Statistics and Probability Letters 142 (2018) 109–117

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

A uniform L^1 law of large numbers for functions of i.i.d. random variables that are translated by a consistent estimator

Pierre Lafaye de Micheaux [a], Frédéric Ouimet [b,*]

[a] School of Mathematics and Statistics, UNSW Sydney, Australia
[b] Département de Mathématiques et de Statistique, Université de Montréal, Canada

ARTICLE INFO

Article history:
Received 2 March 2018
Received in revised form 22 May 2018
Accepted 10 June 2018
Available online 22 June 2018

MSC:
60F25

Keywords:
Uniform law of large numbers
Taylor expansion
M-estimators
Score function

ABSTRACT

We develop a new L^1 law of large numbers where the ith summand is given by a function $h(\cdot)$ evaluated at $X_i - \theta_n$, and where $\theta_n \doteq \theta_n(X_1, X_2, \ldots, X_n)$ is an estimator converging in probability to some parameter $\theta \in \mathbb{R}$. Under broad technical conditions, the convergence is shown to hold uniformly in the set of estimators interpolating between θ and another consistent estimator $\theta_n^\star$. Our main contribution is the treatment of the case where $|h|$ blows up at 0, which is not covered by standard uniform laws of large numbers.

1. Introduction

Let $X_1, X_2, X_3, \ldots$ be a sequence of i.i.d. random variables and consider the statistic $T_n(\theta_n^\star)$ where the random variable

$$T_n(\theta) \doteq T_n(X_1, X_2, \ldots, X_n; \theta) : \Omega \to \mathbb{R}$$

depends on an unknown parameter $\theta \in \mathbb{R}$ for which we have a consistent sequence of estimators $\theta_n^\star \doteq \theta_n^\star(X_1, X_2, \ldots, X_n)$. Assume further that the following first-order Taylor expansion is valid :

$$T_n(\theta_n^\star) = T_n(\theta) + (\theta_n^\star - \theta) \int_0^1 T_n'(\theta + v(\theta_n^\star - \theta))dv, \tag{1.1}$$

where

$$T_n'(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i \neq t\}} h(X_i - t), \tag{1.2}$$

and where $h : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ is a measurable function (possibly nonlinear). In statistics, one is often interested in knowing if estimating a parameter (θ here) has an impact on the asymptotic law of a given statistic. See for example the interesting results of de Wet and Randles (1987) in the context of limiting χ^2 U and V statistics. Eqs. (1.1) and (1.2) provide a natural setting for studying the question of whether or not $T_n(\theta_n^\star) - T_n(\theta) \to 0$ whenever $\theta_n^\star \to \theta$, as $n \to \infty$.

* Corresponding author.
E-mail address: ouimetfr@dms.umontreal.ca (F. Ouimet).

https://doi.org/10.1016/j.spl.2018.06.006

Given some regularity conditions on the behavior of $h(\cdot)$ around the origin and in its tails, proving the convergence to $\mathbb{E}[h(X_1-\theta)]$, in probability say, of the integral on the right-hand side of (1.1) is often possible under weak assumptions by adapting standard uniform laws of large numbers. For instance, one can use (Ferguson, 1996, Theorem 16(a)), which was introduced by LeCam (1953) and Rubin (1956). One can also use entropy conditions: see, e.g., van de Geer (2000, Chapter 3) and van der Vaart and Wellner (1996, Section 2.4). Some of these theorems go back to or evolved from the works of Blum (1955), Dehardt (1971), Vapnik and Červonenkis (1971, 1981), Giné and Zinn (1984), Pollard (1984) and Talagrand (1987). For extensive notes on the origins of the entropy conditions, we refer the interested reader to van de Geer (2000, Section 3.8) and Pollard (1984, pp. 36–38).

However, when $|h|$ blows up at 0, namely when $\limsup_{x\to 0}|h(x)| = \infty$, these results are not applicable because the envelope function $h^{\sup}(x) \doteq \sup_{t:|t-\theta|<\delta} \mathbf{1}_{\{x\neq t\}}|h(x-t)|$ is infinite in any small enough neighborhood of θ and, in particular, $h^{\sup}(X_1)$ is not integrable for the outer measure.

