

# Bounds for the asymptotic normality of the maximum likelihood estimator using the Delta method

Andreas Anastasiou<sup>1</sup> and Christophe Ley<sup>2</sup>

<sup>1</sup> Oxford University, Oxford, UK

E-mail: andreas.anastasiou@jesus.ox.ac.uk

<sup>2</sup> Ghent University, Gent, Belgium

E-mail: christophe.ley@ugent.be

## Abstract

The asymptotic normality of the Maximum Likelihood Estimator (MLE) is a cornerstone of statistical theory. In the present paper, we provide sharp explicit upper bounds on Zolotarev-type distances between the exact, unknown distribution of the MLE and its limiting normal distribution. Our approach to this fundamental issue is based on a sound combination of the Delta method, Stein's method, Taylor expansions and conditional expectations, for the classical situations where the MLE can be expressed as a function of a sum of independent and identically distributed terms. This encompasses in particular the broad exponential family of distributions.

*Key words:* Delta method, Maximum likelihood estimator, Normal approximation, Stein's method

## 1 Introduction

The asymptotic normality of maximum likelihood estimators (MLEs) is one of the best-known and most fundamental results in mathematical statistics. Under certain regularity conditions (given later in this section), we have the following classical theorem, first discussed in Fisher (1925).

**Theorem 1.1** (Asymptotic Normality of the MLE). *Let  $X_1, \dots, X_n$  be i.i.d. random variables with probability density (or mass) function  $f(x_i|\theta)$ , where  $\theta$  is a scalar parameter. Its true value is denoted as  $\theta_0$ . Assume that the MLE exists and it is unique and conditions (R1)-(R4), see below, are satisfied. Then*

$$\sqrt{ni(\theta_0)} \left( \hat{\theta}_n(\mathbf{X}) - \theta_0 \right) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, 1)$$

where  $i(\theta_0)$  is the expected Fisher information quantity and  $\xrightarrow{d}$  means convergence in distribution.

The aim of the present paper is to complement this qualitative result with a quantitative statement, in other words, to find the best possible approximation for the distance, at finite sample size  $n$ , between the distribution of  $\sqrt{ni(\theta_0)} \left( \hat{\theta}_n(\mathbf{X}) - \theta_0 \right)$  on the one hand and  $\mathcal{N}(0, 1)$  on the other hand. In mathematical terms, for  $Z \sim \mathcal{N}(0, 1)$ , we are interested in the quantity

$$\begin{aligned} & d_H \left( \sqrt{ni(\theta_0)} \left( \hat{\theta}_n(\mathbf{X}) - \theta_0 \right), Z \right) \\ &= \sup_{h \in H} \left| \mathbb{E} \left[ h \left( \sqrt{ni(\theta_0)} \left( \hat{\theta}_n(\mathbf{X}) - \theta_0 \right) \right) \right] - \mathbb{E} [h(Z)] \right| \end{aligned} \quad (1.1)$$

with

$$H = \{h : \mathbb{R} \rightarrow \mathbb{R}, \text{ absolutely continuous and bounded}\}. \quad (1.2)$$

Distances of this type are called *Zolotarev-type distances*. Our main focus will lie on the classical situations where the MLE can be expressed as a function of a sum of independent and identically distributed terms.

Consider an i.i.d. sample of observations  $\mathbf{X} = (X_1, \dots, X_n)$ . Writing  $\hat{\theta}_n(\mathbf{X})$  the MLE of the scalar parameter of interest  $\theta \in \Theta \subseteq \mathbb{R}$ , we are interested in settings where there exists a one-to-one twice differentiable mapping  $q : \Theta \rightarrow \mathbb{R}$  such that

$$q\left(\hat{\theta}_n(\mathbf{X})\right) = \frac{1}{n} \sum_{i=1}^n g(X_i) \quad (1.3)$$

for some  $g : \mathbb{R} \rightarrow \mathbb{R}$ . Situations of this kind are all but rare; with  $f(x|\theta)$  the probability density (or mass) function, classical examples include

- the normal distribution with density  $f(x|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right)$ ,  $x \in \mathbb{R}$ , for which  $\mu \in \mathbb{R}$  is our unknown parameter, whereas  $\sigma > 0$  is considered to be known. The MLE for  $\theta = \mu$  is

$$\hat{\theta}_n(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n X_i;$$

- the normal distribution, where now the mean  $\mu$  is known and  $\theta = \sigma^2$  represents the unknown parameter, with

$$\hat{\theta}_n(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2;$$

- 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  $\sigma$  is the unknown scale parameter and  $\alpha > 0$  is fixed. The MLE for  $\theta = \sigma$  is defined through

$$\left(\hat{\theta}_n(\mathbf{X})\right)^\alpha = \frac{1}{n} \sum_{i=1}^n X_i^\alpha;$$

- the Laplace scale model with density  $f(x|\sigma) = \frac{1}{2\sigma} \exp(-|x|/\sigma)$ ,  $\theta = \sigma > 0$ , over  $\mathbb{R}$ , for which

$$\hat{\theta}_n(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n |X_i|.$$

Moreover, the broad one-parameter exponential families do satisfy condition (1.3); see Proposition 3.1 for details. Hence, our results do apply to most of the well-known distributions.

We now present in detail the notation and general assumptions made throughout the paper. We write  $E_\theta[\cdot]$  the expectation under the specific value  $\theta$  of the parameter. In line with the notation used above, the joint density or probability mass function of  $\mathbf{X}$  is written  $f(\mathbf{x}|\theta)$ . The true, unknown value of the parameter is  $\theta_0$  and  $\Theta$  denotes the parameter space. For  $X_i = x_i$  some observed values, the likelihood function is denoted by  $L(\theta; \mathbf{x}) = f(\mathbf{x}|\theta)$  and we denote its natural logarithm, called the log-likelihood function, by  $l(\theta; \mathbf{x})$ . The derivatives of the log-likelihood function with respect to  $\theta$  are  $l'(\theta; \mathbf{x}), l''(\theta; \mathbf{x}), \dots, l^{(j)}(\theta; \mathbf{x})$ , for  $j$  any integer greater than 2, and  $i(\theta)$  denotes the expected Fisher information number for one random variable. Whenever the MLE exists and is also unique, we will write it as before under the form  $\hat{\theta}_n(\mathbf{X})$ . For  $\Theta$  being an open interval, we use the results in Makelainen et al. (1981) to secure the existence and uniqueness of the MLE. Thus, it suffices to assume that:

(A1) The log-likelihood function  $l(\theta; \mathbf{x})$  is a twice continuously differentiable function with respect to  $\theta$  and the parameter varies in an open interval  $(a, b)$ , where  $a, b \in \mathbb{R} \cup \{-\infty, \infty\}$  and  $a < b$ ;

(A2)  $\lim_{\theta \rightarrow a, b} l(\theta; \mathbf{x}) = -\infty$ ;

(A3)  $l''(\theta; \mathbf{x}) < 0$  at every point  $\theta \in (a, b)$  for which  $l'(\theta; \mathbf{x}) = 0$ .

Note that we tacitly assume those conditions in Theorem 1.1 when requiring existence and uniqueness of the MLE. Asymptotic normality further requires the following sufficient regularity conditions:

(R1) the parameter is identifiable, which means 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  $\theta$ , the third derivative is continuous in  $\theta$  and  $\int f(x|\theta) dx$  can be differentiated three times under the integral sign;

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

$$\left| \frac{d^3}{d\theta^3} \log f(x|\theta) \right| \leq M(x) \quad \forall x \in \mathbb{X}, \quad \theta_0 - \epsilon < \theta < \theta_0 + \epsilon,$$

with  $E_{\theta_0}[M(X)] < \infty$ ;

(R4)  $i(\theta) > 0, \forall \theta \in \Theta$ .

