^2\right\}.
\end{aligned} \tag{6.23}$$

Having this closed-form expression for the likelihood allows us to derive the MLE under this local dependence structure. For

$$
\begin{aligned}
l(\theta;\boldsymbol{S}) &:= \log(L(\theta;\boldsymbol{S})) \\
l'(\theta;\boldsymbol{S}) &:= \frac{\mathrm{d}}{\mathrm{d}\theta} l(\theta;\boldsymbol{S}) \\
l''(\theta;\boldsymbol{S}) &:= \frac{\mathrm{d}^2}{\mathrm{d}\theta^2} l(\theta;\boldsymbol{S}) \\
l^{(3)}(\theta;\boldsymbol{S}) &:= \frac{\mathrm{d}^3}{\mathrm{d}\theta^3} l(\theta;\boldsymbol{S}),
\end{aligned}
$$

using (6.23) we obtain that

$$
\begin{aligned}
l(\theta;\boldsymbol{S}) &= \frac{n-1}{2}\log(k+1) - \frac{1}{2}\log(2\pi(k+1)\sigma^2) - \frac{n-1}{2}\log(2\pi k(k+2)\sigma^2) \\
&\quad - \frac{(S_1-(k+1)\theta)^2}{2(k+1)\sigma^2} \\
&\quad - \frac{k+1}{2k(k+2)\sigma^2}\sum_{i=2}^{n}\left(S_i - \left((k+1)\,\theta + \frac{1}{k+1}(S_{i-1}-(k+1)\theta)\right)\right)^2 \\
l'(\theta;\boldsymbol{S}) &= \frac{S_1-(k+1)\theta}{\sigma^2} + \frac{k+1}{(k+2)\sigma^2}\sum_{i=2}^{n}\left\{S_i - \left((k+1)\theta + \frac{1}{k+1}(S_{i-1}-(k+1)\theta)\right)\right\} \\
&= \frac{1}{(k+2)\sigma^2}\left\{k\sum_{i=1}^{n} S_i + S_1 + S_n - (k+1)(nk+2)\theta\right\} \\
l''(\theta;\boldsymbol{S}) &= -\frac{(nk+2)(k+1)}{(k+2)\sigma^2} \\
l^{(3)}(\theta;\boldsymbol{S}) &= 0.
\end{aligned} \tag{6.24}
$$

To find the solution of $l'(\theta,\boldsymbol{S}) = 0$ with respect to θ, using (6.24),

$$
\begin{aligned}
& l'(\theta;\boldsymbol{S}) = 0 \Leftrightarrow k\sum_{i=1}^{n} S_i + S_1 + S_n - (k+1)(nk+2)\theta = 0 \\
& \Leftrightarrow (nk+2)(k+1)\theta = k\sum_{i=1}^{n} S_i + S_1 + S_n \\
& \Leftrightarrow \theta = \hat{\theta}_n(\boldsymbol{S}) = \frac{k\sum_{i=1}^{n} S_i + S_1 + S_n}{(nk+2)(k+1)}.
\end{aligned} \tag{6.25}
$$

The solution is unique and also $l''(\theta;\boldsymbol{S}) < 0$. Hence, the unique MLE for θ is $\hat{\theta}_n(\boldsymbol{S})$ as in (6.25). The following Corollary gives the upper bound on the Wasserstein distance between the distribution of $\hat{\theta}_n(\boldsymbol{S})$ and the normal distribution.

Corollary 6.1. *Let $S_1, S_2, \ldots S_n$ be a 1-dependent sequence of random variables with $S_i \sim \mathrm{N}((k+1)\theta, (k+1)\sigma^2)$. The conditions (A.D.1)-(A.D.4) hold. For $Z \sim \mathrm{N}(0,1)$ and $i_2(\theta_0) = \frac{(k+1)^2}{(k+3)\sigma^2}$,*

$$
\begin{aligned}
&d_W\left(\sqrt{n\, i_2(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{S}) - \theta_0\right), Z\right) \\
&\leq 339(n-5)\left[\frac{k(k+1)(k+2)}{(nk^3+(3n+2)k^2+10k+2)}\right]^{\frac{3}{2}} \\
&\quad + \frac{(k+1)^{\frac{3}{2}}(k+2)^{\frac{3}{2}}}{(nk^3+(3n+2)k^2+10k+2)^{\frac{3}{2}}}\left\{\left(1+3^{\frac{3}{4}}\right)\left(2\sqrt{k+2}(37k+2)+4\sqrt{k}(61k+8)\right)\right. \\
&\qquad\qquad \left. +\sqrt{3}\left(3\sqrt{k+2}(k+1)+\sqrt{k}(91k+18)\right)\right\} \\
&\quad + \left|\left(1-\frac{2}{nk+2}\right)\left[\frac{k+3+\frac{2}{n}+\frac{10}{nk}+\frac{2}{nk^2}}{k+3}\right]^{\frac{1}{2}} - 1\right|.
\end{aligned}
$$

Remark 6.1. *The order of the bound with respect to the sample size is $\frac{1}{\sqrt{n}}$.*

Proof. We first check that the assumptions (A.D.1)-(A.D.4) are satisfied. The first assumption is satisfied from (6.24). Now, from (6.21) and (6.22) the derivatives of $\log f(s_1|\theta)$ and $\log f(s_i|s_{i-1};\theta)$, $i = 2,3,\ldots,n$ with respect to θ are

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(s_1|\theta) &= \frac{s_1-(k+1)\theta}{\sigma^2} \\
\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(s_i|s_{i-1};\theta) &= \frac{k+1}{(k+2)\sigma^2}\left(s_i - k\theta - \frac{s_{i-1}}{k+1}\right).
\end{aligned} \tag{6.26}
$$

Therefore,

$$
\begin{aligned}
\mathrm{E}\left[\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(S_1|\theta)\right] &= \mathrm{E}\left[\frac{S_1-(k+1)\theta}{\sigma^2}\right] = 0 \\
\mathrm{E}\left[\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(S_i|S_{i-1};\theta)\right] &= \mathrm{E}\left[\frac{k+1}{(k+2)\sigma^2}\left(S_i - k\theta - \frac{S_{i-1}}{k+1}\right)\right] = 0
\end{aligned}
$$

and thus (A.D.2) holds. The assumption (A.D.3) is also satisfied since, using (6.21) and (6.25),

$$\mathrm{E}\left[\hat{\theta}_n(\boldsymbol{S})\right] = \frac{nk(k+1)\theta_0 + 2(k+1)\theta_0}{(nk+2)(k+1)} = \theta_0.$$

To show that (A.D.4) holds, we first calculate

$$\begin{aligned}\mathrm{Var}\left[\hat{\theta}_n(\boldsymbol{S})\right] &= \frac{1}{(nk+2)^2(k+1)^2}\mathrm{Var}\left(k\sum_{i=1}^{n} S_i + S_1 + S_n\right) \\ &= \frac{1}{(nk+2)^2(k+1)^2}\left\{k^2\mathrm{Var}\left(\sum_{i=1}^{n} S_i\right) + \mathrm{Var}(S_1) + \mathrm{Var}(S_n) + 2k\mathrm{Cov}\left(S_1, \sum_{i=1}^{n} S_i\right)\right. \\ &\qquad \left. + 2k\mathrm{Cov}\left(S_n, \sum_{i=1}^{n} S_i\right)\right\}. \end{aligned} \tag{6.27}$$

We know from (6.21) that $\mathrm{Var}(S_i) = (k+1)\sigma^2, \forall i \in \{1,2,\ldots,n\}$. In addition, since $\{S_i\}_{i=1,2,\ldots,n}$ is a 1-dependent sequence of random variables,

$$\begin{aligned}\mathrm{Var}\left(\sum_{i=1}^{n} S_i\right) &= n\mathrm{Var}(S_1) + 2(n-1)\mathrm{Cov}(S_1,S_2) = n(k+1)\sigma^2 + 2(n-1)\sigma^2 \\ \mathrm{Cov}\left(S_1, \sum_{i=1}^{n} S_i\right) &= \mathrm{Var}(S_1) + \mathrm{Cov}(S_1,S_2) = (k+2)\sigma^2. \end{aligned} \tag{6.28}$$

Applying the above results of (6.28) to (6.27) gives that

$$\mathrm{Var}\left[\hat{\theta}_n(\boldsymbol{S})\right] = \frac{\sigma^2\left(nk^3 + 3nk^2 + 2k^2 + 10k + 2\right)}{(nk+2)^2(k+1)^2}. \tag{6.29}$$

Therefore,

$$\begin{aligned} i_2(\theta_0) &= \lim_{n\to\infty} \frac{1}{n\mathrm{Var}\left(\hat{\theta}_n(\boldsymbol{S})\right)} = \lim_{n\to\infty} \frac{(nk+2)^2(k+1)^2}{n\sigma^2(nk^3+3nk^2+2k^2+10k+2)} \\ &= \lim_{n\to\infty} \frac{n^2(k+1)^2\left(k^2 + \frac{4k}{n} + \frac{4}{n^2}\right)}{n^2\sigma^2\left(k^3 + 3k^2 + \frac{2k^2}{n} + \frac{10k}{n} + \frac{2}{n}\right)} = \frac{(k+1)^2}{(k+3)\sigma^2} > 0, \end{aligned} \tag{6.30}$$

which shows that (A.D.4) is satisfied. To obtain α as defined in (6.6), we calculate the variance of the score function, which is

$$\begin{aligned}\mathrm{Var}\left[l'(\theta_0;\boldsymbol{S})\right] &= \frac{1}{(k+2)^2\sigma^4}\mathrm{Var}\left(k\sum_{i=1}^{n}S_i+S_1+S_n\right)\\ &= \frac{1}{(k+2)^2\sigma^4}\left\{k^2\mathrm{Var}\left(\sum_{i=1}^{n}S_i\right)+\mathrm{Var}(S_1)+\mathrm{Var}(S_n)+2k\mathrm{Cov}\left(S_1,\sum_{i=1}^{n}S_i\right)\right.\\ &\qquad\left.+2k\mathrm{Cov}\left(S_n,\sum_{i=1}^{n}S_i\right)\right\}.\end{aligned}$$

Using (6.28), we conclude that

$$\begin{aligned}\mathrm{Var}\left[l'(\theta_0;\boldsymbol{S})\right] &= \frac{1}{(k+2)^2\sigma^2}\left[nk^2(k+1)+2k^2(n-1)+2(k+1)+4k(k+2)\right]\\ &= \frac{1}{(k+2)^2\sigma^2}\left[nk^3+3nk^2+2k^2+10k+2\right]. \qquad (6.31)\end{aligned}$$

The above result and (6.29) yield

$$\alpha = \sqrt{\frac{\mathrm{Var}\left[l'(\theta_0;\boldsymbol{S})\right]}{\mathrm{Var}\left[\hat{\theta}_n(\boldsymbol{S})\right]}} = \sqrt{\frac{(nk+2)^2(k+1)^2}{\sigma^4(k+2)^2}} = \frac{(nk+2)(k+1)}{(k+2)\sigma^2}. \qquad (6.32)$$