We faced such a problem when analyzing the convergence of score functions in the context of testing the goodness-of-fit of the Laplace distribution with unknown location and scale parameters (μ,σ). If the family of alternatives is taken to be the asymmetric power distribution (Komunjer, 2007) or the skewness exponential power distribution (Fernández et al., 1995), a score function evaluated at the maximum likelihood estimator $(\mu_n^\star,\sigma_n^\star)$ can be used, in the spirit of Desgagné et al. (2013) and Desgagné and Lafaye de Micheaux (2018). If the score function is expanded around (μ,σ), then a multivariate version of (1.1) is obtained. One of the integrals in the expansion will have an integrand (1.2) where $h(\cdot)$ contains a logarithmic term. Standard uniform laws of large numbers cannot be applied to show the convergence of such integrals because the envelope function of the class of functions $\{\log(\cdot - t)\}_{t:|t-\mu|<\delta}$ is infinite in any small enough neighborhood of μ. In Section 3, we show how the main result of this paper (Theorem 2.6) can be used to prove a crucial part of the problem described above.

More generally, the main result is that, under broad conditions, one obtains

$$\lim_{n\to\infty}\sup_{v\in[0,1]}\mathbb{E}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\{X_i\neq\theta+v(\theta_n^\star-\theta)\}}h(X_i-\theta-v(\theta_n^\star-\theta))-\mathbb{E}\big[h(X_1-\theta)\big]\right| = 0. \tag{1.3}$$

From (1.3) and the setting above, one can conclude that $T_n(\theta_n^\star)-T_n(\theta)\to 0$ in probability as $n\to\infty$.

2. A new uniform L^1 law of large numbers

Throughout the paper, the labels $(X.k)$, $(H.k)$ and $(E.k)$ denote, respectively, assumptions that we will make on X_1, $h(\cdot)$ and θ_n. Fig. 2.1 at the end of the current section illustrates the logical structure of these assumptions and their implications. We start by proving a non-uniform version of Theorem 2.6.

Proposition 2.1. *Let $\theta\in\mathbb{R}$ and let $X_1, X_2, X_3, \ldots$ be a sequence of i.i.d. random variables such that*

(X.1) $\mathbb{P}(X_1=\theta)=0.$

Let $h:\mathbb{R}\setminus\{0\}\to\mathbb{R}$ be a measurable function that satisfies

(H.1) $\mathbb{P}(X_1-\theta\in\mathcal{D}_h)=0$, *where $\mathcal{D}_h$ is the set of discontinuity points of $h(\cdot)$,*
(H.2) $\mathbb{E}\,|h(X_1-\theta)|<\infty.$

Let $\theta_n \doteq \theta_n(X_1,X_2,\ldots,X_n)$ be an estimator that satisfies

(E.1) $\theta_n \xrightarrow{\mathbb{P}} \theta,$
(E.2) *For all $n\in\mathbb{N}$ and all $i\in\{1,2,\ldots,n\}$, $(X_i-\theta_n, X_i-\theta) \overset{law}{=} (X_1-\theta_n, X_1-\theta)$,*
(E.3) *There exists $N_0\in\mathbb{N}$ such that $\big\{\mathbf{1}_{\{X_1\neq\theta_n\}}h(X_1-\theta_n)\big\}_{n\geq N_0}$ is uniformly integrable.*

Then,

$$\mathbb{E}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\{X_i\neq\theta_n\}}h(X_i-\theta_n)-\mathbb{E}\big[h(X_1-\theta)\big]\right| \longrightarrow 0. \tag{2.1}$$

Remark 2.2. Condition (E.2) is satisfied for any estimator that is symmetric with respect to its n variables. For example, this is the case for any maximum likelihood estimator that is based on i.i.d. observations.

Proof of Proposition 2.1. From (X.1) and (E.1), we know that $\mathbf{1}_{\{X_1=\theta_n\}}\xrightarrow{\mathbb{P}} 0$. Indeed, for any $\varepsilon>0$,

- take $\delta\doteq\delta_\varepsilon>0$ such that $\mathbb{P}(|X_1-\theta|<\delta)<\varepsilon/2$, and
- take $N\doteq N_{\delta,\varepsilon}$ such that for all $n\geq N$, we have $\mathbb{P}(|\theta_n-\theta|\geq\delta)<\varepsilon/2$.