These conditions, in particular (R2), ensure that, provided the respective expressions exist,  $E_{\theta}[l'(\theta; \mathbf{X})] = 0$  and  $\text{Var}_{\theta}[l'(\theta; \mathbf{X})] = ni(\theta)$ . These conditions form the basis of Theorem 1.1 above; see page 472 of Casella and Berger (2002) for a basic sketch of the proof.

The first paper to address the problem of assessing the accuracy of the asymptotic normal approximation for MLE is Anastasiou and Reinert (2015). After deriving general bounds on Zolotarev-type distances, they use the *bounded Wasserstein distance*  $d_{bW}$ , which is also known as Fortet-Mourier distance (see, e.g., Nourdin and Peccati (2012)) and is linked to the Kolmogorov distance ( $H$  is the class of indicator functions of half-spaces) via  $d_K(\cdot, \cdot) \leq 2\sqrt{d_{bW}(\cdot, \cdot)}$ . We state in Theorem 2.4 of Section 2 the bound obtained in that paper. For the broad class of distributions satisfying (1.3), our bound is better than, or at least as good as, the Anastasiou and Reinert (2015) bound (hereafter referred to as AR-bound) both in terms of sharpness and simplicity. The tools we use to reach this result are the Delta method, Stein's method for normal approximation, Taylor expansions and conditional expectations.

The paper is organised as follows. Our new upper bound is described, proved and compared to the AR-bound in Section 2. In Section 3 we then apply our results to the class of one-parameter exponential family distributions and treat some specific examples in detail.

## 2 New bounds on the distance to the normal distribution for the MLE

In order to obtain bounds on the aforementioned distance, we partly employ the following lemma. From now on, unless otherwise stated,  $\|\cdot\|$  denotes the infinity norm  $\|\cdot\|_{\infty}$ .

**Lemma 2.1** (Reinert (1998)). Let  $Y_1, \dots, Y_n$  be independent random variables with  $\mathbb{E}(Y_i) = 0$ ,  $\text{Var}(Y_i) = \sigma^2 > 0$  and  $\mathbb{E}[|Y_i|^3] < \infty$ . Let  $W = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i$ , with  $\mathbb{E}(W) = 0$ ,  $\text{Var}(W) = \sigma^2$  and let  $K \sim \mathcal{N}(0, \sigma^2)$ . Then for any function  $h \in H$ , with  $H$  given in (1.2), one has

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

As we shall see below, our strategy consists in benefiting from the special form of  $q(\hat{\theta}_n(\mathbf{X}))$ , which is a sum of random variables and thus allows us to use the sharp bound of this lemma. It is precisely at this point that the Delta method comes into play: abusing notations and language, instead of comparing  $\hat{\theta}_n(\mathbf{X})$  to  $Z \sim \mathcal{N}(0, 1)$  we rather compare  $q(\hat{\theta}_n(\mathbf{X}))$  to  $Z$ , and then bound the distance between  $\hat{\theta}_n(\mathbf{X})$  and  $q(\hat{\theta}_n(\mathbf{X}))$ . The outcome of this approach is the next theorem, the main result of the present paper.

**Theorem 2.2.** Let  $X_1, \dots, X_n$  be i.i.d. random variables with probability density (or mass) function  $f(x_i|\theta)$  and let  $Z \sim \mathcal{N}(0, 1)$ . Assume that (A1)-(A3) and the regularity conditions (R1)-(R4) are satisfied, and hence the MLE  $\hat{\theta}_n(\mathbf{X})$  exists and is unique. Furthermore let  $q : \Theta \rightarrow \mathbb{R}$  be a one-to-one twice differentiable function with  $q'(\theta) \neq 0 \forall \theta \in \Theta$  and such that  $q(\hat{\theta}_n(\mathbf{X})) = \frac{1}{n} \sum_{i=1}^n g(X_i)$ , where the mapping  $g : \mathbb{R} \rightarrow \mathbb{R}$  is such that  $\mathbb{E}[|g(X_1) - q(\theta_0)|^3] < \infty$  for  $\theta_0 \in \Theta$  the true value of the parameter. Also, there exists a positive constant  $0 < \epsilon = \epsilon(\theta_0)$  as in (R3) with  $(\theta_0 - \epsilon, \theta_0 + \epsilon) \subset \Theta$ . Then, for any  $h \in H$  as in (1.2) we have

$$\begin{aligned} & \left| \mathbb{E} \left[ h \left( \sqrt{n i(\theta_0)} \left( \hat{\theta}_n(\mathbf{X}) - \theta_0 \right) \right) \right] - \mathbb{E}[h(Z)] \right| \\ & \leq \frac{\|h'\|}{\sqrt{n}} \left( 2 + \frac{[i(\theta_0)]^{\frac{3}{2}}}{|q'(\theta_0)|^3} \mathbb{E}[|g(X_1) - q(\theta_0)|^3] \right) \\ & + \mathbb{E} \left[ \left( \hat{\theta}_n(\mathbf{X}) - \theta_0 \right)^2 \right] \left( 2 \frac{\|h\|}{\epsilon^2} \mathbb{1}_{\{\exists \theta \in \Theta : q(\theta) \neq \theta\}} + \frac{\|h'\| \sqrt{n i(\theta_0)}}{2 |q'(\theta_0)|} \sup_{\theta: |\theta - \theta_0| \leq \epsilon} |q''(\theta)| \right). \end{aligned} \quad (2.1)$$

*Proof.* The asymptotic normality of the MLE is explicitly stated in Theorem 1.1. Applying the widely known Delta method to this result in combination with the requirement  $q'(\theta_0) \neq 0$  we obtain

$$\frac{\sqrt{n i(\theta_0)}}{q'(\theta_0)} \left( q \left( \hat{\theta}_n(\mathbf{X}) \right) - q(\theta_0) \right) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, 1), \quad (2.2)$$

with  $q \left( \hat{\theta}_n(\mathbf{X}) \right) = \frac{1}{n} \sum_{i=1}^n g(X_i)$ . Using the triangle inequality we get that

$$\begin{aligned} & \left| \mathbb{E} \left[ h \left( \sqrt{n i(\theta_0)} \left( \hat{\theta}_n(\mathbf{X}) - \theta_0 \right) \right) \right] - \mathbb{E}[h(Z)] \right| \\ & \leq \left| \mathbb{E} \left[ h \left( \frac{\sqrt{n i(\theta_0)}}{q'(\theta_0)} \left( q \left( \hat{\theta}_n(\mathbf{X}) \right) - q(\theta_0) \right) \right) \right] - \mathbb{E}[h(Z)] \right| \end{aligned} \quad (2.3)$$

$$+ \left| \mathbb{E} \left[ h \left( \sqrt{n i(\theta_0)} \left( \hat{\theta}_n(\mathbf{X}) - \theta_0 \right) \right) - h \left( \frac{\sqrt{n i(\theta_0)}}{q'(\theta_0)} \left( q \left( \hat{\theta}_n(\mathbf{X}) \right) - q(\theta_0) \right) \right) \right] \right|. \quad (2.4)$$