For $\xi_1 = \frac{\mathrm{d}}{\mathrm{d}\theta}\log f(S_1|\theta)\Big|_{\theta=\theta_0}\sqrt{\frac{n}{\mathrm{Var}[l'(\theta_0;\boldsymbol{S})]}}$, $\xi_i = \frac{\mathrm{d}}{\mathrm{d}\theta}\log f(S_i|S_{i-1};\theta)\Big|_{\theta=\theta_0}\sqrt{\frac{n}{\mathrm{Var}[l'(\theta_0;\boldsymbol{S})]}}$, $i=2,3,\ldots,n$, using (6.31) and (6.26), we get that

$$\xi_1 = \frac{\sqrt{n}(k+2)[S_1-(k+1)\theta_0]}{\sigma\sqrt{nk^3+(3n+2)k^2+10k+2}}$$

and therefore

$$\begin{aligned}\mathrm{E}\left(\xi_1^2\right) &= \frac{n(k+2)^2\mathrm{E}(S_1-(k+1)\theta)^2}{(nk^3+(3n+2)k^2+10k+2)\sigma^2} = \frac{n(k+2)^2(k+1)}{nk^3+(3n+2)k^2+10k+2}\\ \mathrm{E}\left(\xi_1^4\right) &= \frac{n^2(k+2)^4\mathrm{E}(S_1-(k+1)\theta)^4}{(nk^3+(3n+2)k^2+10k+2)^2\sigma^4} = \frac{3n^2(k+2)^4(k+1)^2}{(nk^3+(3n+2)k^2+10k+2)^2}. \qquad (6.33)\end{aligned}$$

Furthermore, for $i = 2,3,\ldots,n$,

$$\xi_i = \frac{\sqrt{n}(k+1)\left[S_i - \left((k+1)\theta + \frac{1}{k+1}\left(S_{i-1} - (k+1)\theta\right)\right)\right]}{\sigma\sqrt{nk^3+(3n+2)k^2+10k+2}}$$

so that for $i = 2,3,\ldots,n$,

$$\begin{aligned}
\mathrm{E}\left(\xi_i^2\right) &= \frac{n(k+1)^2\mathrm{E}\left[S_i - \left((k+1)\theta + \frac{1}{k+1}\left(S_{i-1} - (k+1)\theta\right)\right)\right]^2}{\sigma^2(nk^3+(3n+2)k^2+10k+2)} \\
&= \frac{nk(k+1)(k+2)}{nk^3+(3n+2)k^2+10k+2} \\
\mathrm{E}\left(\xi_i^4\right) &= \frac{n^2(k+1)^4\mathrm{E}\left[S_i - \left((k+1)\theta + \frac{1}{k+1}\left(S_{i-1} - (k+1)\theta\right)\right)\right]^4}{\sigma^4(nk^3+(3n+2)k^2+10k+2)^2} \\
&= \frac{3n^2k^2(k+1)^2(k+2)^2}{(nk^3+(3n+2)k^2+10k+2)^2}.
\end{aligned} \tag{6.34}$$

We are now ready to calculate the first three terms of the general bound (6.7) denoted from now on by

$$\begin{aligned}
Q_v = Q_v(k,n) := \frac{1}{n^{\frac{3}{2}}}\Bigg\{ & 2\sum_{j\in A_v}\sum_{l\in B_v}\left[\mathrm{E}\left((\xi_v)^2\right)\mathrm{E}\left(\left(\xi_j\right)^2\right)\mathrm{E}\left(\left(\xi_l\right)^2\right)\right]^{\frac{1}{2}} \\
& +2\sum_{j\in A_v}\sum_{l\in B_v}\left[\mathrm{E}\left((\xi_v)^4\right)\mathrm{E}\left(\left(\xi_j\right)^4\right)\mathrm{E}\left(\left(\xi_l\right)^4\right)\right]^{\frac{1}{4}} + |A_v|\sum_{j\in A_v}\left[\mathrm{E}\left((\xi_v)^2\right)\mathrm{E}\left(\left(\xi_j\right)^4\right)\right]^{\frac{1}{2}}\Bigg\}.
\end{aligned}$$

We split our approach depending on whether 1 is an element of either A_i or B_i, for $i \in \{1,2,\ldots,n\}$.

Case 1: $i = 6,7,\ldots,n$. Using the results in (6.34) and since $|A_i| \leq 5$, $|B_i| \leq 9$, $\forall i \in \{6,7,\ldots,n\}$, we obtain that

$$\begin{aligned}
Q_i &\leq \frac{1}{n^{\frac{3}{2}}}\left\{90\left[\mathrm{E}\left((\xi_2)^4\right)\right]^{\frac{3}{4}} + 90\left[\mathrm{E}\left((\xi_2)^2\right)\right]^{\frac{3}{2}} + 25\left[\mathrm{E}\left((\xi_2)^2\right)\mathrm{E}\left((\xi_2)^4\right)\right]^{\frac{1}{2}}\right\} \\
&= \left[\frac{k(k+1)(k+2)}{(nk^3+(3n+2)k^2+10k+2)}\right]^{\frac{3}{2}}\left(90(3)^{\frac{3}{4}} + 90 + 25\sqrt{3}\right) \\
&< 339\left[\frac{k(k+1)(k+2)}{(nk^3+(3n+2)k^2+10k+2)}\right]^{\frac{3}{2}}.
\end{aligned} \tag{6.35}$$

Issues arise due to ξ_1 not having the same distribution as ξ_i for $i \in \{2,3,\ldots,n\}$. There are hence five more special cases corresponding to $i = 1,2,\ldots,5$. We treat these cases separately.

Case 2: $i = 1$. For $A_1 = \{1,2,3\}$ and $B_1 = \{1,2,\ldots,5\}$, the results in (6.33) and (6.34) yield

$$\begin{aligned}
Q_1 = \frac{1}{n^{\frac{3}{2}}} \Bigg\{ & 2\left[\left[\mathrm{E}\left((\xi_1)^2\right)\right]^{\frac{3}{2}} + 6\mathrm{E}\left((\xi_1)^2\right)\left[\mathrm{E}\left((\xi_2)^2\right)\right]^{\frac{1}{2}} + 8\mathrm{E}\left((\xi_2)^2\right)\left[\mathrm{E}\left((\xi_1)^2\right)\right]^{\frac{1}{2}}\right] \\
& + 2\left[\left[\mathrm{E}\left((\xi_1)^4\right)\right]^{\frac{3}{4}} + 6\left[\mathrm{E}\left((\xi_1)^4\right)\right]^{\frac{1}{2}}\left[\mathrm{E}\left((\xi_2)^4\right)\right]^{\frac{1}{4}} + 8\left[\mathrm{E}\left((\xi_1)^4\right)\right]^{\frac{1}{4}}\left[\mathrm{E}\left((\xi_2)^4\right)\right]^{\frac{1}{2}}\right] \\
& + 3\left[\left[\mathrm{E}\left((\xi_1)^2\right)\right]^{\frac{1}{2}}\left(\left[\mathrm{E}\left((\xi_1)^4\right)\right]^{\frac{1}{2}} + 2\left[\mathrm{E}\left((\xi_2)^4\right)\right]^{\frac{1}{2}}\right)\right] \Bigg\} \\
= \frac{(k+1)^{\frac{3}{2}}(k+2)^2}{(nk^3+(3n+2)k^2+10k+2)^{\frac{3}{2}}} & \Bigg\{ 2\left[(k+2)+6\sqrt{k(k+2)}+8k\right] \\
& \qquad + 2(3)^{\frac{3}{4}}\left[(k+2)+6\sqrt{k(k+2)}+8k\right] + 3\sqrt{3}(2k+2) \Bigg\} \\
= \frac{2(k+1)^{\frac{3}{2}}(k+2)^2}{(nk^3+(3n+2)k^2+10k+2)^{\frac{3}{2}}} & \Bigg\{ \left(9k+2+6\sqrt{k(k+2)}\right)\left(1+3^{\frac{3}{4}}\right) \\
& \qquad + 3\sqrt{3}(k+1) \Bigg\}. \qquad (6.36)
\end{aligned}$$

Case 3: $i = 2$. For $A_2 = \{1,2,3,4\}$ and $B_2 = \{1,2,\ldots,6\}$, the results in (6.33) and (6.34) yield

$$\begin{aligned}
Q_2 = \frac{1}{n^{\frac{3}{2}}} \Bigg\{ & 2\left[\mathrm{E}\left((\xi_1)^2\right)\left[\mathrm{E}\left((\xi_2)^2\right)\right]^{\frac{1}{2}} + 8\mathrm{E}\left((\xi_2)^2\right)\left[\mathrm{E}\left((\xi_1)^2\right)\right]^{\frac{1}{2}} + 15\left[\mathrm{E}\left((\xi_2)^2\right)\right]^{\frac{3}{2}}\right] \\
& + 2\left[\left[\mathrm{E}\left((\xi_1)^4\right)\right]^{\frac{1}{2}}\left[\mathrm{E}\left((\xi_2)^4\right)\right]^{\frac{1}{4}} + 8\left[\mathrm{E}\left((\xi_2)^4\right)\right]^{\frac{1}{2}}\left[\mathrm{E}\left((\xi_1)^4\right)\right]^{\frac{1}{4}} + 15\left[\mathrm{E}\left((\xi_2)^4\right)\right]^{\frac{3}{4}}\right] \\
& + 4\left[\left[\mathrm{E}\left((\xi_2)^2\right)\right]^{\frac{1}{2}}\left(\left[\mathrm{E}\left((\xi_1)^4\right)\right]^{\frac{1}{2}} + 3\left[\mathrm{E}\left((\xi_2)^4\right)\right]^{\frac{1}{2}}\right)\right] \Bigg\}
\end{aligned}$$

$$= \frac{\sqrt{k}(k+1)^{\frac{3}{2}}(k+2)^{\frac{3}{2}}}{(nk^3+(3n+2)k^2+10k+2)^{\frac{3}{2}}} \left\{ 2\left[k+2+8\sqrt{k(k+2)}+15k\right] \right.$$
$$\left. +2(3)^{\frac{3}{4}}\left[k+2+8\sqrt{k(k+2)}+15k\right]+4\sqrt{3}(4k+2) \right\}$$
$$= \frac{4\sqrt{k}(k+1)^{\frac{3}{2}}(k+2)^{\frac{3}{2}}}{(nk^3+(3n+2)k^2+10k+2)^{\frac{3}{2}}} \left\{ \left(8k+1+4\sqrt{k(k+2)}\right)\left(1+3^{\frac{3}{4}}\right) \right.$$
$$\left. +2\sqrt{3}(2k+1) \right\}. \tag{6.37}$$

Case 4: $i=3$. Following the same steps as in Case 3, now for $A_3=\{1,2,\ldots,5\}$ and $B_3=\{1,2,\ldots,7\}$, the results in (6.33) and (6.34) give that

$$Q_3 = \frac{\sqrt{k}(k+1)^{\frac{3}{2}}(k+2)^{\frac{3}{2}}}{(nk^3+(3n+2)k^2+10k+2)^{\frac{3}{2}}} \left\{ 2\left(25k+2+10\sqrt{k(k+2)}\right)\left(1+3^{\frac{3}{4}}\right) \right.$$
$$\left. +5\sqrt{3}(5k+2) \right\}. \tag{6.38}$$

