IMS Lecture Notes–Monograph Series
Optimality: The Third Erich L. Lehmann Symposium
Vol. 57 (2009) 276–303

DOI: 10.1214/09-LNMS5717

Large Sample Statistical Inference for Skew-Symmetric Families on the Real Line

Rolando Cavazos–Cadena[1,*] **and Graciela González–Farías**[2,*]

Universidad Autónoma Agraria Antonio Narro and
Centro de Investigación en Matemáticas A. C.

Abstract: For a general family of one-dimensional skew-symmetric probability densities, the application of the maximum likelihood method to the estimation of the asymmetry parameter λ is studied. Under mild conditions, the existence and consistency of a sequence $\{\hat{\lambda}_n\}$ of maximum likelihood estimators is established, and the limit distributions of $\{\hat{\lambda}_n\}$ and the sequence of likelihood ratios are determined under the null hypothesis $\mathcal{H}_0 : \lambda = 0$. These latter conclusions, which hold under differential singularity of the likelihood function at $\lambda = 0$, extend to the present framework results recently obtained for general statistical models with null Fisher information.

Contents

1. Introduction

This work concerns likelihood inference for the general one-dimensional skew-symmetric family, which is constructed as follows: Given two symmetric densities f and g on the real line—that is, $f(x) = f(-x)$ and $g(x) = g(-x)$ for all $x \in \mathbb{R}$—let

$$G(x) := \int_{-\infty}^{x} g(z)\, dz \tag{1.1}$$

[1]Departamento de Estadística y Cálculo, Universidad Autónoma Agraria Antonio Narro, Buenavista, Saltillo COAH, 25315, México, e-mail: rcavazos@uaaan.mx
[2]Centro de Investigación en Matemáticas A. C., Apartado Postal 402, Guanajuato, GTO, 36240, México, e-mail: farias@cimat.mx
*This research was supported by the PSF Organization under Grant 2007-3, and by CONACYT under Grants 25357 and 45974-F.
AMS 2000 subject classifications: Primary 62F10; secondary 62F12.
Keywords and phrases: boundedness of maximum likelihood estimators, Kullback's inequality, lateral Taylor series, strong law of large numbers, central limit theorem, asymptotic normality.

be the cumulative distribution function of density g. Notice that $\int_{\mathbb{R}}[1 - G(\lambda w)] \times f(w)\,dw = \int_{\mathbb{R}}[1 - G(-\lambda w)]f(-w)\,dw = \int_{\mathbb{R}} G(\lambda w)f(w)\,dw$ for every $\lambda \in \mathbb{R}$, so that $\int_{\mathbb{R}} 2f(w)G(\lambda w)\,dw = 1$. This argument, due to Azzalini [1], shows that for each $\lambda \in \mathbb{R}$

$$\rho(x; \lambda) := 2f(x)G(\lambda x), \quad x \in \mathbb{R} \tag{1.2}$$

is a genuine density, and the collection

$$S(f, g) := \{\rho(\cdot; \lambda) \mid \lambda \in \mathbb{R}\}, \tag{1.3}$$

is the skew-symmetric family determined by f and g. When the parametrization $\lambda \mapsto \rho(\cdot; \lambda)$ is one-to-one, $f(\cdot) = \rho(\cdot; 0)$ is the unique symmetric density in $S(f, g)$, and λ can be considered as a measure of the asymmetry (or skewness) of density $\rho(\cdot; \lambda)$. The first systematic treatment of a skew-symmetric family was presented in Azzalini [1, 2] for the case in which f and g coincide with the standard normal density φ, and location-scale and regression models based on $S(\varphi, \varphi)$ as well as multivariate extensions have been intensively studied during the last twenty years; see Azzalini and Dalla Valle [4], Azzalini and Capitanio [3], Pewsey [13], Genton [10] and the references therein. The analysis of the maximum likelihood method for the location-scale model based on $S(\varphi, \varphi)$ has proved most challenging since, in that context, the Fisher information matrix has incomplete rank at $\lambda = 0$, a problem that also arises for the skew exponential family, which includes the skew normal location-scale model as a particular case (Azzalini [2], DiCiccio and Monti [9]). The singularity of the information matrix has been analyzed via the centered parametrization introduced in Azzalini [1] and asymptotic results are based on the recent work by Rotzintzky *et al.* [14]. Using the conclusions in this latter paper, rates of convergence for maximum likelihood estimators are derived in Chiogna [8], and Sartori [15] studied the finiteness of the estimator of the asymmetry parameter.

As suggested by the previous comments, family $S(\varphi, \varphi)$ has been intensively studied, and a great effort has been done on generalizations of that model (Genton [10]), so that looking for inference results applicable to a broad class of skew families is, certainly, an interesting problem. This work is *a first step* in this direction since, although no scale or location parameters will be introduced, the maximum likelihood method applied to the estimation of the asymmetry parameter λ will be studied under rather *minimal conditions* on densities f and g, making the likelihood inference problem a very interesting one. *The first objective* of this note is

(i) To establish the existence of a sequence $\{\hat{\lambda}_n\}$ of maximum likelihood estimators of λ and to prove its consistency.

To see the interest behind this problem, denote the (kernel log-) likelihood corresponding to a single observation x by

$$\ell(\lambda; x) := \log(G(\lambda x)), \tag{1.4}$$

and observe that if g is the standard normal density, then $\ell(\cdot; x)$ is strictly concave for $x \neq 0$, a property that yields the existence and consistency of maximum likelihood estimators (Newey and McFadden [12]). However, strict concavity of $\ell(\cdot; x)$ is far from being a general property, and may fail in common cases, for instance, if g is the Laplace density. In this work, problem (i) above will be studied under the minimal assumption that the parametrization $\lambda \mapsto \rho(\cdot; \lambda)$ is identifiable and, since the parameter space is not compact for the $S(f, g)$ family, to the best of the authors' knowledge, in this context the existence and consistency of maximum likelihood estimators can not be directly obtained form general available results. The second problem studied in this work concerns the asymptotic distribution of $\{\hat{\lambda}_n\}$ and, as usual, the analysis below requires differentiability assumptions on $\ell(\cdot; x)$, and then, on density g. This problem will be studied under conditions allowing g to be non smooth at $x = 0$, which can be roughly described as follows:

A1: g is continuous on $\mathbb{R}$, is 'smooth' outside 0, and has lateral derivatives at zero.

Under this requirement, it is not difficult to see that if the true parameter value, say ν, is non-null, then $\ell(\cdot; x)$ is smooth on a neighborhood of ν if $\ell(\nu; x) < \infty$. Thus, when the information number at ν, given by $\mathcal{I}(\nu) = \int_{\mathbb{R}} (\partial_\lambda \ell(\nu; x))^2 \rho(\nu; x)\, dx (> 0)$, is finite, under standard regularity conditions Wald's classical results yield that $\sqrt{n\mathcal{I}(\nu)}(\hat{\lambda}_n - \nu)$ has a standard normal distribution at the limit (Lehmann and Casella [11], Section 6.3, Shao [16], Section 4.4). However, under the null hypothesis $\mathcal{H}_0 : \lambda = 0$ such a direct conclusion is not possible, since $g(\cdot)$ is not necessarily differentiable in a neighborhood of zero under condition A1 above. Also, observe that $\partial_\lambda \ell(0; x) = 2g(0)x$, so that $\mathcal{I}(0) = 4g(0)^2 \int_R x^2 f(x)\, dx$; thus, $\mathcal{I}(0)$ is null if $g(0) = 0$ and, again, in this case the asymptotic distribution of $\{\hat{\lambda}_n\}$ can not be obtained from the results in the aforementioned references. The case of null information at $\lambda = 0$ was recently studied in a general context by Rotnitzky *et al.* [14] under several assumptions, including (a) compactness of the parameter space, and (b) the existence of higher order derivatives $\partial_\lambda^k \ell(\lambda; x)$ in a neighborhood of zero. Since in the present context neither the parameter space is compact, nor higher order derivatives $\partial_\lambda^k \ell(\lambda; x)$ exist around zero, if $g(0) = 0$ then the limiting distribution can not be solved by direct application of available results under condition A1. Therefore, *the second problem* considered in this note is

(ii) to determine both the limit distribution of the (appropriately normalized) sequence $\{\hat{\lambda}_n\}$ under the hypothesis $\mathcal{H}_0 : \lambda = 0$, and the asymptotic null distribution of the likelihood ratio statistic.

The approach used below to study problems (i) and (ii) can be described as follows: The monotonicity of the mapping $\lambda \mapsto \ell(\lambda; x)$ is used to establish the existence of maximizers $\hat{\lambda}_n$ of the observed likelihood when the sample size n is large enough, whereas the proof for the consistency of $\{\hat{\lambda}_n\}$ follows the ideas in Cavazos-Cadena and Gonzalez-Farías [7] where, under mild conditions, it is shown that sequence $\{\hat{\lambda}_n\}$ is consistent if and only if it is bounded with probability 1; here, after establishing the boundedness property, a direct proof of consistency is given via a simple consequence of Kullback's inequality. Concerning problem (ii), as in the results presented in Lehmann and Casella ([11], Section 6.3), Shao ([16], Section 4.4), or in Rotnitzky *et al.* [14], the analysis is based on Taylor series expansions for the observed likelihood and its derivative around zero, which in the present context are *lateral expansions*; they are used to show that the maximum likelihood estimator is no null with probability increasing to 1, and the asymptotic distributions are obtained via the central limit theorem and the strong law of large numbers.

The organization of the paper is as follows: Problem (i) is analyzed in the following three sections. Thus, in Section 2 the identification property of the parametrization $\lambda \mapsto \rho(\lambda; \cdot)$ is discussed, the existence of a sequence $\{\hat{\lambda}_n\}$ of maximum likelihood estimators is established in Section 3, and the consistency of $\{\hat{\lambda}_n\}$ is proved in Section 4. After this point, the reminder of the paper concerns problem (ii). In Section 5 the basic dominance and smoothness assumptions are formally introduced, and the main asymptotic result is stated as Theorem 5.1. Next, in Section 6 the necessary technical tools concerning *lateral Taylor series* for the likelihood function and its first derivative are established and, finally, the exposition concludes in Section 7 with a proof of Theorem 5.1.

Notation. Throughout the remainder $X_1, X_2, X_3, \ldots$ stands for a sequence of independent and identically distributed random variables with common density belonging to family $S(f,g)$ in (1.3): For each $n = 1, 2, \ldots$, set

$$X_1^n := (X_1, \ldots, X_n). \tag{1.5}$$

The distribution of the sequence $(X_1, X_2, \ldots)$ when ν is the true parameter value is denoted by $P_\nu[\,\cdot\,]$, whereas $E_\nu[\,\cdot\,]$ stands for the corresponding expectation operator. On the other hand, if $H(\cdot)$ is a function defined around zero, $H(0+) := \lim_{x\searrow 0} H(x)$, and $H(0-) := \lim_{x\nearrow 0} H(x)$, whereas given $A \subset \mathbb{R}$, I_A is the indicator function of set A, that is, $I_A(x) = 1$ if $x \in A$, whereas $I_A(x) = 0$ when $x \notin A$. Finally, the following convention is enforced: $\sum_{i=a}^{b} C_i = 0$ when $b < a$.

2. Identifiability

In this section the identifiability of the parametrization $\lambda \mapsto \rho(\cdot;\lambda)$ is briefly discussed. This condition establishes that different parameters correspond to different densities, and plays a fundamental role in parametric estimation (Newey and McFadden [12]).

Assumption 2.1. *The mapping $\lambda \mapsto \rho(\cdot;\lambda)$ is one-to-one, that is,*

$$\int_{\mathbb{R}} |\rho(x;\lambda) - \rho(x;\nu)|\,dx = \int_{\mathbb{R}} 2f(x)|G(\lambda x) - G(\nu x)|\,dx \neq 0, \quad \text{if } \lambda \neq \nu;$$

see (1.2).

Some primitive conditions ensuring this requirement are now given.

Lemma 2.1. *Assumption 2.1 holds under either of the following conditions (i)–(iii).*

(i) For every nonempty open interval $J \subset \mathbb{R}$, $\int_J f(x)\,dx > 0$;
(ii) $\int_J g(x)\,dx > 0$ for each nonempty open interval $J \subset \mathbb{R}$;
(iii) There exists $\delta > 0$ such that

$$\int_J g(x)\,dx > 0 \quad \text{and} \quad \int_J f(x)\,dx > 0,$$

for each nonempty open interval $J \subset (0,\delta)$.

Proof. Let λ and ν be fixed and *different* real numbers, and suppose that

$$\int_{\mathbb{R}} f(x)|G(\lambda x) - G(\nu x)|\,dx = 0. \tag{2.1}$$

A contradiction will be obtained under each of the three conditions in the lemma.

