Parametric Maximum Likelihood Estimators and Resampling

Yuri K. Belyaev\textsuperscript{1} and Leif Nilsson\textsuperscript{1}

Department of Mathematical Statistics
Umeå University
S-901 87 Umeå, Sweden

Abstract

We considered statistical data consisting of observations of independent but not necessarily identically distributed random variables. The distributions of the random variables are allowed to belong to different parametric families with the same parametric set. The aim of this paper is to justify that resampling methods can be used to estimate the distribution laws of the deviations of maximum likelihood estimators, from true unknown parameters. The ordinary definition of maximum likelihood estimators is slightly generalized and a similar definition for maximum likelihood estimators is made based on resampling. Under the same rather general assumptions, existence and consistency of these two types of estimators of the true unknown parameters are proved. The main result is that the distribution laws of the deviations of maximum likelihood estimators from the true unknown parameters, can be estimated by the resampling method so that the estimates are asymptotically accurate. To avoid assuming the existence of limiting distribution laws, the notion of weakly approaching sequences is used.

Key words: maximum likelihood estimators, resampling, bootstrap, consistency, weakly approaching sequences of distribution laws, central limit resampling theorem.


\textsuperscript{1}Research was supported by The Bank of Sweden Tercentenary Foundation.
1 Introduction

Most of the theory related to asymptotic properties of point estimators of unknown parameters concerns statistical data which are observations of independent and identically distributed (i.i.d.) random variables (r.v.'s). In many applications, there is a great need to analyze statistical data which contain values of independent but not identically distributed (i.n.i.d.) r.v.'s. For example, reliability and biomedical data often contain, together with observations of r.v.'s (e.g. life times), so-called explanatory variables (covariates) which influence the distribution laws (d.l.'s) of observed r.v.'s. Statistical models describing data with explanatory variables usually include d.l.'s of many unknown parameters (e.g. loading coefficients in reliability theory) which have to be estimated by using observed r.v.'s. In some cases the distribution laws can be different because of differently planned experiments, and still have the same parametric set; e.g. when the r.v.'s have the same true parameters and different types of censorings are used (see e.g. Qiying [19]). Point estimators obtained for some vector valued parameters under interest can be used depending on their performance. The performance of the point estimators can be measured using the d.l.'s of the deviations of the estimators from the true unknown parameters, or by related characteristics such as bias and variance. The d.l.'s of deviations are unknown and have to be estimated using the same statistical data as used for point estimations. Estimating the d.l.'s of deviations is a more difficult problem than point estimation of the parameters, especially in the case when the data are observations of i.n.i.d. r.v.'s.

In this paper, we consider problems related to the estimation of the performance of maximum likelihood estimators (ML-estimators) of the parameters under interest, in the case where the data are observations of i.n.i.d. r.v.'s. The statement of the problem is asymptotic; that is we assume that the number of observations $n$ is unbounded and increasing ($n \to \infty$). Assuming that certain regularity conditions are fulfilled, ML-estimators possess asymptotic efficiency in contrast to many other point estimators. Therefore, ML-estimators are used widely in the analysis of statistical data. The case of i.i.d. observations has been very well established, and nowadays the asymptotic theory of ML-estimators is included in many graduate text books in mathematical statistics; see e.g. Lehmann [16] and Ferguson [11]. There are also some results related to the case of i.n.i.d. observations; see e.g. Bradley and Gart [6], Hoadley [13], Philippou and Roussas [18] and Weiss [23], [24]. These results address only the question of consistency of the ML-estimators, and the asymptotic normality of the deviations of ML-estimators from the true unknown parameters. In Reeds [20], the Jackknife method was used in the i.i.d. case to estimate the d.l.'s of the deviations of ML-estimators from the true unknown parameters. Another promising method for estimating the d.l.'s of deviations is the method of resampling. Burke and Gom-
bay [7] obtained consistent estimators of d.l.’s of deviations using the method of
resampling in the i.i.d. case. We do not know of any results related to obtaining
asymptotically accurate estimators of the d.l.’s of deviations of the ML-estimators
from the true unknown parameters in the i.n.i.d. case. Our aim in this report is
to use the resampling method to obtain asymptotically accurate estimators for
these d.l.’s of deviations in the i.n.i.d. case. From the point of view of researchers
in statistical data analysis, the method of resampling is very attractive because
it is universal and most of the numerous calculations inherited with it can easily
be made by computers. Computer intensive methods in statistical analysis are of
increasing popularity and they are thoroughly described in many books devoted
to the bootstrap method (see e.g. Efron and Tibshirani [10] and Hjort [12]).
To avoid restricting ourselves to estimators having a fixed limit distribution,
a generalization of the ordinary weak convergence introduced by Belyaev [2], and
generalized in Belyaev and Sj"oestedt [4], will be used.

The paper is organized as follows: In Section 2 and 3, we prove the exis-
tence and the consistency of ML-estimators and resampled ML-estimators (RML-
estimators) under the same assumptions. These assumptions are slightly different
from those made by Hoadley [13]. We can avoid assuming the existence of mo-
moments of order higher than the first of the log likelihood ratio. In our approach to
obtaining asymptotically accurate estimators of d.l.’s of deviations, it is essential
that consistent ML- and RML-estimators exist, simultaneously. We prove that
under some assumptions using the resampling method, the d.l.’s of deviations of
ML-estimators from the true unknown parameters can be estimated asymptoti-
cally accurate in the sense of weakly approaching sequences. These assumptions
are close to the regularity conditions figured in text books of mathematical statis-
tics. The notion of weakly approaching sequences is restated in Section 4.

The following notations are used here. $\mathbb{R}^g$ and $\mathbb{R}^{g \times h}$ denote the Euclidian
g-dimensional space and the space of all real valued matrices with $g$ rows and
$h$ columns, respectively. Bold fonts are used for column vectors, and contour
fonts such as $\mathcal{A}$ are used for matrices (or sets of vectors). A column vector
$\mathbf{A} = (A_1 A_2 \cdots A_g)^T \in \mathbb{R}^g$ is a real valued $g$-dimensional vector, and a real valued
matrix $\mathcal{A} = (A_1 A_2 \cdots A_h) \in \mathbb{R}^{g \times h}$ is a matrix of $g$ rows and $h$ columns containing
the elements $A_{ij}$, $i = 1, 2, \ldots, g$, $j = 1, 2, \ldots, h$, where $T$ denotes transposition.
In particular, $\mathbf{0}_g = (0, 0, \ldots, 0)^T \in \mathbb{R}^g$. The Frobenius norm of a matrix $\mathcal{A} \in \mathbb{R}^{g \times h}$;
that is $\sqrt{\sum_{i=1}^g \sum_{j=1}^h A_{ij}^2}$, is denoted by $\| \mathcal{A} \|_F$ and the determinant by $\text{det}(\mathcal{A})$. The
notation $\| \cdot \|_2$ is also used for the Euclidian norm of a vector. $a \vee b$ and $a \wedge b$
stand for $\max(a, b)$ and $\min(a, b)$, respectively, and $a := \text{expression}$ means that
$a$ is defined by $\text{expression}$. An open ball with center at $\mathbf{a}$ and radius $\rho$ in a
set $\mathcal{A}$ is denoted by $\mathcal{B}_\rho(\mathbf{a})$; that is $\mathcal{B}_\rho(\mathbf{a}) = \{ \mathbf{a}' : \| \mathbf{a}' - \mathbf{a} \|_2 < \rho, \mathbf{a}' \in \mathcal{A} \}$. The
Borel $\sigma$-algebra generated by all open balls in a set $\mathcal{A}$ is denoted with $\mathcal{B}(\mathcal{A})$.
The interior of a set $\mathcal{A}$; that is the union over all open sets belonging to $\mathcal{A}$ is
denoted by $\mathcal{A}$. The closure of a set $\mathcal{A}$; that is the intersection over all closed sets
2 Consistency for maximum likelihood estimators

Let us consider the following scheme of series of r.v.'s

\[ \mathcal{Y}_n = \{ Y_{1n}, Y_{2n}, \ldots, Y_{nn} \}, \quad n = 1, 2, \ldots \]

Each r.v. \( Y_{rn} \) takes values in some measurable space \( (\mathcal{Y}_{rn}, \mathcal{F}(\mathcal{Y}_{rn})) \), where \( \mathcal{F}(\mathcal{Y}_{rn}) \) is a \( \sigma \)-algebra of events, \( (r, n) \in \mathcal{I} := \{(r', n': 1 \leq r' \leq n', n' = 1, 2, \ldots \} \). The r.v.'s \( Y_{1n}, Y_{2n}, \ldots, Y_{nn} \) are assumed to be independent for each \( n \). Usually each value of \( Y_{rn} \) consists of responses and explanatory variables; that is some of its elements can be nonrandom and they can influence the d.l. of \( Y_{rn} \). Let

\[ P_{rn} := \{ P_{\theta_{rn}}, \theta \in \Theta \}, \quad (r, n) \in \mathcal{I}, \]

be families of d.l.'s, where \( \Theta \) is an open subset of \( \mathbb{R}^k \). Each d.l. \( P_{\theta_{rn}} \) is defined on \( \mathcal{F}(\mathcal{Y}_{rn}) \), and \( P_{\theta_{rn}} \neq P_{\theta'_{rn}} \) if \( \theta 
eq \theta' \), \( (r, n) \in \mathcal{I} \). Assume that for each \( \theta \in \Theta \) there exist probability density functions

\[ p_{rn}(y_{rn}, \theta) := \frac{dP_{\theta_{rn}}(y_{rn})}{d\mu_{rn}}, \quad y_{rn} \in \mathcal{Y}_{rn}, \]

with respect to some \( \sigma \)-finite measure \( \mu_{rn} \) on \( \mathcal{F}(\mathcal{Y}_{rn}) \). Furthermore, assume that the set \( \{ y_{rn} : p_{rn}(y_{rn}, \theta) > 0 \} \subset \mathcal{Y}_{rn} \) does not depend on \( \theta \in \Theta \), and that each r.v. \( Y_{rn} \) has the d.l. \( P_{\theta_0_{rn}} \), where \( \theta_0 \in \Theta \) will be called the true parameter.

The outcome \( y_{rn} \), of the r.v. \( Y_{rn} \), can be interpreted as the result of the \( r \)th experiment in a series of \( n \) experiments.

Henceforth, all statistical data are considered as random and the capital letters \( Y_{1n}, Y_{2n}, \ldots, Y_{nn} \) will be used. Moreover, all variables obtained by using the operators inf and sup are assumed to be r.v.'s.

One of the mostly frequently used approaches to estimating \( \theta_0 \) from a random data set \( \mathcal{Y}_n \), is the maximum likelihood method. The maximum likelihood
estimator is obtained as the $\theta$-value which maximizes the loglikelihood function

$$l_n(\theta, Y_n) := \sum_{r=1}^{n} \log(p_r(Y_{rn}, \theta)).$$

The loglikelihood function of the $r$th experiment will be denoted

$$l_r(\theta, Y_r) := \log(p_r(Y_{rn}, \theta)),$$

and it can be considered as a random process with “time parameter” $\theta \in \Theta$.

A more formal definition of the maximum likelihood estimators follows.

**Definition 1** A statistic $\hat{\theta}_n(\mathcal{A}) : Y_{1n} \times Y_{2n} \times \cdots \times Y_{nn} \to \Theta$, where $\mathcal{A} \in \mathfrak{A}(\Theta)$ is called a maximum likelihood estimator on $\mathcal{A}$ (ML($\mathcal{A}$)-estimator) of the true parameter $\theta_0$, based on observations $Y_n$, if and only if

$$\prod_{r=1}^{n} p_r(Y_{rn}, \hat{\theta}_n(\mathcal{A})) := \sup_{\theta \in \mathcal{A}} \prod_{r=1}^{n} p_r(Y_{rn}, \theta).$$

The statistic $\hat{\theta}_n = \hat{\theta}_n(\mathcal{A})$ is called the ML-estimator of $\theta_0$ if there exist a sequence of open sets $\{B_k\}_{k \geq 1}$, such that $B_1 \subset B_2 \subset \cdots$, $\bigcup_{j \geq 1} B_j = \Theta$, and for any two sufficiently large $k_1$ and $k_2$, $\hat{\theta}_n(B_{k_1}) = \hat{\theta}_n(B_{k_2})$ for all $n \geq n(k_1, k_2, \theta_0)$.