Case 5: $i=4$. For $A_4=\{2,3,\ldots,6\}, B_4=\{1,2,\ldots,8\}$, the results in (6.33) and (6.34) yield

$$Q_4 = \frac{1}{n^{\frac{3}{2}}} \left\{ 2\left[5\mathrm{E}\left((\xi_2)^2\right)\left[\mathrm{E}\left((\xi_1)^2\right)\right]^{\frac{1}{2}}+35\left[\mathrm{E}\left((\xi_2)^2\right)\right]^{\frac{3}{2}}\right] \right.$$
$$+2\left[5\left[\mathrm{E}\left((\xi_2)^4\right)\right]^{\frac{1}{2}}\left[\mathrm{E}\left((\xi_1)^4\right)\right]^{\frac{1}{4}}+35\left[\mathrm{E}\left((\xi_2)^4\right)\right]^{\frac{3}{4}}\right]$$
$$\left. +25\left[\left[\mathrm{E}\left((\xi_2)^2\right)\right]^{\frac{1}{2}}\left[\mathrm{E}\left((\xi_2)^4\right)\right]^{\frac{1}{2}}\right] \right\}$$
$$= \frac{5k(k+1)^{\frac{3}{2}}(k+2)^{\frac{3}{2}}}{(nk^3+(3n+2)k^2+10k+2)^{\frac{3}{2}}} \left\{ 2\left[\sqrt{k+2}+7\sqrt{k}\right]\left(1+3^{\frac{3}{4}}\right)+5\sqrt{3k} \right\}. \tag{6.39}$$

Case 6: $i = 5$. Following the same steps as in Case 5, now for $A_5 = \{3,4,\ldots,7\}$ and $B_5 = \{1,2,\ldots,9\}$, the results in (6.33) and (6.34) yield

$$Q_5 = \frac{5k(k+1)^{\frac{3}{2}}(k+2)^{\frac{3}{2}}}{(nk^3+(3n+2)k^2+10k+2)^{\frac{3}{2}}}\left\{2\left[\sqrt{k+2}+8\sqrt{k}\right]\left(1+3^{\frac{3}{4}}\right)+5\sqrt{3k}\right\}. \quad (6.40)$$

The sum of the results of (6.36), (6.37), (6.38), (6.39) and (6.40) with $(n-5)$ times the bound in (6.35) consists an upper bound for the first three terms of the general upper bound as expressed in (6.7). For the fourth term of the general upper bound, (6.30), (6.31) and (6.32) yield

$$\begin{aligned}\left|\frac{\sqrt{n\,i_2(\theta_0)\mathrm{Var}[l'(\theta_0;\boldsymbol{X})]}}{\alpha}-1\right| &= \left|\frac{\left[\frac{n(k+1)^2\left(nk^3+3nk^2+2k^2+10k+2\right)}{k+3}\right]^{\frac{1}{2}}}{(nk+2)(k+1)}-1\right| \\ &= \left|\frac{nk}{nk+2}\left[\frac{k+3+\frac{2}{n}+\frac{10}{nk}+\frac{2}{nk^2}}{k+3}\right]^{\frac{1}{2}}-1\right| \\ &= \left|\left(1-\frac{2}{nk+2}\right)\left[\frac{k+3+\frac{2}{n}+\frac{10}{nk}+\frac{2}{nk^2}}{k+3}\right]^{\frac{1}{2}}-1\right|. \end{aligned} \quad (6.41)$$

The fifth term of the bound in (6.7) involves the calculation of $S_d(n)$, which is equal to zero from (6.24). Therefore, the fifth term of the general upper bound vanishes for this example. For the last term we have from (6.24) that

$$\mathrm{E}\left[l''(\theta_0;\boldsymbol{S})\right] = -\frac{(nk+2)(k+1)}{(k+2)\sigma^2} = -\alpha$$

and therefore

$$\sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{S})-\theta_0\right)^2\right]\mathrm{E}\left[\left(l''(\theta_0;\boldsymbol{S})+\alpha\right)^2\right]} = \sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{S})-\theta_0\right)^2\right]\mathrm{Var}\left[l''(\theta_0;\boldsymbol{S})\right]} = 0.$$

The results of Case 1 - Case 6 and (6.41) give the assertion of the corollary. □

6.3 An alternative bound

In this section, under stricter regularity conditions we give the asymptotic normality result for the distribution of the MLE obtained from locally dependent random variables. Under these new conditions, an alternative bound on the Wasserstein distance between the exact distribution of the MLE and the normal distribution for locally dependent random variables is given.

We work under the following regularity conditions, taken from Bhat (1974):

(R.D.1) The range of x_k given $x_1, x_2, \ldots, x_{k-1}$, for $k = 2, 3, \ldots, n$, does not depend on θ and differentiation with respect to θ may be carried out under the integral sign, integration being with respect to x_k, up to the third order, for $\log f(x_k|x_{k-1}, \ldots, x_1; \theta)$. These derivatives are also assumed to be continuous in $\theta \in \Theta$. In addition,

$$\mathrm{E}_{k-1}\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_k|X_{k-1},\ldots,X_1;\theta)\right)\Big|_{\theta=\theta_0} = 0, \quad \text{almost surely}$$

and we denote by

$$i_k(\theta_0) := \mathrm{E}_{k-1}\left[\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_k|X_{k-1},\ldots,X_1;\theta)\right)^2\right]\Big|_{\theta=\theta_0}$$

where E_{k-1} denotes the conditional expectation given $X_1, X_2, \ldots, X_{k-1}$.

(R.D.2) The third derivative $\frac{\mathrm{d}^3}{\mathrm{d}\theta^3}\log f(X_k|X_{k-1},\ldots,X_1;\theta)$ is bounded in probability uniformly in $\theta \in \Theta$, $(X_1, X_2, \ldots, X_k)$.

(R.D.3) $\forall \theta_0 \in \Theta, \bar{i}_n(\theta_0) = \frac{1}{n}\sum_{j=1}^{n} i_j(\theta_0) \to \bar{i}(\theta_0) > 0$ as $n \to \infty$, where $\bar{i}(\theta_0)$ is a constant almost surely and θ_0 is the true value of θ.

(R.D.4) $\frac{1}{n^{1+\frac{\delta}{2}}}\sum_{k=2}^{n}\mathrm{E}_{k-1}\left|\frac{\partial \log f(x_k|x_{k-1},\ldots,x_1;\theta)}{\partial\theta}\right|^{2+\delta} \to 0$ a.s. for some $\delta > 0$.

(R.D.5) $\limsup_{n\to\infty}\left\{\frac{1}{n^2}\sum_{k=1}^{n}\mathrm{Var}\left(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(X_k|X_{k-1},\ldots,X_1;\theta)\right)\Big|_{\theta=\theta_0}\right\}<\infty.$

The condition (R.D.3) implies that $\sum_{k=1}^{n} i_k(\theta_0)\to\infty$ as $n\to\infty$, which is essential for the CLT to hold for the dependent random variables $\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_k|X_{k-1},\ldots,X_1;\theta)\Big|_{\theta=\theta_0}$. From now on, unless otherwise stated

$$\bar{i}(\theta_0) := \lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n} i_j(\theta_0). \tag{6.42}$$

Also, if $i_k(\theta_0)$ is an absolute constant $\forall k\in\{1,2,\ldots,n\}$ then (R.D.4) is not necessary as shown in Silvey (1961).

Assuming that the MLE exists and is unique and that the conditions (R.D.1)-(R.D.5) are satisfied, then Theorem 3 in Bhat (1974) gives the result for the asymptotic normality of the MLE in the case of dependent random variables. The proof can be found in Bhat (1974).

Theorem 6.2. *(Bhat, 1974). Let $X_1,X_2,\ldots,X_n$ be dependent and identically distributed random variables with probability density (or mass) functions $f(x_1|\theta)$, $f(x_i|x_{i-1},\ldots,x_1;\theta)$, $i=2,3,\ldots,n$ where $\theta\in\Theta\subset\mathbb{R}$, with Θ denoting the open parameter space. Assume that the regularity conditions (R.D.1)-(R.D.5) hold. Let $Z\sim\mathrm{N}(0,1)$. Then,*

$$\sqrt{n\bar{i}(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)\xrightarrow[n\to\infty]{\mathrm{d}} Z. \tag{6.43}$$

The following theorem gives an upper bound on the Wasserstein distance between the distribution of the MLE and its limiting, under (R.D.1)-(R.D.5), normal distribution in the case of *m*-dependent sequence of random variables.

Theorem 6.3. *Let $\{X_i, i=1,2,\ldots,n\}$ be an m-dependent sequence of identically distributed random variables with probability density (or mass) functions $f(x_1|\theta)$, $f(x_i|x_{i-1},\ldots,x_1;\theta)$, $i=2,3,\ldots,n$, where $\theta\in\Theta\subset\mathbb{R}$ and $(X_1,X_2,\ldots,X_n)\in S$, the support of the joint probability density (or mass) function. Assume that the regularity*

conditions (R.D.1)-(R.D.5) are satisfied and also that $\hat{\theta}_n(\boldsymbol{X})$ *exists and is unique. Let* $\bar{i}(\theta_0)$ *as in* (6.42)*,* $\mathrm{Var}\left[l'(\theta_0;\boldsymbol{X})\right] > 0$*,*

$$\xi_1 = \frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta)\Big|_{\theta=\theta_0}\sqrt{\frac{n}{\mathrm{Var}\left[l'(\theta_0;\boldsymbol{X})\right]}}$$

and for $i = 2,3,\ldots,n$*,*

$$\xi_i = \frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_i|X_{i-1},\ldots,X_{i-m};\theta)\Big|_{\theta=\theta_0}\sqrt{\frac{n}{\mathrm{Var}\left[l'(\theta_0;\boldsymbol{X})\right]}}.$$

With $S_d(n)$ *as in* (6.3) *and for* $Z \sim \mathrm{N}(0,1)$*,*

$$\begin{aligned}
&d_W\left(\sqrt{n\,\bar{i}(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right),Z\right)\\
&\leq \frac{2}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j\in A_i}\sum_{k\in B_i}\left[\mathrm{E}\left((\xi_i)^4\right)\mathrm{E}\left((\xi_j)^4\right)\mathrm{E}\left((\xi_k)^4\right)\right]^{\frac{1}{4}}\\
&+\frac{2}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j\in A_i}\sum_{k\in B_i}\left[\mathrm{E}\left((\xi_i)^2\right)\mathrm{E}\left((\xi_j)^2\right)\mathrm{E}\left((\xi_k)^2\right)\right]^{\frac{1}{2}}\\
&+\frac{1}{n^{\frac{3}{2}}}\sum_{i=1}^{n}|A_i|\sum_{j\in A_i}\left[\mathrm{E}\left((\xi_i)^2\right)\mathrm{E}\left((\xi_j)^4\right)\right]^{\frac{1}{2}}\\
&+\left|\frac{\sqrt{\mathrm{Var}\left[l'(\theta_0;\boldsymbol{X})\right]}}{\sqrt{n\,\bar{i}(\theta_0)}}-1\right|+\frac{S_d(n)}{2\sqrt{n\,\bar{i}(\theta_0)}}\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]\\
&+\frac{1}{\sqrt{n\,\bar{i}(\theta_0)}}\sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]}\sqrt{\mathrm{E}\left[n\,\bar{i}(\theta_0)+l''(\theta_0;\boldsymbol{X})\right]^2}. \qquad (6.44)
\end{aligned}$$

Remark 6.2. *In order for the above bound to approach zero with rate* $\sqrt{n}$ *as the sample size increases, we require that* $\left|\frac{\sqrt{\mathrm{Var}(l'(\theta_0;\boldsymbol{X}))}}{\sqrt{n\,\bar{i}(\theta_0)}}-1\right| = o\left(\frac{1}{\sqrt{n}}\right)$.