Assume that condition (i) holds. Given an open interval $J \subset \mathbb{R}$ with positive length, (2.1) and $\int_J f(x)\,dx > 0$ together yield that $|G(\lambda x) - G(\nu x)| = 0$ for some $x \in J$, so that $\{x \,|\, G(\lambda x) - G(\nu x) = 0\}$ is dense in $\mathbb{R}$. Since $G(\cdot)$ is continuous, this set is also closed, so that

$$G(\lambda x) = G(\nu x) \quad \text{for all } x \in \mathbb{R}. \tag{2.2}$$

This fact implies that $\lambda \neq 0$ and $\nu \neq 0$. Indeed, if $\nu = 0$, it follows that $\lambda \neq 0$ and the right-hand side equals $1/2$ for all x, whereas the values of left-hand side cover the whole interval $(0,1)$ as x moves on $\mathbb{R}$. Thus, $\nu \neq 0$ and, similarly, $\lambda \neq 0$. Moreover, (2.2) also yields that λ and ν have the same sign since, otherwise, as $x \to \infty$ one side of the equality converges to 1 and the other converges to 0. Therefore, recalling that $\lambda \neq \nu$, it follows that $|\lambda| \neq |\nu|$, and without loss of generality it can be assumed that $\beta = |\lambda|/|\nu| = \lambda/\nu \in (0,1)$. Replacing x by y/ν, (2.2) yields that, for all $y \in \mathbb{R}$, $G(y) = G(\beta y)$, and then

$$G(y) = G(\beta^n y), \quad y \in \mathbb{R}, \quad n = 1, 2, 3, \ldots.$$

Letting n go to ∞, the continuity of $G(\cdot)$ and the inclusion $\beta \in (0,1)$ yield that $G(y) = G(0) = 1/2$ for all $y \in \mathbb{R}$, in contradiction with the basic properties of a distribution function.

Under condition (ii), $G(b) - G(a) = \int_a^b g(x)\,dx > 0$ for $a < b$, so that $G(\cdot)$ is strictly increasing. Thus, since $\lambda \neq \nu$, $|G(\lambda x) - G(\nu x)| > 0$ for all $x \neq 0$, and it follows that (2.1) is equivalent to $\int_{\mathbb{R}} f(x)\,dx = 0$, which is not possible, since $f(\cdot)$ is a density.

Suppose that condition (iii) occurs. In this context, $G(\cdot)$ is strictly increasing on the interval $(0,\delta)$ and then on $(-\delta,\delta)$, by symmetry. Next, define $\delta_1 := \delta/(|\lambda| + |\nu| + 1)$ and, recalling that $\lambda \neq \nu$, notice that if $x \in [\delta_1/2, \delta_1]$ then λx and νx are different points in $(-\delta,\delta)$, so that $|G(\lambda x) - G(\nu x)| > 0$. Thus, by continuity of $G(\cdot)$, $\min_{x\in[\delta_1/2,\delta_1]} |G(\lambda x) - G(\nu x)| =: \varepsilon > 0$. Consequently,

$$\begin{aligned} &\int_{[\delta_1/2,\delta_1]} f(x)|G(\lambda x) - G(\nu x)|\,dx \\ &\qquad \geq \varepsilon \int_{[\delta_1/2,\delta_1]} f(x)\,dx > 0, \end{aligned}$$

where the inclusion $[\delta_1/2, \delta_1] \subset (0,\delta)$ was used to set the last inequality. Therefore, (2.1) can not occur under condition (iii). □

According to the previous result, Assumption 2.1 is valid under mild requirements on densities f and g, and it is interesting to observe that if conditions (i)–(iii) in Lemma 2.1 do not hold, then identifiability may fail.

Example 2.1. Let the symmetric densities f and g be such that $f(x) = 0$ for $x \in [-1,1]$, whereas $g(x) = 0$ for $|x| > 1$. In this case, it is not difficult to see that the general density $\rho(\cdot;\lambda) \in S(f,g)$ satisfies $\rho(\cdot;\lambda) = \rho(\cdot;1)$ for $\lambda \geq 1$, and $\rho(\cdot;\lambda) = \rho(\cdot;-1)$ when $\lambda \leq -1$, so that Assumption 2.1 does not hold.

The basic consequence of Assumption 2.1, which plays a central role in the subsequent development, is established in the following lemma. Firstly, recall that $G(\cdot)$ is increasing and notice that if $\lambda_2 > \lambda_1$ then

$$\int_0^\infty f(x)|G(\lambda_2 x) - G(\lambda_1 x)|\,dx = \int_0^\infty f(x)(G(\lambda_2 x) - G(\lambda_1 x))\,dx,$$

and

$$\int_{-\infty}^{0} f(x)|G(\lambda_2 x) - G(\lambda_1 x)|\,dx = \int_{-\infty}^{0} f(x)(G(\lambda_1 x) - G(\lambda_2 x))\,dx$$
$$= \int_{0}^{\infty} f(x)(G(\lambda_2 x) - G(\lambda_1 x))\,dx,$$

where the second equality comes from $\int_{\mathbb{R}} f(x)G(\lambda_i x)\,dx = 1/2$ for $i = 1, 2$, so that

$$\text{(2.3)} \qquad \int_R f(x)|G(\lambda_2 x) - G(\lambda_1 x)|\,dx$$
$$= 2\int_0^{\infty} f(x)(G(\lambda_2 x) - G(\lambda_1 x))\,dx$$
$$= 2\int_{-\infty}^{0} f(x)(G(\lambda_1 x) - G(\lambda_2 x))\,dx, \quad \lambda_1 < \lambda_2.$$

Lemma 2.2. *Under the identifiability Assumption 2.1, the following assertions (i) and (ii) hold:*

(i) For each $\lambda \in \mathbb{R}$, $\int_0^\infty f(x)G(\lambda x)\,dx > 0$ and $\int_{-\infty}^0 f(x)G(\lambda x)\,dx > 0$.
(ii) There exists a function $c : \mathbb{R} \to (0, \infty)$ such that

$$P_\lambda[X_1 < -c(\lambda)] > 0 \quad \textit{and} \quad P_\lambda[X_1 > c(\lambda)] > 0, \quad \lambda \in \mathbb{R}.$$

Proof. (i) Notice that

$$0 \le \int_0^\infty f(x)G(\lambda_1 x)\,dx \le \int_0^\infty f(x)G(\lambda_2 x)\,dx, \quad \lambda_1 < \lambda_2,$$

since $G(\cdot)$ is increasing. Now, suppose that $\int_0^\infty f(x)G(\lambda_2 x)\,dx = 0$ for some $\lambda_2 \in \mathbb{R}$. In this case the above display yields $\int_0^\infty f(x)G(\lambda_1 x)\,dx = 0$ for every $\lambda_1 \le \lambda_2$, and then,

$$\int_R f(x)|G(\lambda_2 x) - G(\lambda_1 x)|\,dx = 0, \quad \lambda_1 \le \lambda_2,$$

by the first equality in (2.3), contradicting Assumption 2.1; consequently, $\int_0^\infty f(x) \times G(\lambda x)\,dx > 0$ for all $\lambda \in \mathbb{R}$, whereas the other part of the conclusion can be obtained along similar lines.

(ii) By the monotone convergence theorem, as $\varepsilon \searrow 0$, $\int_\varepsilon^\infty f(x)G(\lambda x)\,dx \nearrow \int_0^\infty f(x)G(\lambda x)\,dx$ and $\int_{-\infty}^{-\varepsilon} f(x)G(\lambda x)\,dx \nearrow \int_{-\infty}^0 f(x)G(\lambda x)\,dx$ for each λ. Therefore, by part (i), there exists $c(\lambda) > 0$ such that $\int_{c(\lambda)}^\infty f(x)G(\lambda x)\,dx > 0$ and $\int_{-\infty}^{-c(\lambda)} f(x)G(\lambda x)\,dx > 0$, and the conclusion follows. □

3. Existence of Maximum Likelihood Estimators

The objective of this section is to establish the existence of a sequence of maximum likelihood estimators, an idea that is formally stated below. To begin with, given a fixed sample size $n > 0$, for each possible sample $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ define the average (kernel log-)likelihood function $L_n(\cdot; \mathbf{x})$ by

$$\text{(3.1)} \qquad L_n(\lambda; \mathbf{x}) := \frac{1}{n}\sum_{k=1}^{n} \ell(\lambda; x_k) = \frac{1}{n}\sum_{k=1}^{n} \log(G(\lambda x_k)), \quad \lambda \in \mathbb{R},$$

(see (1.1) and (1.4)), where the usual convention $\log(0) := -\infty$ is enforced. Since $G(\cdot)$ is continuous and takes values in $[0,1]$, $L_n(\cdot;\cdot)$ is continuous function from $\mathbb{R}\times\mathbb{R}^n$ into $[-\infty,0]$; moreover, for every $Q\subset\{1,2,\ldots,n\}$

$$L_n(\lambda;\mathbf{x}) \le \frac{1}{n}\sum_{i\,:i\in Q}\log(G(\lambda x_i)) \le 0, \quad \lambda\in\mathbb{R}, \quad \emptyset\ne Q. \tag{3.2}$$

Definition 3.1. Let $\{\lambda_n : \mathbb{R}^n \to \mathbb{R}\}$ be a sequence of (Borel) measurable functions and set

$$\hat\lambda_n := \lambda_n(X_1^n).$$

In this case, $\{\hat\lambda_n\}$ is *a sequence of maximum likelihood of estimators of* λ if

$$P_{\lambda_0}\left[\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}[L_k(\hat\lambda_k;X_1^k)\ge L_k(\lambda;X_1^k) \text{ for all } \lambda]\right] = 1, \quad \lambda_0\in\mathbb{R}. \tag{3.3}$$

Remark 3.1. (i) By continuity, $L_n(\hat\lambda_n;X_1^n)\ge L_n(\lambda;X_1^n)$ occurs for every $\lambda\in\mathbb{R}$ if and only if it holds for each rational number, so that $[L_n(\hat\lambda_n;X_1^n)\ge L_n(\lambda;X_1^n),\ \lambda\in\mathbb{R}]$ is an event.

(ii) In words, $\{\hat\lambda_n\}$ is a sequence of maximum likelihood estimators of λ if, with probability 1 and regardless of the true parameter value, $\hat\lambda_n$ maximizes the observed average likelihood function $L_n(\cdot;X_1^n)$ whenever n is large enough. The event within brackets in (3.3) is the inferior limit of the events $[L_k(\hat\lambda_k;X_1^k)\ge L_k(\lambda;X_1^k),\ \lambda\in\mathbb{R}]$, and when (3.3) holds then, as $k\to\infty$, $P_{\lambda_0}[L_k(\hat\lambda_k;X_1^k)\ge L_k(\lambda;X_1^k) \text{ for all } \lambda]\to 1$; see, for instance, Billingsley ([6], Section 4).

The main objective of this section is to prove the following result.

Theorem 3.1. *Under Assumption 2.1, there exists a sequence of maximum likelihood estimators of* λ*.*

The proof of this theorem has been divided into three simple lemmas involving the following notation: For each $\mathbf{x}\in\mathbb{R}^n$, set

$$m_n(\mathbf{x}) := \sup_{\lambda\in\mathbb{R}} L_n(\lambda;\mathbf{x}), \tag{3.4}$$

so that (3.1) and (3.2) lead to

$$0\ge m_n(\mathbf{x})\ge L_n(0;\mathbf{x}) = -\log(2), \quad \mathbf{x}\in\mathbb{R}^n, \tag{3.5}$$

since $G(0)=1/2$. Next, define

$$\mathcal{M}_n(\mathbf{x}) := \{\nu\in\mathbb{R}\,|\,L_n(\nu;\mathbf{x}) = m_n(\mathbf{x})\}, \quad \mathbf{x}\in\mathbb{R}^n, \tag{3.6}$$

which is a closed subset of $\mathbb{R}$, by the continuity of $L_n(\cdot;\mathbf{x})$. As can be seen from the monotonicity of $\ell(\cdot;x)$, the set $\mathcal{M}_n(\mathbf{x})$ may be empty if the observed sample $\mathbf{x}\in\mathbb{R}^n$ does not contain observations of different sign. The first step to the proof of Theorem 3.1 is the following lemma, showing that $\mathcal{M}_n(\mathbf{x})$ is nonempty and compact if $\mathbf{x}\in\mathbb{R}^n$ contains components with opposite signs. For each integer $n\ge 2$, set

$$\mathcal{S}_n := \{\mathbf{x}\in\mathbb{R}^n\,|\,x_ix_j<0 \text{ for some } i \text{ and } j \text{ with } 1\le i\ne j\le n\} \tag{3.7}$$

and observe that $\mathcal{S}_n$ is an open subset of $\mathbb{R}^n$.