We get, for all $n\geq N$,

$$\mathbb{P}(X_1=\theta_n)\leq\mathbb{P}(X_1=\theta_n, |\theta_n-\theta|<\delta)+\mathbb{P}(|\theta_n-\theta|\geq\delta)<\varepsilon.$$

In particular, this shows $\mathbf{1}_{\{X_1=\theta_n\}}|h(X_1-\theta)| \xrightarrow{\mathbb{P}} 0$. Since this sequence is uniformly integrable by (H.2), we also have the L^1 convergence. By using Jensen's inequality and (E.2), we deduce

$$\mathbb{E}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\{X_i=\theta_n\}}h(X_i-\theta)\right| \leq \mathbb{E}\left[\mathbf{1}_{\{X_1=\theta_n\}}|h(X_1-\theta)|\right] \longrightarrow 0. \tag{2.2}$$

By (H.2) and the law of large numbers in L^1 (see, e.g., Theorem 1.2.6 in Stroock (2011)), we also know that

$$\mathbb{E}\left|\frac{1}{n}\sum_{i=1}^{n}h(X_i-\theta) - \mathbb{E}\left[h(X_1-\theta)\right]\right| \longrightarrow 0. \tag{2.3}$$

By combining (2.2) and (2.3), we have shown

$$\mathbb{E}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\{X_i\neq\theta_n\}}h(X_i-\theta) - \mathbb{E}\left[h(X_1-\theta)\right]\right| \longrightarrow 0. \tag{2.4}$$

To conclude the proof, we show that

$$Y_n \mathrel{\hat{=}} \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\{X_i\neq\theta_n\}}h(X_i-\theta_n) - \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\{X_i\neq\theta_n\}}h(X_i-\theta) \xrightarrow{L^1} 0.$$

From Jensen's inequality and (E.2), we have

$$\mathbb{E}|Y_n| \leq \mathbb{E}\left[\mathbf{1}_{\{X_1\neq\theta_n\}}\left|h(X_1-\theta_n)-h(X_1-\theta)\right|\right]. \tag{2.5}$$

The sequence $\{\mathbf{1}_{\{X_1\neq\theta_n\}}|h(X_1-\theta_n)-h(X_1-\theta)|\}_{n\in\mathbb{N}}$ converges to 0 in probability by (H.1), (E.1) and the continuous mapping theorem (van der Vaart, 1998, Theorem 2.3). Furthermore, the sequence is uniformly integrable for $n \geq N_0$ by (H.2), (E.3) and the fact that the sums of random variables coming (respectively) from two uniformly integrable sequences form a uniformly integrable sequence. Hence, $Y_n \to 0$ in L^1. □

Since the distribution of $X_1-\theta_n$ is rarely known, condition (E.3) in Proposition 2.1 is impractical to verify. The next lemma fix this problem.

Lemma 2.3. *Let $\theta \in \mathbb{R}$. Let $X_1, X_2, X_3, \ldots$ be a sequence of i.i.d. random variables. Let $h : \mathbb{R}\setminus\{0\} \to \mathbb{R}$ be a measurable function. Let $\theta_n \mathrel{\hat{=}} \theta_n(X_1, X_2, \ldots, X_n)$ be an estimator that satisfies*

(E.4) *If $\limsup_{x\to 0}|h(x)| < \infty$, we impose no condition. Otherwise, assume that there exist $N_1 \in \mathbb{N}$, $\alpha_0 > 0$ and a constant $C_{\alpha_0} > 0$ such that*

$$\sup_{n\geq N_1}\ \sup_{A\in\mathcal{B}_{>0}([-\alpha_0,\alpha_0])} \frac{\mathbb{P}(X_1-\theta_n\in A)}{Lebesgue(A)} \leq C_{\alpha_0} < \infty,$$

where $\mathcal{B}_{>0}([-\alpha_0,\alpha_0])$ denotes the Borel sets of positive Lebesgue measure on the interval $[-\alpha_0,\alpha_0]$.