Proof. By the definition of the MLE and (R.D.1), $l'\left(\hat{\theta}_n(\boldsymbol{x});\boldsymbol{x}\right) = 0$. A similar approach as the one in Chapter 3, based on a second order Taylor expansion, yields

$$\sqrt{n\,\bar{i}(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{x})-\theta_0\right) = \frac{l'(\theta_0;\boldsymbol{x})+R_1(\theta_0;\boldsymbol{x})+R_2(\theta_0;\boldsymbol{x})}{\sqrt{n\,\bar{i}(\theta_0)}}, \qquad (6.45)$$

for $\bar{i}(\theta_0)$ as in (6.42). Also, $R_1(\theta_0;\boldsymbol{x})$ and $R_2(\theta_0;\boldsymbol{x})$ are as in (3.9) and (3.10), respectively. The triangle inequality gives that

$$\begin{aligned}
&\left|\mathrm{E}\left[h\left(\sqrt{n\bar{i}(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)\right)\right]-\mathrm{E}[h(Z)]\right| \\
&\leq \left|\mathrm{E}\left[h\left(\frac{l'(\theta_0;\boldsymbol{X})}{\sqrt{n\bar{i}(\theta_0)}}\right)\right]-\mathrm{E}[h(Z)]\right| \qquad (6.46)\\
&\quad+\left|\mathrm{E}\left[h\left(\sqrt{n\bar{i}(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)\right)-h\left(\frac{l'(\theta_0;\boldsymbol{X})}{\sqrt{n\bar{i}(\theta_0)}}\right)\right]\right|. \qquad (6.47)
\end{aligned}$$

Step 1: Bound for (6.46). With W as in (6.11), the steps are the same as in the proof of Theorem 6.1 and lead to

$$\begin{aligned}
(6.46) &\leq \left|\mathrm{E}[h(W)]-\mathrm{E}[h(Z)]\right| \qquad (6.48)\\
&\quad+\left|\mathrm{E}\left[h\left(\frac{l'(\theta_0;\boldsymbol{X})}{\sqrt{n\bar{i}(\theta_0)}}\right)-h(W)\right]\right|. \qquad (6.49)
\end{aligned}$$

The assumptions of Lemma 2.3 hold for (6.48) and thus we directly use (2.15) to bound this term. For (6.49) using a first order Taylor expansion of $h\left(\frac{l'(\theta_0;\boldsymbol{X})}{\sqrt{n\bar{i}(\theta_0)}}\right)$ about W and the Cauchy-Schwarz inequality, we have that

$$\begin{aligned}
(6.49) &\leq \|h'\|\left|\frac{1}{\sqrt{n\bar{i}(\theta_0)}}-\frac{1}{\sqrt{\mathrm{Var}(l'(\theta_0;\boldsymbol{X}))}}\right|\mathrm{E}|l'(\theta_0;\boldsymbol{X})| \\
&\leq \|h'\|\left|\frac{\sqrt{\mathrm{Var}(l'(\theta_0;\boldsymbol{X}))}}{\sqrt{n\bar{i}(\theta_0)}}-1\right|. \qquad (6.50)
\end{aligned}$$

Step 2: Bound for (6.47). The approach is similar to the proof of Theorem 6.1. Let

$$\begin{aligned}
C_d = C_d(h,\theta_0;\boldsymbol{X}) &:= h\left(\sqrt{n\bar{i}(\theta_0)}(\hat{\theta}_n(\boldsymbol{X})-\theta_0)\right)-h\left(\frac{l'(\theta_0;\boldsymbol{X})}{\sqrt{n\bar{i}(\theta_0)}}\right) \\
&= h\left(\frac{l'(\theta_0;\boldsymbol{X})+R_1(\theta_0;\boldsymbol{X})+R_2(\theta_0;\boldsymbol{X})}{\sqrt{n\bar{i}(\theta_0)}}\right)-h\left(\frac{l'(\theta_0;\boldsymbol{X})}{\sqrt{n\bar{i}(\theta_0)}}\right)
\end{aligned}$$

using (6.45). For $t(\boldsymbol{X})$ between $\frac{l'(\theta_0;\boldsymbol{X})}{\sqrt{n\,\bar{i}(\theta_0)}}$ and $\frac{l'(\theta_0;\boldsymbol{X})+R_1(\theta_0;\boldsymbol{x})+R_2(\theta_0;\boldsymbol{X})}{\sqrt{n\,\bar{i}(\theta_0)}}$, a first order Taylor expansion of $h\left(\frac{l'(\theta_0;\boldsymbol{X})+R_1(\theta_0;\boldsymbol{X})+R_2(\theta_0;\boldsymbol{X})}{\sqrt{n\,\bar{i}(\theta_0)}}\right)$ about $\frac{l'(\theta_0)}{\sqrt{n\,\bar{i}(\theta_0)}}$ yields

$$
\begin{aligned}
(6.47) = |\mathrm{E}[C_d]| &\leq \left[\mathrm{E}\left|\frac{R_1(\theta_0,\boldsymbol{X})+R_2(\theta_0,\boldsymbol{X})}{\sqrt{n\,\bar{i}(\theta_0)}}h'(t(\boldsymbol{X}))\right|\right] \\
&\leq \frac{\|h'\|}{2\sqrt{n\,\bar{i}(\theta_0)}}\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\left|l^{(3)}(\theta^*;\boldsymbol{X})\right|\right] \\
&\quad + \frac{\|h'\|}{\sqrt{n\,\bar{i}(\theta_0)}}\mathrm{E}\left[\left|\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)\left(n\,\bar{i}(\theta_0)+l''(\theta_0;\boldsymbol{X})\right)\right|\right].
\end{aligned}
$$

Using now (R.D.2), the Cauchy-Schwarz inequality and (6.3) give

$$
\begin{aligned}
(6.47) \leq{}& \frac{\|h'\|S_d(n)}{2\sqrt{n\,\bar{i}(\theta_0)}}\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right] \\
&+ \frac{\|h'\|}{\sqrt{n\,\bar{i}(\theta_0)}}\sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]}\sqrt{\mathrm{E}\left[n\,\bar{i}(\theta_0)+l''(\theta_0;\boldsymbol{X})\right]^2}. \qquad (6.51)
\end{aligned}
$$

We conclude that (6.50), (6.51) and the application of Lemma 2.3 to (6.48) gives for $h \in H_W$ the result of the theorem as expressed in (6.44). □

Remark 6.3. *The bound in* (6.44) *is similar to the result in* (6.7) *and holds under the rather restrictive conditions (R.D.1)-(R.D.5). Notice the difference between $i_2(\theta_0)$ used in Theorem 6.1 with $\bar{i}(\theta_0)$ of Theorem 6.3 as defined in (R.D.3). The limit $i_2(\theta_0)$ requires the calculation of* $\mathrm{Var}\left(\hat{\theta}_n(\boldsymbol{X})\right)$*, while $\bar{i}(\theta_0)$ is independent of the MLE. However, $\bar{i}(\theta_0)$ is not necessarily easy to calculate.*

6.4 Results for not analytically known MLE

The results of Theorems 6.1 and 6.3 can be used only in cases where a closed-form expression of the MLE is available. Thus, as in Section 3.5, a solution for situations where the MLE exists, but does not have a closed-form expression is desirable. Further assumptions are made in order to ensure both the validity of the procedure to be followed

and that a meaningful upper bound is obtained. Depending on these assumptions we split the results of this section into two categories. The first category requires the support of the distribution to be bounded, while for the second one we put the boundedness condition on the parameter space instead of the support of the distribution.

6.4.1 Bounded support

In addition to the regularity conditions (R.D.1)-(R.D.5), the results presented in this subsection hold under the following assumptions:

(F.A.1) The support of the distribution is bounded;

(F.A.2) The sample size, n, satisfies

$$n > \frac{[\|x\| S_d(n)]^{\frac{2}{3}}}{\bar{i}(\theta_0)}.$$

This ensures that $1 - \frac{\|x\| S_d(n)}{[n\bar{i}(\theta_0)]^{\frac{3}{2}}} > 0$ for $S_d(n)$ as in (6.3).

For ease of presentation, we denote by

$$\begin{aligned}
D_1 &:= D_1(\theta_0, x, n) = 1 - \frac{\|x\| S_d(n)}{[n\bar{i}(\theta_0)]^{\frac{3}{2}}} \\
\gamma_1 &:= \gamma_1(\theta_0, x, n) = \frac{4\|x\|}{n^{\frac{3}{2}}} \sum_{i=1}^{n} \sum_{j \in A_i} \sum_{k \in B_i} \left[\mathrm{E}\left((\xi_i)^4 \right) \mathrm{E}\left((\xi_j)^4 \right) \mathrm{E}\left((\xi_k)^4 \right) \right]^{\frac{1}{4}} \\
&+ \frac{4\|x\|}{n^{\frac{3}{2}}} \sum_{i=1}^{n} \sum_{j \in A_i} \sum_{k \in B_i} \left[\mathrm{E}\left((\xi_i)^2 \right) \mathrm{E}\left((\xi_j)^2 \right) \mathrm{E}\left((\xi_k)^2 \right) \right]^{\frac{1}{2}} \\
&+ \frac{2\|x\|}{n^{\frac{3}{2}}} \sum_{i=1}^{n} |A_i| \sum_{j \in A_i} \left[\mathrm{E}\left((\xi_i)^2 \right) \mathrm{E}\left((\xi_j)^4 \right) \right]^{\frac{1}{2}} + 2\|x\| \left| \frac{\sqrt{\mathrm{Var}(l'(\theta_0; \boldsymbol{X}))}}{\sqrt{n\bar{i}(\theta_0)}} - 1 \right|. \quad (6.52)
\end{aligned}$$

The theorem that follows gives an upper bound for the MSE of the MLE, which is then used to bound the Wasserstein distance between the distribution of the MLE and the normal distribution in cases where the MLE is not analytically known and (F.A.1), (F.A.2) hold.

Theorem 6.4. *Let $\{X_i, i = 1,2,\ldots,n\}$ be an m-dependent sequence of random variables with joint probability density (or mass) function $f(\boldsymbol{x}|\theta)$ as in* (6.1)*. Assume that the regularity conditions (R.D.1)-(R.D.5), as well as the assumptions (F.A.1), (F.A.2) are satisfied. Assume that the MLE exists and is unique. Then $A_d = A_d(\theta_0, n)$ is an upper bound for $\sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right)^2\right]}$, where*

$$A_d = \frac{1}{D_1}\left\{ \frac{\|x\|\sqrt{\mathrm{E}\left[\left(l''(\theta_0;\boldsymbol{X}) + n\bar{i}(\theta_0)\right)^2\right]}}{[n\bar{i}(\theta_0)]^{\frac{3}{2}}} + \left[\frac{\|x\|^2 \mathrm{E}\left[\left(l''(\theta_0;\boldsymbol{X}) + n\bar{i}(\theta_0)\right)^2\right]}{[n\bar{i}(\theta_0)]^3} + \frac{D_1(1+\gamma_1)}{n\bar{i}(\theta_0)}\right]^{\frac{1}{2}} \right\},$$

with D_1 and γ_1 as in (6.52)*. Assuming that $\mathrm{Var}\left[l'(\theta_0;\boldsymbol{X})\right] > 0$, then for*

$$\xi_1 = \frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta)\Big|_{\theta=\theta_0}\sqrt{\frac{n}{\mathrm{Var}\left[l'(\theta_0;\boldsymbol{X})\right]}}$$

and

$$\xi_i = \frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_i|X_{i-1},\ldots,X_{i-m};\theta)\Big|_{\theta=\theta_0}\sqrt{\frac{n}{\mathrm{Var}\left[l'(\theta_0;\boldsymbol{X})\right]}}, i = 2,3,\ldots,n,$$

we have that

$$\begin{aligned}
&d_W\left(\sqrt{n\bar{i}(\theta_0)}(\hat{\theta}_n(\boldsymbol{X}) - \theta_0), Z\right) \\
&\leq \frac{2}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j\in A_i}\sum_{k\in B_i}\left[\mathrm{E}\left((\xi_i)^4\right)\mathrm{E}\left((\xi_j)^4\right)\mathrm{E}\left((\xi_k)^4\right)\right]^{\frac{1}{4}} \\
&+ \frac{2}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j\in A_i}\sum_{k\in B_i}\left[\mathrm{E}\left((\xi_i)^2\right)\mathrm{E}\left((\xi_j)^2\right)\mathrm{E}\left((\xi_k)^2\right)\right]^{\frac{1}{2}} \\
&+ \frac{1}{n^{\frac{3}{2}}}\sum_{i=1}^{n}|A_i|\sum_{j\in A_i}\left[\mathrm{E}\left((\xi_i)^2\right)\mathrm{E}\left((\xi_j)^4\right)\right]^{\frac{1}{2}} + \left|\frac{\sqrt{\mathrm{Var}\left[l'(\theta_0;\boldsymbol{X})\right]}}{\sqrt{n\bar{i}(\theta_0)}} - 1\right| \\
&+ \frac{A_d}{\sqrt{n\bar{i}(\theta_0)}}\left(\sqrt{\mathrm{E}\left(l''(\theta_0;\boldsymbol{X}) + n\bar{i}(\theta_0)\right)^2} + \frac{S_d(n)A_d}{2}\right). \qquad (6.53)
\end{aligned}$$