This definition is a slight generalization of the ordinary ML-estimators, defined as the $\hat{\theta}(\mathcal{A})$-value satisfying (1) with $\mathcal{A} = \Theta$. Mixtures of density functions similar to Example 5.6, page 442 in Lehmann [16], are examples of situations where the ML-estimators exist, while the ordinary ML-estimators do not exist.

Before giving the theory related to consistency of maximum likelihood estimators, some definitions are needed. A function which plays a central role in our proof of consistency of the maximum likelihood estimators is the loglikelihood ratio

$$R_{rn}(\theta_0, \theta) := \log\left(\frac{p_r(Y_{rn}, \theta_0)}{p_r(Y_{rn}, \theta)}\right) = l_r(\theta_0, Y_{rn}) - l_r(\theta, Y_{rn}), \quad (r, n) \in \mathcal{I}.$$

For a set $\mathcal{B} \in \mathfrak{A}(\Theta)$, let

$$R_{rn}(\theta_0, \mathcal{B}) := \inf_{\theta \in \mathcal{B}} R_{rn}(\theta_0, \theta).$$

The existence and the consistency of maximum likelihood estimators shall be proved under the following assumptions:

**A1($\rho_1$):** For each $y_{rn} \in Y_{rn}$, $p_r(y_{rn}, \cdot)$ is a continuous function on $\Theta$, and for each $\theta_0, \theta \in \Theta$, where $\theta \neq \theta_0$, there exist two finite values $b(\theta_0, \theta) > 0$ and $\rho_1 = \rho_1(\theta_0, \theta) > 0$ such that

$$\liminf_{n} \min_{1 \leq r \leq n} \mathbb{E}_{\theta_0}[R_{rn}(\theta_0, \mathcal{B}_{\rho_1}(\theta))] > b(\theta_0, \theta).$$
Remark Instead of assuming continuity of \( p_{r_n}(y_{r_n}, \cdot) \) it is possible to assume semicontinuity from above; that is for any \( \theta \in \Theta \),
\[
\sup_{\theta' \in B_i(\theta)} p_{r_n}(y_{r_n}, \theta') \downarrow p_{r_n}(y_{r_n}, \theta) \quad \text{as} \quad \epsilon \to 0.
\]

Remark The function \( R_{r_n}(\theta_0, \cdot) \) possesses the property
\[
\mathbb{E}_{\theta_0}[R_{r_n}(\theta_0, \cdot)] > 0 \quad \text{if} \quad \theta \neq \theta_0
\]
(see e.g. Ferguson [11], page 113), which can be used to check assumption A1(\( \rho_1 \)).

A2(\( \rho_2 \)): For each \( \theta_0, \theta \in \Theta \) and any arbitrarily small \( \epsilon > 0 \), there exist positive \( c_2(\epsilon) = c_2(\epsilon, \theta_0, \theta) \) and \( \rho_2 = \rho_2(\epsilon, \theta_0, \theta) \) such that
\[
\limsup_n \max_{1 \leq r \leq n} \mathbb{E}_{\theta_0}[|R_{r_n}(\theta_0, \theta)|] I(|R_{r_n}(\theta_0, \theta)| > c_2(\epsilon)) < \epsilon, \quad (2)
\]
and
\[
\limsup_n \max_{1 \leq r \leq n} \mathbb{E}_{\theta_0}[|R_{r_n}(\theta_0, \theta)|^2] I(|R_{r_n}(\theta_0, \theta)| > c_2(\epsilon)) < \epsilon. \quad (3)
\]

For the proof of strong consistency of ML-estimators, the following alternative assumption of A2(\( \rho_2 \)) is needed:

A3(\( \gamma, \rho_3 \)): For each \( \theta_0, \theta \in \Theta \) and some \( \gamma > 1 \), there exist positive \( c_3(\gamma, \theta_0, \theta) \) and \( \rho_3 = \rho_3(\gamma, \theta_0, \theta) \) such that
\[
\limsup_n \max_{1 \leq r \leq n} \mathbb{E}_{\theta_0}[|R_{r_n}(\theta_0, \theta)|^\gamma] < c_3(\gamma, \theta_0, \theta),
\]
and
\[
\limsup_n \max_{1 \leq r \leq n} \mathbb{E}_{\theta_0}[|R_{r_n}(\theta_0, \theta)|^\gamma] < c_3(\gamma, \theta_0, \theta).
\]

Remark It is shown in the Appendix that assumption A3(\( \gamma, \rho_3 \)) with \( \gamma > 1 \) implies assumption A2(\( \rho_2 \)).

Theorem 1 Suppose that assumptions A1(\( \rho_1 \)) and A2(\( \rho_2 \)) hold with \( \rho_1 > 0 \) and \( \rho_2 > 0 \). Then, for any compact set \( \mathcal{K} \in \mathfrak{B}(\Theta) \) such that \( \theta_0 \in \mathcal{K} \), there exists with probability tending to one, at least one ML(\( \mathcal{K} \))-estimator \( \hat{\theta}_n(\mathcal{K}) \), and all such ML(\( \mathcal{K} \))-estimators converge in probability to \( \theta_0 \) as \( n \to \infty \).

In the case when assumptions A1(\( \rho_1 \)) and A3(\( \gamma, \rho_3 \)) with \( \gamma > 2 \) hold, the following statement is valid for ML(\( \mathcal{K} \))-estimators.
Theorem 2  Suppose that assumptions A1($\rho_1$) and A3($\gamma, \rho_3$) hold with $\rho_1 > 0$, $\rho_3 > 0$, and $\gamma > 2$. Then, for any compact set $\mathcal{K} \in \mathfrak{B}(\Theta)$ such that $\theta_0 \in \mathcal{K}$, there exists with probability one for all sufficiently large $n$, at least one ML($\mathcal{K}$)-estimator $\hat{\theta}_n(\mathcal{K})$, and all such ML($\mathcal{K}$)-estimators converge almost surely to $\theta_0$ as $n \to \infty$.

Proofs of Theorem 1 and Theorem 2 are given in the appendix.

If there exists only one ML-estimate $\hat{\theta}_n(\Theta)$, then one can use the following corollaries.

Corollary 1 Assume that assumptions A1($\rho_1$) and A2($\rho_2$) hold with $\rho_1 > 0$, $\rho_2 > 0$, and that with probability tending to one there exists only one local maximum $\hat{\theta}_n$ of $\prod_{r=1}^{n} p_{r}(Y_{rn}, \cdot)$. Then, $\hat{\theta}_n = \hat{\theta}_n(\Theta)$ converges in probability to $\theta_0$ as $n \to \infty$.

Corollary 2 Assume that assumptions A1($\rho_1$) and A3($\gamma, \rho_3$) hold with $\rho_1 > 0$, $\rho_3 > 0$, $\gamma > 2$, and that with probability one there exists for each sufficiently large $n$ only one point $\hat{\theta}_n$ which is a local maximum of $\prod_{r=1}^{n} p_{r}(Y_{rn}, \cdot)$. Then, $\hat{\theta}_n = \hat{\theta}_n(\Theta)$ converges almost surely to $\theta_0$ as $n \to \infty$.

3 Consistency for resampled maximum likelihood estimators

We define an analogue of the ML-estimator which is based on resampled copies of the data $\mathcal{Y}_n$. Let $\mathcal{J}_n^* := \{J_{n1}^*, J_{n2}^*, \ldots, J_{nn}^*\}$ be $n$ i.i.d. discrete uniform r.v.'s. with sample space $\{1, 2, \ldots, n\}$.

These r.v.'s are independent of the data $\mathcal{Y}_n$ and can be obtained for example by independent simulations for each $n = 1, 2, \ldots$. The resampled copy of the data $\mathcal{Y}_n$ is then the r.v.

$$\mathcal{Y}_n^* := \{Y_{J_{n1}^*}, Y_{J_{n2}^*}, \ldots, Y_{J_{nn}^*}\},$$

possessing two types of randomness, namely, the original randomness of $\mathcal{Y}_n$ and an additional randomness induced by using (simulated) values of the r.v. $\mathcal{J}_n^*$.

The mathematical expectations and probabilities related only to the original randomness of $\mathcal{Y}_n$ are denoted as usual by $E_{\theta_0n}[]$ and $P_{\theta_0n}(\cdot)$. Those related only to $\mathcal{J}_n^*$ are denoted by $E_{\mathcal{J}_n^*}[]$ and $P_{\mathcal{J}_n^*}(\cdot)$. The mathematical expectations and probabilities related to both $\mathcal{Y}_n$ and $\mathcal{J}_n^*$, are denoted by $E_{\theta_0n}^{(\cdot)}[]$ and $P_{\theta_0n}^{(\cdot)}(\cdot)$.

Formally, $E_{\mathcal{Y}_n}[] = E_{\theta_0n}^{(\cdot)}[\cdot | \mathcal{J}_n^*]$, $P_{\mathcal{Y}_n}(\cdot) = P_{\theta_0n}^{(\cdot)}(\cdot | \mathcal{J}_n^*)$, $E_{\mathcal{Y}_n}[] = E_{\theta_0n}[\cdot | \mathcal{Y}_n]$, and $P_{\mathcal{Y}_n}(\cdot) = P_{\theta_0n}(\cdot | \mathcal{Y}_n)$. A function of both $\mathcal{Y}_n$ and $\mathcal{J}_n^*$, converging in probability (almost surely) with respect to both $\mathcal{Y}_n$ and $\mathcal{J}_n^*$, is said to be converging in $P_{\theta_0}^{(\cdot)}$-probability ($P_{\theta_0}^{(\cdot)}$-almost surely). In the same manner, $o_{\mathcal{P}}(\cdot)$ and $O_{\mathcal{P}}(\cdot)$ which
are related to both $\mathbb{Y}_n$ and $\mathbf{J}_n^*$ will be denoted $\sigma_p(\cdot)$ and $O_p(\cdot)$, respectively. Henceforth, we shall drop the index $n$ of expectations and probabilities.

Let us introduce the following r.v.'s similar to the loglikelihood function and the loglikelihood ratio, introduced in Section 2:

$$I_n^*(\theta, \mathbf{Y}_{rn}) := \sum_{r=1}^n M_{rn}^* l_{rn}(\theta, \mathbf{Y}_{rn}),$$

$$R_n^*(\theta_0, \theta) := \sum_{r=1}^n M_{rn}^* R_{rn}(\theta_0, \theta).$$

where $M_{rn}^* := \sum_{i=1}^n I(J_{rn}^* = r)$ and $\mathcal{B} \in \mathfrak{B}(\Theta)$. The likelihood function of the resampled data $\mathbb{Y}_n^*$ is given by

$$\prod_{r=1}^n p_{rn}(\mathbf{Y}_{J_{rn}^*}, \theta) = \prod_{r=1}^n (p_{rn}(\mathbf{Y}_{rn}, \theta))^{M_{rn}^*}, \quad \theta \in \Theta.$$ 

Formally, we can apply the maximum likelihood principle to obtain an estimator for $\theta_0$, by using this artificial likelihood. Let us introduce the following definition.