We first start to obtain an upper bound for (2.3) using (indirectly) Stein's method via Lemma 2.1. Some simple rewriting yields

$$\begin{aligned} \frac{\sqrt{n i(\theta_0)}}{q'(\theta_0)} \left( q \left( \hat{\theta}_n(\mathbf{X}) \right) - q(\theta_0) \right) &= \frac{\sqrt{n i(\theta_0)}}{q'(\theta_0)} \left( \frac{1}{n} \sum_{i=1}^n g(X_i) - q(\theta_0) \right) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \frac{\sqrt{i(\theta_0)}}{q'(\theta_0)} (g(X_i) - q(\theta_0)) \right\} = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i, \end{aligned}$$

where  $Y_i = \frac{\sqrt{i(\theta_0)}}{q'(\theta_0)} (g(X_i) - q(\theta_0))$ ,  $i = 1, 2, \dots, n$  and, obviously, the  $Y_i$ 's are independent and identically distributed random variables. The Central Limit Theorem applied to  $\frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i$  implies  $\frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_i - \mathbb{E}(Y_1)) \xrightarrow{d} \mathcal{N}(0, \text{Var}(Y_1))$ . From (2.2) we know however that  $\frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i \xrightarrow{d} \mathcal{N}(0, 1)$ ; comparing the two asymptotic results reveals that, necessarily (since two normal distributions can only be equal if their expectations and variances are the same), we have

$$\mathbb{E}[g(X_1)] = q(\theta_0) \quad \text{and} \quad \text{Var}[g(X_1)] = \frac{(q'(\theta_0))^2}{i(\theta_0)},$$

where condition (R4) allows to divide by  $i(\theta_0)$ . Hence,  $\mathbb{E}[Y_i] = 0$  and  $\text{Var}[Y_i] = 1$  (as Lemma 2.1 requires). Applying the result of the lemma we get

$$\begin{aligned} &\left| \mathbb{E} \left[ h \left( \frac{\sqrt{n i(\theta_0)}}{q'(\theta_0)} \left( q \left( \hat{\theta}_n(\mathbf{X}) \right) - q(\theta_0) \right) \right) \right] - \mathbb{E}[h(Z)] \right| \\ &\leq \frac{\|h'\|}{\sqrt{n}} \left( 2 + \frac{[i(\theta_0)]^{\frac{3}{2}}}{|q'(\theta_0)|^3} \mathbb{E} \left[ |g(X_1) - q(\theta_0)|^3 \right] \right). \end{aligned} \quad (2.5)$$

Now we are searching for an upper bound on (2.4). Since the case  $q(\theta) = \theta$  is obvious, we from here on assume that  $q(\theta) \neq \theta$ . To do so, we denote by

$$\begin{aligned} A := A(q, \theta_0, \mathbf{X}) &:= h \left( \sqrt{n i(\theta_0)} \left( \hat{\theta}_n(\mathbf{X}) - \theta_0 \right) \right) \\ &\quad - h \left( \frac{\sqrt{n i(\theta_0)}}{q'(\theta_0)} \left( q \left( \hat{\theta}_n(\mathbf{X}) \right) - q(\theta_0) \right) \right) \end{aligned}$$

and our scope is to find an upper bound for  $|\mathbb{E}[A]|$ . Using the law of total expectation related to conditioning on  $|\hat{\theta}_n(\mathbf{X}) - \theta_0| > \epsilon$  or  $|\hat{\theta}_n(\mathbf{X}) - \theta_0| \leq \epsilon$  and the triangle inequality we obtain

$$\begin{aligned} |\mathbb{E}[A]| &= \left| \mathbb{E} \left[ A \mid \left| \hat{\theta}_n(\mathbf{X}) - \theta_0 \right| > \epsilon \right] \mathbb{P} \left( \left| \hat{\theta}_n(\mathbf{X}) - \theta_0 \right| > \epsilon \right) \right. \\ &\quad \left. + \mathbb{E} \left[ A \mid \left| \hat{\theta}_n(\mathbf{X}) - \theta_0 \right| \leq \epsilon \right] \mathbb{P} \left( \left| \hat{\theta}_n(\mathbf{X}) - \theta_0 \right| \leq \epsilon \right) \right| \\ &\leq \left| \mathbb{E} \left[ A \mid \left| \hat{\theta}_n(\mathbf{X}) - \theta_0 \right| > \epsilon \right] \right| \mathbb{P} \left( \left| \hat{\theta}_n(\mathbf{X}) - \theta_0 \right| > \epsilon \right) \\ &\quad + \left| \mathbb{E} \left[ A \mid \left| \hat{\theta}_n(\mathbf{X}) - \theta_0 \right| \leq \epsilon \right] \right| \mathbb{P} \left( \left| \hat{\theta}_n(\mathbf{X}) - \theta_0 \right| \leq \epsilon \right) \\ &\leq \mathbb{E} \left[ |A| \mid \left| \hat{\theta}_n(\mathbf{X}) - \theta_0 \right| > \epsilon \right] \mathbb{P} \left( \left| \hat{\theta}_n(\mathbf{X}) - \theta_0 \right| > \epsilon \right) \\ &\quad + \mathbb{E} \left[ |A| \mid \left| \hat{\theta}_n(\mathbf{X}) - \theta_0 \right| \leq \epsilon \right] \mathbb{P} \left( \left| \hat{\theta}_n(\mathbf{X}) - \theta_0 \right| \leq \epsilon \right). \end{aligned}$$

Markov's inequality and the elementary results of  $\mathbb{P}\left(\left|\hat{\theta}_n(\mathbf{X}) - \theta_0\right| \leq \epsilon\right) \leq 1$  and  $|A| \leq 2\|h\|$  further yield

$$|E[A]| \leq 2\|h\| \frac{E\left[\left(\hat{\theta}_n(\mathbf{X}) - \theta_0\right)^2\right]}{\epsilon^2} + E\left[|A| \mid \left|\hat{\theta}_n(\mathbf{X}) - \theta_0\right| \leq \epsilon\right]. \quad (2.6)$$

We now focus on the conditional expectation on the right-hand side of (2.6). A second-order Taylor expansion of  $q\left(\hat{\theta}_n(\mathbf{x})\right)$  about  $\theta_0$  gives

$$q\left(\hat{\theta}_n(\mathbf{x})\right) = q(\theta_0) + \left(\hat{\theta}_n(\mathbf{x}) - \theta_0\right) q'(\theta_0) + \frac{1}{2} \left(\hat{\theta}_n(\mathbf{x}) - \theta_0\right)^2 q''(\theta^*), \quad (2.7)$$

for  $\theta^*$  between  $\hat{\theta}_n(\mathbf{x})$  and  $\theta_0$ . Since we assume that  $q'(\theta) \neq 0 \forall \theta \in \Theta$ , we can multiply both sides in (2.7) with  $\frac{\sqrt{ni(\theta_0)}}{q'(\theta_0)}$ . Rearranging the terms, we obtain

$$\frac{\sqrt{ni(\theta_0)}\left(q\left(\hat{\theta}_n(\mathbf{x})\right) - q(\theta_0)\right)}{q'(\theta_0)} = \sqrt{ni(\theta_0)}\left(\hat{\theta}_n(\mathbf{x}) - \theta_0\right) + \frac{\sqrt{ni(\theta_0)}}{2q'(\theta_0)} q''(\theta^*) \left(\hat{\theta}_n(\mathbf{x}) - \theta_0\right)^2.$$