Remark 6.4. *The order of A_d in terms of the sample size is $\frac{1}{\sqrt{n}}$. Assuming that $\left|\frac{\sqrt{\mathrm{Var}(l'(\theta_0;\boldsymbol{X}))}}{\sqrt{n\bar{i}(\theta_0)}}-1\right| = \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$, then the order of the final upper bound in* (6.53) *is also $\frac{1}{\sqrt{n}}$.*

Proof. Using similar steps as in the proof of Theorem 6.3, but keeping the result in a general form and not specifying in terms of the Wasserstein distance, we obtain that for $h:\mathbb{R}\to\mathbb{R}$ being any absolutely continuous function,

$$\begin{aligned}
&\left|\mathrm{E}\left[h\left(\sqrt{n\bar{i}(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)\right)\right]-\mathrm{E}[h(Z)]\right| \\
&\leq \frac{2\|h'\|}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j\in A_i}\sum_{k\in B_i}\left[\mathrm{E}\left((\xi_i)^4\right)\mathrm{E}\left((\xi_j)^4\right)\mathrm{E}\left((\xi_k)^4\right)\right]^{\frac{1}{4}} \\
&+\frac{2\|h'\|}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j\in A_i}\sum_{k\in B_i}\left[\mathrm{E}\left((\xi_i)^2\right)\mathrm{E}\left((\xi_j)^2\right)\mathrm{E}\left((\xi_k)^2\right)\right]^{\frac{1}{2}} \\
&+\frac{\|h'\|}{n^{\frac{3}{2}}}\sum_{i=1}^{n}|A_i|\sum_{j\in A_i}\left[\mathrm{E}\left((\xi_i)^2\right)\mathrm{E}\left((\xi_j)^4\right)\right]^{\frac{1}{2}}+\|h'\|\left|\frac{\sqrt{\mathrm{Var}(l'(\theta_0;\boldsymbol{X}))}}{\sqrt{n\bar{i}(\theta_0)}}-1\right| \\
&+\frac{\|h'\|S_d(n)}{2\sqrt{n\bar{i}(\theta_0)}}\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right] \\
&+\frac{\|h'\|}{\sqrt{n\bar{i}(\theta_0)}}\sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]}\sqrt{\mathrm{E}\left[\left(l''(\theta_0;\boldsymbol{X})+n\bar{i}(\theta_0)\right)^2\right]}. \qquad (6.54)
\end{aligned}$$

The bound depends on $\hat{\theta}_n(\boldsymbol{X})$ only through the MSE. With γ_1 as in (6.52) we denote by

$$\begin{aligned}
D_{x^2} = \gamma_1 &+ \frac{2\|x\|}{\sqrt{n\bar{i}(\theta_0)}}\sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]}\sqrt{\mathrm{E}\left[\left(l''(\theta_0;\boldsymbol{X})+n\bar{i}(\theta_0)\right)^2\right]} \\
&+\frac{\|x\|S_d(n)}{\sqrt{n\bar{i}(\theta_0)}}\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]
\end{aligned}$$

the bound in (6.54) for the case of $h(x)=x^2$. Following the same approach as in (3.56),

$$\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right] \leq \frac{1}{n\bar{i}(\theta_0)}\left(D_{x^2}+1\right) \qquad (6.55)$$

and since D_{x^2} is an expression of the mean squared error and its positive root, our next step is to solve the quadratic inequality in (6.55) with unknown

$$U_d := U_d(\theta_0) := \sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right)^2\right]}. \tag{6.56}$$

We have that

$$\begin{aligned} U_d^2 &\leq \frac{\|x\| S_d(n)}{[n\bar{i}(\theta_0)]^{\frac{3}{2}}} U_d^2 + \frac{2\|x\| \sqrt{\mathrm{E}\left[\left(l''(\theta_0;\boldsymbol{X}) + n\bar{i}(\theta_0)\right)^2\right]}}{[n\bar{i}(\theta_0)]^{\frac{3}{2}}} U_d + \frac{1+\gamma_1}{n\bar{i}(\theta_0)} \\ \Leftrightarrow &\left(1 - \frac{\|x\| S_d(n)}{[n\bar{i}(\theta_0)]^{\frac{3}{2}}}\right) U_d^2 - \frac{2\|x\| \sqrt{\mathrm{E}\left[\left(l''(\theta_0;\boldsymbol{X}) + n\bar{i}(\theta_0)\right)^2\right]}}{[n\bar{i}(\theta_0)]^{\frac{3}{2}}} U_d - \frac{1+\gamma_1}{n\bar{i}(\theta_0)} \leq 0 \end{aligned} \tag{6.57}$$

and the two solutions of this inequality are

$$\begin{aligned} U_{d_{1,2}} = \frac{1}{2D_1} \Bigg\{ &\frac{2\|x\| \sqrt{\mathrm{E}\left[\left(l''(\theta_0;\boldsymbol{X}) + n\bar{i}(\theta_0)\right)^2\right]}}{[n\bar{i}(\theta_0)]^{\frac{3}{2}}} \\ &\pm \left[4 \frac{\|x\|^2 \mathrm{E}\left[\left(l''(\theta_0;\boldsymbol{X}) + n\bar{i}(\theta_0)\right)^2\right]}{[n\bar{i}(\theta_0)]^3} + 4 \frac{D_1 (1+\gamma_1)}{n\bar{i}(\theta_0)} \right]^{\frac{1}{2}} \Bigg\}. \end{aligned} \tag{6.58}$$

To have meaningful results we need the quantity

$$4 \frac{\|x\|^2}{[n\bar{i}(\theta_0)]^3} \mathrm{E}\left[\left(l''(\theta_0;\boldsymbol{X}) + n\bar{i}(\theta_0)\right)^2\right] + 4 \frac{D_1 (1+\gamma_1)}{n\bar{i}(\theta_0)},$$

which is in the square root above, to be non-negative. This is satisfied through (F.A.2) which ensures that D_1 is positive. It follows that the solution with the negative square root in (6.58) is negative, while the other with the positive square root is positive and the solution of (6.57) is

$$0 \leq U_d = \sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right)^2\right]} \leq A_d.$$

Applying this result to the bound in (6.44) gives the second result of the theorem. □

6.4.2 Bounded parameter space

In a number of situations the support of the distribution is not bounded and (F.A.1) is not satisfied. This subsection gives another approach to the problem of bounding the MSE of the MLE without using (F.A.1). Apart from (R.D.1)-(R.D.5), we still make two further assumptions though, and these are

(Fur.Ass.1) The parameter space is a closed (or semiclosed) interval of the form $[a,b]$ (or $[a,b)$, $(a,b]$), with $a < b \in \mathbb{R}$. From now on, $\kappa = b - a$.

(Fur.Ass.2) With κ as in (Fur.Ass.1) $\exists N \in \mathbb{N}$ such that $\forall n \geq N$

$$n > \frac{\kappa S_d(n)}{2\bar{i}(\theta_0)}.$$

Denote by

$$\begin{aligned}
D_2 &:= D_2(\theta_0, x, n) = 1 - \frac{\kappa S_d(n)}{2n\bar{i}(\theta_0)} \\
\gamma_2 &:= \gamma_2(\theta_0, x, n) = \frac{2}{n^{\frac{3}{2}}} \sum_{i=1}^{n} \sum_{j \in A_i} \sum_{k \in B_i} \left[\mathrm{E}\left((\xi_i)^4 \right) \mathrm{E}\left((\xi_j)^4 \right) \mathrm{E}\left((\xi_k)^4 \right) \right]^{\frac{1}{4}} \\
&+ \frac{2}{n^{\frac{3}{2}}} \sum_{i=1}^{n} \sum_{j \in A_i} \sum_{k \in B_i} \left[\mathrm{E}\left((\xi_i)^2 \right) \mathrm{E}\left((\xi_j)^2 \right) \mathrm{E}\left((\xi_k)^2 \right) \right]^{\frac{1}{2}} \\
&+ \frac{1}{n^{\frac{3}{2}}} \sum_{i=1}^{n} |A_i| \sum_{j \in A_i} \left[\mathrm{E}\left((\xi_i)^2 \right) \mathrm{E}\left((\xi_j)^4 \right) \right]^{\frac{1}{2}} + \left| \frac{\sqrt{\mathrm{Var}\left[l'(\theta_0; \boldsymbol{X}) \right]}}{\sqrt{n\bar{i}(\theta_0)}} - 1 \right|. \qquad (6.59)
\end{aligned}$$

The following theorem gives a bound on the Wasserstein distance related to the MLE in cases where it is not known analytically and the support of the distribution is not necessarily bounded. The approach is again based on finding an upper bound for the MSE of the MLE, which we then apply to the expression of the general upper bound.