**Definition 2** A statistic $\hat{\theta}_n^*(\mathcal{A}) : \mathcal{Y}_{J_{rn}^*} \times \mathcal{Y}_{J_{rn}^*} \times \cdots \times \mathcal{Y}_{J_{rn}^*} \to \Theta$, where $\mathcal{A} \in \mathfrak{B}(\Theta)$, is called a resampled maximum likelihood estimator for $\theta_0$ on $\mathcal{A}$ (RML($\mathcal{A}$)-estimator) based on the resampling copy $\mathbb{Y}_n^*$ of the data $\mathbb{Y}_n$ if and only if

$$\prod_{r=1}^n \left( p_{rn}(\mathbf{Y}_{rn}, \hat{\theta}_n^*(\mathcal{A})) \right)^{M_{rn}^*} = \sup_{\theta \in \mathcal{A}} \prod_{r=1}^n \left( p_{rn}(\mathbf{Y}_{rn}, \theta) \right)^{M_{rn}^*}. \quad (4)$$

The statistic $\hat{\theta}_n^* = \hat{\theta}_n^*(\Theta)$ is called the RML-estimator of $\theta_0$ based on the resampling copy $\mathbb{Y}_n^*$ if there exist a sequence of open sets $\{\mathcal{B}_j\}_{j \geq 1}$ such that, $\mathcal{B}_1 \subset \mathcal{B}_2 \subset \cdots$, $\bigcup_{j \geq 1} \mathcal{B}_j = \Theta$, and for any two sufficiently large $k_1$ and $k_2$, $\hat{\theta}_n^*(\mathcal{B}_{k_1}) = \hat{\theta}_n^*(\mathcal{B}_{k_2})$ for all $n \geq n(k_1, k_2, \theta_0)$.

The next theorems give sufficient conditions which guarantee existence and consistency of RML($\mathcal{A}$)-estimators.

**Theorem 3** Suppose that assumptions $A1(\rho_1)$ and $A2(\rho_2)$ hold with $\rho_1 > 0$ and $\rho_2 > 0$. Then, for any compact set $\mathcal{K} \in \mathfrak{B}(\Theta)$, such that $\theta_0 \in \mathcal{K}$, there exists with $P_{\theta_0}^{(\ast)}$-probability tending to one, at least one RML($\mathcal{K}$)-estimator $\hat{\theta}_n^*(\mathcal{K})$, and all such estimators converge in $P_{\theta_0}^{(\ast)}$-probability to $\theta_0$ as $n \to \infty$.

**Theorem 4** Suppose that assumptions $A1(\rho_1)$ and $A3(\gamma, \rho_3)$ hold with $\rho_1 > 0$, $\rho_3 > 0$, and $\gamma > 2$. Then, for any compact set $\mathcal{K} \in \mathfrak{B}(\Theta)$, such that $\theta_0 \in \mathcal{K}$, there exists with $P_{\theta_0}^{(\ast)}$-probability one for all sufficiently large $n$, at least one RML($\mathcal{K}$)-estimator $\hat{\theta}_n^*(\mathcal{K})$, and all such RML($\mathcal{K}$)-estimators converge $P_{\theta_0}^{(\ast)}$-almost surely to $\theta_0$ as $n \to \infty$. 

7
4 Weakly approaching d.l.’s

Let us now consider distribution laws of the deviations of the ML-estimator from the true parameter, and the RML-estimator from the corresponding ML-estimator. Before we do that, however let us restate the definitions of weakly approaching sequences of d.l.’s introduced by Belyaev and Sjöstedt [4].

Let $C_b(\mathbb{R}^k)$ denote all real valued bounded functions on $\mathbb{R}^k$.

**Definition 3** Two sequences of distribution laws $\{\mathcal{L}(X_n)\}_{n \geq 1}$ and $\{\mathcal{L}(V_n)\}_{n \geq 1}$, $X_n, V_n \in \mathbb{R}^k$, are said to weakly approach each other if and only if for any function $h(\cdot) \in C_b(\mathbb{R}^k)$, $E[h(X_n)] - E[h(V_n)] \to 0$, as $n \to \infty$. This type of convergence is denoted by

$$\mathcal{L}(X_n) \xrightarrow{w} \mathcal{L}(V_n) \text{ as } n \to \infty.$$  

**Remark** If $\mathcal{L}(Z_n) = \mathcal{L}(Z_0)$ for some fixed random variable $Z_0$, $n = 1, 2, \ldots$, then Definition 3 and the definition of the widely used weak convergence are equivalent.

Since the distribution laws of RML-estimators are themselves random distribution laws, we have to use the following definition for random distribution laws.

**Definition 4** Let $\{Z_n, W_n\}_{n \geq 1}$ and $\{V_n\}_{n \geq 1}$ be two sequences of random variables, $Z_n, V_n \in \mathbb{R}^k$, where the random elements $W_n$ belong to some space $W_n$, and $Z_n, W_n$ are defined on the same probability space for each $n$. Then the sequence of conditional distribution laws $\{\mathcal{L}(Z_n | W_n)\}_{n \geq 1}$, given $W_n$, weakly approaches $\{\mathcal{L}(V_n)\}_{n \geq 1}$ in probability if and only if for any function $h(\cdot) \in C_b(\mathbb{R}^k)$, $E[h(Z_n) | W_n] - E[h(V_n)] \to 0$ as $n \to \infty$, in probability. This type of convergence is denoted by

$$\mathcal{L}(Z_n | W_n) \xrightarrow{w(P)} \mathcal{L}(V_n) \text{ as } n \to \infty.$$  

If $l_n(\theta, Y_n)$ is differentiable in $\theta$, then the mean value of the score vectors at $\theta$ is

$$\bar{U}_n(\theta_n, W_n) := n^{-1} \sum_{r=1}^{n} U_{rn}(\theta_n, Y_{rn}),$$

where

$$U_{rn}(\theta, Y_{rn}) := \left( \frac{\partial l_{rn}(\theta, Y_{rn})}{\partial \theta_1}, \frac{\partial l_{rn}(\theta, Y_{rn})}{\partial \theta_2}, \ldots, \frac{\partial l_{rn}(\theta, Y_{rn})}{\partial \theta_k} \right)^T.$$  

If $\theta_0 \in \tilde{\mathcal{K}} \subseteq \Theta$ and if the ML($\mathcal{K}$)-estimator $\hat{\theta}_n(\mathcal{K})$ exists, then it can be found as one of possibly many solutions to the likelihood equation

$$\bar{U}_n(\theta_n, W_n) = 0_k, \quad \theta_n \in \mathcal{K}. \quad (5)$$
Similarly, the RML(\(\mathcal{K}\))-estimator \(\hat{\theta}_n^{*}(\mathcal{K})\) can be found as one of possibly many solutions to the equation

\[
\hat{U}_n^{*}(\theta_n, Y_n) := n^{-1} \sum_{r=1}^{n} M_{r,n}^{*} U_{r,n}(\theta_n, Y_n) = 0 \quad \theta_n \in \mathcal{K},
\]

(6)

By considering growing sequences of compact sets \(\{\mathcal{K}_j\}_{j \geq 1}\) where \(\bigcup_{j \geq 1} \mathcal{K}_j = \Theta\), the ML-estimator \(\hat{\theta}_n = \hat{\theta}_n(\Theta)\) and the RML-estimator \(\hat{\theta}_n^{*} = \hat{\theta}_n^{*}(\Theta)\) can be obtained.

Let us introduce the following regularity assumptions:

**B1:** For almost all \(y_{r,n} \in Y_{r,n}\), the partial derivatives

\[
\frac{\partial l_{r,n}(\theta, y_{r,n})}{\partial \theta_i} \quad \text{and} \quad \frac{\partial^2 l_{r,n}(\theta, y_{r,n})}{\partial \theta_i \partial \theta_j}
\]

exist and are continuous functions of \(\theta \in \Theta, (r, n) \in \mathcal{I}\).

**B2(\(\beta\))**; For each \(\theta \in \Theta\), there exists \(\beta > 0\), such that,

\[
E_{\Theta} \left[ \left| \frac{\partial l_{r,n}(\theta, Y_{r,n})}{\partial \theta_i} \right|^{2+\beta} \right] \leq d_2(\beta, \theta),
\]

where \(d_2(\beta, \theta)\) is a finite constant, \(i = 1, 2, \ldots, k, (r, n) \in \mathcal{I}\).

**B3(\(\beta\))**; For each \(\theta \in \Theta\),

\[
E_{\Theta} \left[ \frac{\partial l_{r,n}(\theta, Y_{r,n})}{\partial \theta_i} \frac{\partial l_{r,n}(\theta, Y_{r,n})}{\partial \theta_j} \right] = -E_{\Theta} \left[ \frac{\partial^2 l_{r,n}(\theta, Y_{r,n})}{\partial \theta_i \partial \theta_j} \right],
\]

and there exists \(\beta > 0\), such that,

\[
E_{\Theta} \left[ \left| \frac{\partial^2 l_{r,n}(\theta, Y_{r,n})}{\partial \theta_i \partial \theta_j} \right|^{1+\beta} \right] \leq d_3(\beta, \theta),
\]

where \(d_3(\beta, \theta)\) is a finite constant, \(i, j = 1, 2, \ldots, k, (r, n) \in \mathcal{I}\).

For any \(\theta \in \Theta\) and \(\alpha > 0\), denote

\[
Q_{i,j,n}(\alpha, \theta, Y_{r,n}) := \sup_{\theta' \in B_{\alpha}(\theta)} \left| \frac{\partial^2 l_{r,n}(\theta, Y_{r,n})}{\partial \theta_i \partial \theta_j} - \frac{\partial^2 l_{r,n}(\theta', Y_{r,n})}{\partial \theta_i \partial \theta_j} \right|.
\]

(7)
B4(\(c, \delta\)): For any \(\theta \in \Theta\) and for any arbitrarily small \(\delta > 0\) there exists a small \(\epsilon(\delta, \theta)\) such that, for arbitrarily small positive \(\epsilon \leq \epsilon(\delta, \theta)\)

\[
E_\theta [Q_{t;j,n}(\epsilon, \theta, Y_{rn})] \leq \delta^2,
\]

\(i, j = 1, 2, \ldots, k, (r, n) \in \mathcal{I}\).

B5: For each \(\theta \in \Theta\),

\[
E_\theta \left[ \frac{\partial l_{rn}(\theta, Y_{rn})}{\partial \theta_i} \right] = 0_k,
\]

\(i = 1, 2, \ldots, k, (r, n) \in \mathcal{I}\).

For each \(\theta \in \Theta\), let

\[
\mathbb{F}_n(\theta) = n^{-1} \sum_{r=1}^{n} \mathbb{F}_{rn}(\theta),
\]

where \(\mathbb{F}_{rn}(\theta)\) is the Fisher information matrix with elements

\[
F_{ijrn}(\theta) := -E_\theta \left[ \frac{\partial^2 l_{rn}(\theta, Y_{rn})}{\partial \theta_i \partial \theta_j} \right],
\]

\(i, j = 1, 2, \ldots, k, (r, n) \in \mathcal{I}\).

B6: For each \(\theta \in \Theta\), there exists a constant \(d_0(\theta) > 0\), such that,

\[
\inf_n |\text{det} (\mathbb{F}_n(\theta))| \geq d_0(\theta).
\]

We now state theorems concerning the distribution laws of the deviations of the ML- and RML-estimators from the true parameter.