Using the above result along with another first-order Taylor expansion (recall that  $\sqrt{n}\left(\hat{\theta}_n(\mathbf{X}) - \theta_0\right)^2 = o_P(1)$  as  $n \rightarrow \infty$ ) gives

$$\begin{aligned} h\left(\sqrt{ni(\theta_0)}\left(\hat{\theta}_n(\mathbf{x}) - \theta_0\right)\right) - h\left(\frac{\sqrt{ni(\theta_0)}}{q'(\theta_0)}\left(q\left(\hat{\theta}_n(\mathbf{x})\right) - q(\theta_0)\right)\right) \\ = -\frac{\sqrt{ni(\theta_0)}}{2q'(\theta_0)} q''(\theta^*) \left(\hat{\theta}_n(\mathbf{x}) - \theta_0\right)^2 h'(t(\mathbf{x})), \end{aligned} \quad (2.8)$$

where  $t(\mathbf{x})$  is between  $\sqrt{ni(\theta_0)}\left(\hat{\theta}_n(\mathbf{x}) - \theta_0\right)$  and  $\frac{\sqrt{ni(\theta_0)}}{q'(\theta_0)}\left(q\left(\hat{\theta}_n(\mathbf{x})\right) - q(\theta_0)\right)$ . Equality (2.8) combined with Lemma 2.1 in [Anastasiou and Reinert \(2015\)](#) related to conditional expectations yields

$$\begin{aligned} & E\left[|A| \mid \left|\hat{\theta}_n(\mathbf{X}) - \theta_0\right| \leq \epsilon\right] \\ &= E\left[\left|-\frac{\sqrt{ni(\theta_0)}}{2q'(\theta_0)} q''(\theta^*) \left(\hat{\theta}_n(\mathbf{X}) - \theta_0\right)^2 h'(t(\mathbf{X}))\right| \mid \left|\hat{\theta}_n(\mathbf{X}) - \theta_0\right| \leq \epsilon\right] \\ &\leq \frac{\|h'\| \sqrt{ni(\theta_0)}}{2|q'(\theta_0)|} E\left[|q''(\theta^*)| \left(\hat{\theta}_n(\mathbf{X}) - \theta_0\right)^2 \mid \left|\hat{\theta}_n(\mathbf{X}) - \theta_0\right| \leq \epsilon\right] \\ &\leq \frac{\|h'\| \sqrt{ni(\theta_0)}}{2|q'(\theta_0)|} \sup_{\theta: |\theta - \theta_0| \leq \epsilon} |q''(\theta)| E\left[\left(\hat{\theta}_n(\mathbf{X}) - \theta_0\right)^2 \mid \left|\hat{\theta}_n(\mathbf{X}) - \theta_0\right| \leq \epsilon\right] \\ &\leq \frac{\|h'\| \sqrt{ni(\theta_0)}}{2|q'(\theta_0)|} \sup_{\theta: |\theta - \theta_0| \leq \epsilon} |q''(\theta)| E\left[\left(\hat{\theta}_n(\mathbf{X}) - \theta_0\right)^2\right]. \end{aligned} \quad (2.9)$$

Combining the bounds in (2.5), (2.6) and (2.9) gives the result of the theorem.  $\blacksquare$

*Remark 2.3.* (1) The convergence of the second and third terms is governed by the asymptotic behaviour of  $\mathbb{E} \left[ \left( \hat{\theta}_n(\mathbf{X}) - \theta_0 \right)^2 \right]$ , whose rate of convergence is  $\mathcal{O} \left( \frac{1}{n} \right)$ . This result is obtained using the decomposition

$$\mathbb{E} \left[ \left( \hat{\theta}_n(\mathbf{X}) - \theta_0 \right)^2 \right] = \text{Var}[\hat{\theta}_n(\mathbf{X})] + \text{bias}^2[\hat{\theta}_n(\mathbf{X})]. \quad (2.10)$$

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

$$n \text{Var}[\hat{\theta}_n(\mathbf{X})] \xrightarrow{n \rightarrow \infty} [i(\theta_0)]^{-1},$$

and hence the variance of the MLE is of order  $\frac{1}{n}$ . In addition, from Theorem 1.1 the bias of the MLE is of order  $\frac{1}{\sqrt{n}}$ ; see also Cox and Snell (1968), where no explicit conditions are given. Combining these two results and using (2.10) shows that the mean squared error of the MLE is of order  $\frac{1}{n}$ .

(2) In the simplest possible situation where  $\hat{\theta}_n(\mathbf{X})$  is already a sum of i.i.d. terms,  $q(x) = x$  and hence our upper bound simplifies to

$$\left| \mathbb{E} \left[ h \left( \sqrt{n i(\theta_0)} \left( \hat{\theta}_n(\mathbf{X}) - \theta_0 \right) \right) \right] - \mathbb{E} [h(Z)] \right| \leq \frac{\|h'\|}{\sqrt{n}} \left( 2 + [i(\theta_0)]^{\frac{3}{2}} \mathbb{E} \left[ |g(X_1) - \theta_0|^3 \right] \right),$$

which is equivalent to Lemma 2.1.

In order to appreciate the sharpness and simplicity of our bound (2.1), we compare it to the AR-bound. To this end, we now state the main result of Anastasiou and Reinert (2015).

**Theorem 2.4** (Anastasiou and Reinert (2015)). *Let  $X_1, X_2, \dots, X_n$  be i.i.d. random variables with density or frequency function  $f(x_i|\theta)$  such that the regularity conditions (R1)-(R4) are satisfied and that the MLE,  $\hat{\theta}_n(\mathbf{X})$ , exists and it is unique. Assume that  $\mathbb{E} \left[ \left| \left( \frac{d}{d\theta} \log f(X_1|\theta) \right)_{\theta=\theta_0} \right|^3 \right] < \infty$  and that  $\mathbb{E} \left[ \left( \hat{\theta}_n(\mathbf{X}) - \theta_0 \right)^4 \right] < \infty$ . Let  $0 < \epsilon = \epsilon(\theta_0)$  be such that  $(\theta_0 - \epsilon, \theta_0 + \epsilon) \subset \Theta$  as in (R3) and let  $Z \sim \mathcal{N}(0, 1)$ . Then for any function  $h$  that is absolutely continuous and bounded,*

$$\begin{aligned} & \left| \mathbb{E} \left[ h \left( \sqrt{n i(\theta_0)} \left( \hat{\theta}_n(\mathbf{X}) - \theta_0 \right) \right) \right] - \mathbb{E} [h(Z)] \right| \\ & \leq \frac{\|h'\|}{\sqrt{n}} \left( 2 + \frac{1}{[i(\theta_0)]^{\frac{3}{2}}} \mathbb{E} \left[ \left| \left( \frac{d}{d\theta} \log f(X_1|\theta) \right)_{\theta=\theta_0} \right|^3 \right] \right) \\ & + 2\|h\| \frac{\mathbb{E} \left[ \left( \hat{\theta}_n(\mathbf{X}) - \theta_0 \right)^2 \right]}{\epsilon^2} + \frac{\|h'\|}{\sqrt{n i(\theta_0)}} \left\{ \mathbb{E} \left[ |R_2(\theta_0; \mathbf{X})| \mid |\hat{\theta}_n(\mathbf{X}) - \theta_0| \leq \epsilon \right] \right. \\ & \left. + \frac{1}{2} \left[ \mathbb{E} \left[ \left( \sup_{\theta: |\theta - \theta_0| \leq \epsilon} |l^{(3)}(\theta; \mathbf{X})| \right)^2 \mid |\hat{\theta}_n(\mathbf{X}) - \theta_0| \leq \epsilon \right] \right]^{\frac{1}{2}} \left[ \mathbb{E} \left[ \left( \hat{\theta}_n(\mathbf{X}) - \theta_0 \right)^4 \right] \right]^{\frac{1}{2}} \right\}, \end{aligned}$$

where

$$R_2(\theta_0; \mathbf{x}) = (\hat{\theta}_n(\mathbf{x}) - \theta_0) (l''(\theta_0; \mathbf{x}) + n i(\theta_0)). \quad (2.11)$$

Obvious observations are that the AR-bound requires finiteness of the fourth moment of  $\hat{\theta}_n(\mathbf{X}) - \theta_0$  and that this bound is more complicated than ours. Let us now comment on the bounds term by term.