Theorem 6.5. *Let $\{X_i, i = 1,2,\ldots,n\}$ be an m-dependent sequence of random variables with joint probability density (or mass) function $f(\boldsymbol{x}|\theta)$ as in* (6.1)*. Assume that the regularity conditions (R.D.1)-(R.D.5), as well as the assumptions (Fur.Ass.1), (Fur.Ass.2) are satisfied. Assume that the MLE exists and is unique. Then $A_2 = A_2(\theta_0, n)$ is an upper bound for $\sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X}) - \theta_0\right)^2\right]}$, where*

$$A_2 = \frac{1}{2D_2}\left\{ \frac{\kappa\sqrt{\mathrm{E}\left[\left(l''(\theta_0;\boldsymbol{X}) + n\bar{i}(\theta_0)\right)^2\right]}}{n\bar{i}(\theta_0)} + \left[\frac{\kappa^2 \mathrm{E}\left[\left(l''(\theta_0;\boldsymbol{X}) + n\bar{i}(\theta_0)\right)^2\right]}{[n\bar{i}(\theta_0)]^2} + 4\frac{D_2\kappa\left(\gamma_2 + \sqrt{\frac{2}{\pi}}\right)}{\sqrt{n\bar{i}(\theta_0)}} \right]^{\frac{1}{2}} \right\}. \quad (6.60)$$

with D_2 and γ_2 as in (6.59) *and κ as in (Fur.Ass.1). Then if* $\mathrm{Var}\left[l'(\theta_0;\boldsymbol{X})\right] > 0$*,*

$$\xi_1 = \frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_1|\theta)\Big|_{\theta=\theta_0}\sqrt{\frac{n}{\mathrm{Var}\left[l'(\theta_0;\boldsymbol{X})\right]}}$$

and

$$\xi_i = \frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_i|X_{i-1},\ldots,X_{i-m};\theta)\Big|_{\theta=\theta_0}\sqrt{\frac{n}{\mathrm{Var}\left[l'(\theta_0;\boldsymbol{X})\right]}}, i = 2,3,\ldots,n,$$

we have that

$$\begin{aligned} d_W\left(\sqrt{n\bar{i}(\theta_0)}(\hat{\theta}_n(\boldsymbol{X}) - \theta_0), Z\right) &\leq \frac{2}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j\in A_i}\sum_{k\in B_i}\left[\mathrm{E}\left((\xi_i)^4\right)\mathrm{E}\left((\xi_j)^4\right)\mathrm{E}\left((\xi_k)^4\right)\right]^{\frac{1}{4}} \\ &+ \frac{2}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j\in A_i}\sum_{k\in B_i}\left[\mathrm{E}\left((\xi_i)^2\right)\mathrm{E}\left((\xi_j)^2\right)\mathrm{E}\left((\xi_k)^2\right)\right]^{\frac{1}{2}} \\ &+ \frac{1}{n^{\frac{3}{2}}}\sum_{i=1}^{n}|A_i|\sum_{j\in A_i}\left[\mathrm{E}\left((\xi_i)^2\right)\mathrm{E}\left((\xi_j)^4\right)\right]^{\frac{1}{2}} + \left|\frac{\sqrt{\mathrm{Var}\left[l'(\theta_0;\boldsymbol{X})\right]}}{\sqrt{n\,\bar{i}(\theta_0)}} - 1\right| \\ &+ \frac{A_2}{\sqrt{n\bar{i}(\theta_0)}}\left(\sqrt{\mathrm{E}\left[\left(l''(\theta_0;\boldsymbol{X}) + n\bar{i}(\theta_0)\right)^2\right]} + \frac{S_d(n)A_2}{2}\right). \end{aligned} \quad (6.61)$$

Proof. We use (Fur.Ass.1) to get that

$$\begin{aligned}\frac{\left|\hat{\theta}_n(X)-\theta_0\right|}{\kappa}\leq 1 \quad &\Rightarrow \quad \frac{1}{\kappa^2}\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]\leq \frac{1}{\kappa}\mathrm{E}\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right| \\ &\Rightarrow \quad \mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]\leq \kappa\mathrm{E}\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right|. \end{aligned} \tag{6.62}$$

For $h(x)=|x|$, (6.62) yields

$$\begin{aligned}&\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]\leq \kappa\mathrm{E}\left|\hat{\theta}_n(\boldsymbol{X})-\theta_0\right| \\ &=\frac{\kappa}{\sqrt{n\bar{i}(\theta_0)}}\left|\mathrm{E}\left[h\left(\sqrt{n\bar{i}(\theta_0)}\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)\right)\right]-\mathrm{E}[h(Z)]+\mathrm{E}[h(Z)]\right| \\ &\leq \frac{\kappa}{\sqrt{n\bar{i}(\theta_0)}}\left(B_{|x|}+\sqrt{\frac{2}{\pi}}\right), \end{aligned} \tag{6.63}$$

where

$$\begin{aligned}B_{|x|}=\gamma_2+\frac{S_d(n)}{2\sqrt{n\bar{i}(\theta_0)}}\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right] \\ +\frac{\sqrt{\mathrm{E}\left[\left(l''(\theta_0;\boldsymbol{X})+n\bar{i}(\theta_0)\right)^2\right]}}{\sqrt{n\bar{i}(\theta_0)}}\sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]}\end{aligned} \tag{6.64}$$

is the bound in (6.54) for $h(x)=|x|$. Obviously, in this case $\|h'\|=1$. Thus, for U_d as in (6.56) we can easily see that (6.63) and (6.64) yield

$$\begin{aligned}&U_d^2\leq \frac{\kappa S_d(n)}{2n\bar{i}(\theta_0)}U_d^2+\frac{\kappa\sqrt{\mathrm{E}\left[\left(l''(\theta_0;\boldsymbol{X})+n\bar{i}(\theta_0)\right)^2\right]}}{n\bar{i}(\theta_0)}U_d+\frac{\kappa\left(\gamma_2+\sqrt{\frac{2}{\pi}}\right)}{\sqrt{n\bar{i}(\theta_0)}} \\ &\Leftrightarrow \left(1-\frac{\kappa S_d(n)}{2n\bar{i}(\theta_0)}\right)U_d^2-\frac{\kappa\sqrt{\mathrm{E}\left[\left(l''(\theta_0;\boldsymbol{X})+n\bar{i}(\theta_0)\right)^2\right]}}{n\bar{i}(\theta_0)}U_d-\frac{\kappa\left(\gamma_2+\sqrt{\frac{2}{\pi}}\right)}{\sqrt{n\bar{i}(\theta_0)}}\leq 0\end{aligned} \tag{6.65}$$

with the solutions of this simple quadratic inequality found to be

$$U_{d_{1,2}} = \frac{1}{2D_2}\left\{ \frac{\kappa\sqrt{\mathrm{E}\left[\left(l''(\theta_0;\boldsymbol{X})+n\bar{i}(\theta_0)\right)^2\right]}}{n\bar{i}(\theta_0)} \right.$$
$$\left. \pm \left[\frac{\kappa^2\mathrm{E}\left[\left(l''(\theta_0;\boldsymbol{X})+n\bar{i}(\theta_0)\right)^2\right]}{[n\bar{i}(\theta_0)]^2} + 4\frac{D_2\kappa\left(\gamma_2+\sqrt{\frac{2}{\pi}}\right)}{\sqrt{n\bar{i}(\theta_0)}} \right]^{\frac{1}{2}} \right\}. \quad (6.66)$$

For meaningful results, D_2 has to be positive which is ensured through (Fur.Ass.2). This leads to one of the solutions in (6.66) being negative and the other one being positive. The positive solution of (6.65) is

$$0 \leq U_d = \sqrt{\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]} \leq A_2,$$

with A_2 as in (6.60). Using this result in the expression of the upper bound in (6.44) gives the second result of the theorem as expressed in (6.61). □

A prominent example where Theorem 6.5 may be useful is parameter estimation in MA(q) time series; see Subsection 7.2.1 for more details.

Chapter 7

Conclusion

This chapter summarises the work carried out in this thesis and places the research in a wider context. An outline of what we consider as open problems on which we plan to work in the future is also provided. We subdivide these future work ideas into two parts. The first one is concerned with further assessing the performance of our bounds through various, possibly complicated applications, while in the second part we go beyond the MLE and the derivation of new bounds for other asymptotic results is discussed.

7.1 Summary

In this thesis, we have derived under various settings upper bounds on the distributional distance between the exact, unknown distribution of the MLE and the normal distribution. These bounds are explicit and novel. Among other famous mathematical tools, such as Taylor expansions, our work has been partly based on an elegant probabilistic technique; Stein's method. The purpose and the motivation for this research project have been set out in the Introduction.

In Chapter 2, we gave insights to the CLT; the basic theory and properties related

to the MLE; as well as a concise introduction to Stein's method. The assumptions ensuring existence and uniqueness of the MLE along with sufficient regularity conditions for the asymptotic normality of a scalar MLE for i.i.d. random variables were given. After presenting and proving the asymptotic result for the distribution of the MLE, Stein's method was concisely reviewed with Lemma 2.2 being one of the main results we used in Chapters 3 and 5 while Lemma 2.3 was used in Chapter 6 in order to obtain bounds related to single-parameter distributions.

In Chapter 3, we introduced the bounded Wasserstein distance, which is the metric used throughout the chapter and in Chapter 5. A connection of this metric with the Kolmogorov distance allowed us to give insights to confidence intervals. The main result of the chapter is given in Theorem 3.1 with the derived bound being, in terms of the sample size, of order $\frac{1}{\sqrt{n}}$. After the general results, we turned our focus on single-parameter exponential families, with the bound applied on the exponential distribution presented in both its canonical and non-canonical form. We explained why the results in the canonical case are simpler and sharper and this is illustrated on simulated data. A slightly more complicated case was then treated based on the issue of the MLE having positive probability of being on the boundary of the parameter space for discrete distributions. The solution of this problem relies on a perturbation method which introduces a perturbed parameter and a perturbed MLE that are now interior to the parameter space. As probably expected, this procedure created error terms which were also bounded, leading to an upper bound for such cases of order still $\frac{1}{\sqrt{n}}$. The Poisson distribution with parameter space $[0,\infty)$ was the example we attacked. In the last part of Chapter 3, we focused on bounding the mean squared error, $\mathrm{E}\left[\left(\hat{\theta}_n(\boldsymbol{X})-\theta_0\right)^2\right]$, by a quantity of which the calculation does not require to know the MLE explicitly. This approach gave us the opportunity of presenting a new bound on the bounded Wasserstein distance between the distribution of the MLE and the normal distribution which holds even in cases where an analytic expression of the MLE is not

available. The Beta distribution served as an illustration of these results.

In Chapter 4, our interest shifted to the multi-parameter setting. We gave upper bounds related to the distribution of a vector MLE and the multivariate normal. We distinguished between two cases depending on whether the random vectors are assumed to be heterogeneous or not. Different regularity conditions were given in each case in order for the asymptotic normality of the MLE to hold. Keeping the dimensionality of the parameter constant, the bounds are of the same order as those in the single-parameter case. The general family of linear regression models was treated separately, with the results being simple and easy to calculate as can be seen from the example of the straight-line regression. The normal distribution with $\boldsymbol{\theta_0} = (\mu, \sigma^2)$ served as a test case for identically distributed random variables. The idea of deriving bounds for situations where the MLE does not have a closed-form expression has been expanded to the multi-parameter setting. An upper bound on $\mathrm{E}\left[\sum_{j=1}^{d}\left(\hat{\theta}_n(\boldsymbol{X})_j - \theta_{0_j}\right)^2\right]$ that does not depend on the MLE was derived and applied on the general bound, which gives an upper bound of which the calculation does not rely upon $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})$. As an example we used again the Beta distribution but now with both shape parameters unknown.

Chapter 5 presents an exploration of the way that the Delta method can be used to obtain bounds on the distributional distance of interest, constrained to the scalar parameter case. The bounds given are of the same order and similar to those in Chapter 3. However, in some cases they have a simpler representation and are also sharper. A special interest was again put on single-parameter exponential family distributions, which made the comparison to the results in Chapter 3 straightforward. The generalised gamma distribution with the scale parameter considered unknown was treated.

In Chapter 6, the independence assumption between the random variables has been relaxed and we introduced a local dependence structure. The notion of $m-$dependence

was explained and employing the results from Stein's method as in Lemma 2.3, we achieved to obtain bounds on the Wasserstein distance between the distribution of the MLE and the normal distribution. An example from locally dependent normal random variables showed that the bound behaves well in terms of simplicity and order; it is still of order $\frac{1}{\sqrt{n}}$. Under further assumptions that require boundedness of the support of the distribution or of the parameter space, we obtained as in Chapter 3 an upper bound for cases where the MLE can not be explicitly derived.