**Theorem 5** If assumptions B1-B6 hold and there exists with \(P_{\theta_0}\)-probability tending to one, as \(n \to \infty\), a consistent ML-estimator \(\hat{\theta}_n\) for \(\theta_0\), then

\[
\mathcal{L} \left( \sqrt{n}(\hat{\theta}_n - \theta_0) \right) \xrightarrow{w} \mathcal{N}_k(0, \mathbb{F}_n^{-1}(\theta_0)), \quad n \to \infty.
\]

**Theorem 6** If assumptions B1-B6 hold and there exists with \(P_{\theta_0}^{(\star)}\)-probability tending to one, as \(n \to \infty\), a consistent RML-estimator \(\hat{\theta}_n\) for \(\theta_0\), then

\[
\mathcal{L} \left( \sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) \right| Y_n) \xrightarrow{w(P)} \mathcal{N}_k(0, \mathbb{F}_n^{-1}(\theta_0)), \quad n \to \infty.
\]

Proofs for these assertions are included in the appendix. From Theorems 5 and 6 the following corollary is immediate:

**Corollary 3** If assumptions B1-B6 hold and there exist with \(P_{\theta_0}^{(\star)}\)-probability tending to one, as \(n \to \infty\), a consistent ML-estimator \(\hat{\theta}_n\) and a consistent RML-estimator \(\hat{\theta}_n^*\) for \(\theta_0\), then

\[
\mathcal{L} \left( \sqrt{n}(\hat{\theta}_n - \theta_0) \right| Y_n) \xrightarrow{w(P)} \mathcal{L} \left( \sqrt{n}(\hat{\theta}_n^* - \theta_0) \right), \quad n \to \infty.
\]
5 Conclusions

The assumptions introduced in Sections 2 and 4 guarantee that the d.l.’s of the deviations of ML-estimators from the true unknown parameter can be estimated accurately by resampling when the size of the data is increasing. For many statistical models, used in routine statistical data analysis, these assumptions are fulfilled. In our forthcoming research report, two statistical models will be considered: the exponential regression model and the linear logistic regression model. The exponential regression model is related to reliability and survival data (see e.g. Andersen et al. Sec VII.6.1, Lawless [15] Sec. 6.3 and Kalbfleisch and Prentice [14] Sec 2.3). The linear logistic regression model is often used in analysis of biomedical data (see e.g. Collett [8] Sec 3.6 and McCullagh and Nelder [17] sec 4.3). Under some natural restrictions to the explanatory variables (covariates), these models fulfill the assumptions of Sections 2 and 4.

The above used notion of weakly approaching sequences of d.l.’s has some advantages in comparison to that of weak convergence. Especially in design of sequential experiments where some current “optimal” values of explanatory variables have to be found, and where the variability of these values essentially influences the d.l.’s of the deviations. At the same time, it is not necessary to find the distribution of the stabilizing transformations which can be difficult.

Our preliminary numerical experiments are in correspondence with the suggested theory. We observed increasing accuracy for the estimates of d.l.’s of the deviations when the number of observations was increasing. We also found that when the number of observations was relatively low, the accuracy could depend heavily on the observations. With an increase in number of observations, we saw this accuracy increase. A topic for further research is to find some “tools for acceleration”, which increase this accuracy for a fixed number of observations.

6 Acknowledgements

We want to thank Dr. Magnus Ekström for constructive criticism of a draft version of this paper. We also want to thank Dr. Paul Haemig and Dr. Sean McGuiness for improving the English language.

Appendix

This appendix contains proofs for all lemmas and theorems stated in Sections 2, 3 and 4. All notational conventions adopted in the previous sections are also
applicable here.

**A.1 Proof of theorems in Section 2**

Before we prove Theorem 1 and Theorem 2, some definitions and preliminary results are needed. For a set $B \in \mathfrak{B}(\Theta)$, we define the following r.v.'s related to the loglikelihood ratio:

$$R_n(\theta_0, B) := \sum_{r=1}^{n} R_{rn}(\theta_0, B), \quad R_n(\theta_0, \mathcal{B}) := \inf_{\theta \in B} R_n(\theta_0, \theta),$$

$$R^0_{rn}(\theta_0, B) := R_{rn}(\theta_0, B) - \mathbb{E}_\theta [R_{rn}(\theta_0, B)], \quad R^0_n(\theta_0, B) := \sum_{r=1}^{n} R^0_{rn}(\theta_0, B).$$

Equality (1) can now be written in the equivalent form

$$R_n(\theta_0, \hat{\theta}_n(\mathcal{A})) = R_n(\theta_0, \mathcal{A}).$$

Three relations which $R_n(\theta_0, \cdot)$ satisfies will now be stated:

$$R_n(\theta_0, B) = \inf_{\theta \in B} \sum_{r=1}^{n} R_{rn}(\theta_0, \theta) \geq \sum_{r=1}^{n} \inf_{\theta \in B} R_{rn}(\theta_0, \theta) = R_n(\theta_0, B); \quad (9)$$

if $B \subset A$, where $\mathcal{A}, \mathcal{B} \in \mathfrak{B}(\Theta)$, then

$$R_n(\theta_0, B) \geq R_n(\theta_0, A); \quad (10)$$

and if each $B_j \in \mathfrak{B}(\Theta)$, then

$$R_n \left( \theta_0, \bigcup_{j=1}^{m} B_j \right) = \min_{1 \leq j \leq m} R_n(\theta_0, B_j). \quad (11)$$

If $B \subset A$, where $\mathcal{A}, \mathcal{B} \in \mathfrak{B}(\Theta)$, then for each $(r, n) \in \mathcal{T}$,

$$R_{rn}(\theta_0, \mathcal{B}) \geq R_{rn}(\theta_0, \mathcal{A}). \quad (12)$$

By Lyapunov's inequality, we have that

$$\left( \mathbb{E} \left[ \left| R^0_{rn}(\theta_0, \mathcal{B}) \right| ^\gamma \right] \right)^\gamma \leq \mathbb{E} \left[ \left| R^0_{rn}(\theta_0, \mathcal{B}) \right| ^\gamma \right], \quad \gamma \geq 1. \quad (13)$$

From the elementary inequality

$$|a + b|^\eta \leq 2^{\eta-1} (|a|^\eta + |b|^\eta), \quad \eta \geq 1, \quad (14)$$

and Lyapunov's inequality, we get that

$$\mathbb{E} \left[ \left| R^0_{rn}(\theta_0, \mathcal{B}) \right| ^\gamma \right] \leq 2^\gamma \mathbb{E} \left[ \left| R_{rn}(\theta_0, \mathcal{B}) \right| ^\gamma \right], \quad \gamma \geq 1. \quad (15)$$
The following inequality will be needed. If $X_1, X_2, \ldots, X_n$ are $n$ independent univariate random variables with zero mean and $E[|X_i|^\gamma] < \infty$, $\alpha \geq 1$, $r = 1, 2, \ldots, n$, then
\[
E \left[ \sum_{r=1}^{n} X_r^0 \right] \leq k(\alpha) n^{\max(\alpha/2-1, 0)} \sum_{r=1}^{n} E[|X_r|^\gamma],
\]
where $k(\alpha)$ is some finite constant. It has been shown that $k(\alpha) \leq 2$ if $1 \leq \alpha \leq 2$ (von Bahr and Esseen [22]). It is also possible to show, by using the results of Dharmadhikari and Jogdeo [9], that $k(\alpha) \leq 5$ if $2 \leq \alpha \leq 3$. Of course, if $\alpha = 2$, then $k(2) = 1$.

**Lemma 1** Assumption A3($\gamma, \rho_3$), with $\gamma > 1$ and $\rho_3 > 0$, implies assumption A2($\rho_2$) with $\rho_2 = \rho_3$.

*Proof of Lemma 1.* By applying Hölder’s inequality, Chebyshev’s inequality, and assumption A3($\gamma, \rho_3$), we obtain for any arbitrarily small $\epsilon > 0$ that
\[
E_{\theta_0}( |R_n (\theta_0, \theta) | I ( |R_n (\theta_0, \theta) | > c_2(\epsilon, \theta_0, \theta) ) ] \\
\leq E_{\theta_0}( |R_n (\theta_0, \theta) |^\gamma ) P_{\theta_0}( |R_n (\theta_0, \theta) | > c_2(\epsilon, \theta_0, \theta) )^{1/\gamma} \\
\leq (c_3(\gamma, \theta_0, \theta))^{1/\gamma} \left( \frac{E_{\theta_0}( |R_n (\theta_0, \theta) |^\gamma )}{(c_2(\epsilon, \theta_0, \theta))^\gamma} \right)^{1/\gamma} \leq \frac{c_3(\gamma, \theta_0, \theta)}{(c_2(\epsilon, \theta_0, \theta))^\gamma-1}.
\]

If $c_2(\epsilon, \theta_0, \theta) > \left( \frac{c_3(\gamma, \theta_0, \theta)}{\epsilon} \right)^{1/\gamma}$, then the right hand side of (17) is less than $\epsilon$. Thus, (2) holds.

Likewise, if $c_2(\epsilon, \theta_0, \theta) > \left( \frac{c_3(\gamma, \theta_0, \theta)}{\epsilon} \right)^{1/\gamma}$, we find that
\[
E_{\theta_0}( |R_n (\theta_0, \theta_2(\theta) ) | I ( |R_n (\theta_0, \theta_2(\theta) ) | > c_2(\epsilon, \theta_0, \theta) ) ] < \epsilon,
\]
where $\rho_2(\epsilon, \theta_0, \theta) = \rho_3(\gamma, \theta_0, \theta)$. Hence, (3) holds and the desired result is obtained.

**Lemma 2** Let $X_{1n}, X_{2n}, \ldots, X_{mn}$ be independent r.v.’s with $E[|X_{rn}|] < \infty$, $(r, n) \in \Gamma$. Assume that for any arbitrarily small $\epsilon > 0$ there exists a constant $c_\epsilon$, $0 < c_\epsilon < \infty$, such that
\[
\limsup_n \max_{1 \leq r \leq n} E[|X_{rn}| I(|X_{rn}| > c_\epsilon)] < \epsilon.
\]

Then
\[
n^{-1} \sum_{r=1}^{n} X_{rn}^0 = o_P(1), \quad n \to \infty,
\]
where $X_{rn}^0 := X_{rn} - E[X_{rn}]$. 

}\
Proof of Lemma 2. Let us define for any $c > 0$ the following r.v.’s:

$$V_{rn}(c) := X_{rn} I(\{|X_{rn}| \leq c\}), \quad V_{rn}^o(c) := V_{rn}(c) - E[V_{rn}(c)],$$
$$W_{rn}(c) := X_{rn} I(\{|X_{rn}| > c\}), \quad W_{rn}^o(c) := W_{rn}(c) - E[W_{rn}(c)].$$

Since $|V_{rn}(c)| \leq c$, we have from relation (14) and Lyapunov’s inequality that $E[|V_{rn}^o(c)|^\alpha] \leq 2^\alpha c^\alpha < \infty$ for any $\alpha > 0$. By (18) and Lyapunov’s inequality, we have for all sufficiently large $n$ that $E[|W_{rn}^o(c)|] < 2\epsilon$. Let $\delta > 0$ be arbitrarily small, $\epsilon = \delta^2$, and $1 < \alpha \leq 2$. Then by Chebyshev’s inequality and (16), it follows for all sufficiently large $n$ that

$$P\left( \left| \frac{1}{n} \sum_{r=1}^{n} X_{rn}^o \right| > \delta \right)$$
$$\leq P\left( \left| \frac{1}{n} \sum_{r=1}^{n} V_{rn}^o(c) \right| > \delta/2 \right) + P\left( \left| \frac{1}{n} \sum_{r=1}^{n} W_{rn}^o(c) \right| > \epsilon/2 \right)$$
$$\leq \left( \frac{2}{\delta n} \right)^\alpha E \left[ \left| \sum_{r=1}^{n} V_{rn}^o(c) \right|^\alpha \right] + \frac{2}{\delta n} E \left[ \left| \sum_{r=1}^{n} W_{rn}^o(c) \right|^\alpha \right]$$
$$\leq \left( \frac{2}{\delta n} \right)^\alpha \left( 2n \sum_{r=1}^{n} E \left[ |V_{rn}^o(c)|^\alpha \right] \right) + \frac{2}{\delta n} \sum_{r=1}^{n} E \left[ |W_{rn}^o(c)|^\alpha \right]$$
$$\leq \left( \frac{2}{\delta n} \right)^\alpha \frac{2n^2 \delta^\alpha}{\alpha} + \frac{2}{\delta n} \sum_{r=1}^{n} E \left[ |W_{rn}^o(c)|^\alpha \right]$$