- In the first term of the bounds, the different positioning of the expected Fisher information number is explained by the fact that we apply Lemma 2.1 to the standardised version of  $g(X_1), g(X_2), \dots, g(X_n)$ , which have variance  $\frac{[q'(\theta_0)]^2}{i(\theta_0)}$ , while Anastasiou and Reinert (2015) obtain the result by applying the lemma after standardising  $l'(\theta_0; X_1), l'(\theta_0; X_2), \dots, l'(\theta_0; X_n)$ , which have variance equal to  $i(\theta_0)$ .
- The second and third terms vanish in our bound when  $q(\theta) = \theta \forall \theta \in \Theta$ , while the AR-bound does not take this simplification into account. In addition, when  $q(\theta) \neq \theta$  the second term is the same in both bounds, whereas the third term in our bound reads

$$\mathbb{E} \left[ \left( \hat{\theta}_n(\mathbf{X}) - \theta_0 \right)^2 \right] \frac{\|h'\| \sqrt{n i(\theta_0)}}{2 |q'(\theta_0)|} \sup_{\theta: |\theta - \theta_0| \leq \epsilon} |q''(\theta)| \quad (2.12)$$

and is to be compared to

$$\begin{aligned} & \frac{\|h'\|}{2\sqrt{n i(\theta_0)}} \left[ \mathbb{E} \left[ \left( \sup_{\theta: |\theta - \theta_0| \leq \epsilon} |l^{(3)}(\theta; \mathbf{X})| \right)^2 \middle| |\hat{\theta}_n(\mathbf{X}) - \theta_0| \leq \epsilon \right] \right]^{\frac{1}{2}} \left[ \mathbb{E} \left[ \left( \hat{\theta}_n(\mathbf{X}) - \theta_0 \right)^4 \right] \right]^{\frac{1}{2}} \\ & + \frac{\|h'\|}{\sqrt{n i(\theta_0)}} \mathbb{E} \left[ |R_2(\theta_0; \mathbf{X})| \middle| |\hat{\theta}_n(\mathbf{X}) - \theta_0| \leq \epsilon \right] \end{aligned} \quad (2.13)$$

where  $R_2(\theta_0; \mathbf{x}) = (\hat{\theta}_n(\mathbf{x}) - \theta_0) (l''(\theta_0; \mathbf{x}) + n i(\theta_0))$ . The second derivative,  $q''(\theta)$ , plays in our bound the role of  $l^{(3)}(\theta; \mathbf{X})$ , up to an important difference:  $l^{(3)}(\theta; \mathbf{X})$  is a sum. Consequently, the first term in (2.13) has  $\sqrt{n}$  in its numerator, exactly as in (2.12). The distinct positioning of the information quantity  $i(\theta_0)$  has the same reason as explained above. Besides the obvious additional term in the AR bound (the second term in (2.13)), our bound is also clearly sharper at the level of moments of  $\hat{\theta}_n(\mathbf{X}) - \theta_0$  since

$$\mathbb{E} \left[ \left( \hat{\theta}_n(\mathbf{X}) - \theta_0 \right)^2 \right] \leq \left[ \mathbb{E} \left[ \left( \hat{\theta}_n(\mathbf{X}) - \theta_0 \right)^4 \right] \right]^{\frac{1}{2}}$$

by the Cauchy-Schwarz inequality.

From this comparison one sees that our new bound is simpler and, moreover, has one term less. This is particularly striking in the simplest possible setting where  $\hat{\theta}_n(\mathbf{X})$  is a sum of i.i.d. terms, where our bound clearly improves on the AR-bound. An advantage of the AR-bound is its wider applicability as it works for all MLE settings, even when an analytic expression of the MLE is not known.

### 3 Calculation of the bound in different scenarios

In this section we shall consider different examples for which we explicitly calculate our upper bound from Theorem 2.2 and compare it to the AR-bound. To further assess its accuracy, we simulate data from various distributions and compare our bound to the actual distance between the unknown exact law of the MLE and its asymptotic normal law, for distinct values of the sample size  $n$ .



### 3.1 Bounds for one-parameter exponential families

The probability density (or mass) function for one-parameter exponential families is given by

$$f(x|\theta) = \exp \{k(\theta)T(x) - A(\theta) + S(x)\} \mathbb{1}_{\{x \in B\}}, \quad (3.1)$$

where the set  $B = \{x : f(x|\theta) > 0\}$  is the support of the density and does not depend on  $\theta$ ;  $k(\theta)$  and  $A(\theta)$  are functions of the parameter;  $T(x)$  and  $S(x)$  are functions only of the data. Whenever  $k(\theta) = \theta$  we have the so-called *canonical case*, where  $\theta$  and  $T(X)$  are called the *natural parameter* and *natural observation* (Casella and Berger, 2002). The identifiability constraint in (R1) entails that  $k'(\theta) \neq 0$  (Geyer, 2013), an important detail for the following investigation.

**Proposition 3.1.** *Suppose  $X_1, \dots, X_n$  are i.i.d. with probability density (or mass) function that can be expressed in the form of (3.1). Assume that (A1)-(A3) and the regularity conditions (R1)-(R4) are satisfied, and hence the MLE  $\hat{\theta}_n(\mathbf{X})$  exists and is unique. Let  $\theta \mapsto D(\theta) = \frac{A'(\theta)}{k'(\theta)}$  be invertible. Then  $q(\cdot) = D(\cdot)$ , with  $q : \Theta \rightarrow \mathbb{R}$  as in Theorem 2.2.*

*Proof.* Using (3.1), we have that

$$\begin{aligned} L(\theta; \mathbf{x}) &= \prod_{i=1}^n f(x_i|\theta) = \exp \left\{ k(\theta) \sum_{i=1}^n T(x_i) - n A(\theta) + \sum_{i=1}^n S(x_i) \right\}, \\ l(\theta; \mathbf{x}) &= k(\theta) \sum_{i=1}^n T(x_i) - n A(\theta) + \sum_{i=1}^n S(x_i), \end{aligned}$$

and hence

$$l'(\theta; \mathbf{x}) = k'(\theta) \sum_{i=1}^n T(x_i) - n A'(\theta) = 0 \iff D(\theta) = \frac{1}{n} \sum_{i=1}^n T(x_i),$$

which means that  $\hat{\theta}_n(\mathbf{X}) = D^{-1}(\frac{1}{n} \sum_{i=1}^n T(X_i))$  under the invertibility assumption for  $D(\theta)$ . The claim readily follows.  $\blacksquare$

This result hence shows that, as announced in the Introduction, the broad one-parameter exponential families do satisfy (1.3). Consequently, Theorem 2.2 can be applied to (3.1), resulting in