7.2 Open problems for future exploration

Having developed the theory related to upper bounds on the distributional distance between the distribution of the MLE and the normal distribution, a next step is to cover a number of important applications and also develop similar results for different kinds of random variables–either estimators or not–which follow asymptotically a known distribution that is not necessarily the normal. For ease of presentation this section is split into two parts depending on whether we discuss further applications or theory.

7.2.1 Further applications

Time Series is the first broad area we would like to discuss. A time series can be seen as a sequence of data, which has traditionally been of great importance when it comes to statistical analysis and forecasting. Maximum likelihood estimation is among the two most used estimation methods–the other one is least squares–for the parameters in time series models. Under sufficient conditions (Theorem 10.8.1 from Brockwell and Davis (1991)), the asymptotic normality of the MLE holds and its proof for general ARMA(p,q) processes is given in Brockwell and Davis (1991), p.390. We have carried out preliminary work on the MA(q) process $\{X_t, t = 0, \pm 1, \pm 2, \ldots\}$, where

the relationship

$$X_t = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q} \tag{7.1}$$

holds, with the process $\{Z_t\} \sim \text{WN}(0, \sigma^2)$. The notation $\text{WN}(0, \sigma^2)$ stands for white noise and the random variables Z_t have mean zero, variance σ^2 and also $\text{Cov}(Z_t, Z_{t+a}) = 0$, for $a \in \mathbb{Z} \setminus \{0\}$. The MA($q$) process gives q-dependent random variables and (7.1) defines the local dependence structure between the random variables X_t. More specifically, from Brockwell and Davis (1991) for $\theta_0 = 1$,

$$\text{Cov}(X_t, X_{t+a}) = \begin{cases} \sigma^2 \sum_{j=0}^{q-|a|} \theta_j \theta_{j+|a|}, & \text{if } |a| \leq q \\ 0, & \text{if } |a| > q. \end{cases} \tag{7.2}$$

For the MA(1) process, assume that Z_t are normally distributed and denote by $\Sigma = \text{E}(\boldsymbol{X}\boldsymbol{X}^\intercal)$ the $n \times n$ covariance matrix of $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$. The likelihood function of $\boldsymbol{X}$ is

$$L(\theta; \boldsymbol{X}) = \frac{1}{(2\pi)^{\frac{n}{2}} [\det(\Sigma)]^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} \boldsymbol{X}^\intercal \Sigma^{-1} \boldsymbol{X} \right\}.$$

The covariance matrix Σ is obtained using (7.2) and

$$\Sigma = \sigma^2 \begin{pmatrix} 1+\theta^2 & \theta & 0 & 0 & 0 & \ldots & 0 \\ \theta & 1+\theta^2 & \theta & 0 & 0 & \ldots & 0 \\ 0 & \theta & 1+\theta^2 & \theta & 0 & \ldots & 0 \\ 0 & 0 & \theta & 1+\theta^2 & \theta & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \theta & 1+\theta^2 & \theta \\ 0 & 0 & 0 & 0 & 0 & \theta & 1+\theta^2 \end{pmatrix}.$$

The MLE for θ exists but there is not an analytic expression available and numerical methods are used to maximize the likelihood. We could employ the results of Section

6.4 to find an upper bound on the distributional distance of interest under this time series framework of locally dependent random variables.

Another open problem is to obtain bounds related to the popular and rapidly expanding area of networks. Such a network is the Erdös-Renyi (Bernoulli) random graph, where an edge between two vertices is drawn with probability p, independently of the other edges. Suppose we have n vertices. The degree of a randomly picked vertex follows the Binomial$(n-1,p)$ distribution with unknown parameter p. Thus, let $X_1, X_2, \ldots, X_n \sim \text{Bin}(n-1,p)$ be i.i.d. random variables for the degree of the vertices $1,2,\ldots,n$ respectively with $p \in [0,1]$. The MLE for the parameter p in this case is $\hat{\theta}_n(\boldsymbol{X}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} X_i$. Since we have a discrete distribution with compact parameter space, there is positive probability of the MLE being on the boundary of the parameter space. Thus, the perturbation approach as explained in Section 3.4 needs to be adapted before finding the required upper bound. Treating this case could lead to work on more complex examples, one of which is the geometric random graph. Two vertices are connected if and only if the distance between them is at most a radius r. The idea would be to estimate this radius and apply the theory for locally dependent random variables to the distribution of the estimator. In Raič (2004), relevant bounds on the distributional distance between the distribution of $W = \sum_{i=1}^{n} X_i$ and the normal have been obtained in the case of a dependence graph, with $X_i, i = 1,2,\ldots,n$ the degree of the vertex i. These results could be useful when it comes to bounding the relevant distributional distance for the MLE which can be of any form and not restricted to a sum of random variables.

The extension of the Delta method approach in order to treat multi-parameter distributions is also an interesting idea for future work. For this purpose, let $q : \Theta \to \mathbb{R}^d$, such as all the entries of the function q have continuous partial derivatives with respect to $\boldsymbol{\theta}$. In addition, we require that $q(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}(\boldsymbol{X})) = \frac{1}{n} \sum_{i=1}^{n} g(X_i)$, for some function $g : \mathbb{R} \to \mathbb{R}^d$. We use a multivariate generalisation of the Delta method on the result of Theorem 4.1 for

the asymptotic normality of the MLE. We denote by $K := [\nabla q(\boldsymbol{\theta_0})][I(\boldsymbol{\theta_0})]^{-1}[\nabla q(\boldsymbol{\theta_0})]^\intercal$, where $\nabla q(\boldsymbol{\theta_0})$ is just the $d \times d$ Jacobian matrix of the function q evaluated at $\boldsymbol{\theta_0}$. The Delta method gives

$$\sqrt{n}\left(q\left(\hat{\boldsymbol{\theta}}_n(\boldsymbol{X})\right) - q(\boldsymbol{\theta_0})\right) \xrightarrow[n\to\infty]{\mathrm{d}} K^{\frac{1}{2}}\boldsymbol{Z}.$$

Assuming that the Jacobian matrix of q is invertible, preliminary work has shown that the approach to be followed in order to find an upper bound for the quantity of interest is a combination of the steps explained in Chapters 4 and 5.

7.2.2 Beyond the classical framework of the MLE

The asymptotic theory for $\hat{\theta}_n(\boldsymbol{X})$ allows to Taylor expand its bias and variance terms using expressions based on the sample size, n. For instance, according to Ferrari et al. (1996), expansions for the bias, $B(\hat{\theta}_n(\boldsymbol{X}))$, and variance of the MLE to order n^{-2} can be written as

$$B(\hat{\theta}_n(\boldsymbol{X})) = \frac{B_1(\hat{\theta}_n(\boldsymbol{X}))}{n} + \frac{B_2(\hat{\theta}_n(\boldsymbol{X}))}{n^2}, \quad \mathrm{Var}(\hat{\theta}_n(\boldsymbol{X})) = \frac{V_1(\hat{\theta}_n(\boldsymbol{X}))}{n} + \frac{V_2(\hat{\theta}_n(\boldsymbol{X}))}{n^2},$$

where closed formulas for the functions B_1, B_2, V_1 and V_2 can be easily derived. Using the above expansions, one can define various types of bias-corrected estimators, with the most straightforward being

$$\tilde{\theta}_1(\boldsymbol{X}) = \hat{\theta}_n(\boldsymbol{X}) - \frac{B_1(\hat{\theta}_n(\boldsymbol{X}))}{n}, \quad \tilde{\theta}_2(\boldsymbol{X}) = \hat{\theta}_n(\boldsymbol{X}) - \frac{B_1(\hat{\theta}_n(\boldsymbol{X}))}{n} - \frac{B_2(\hat{\theta}_n(\boldsymbol{X}))}{n^2}.$$

These modified estimators are bias-free to order $\frac{1}{n}$ as shown in Ferrari et al. (1996). An open problem raised directly from the results of the current thesis is to find upper bounds for the distributional distance related to these bias-corrected estimators. We would expect the constant of the bound to be better than the one linked with the MLE.

In addition, the order of the new bounds with respect to the sample size should also be examined; is it better than the one for the bound related to the MLE?

Another interesting idea for future work is to expand the approach explained in this thesis to a broad family of parametric estimators, called M-estimators. These are robust estimators defined as a root of an estimating function. This estimating function is usually the derivative of another statistical function, $\rho(x,\theta)$. An M-estimator is the minimum of $\rho(x,\theta)$. To set the scene in the case of independent random variables $X_i, i = 1,2,\ldots,n$, the M-estimator $\hat{\theta}_n^M(\boldsymbol{X})$ is the solution that minimises $\sum_{i=1}^n \rho(X_i,\theta)$. The choice of the function ρ differs–for example $\rho(x_i,\theta) = -L(\theta;x_i)$ in the case of the MLE–and when it is differentiable, the computation of $\hat{\theta}_n^M(\boldsymbol{X})$ comes from deriving the critical points of $\frac{\mathrm{d}}{\mathrm{d}\theta}\sum_{i=1}^n \rho(x_i,\theta)$. Under again some sufficient regularity conditions, it has been shown that the distribution of the M-estimators is asymptotically normal; see Huber (1964) for the first proof related to location parameters. The assessment of the quality of this asymptotic result remains an open problem. One could follow an approach based again on Taylor expansions of $0 = \sum_{i=1}^n \frac{\mathrm{d}}{\mathrm{d}\theta}\rho\left(x_i, \hat{\theta}_n^M(\boldsymbol{x})\right)$ about the true value of the unknown parameter θ_0. Then, Stein's method for normal approximation can be employed in order to find an upper bound on the distributional distance between the distribution of $W_M = \sum_{i=1}^n \frac{\mathrm{d}}{\mathrm{d}\theta}\rho\left(X_i,\theta_0\right)$ and the normal distribution. Alternative techniques can then be developed to find upper bounds for all the remainder terms.

A different, interesting direction is towards Bayesian Statistics. It has already been shown in the statistical literature (see for example van der Vaart (1998)) that under specific assumptions both the posterior measures and the Bayes point estimators are asymptotically normal. Specifically, according to the Bernstein-von Mises theorem as expressed in van der Vaart (1998), under remarkably weak conditions, the posterior distribution of $\sqrt{n}(\Theta_n - \theta_0)$ is under θ_0 asymptotically equal to $\mathrm{N}\left(Y, \frac{1}{i(\theta_0)}\right)$. In this case, Y follows the $\mathrm{N}\left(0, \frac{1}{i(\theta_0)}\right)$ distribution and Θ_n is the "random" parameter,

which in the Bayesian perspective is a random variable. Furthermore, under stronger regularity conditions we obtain that the distribution of Θ_n is approximately equal to $\mathrm{N}\left(\hat{\theta}_n(\boldsymbol{X}), \frac{1}{n\,i(\theta_0)}\right)$, where $\hat{\theta}_n(\boldsymbol{X})$ is the MLE. This is again an asymptotic result of which the quality could be assessed through upper bounds on the distributional distance between the distribution of $\sqrt{n}(\Theta_n - \theta_0)$ and the normal. We would expect that new horizons in applications will be opened if such results in the area of Bayesian Statistics become available in the near future.