(19)

Then for all sufficiently large $n > \left( \frac{2^\alpha}{\alpha \delta n} \right)^{1/\alpha}$, (19) is less or equal to $6\delta$. The desired result follows, since $\delta$ is arbitrarily small and $\epsilon = \delta^2$. \qed

Lemma 3 Suppose that assumption A1($\rho_1$) holds with $\rho_1 > 0$. Then for each $\theta_0, \theta \in \Theta$, where $\theta_0 \neq \theta$, for each positive $\rho_0 \leq \rho_1$, and all sufficiently large $n \geq n(\rho_0, \theta_0, \theta)$

$$\min_{1 \leq r \leq n} E_{\theta_0} [R_{rn}(\theta_0, B_{\rho_0}(\theta))] > b(\theta_0, \theta) > 0.$$  (20)

Proof of Lemma 3. By relation (12) we have for any $\rho \leq \rho'$ that $R_{rn}(\theta_0, B_{\rho}(\theta)) \leq R_{rn}(\theta_0, B_{\rho'}(\theta))$. Then it follows from assumption A1($\rho_1$) that

$$\min_{1 \leq r \leq n} E_{\theta_0} [R_{rn}(\theta_0, B_{\rho_0}(\theta))] \geq \min_{1 \leq r \leq n} E_{\theta_0} [R_{rn}(\theta_0, B_{\rho_1}(\theta))] > b(\theta_0, \theta) > 0,$$

for all sufficiently large $n$. Hence, inequality (20) holds and the proof is complete. \qed

Lemma 4 Suppose that assumption A3($\gamma, \rho_3$) holds with $\gamma > 2$ and $\rho_3 > 0$. Then for each $\theta_0, \theta \in \Theta$, where $\theta_0 \neq \theta$, for each positive $\rho_0 \leq \rho_3$, and all sufficiently large $n \geq n(\gamma, \rho_0, \theta_0, \theta)$

$$\max_{1 \leq r \leq n} E_{\theta_0} [R_{rn}(\theta_0, B_{\rho_0}(\theta))] < 2c_3(\gamma, \theta_0, \theta) < \infty.$$  (21)
Proof of Lemma 4. Let us consider the negative and positive parts of the r.v.’s \( R_{\epsilon n}(\theta_0, B_{\rho_0}(\theta)) \); that is,

\[
R^{-\epsilon}_{\epsilon n}(\theta_0, B_{\rho_0}(\theta)) := - (R_{\epsilon n}(\theta_0, B_{\rho_0}(\theta)) \wedge 0),
\]

and

\[
R^{\epsilon}_{\epsilon n}(\theta_0, B_{\rho_0}(\theta)) := R_{\epsilon n}(\theta_0, B_{\rho_0}(\theta)) \lor 0.
\]

By relation (12) we also have that

\[
R^{-\epsilon}_{\epsilon n}(\theta_0, B_{\rho_0}(\theta)) \leq R^{-\epsilon}_{\epsilon n}(\theta_0, B_{\rho_0}(\theta)) \quad \text{and} \quad R^{\epsilon}_{\epsilon n}(\theta_0, B_{\rho_0}(\theta)) \leq R^{\epsilon}_{\epsilon n}(\theta_0, \theta).
\]

It follows, from (22) and A3(\( \gamma, \rho_3 \)), that

\[
E_{\rho_0}[\left|R^{-\epsilon}_{\epsilon n}(\theta_0, B_{\rho_0}(\theta))\right|^r] + E_{\rho_0}[\left|R^{\epsilon}_{\epsilon n}(\theta_0, B_{\rho_0}(\theta))\right|^r] 
\leq 2E_{\rho_0}[\left|R^{-\epsilon}_{\epsilon n}(\theta_0, B_{\rho_0}(\theta))\right|^r] + E_{\rho_0}[\left|R^{-\epsilon}_{\epsilon n}(\theta_0, \theta)\right|^r] < \infty,
\]

for all sufficiently large \( n \) and any \( r \), \( 1 \leq r \leq n \). Hence, the inequality (21) holds and the proof is complete.

We are now ready to prove Theorem 1 and Theorem 2.

Proof of Theorem 1. Let \( \epsilon > 0 \) be arbitrarily small and such that \( B_{\epsilon}(\theta_0) \subset \mathcal{K} \).

For each \( \theta \in \mathcal{K} \setminus B_{\epsilon}(\theta_0) \), a positive \( \rho_\epsilon = \rho_\epsilon(\theta_0, \theta) \leq \rho_1 \land \rho_2 \) can be found such that \( \theta_0 \not\in B_{\rho_\epsilon}(\theta) \).

Thus,

\[
\mathcal{K} \setminus B_{\epsilon}(\theta_0) \subset \bigcup_{\theta \in \mathcal{K} \setminus B_{\epsilon}(\theta_0)} B_{\rho_\epsilon}(\theta).
\]

Since the set \( \mathcal{K} \setminus B_{\epsilon}(\theta_0) \) is compact, a finite subcover

\[
\bigcup_{j=1}^{m_\epsilon} B_j \supset \mathcal{K} \setminus B_{\epsilon}(\theta_0),
\]

where \( B_j = B_{\rho_\epsilon(\theta_0, \theta_j)}(\theta_j) \), can be selected. Here we stress that \( m_\epsilon \) does not depend on \( n \).

Let

\[
A_n := \{ R_n(\theta_0, \bigcup_{j=1}^{m_\epsilon} B_j) > 0 \}.
\]

For all sufficiently large \( n \), the following sequence of inequalities is obtained by using the relations (11), (9), and (20):

\[
P_{\rho_0}(A_n) = P_{\rho_0}\left( \bigcap_{j=1}^{m_\epsilon} (R_n(\theta_0, B_j) > 0) \right) = 1 - P_{\rho_0}\left( \bigcup_{j=1}^{m_\epsilon} (R_n(\theta_0, B_j) \leq 0) \right) 
\geq 1 - \sum_{j=1}^{m_\epsilon} P_{\rho_0}(R_n(\theta_0, B_j) \leq 0) \geq 1 - \sum_{j=1}^{m_\epsilon} P_{\rho_0}(R_n(\theta_0, B_j) \leq 0)
\]
\[ 1 - \sum_{j=1}^{m_r} P_{\theta_0} \left( R_{n}(\theta_0, B_j) \leq -\sum_{r=1}^{n} E_{\theta_0}[R_{\infty}(\theta_0, B_j)] \right) \]
\[ \geq 1 - \sum_{j=1}^{m_r} P_{\theta_0} \left( R_{n}(\theta_0, B_j) < -n\delta(\theta_0, \theta) \right). \]

Then, by using assumption A2(\rho_2) and Lemma 2 with \( X_m = R_{\infty}(\theta_0, B_j), j = 1, 2, \ldots, m, \) it follows from (23) that
\[ P_{\theta_0}(A_n) \geq 1 - \sum_{j=1}^{m_r} P_{\theta_0} \left( R_{n}(\theta_0, B_j) < -n\delta(\theta_0, \theta) \right) = 1 - o(1), \quad n \to \infty. \]

By relation (10) we obtain that
\[ P_{\theta_0}(R_{n}(\theta_0, \mathcal{K} \setminus B_i(\theta_0)) > 0) \geq P_{\theta_0}(A_n). \]

Since \( \partial B_i(\theta_0) \cap \mathcal{K} \subset \cup_{j=1}^{m} B_j, \) we have that
\[ P_{\theta_0}(R_{n}(\theta_0, \partial B_i(\theta_0) \cap \mathcal{K}) > 0) \geq P_{\theta_0}(A_n). \]

Moreover, \( R_{n}(\theta_0, \theta_0) = 0 \) and according to assumption A1(\rho_1), \( R_{n}(\theta_0, \cdot) \) is a continuous function. Then if \( A_n \) occurs, \( R_{n}(\theta_0, \cdot) \) has to attain its minimum value inside the set \( B_i(\theta_0) \cap \mathcal{K} = B_i(\theta_0). \) Hence, there exists at least one ML(\mathcal{K})-estimator \( \hat{\theta}_n(\mathcal{K}) \in B_i(\theta_0) \) with a probability tending to one, since \( P_{\theta_0}(A_n) \to 1 \) as \( n \to \infty. \) Furthermore, if \( A_n \) occurs, then \( \| \hat{\theta}_n(\mathcal{K}) - \theta_0 \|_2 < \epsilon. \) Since \( \epsilon \) can be chosen arbitrarily small and \( P_{\theta_0}(A_n) \to 1 \) as \( n \to \infty, \) we obtain that \( \hat{\theta}_n(\mathcal{K}) \xrightarrow{p} \theta_0 \) as \( n \to \infty. \) This completes the proof of Theorem 1. \( \square \)

**Proof of Theorem 2.** The main idea of this proof is the same as the one of Theorem 1. Here we use \( \rho_1 = \rho_1(\theta_0, \theta) \leq \rho_1 \wedge \rho_2 \) in the construction of a finite subcover \( \cup_{j=1}^{m} B_j. \) We can now estimate \( P_{\theta_0}(A_n) \) as in (23), and prolong the estimation by using the Chebyshev inequality and inequalities (16), (15) and (21). We obtain that
\[ P_{\theta_0}(A_n) \leq 1 - \sum_{j=1}^{m_r} P_{\theta_0} \left( |R_{n}(\theta_0, B_j)| \geq n\delta(\theta_0, \theta) \right) \]
\[ \geq 1 - \sum_{j=1}^{m_r} E_{\theta_0} \left[ |R_{n}(\theta_0, B_j)| \right] / (n\delta(\theta_0, \theta)) \]
\[ \geq 1 - \sum_{j=1}^{m_r} k(\gamma)n^{\gamma/2 - 1} \sum_{r=1}^{n} E_{\theta_0} \left[ |R_{n}(\theta_0, B_j)| \right] / (n\delta(\theta_0, \theta)) \]
\[ \geq 1 - \sum_{j=1}^{m_r} k(\gamma)n^{\gamma/2 - 1} \sum_{r=1}^{m} 2^{\gamma} E_{\theta_0} \left[ |R_{\infty}(\theta_0, B_j)| \right] / (n\delta(\theta_0, \theta)) \]
\[ \geq 1 - 2^{\gamma} k(\gamma)n^{-\gamma/2} \sum_{j=1}^{m_r} \frac{2\epsilon(\gamma, \theta_0, \theta_j)}{\delta(\theta_0, \theta)}. \]

(24)
The conclusion is drawn from the convergence of the series

\[
\sum_{n=1}^{\infty} P_{\theta_0}(R_n(\theta_0, \mathcal{K} \setminus \mathcal{B}_c(\theta_0))) \leq \sum_{n=1}^{\infty} (1 - P_{\theta_0}(A_n)) \leq k'(\gamma) \sum_{n=1}^{\infty} n^{-\gamma/2} < \infty,
\]

where \( k'(\gamma) = 2^{\gamma+1} k(\gamma) \sum_{j=1}^{m_r} \frac{\gamma(\gamma,\theta_0,\theta_1)}{\gamma(\theta_0,\theta_1)} \) < \infty. \( \square \)

### A.2 Proof of theorems in Section 3

Before we prove Theorem 3 and Theorem 4, let us give some preliminary results and define some r.v.’s similar to those introduced in the previous subsection.