Lemma 3.1. *For each integer $n \geq 2$ and $\mathbf{x} \in \mathcal{S}_n$, the set of maximizers $\mathcal{M}_n(\mathbf{x})$ is nonempty and compact.*

Proof. Given $\mathbf{x} \in \mathcal{S}_n$, select indexes i^* and j^* such that $x_{i^*} < 0$ and $x_{j^*} > 0$, so that $\lim_{\lambda\to\infty} \log(G(\lambda x_{i^*})) = -\infty = \lim_{\lambda\to-\infty} \log(G(\lambda x_{j^*}))$. After setting $Q = \{i^*\}$ and $Q = \{j^*\}$ in (3.2), these convergences yield that

$$\lim_{|\lambda|\to\infty} L_n(\lambda; \mathbf{x}) = -\infty,$$

and it follows that the set $\mathcal{M}_n(\mathbf{x})$—consisting of the maximizers of the continuous function $L_n(\cdot; \mathbf{x})$—is nonempty and compact. □

Next, for each $n \geq 2$, define $\lambda_n^+ : \mathcal{S}_n \to \mathbb{R}$ by

$$\lambda_n^+(\mathbf{x}) := \max \mathcal{M}_n(\mathbf{x}), \quad \mathbf{x} \in \mathcal{S}_n, \tag{3.8}$$

so that $\lambda_n^+(\mathbf{x})$ is the largest element in $\mathcal{M}_n(\mathbf{x})$. Observe that $\lambda_n^+(\mathbf{x}) \in \mathcal{M}_n(\mathbf{x})$ is a well-defined finite number for each $\mathbf{x} \in \mathcal{S}_n$, by Lemma 3.1. As it is shown below, this function $\lambda_n^+(\cdot)$ is upper semi-continuous.

Lemma 3.2. *Let the integer $n \geq 2$ be arbitrary but fixed, and suppose that $\{\mathbf{x}_k\} \subset \mathcal{S}_n$ is such that $\lim_{k\to\infty} \mathbf{x}_k = \mathbf{y} \in \mathcal{S}_n$. In this context,*

(i) If $\nu_k \in \mathcal{M}_n(\mathbf{x}_k)$ for each k, then sequence $\{\nu_k\}$ is bounded, and
(ii) Every limit point of $\{\nu_k\}$ belongs to $\mathcal{M}_n(\mathbf{y})$.
Consequently,
(iii) $\lambda_n^+(\cdot)$ is upper semi-continuous.

Proof. To begin with, write $\mathbf{x}_k = (x_{k1}, \ldots, x_{kn})$ and notice that, since $\nu_k \in \mathcal{M}_n(\mathbf{x}_k)$,

$$\frac{1}{n}\sum_{i=1}^{n} \log(G(\lambda x_{ki})) = L_n(\lambda; \mathbf{x}_k) \leq L_n(\nu_k; \mathbf{x}_k) = \frac{1}{n}\sum_{i=1}^{n} \log(G(\nu_k x_{ki})), \tag{3.9}$$

for ever $\lambda \in \mathbb{R}$; in particular,

$$-\log(2) = L_n(0; \mathbf{x}_k) \leq \frac{1}{n}\sum_{i=1}^{n} \log(G(\nu_k x_{ki})). \tag{3.10}$$

(i) *Assume* now that $\limsup_{k\to\infty} \nu_k = \infty$ and, taking a subsequence if necessary, without loss of generality suppose that $\nu_k \to \infty$. In this context, select an index i^* such that $y_{i^*} < 0$, which is possible since $\mathbf{y} \in \mathcal{S}_n$ (see (3.7)) and observe that the convergence $x_{ki^*} \to y_{i^*} < 0$ leads to $\nu_k x_{ki^*} \to -\infty$ as $k \to \infty$, so that $\log(G(\nu_k x_{ki^*})) \to -\infty$, and then, via (3.2) with $Q = \{i^*\}$, this yields

$$\lim_{k\to\infty} \frac{1}{n}\sum_{i=1}^{n} \log(G(\nu_k x_{ki})) = -\infty,$$

which contradicts (3.10); it follows that $\limsup_k \nu_k < \infty$, whereas the inequality $\liminf_k \nu_k > -\infty$ can be established along similar lines.

(ii) Let ν^* be an arbitrary limit point of $\{\nu_k\}$, and notice that, by part (i), ν^* is finite; selecting a subsequence, if necessary, assume that $\nu_k \to \nu^*$. In this case, taking the limit as k goes to ∞ in (3.9), it follows that

$$L_n(\lambda; \mathbf{y}) = \frac{1}{n}\sum_{i=1}^{n} \log(G(\lambda y_k)) \leq \frac{1}{n}\sum_{i=1}^{n} \log(G(\nu^* y_k)) = L_n(\nu^*; \mathbf{y}), \quad \lambda \in \mathbb{R},$$

i.e., $\nu^* \in \mathcal{M}_n(\mathbf{y})$.

(iii) Let $\mathbf{y} \in \mathcal{S}_n$ be arbitrary. If $\{\mathbf{x}_k\} \subset \mathcal{S}_n$ is such that $\lim_k \mathbf{x}_k = \mathbf{y}$, recalling that $\lambda_n^+(\mathbf{x}_k) \in \mathcal{M}_n(\mathbf{x}_k)$, part (ii) with $\nu_k = \lambda_n^+(\mathbf{x}_k)$ yields that $\limsup_k \lambda_n^+(\mathbf{x}_k) \in \mathcal{M}_n(\mathbf{y})$, and then $\limsup_k \lambda_n^+(\mathbf{x}_k) \leq \lambda_n^+(\mathbf{y})$, by (3.8), so that $\lambda_n^+(\cdot)$ is upper semicontinuous. □

The last step before the proof of Theorem 3.1 is the following consequence of Lemma 2.2(i).

Lemma 3.3. *Under Assumption 2.1,*

$$\lim_{n\to\infty} P_\nu\left[\bigcap_{k=n}^{\infty}[X_1^k \in \mathcal{S}_k]\right] = 1, \quad \nu \in \mathbb{R};$$

see (1.5) for notation.

Proof. Let $\nu \in \mathbb{R}$ be fixed and observe that for every $i = 1, 2, \ldots,$

$$P_\nu[X_i \leq 0] = 1 - P_\nu[X_i > 0] = 1 - 2\int_0^\infty f(x)G(\nu x)\,dx =: \rho_-(\nu) \in [0,1),$$

where the inclusion stems from Lemma 2.2(i), so that for each $k > 0$, $\rho_-(\nu)^k = P_\nu[X_i \leq 0,\ 1 \leq i \leq k]$; similarly, $P_\nu[X_i \geq 0,\ 1 \leq i \leq k] = \rho_+(\nu)^k$ for some $\rho_+(\nu) \in [0,1)$. Since

$$[X_1^k \in \mathcal{S}_k]^c = [X_i \leq 0,\ 1 \leq i \leq k] \cup [X_i \geq 0,\ 1 \leq i \leq k],$$

it follows that $P_\nu[[X_1^k \in \mathcal{S}_k]^c] \leq 2\rho(\nu)^k$, where $\rho(\nu) \in [0,1)$ is given by $\rho(\nu) := \max\{\rho_-(\nu), \rho_+(\nu)\}$. Observing that (1.5) and (3.7) together yield that $[X_1^n \in \mathcal{S}_n] \subset [X_1^k \in \mathcal{S}_k]$ for $k \geq n$, it follows that

$$P_\nu\left[\bigcap_{k=n}^{\infty}[X_1^k \in \mathcal{S}_k]\right] \geq P_\nu\left[X_1^n \in \mathcal{S}_n\right] \geq 1 - 2\rho(\nu)^n$$

and the conclusion is obtained taking the limit as $n \to \infty$. □

Notice that the last display and Lemma 3.1 together shows that, with probability increasing to 1 at a geometric rate, the function $L_n(\cdot; X_1^n)$ has a maximizer when n is large enough.

Proof of Theorem 3.1. Select a point $\lambda^* \in \mathbb{R}$ and for each positive integer k define $\lambda_k : \mathbb{R}^k \to \mathbb{R}$ as follows: $\lambda_1(\cdot) \equiv \lambda^*$, whereas, for $k \geq 2$, $\lambda_k(\mathbf{x}) := \lambda^*$ if $\mathbf{x} \in \mathbb{R}^k \setminus \mathcal{S}_k$, and $\lambda_k(\mathbf{x}) := \lambda_k^+(\mathbf{x})$ if $\mathbf{x} \in \mathcal{S}_k$. Since the $\mathcal{S}_k$'s are open sets, Lemma 3.2(iii) implies that each function $\lambda_k(\cdot)$ is measurable, and then $\hat{\lambda}_k = \lambda_k(X_1^k)$ is a genuine statistic. Using (3.4)–(3.6) and (3.8), this specification yields $[X_1^k \in \mathcal{S}_k] \subset [L_k(\hat{\lambda}_k; X_1^k) \geq L_k(\lambda; X_1^k), \lambda \in \mathbb{R}]$ for $k \geq 2$. Therefore, for each integer $m \geq 2$,

$$\begin{aligned}\bigcap_{k=m}^{\infty}[X_1^k \in \mathcal{S}_k] &\subset \bigcap_{k=m}^{\infty}[L_k(\hat{\lambda}_k; X_1^k) \geq L_k(\lambda; X_1^k), \lambda \in \mathbb{R}] \\ &\subset \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}[L_k(\hat{\lambda}_k; X_1^k) \geq L_k(\lambda; X_1^k), \lambda \in \mathbb{R}],\end{aligned}$$

a relation that yields that, for every parameter ν,

$$P_\nu\left[\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}[L_k(\hat{\lambda}_k; X_1^k) \geq L_k(\lambda; X_1^k), \lambda \in \mathbb{R}]\right] \geq P_\nu\left[\bigcap_{k=m}^{\infty}[X_1^k \in \mathcal{S}_k]\right].$$

After taking the limit as $m \to \infty$, an application of Lemma 3.3 leads to

$$P_\nu\left[\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}[L_k(\hat{\lambda}_k; X_1^k) \geq L_k(\lambda; X_1^k), \lambda \in \mathbb{R}]\right] = 1, \quad \nu \in \mathbb{R},$$

so that, by Definition 3.1, $\{\hat{\lambda}_n\}$ is a sequence of maximum likelihood estimators. □

4. Consistency

The objective of this section is to show that a sequence $\{\hat{\lambda}_n\}$ of maximum likelihood estimators of λ is consistent, *i.e.*, that $\{\hat{\lambda}_n\}$ converges to the true parameter value with probability 1.

Theorem 4.1. *Suppose that Assumption 2.1 holds and let $\{\hat{\lambda}_n\}$ be a sequence of maximum likelihood estimators. In this context,*

$$P_\nu\left[\lim_{n\to\infty}\hat{\lambda}_n = \nu\right] = 1, \quad \nu \in \mathbb{R}.$$

The proof of this result relies on the two lemmas stated below and involves the following notation: Throughout the remainder of the section $\{\hat{\lambda}_n\}$ is a given sequence of maximum likelihood estimators of λ and the event Ω^* is given by

$$\Omega^* := \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}[L_k(\hat{\lambda}_k; X_1^k) \geq L_k(\lambda; X_1^k),\ \lambda \in \mathbb{R}]. \tag{4.1}$$

Also, for each $\nu \in \mathbb{R}$,

$$\begin{aligned}\Omega_\nu^- &:= \left[\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} I[X_i \leq -c(\nu)] = P_\nu[X \leq -c(\nu)]\right],\\ \Omega_\nu^+ &:= \left[\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} I[X_i \geq c(\nu))] = P_\nu[X \geq c(\nu)]\right],\end{aligned} \tag{4.2}$$

where $c(\nu) > 0$ is as in Lemma 2.2(ii). Notice that the strong law of large numbers and Definition 3.1 yield

$$P_\nu[\Omega^*] = P_\nu[\Omega_\nu^+] = 1 = P_\nu[\Omega_\nu^-], \quad \nu \in \mathbb{R}. \tag{4.3}$$

The core of the proof of Theorem 4.1 is the following boundedness property.