The log-likelihood ratio statistic is used in the so-called likelihood ratio test where the aim is to distinguish between two nested models. In the case of a scalar parameter, aiming to test whether $\theta = \theta_0$ or not, it is well known that

$$2(l(\hat{\theta}_n(\boldsymbol{X}); \boldsymbol{X}) - l(\theta_0; \boldsymbol{X})) \xrightarrow{\mathrm{d}} \chi^2_1. \tag{7.3}$$

In the case of a vector parameter and composite hypothesis testing of whether $\boldsymbol{\theta} \in \Theta_0$ or not, for Θ_0 a subset of the parameter space Θ, we get that

$$2(l(\hat{\boldsymbol{\theta}}_{\boldsymbol{2}}(\boldsymbol{X}); \boldsymbol{X}) - l(\hat{\boldsymbol{\theta}}_{\boldsymbol{1}}(\boldsymbol{X}); \boldsymbol{X})) \xrightarrow{\mathrm{d}} \chi^2_d, \tag{7.4}$$

where $\hat{\boldsymbol{\theta}}_{\boldsymbol{1}}(\boldsymbol{X})$ and $\hat{\boldsymbol{\theta}}_{\boldsymbol{2}}(\boldsymbol{X})$ are the MLEs for $\boldsymbol{\theta} \in \Theta_0$ and $\boldsymbol{\theta} \in \Theta$, respectively. In addition, d is the difference in the dimensionality of Θ and Θ_0. The result in (7.4) is known as Wilks's theorem and it was first proved in Wilks (1938). One could work on assessing the quality of the asymptotic results in (7.3) and (7.4). An idea of the process we have in mind is to split again the quantity of interest into two easier to handle terms. The first of these two terms could be upper bounded using Stein's method for χ^2 approximation; specifically the results in Gaunt et al. (2015) can be useful, where explicit bounds of order $\frac{1}{n}$ for smooth test functions are given. For the second term, one could exploit further expansions and known probability inequalities, such as the Markov and Cauchy-Schwarz inequalities.

References

R. Arratia, L. Goldstein, and L. Gordon. Poisson Approximation and the Chen-Stein Method. *Statistical Science*, **5**:403–434, 1990.

A. D. Barbour. Stein's method for diffusion approximations. *Probability Theory and Related Fields*, **84**:297–322, 1990.

A. D. Barbour and L. Holst. Some applications of the Stein-Chen method for proving Poisson convergence. *Advances in Applied Probability*, **21**:74–90, 1989.

A. D. Barbour, L. Holst, and S. Janson. *Poisson Approximation*. Oxford University Press, 1992.

A. D. Barbour, H. L. Gan, and A. Xia. Stein factors for negative binomial approximation in wasserstein distance. *Bernoulli*, **21**:1002–1013, 2015.

R. H. Berk. Consistency and asymptotic normality of MLE's for exponential models. *The Annals of Mathematical Statistics*, **43**:193–204, 1972.

B. R. Bhat. On the Method of Maximum-Likelihood for Dependent Observations. *Journal of the Royal Statistical Society. Series B (Methodological)*, **36**:48–53, 1974.

P. Billingsley. Statistical Methods in Markov Chains. *The Annals of Mathematical Statistics*, **32**:12–40, 1961.

P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer Series in Statistics. Springer-Verlag, New York, second edition, 1991.

G. Casella and R. L. Berger. *Statistical Inference*. Brooks/Cole, Cengage Learning, Duxbury, Pacific Grove, second edition, 2002.

S. Chatterjee and E. Meckes. Multivariate normal approximation using exchangeable pairs. *ALEA, Latin American Journal of Probability and Mathematical Statistics*, **4**: 257–283, 2008.

S. Chatterjee, J. Fulman, and A. Röllin. Exponential approximation by exchangeable pairs and spectral graph theory. *ALEA, Latin American Journal of Probability and Mathematical Statistics*, **8**:1–27, 2011.

L. H. Y. Chen. Poisson approximation for dependent trials. *Annals of Probability*, **3**: 534–545, 1975.

L. H. Y. Chen, L. Goldstein, and Q. M. Shao. *Normal Approximation by Stein's Method*. Springer-Verlag, Berlin Heidelberg, 2011.

D. R. Cox and E. J. Snell. A General Definiton of Residuals. *Journal of the Royal Statistical Society*, **30**:248–275, 1968.

A. C. Davison. *Statistical Models*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2008.

P. Diaconis. *An example for Stein's method*. Stanford University, Department of Statistics, Technical Report, 1989.

L. H. Dicker and M. A. Erdogdu. Flexible results for quadratic forms with applications to variance components estimation. *Annals of Statistics. To appear*, 2016+.

C. Döbler. Stein's method of exchangeable pairs for the Beta distribution and generalizations. *Electronic Journal of Probability*, **20**:1–34, 2015.

W. Ehm. Binomial approximation to the Poisson binomial distribution. *Statistics & Probability Letters*, **11**:7–16, 1991.

L. Fahrmeir and H. Kaufmann. Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *The Annals of Statistics*, **13**: 342–368, 1985.

S. L. P. Ferrari, D. A. Botter, G. M. Cordeiro, and F. Cribari-Neto. Second and third order bias reduction for one-parameter family models. *Statistics and Probability Letters*, **30**:339–345, 1996.

A. M. Fink and M. Jodeit. On Chebyshev's other inequality. *Lecture Notes-Monograph Series, Inequalities in Statistics and Probability*, **5**:115–120, 1984.

R. A. Fisher. Theory of Statistical Estimation. *Mathematical Proceedings of the Cambridge Philosophical Society*, **22**:700–725, 1925.

R. E. Gaunt. *Rates of Convergence of Variance-Gamma Approximations via Stein's Method*. PhD thesis, University of Oxford, 2013.

R. E. Gaunt, A. Pickett, and G. Reinert. Chi-square approximation by Stein's method with application to Pearson's statistic. http://arxiv.org/pdf/1507.01707v1.pdf, 2015.

C. J. Geyer. Asymptotics of exponential families. University lecture notes, http://www.stat.umn.edu/geyer/8112/notes/expfam.pdf, 2013.

L. Goldstein and G. Reinert. Stein's method for the beta distribution and the Polya-Eggenberger urn. *Journal of Applied Probability*, **50**:1187–1205, 2013.

F. Götze. On the rate of convergence in the multivariate CLT. *Annals of Probability*, **19**:724–739, 1991.

L. Heinrich. A Method for the Derivation of Limit Theorems for Sums of m-dependent Random Variables. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **60**:501–515, 1982.

B. Hoadley. Asymptotic Properties of Maximum Likelihood Estimators for the Independent Not Identically Distributed Case. *The Annals of Mathematical Statistics*, **42**: 1977–1991, 1971.

P. J. Huber. Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, **35**:73–101, 1964.

M. G. Kendall and A. Stuart. *The Advanced Theory of Statistics*, volume 1 Distribution Theory. Charles Griffin and Company Limited, London, third edition, 1969.

N. M. Kiefer. Maximum likelihood estimation (mle). University Lecture, https://courses.cit.cornell.edu/econ620/reviewm5.pdf, 2008.

S. Lauritzen. *Extremal Families and Systems of Sufficient Statistics*. Lecture Notes in Statistics, No.49. Springer-Verlag, Berlin-Heidelberg-New York, 1988.

C. Ley and Y. Swan. Stein's density approach and information inequalities. *Electronic Communications in Probability*, **18**:1–14, 2013a.

C. Ley and Y. Swan. Local Pinsker inequalities via Stein's discrete density approach. *IEEE Transactions on Information Theory*, **59**:5584–5591, 2013b.

C. Ley, G. Reinert, and Y. Swan. Approximate computation of expectations: a canonical Stein operator. http://arxiv.org/pdf/1408.2998.pdf, 2014.

C. Ley, G. Reinert, and Y. Swan. Distances between nested densities and a measure of the impact of the prior in Bayesian Statistics. http://arxiv.org/pdf/1510.05826v1.pdf, 2015.

H. M. Luk. *Stein's method for the gamma distribution and related statistical applications*. PhD thesis, University of Southern California, Los Angeles, USA, 1994.

T. Makelainen, K. Schmidt, and G. P. H. Styan. On the existence and uniqueness of the maximum likelihood estimate of a vector-valued parameter in fixed-size samples. *The Annals of Statistics*, **9**:758–767, 1981.

E. Meckes. On Stein's method for multivariate normal approximation. *IMS Collections, High Dimensional Probability V: The Luminy Volume*, **5**:153–178, 2009.

I. Nourdin and G. Peccati. Stein's method on Wiener chaos. *Probability Theory and Related Fields*, **145**:75–118, 2009.

I. Nourdin and G. Peccati. *Normal Approximations with Malliavin Calculus. From Stein's method to universality*. Cambridge Tracts in Mathematics, No.192. Cambridge University Press, 2012.

I. Nourdin, G. Peccati, and Y. Swan. Integration by parts and representation of information functionals. *Proceedings of the 2014 IEEE International Symposium on Information Theory (ISIT)*, pages 2217–2221, 2014.

E. Peköz. Stein's method for geometric approximation. *Journal of Applied Probability*, **33**:707–713, 1996.

E. Peköz and A. Röllin. New rates for exponential approximation and the theorems of Rényi and Yaglom. *The Annals of Probability*, **39**:587–608, 2011.

A. Pickett. *Rates of Convergence of χ^2 approximations via Stein's method*. PhD thesis, University of Oxford, 2004.

J. Pike and H. Ren. Stein's method and the Laplace distribution. *ALEA, Latin American Journal of Probability and Mathematical Statistics*, **11**:571–587, 2014.

I. Pinelis and R. Molzon. Optimal-order bounds on the rate of convergence to normality in the multivariate delta method. *Electronic Journal of Statistics*, **10**:1001–1063, 2016.

M. Raič. A Multivariate CLT for Decomposable Random Vectors with Finite Second Moments. *Journal of Theoretical Probability*, **17**:573–603, 2004.

G. Reinert. A Weak Law of Large Numbers for Empirical Measures via Stein's Method. *The Annals of Probability*, **23**:334–354, 1995.

G. Reinert. Couplings for normal approximations with Stein's method. *Microsurveys in Discrete Probability. D. Aldous, J. Propp eds., Dimacs series. AMS*, pages 193–207, 1998.

G. Reinert and A. Röllin. Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition. *The Annals of Probability*, **37**:2150–2173, 2009.

N. Ross. Fundamentals of Stein's method. *Probability Surveys*, **8**:210–293, 2011.

I. Shevtsova. On the absolute constants in the Berry Esseen type inequalities. *Doklady Mathematics*, **89**:378–381, 2014.

S. D. Silvey. A Note on Maximum-Likelihood in the Case of Dependent Random Variables. *Journal of the Royal Statistical Society. Series B (Methodological)*, **23**: 444–452, 1961.

E. C. Somayya. A Method for finding a Square Root of a 2x2 Matrix. *The Mathematics Education*, **XXXI**:52–54, 1997.

C. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, volume **2**, pages 586–602. Berkeley: University of California Press, 1972.

C. Stein. *Approximate Computation of Expectations*. Lecture Notes-Monograph Series. Institute of Mathematical Statistics, Hayward, California, 1986.

C. Stein, P. Diaconis, S. Holmes, and G. Reinert. Use of exchangeable pairs in the analysis of simulations. *IMS Lecture Notes–Monograph Series*, **46**:1–25, 2004.

F. Tricomi and A. Erdélyi. The asymptotic expansion of a ratio of gamma functions. *Pacific Journal of Mathematics*, **1**:133–142, 1951.

A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.

S. S. Wilks. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *The Annals of Mathematical Statistics*, **9**:60–62, 1938.