\[
R^*_m(\theta_0, \mathcal{B}) := \inf_{\theta \in \mathcal{B}} R^*_m(\theta_0, \theta), \quad R^*_n(\theta_0, \mathcal{B}) := \inf_{\theta \in \mathcal{B}} R^*_n(\theta_0, \theta),
\]

\[
R^*_n(\theta_0, \mathcal{B}) := \sum_{r=1}^{n} M^*_r R^*_n(\theta_0, \mathcal{B}), \quad R^*_0(\theta_0, \mathcal{B}) := \sum_{r=1}^{n} M^*_r R^*_n(\theta_0, \mathcal{B}),
\]

where \( \mathcal{B} \in \mathfrak{B}(\Theta) \). For these r.v.’s, similar properties as in (9), (10), and (11) exist, namely

\[
R^*_n(\theta_0, \mathcal{B}) \geq \sum_{r=1}^{n} M^*_r \inf_{\theta \in \mathcal{B}} R^*_n(\theta_0, \theta) = R^*_n(\theta_0, \mathcal{B}), \quad (25)
\]

\[
R^*_n(\theta_0, \mathcal{B}) \geq R^*_n(\theta_0, \mathcal{A}) \quad \text{if} \quad \mathcal{A} \subseteq \mathcal{B}, \quad \mathcal{A}, \mathcal{B} \in \mathfrak{B}(\Theta), \quad (26)
\]

\[
R^*_n(\theta_0, \cup_{i=1}^{m} \mathcal{B}_i) = \min_{1 \leq m \leq n} R^*_n(\theta_0, \mathcal{B}_i), \quad \mathcal{B}_i \in \mathfrak{B}(\Theta), \quad i = 1, 2, \ldots, m. \quad (27)
\]

Furthermore, equality (4) can be written in the equivalent form

\[
R^*_n(\theta_0, \hat{\gamma}^	op(\mathcal{A})) = R^*_n(\theta_0, \mathcal{A}).
\]

Since \( M^*_r \) is a sum of indicator functions which are independent and identically distributed, it is binomially distributed, and therefore we have that

\[
\sum_{r=1}^{n} M^*_r = n, \quad \mathbb{E}^*[M^*_r] = 1, \quad \mathbb{E}^*[(M^*_r)^2] = 2 - 1/n.
\]

**Remark** For higher moments Belyaev and Rydén [3] proved that

\[
\mu(\gamma) := \sup_{n} \mathbb{E}^*[\gamma(M^*_r)^\gamma] \leq \left( \sum_{j=0}^{\gamma} \sigma(\gamma, j) \right)^{\gamma} < \infty, \quad (28)
\]

where \( \sigma(k, j), \ j = 0, 1, \ldots, k, \) are the Stirling numbers of second order (see e.g. Sachkov [21]), and \( \gamma \) is the smallest integer not less than \( \gamma \).
Lemma 5 Let $X_{1n}, X_{2n}, \ldots, X_{nn}$ be independent r.v.'s with $E[|X_{rn}|] < \infty$, $(r, n) \in \mathcal{F}$. Assume that for any arbitrarily small $\epsilon > 0$, there exists a constant $c_{\epsilon} > 0$, such that

$$\limsup_{n} \max_{1 \leq r \leq n} E[|X_{rn}|(|X_{rn}| > c_{\epsilon})] < \epsilon.$$  

Then

$$n^{-1} \sum_{r=1}^{n} X_{rn}^{0} = o_{p}(1), \quad n \to \infty,$$

where $X_{rn}^{0} := X_{rn} - E[X_{rn}]$.

Remark $\sum_{r=1}^{n} X_{rn}^{0} = \sum_{r=1}^{n} M_{rn}^{*} X_{rn}^{0}$, where $X_{rn}^{0} = X_{rn} - E[X_{rn}]$.

Proof of Lemma 5. Let us use the r.v.'s $V_{rn}^{0}(c)$ and $W_{rn}^{0}(c)$ defined in the proof of Lemma 2. From that proof, which is made under the same assumptions as this lemma, we have that $E[|V_{rn}^{0}(c)|^{\alpha}] \leq 2^{\alpha} c_{\alpha} < \infty$ for any $\alpha > 0$ and any $c > 0$. By Lyapunov’s inequality, we have for any $0 < \alpha \leq 2$ that

$$E^{*}[\left(M_{rn}^{*}\right)^{\alpha}] \leq E^{*}[\left(M_{rn}^{*} V_{rn}^{0}(c_{\epsilon})\right)^{\alpha/2}] \leq (2 - 1/n)^{\alpha/2} \leq 2.$$  

Let $\delta > 0$ be arbitrarily small, $\epsilon = \delta^{2}$, and $1 < \alpha \leq 2$. Then by Chebyshev’s inequality and (16), it follows for all sufficiently large $n$ that

$$p^{(\alpha)}\left(\left|\sum_{r=1}^{n} X_{rn}^{0}\right| > \delta\right) = p^{(\alpha)}\left(\left|\sum_{r=1}^{n} M_{rn}^{*} X_{rn}^{0}\right| > \delta\right) \leq p^{(\alpha)}\left(\left|\sum_{r=1}^{n} M_{rn}^{*} V_{rn}^{0}(c_{\epsilon})\right| > \delta/2\right) + p^{(\alpha)}\left(\left|\sum_{r=1}^{n} M_{rn}^{*} W_{rn}^{0}(c_{\epsilon})\right| > \epsilon/2\right)$$

$$\leq \left(\frac{2}{\delta n}\right)^{\alpha} E^{*}\left[\sum_{r=1}^{n} M_{rn}^{*} V_{rn}^{0}(c_{\epsilon})\right]^{\alpha} \frac{2}{\delta n} E^{*}\left[\sum_{r=1}^{n} M_{rn}^{*} W_{rn}^{0}(c_{\epsilon})\right]^{\alpha} \leq \left(\frac{2}{\delta n}\right)^{\alpha} E^{*}\left[\sum_{r=1}^{n} M_{rn}^{*} V_{rn}^{0}(c_{\epsilon})\right]^{\alpha} \frac{2}{\delta n} E^{*}\left[\sum_{r=1}^{n} M_{rn}^{*} W_{rn}^{0}(c_{\epsilon})\right]^{\alpha} \leq \left(\frac{2}{\delta n}\right)^{\alpha} \sum_{r=1}^{n} E^{*}[\left(M_{rn}^{*}\right)^{\alpha}] 2^{\alpha} c_{\epsilon}^{\alpha} + \frac{2}{\delta n} n2\epsilon \leq 2\left(\frac{2^{1+2\alpha}}{n^{\alpha-1}} + 2\delta\right).$$  

Then for all sufficiently large $n > \left(\frac{2^{1+2\alpha}}{n^{\alpha-1}}\right)^{1/\alpha}$, (31) is less or equal to $6\delta$. The desired result follows, since $\delta$ is arbitrarily small and $\epsilon = \delta^{2}$.

We are now ready to prove Theorem 3.

Proof of Theorem 3. The idea of this proof is similar to the proof of Theorem 1. Let us consider the same compact set $\mathcal{K} \setminus \mathcal{B}_{c}(\theta_{0})$ and its finite cover by open
with \(E\) \(i\) \(B\) \(R\) Since \(\frac{1}{2}\) \(\frac{1}{5}\) \(\frac{1}{0}\) we obtain following sequence of Theorem \(2\). Here, we also use \(\rho_2(\theta_0, \theta) \leq \rho_1 \wedge \rho_3\) in the construction of a finite subcover \(\cup_{j=1}^{m_\epsilon} \mathcal{B}_j\). By inequality (15) and (21), we have for \(\gamma > 2\) that

\[
E_{\theta_0} \left[ \left| M_{r_n}^* H_{r_n}^2 (\theta_0, \mathcal{B}_i) \right| \right] = (M_{r_n}^*)^\gamma E_{\theta_0} \left[ \left| H_{r_n}^2 (\theta_0, \mathcal{B}_i) \right| \right] 
\leq 2^{\gamma + 1} (M_{r_n}^*)^\gamma c_3(\gamma, \theta_0, \theta_i). \tag{33}
\]
Hence, by applying relations (16), (33), and (28) it follows that

\[
E_{\theta_0}^{(s)} \left[ P_{n}^{(t)}(\theta_0, B_i) \right] = E^{*} \left[ E_{\theta_0} \left[ \sum_{r=1}^{n} M_{r,n}^{*} P_{n}^{(t)}(\theta_0, B_i) \right] \right] \\
\leq E^{*} \left[ k(\gamma) n^{\gamma/2} \sum_{r=1}^{n} E_{\theta_0} \left[ M_{r,n}^{*} P_{n}^{(t)}(\theta_0, B_i) \right] \right] \\
\leq k(\gamma) n^{\gamma/2} 2^{-\gamma + 1} E^{*} \left[ (M_{r,n}^{*})^{\gamma} \right] c_3(\gamma, \theta_0, \theta_i) \\
\leq k'(\gamma) c_3(\gamma, \theta_0, \theta_i) n^{\gamma/2}, \quad \text{(34)}
\]

where \( k'(\gamma) = 2^{\gamma + 1} k(\gamma) \mu(\gamma). \)

We can now estimate \( P_{\theta_0}^{(s)}(A_n^{*}) \) as in (32) and prolong the estimation by using the Chebyshev inequality and inequalities (34),

\[
P_{\theta_0}^{(s)}(A_n^{*}) \geq 1 - \sum_{i=1}^{m_1} P_{\theta_0}^{(s)} \left[ P_{n}^{(t)}(\theta_0, B_i) \geq n b(\theta_0, \theta_i) \right] \\
\geq 1 - \sum_{i=1}^{m_1} E_{\theta_0}^{(s)} \left[ \left( n b(\theta_0, \theta_i) \right)^{\gamma} \right] \geq 1 - \sum_{i=1}^{m_1} k'(\gamma) c_3(\gamma, \theta_0, \theta_i) n^{-\gamma/2} \\
\geq 1 - k'(\gamma) \left( \sum_{i=1}^{m_1} \frac{c_3(\gamma, \theta_0, \theta_i)}{(n b(\theta_0, \theta_i)^{\gamma})} \right) n^{-\gamma/2}. \quad \text{(35)}
\]

The conclusion is drawn from the convergence of the series

\[
\sum_{n=1}^{\infty} \sum_{i=1}^{m_1} P_{\theta_0}^{(s)}(A_n^{*}) \leq \sum_{n=1}^{\infty} (1 - P_{\theta_0}^{(s)}(A_n^{*})) \leq \kappa(\gamma) \sum_{n=1}^{\infty} n^{-\gamma/2} < \infty,
\]

where \( \kappa(\gamma) = 2^{\gamma + 1} k'(\gamma) \sum_{i=1}^{m_1} \frac{c_3(\gamma, \theta_0, \theta_i)}{(n b(\theta_0, \theta_i)^{\gamma})} < \infty. \)

\[\square\]

### A.3 Proofs of theorems in Section 4

The proofs of Theorems 5 and 6 will be based on the central limit resampling theorem stated in Belyaev and Sjöstedt [4]. This theorem is restated in Proposition 1 below.

**Proposition 1** Assume that \( \mathcal{X} = \{X_{1n}, X_{2n}, \ldots, X_{rn}\} \) is an array of \( n \) independent random vectors \( X_{rn} = (X_{1n}, X_{2n}, \ldots, X_{rn})^T \) with zero means, and that there exists a \( \zeta > 0 \) such that \( E \left[ \sqrt{n} X_{rn} \right]^{2 + \zeta} \leq d(\zeta), \ i = 1, 2, \ldots, k, \ r = 1, 2, \ldots, n, \) where \( d(\zeta) \) is some finite constant, \( n = 1, 2, \ldots. \) Then as \( n \to \infty, \)

\[
\begin{align*}
(i) & \quad \mathcal{L} \left( X_{n} \right) \xrightarrow{w} \mathcal{N}_k \left( \mathbf{0}, \mathbb{E}_n \right), \\
(ii) & \quad \mathcal{L} \left( \sum_{r=1}^{n} (M_{r,n}^{*} - 1) X_{rn} \mid \mathcal{X}_n \right) \xrightarrow{w(F)} \mathcal{N}_k \left( \mathbf{0}, \mathbb{E}_n \right),
\end{align*}
\]

where \( \mathbb{E}_n := n^{-1} \sum_{r=1}^{n} E[X_{rn}X_{rn}^T]. \)
Next, an auxiliary lemma is needed for proving Theorem 5 and Theorem 6.