Lemma 4.1. *Under Assumption 2.1, if $\{\hat{\lambda}_n\}$ is a sequence of maximum likelihood estimators of λ, then*

$$P_\nu\left[\limsup_{n\to\infty}|\hat{\lambda}_n| < \infty\right] = 1, \quad \nu \in \mathbb{R}.$$

Proof. It will be shown, by contradiction, that the event

$$\left[\limsup_{n\to\infty} \hat\lambda_n = \infty\right] \cap \Omega^* \cap \Omega_\nu^- \quad \text{is empty.} \tag{4.4}$$

To achieve this goal, *suppose* that the sample trajectory $X_1, X_2, \ldots$ is such that the above intersection occurs, and observe that *along this path* assertions (a)–(c) below hold:

(a) Since the event $[\limsup_n \hat\lambda_n = \infty]$ occurs when $X_1, X_2, \ldots$ is observed, there exist a (trajectory dependent) subsequence $\{n_k\}$ such that

$$n_k \geq k \quad \text{and} \quad \hat\lambda_{n_k} \geq k$$

for all positive integers k;

(b) Using that $X_1, X_2, \ldots$ is such that Ω^* occurs, it follows that $\hat\lambda_n$ maximizes $L_n(\cdot; X_1^n)$ for n large enough, so that there exists a positive integer M such that

$$L_n(\hat\lambda_n; X_1^n) \geq L_n(0; X_1^n) = -\log(2), \quad n \geq M;$$

(c) Since the observation of $X_1, X_2, \ldots$ implies that Ω_ν^- occurs,

$$\frac{1}{n}\sum_{i=1}^n I[X_i \leq -c(\nu)] \to P_\nu[X_1 \leq -c(\nu)] \quad \text{as } n \to \infty.$$

Notice now that for each integer $M_1 > M$ and $k > M_1$, (a) and (b) together yield

$$\begin{aligned}
-\log(2) \leq L_{n_k}(\hat\lambda_{n_k}; X_1^{n_k}) &= \frac{1}{n_k}\sum_{i=1}^{n_k} \log(G(\hat\lambda_{n_k} X_i)) \\
&\leq \frac{1}{n_k}\sum_{i=1}^{n_k} \log(G(\hat\lambda_{n_k} X_i)) I[X_i \leq -c(\nu)] \\
&\leq \frac{1}{n_k}\sum_{i=1}^{n_k} \log[G(-M_1 c(\nu))] I[X_i \leq -c(\nu)]
\end{aligned}$$

where (3.2) with $Q = \{i : i \leq n_k,\ X_i \leq -c(\nu)\}$ was used to set the second inequality, and the third one follows from the monotonicity of $\log(G(\cdot))$, since $\lambda_{n_k} > k > M_1$. From this point, letting k go to ∞, (c) leads to

$$-\log(2) \leq \log(G(-M_1 c(\nu))) P_\nu[X_1 \leq -c(\nu)];$$

and, recalling that $\lim_{x\to-\infty} \log(G(x)) = -\infty$ and that $P_\nu[X_1 \leq -c(\nu)]$ and $c(\nu)$ are both positive (by Lemma 2.2(ii)), taking the limit as $M_1 \to \infty$, it follows that $-\log(2) \leq -\infty$, which is a contradiction, establishing (4.4). Therefore, $[\limsup_{n\to\infty} \hat\lambda_n = \infty] \subset (\Omega^*)^c \cup (\Omega_\nu^-)^c$, inclusion that yields

$$P_\nu\left[\limsup_{n\to\infty} \hat\lambda_n = \infty\right] = 0, \quad \nu \in \mathbb{R},$$

by (4.3). Similarly, it can be established that $[\liminf_{n\to\infty} \hat\lambda_n = -\infty] \subset (\Omega^*)^c \cup (\Omega_\nu^+)^c$, so that $P_\nu[\liminf_{n\to\infty} \hat\lambda_n = -\infty] = 0$ for all $\nu \in \mathbb{R}$; the conclusion follows combining this fact with the above display. □

To continue, observe that since $\log(G(\cdot)) \le 0$, $E_\nu[\log(G(\lambda X_1))]$ is always a well-defined non positive number, where the expectation may assume the value $-\infty$. Also, since $x \mapsto x\log(x)$ is bounded on $(0,1)$, $E_\nu[\log(G(\nu x))] = \int_{\mathbb{R}} 2\log(G(\nu x))G(\nu x)f(x)\,dx$ is finite.

Lemma 4.2. *Suppose that Assumption 2.1 holds and let $\nu \in \mathbb{R}$ be arbitrary but fixed.*

(i) [Kullback's inequality.] *For each $\lambda \in \mathbb{R} \setminus \{\nu\}$,*

$$E_\nu[\log(G(\lambda X_1))] < E_\nu[\log(G(\nu X_1))].$$

(ii) Assume that $\{r_k\}$ and $\{s_k\}$ are two real sequences such that, for some $\nu^* \in \mathbb{R}$,

$$r_k \searrow \nu^* \quad \textit{and} \quad s_k \nearrow \nu^* \quad \textit{as } k \to \infty$$

and suppose that, for every $k = 1, 2, 3, \ldots$, the following inequality holds:

$$\begin{aligned} &E_\nu[\log(G(\nu X))] \\ &\quad \le E_\nu[\log(G(r_k X))I[X \ge 0]] + E_\nu[\log(G(s_k X))I[X < 0]]. \end{aligned} \tag{4.5}$$

In this case, $\nu = \nu^*$.

Proof. (i) If $\lambda \neq \nu$, Assumption 2.1 and the *strict concavity* of the logarithmic function yield, via Jensen's inequality, that

$$\begin{aligned} \int_{\mathbb{R}} \log\left(\frac{\rho(x;\lambda)}{\rho(x;\nu)}\right)\rho(x;\nu)\,dx &< \log\left(\int_{\mathbb{R}} \frac{\rho(x;\lambda)}{\rho(x;\nu)}\rho(x;\nu)\,dx\right) \\ &= \log\left(\int_{\mathbb{R}} \rho(x;\lambda)\,dx\right) = 0. \end{aligned}$$

Observing that $\rho(x;\lambda)/\rho(x;\nu) = G(\lambda x)/G(\nu x)$ when $\rho(x;\nu)$ is positive, the above inequality can be written as $E_\nu[\log(G(\lambda X_1)) - \log(G(\nu X_1))] < 0$, which yields the desired conclusion since, as already noted, $E_\nu[\log(G(\nu X_1))]$ is finite.

(ii) Notice that $r_k \searrow \nu^*$ leads to $r_k X \searrow \nu^* X$ on $[X \ge 0]$, whereas $s_k \nearrow \nu^*$ implies that $s_k X \searrow \nu^* X$ on $[X < 0]$. Therefore, since $-\log(G(\cdot))$ is decreasing and nonnegative,

$$0 \le -\log(G(r_k X))I[X \ge 0] \nearrow -\log(G(\nu^* X))I[X \ge 0]$$

and

$$0 \le -\log(G(s_k X))I[X < 0] \nearrow -\log(G(\nu^* X))I[X < 0].$$

Now, an application of the monotone convergence theorem yields

$$E_\nu[\log(G(r_k X))I[X \ge 0]] \searrow E_\nu[\log(G(\nu^* X))I[X \ge 0]]$$

and

$$E_\nu[\log(G(s_k X))I[X < 0]] \searrow E_\nu[\log(G(\nu^* X))I[X < 0]],$$

convergences that, after taking the limit as k goes to ∞ in (4.5), lead to

$$E_\nu[\log(G(\nu X))] \le E_\nu[\log(G(\nu^* X))],$$

and then $\nu = \nu^*$, by part (i). ∎

After these preliminaries, the proof of Theorem 4.1 is presented below. The argument uses the following notation, where $\mathcal{Q}$ stands for the set of rational numbers: For each $\nu \in \mathbb{R}$, define

$$\Omega_\nu^0 := \left[\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \log(G(\nu X_i)) = E_\nu[\log(G(\nu X_1))\right]; \tag{4.6}$$

$$\Omega_\nu^1 := \left[\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \log(G(\lambda X_i))I[X_i \ge 0] \right. \\ \left. = E_\nu[\log(G(\lambda X_1))I[X_1 \ge 0]], \quad \lambda \in \mathcal{Q}\right], \tag{4.7}$$

and

$$\Omega_\nu^2 := \left[\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \log(G(\lambda X_i))I[X_i < 0] \right. \\ \left. = E_\nu[\log(G(\lambda X_1))I[X_1 < 0]], \quad \lambda \in \mathcal{Q}\right]. \tag{4.8}$$

Since $\mathcal{Q}$ is denumerable, the strong law of large numbers yields $P_\nu[\Omega_\nu^i] = 1$ for $i = 0, 1, 2$, and then, setting

$$\Omega_\nu = \Omega_\nu^0 \cap \Omega_\nu^1 \cap \Omega_\nu^2, \tag{4.9}$$

it follows that

$$P_\nu[\Omega_\nu] = 1. \tag{4.10}$$

Proof of Theorem 4.1. Given $\nu \in \mathbb{R}$, it is sufficient to show that

$$\left[\limsup_n |\hat\lambda_n| < \infty\right] \cap \Omega^* \cap \Omega_\nu \subset \left[\lim_n \hat\lambda_n = \nu\right], \tag{4.11}$$

where Ω^* and Ω_ν are specified in (4.1) and (4.9), respectively. Indeed, if this inclusion is valid, then (4.3), (4.10) and Lemma 4.1 together imply that $P_\nu[\lim_n \hat\lambda_n = \nu] = 1$. To establish (4.11) let $X_1, X_2, X_3, \ldots$ be a fixed trajectory such that $[\limsup_n |\hat\lambda_n| < \infty] \cap \Omega^* \cap \Omega_\nu$ occurs, select an arbitrary limit point ν^* of the associated sequence $\{\hat\lambda_n\}$, and observe the following facts (a)–(c):

(a) ν^* is finite, since $\limsup_n |\hat\lambda_n| < \infty$ holds when $X_1, X_2, X_3, \ldots$ is observed. Let r and s be arbitrary rational numbers satisfying

$$s < \nu^* < r \tag{4.12}$$

and select a sequence $\{n_k\}$ of positive integers such that

$$n_k > k, \quad s < \hat\lambda_{n_k} < r, \quad k = 1, 2, \ldots. \tag{4.13}$$

(b) Since the path $X_1, X_2, X_3, \ldots$ is such that Ω^* occurs, $\hat\lambda_n$ is a maximizer of $L_n(\cdot; X_1^n)$ when n is large enough, say $n > M$; see (4.1). In particular, $L_n(\nu; X_1^n) \le$

$L_n(\hat{\lambda}_n; X_1^n)$ when $n > M$, and then, replacing n by n_k in this inequality and using the monotonicity of $\log(G(\cdot))$ as well as (4.13), it follows that

$$\begin{aligned}
\frac{1}{n_k}\sum_{i=1}^{n_k}\log(G(\nu X_i)) &= L_{n_k}(\nu; X_1^{n_k}) \\
&\le L_{n_k}(\hat{\lambda}_{n_k}; X_1^{n_k}) \\
&= \frac{1}{n_k}\sum_{i=1}^{n_k}\log(G(\hat{\lambda}_{n_k} X_i)) \\
&\le \frac{1}{n_k}\sum_{i=1}^{n_k}\log(G(rX_i))I[X_i \ge 0] \\
&\qquad + \frac{1}{n_k}\sum_{i=1}^{n_k}\log(G(sX_i))I[X_i < 0], \quad k > M.
\end{aligned} \tag{4.14}$$

(c) Since the trajectory $X_1, X_2, \ldots$ is such that Ω_ν occurs, a glance to (4.6)–(4.9) immediately yields that, as $k \to \infty$,

$$\frac{1}{n_k}\sum_{i=1}^{n_k}\log(G(\nu X_i)) \to E_\nu[\log(G(\nu X))],$$
$$\frac{1}{n_k}\sum_{i=1}^{n_k}\log(G(rX_i))I[X_i \ge 0] \to E_\nu[\log(G(rX)I[X \ge 0]], \text{ and}$$
$$\frac{1}{n_k}\sum_{i=1}^{n_k}\log(G(sX_i))I[X_i < 0] \to E_\nu[\log(G(sX)I[X < 0]].$$

After taking the limit as $k \to \infty$ in (4.14), these convergences yield that

$$E_\nu[\log(G(\nu X))] \le E_\nu[\log(G(rX)I[X \ge 0]] + E_\nu[\log(G(sX)I[X < 0]],$$

and then, since r and s are arbitrary rational numbers satisfying (4.12), from Lemma 4.2(ii) it follows that $\nu^* = \nu$. In short, it has been proved that along an arbitrary path $X_1, X_2, \ldots$ for which the intersection $[\limsup_n |\hat{\lambda}_n| < \infty] \cap \Omega^* \cap \Omega_\nu$ occurs, the corresponding sequence $\{\hat{\lambda}_n\}$ has ν as its unique limit point, so that $\hat{\lambda}_n \to \nu$ as $n \to \infty$. This establishes (4.11) and, as already noted, completes the proof. □

5. Asymptotic Distribution

The remainder of the paper concerns the asymptotic behavior of a consistent sequence $\{\hat{\lambda}_n\}$ of maximum likelihood estimators of λ, whose existence is guaranteed by Assumption 2.1. As already mentioned, the large sample properties of $\{\hat{\lambda}_n\}$ will be studied under the null hypothesis $\mathcal{H}_0 : \lambda = 0$, and the analysis below requires two properties on the densities g and f generating the family $S(f, g)$, namely, (i) smoothness of density g outside of $\{0\}$ and a 'good' behavior of its derivatives around $\lambda = 0$, and (ii) a moment-dominance condition involving both densities f and g. After a formal presentation of these assumptions, the main result is stated at the end of the section, and the corresponding proof is given after establishing the necessary technical preliminaries.

Assumption 5.1. *For some nonnegative integer r—hereafter referred to as the critical order— the following conditions hold:*
(i) The symmetric density $g(x)$ is continuous on $\mathbb{R}$ and has derivatives up to order $2r+2$ on the interval $(0,\infty)$;
(ii) $D^k g(0+) = \lim_{x\searrow 0} D^k g(x)$ exists and is finite for $k = 0,1,2,\ldots,2r+1$, and
(iii) $D^r g(0+) \neq 0$, whereas $D^s g(0+) = 0$ for $0 \le s < r$.