**Corollary 3.2.** *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 (A1)-(A3) and (R1)-(R4) are satisfied, and hence the MLE  $\hat{\theta}_n(\mathbf{X})$  exists and is unique. With  $Z \sim \mathcal{N}(0, 1)$ ,  $h \in H$  as defined in (1.2) and  $0 < \epsilon = \epsilon(\theta_0) : (\theta_0 - \epsilon, \theta_0 + \epsilon) \subset \Theta$  as in (R3), we obtain*

$$\begin{aligned} & \left| \mathbb{E} \left[ h \left( \sqrt{n} i(\theta_0) \left( \hat{\theta}_n(\mathbf{X}) - \theta_0 \right) \right) \right] - \mathbb{E} [h(Z)] \right| \\ & \leq \frac{\|h'\|}{\sqrt{n}} \left( 2 + \frac{|k'(\theta_0)|^3 \mathbb{E} \left[ |T(X_1) - D(\theta_0)|^3 \right]}{|A''(\theta_0) - k''(\theta_0)D(\theta_0)|^{\frac{3}{2}}} \right) \\ & + \mathbb{E} \left[ \left( \hat{\theta}_n(\mathbf{X}) - \theta_0 \right)^2 \right] \left( 2 \frac{\|h\|}{\epsilon^2} \mathbb{1}_{\{\exists \theta \in \Theta : D(\theta) \neq \theta\}} \right. \\ & \quad \left. + \frac{\|h'\| \sqrt{n} |k'(\theta_0)|}{2 \sqrt{|A''(\theta_0) - k''(\theta_0)D(\theta_0)|}} \sup_{\theta : |\theta - \theta_0| \leq \epsilon} |D''(\theta)| \right). \end{aligned}$$

*Proof.* We readily have

$$i(\theta_0) = \mathbb{E}[-l''(\theta_0; X_1)] = A''(\theta_0) - k''(\theta_0)\mathbb{E}[T(X_1)] = \frac{A''(\theta_0)k'(\theta_0) - k''(\theta_0)A'(\theta_0)}{k'(\theta_0)}$$

and  $q'(\theta_0) = \frac{A''(\theta_0)k'(\theta_0) - k''(\theta_0)A'(\theta_0)}{[k'(\theta_0)]^2}$ . Combining these two results,

$$\frac{\sqrt{i(\theta_0)}}{|q'(\theta_0)|} = \frac{|k'(\theta_0)|^{\frac{3}{2}}}{\sqrt{|A''(\theta_0)k'(\theta_0) - k''(\theta_0)A'(\theta_0)|}} = \frac{|k'(\theta_0)|}{\sqrt{|A''(\theta_0) - k''(\theta_0)D(\theta_0)|}}.$$

This result, along with the fact that  $g(x) = T(x)$  and  $q(\theta) = D(\theta)$  by Proposition 3.1, allows to deduce the announced upper bound from Theorem 2.2.  $\blacksquare$

*Remark 3.3.* It is particularly interesting to spell out this bound in the canonical case  $k(\theta) = \theta$ . Since then  $k'(\theta) = 1$ ,  $D(\theta) = A'(\theta)$ , we find

$$\begin{aligned} \left| \mathbb{E} \left[ h \left( \sqrt{n i(\theta_0)} \left( \hat{\theta}_n(\mathbf{X}) - \theta_0 \right) \right) \right] - \mathbb{E} [h(Z)] \right| &\leq \frac{\|h'\|}{\sqrt{n}} \left( 2 + \frac{\mathbb{E} \left[ |T(X_1) - A'(\theta_0)|^3 \right]}{|A''(\theta_0)|^{\frac{3}{2}}} \right) \\ &+ \mathbb{E} \left[ \left( \hat{\theta}_n(\mathbf{X}) - \theta_0 \right)^2 \right] \left( 2 \frac{\|h\|}{\epsilon^2} \mathbb{1} \{ \exists \theta \in \Theta : A'(\theta) \neq \theta \} \right. \\ &\quad \left. + \frac{\|h'\| \sqrt{n}}{2 \sqrt{|A''(\theta_0)|}} \sup_{\theta: |\theta - \theta_0| \leq \epsilon} |A'''(\theta)| \right). \end{aligned}$$

As  $i(\theta) = A''(\theta)$  and  $l''(\theta; \mathbf{X}) = -nA''(\theta)$ ,  $R_2(\theta; \mathbf{x}) = 0$  and straightforward manipulations show that all terms in the AR-bound coincide with those in our bound, except for  $\left[ \mathbb{E} \left[ \left( \hat{\theta}_n(\mathbf{X}) - \theta_0 \right)^4 \right] \right]^{\frac{1}{2}}$ , making the AR-bound less sharp than ours. However, Anastasiou and Reinert (2015) have shown that, in the canonical exponential setting, their bound can actually have an  $\mathbb{E} \left[ \left( \hat{\theta}_n(\mathbf{X}) - \theta_0 \right)^2 \right]$  factor, implying that both bounds are exactly the same in the canonical case. In order to get an idea of how our bound improves on the AR-bound in non-canonical cases, we treat the exponential distribution under a non-canonical parametrisation in Subsection 3.3.

### 3.2 Bounds for the Generalised Gamma distribution

Let us consider  $X_1, \dots, X_n$  i.i.d. random variables from the Generalized Gamma  $\text{GG}(\theta, d, p)$  distribution, where the shape parameters  $d, p > 0$  are considered to be known and the scale parameter  $\theta$  is the unknown parameter of interest. The Generalised Gamma distribution includes many other known distributions as special cases: the Weibull for  $d = p$ , the Gamma when  $p = 1$ , and the negative exponential when  $p = d = 1$ . Indeed, with  $\Gamma(\cdot)$  denoting the Gamma function, the probability density function for  $x > 0$  is

$$\begin{aligned} f(x|\theta) &= \frac{px^{d-1} \exp \left\{ - \left( \frac{x}{\theta} \right)^p \right\}}{\theta^d \Gamma \left( \frac{d}{p} \right)} \\ &= \exp \left\{ - \frac{x^p}{\theta^p} + \log p - d \log \theta + (d-1) \log x - \log \left( \Gamma \left( \frac{d}{p} \right) \right) \right\} \end{aligned}$$

where, in the terminology of one-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

$$l(\theta_0; \mathbf{x}) = -\frac{1}{\theta_0^p} \sum_{i=1}^n x_i^p + n \log p - nd \log \theta_0 + (d-1) \log \left( \prod_{i=1}^n x_i \right) - n \log \left( \Gamma \left( \frac{d}{p} \right) \right)$$

$$l'(\theta_0; \mathbf{x}) = \frac{p}{\theta_0^{p+1}} \sum_{i=1}^n x_i^p - n \frac{d}{\theta_0} = 0 \Leftrightarrow \hat{\theta}_n(\mathbf{x}) = \left( \frac{p}{nd} \sum_{i=1}^n x_i^p \right)^{\frac{1}{p}}.$$

It is easy to verify that indeed  $l''(\hat{\theta}_n(\mathbf{x}); \mathbf{x}) = -n \frac{pd}{[\hat{\theta}_n(\mathbf{x})]^2} < 0$ , which shows that the MLE exists and is unique. The regularity conditions (R1)-(R4) are also satisfied and using Corollary 3.2 for  $\epsilon = \frac{\theta_0}{2}$  we obtain

$$\begin{aligned} & \left| \mathbb{E} \left[ h \left( \sqrt{n} i(\theta_0) \left( \hat{\theta}_n(\mathbf{X}) - \theta_0 \right) \right) \right] - \mathbb{E} [h(Z)] \right| \leq \frac{\|h'\|}{\sqrt{n}} \left( 2 + \left( 3 + 6 \frac{p}{d} \right)^{\frac{3}{4}} \right) \\ & + \left( 1 - 2 \left( \frac{p}{nd} \right)^{\frac{1}{p}} \frac{\Gamma \left( \frac{nd+1}{p} \right)}{\Gamma \left( \frac{nd}{p} \right)} + \left( \frac{p}{nd} \right)^{\frac{2}{p}} \frac{\Gamma \left( \frac{nd+2}{p} \right)}{\Gamma \left( \frac{nd}{p} \right)} \right) \mathbb{1} \{ \{d \neq 1\} \cup \{p \neq 1\} \} \\ & \times \left[ 8 \|h\| + \frac{\|h'\| \sqrt{ndp} |p-1|}{2} \left( \frac{1}{2^{p-2}} \mathbb{1} \{p < 2\} + \left( \frac{3}{2} \right)^{p-2} \mathbb{1} \{p \geq 2\} \right) \right]. \end{aligned} \quad (3.2)$$