**Lemma 6** Let \( \{X_n\}_{n \geq 1}, \{Z_n \mid W_n\}_{n \geq 1}, \) and \( \{V_n\}_{n \geq 1} \) be three sequences of \( k \)-dimensional random column vectors, introduced in Definition 3 and Definition 4, satisfying

\[
\begin{align*}
\mathcal{L}(X_n) &\xrightarrow{w} \mathcal{L}(V_n) \quad \text{as} \quad n \to \infty, \\
\mathcal{L}(Z_n \mid W_n) &\xrightarrow{w(P)} \mathcal{L}(V_n) \quad \text{as} \quad n \to \infty,
\end{align*}
\]

where \( \{V_n\}_{n \geq 1} \) is uniformly tight. If there exists a constant \( \kappa < \infty \) such that \( \sup_n \|W_n\|_2 \leq \kappa \), where each \( W_n \) is a \( m \times k \) nonrandom matrix, then

\[
\begin{align*}
\mathcal{L}(E_n X_n) &\xrightarrow{w} \mathcal{L}(E_n V_n) \quad \text{as} \quad n \to \infty, \\
(36) \\
\mathcal{L}(E_n Z_n \mid W_n) &\xrightarrow{w(P)} \mathcal{L}(E_n V_n) \quad \text{as} \quad n \to \infty. \\
(37)
\end{align*}
\]

**Proof of Lemma 6.** From Definitions 3 and 4, we have for any function \( h(\cdot) \in \mathcal{C}_b(\mathbb{R}^k) \), that

\[
\begin{align*}
E[h(X_n)] - E[h(V_n)] &\longrightarrow 0 \quad \text{as} \quad n \to \infty, \\
(38) \\
E[h(Z_n) \mid W_n] - E[h(V_n)] &\xrightarrow{P} 0 \quad \text{as} \quad n \to \infty. \\
(39)
\end{align*}
\]

If \( \{E_n V_n\}_{n \geq 1} \) is uniformly tight, Theorems 3 and 4 in Belyaev and Sjöstedt [4] tell us that (36) and (37) hold, if and only if for each \( t \in \mathbb{R}^n \),

\[
\begin{align*}
E[e^{it^T X_n}] - E[e^{it^T V_n}] &\longrightarrow 0 \quad \text{as} \quad n \to \infty, \\
(40) \\
E[e^{it^T Z_n} \mid W_n] - E[e^{it^T V_n}] &\xrightarrow{P} 0 \quad \text{as} \quad n \to \infty, \\
(41)
\end{align*}
\]

where \( i = \sqrt{-1} \).

Since \( \sup_n \|W_n\|_2 \leq \kappa \), we have that \( \{E_n V_n\}_{n \geq 1} \) is uniformly tight. Moreover, for any \( t' \in \mathbb{R}^1 \), \( E[e^{it}] = E[\cos(t')] + iE[\sin(t')] \), the expectations in (40) and (41) can be represented as expectations of real valued bounded functions of r.v.'s in \( \mathbb{R}^1 \). The desired result follows from (38) and (39). \( \square \)

Before we prove Theorems 5 and 6, let us define the following r.v.'s: If \( l_n(\theta, \psi_n) \) is twice differentiable in \( \theta \)

\[
\begin{align*}
\overline{h}_n(\theta, \psi_n) := n^{-1} \sum_{r=1}^n \overline{h}_r(\theta, \psi_n), \quad \text{where} \quad \overline{h}_r(\theta, \psi_n) := \left( \frac{\partial^2 l_r(\theta, Y_{rn})}{\partial \theta_i \partial \theta_j} \right)_{i,j=1,2,\ldots,k}.
\end{align*}
\]

**Proof of Theorem 5.** To get shorter notations, let us drop the arguments \( \psi_n \) and \( Y_{rn} \) of all functions in Section 4, e.g. \( \overline{U}_n(\hat{\theta}_n) = \overline{U}_n(\hat{\theta}_n, \psi_n) \) and \( U_r(\hat{\theta}_n) = U_r(\hat{\theta}_n, Y_{rn}) \).
The first order Taylor expansion of equation system (5) with \( \theta_n = \hat{\theta}_n \), about \( \theta_0 \), gives us the following equivalent relation

\[
\sqrt{n} \hat{U}_n(\hat{\theta}_n) = \sqrt{n} \hat{U}_n(\theta_0) + \left( \frac{\hat{\theta}_n - \theta_0}{n^{1/2}} \right) \sqrt{n}(\hat{\theta}_n - \theta_0) = 0_k, \quad (42)
\]

where

\[
\frac{\hat{\theta}_n - \theta_0}{n^{1/2}} = \left( \frac{\hat{\theta}_n - \theta_0}{\sqrt{n}} \right),
\]

and \( \hat{\theta}_n = \theta_0 + \alpha_n(\hat{\theta}_n - \theta_0) \), for some \( \alpha_n = \alpha_n(\theta_0, \hat{\theta}_n) \in [0, 1] \).

Let us start to prove that each element \( \frac{\hat{\theta}_n - \theta_0}{n^{1/2}} \) of matrix \( \frac{\hat{\theta}_n - \theta_0}{n^{1/2}} \) is converging to zero in probability, as \( n \to \infty \). If \( \hat{\theta}_n \in B_\alpha(\theta_0) \), where \( \alpha > 0 \), then we have from (7) that

\[
\left| \frac{\hat{\theta}_n - \theta_0}{n^{1/2}} \right| \leq n^{-1} \sum_{r=1}^{n} Q_{ij,r} \alpha_n(\theta_0) \leq \delta.
\]

By assumption \( B_4(\epsilon, \delta) \) with any arbitrarily small positive \( \epsilon \leq \epsilon(\delta, \theta_0) \) and Chebyshev’s inequality, it then follows that when \( \hat{\theta}_n \in B_\epsilon(\theta_0) \)

\[
P_{\theta_0} \left( \left\{ \left| \frac{\hat{\theta}_n - \theta_0}{n^{1/2}} \right| > \delta \right\} \mid \left\{ \hat{\theta}_n \in B_\epsilon(\theta_0) \right\} \right) = P_{\theta_0} \left( \left| \frac{\hat{\theta}_n - \theta_0}{n^{1/2}} \right| > \delta \mid \hat{\theta}_n \in B_\epsilon(\theta_0) \right) \leq \delta^{-1} E_{\theta_0} \left[ \left| \frac{\hat{\theta}_n - \theta_0}{n^{1/2}} \right| \mid \hat{\theta}_n \in B_\epsilon(\theta_0) \right] \leq \delta.
\]

Further, since \( \hat{\theta}_n = \theta_0 + \alpha_n(\hat{\theta}_n - \theta_0) \) and \( \hat{\theta}_n \) is a consistent estimator of \( \theta_0 \), we have for sufficiently large \( n \), say \( n \geq N(\epsilon, \delta) \), that

\[
P_{\theta_0} \left( \hat{\theta}_n \notin B_\epsilon(\theta_0) \right) \leq \delta.
\]

This together with (43) implies that

\[
P_{\theta_0} \left( \left| \frac{\hat{\theta}_n - \theta_0}{n^{1/2}} \right| > \delta \right) \leq P_{\theta_0} \left( \left\{ \left| \frac{\hat{\theta}_n - \theta_0}{n^{1/2}} \right| > \delta \right\} \cap \left\{ \hat{\theta}_n \in B_\epsilon(\theta_0) \right\} \right) + P_{\theta_0} \left( \left\{ \left| \frac{\hat{\theta}_n - \theta_0}{n^{1/2}} \right| > \delta \right\} \cap \left\{ \hat{\theta}_n \notin B_\epsilon(\theta_0) \right\} \right) \leq 2\delta,
\]

that is, as \( n \to \infty \)

\[
\frac{\hat{\theta}_n - \theta_0}{n^{1/2}} = o_P(1), \quad i, j = 1, 2, \ldots, k.
\]

Let us now consider \( \frac{\hat{\theta}_n - \theta_0}{n^{1/2}} \). Note that by (8) we have that

\[
E_{\theta_0} [\frac{\hat{\theta}_n - \theta_0}{n^{1/2}}] = -\frac{\hat{\theta}_n - \theta_0}{n^{1/2}}.
\]
From assumption B3(β) and the Lyapunov inequality, we have for each element \( H_{ij}(\theta_0) \) of \( \mathbb{R}_n(\theta_0) \) that
\[
|F_{ij}(\theta_0)| = |E_{\theta_0}[H_{ij}(\theta_0)]| \\
\leq E_{\theta_0}\left[\left|\frac{\partial^2 l_{\theta_0}}{\partial \theta_i \partial \theta_j}\right|^{1+\beta}\right]^{1/(1+\beta)} \\
\leq (d_3(\beta, \theta_0))^{1/\beta} < \infty. \quad (45)
\]

Without loss of generality, let \( 0 < \beta \leq 1 \). Then by assumption B3(β) and inequalities (16), (14) and (45), it follows that, for all sufficiently large \( n \),
\[
E_{\theta_0}\left[|\Pi_{ij}(\theta_0) + F_{ij}(\theta_0)|^{1+\beta}\right] \\
\leq 2n^{-(1+\beta)} \sum_{r=1}^{n} E_{\theta_0}\left[\left|\frac{\partial^2 l_{\theta_0}}{\partial \theta_i \partial \theta_j}\right|^{1+\beta} + |F_{ij}(\theta_0)|^{1+\beta}\right] \\
\leq 2n^{-(1+\beta)} \sum_{r=1}^{n} 2^\beta \left[E_{\theta_0}\left[\left|\frac{\partial^2 l_{\theta_0}}{\partial \theta_i \partial \theta_j}\right|^{1+\beta}\right]ight] \\
\leq 2^{2+\beta} d_3(\beta, \theta_0)n^{-\beta} < \infty.
\]

Then, by the Chebyshev inequality with any arbitrarily small \( \epsilon > 0 \), we have
\[
P_{\theta_0}\left(|\Pi_{ij}(\theta_0) + F_{ij}(\theta_0)| > \epsilon\right) \leq \frac{2^{2+\beta} d_3(\beta, \theta_0)n^{-\beta}}{\epsilon^{1+\beta}} \rightarrow 0, \quad n \rightarrow \infty.
\]

Hence as \( n \rightarrow \infty \)
\[
\Pi_{ij}(\theta_0) + F_{ij}(\theta_0) = o_P(1). \quad (46)
\]

Furthermore, from (45) it follows that each \( F_{ij}(\theta_0) = O(1) \).