Remark 5.1. (i) Under this Assumption, $g(\cdot)$ statisfies that $D^r g(0+) = \lim_{x\searrow 0} r!g(x)/x^r \ge 0$, by continuity of g if $r = 0$, or by L'Hopital's rule, if $r > 0$, so that $D^r g(0+) > 0$, since $D^r g(0+)$ is no null. It follows that a density g satisfying Assumption 5.1 can be expressed as $g(x) = |x|^r h(|x|)$ where r is a nonnegative integer (the critical order), $h : [0,\infty) \to [0,\infty)$ is continuous with $h(0) > 0$, and has derivatives of order up to $2r+2$ on $(0,\infty)$, which are 'well-behaved' near zero so that the required lateral limits of the derivatives of g exist at $x = 0$. Thus, besides the smoothness requirement on the whole interval $(0,\infty)$, the core of Assumption 5.1 essentially concerns the local behavior of g around the origin.

(ii) Under Assumption 5.1, density $g(\cdot)$ is continuous, so that $G(\cdot)$ has continuous derivative, and then $\partial_\lambda \ell(\lambda;x) = \partial_\lambda \log(G(\lambda x))$ exists if $\ell(\lambda;x)$ is finite. Suppose now that $\hat\lambda_n$ maximizes $L_n(\cdot;X_1^n)$. In this case $-\log(2) = L_n(0;X_1^n) \le L_n(\hat\lambda_n;X_1^n)$ implies that $\log(G(\hat\lambda_n X_i)$ is finite for every $i = 1,2,\ldots,n$, by (3.1), so that $L_n(\cdot;X_1^n)$ is differentiable at $\hat\lambda_n$, and then the likelihood equation holds: $\partial_\lambda L_n(\hat\lambda_n;X_1^n) = 0$.

By symmetry, Assumption 5.1 yields that $g(\cdot)$ also has derivatives up to order $2r+2$ on the interval $(-\infty,0)$; indeed, if $k \le 2r+2$ then $D^k g(x) = (-1)^k D^k g(-x)$ for $x \neq 0$, so that

$$D^k g(0-) = \lim_{x\nearrow 0} D^k g(x) = (-1)^k D^k g(0+), \quad k = 0,1,2,\ldots,2r+1, \tag{5.1}$$

and the nullity of $D^k g(0+)$ for $0 \le k < r$ implies that g has null (bilateral) derivative at $x = 0$ of any order less that r. On the other hand, under Assumption 5.1 the cumulative distribution function G in (1.1) has derivatives up to order $2r+3$ on $\mathbb{R}\setminus\{0\}$, and using the relation $D^k G(x) = D^{k-1} g(x)$ for $x \neq 0$ and $k > 0$ it follows that

$$D^k G(0) = 0, \quad 1 \le k < r+1 \tag{5.2}$$

and

$$D^k G(0+) = D^{k-1} g(0+), \quad D^k G(0-) = D^{k-1} g(0-), \quad r+1 \le k \le 2r+2. \tag{5.3}$$

Next, define

$$H(x) := \log(G(x)), \quad x \in \mathbb{R}, \tag{5.4}$$

and observe the equalities

$$\begin{aligned} \ell(\lambda;x) &= H(\lambda x), \quad x \in \mathbb{R}, \\ \partial_\lambda^k \ell(\lambda;x) &= D^k H(\lambda x) x^k, \quad x \neq 0, \quad 1 \le k \le 2r+3 \end{aligned} \tag{5.5}$$

see (1.4). It follows that the lateral limits of $\partial_\lambda^k \ell(\cdot;x)$ at zero are given by

$$\partial_\lambda^k \ell(0+;x) = \begin{cases} D^k H(0+) x^k, \text{ if } x > 0; \\ D^k H(0-) x^k, \text{ if } x < 0 \end{cases} \tag{5.6}$$

and

$$\partial_\lambda^k \ell(0-;x) = \begin{cases} D^k H(0-)x^k, \text{ if } x > 0; \\ D^k H(0+)x^k, \text{ if } x < 0. \end{cases} \tag{5.7}$$

The analysis below uses (lateral) Taylor expansions of order $2r+2$ for $L_n(\cdot; X_1^n)$ around zero, and it is necessary to have an integrable bound for the residual as well as finite second moments for the coefficients. For this reason, the following conditions will be enforced.

Assumption 5.2. *Conditions (i) and (ii) below hold, where r is the the critical order in Assumption 5.1:*

(i) $E_0[X_1|^{4r+2}] = \int_{\mathbb{R}} x^{4r+2} f(x)\,dx < \infty$;

(ii) There exists a function $W : \mathbb{R} \to [0,\infty)$ and $\delta > 0$ such that

$$|\partial_\lambda^{2r+3}\ell(\lambda;\cdot)| \le W(\cdot), \quad 0 < |\lambda| \le \delta, \quad \textit{and} \quad \int_{\mathbb{R}} W(x)f(x)\,dx < \infty. \tag{5.8}$$

Remark 5.2. The moment requirement in Assumption 5.2(i) concerns only density f, whereas the dominance condition in the second part involves a relation between g and f. Using (5.4), it can be shown by induction that, for $x \neq 0$, $D^k H(x)$ is a polynomial in $D^s g(x)/G(x), s = 0, 1, 2, \ldots, k-1$, and then, setting

$$M_r := \max\{|D^k g(x)|/G(x) : 0 \le k \le 2r+2,\ x \neq 0\},$$

via (5.5) it follows that there exists a constant B such that $|\partial_\lambda^{2r+3}\ell(\lambda, x)| \le BM_r|x|^{2r+3}$ so that, if $M_r < \infty$, then Assumption 5.2 holds entirely if the moment condition in part (i) is valid. It is not difficult to see that M_r is finite when there exists $x_0 > 0$ for which (a) or (b) below occur:

(a) $g(x)$ is a rational function for $x > x_0$, as it is the case if $g(\cdot)$ is a multiple of a t-density for x large enough;

(b) $g(x) = p(x)e^{-\beta x}$ on (x_0, ∞), where β is a positive constant and $p(x)$ is a polynomial or, more generally, a linear combination of terms of the form x^s; this occurs, for instance, when $g(\cdot)$ is proportional to a mixture of gamma densities on (x_0, ∞).

To state the result on the large sample distribution of maximum likelihood estimators, set

$$V_{r+1} := 4\left(\frac{D^r g(0+)}{(r+1)!}\right)^2 E_0\left[X_1^{2r+2}\right] > 0 \tag{5.9}$$

which, as it will be shown later, is the variance of $\partial_\lambda^{r+1}\ell(0+, X_i)/(r+1)!$; for the strict inequality, see Remark 5.1(i).

Theorem 5.1. *Let $\{\hat\lambda_n\}$ be a consistent sequence of maximum likelihood estimators of λ, and suppose that Assumptions 5.1 and 5.2 hold. In this context, under the hypothesis $\mathcal{H}_0 : \lambda = 0$, the following convergences (i) and (ii) occur as $n \to \infty$, where r is the critical order in Assumption 5.1, and Z is a random variable with standard normal distribution:*

$$(nV_{r+1})^{1/(2(r+1))}\,\hat\lambda_n \xrightarrow{\mathrm{d}} |Z|^{1/(r+1)}\mathrm{sign}(Z), \tag{i}$$

and

$$2n[L_n(\hat\lambda_n; X_1^n) - L_n(0; X_1^n)] \xrightarrow{\mathrm{d}} Z^2. \tag{ii}$$

Remark 5.3. (i) Suppose that the critical index r is null. In this case Theorem 5.1 (i) yields that $(nV_1)^{1/2}\,\hat\lambda_n \xrightarrow{\mathrm{d}} |Z|\mathrm{sign}(Z) \stackrel{\mathrm{d}}{=} Z$. This conclusion coincides with that obtained from the general classical results presented, for instance, in Lehmann and Casella (1998, Section 6.3), or Shao ([16], Section 4.4), where derivatives up to order 2 are required for $g(\cdot)$ around zero; under Assumption 5.1, only the lateral limits of Dg and D^2g exist at zero.

(ii) Suppose that the critical order r is positive and that $g(\cdot)$ has (bilateral) derivative of order r at zero, so that $D^r g(0+) = D^r g(0-)$. Since $D^r g(0+) \neq 0$ it follows from (5.1) that r is an even integer, and $g(\cdot)$ has derivatives up to order r on the real line. Thus, setting $s = r+1$, s is odd, $\ell(\cdot; x)$ has derivatives up to order s on $\mathbb{R}$, and $\partial_\lambda^k \ell(0;\cdot) = 0$ for $1 \le k < s$, whereas $\partial_\lambda^s \ell(0, x) \neq 0$ for $x \neq 0$. If, among other conditions, $g(\cdot)$ has derivatives up to order $2s$ at zero, an application of Theorem 1 in Rotnitzky *et al.* [14] yields the conclusions in Theorem 5.1; notice, however, that Assumption 5.1 only ensures the existence of the lateral limits $D^k g(0\pm)$ for $s < k \le 2s$, so that Theorem 5.1 extends Theorem 1 in Rotnitzky *et al.* [14] to the framework of this work.

The rather technical proof of Theorem 5.1, requiring some explicit computations for the lateral limits $\partial_\lambda^k \ell(0\pm; x)$ in terms of density $g(\cdot)$, will be given after the preliminaries established in the following section.

6. Technical Preliminaries

This section is dedicated to establish the basic tools that will be used to prove Theorem 5.1, namely, lateral Taylor expansions around the origin for the average likelihood $L_n(\cdot; X_1^n)$ and its first derivative; via the absolute value function, such expansions are stated below as single equations. The following notation will be used:

$$\Delta(x) := \partial_\lambda^{r+1}\ell(0-; x) - \partial_\lambda^{r+1}\ell(0+; x), \quad x \in \mathbb{R}. \tag{6.1}$$

Theorem 6.1. *Suppose that Assumptions 5.1 and 5.2 hold. In this case, the assertions (i)—(iii) below occur, where r is the critical order in Assumption 5.1, $\delta > 0$ and $W(\cdot)$ are as in Assumption 5.2, and*

$$\Delta_n := \frac{1}{n}\sum_{i=1}^{n} \Delta(X_i);$$

see (6.1).

(i) For each positive integer n and $\alpha \in (-\delta, \delta)$,

$$\begin{aligned} &L_n(\alpha; X_1^n) - L_n(0; X_1^n) \\ &= |\alpha|^r \alpha \Bigg\{ \sum_{k=r+1}^{2r+1} \frac{\partial_\lambda^k L_n(0+; X_1^n)}{k!} |\alpha|^{k-r-1} \\ &+ \left(\frac{\partial_\lambda^{2r+2} L_n(0+; X_1^n) + \Delta_n I_{(-\infty,0)}(\alpha)}{(2r+2)!} + \frac{W_n^*(\alpha)}{(2r+3)!}\alpha \right) |\alpha|^r \alpha \Bigg\}, \end{aligned} \tag{6.2}$$

and

$$
\begin{aligned}
&\partial_\lambda L_n(\alpha, X_1^n)\\
&= |\alpha|^r \Bigg\{ \sum_{k=r}^{2r} \frac{\partial_\lambda^{k+1} L_n(0+; X_1^n)}{k!} |\alpha|^{k-r}\\
&+ \left(\frac{\partial_\lambda^{2r+2} L_n(0+; X_1^n) + \Delta_n I_{(-\infty,0)}(\alpha)}{(2r+1)!} + \frac{\tilde{W}_n(\alpha)}{(2r+2)!}\alpha \right) |\alpha|^r \alpha \Bigg\},
\end{aligned} \tag{6.3}
$$

where the random variables $W_n^*(\alpha)$ *and* $\tilde{W}_n(\alpha)$ *satisfy*

$$
|W_n^*(\alpha)|,\ |\tilde{W}_n(\alpha)| \le W_n := \frac{1}{n} \sum_{i=1}^n W(X_i). \tag{6.4}
$$

(ii) Under $\mathcal{H}_0 : \lambda = 0$ *for each* $k = r+1, \ldots, 2r+1$*, the following convergences hold as* $n \to \infty$*:*

$$
\sqrt{n}\, \partial_\lambda^k L_n(0+; X_1^n) \xrightarrow{\mathrm{d}} \mathcal{N}(0, v_k), \quad \textit{where } v_k = E_0 \left[\left(\partial_\lambda^k \ell(0+; X_1) \right)^2 \right], \tag{6.5}
$$

whereas

$$
\Delta_n \to 0 \quad \textit{and} \quad 2\frac{\partial_\lambda^{2r+2} L_n(0+; X_1^n)}{(2r+2)!} \to -V_{r+1} \quad P_0\textit{-a.s.}; \tag{6.6}
$$

see (5.9).

The proof of this theorem relies on explicit formulas for $\partial_\lambda^k \ell(0\pm; x)$ in terms of density $g(\cdot)$ and, in this direction, the following lemma concerning the lateral limits at zero of the derivatives of function $H(\cdot)$ in (5.4) will be useful.