Let us briefly show how to obtain this bound. For the Generalised Gamma distribution,  $D(\theta_0) = q(\theta_0) = \frac{d}{p} \theta_0^p$  and thus  $\mathbb{E} \left[ |T(X) - D(\theta_0)|^3 \right] = \mathbb{E} \left[ \left| X^p - \frac{d}{p} \theta_0^p \right|^3 \right]$ . This third absolute moment is very complicated to calculate. Therefore, we use Hölder's inequality and the fact  $X \sim \text{GG}(\theta_0, d, p) \Rightarrow X^p \sim \text{Gamma} \left( \frac{d}{p}, \frac{1}{\theta_0^p} \right)$  to get

$$\begin{aligned} & \mathbb{E} \left[ \left| X^p - \frac{d}{p} \theta_0^p \right|^3 \right] \leq \left[ \mathbb{E} \left[ \left( X^p - \frac{d}{p} \theta_0^p \right)^4 \right] \right]^{\frac{3}{4}} \\ & = \left[ \mathbb{E} [X^{4p}] + \left( \frac{d\theta_0^p}{p} \right)^4 + 6 \left( \frac{d\theta_0^p}{p} \right)^2 \mathbb{E} [X^{2p}] - 4 \left( \frac{d\theta_0^p}{p} \right)^3 \mathbb{E} [X^p] - 4 \frac{d}{p} \theta_0^p \mathbb{E} [X^{3p}] \right]^{\frac{3}{4}} \\ & = \left[ \theta_0^{4p} \frac{d}{p} \left( 6 + 3 \frac{d}{p} \right) \right]^{\frac{3}{4}} = \theta_0^{3p} \left( \frac{d}{p} \right)^{\frac{3}{4}} \left( 6 + 3 \frac{d}{p} \right)^{\frac{3}{4}}. \end{aligned} \quad (3.3)$$

Simpler calculations yield

$$\frac{|k'(\theta_0)|}{\sqrt{|A''(\theta_0) - k''(\theta_0)D(\theta_0)|}} = \frac{\sqrt{p}}{\sqrt{d\theta_0^p}}. \quad (3.4)$$

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

$$\begin{aligned}
\mathbb{E}\left[\left(\hat{\theta}_n(\mathbf{X}) - \theta_0\right)^2\right] &= \left(\frac{p}{nd}\right)^{\frac{2}{p}} \mathbb{E}\left[\left(\sum_{i=1}^n X_i^p\right)^{\frac{2}{p}}\right] + \theta_0^2 - 2\left(\frac{p}{nd}\right)^{\frac{1}{p}} \theta_0 \mathbb{E}\left[\left(\sum_{i=1}^n X_i^p\right)^{\frac{1}{p}}\right] \\
&= \theta_0^2 \left(\frac{p}{nd}\right)^{\frac{2}{p}} \frac{\Gamma\left(\frac{nd+2}{p}\right)}{\Gamma\left(\frac{nd}{p}\right)} + \theta_0^2 - 2\theta_0^2 \left(\frac{p}{nd}\right)^{\frac{1}{p}} \frac{\Gamma\left(\frac{nd+1}{p}\right)}{\Gamma\left(\frac{nd}{p}\right)} \\
&= \theta_0^2 \left(1 - 2\left(\frac{p}{nd}\right)^{\frac{1}{p}} \frac{\Gamma\left(\frac{nd+1}{p}\right)}{\Gamma\left(\frac{nd}{p}\right)} + \left(\frac{p}{nd}\right)^{\frac{2}{p}} \frac{\Gamma\left(\frac{nd+2}{p}\right)}{\Gamma\left(\frac{nd}{p}\right)}\right). \tag{3.5}
\end{aligned}$$

Regarding  $\sup_{\theta:|\theta-\theta_0|\leq\epsilon} |D''(\theta)|$ , one has to be careful as the supremum depends on the value of  $p$ :

$$\begin{aligned}
\sup_{\theta:|\theta-\theta_0|\leq\epsilon} |d(p-1)\theta^{p-2}| &= d|p-1| \sup_{\theta:|\theta-\theta_0|\leq\epsilon} |\theta^{p-2}| \\
&= d|p-1| \begin{cases} (\theta_0 - \epsilon)^{p-2}, & \text{if } 0 < p < 2 \\ (\theta_0 + \epsilon)^{p-2}, & \text{if } p \geq 2. \end{cases} \tag{3.6}
\end{aligned}$$

Thus, applying now the results of (3.3), (3.4), (3.5) and (3.6) on the general expression for the upper bound in Corollary 3.2, we obtain the result in (3.2).

*Remark 3.4.* (1) The bound in (3.2) is  $\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$ . This is not obvious as we need to comment on the order of 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 \(1951\)](#))

$$\frac{\Gamma(z+a)}{\Gamma(z+b)} = z^{a-b} \left(1 + \frac{(a-b)(a+b-1)}{2z} + \mathcal{O}(|z|^{-2})\right)$$

for large  $z$  (here,  $nd/p$ ) 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 (3.2) comes from the fact that  $q(\theta) = \theta \forall \theta \in \Theta \iff d, p = 1$ .

### 3.3 Bounds for the (negative) exponential distribution

In this subsection, we consider the most famous special case of the Generalised Gamma distribution: the (negative) exponential distribution. First we will treat the canonical form of the distribution and then we will change the parameterisation to discuss the more interesting non-canonical setting.

#### 3.3.1 The canonical case: $\text{Exp}(\theta)$

We start with  $X_1, \dots, X_n$  exponentially distributed i.i.d. random variables with scale parameter  $\theta > 0$  and probability density function  $f(x|\theta) = \theta \exp\{-\theta x\} = \exp\{\log \theta - \theta x\}$  for  $x > 0$ , which

we write  $Exp(\theta)$ . In terms of (3.1), this means  $B = (0, \infty)$ ,  $\Theta = (0, \infty)$ ,  $T(x) = -x$ ,  $k(\theta) = \theta$ ,  $A(\theta) = -\log \theta$  and  $S(x) = 0$ . Further we have that

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

the unique MLE is given by  $\hat{\theta}_n(\mathbf{X}) = 1/\bar{X}$  with  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  and (R1)-(R4) are satisfied.

With this in hand, we can easily see that  $D(\theta_0) = q(\theta_0) = \frac{A'(\theta_0)}{k'(\theta_0)} = -\frac{1}{\theta_0}$  and

$$\frac{|k'(\theta_0)|}{\sqrt{|A''(\theta_0) - k''(\theta_0)D(\theta_0)|}} = \theta_0. \quad (3.7)$$

Simple calculations allow us here to bypass the Hölder inequality used for the Generalized Gamma case and to obtain  $E[|T(X) - D(\theta_0)|^3] = E\left[\left|\frac{1}{\theta_0} - X\right|^3\right] \leq \frac{2.41456}{\theta_0^3}$ . Since  $X_i \sim \text{Exp}(\theta)$ ,  $\forall i \in \{1, 2, \dots, n\}$  then  $\bar{X} \sim \text{Gam}(n, n\theta)$ , with  $\text{Gam}(\alpha, \beta)$  being the Gamma distribution with shape parameter  $\alpha$  and rate parameter  $\beta$ . Consequently,

$$E[(\hat{\theta}_n(\mathbf{X}) - \theta_0)^2] = \frac{(n\theta_0)^2}{(n-1)(n-2)} - \frac{2n\theta_0^2}{n-1} + \theta_0^2 = \frac{(n+2)\theta_0^2}{(n-1)(n-2)}.$$