Concerning \( \hat{U}_{ij}(\theta_n) \), it follows from assumption B3(β) that each element of the covariance matrix
\[
E_{\theta_0}\left[n\hat{U}_{ij}(\theta_0)\hat{U}_{ij}(\theta_0)^\top\right]
\]
is
\[
n^{-1} \sum_{r=1}^{n} E_{\theta_0}\left[\frac{\partial l_{\theta_0}}{\partial \theta_i} \frac{\partial l_{\theta_0}}{\partial \theta_j}\right]_{\theta=\theta_0} = -n^{-1} \sum_{r=1}^{n} E_{\theta_0}\left[\frac{\partial^2 l_{\theta_0}}{\partial \theta_i \partial \theta_j}\right] = F_{ij}(\theta_0).
\]

Because of assumptions B2(β) and B5, Proposition 1 (i) can now be applied on \( \sqrt{n}\hat{U}_{ij}(\theta_0) \), that is,
\[
\mathcal{L}\left(\sqrt{n}\hat{U}_{ij}(\theta_0)\right) \overset{w}{\longrightarrow} N(0, \Sigma_n(\theta_0)), \quad n \rightarrow \infty. \quad (47)
\]
If two sequences of d.l.’s are weakly approaching each other and one of them is uniformly tight, then the other one is also uniformly tight (see Lemma 4 in
Belyaev and Sjöstedt [5]). Then, since the right hand side of (47) is uniformly tight, we have that

\[ \sqrt{n} \mathcal{U}_{n}(\theta_0) = O_P(1), \quad n \to \infty. \]  

(48)

By considering equation (42) and using the relations (44), (45), (46) and (48), we are led to the conclusion that

\[ \sqrt{n}(\hat{\theta}_n - \theta_0) = O_P(1), \quad n \to \infty. \]  

(49)

It follows from relation (44) that

\[ \tilde{\pi}_n(\hat{\theta}_n, \theta_0) \sqrt{n}(\hat{\theta}_n - \theta_0) = O_P(1), \quad n \to \infty, \]  

(50)

and from relation (46) that

\[ \tilde{\pi}_n(\theta_0) \sqrt{n}(\hat{\theta}_n - \theta_0) = -\tilde{\pi}_n(\theta_0) \sqrt{n}(\hat{\theta}_n - \theta_0) + O_P(1), \quad n \to \infty. \]  

(51)

With help of relations (50) and (51), equation (42) can be rewritten as

\[ \tilde{\pi}_n(\theta_0) \sqrt{n}(\hat{\theta}_n - \theta_0) = \sqrt{n} \mathcal{U}_{n}(\theta_0) + O_P(1), \quad n \to \infty. \]  

(52)

Lemma 5 in Belyaev and Sjöstedt [5] now tells us that in spite of the presence of \( O_P(1) \), the d.l. of the right hand side of (52) is weakly approaching \( \mathcal{L}(\sqrt{n} \mathcal{U}_{n}(\theta_0)) \) as \( n \to \infty \). It can be proved under assumption B2(\( \beta \)) and B6 that \( \sup_n \| \tilde{\pi}_n \|_2 \) is bounded from above. Then, Lemma 6 and relation (47) give us the desired result and the proof is completed.

\[ \square \]

Proof of Theorem 6. This proof will be constructed in the same manner as the proof of Theorem 5. To shorten notation, we drop as above the arguments \( \mathcal{Y}_n \) and \( \mathcal{Y}_n \) of all the functions in Section 4.

The first order Taylor expansion of equation system (6), with \( \theta_n = \hat{\theta}_n \) about \( \theta_0 \), yields

\[ \sqrt{n} \mathcal{U}_{n}^* (\hat{\theta}_n) = \sqrt{n} \mathcal{U}_{n}^* (\theta_0) + \left( \tilde{\pi}_n^* (\theta_0) + \tilde{\pi}_n^* (\hat{\theta}_n, \theta_0) \right) \sqrt{n}(\hat{\theta}_n^* - \theta_0) = O_k, \]  

(53)

where

\[ \tilde{\pi}_n^* (\cdot) = n^{-1} \sum_{i=1}^{n} M_{r_n} \mathcal{Y}_n (\cdot), \quad \tilde{\pi}_n^* (\hat{\theta}_n^*, \theta_0) = \left( \tilde{\pi}_n^* (\hat{\theta}_n^*) - \tilde{\pi}_n^* (\theta_0) \right), \]

and \( \hat{\theta}_n^* = \theta_0 + \alpha_n^* (\hat{\theta}_n^* - \theta_0) \), for some \( \alpha_n^* = \alpha_n^*(\theta_0, \hat{\theta}_n^*) \in [0, 1] \).

Let us begin by proving that each element of the matrix \( \tilde{\mathcal{F}}_n^*(\hat{\theta}_n^*, \theta_0) \) is converging to zero, in \( P_{\theta_0}^{(+)} \)-probability as \( n \to \infty \). From assumption B4(\( \epsilon, \delta \)) with any arbitrarily small positive \( \epsilon \leq \epsilon(\delta, \theta_0) \), we have that

\[ E_{\theta_0}^*[|\tilde{\mathcal{F}}_n^*(\hat{\theta}_n^*, \theta_0)| I (\hat{\theta}_n^* \in \mathcal{B}(\theta_0))] \leq n^{-1} \sum_{r=1}^{n} E_{\theta_0}^*[M_{r_n} Q_{r_n}(\epsilon, \theta_0)] \]

\[ \leq n^{-1} \sum_{r=1}^{n} E_{\theta_0}^*[M_{r_n} E_{\theta_0}^*[Q_{r_n}(\epsilon, \theta_0)]] \leq n^{-1} \sum_{r=1}^{n} E_{\theta_0}^*[M_{r_n}^2] \delta^2 = \delta^2. \]

24
Then by Chebyshev’s inequality, we have that
\[
P_{\theta_0}^{(\ast)} \left( \left\{ \left| \xi^*_{ijn} (\hat{\theta}_n, \theta_0) \right| > \delta \right\} \cap \left\{ \hat{\theta}_n^* \in B_\varepsilon (\theta_0) \right\} \right) \leq \delta. \tag{54}
\]

Since \( \hat{\theta}_n^* = \theta_0 + \alpha_n^* (\hat{\theta}_n - \theta_0) \), and \( \hat{\theta}_n^* \) is a consistent estimator of \( \theta_0 \), we have for sufficiently large \( n \), say \( n \geq N(\epsilon, \delta) \), that
\[
P_{\theta_0}^{(\ast)} \left( \hat{\theta}_n^* \not\in B_\varepsilon (\theta_0) \right) \leq \delta.
\]

From this and relation (54) we have, by the same arguments as in the proof of Theorem 5, that
\[
P_{\theta_0}^{(\ast)} \left( \left| \xi^*_{ijn} (\hat{\theta}_n, \theta_0) \right| > \delta \right) \leq 2\delta.
\]

That is, each
\[
\xi^*_{ijn} (\hat{\theta}_n, \theta_0) \sim o_{P^{(\ast)}} (1), \quad n \to \infty. \tag{55}
\]

Let us now consider \( \Xi^*_n (\theta_0^* \). For each element of \( \Xi^*_n (\theta_0^* \), we have that
\[
E_{\theta_0}^{(\ast)} [M_{ij}^* H_{ij}(\theta_0)] = E_{\theta_0}^{(\ast)} \left[ M_{ij}^* E_{\theta_0} \left[ \frac{\partial^2 L_{ij}(\theta_0)}{\partial \theta_i \partial \theta_j} \right] \right] = -F_{ij}(\theta_0).
\]

Without loss of generality, let \( 0 < \beta \leq 1 \). Then by inequalities (16), (14), (45) and (30), we have that
\[
E_{\theta_0}^{(\ast)} \left[ n^{-1} \sum_{r=1}^n M_{ij}^* \left( H_{ij}(\theta_0) + F_{ij}(\theta_0) \right) \right]^{1+\beta} = E_{\theta_0}^{(\ast)} \left[ \left( \sum_{r=1}^n M_{ij}^* \left( H_{ij}(\theta_0) + F_{ij}(\theta_0) \right) \right)^{1+\beta} \right] \leq \left( \sum_{r=1}^n M_{ij}^* \right)^{1+\beta} E_{\theta_0} \left[ \left( \frac{\partial^2 L_{ij}(\theta_0)}{\partial \theta_i \partial \theta_j} + F_{ij}(\theta_0) \right) \right]^{1+\beta} \leq n^{-1+\beta} \sum_{r=1}^n M_{ij}^* \left( \frac{\partial^2 L_{ij}(\theta_0)}{\partial \theta_i \partial \theta_j} + F_{ij}(\theta_0) \right)^{1+\beta} \leq 2^{1+\beta} n^{-1+\beta} \sum_{r=1}^n M_{ij}^* \left( \frac{\partial^2 L_{ij}(\theta_0)}{\partial \theta_i \partial \theta_j} + F_{ij}(\theta_0) \right)^{1+\beta} \leq 2^{1+\beta} n^{-\beta} d_3 (\beta, \theta_0) \to 0, \quad n \to \infty.
\]

Applying Chebyshev’s inequality, it follows that, as in the proof of Theorem 5,
\[
\hat{H}_{ij}(\theta_0) + F_{ij}(\theta_0) = o_{P^{(\ast)}} (1), \quad n \to \infty. \tag{56}
\]

According to relations (55) and (56), equation (53) can now be rewritten as
\[
\sqrt{n} \xi^*_m (\theta_0) = (\Xi^*_n (\theta_0) + \Xi^*_n (\theta_0)) \sqrt{n} \left( \hat{\theta}_n^* - \theta_n + \hat{\theta}_n - \theta_0 \right), \tag{57}
\]

where each element of \( \Xi^*_n (\theta_0) := \Xi^*_n (\theta_0) - \Xi^*_n (\theta_0) + \Xi^*_n (\hat{\theta}_n^* - \theta_0) \) is \( o_{P^{(\ast)}} (1) \).
Since all assumptions of Theorem 5 are fulfilled, we have from relations (49) and (52) that (57) can be rewritten as
\[
\sqrt{n} \left( \bar{U}_n^n(\theta_0) - \bar{U}_n^n(\theta_0) \right) = (\bar{\pi}_n(\theta_0) + \bar{\pi}_n(\theta_0)) \sqrt{n} \left( \hat{\theta}_n - \hat{\theta}_n \right) + o_{\text{PP}}(1). \tag{58}
\]

Proposition 1 (ii) can be applied to the left hand side of (58) yielding
\[
\mathcal{L} \left( \sqrt{n} \left( \bar{U}_n^n(\theta_0) - \bar{U}_n^n(\theta_0) \right) \mid \mathcal{Y}_n \right) \overset{w}{\underset{n \to \infty}\longrightarrow} \mathcal{N}(0, \bar{\pi}_n(\theta_0)), \quad n \to \infty. \tag{59}
\]
If a sequence of d.l.’s and a sequence of conditional d.l.’s are weakly approaching each other in probability, and if the first sequence is uniformly tight, then the other is uniformly tight in probability (see Lemma 6 in Belyaev and Sjöstedt [5]). Then, since the right hand side of (59) is uniformly tight,
\[
\sqrt{n} \left( \bar{U}_n^n(\theta_0) - \bar{U}_n^n(\theta_0) \right) = O_{\text{PP}}(1), \quad n \to \infty. \tag{60}
\]

By considering equation (58) and using relations (45), (55), (56) and (60), we are led to the conclusion that
\[
\sqrt{n} \left( \hat{\theta}_n - \theta_0 \right) = O_{\text{PP}}(1), \quad n \to \infty,
\]
and it follows that
\[
\bar{\pi}_n(\theta_0) \sqrt{n} \left( \hat{\theta}_n - \theta_0 \right) = o_{\text{PP}}(1), \quad n \to \infty.
\]
Equation (58) can now be rewritten as
\[
\bar{\pi}_n(\theta_0) \sqrt{n} \left( \hat{\theta}_n - \hat{\theta}_n \right) = \sqrt{n} \left( \bar{U}_n^n(\theta_0) - \bar{U}_n^n(\theta_0) \right) + o_{\text{PP}}(1), \quad n \to \infty. \tag{61}
\]

Lemma 7 in Belyaev and Sjöstedt [5] now tells us that in spite of the presence of $o_{\text{PP}}(1)$, the conditional d.l. of the right hand side of (61), given $\mathcal{Y}_n$, is weakly approaching in probability, $\mathcal{L} \left( \sqrt{n} \left( \bar{U}_n^n(\theta_0) - \bar{U}_n^n(\theta_0) \right) \mid \mathcal{Y}_n \right)$ as $n \to \infty$. Finally, Lemma 6 and the relations (59) give us the desired result, and the proof is completed. \qed

References