Lemma 6.1. *Suppose that Assumption 5.1 holds. In this case, the lateral limits at 0 of the derivatives of function* $H(\cdot)$ *in (5.4) satisfy the following relations (i)–(iii):*

(i) $D^k H(0+) = D^k H(0-) = 0$, $1 \le k < r+1$*;*
(ii) If $r+1 \le k < 2r+2$*, then*

$$
D^k H(0+) = 2D^{k-1} g(0+) \quad \textit{and} \quad D^k H(0-) = 2D^{k-1} g(0-).
$$

(iii)
$$
D^{2r+2} H(0+) = 2D^{2r+1} g(0+) - \frac{1}{2} \binom{2r+2}{r+1} \left(D^{r+1} H(0+) \right)^2,
$$

and

$$
D^{2r+2} H(0-) = 2D^{2r+1} g(0-) - \frac{1}{2} \binom{2r+2}{r+1} \left(D^{r+1} H(0-) \right)^2.
$$

Proof. Recalling that the distribution function $G(x)$ is continuous and has derivatives up to order $2r+3$ on $\mathbb{R} \setminus \{0\}$; from (5.4) it follows that $G(x) DH(x) = DG(x)$ and, via Leibinitz' formula,

$$
G(x) D^k H(x) + \sum_{i=1}^{k-1} \binom{k-1}{i} D^i G(x) D^{k-i} H(x) = D^k G(x)
$$

for $x \neq 0$ and $2 \le k \le 2r+3$; since $G(\cdot)$ is continuous and $G(0) = 1/2$, taking lateral limit as x approaches to zero these equalities lead to

$$
DH(0\pm) = 2DG(0)
$$

and, for $2 \le k \le 2r+2$,

$$D^k H(0\pm) + 2\sum_{i=1}^{k-1} \binom{k-1}{i} D^i G(0\pm) D^{k-i} H(0\pm) = 2D^k G(0\pm).$$

Since $D^k G(0\pm) = 0$ when $1 \le k \le r$, by (5.2), these relations yield $D^k H(0\pm) = 0$ for $1 \le k \le r$, establishing part(i), as well as

$$D^{r+1} H(0\pm) = 2D^{r+1} G(0\pm), \tag{6.7}$$

and

$$\text{for } r+1 < k \le 2r+2, \quad D^k H(0\pm) + 2\sum_{i=r+1}^{k-1} \binom{k-1}{i} D^i G(0\pm) D^{k-i} H(0\pm) = 2D^k G(0\pm). \tag{6.8}$$

To prove part (ii), select an integer k such that $r+1 < k < 2r+2$. In this case, if $k > i \ge r+1$ then $1 \le k-i < r+1$, and then $D^{k-i}H(0\pm) = 0$, by part (i), so that the summation in the above display is null. Therefore, $D^k H(0\pm) = 2D^k G(0\pm)$, and combining this with (6.7) it follows that

$$D^k H(0\pm) = 2D^k G(0\pm), \quad r+1 \le k < 2r+2,$$

equalities that yield part (ii) via (5.3). To conclude, observe that if $k = 2r+2$ then $2r+1 \ge i > r+1$ implies that $1 \le k-i < r+1$, and in this case $D^{k-i}H(0\pm) = 0$, by part (i), so that the terms in the summation in (6.8) with $k = 2r+2$ are null when $i > r+1$. Consequently,

$$D^{2r+2} H(0\pm) + 2\binom{2r+1}{r+1} D^{r+1} G(0\pm) D^{r+1} H(0\pm) = 2D^{2r+2} G(0\pm);$$

since

$$2D^{r+1}G(0\pm) = 2D^r g(0\pm) = D^{r+1} H(0\pm)$$

and $D^{2r+2}G(0\pm) = D^{2r+1}g(0\pm)$, by (5.3) and part (ii), respectively, the conclusion follows observing that $\binom{2r+1}{r+1} = 2^{-1}\binom{2r+2}{r+1}$. □

The expressions in the previous lemma are used below to determine the lateral limits of $\partial_\lambda^k \ell(\cdot; x)$ at zero in terms of density $g(\cdot)$.

Lemma 6.2. *Under Assumption 5.1, assertions (i)–(v) below hold:*

(i) $\partial_\lambda^k \ell(0+;\cdot) = 0 = \partial_\lambda^k \ell(0-;\cdot)$ *for* $1 \le k \le r$.
(ii) For each $x \in \mathbb{R}$ *and* $r+1 \le k < 2r+2$,

$$\partial_\lambda^k \ell(0+;x) = 2D^{k-1} g(0+)|x|^{k-1} x,$$
$$\partial_\lambda^k \ell(0-;x) = 2D^{k-1} g(0-)|x|^{k-1} x.$$

(iii) $\partial_\lambda^k \ell(0-;x) = (-1)^{k-1} \partial_\lambda^k \ell(0+;x)$ *for* $r+1 \le k < 2r+2$ *and* $x \in \mathbb{R}$.
(iv) For each $x \in \mathbb{R}$

$$\partial_\lambda^{2r+2} \ell(0+;x) = 2D^{2r+1} g(0+)|x|^{2r+1} x - \frac{1}{2}\binom{2r+2}{r+1}\left(\partial_\lambda^{r+1}\ell(0+;x)\right)^2, \tag{6.9}$$

and

$$\partial_\lambda^{2r+2}\ell(0-;x) = 2D^{2r+1}g(0-)|x|^{2r+1}x - \frac{1}{2}\binom{2r+2}{r+1}\left(\partial_\lambda^{r+1}\ell(0-;x)\right)^2. \tag{6.10}$$

Consequently,

(v) The difference between $\partial_\lambda^{2r+2}\ell(0-;x)$ *and* $\partial_\lambda^{2r+2}\ell(0+;x)$ *is given by*

$$\Delta(x) = -4D^{2r+1}g(0+)\ |x|^{2r+1}x, \quad x \in \mathbb{R}; \tag{6.11}$$

see (6.1).

Proof. From Lemma 6.1(i), part (i) follows via (5.6) and (5.7), whereas these latter equalities and Lemma 6.1(ii) together yield that, for $r+1 \leq k \leq 2r+1$, (a) and (b) below hold:

(a) For $x \geq 0$,

$$\partial_\lambda^k\ell(0+;x) = 2D^{k-1}g(0+)x^k \quad \text{and} \quad \partial_\lambda^k\ell(0-;x) = 2D^{k-1}g(0-)x^k;$$

(b) If $x < 0$,

$$\begin{aligned}\partial_\lambda^k\ell(0+;x) &= 2D^{k-1}g(0-)x^k\\ &= 2D^{k-1}g(0+)(-1)^{k-1}x^k = 2D^{k-1}g(0+)|x|^{k-1}x,\\ \partial_\lambda^k\ell(0-;x) &= 2D^{k-1}g(0+)x^k\\ &= 2D^{k-1}g(0-)(-1)^{k-1}x^k = 2D^{k-1}g(0-)|x|^{k-1}x,\end{aligned}$$

where (5.1) was used to set the second equalities. These facts (a) and (b) together lead to part (ii), which implies part (iii) via (5.1). To establish part (iv), notice that Lemma 6.1(iii) and (5.6) together imply that:

For $x \geq 0$

$$\begin{aligned}\partial_\lambda^{2r+2}\ell(0+;x) &= 2D^{2r+1}g(0+)x^{2r+2} - \frac{1}{2}\binom{2r+2}{r+1}\left(D^{r+1}H(0+)x^{r+1}\right)^2\\ &= 2D^{2r+1}g(0+)x^{2r+2} - \frac{1}{2}\binom{2r+2}{r+1}\left(\partial_\lambda^{r+1}\ell(0+;x)\right)^2,\end{aligned}$$

showing that (6.9) holds for $x \geq 0$, whereas combining Lemma 6.1(iii) with relations (5.6) and (5.1) it follows that if $x < 0$, then

$$\begin{aligned}&\partial_\lambda^{2r+2}\ell(0+;x)\\ &\quad = 2D^{2r+1}g(0-)x^{2r+2} - \frac{1}{2}\binom{2r+2}{r+1}\left(D^{r+1}H(0-)x^{r+1}\right)^2\\ &\quad = 2D^{2r+1}g(0+)(-1)^{2r+1}x^{2r+1}x - \frac{1}{2}\binom{2r+2}{r+1}\left(\partial_\lambda^{r+1}\ell(0+;x)\right)^2\\ &\quad = 2D^{2r+1}g(0+)|x|^{2r+1}x - \frac{1}{2}\binom{2r+2}{r+1}\left(\partial_\lambda^{r+1}\ell(0+;x)\right)^2,\end{aligned}$$

and then (6.9) also holds for $x < 0$. Equality (6.10) can be established along similar lines and, finally, observing that $\partial_\lambda^{r+1}\ell(0+;x)$ and $\partial_\lambda^{r+1}\ell(0-;x)$ have the same absolute value, by part (iii), via part (iv), (6.11) follows immediately from (5.1) and (6.1). □

Next, the above expressions will be used to write lateral Taylor expansions for $\ell(\cdot;x)$ and $\partial_\lambda \ell(\cdot;x)$ around the origin.

Lemma 6.3. *Suppose that Assumptions 5.1 and 5.2 hold. In this case, the following expansions are valid for $x \in \mathbb{R}$ and $\alpha \in (-\delta,\delta)\setminus\{0\}$:*

$$\begin{aligned}&\ell(\alpha,x)-\ell(0;x)\\ &= |\alpha|^r\alpha\left\{\sum_{k=r+1}^{2r+1}\frac{\partial_\lambda^k\ell(0+;x)}{k!}|\alpha|^{k-r-1}\right.\\ &\qquad\left.+\left(\frac{\partial_\lambda^{2r+2}\ell(0+;x)+I_{(-\infty,0)}(\alpha)\Delta(x)}{(2r+2)!}+\frac{W^*(\alpha,x)}{(2r+3)!}\alpha\right)|\alpha|^r\alpha\right\},\end{aligned}$$

and

$$\begin{aligned}&\partial_\lambda\ell(\alpha,x)\\ &= |\alpha|^r\left\{\sum_{k=r}^{2r}\frac{\partial_\lambda^{k+1}\ell(0+;x)}{k!}|\alpha|^{k-r}\right.\\ &\qquad\left.+\left(\frac{\partial_\lambda^{2r+2}\ell(0+;x)+I_{(-\infty,0)}(\alpha)\Delta(x)}{(2r+1)!}+\frac{\tilde W(\alpha,x)}{(2r+2)!}\alpha\right)|\alpha|^r\alpha\right\},\end{aligned}$$

where $\Delta(\cdot)$ is as in (6.1), and

$$|W^*(\alpha,x)|,\ |\tilde W(\alpha,x)|\le W(x). \tag{6.12}$$

Proof. Select $\alpha_0\neq 0$ with the same sign as α and $|\alpha_0|<|\alpha|$, so that the closed interval joining α_0 and α is contained in $(-\delta,\delta)\setminus\{0\}$. Since $\ell(\cdot;x)$ has derivatives up to order $2r+3$ outside 0, there exist points α^* and $\tilde\alpha$ between α_0 and α such that the following Taylor expansions hold:

$$\ell(\alpha,x)-\ell(\alpha_0;x)=\sum_{k=1}^{2r+2}\frac{\partial_\lambda^k\ell(\alpha_0;x)}{k!}(\alpha-\alpha_0)^k+\frac{\partial_\lambda^{2r+3}\ell(\alpha^*;x)}{(2r+3)!}(\alpha-\alpha_0)^{2r+3}, \tag{6.13}$$

and

$$\partial_\lambda\ell(\alpha,x)=\sum_{k=0}^{2r+1}\frac{\partial_\lambda^{k+1}\ell(\alpha_0;x)}{k!}(\alpha-\alpha_0)^k+\frac{\partial_\lambda^{2r+3}\ell(\tilde\alpha;x)}{(2r+2)!}(\alpha-\alpha_0)^{2r+2}, \tag{6.14}$$

where

$$\left|\partial_\lambda^{2r+3}\ell(\alpha^*;x)\right|,\left|\partial_\lambda^{2r+3}\ell(\tilde\alpha;x)\right|\le W(x), \tag{6.15}$$

by Assumption 5.2(ii). Next, the conclusions in the lemma will be obtained taking lateral limits as α_0 goes to zero. Recall that $\ell(\cdot;x)$ is continuous and consider the following exhaustive cases:

Case 1: $\alpha>0$. taking the limit as α_0 decreases to zero, the above displayed relations and Lemma 6.2(i) together yield

$$\begin{aligned}&\ell(\alpha,x)-\ell(0;x)\\ &\quad=\sum_{k=r+1}^{2r+2}\frac{\partial_\lambda^k\ell(0+;x)}{k!}\alpha^k+\frac{W^*(\alpha;x)}{(2r+3)!}\alpha^{2r+3}\end{aligned}$$

$$= \sum_{k=r+1}^{2r+1} \frac{\partial_\lambda^k \ell(0+;x)}{k!}\alpha^k + \frac{\partial_\lambda^{2r+2}\ell(0+;x)}{(2r+2)!}\alpha^{2r+2} + \frac{W^*(\alpha;x)}{(2r+3)!}\alpha^{2r+3}$$
$$= \alpha^{r+1}\left\{\sum_{k=r+1}^{2r+1} \frac{\partial_\lambda^k \ell(0+;x)}{k!}\alpha^{k-r-1} \right.$$
$$\left. + \left(\frac{\partial_\lambda^{2r+2}\ell(0+;x)}{(2r+2)!} + \frac{W^*(\alpha;x)}{(2r+3)!}\alpha\right)\alpha^{r+1}\right\},$$

and

$$\partial_\lambda \ell(\alpha;x) = \sum_{k=r}^{2r+1} \frac{\partial_\lambda^{k+1}\ell(0+;x)}{k!}\alpha^k + \frac{\tilde{W}(\alpha;x)}{(2r+2)!}\alpha^{2r+2}$$
$$= \sum_{k=r}^{2r} \frac{\partial_\lambda^{k+1}\ell(0+;x)}{k!}\alpha^k + \frac{\partial_\lambda^{2r+2}\ell(0+;x)}{(2r+1)!}\alpha^{2r+1} + \frac{\tilde{W}(\alpha;x)}{(2r+2)!}\alpha^{2r+2}$$
$$= \alpha^r\left\{\sum_{k=r}^{2r} \frac{\partial_\lambda^{k+1}\ell(0+;x)}{k!}\alpha^{k-r} \right.$$
$$\left. + \left(\frac{\partial_\lambda^{2r+2}\ell(0+;x)}{(2r+1)!} + \frac{\tilde{W}(\alpha;x)}{(2r+2)!}\alpha\right)\alpha^{r+1}\right\},$$

where $W^*(\alpha,x)$ is given by $W^*(\alpha,x) := \lim_{\alpha_0\searrow 0}\partial_\lambda^{2r+3}\ell(\alpha^*;x)$ and, similarly, $\tilde{W}(\alpha,x) := \lim_{\alpha_0\searrow 0}\partial_\lambda^{2r+3}\ell(\tilde{\alpha};x)$, so that

$$|W^*(\alpha,x)|, |\tilde{W}(\alpha,x)| \le W(x),$$

by (6.15); since α is positive, so that $I_{(-\infty,0)}(\alpha) = 0$, these last three displays are equivalent to (6.12)–(6.12).

Case 2: $\alpha < 0$. In this context, taking the limit as α_0 increases to zero in (6.13) and (6.14), Lemma 6.2(i) yields that

$$\ell(\alpha,x) - \ell(0;x) = \sum_{k=r+1}^{2r+2} \frac{\partial_\lambda^k \ell(0-;x)}{k!}\alpha^k + \frac{W^*(\alpha;x)}{(2r+3)!}\alpha^{2r+3}$$

and

$$\partial_\lambda \ell(\alpha;x) = \sum_{k=r}^{2r+1} \frac{\partial_\lambda^{k+1}\ell(0-;x)}{k!}\alpha^k + \frac{\tilde{W}(\alpha;x)}{(2r+2)!}\alpha^{2r+2},$$

where, analogously to the previous case, $W^*(\alpha,x) := \lim_{\alpha_0\nearrow 0}\partial_\lambda^{2r+3}\ell(\alpha^*;x)$ and $\tilde{W}(\alpha,x) := \lim_{\alpha_0\nearrow 0}\partial_\lambda^{2r+3}\ell(\tilde{\alpha};x)$ so that, again, (6.15) implies that (6.12) is valid. Observe now that Lemma 6.2(iii) allows to write

$$\sum_{k=r+1}^{2r+2} \frac{\partial_\lambda^k \ell(0-;x)}{k!}\alpha^k + \frac{W^*(\alpha;x)}{(2r+3)!}\alpha^{2r+3}$$
$$= \sum_{k=r+1}^{2r+1} \frac{\partial_\lambda^k \ell(0+;x)}{k!}(-1)^{k-1}\alpha^k + \frac{\partial_\lambda^{2r+2}\ell(0-;x)}{(2r+2)!}\alpha^{2r+2} + \frac{W^*(\alpha;x)}{(2r+3)!}\alpha^{2r+3}$$
$$= \sum_{k=r+1}^{2r+1} \frac{\partial_\lambda^k \ell(0+;x)}{k!}|\alpha|^{k-1}\alpha + \left(\frac{\partial_\lambda^{2r+2}\ell(0-;x)}{(2r+2)!} + \frac{W^*(\alpha;x)}{(2r+3)!}\alpha\right)(|\alpha|^r\alpha)^2$$

$$= |\alpha|^r \alpha \left\{ \sum_{k=r+1}^{2r+1} \frac{\partial_\lambda^k \ell(0+;x)}{k!} |\alpha|^{k-r-1} + \left(\frac{\partial_\lambda^{2r+2}\ell(0-;x)}{(2r+2)!} + \frac{W^*(\alpha;x)}{(2r+3)!}\alpha \right) |\alpha|^r \alpha \right\}$$

and

$$\sum_{k=r}^{2r+1} \frac{\partial_\lambda^{k+1}\ell(0-;x)}{k!}\alpha^k + \frac{\tilde W(\alpha;x)}{(2r+2)!}\alpha^{2r+2}$$
$$= \sum_{k=r}^{2r} \frac{\partial_\lambda^{k+1}\ell(0+;x)}{k!}(-1)^k\alpha^k + \frac{\partial_\lambda^{2r+2}\ell(0-;x)}{(2r+1)!}\alpha^{2r+1} + \frac{\tilde W(\alpha;x)}{(2r+2)!}\alpha^{2r+2}$$
$$= \sum_{k=r}^{2r} \frac{\partial_\lambda^{k+1}\ell(0+;x)}{k!}|\alpha|^k + |\alpha|^r \left(\frac{\partial_\lambda^{2r+2}\ell(0-;x)}{(2r+1)!} + \frac{\tilde W(\alpha;x)}{(2r+2)!}\alpha \right) |\alpha|^r \alpha$$
$$= |\alpha|^r \left\{ \sum_{k=r}^{2r} \frac{\partial_\lambda^{k+1}\ell(0+;x)}{k!}|\alpha|^{k-r} + \left(\frac{\partial_\lambda^{2r+2}\ell(0-;x)}{(2r+1)!} + \frac{\tilde W(\alpha;x)}{(2r+2)!}\alpha \right) |\alpha|^r \alpha \right\};$$

using that $\partial_\lambda^{2r+2}\ell(0-;x) = \partial_\lambda^{2r+2}\ell(0+;x) + I_{(-\infty,0)}(\alpha)\Delta(x)$, by (6.1), the last four displays together yield that (6.12) and (6.12) are also valid for $\alpha < 0$. □

After the above preliminaries, the main result of this section can be established as follows.

Proof of Theorem 6.1. (i) Since $L_n(\cdot;X_1^n)$ is the average of $\ell(\cdot;X_i)$, $i = 1, 2, \ldots, n$, by Lemma 6.3 the two indicated expansions hold with

$$W_n^*(\alpha) = \frac{1}{n}\sum_{i=1}^n W^*(\alpha, X_i) \quad \text{and} \quad \tilde W_n(\alpha) = \frac{1}{n}\sum_{i=1}^n \tilde W(\alpha, X_i),$$

so that (6.4) is satisfied, by (6.12).

(ii) Under $\mathcal{H}_0 : \lambda = 0$, X_i has symmetric distribution around zero with finite moment of order $4r+2$, by Assumption 5.2(i), so that a random variable of the form $|X_i|^{k-1}X_i$ has zero expectation and finite second moment when $1 \le k < 2r+2$. Thus, from Lemma 6.2, for all $k = r+1, \ldots, 2r+1$,

$$E_0[\partial_\lambda^k \ell(0+;X_i)] = 0 \quad \text{and} \quad E_0[(\partial_\lambda^k \ell(0+;X_i))^2] = v_k < \infty,$$

as well as $E_0[\Delta(X_1)] = 0$. From this point, (5.9) and (6.9) together yield

$$\frac{2}{(2r+2)!}E[\partial_\lambda^{2r+2}\ell(0+;X_i)] = -\left(\frac{1}{(r+1)!}\right)^2 E[(\partial_\lambda^{r+1}\ell(0+;X_i))^2]$$
$$= -4\left(\frac{D^r g(0+)}{(r+1)!}\right)^2 E[X_i^{2r+2}] = -V_{r+1}.$$

where Lemma 6.2(ii) was used to set the second equality (notice that this shows that V_{r+1} is the variance of $\partial_\lambda^{r+1}\ell(0+;X_i)/(r+1)!$). Now, (6.5) and (6.6) follow from the central limit theorem and the strong law of large numbers, respectively. □

7. Proof of Theorem 5.1

After the previous preliminaries, Theorem 5.1 is finally established in this section. The core of the argument has been decoupled into two lemmas showing that, under $\mathcal{H}_0 : \lambda = 0$, along a consistent sequence $\{\hat{\lambda}_n\}$ of maximum likelihood estimators, (i) the expansions in Theorem 6.1 can be simplified substantially, and (ii) that $\hat{\lambda}_n$ is no null that with probability converging to 1.

Lemma 7.1. *Suppose that Assumptions 5.1 and 5.2 hold and let $\{\hat{\lambda}_n\}$ be a consistent sequence of maximum likelihood estimators of λ. In this case, assertions (i) and (ii) below occur under the null hypothesis $\mathcal{H}_0 : \lambda = 0$.*

(i) On the event $[\,|\hat{\lambda}_n| < \delta]$, the following expressions are valid:

$$(7.1)\qquad L_n(\hat{\lambda}_n; X_1^n) - L_n(0; X_1^n) = \hat{\lambda}_n|\hat{\lambda}_n|^r \left[\frac{\partial_\lambda^{r+1} L_n(0+; X_1^n)}{(r+1)!} + A_n - \frac{B_n}{2}\hat{\lambda}_n|\hat{\lambda}_n|^r\right],$$

and

$$(7.2)\qquad \frac{\partial_\lambda L_n(\hat{\lambda}_n; X_1^n)}{r+1} = |\hat{\lambda}_n|^r \left[\frac{\partial_\lambda^{r+1} L_n(0+; X_1^n)}{(r+1)!} + \tilde{A}_n - \tilde{B}_n\hat{\lambda}_n|\hat{\lambda}_n|^r\right],$$

where

$$(7.3)\qquad A_n = O_p\left(\frac{\hat{\lambda}_n}{\sqrt{n}}\right), \quad \tilde{A}_n = O_p\left(\frac{\hat{\lambda}_n}{\sqrt{n}}\right),$$

and

$$(7.4)\qquad \lim_{n\to\infty} B_n = \lim_{n\to\infty} \tilde{B}_n = V_{r+1} \quad P_0\text{-}a.s.$$

(ii) Consequently,

$$2n[L_n(\hat{\lambda}_n; X_1^n) - L_n(0; X_1^n)] = 2n\hat{\lambda}_n|\hat{\lambda}_n|^r \left[\frac{\partial_\lambda^{r+1} L_n(0+; X_1^n)}{(r+1)!} + A_n - \frac{B_n}{2}\hat{\lambda}_n|\hat{\lambda}_n|^r\right] + o_p(1).$$

Proof. (i) Setting

$$A_n := \sum_{k=r+2}^{2r+1} \frac{\partial_\lambda^k L_n(0+; X_1^n)}{k!}|\hat{\lambda}_n|^{k-r-1}$$

and

$$B_n := -\left(\frac{2\partial_\lambda^{2r+2} L_n(0+; X_1^n) + 2\Delta_n I_{(-\infty,0)}(\hat{\lambda}_n)}{(2r+2)!} + \frac{2W_n^*(\hat{\lambda}_n)}{(2r+3)!}\hat{\lambda}_n\right),$$