Moreover, for  $\epsilon > 0$  such that  $0 < \epsilon < \theta_0$ , we obtain  $\sup_{\theta: |\theta - \theta_0| \leq \epsilon} |D''(\theta)| = \frac{2}{(\theta_0 - \epsilon)^3}$ . Choosing  $\epsilon = \frac{\theta_0}{2}$ , we get  $\sup_{\theta: |\theta - \theta_0| \leq \epsilon} |D''(\theta)| = \frac{16}{\theta_0^3}$ . Using this result and (3.7), Corollary 3.2 gives

$$\begin{aligned} \left| E \left[ h \left( \sqrt{n i(\theta_0)} \left( \hat{\theta}_n(\mathbf{X}) - \theta_0 \right) \right) \right] - E[h(Z)] \right| &\leq 4.41456 \frac{\|h'\|}{\sqrt{n}} + 8\|h\| \frac{(n+2)}{(n-1)(n-2)} \\ &\quad + 8\|h'\| \frac{\sqrt{n}(n+2)}{(n-1)(n-2)}. \end{aligned} \quad (3.8)$$

This bound is of order  $\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$  and coincides, as discussed in Remark 3.3, with the AR-bound.

### 3.3.2 The non-canonical case: $\text{Exp}\left(\frac{1}{\theta}\right)$

We now proceed to examine the more interesting case where  $X_1, \dots, X_n$  are i.i.d. random variables from  $Exp\left(\frac{1}{\theta}\right)$ . The probability density function is

$$f(x|\theta) = \frac{1}{\theta} \exp\left\{-\frac{1}{\theta}x\right\} = \exp\left\{-\log\theta - \frac{1}{\theta}x\right\}$$

corresponding to  $B = (0, \infty)$ ,  $\Theta = (0, \infty)$ ,  $T(x) = x$ ,  $k(\theta) = -\frac{1}{\theta}$ ,  $A(\theta) = \log\theta$  and  $S(x) = 0$ . As before, simple steps give that the MLE exists, it is unique and equal to  $\hat{\theta}_n(\mathbf{X}) = \bar{X}$ . The regularity conditions are satisfied and for  $\epsilon = \frac{\theta_0}{2}$  we obtain using Corollary 3.2 that

$$\left| E \left[ h \left( \sqrt{n i(\theta_0)} \left( \hat{\theta}_n(\mathbf{X}) - \theta_0 \right) \right) \right] - E[h(Z)] \right| \leq 4.41456 \frac{\|h'\|}{\sqrt{n}}. \quad (3.9)$$

Indeed,  $D(\theta_0) = q(\theta_0) = \theta_0$ , making the last two terms of the bound in Corollary 3.2 vanish. The result then follows from  $E[|T(X) - D(\theta_0)|^3] = E[|X - \theta_0|^3] \leq 2.41456\theta_0^3$  and

$$\frac{|k'(\theta_0)|}{\sqrt{|A''(\theta_0) - k''(\theta_0)D(\theta_0)|}} = \frac{1}{\theta_0}.$$

*Remark 3.5.* (1) The order of the bound in terms of the sample size is, as expected,  $\frac{1}{\sqrt{n}}$ , corresponding to the order obtained for the Generalized Gamma distribution. The constant here is better than the one inherited from (3.2) for  $p = d = 1$ , thanks to a sharper bound for  $E \left[ |T(X) - D(\theta_0)|^3 \right]$ .

(2) The AR-bound is given by

$$4.41456 \frac{\|h'\|}{\sqrt{n}} + 8 \frac{\|h\|}{n} + 2 \frac{\|h'\|}{\sqrt{n}} + 80 \frac{\|h'\|}{\sqrt{n}} \left( \frac{6}{n} + 3 \right)^{1/2},$$

showing that our new bound is an improvement.

### 3.3.3 Empirical results

For a more complete picture, we also assess the accuracy of our results using simulation-based data. The process we follow is quite simple. We generate 10000 trials of  $n = 10, 100, 1000, 10000$  and 100000 random i.i.d. observations from the exponential distribution  $\text{Exp}(\frac{1}{2})$  (non-canonical case). As function  $h$  we choose  $h(x) = \frac{1}{x^2+2}$  with  $h \in H$ ,  $\|h\| = 0.5$  and  $\|h'\| = \frac{3\sqrt{1.5}}{16}$ . Simple calculations yield  $E[h(Z)] = 0.379$  and each trial gives an MLE  $\hat{\theta}_n(\mathbf{X})$ , hence we have 10000 empirical values of  $h \left( \sqrt{n i(\theta_0)} \left( \hat{\theta}_n(\mathbf{X}) - \theta_0 \right) \right)$  to compare to 0.379. Taking the average hence provides a simulated estimation of  $\left| E \left[ h \left( \sqrt{n i(\theta_0)} \left( \hat{\theta}_n(\mathbf{X}) - \theta_0 \right) \right) \right] - E[h(Z)] \right|$ , which we compare to the upper bound given in (3.9). Our bound provides a very strong improvement on the AR-bound (see Table 3.1). Of course, this estimated distance is only a lower bound to the true distance, as we have chosen a particular function  $h$  instead of the supremum over all functions  $h \in H$ , but its calculation still provides an idea of the accuracy of our bounds. This closeness logically increases with the sample size and becomes quite sharp for  $n \geq 100$ .

Table 3.1: Simulation results for the  $\text{Exp}(\frac{1}{2})$  distribution treated as a non-canonical exponential family

$n$	$\left  \hat{E} \left[ h \left( \sqrt{n i(\theta_0)} \left( \hat{\theta}_n(\mathbf{X}) - \theta_0 \right) \right) \right] - E[h(Z)] \right $	New Bound	AR-bound
10	0.0034	0.321	11.888
100	0.0022	0.101	3.401
1000	0.0012	0.032	1.058
10000	0.0008	0.010	0.333
100000	0.0004	0.003	0.105

### ACKNOWLEDGMENTS

The authors would like to thank Gesine Reinert for various insightful comments and suggestions.

### References

- A. Anastasiou and G. Reinert. Bounds for the normal approximation of the maximum likelihood estimator. *Bernoulli*, forthcoming (2015).
- G. Casella and R. L. Berger. *Statistical Inference*. Brooks/Cole, Cengage Learning, second edition (2002).

- D. R. Cox and E. J. Snell. A general definition of residuals. *Journal of the Royal Statistical Society Series B* **30**, 248–275 (1968).
- R. A. Fisher. Theory of Statistical Estimation. *Mathematical Proceedings of the Cambridge Philosophical Society* **22**, 700–725 (1925).
- C. J. Geyer. Asymptotics of maximum likelihood without the LLN or CLT or sample size going to infinity. In *Advances in Modern Statistical Theory and Applications: A Festschrift in honor of Morris L. Eaton*, pages 1–24. Eds.: G. Jones and X. Shen, Beachwood, OH: Institute of Mathematical Statistics (2013).
- 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. *Annals of Statistics* **9**, 758–767 (1981).
- I. Nourdin and G. Peccati. *Normal Approximations with Malliavin Calculus*. Cambridge Tracts in Mathematics, No.192. Cambridge University Press (2012).
- G. Reinert. Couplings for normal approximations with Stein’s method. In *Microsurveys in Discrete Probability*, pages 193–207. Eds.: D. Aldous, J. Propp, DIMACS series. AMS (1998).
- F. Tricomi and A. Erdélyi. The asymptotic expansion of a ratio of gamma functions. *Pacific Journal of Mathematics* **1**, 133–142 (1951).