(7.1) is equivalent to (6.2) with $\alpha = \hat{\lambda}_n$. Similarly, defining

$$\tilde{A}_n := \frac{1}{r+1}\sum_{k=r+1}^{2r} \frac{\partial_\lambda^{k+1} L_n(0+; X_1^n)}{k!}|\hat{\lambda}_n|^{k-r}$$

and

$$\tilde{B}_n := -\frac{1}{r+1}\left(\frac{\partial_\lambda^{2r+2}L_n(0+;X_1^n)+\Delta_n I_{(-\infty,0)}(\hat{\lambda}_n)}{(2r+1)!}+\frac{\tilde{W}_n(\hat{\lambda}_n)}{(2r+2)!}\hat{\lambda}_n\right)$$
$$= -\left(\frac{2\partial_\lambda^{2r+2}L_n(0+;X_1^n)+2\Delta_n I_{(-\infty,0)}(\hat{\lambda}_n)}{(2r+2)!}+\frac{\tilde{W}_n(\hat{\lambda}_n)}{(r+1)(2r+2)!}\hat{\lambda}_n\right)$$

it follows that (7.2) is equivalent to (6.3) with $\alpha = \hat{\lambda}_n$. Therefore, since (6.2) and (6.3) are valid for $|\alpha| < \delta$, (7.1) and (7.2) hold on the event $[\,|\hat{\lambda}_n| < \delta]$. To conclude, it will be shown that (7.3) and (7.4) are satisfied. First, notice that A_n and $\tilde{A}_n$ defined above are null for $r = 0$, so that (7.3) certainly occurs in this case. On the other hand, if $r > 0$, then convergences (6.5) established in Theorem 6.1(ii) yield that $\partial_\lambda^k L_n(0+;X_1^n) = O_p(1/\sqrt{n})$ for $r+1 \le k < 2r+1$ and then (7.3) follows, since the above expressions for A_n and $\tilde{A}_n$ involve factors $|\hat{\lambda}_n|^s$ with $s \ge 1$ and

$$P_0[\hat{\lambda}_n \to 0] = 1, \tag{7.7}$$

by consistency. Next, observe that

$$|W_n^*(\hat{\lambda}_n)|,\ |\tilde{W}_n(\hat{\lambda}_n)| \le W_n = \frac{1}{n}\sum_{i=1}^n W(X_i),$$

by (6.4), and then the strong law of large numbers and Assumption 5.2(ii) yield that

$$\limsup_{n\to\infty}|W_n^*(\hat{\lambda}_n)|,\ \limsup_{n\to\infty}|\tilde{W}_n(\hat{\lambda}_n)| \le \int_{\mathbb{R}} W(x)f(x),\, dx < \infty \quad P_0\text{-a.s.},$$

so that

$$\lim_{n\to\infty} W_n^*(\hat{\lambda}_n)\hat{\lambda}_n = 0 = \lim_{n\to\infty}\tilde{W}_n(\hat{\lambda}_n)\hat{\lambda}_n \quad P_0\text{-a.s.},$$

by (7.7). From this point, (6.6) in Theorem 6.1(ii) and the specifications of B_n and $\tilde{B}_n$ lead to (7.4).

(ii) Since expansion (7.1) is valid on $[\,|\hat{\lambda}_n| < \delta]$, the conclusion follows from (7.7). □

Lemma 7.2. *Suppose that Assumptions 5.1 and 5.2 are valid, let $\{\hat{\lambda}_n\}$ be a sequence of maximum likelihood estimators of λ, and define*

$$\Omega_n^{**} := \left[L_n(\hat{\lambda}_n;X_1^n) \ge L_n(\lambda;X_1^n),\ \lambda\in\mathbb{R}\right] \cap \left[\partial_\lambda^{r+1}L_n(0+;X_1^n) \ne 0\right].$$

With this notation, assertions (i) and (ii) below occur.

(i) $\hat{\lambda}_n \ne 0$ on Ω_n^{**}.
Consequently,
(ii) $P_{\mathbf{0}}[\hat{\lambda}_n \ne 0] \to 1$ as $n \to \infty$.

Proof. (i) The expansion for $L_n(\cdot;X_1^n) - L_n(0;X_1^n)$ in Theorem 6.1(i) yields that

$$\lim_{\alpha\to 0}\frac{L_n(\alpha;X_1^n)-L_n(0;X_1^n)}{|\alpha|^r\alpha} = \partial_\lambda^{r+1}L_n(0+;X_1^n).$$

It follows that if $\partial_\lambda^{r+1} L_n(0+; X_1^n) > 0$, then $L_n(\alpha; X_1^n) - L_n(0; X_1^n) > 0$ when α is positive and small enough, whereas if $\partial_\lambda^{r+1} L_n(0+; X_1^n) < 0$, then $L_n(\alpha; X_1^n) - L_n(0; X_1^n) > 0$ when $\alpha < 0$ and $|\alpha|$ is sufficiently small. Thus, $\partial_\lambda^{r+1} L_n(0+; X_1^n) \neq 0$ implies that 0 is not a maximizer of $L_n(\cdot; X_1^n)$, so that, if $\hat\lambda_n$ maximizes the average likelihood $L_n(\cdot; X_1^n)$ and $\partial_\lambda^{r+1} L_n(0+; X_1^n) \neq 0$ then $\hat\lambda_n$ is no null, *i.e.*, $\Omega_n^{**} \subset [\hat\lambda_n \neq 0]$.

(ii) Since $D^r g(0+) \neq 0$, it follows that $\partial_\lambda^{r+1} \ell(0+; X_i) = 2D^r g(0+)|X_i|^r X_i$ has a density, and then their average $\partial_\lambda^{r+1} L_n(0+; X_1^n)$ is absolutely continuous. It follows that $P_0[\partial_\lambda^{r+1} L_n(0+; X_1^n) \neq 0] = 1$, and then $P_0[\Omega_n^{**}] \to 1$, by Definition 3.1 (see Remark 3.1(ii)) and, via part (i), the conclusion follows. □

Proof of Theorem 5.1. Suppose that Assumptions 5.1 and 5.2 hold, that the hypothesis $\mathcal{H}_0 : \lambda = 0$ occurs, and let $\{\hat\lambda_n\}$ be a consistent sequence of maximum likelihood estimators. In this context, define

$$\Omega_{n,*} := \big[L_n(\hat\lambda_n; X_1^n) \geq L_n(\lambda; X_1^n),\ \lambda \in \mathbb{R}\big] \cap \big[\, 0 < |\hat\lambda_n| < \delta\big],$$

and notice that the conclusions in Lemma 7.1 occur on this event, since $\Omega_{n,*} \subset [\,|\hat\lambda| < \delta]$. Also, the consistency of $\{\hat\lambda_n\}$, Definition 3.1 and Lemma 7.2(ii) together imply that

$$\lim_{n\to\infty} P_0[\Omega_{n,*}] = 1. \tag{7.8}$$

Observe now that, on the event $\Omega_{n,*}$, the estimator $\hat\lambda_n$ is no null and maximizes $L_n(\cdot; X_1^n)$, so that the likelihood equation $\partial_\lambda L_n(\hat\lambda_n; X_1^n) = 0$ holds; see Remark 5.1(ii). Via (7.2) it follows that

$$\text{on } \Omega_{n,*}, \quad \tilde B_n \sqrt{n}\hat\lambda_n |\hat\lambda_n|^r = \sqrt{n}\frac{\partial_\lambda^{r+1} L_n(0+; X_1^n)}{(r+1)!} + \sqrt{n}\tilde A_n,$$

and then, from (7.8)

$$\begin{aligned}
\tilde B_n \sqrt{n}\hat\lambda_n |\hat\lambda_n|^r &= \sqrt{n}\frac{\partial_\lambda^{r+1} L_n(0+; X_1^n)}{(r+1)!} + \sqrt{n}\tilde A_n + o_p(1)\\
&= \sqrt{n}\frac{\partial_\lambda^{r+1} L_n(0+; X_1^n)}{(r+1)!} + O_p(\hat\lambda_n) + o_p(1)\\
&= \sqrt{n}\frac{\partial_\lambda^{r+1} L_n(0+; X_1^n)}{(r+1)!} + o_p(1),
\end{aligned}$$

where (7.3) was used to set the second equality, and the third one stems from $P_0[\hat\lambda_n \to 0] = 1$, by consistency; via (6.5) this yields that $\tilde B_n \sqrt{n}\hat\lambda_n |\hat\lambda_n|^r = O_p(1)$, and then (7.4) leads to

$$V_{r+1}\sqrt{n}\ \hat\lambda_n |\hat\lambda_n|^r = \sqrt{n}\frac{\partial_\lambda^{r+1} L_n(0+; X_1^n)}{(r+1)!} + o_p(1). \tag{7.9}$$

(i) Using that $\sqrt{n}\dfrac{\partial_\lambda^{r+1} L_n(0+; X_1^n)}{(r+1)!} \xrightarrow{\mathrm{d}} \mathcal{N}(0, V_{r+1})$ (see (5.9) and (6.5)), the above display yields

$$\sqrt{nV_{r+1}}\ \hat\lambda_n |\hat\lambda_n|^r \xrightarrow{\mathrm{d}} Z \quad \text{where } Z \text{ has standard normal distribution;} \tag{7.10}$$

since the inverse of the function $x \mapsto x|x|^r$ is the continuous mapping $x \mapsto |x|^{1/(r+1)}$ sign(x), it follows that

$$(nV_{r+1})^{1/(2(r+1))}\hat{\lambda}_n \xrightarrow{\mathrm{d}} |Z|^{1/(r+1)}\mathrm{sign}(Z).$$

(ii) Since $\sqrt{n}\hat{\lambda}_n|\hat{\lambda}_n|^r = O_p(1)$, by part (i), Lemma 7.1(ii), (7.3), (7.4) and (7.9) together yield

$$\begin{aligned}
&2n[L_n(\hat{\lambda}_n; X_1^n) - L_n(0; X_1^n)] \\
&\quad = 2n\hat{\lambda}_n|\hat{\lambda}_n|^r \left[\frac{\partial_\lambda^{r+1} L_n(0+; X_1^n)}{(r+1)!} + A_n - \frac{B_n}{2}\hat{\lambda}_n|\hat{\lambda}_n|^r\right] + o_p(1) \\
&\quad = 2\sqrt{n}\hat{\lambda}_n|\hat{\lambda}_n|^r \left[\sqrt{n}\frac{\partial_\lambda^{r+1} L_n(0+; X_1^n)}{(r+1)!} + \sqrt{n}A_n - \frac{B_n}{2}\sqrt{n}\hat{\lambda}_n|\hat{\lambda}_n|^r\right] + o_p(1) \\
&\quad = 2\sqrt{n}\hat{\lambda}_n|\hat{\lambda}_n|^r \left[\sqrt{n}\frac{\partial_\lambda^{r+1} L_n(0+; X_1^n)}{(r+1)!} + O_p(\hat{\lambda}_n)\right. \\
&\qquad\qquad\qquad \left. - \frac{V_{r+1} + o_p(1)}{2}\sqrt{n}\hat{\lambda}_n|\hat{\lambda}_n|^r\right] + o_p(1) \\
&\quad = 2\sqrt{n}\hat{\lambda}_n|\hat{\lambda}_n|^r \left[\sqrt{n}\frac{\partial_\lambda^{r+1} L_n(0+; X_1^n)}{(r+1)!} - \frac{V_{r+1}}{2}\sqrt{n}\hat{\lambda}_n|\hat{\lambda}_n|^r\right] + o_p(1) \\
&\quad = 2\sqrt{n}\hat{\lambda}_n|\hat{\lambda}_n|^r \left[V_{r+1}\sqrt{n}\hat{\lambda}_n|\hat{\lambda}_n - \frac{V_{r+1}}{2}\sqrt{n}\hat{\lambda}_n|\hat{\lambda}_n|^r\right] + o_p(1) \\
&\quad = \left(\sqrt{nV_{r+1}}\ \hat{\lambda}_n|\hat{\lambda}_n|^r\right)^2 + o_p(1);
\end{aligned}$$

together with (7.10), this yields that $2n[L_n(\hat{\lambda}_n; X_1^n) - L_n(0; X_1^n)] \xrightarrow{\mathrm{d}} Z^2$, completing the proof. □

References

[1] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.* **12** 171–178.

[2] Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistica* **46** 199–208.

[3] Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *J. Roy. Stat. Soc. Ser. B* **61** 579–602.

[4] Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika* **83** 715–726.

[5] Azzalini, A. (2005). The skew-normal distribution and related multivariate families (with discussion). *Scand. J. Statist.* **32** 159–188 (CR: 189–200).

[6] Billingsley, P. (1995). *Probability and Measure*, 3rd ed. Wiley, New York.

[7] Cavazos-Cadena, R. and González-Farías, G. (2007). Necessary and sufficient conditions for the consistency of maximum likelihood estimators. Submitted.

[8] Chiogna, M. (2005). A note on the asymptotic distribution of the maximum likelihood estimator for the scalar skew-normal distribution. *Stat. Meth. Appl.* **14** 331–341.

[9] DiCiccio, T. and Monti, A. C. (2004). Inferential aspects of the skew exponential power distribution. *J. Amer. Statist. Assoc.* **99** 439–450.

[10] Genton, M. G., ed. (2004). *Skew Elliptical Distributions and Their Applications.* Chapman & Hall, London.

[11] Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*, 2nd ed. Springer, New York.

[12] Newey, W. and McFadden, D. (1993). Estimation in large samples. In *Handbook of Econometrics*, Vol. 4. (D. McFadden and R. Engler, eds.). North-Holland, Amsterdam.

[13] Pewsey, A. (2000). Problems of inference for Azzalini's skew-normal distribution. *J. Appl. Statist.* **27** 859–870.

[14] Rotnitzky, A., Cox, D. R., Bottai, M. and Robins, J. (2000). Likelihood-based inference with singular information matrix. *Bernoulli* **6** 243–284.

[15] Sartori, N. (2006). Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions. *J. Statist. Plann. Inference* **136** 4259–4275.

[16] Shao, J. (1999). *Mathematical Statistics*. Springer, New York.