(E.5) *There exist $N_2 \geq 2$, $C, \gamma, p > 0$ and $\beta_0 > \gamma$ such that, for $\mathbb{P}(X_1-\theta\in\cdot)$-almost-all $x\in\mathbb{R}$, we have*

- *For all $u \geq (x+\gamma)\vee\beta_0$ and for all $n \geq N_2$,*
  $$\mathbb{P}(\theta_n-\theta\leq x-u\,|\,X_1-\theta=x) \leq Ce^{-|x-u|^p}.$$
- *For all $u \leq (x-\gamma)\wedge(-\beta_0)$ and for all $n \geq N_2$,*
  $$\mathbb{P}(\theta_n-\theta\geq x-u\,|\,X_1-\theta=x) \leq Ce^{-|x-u|^p}.$$

(E.6) *There exists $N_3 \in \mathbb{N}$ such that for all $n \geq N_3$, there exists $A_n \in \mathcal{B}(\mathbb{R})$ such that $\mathbb{P}(X_1-\theta\in A_n) = 1$ and, for all $x \in A_n$, the conditional measure $\mathbb{P}(x-(\theta_n-\theta)\in\cdot\,|\,X_1-\theta=x)$, when restricted to $\{u\in\mathbb{R} : |u|\geq\beta_0, |x-u|>\gamma\}$, is absolutely continuous with respect to the Lebesgue measure.*

Assume that $h(\cdot)$ satisfies

(H.3) *For all $x_0 \in \mathbb{R}\setminus\{0\}$, $\limsup_{x\to x_0}|h(x)| < \infty$,*
(H.4) *$\int_{|u|\leq\alpha_0}|h(u)|du < \infty$,*
(H.5) **1.** *$h(\cdot)$ is absolutely continuous on bounded sub-intervals of $(-\infty,-\beta_0)\cup(\beta_0,+\infty)$;*
2. *There exists an integrable random variable M such that $\sup_{|t|\leq\gamma}|h(X_1-\theta-t)|\,\mathbf{1}_{\{|X_1-\theta-t|\geq\beta_0\}} \leq M$ $\mathbb{P}$-almost-surely;*
3. *$\lim_{|\beta|\to\infty}|h(\beta)|e^{-|x-\beta|^p} = 0$ for $\mathbb{P}(X_1-\theta\in\cdot)$-almost-all $x\in\mathbb{R}$, and $\{|h(\beta)|e^{-|X_1-\theta-\beta|^p}\}_{|\beta|\geq\beta_0}$ is uniformly integrable;*
4. *$\int_{|u|\geq\beta_0}\mathbb{E}\left[|h'(u)|\,e^{-|X_1-\theta-u|^p}\right]du < \infty$;*
5. *For almost-all $|u|\geq\beta_0$, we have $-\mathrm{sign}(u)\mathrm{sign}(h(u))h'(u) \leq 0$.*

Then, $(E.3)$ from Proposition 2.1 is satisfied, namely $\left\{\mathbf{1}_{\{X_1\neq\theta_n\}}h(X_1-\theta_n)\right\}_{n\geq N_0}$ is uniformly integrable, where $N_0 \mathrel{\hat{=}} N_1\vee N_2\vee N_3$.

Remark 2.4. If $X_1 - \theta_n$ has a density for n large enough and, in a neighborhood of 0, those densities are uniformly bounded from above by the same positive constant, then (E.4) is satisfied. In general, when θ_n is even only slightly non-trivial, we rarely know the distribution of $X_1 - \theta_n$. However, if θ_n concentrates more and more around θ as $n \to \infty$ (like most maximum likelihood estimators for instance), then we expect the weight of the distribution of X_1 around θ to dominate the weight of the distribution of $X_1 - \theta_n$ around 0. In that case, we can expect (E.4) to be satisfied when X_1 has a regular enough distribution around θ. Condition (E.5) is a way to control the tail behavior of θ_n's distribution for the above heuristic to work. Since the lemma is intended to be used when $|h|$ blows up at 0, condition (E.4) is there to control the distribution of $X_1 - \theta_n$ around 0.

Proof. We want to prove that for $N_0 \doteq N_1 \vee N_2 \vee N_3$, we have

$$\lim_{K\to\infty} \sup_{n\geq N_0} \mathbb{E}\Big[\big|h(X_1-\theta_n)\big|\, \mathbf{1}_{\{X_1\neq\theta_n\}\cap\{|h(X_1-\theta_n)|\geq K\}}\Big] = 0.$$

By (H.3), $h(\cdot)$ is uniformly bounded on compact subsets of $\mathbb{R}\setminus\{0\}$. It is therefore sufficient to show both

$$\lim_{\alpha\to 0} \sup_{n\geq N_0} \mathbb{E}\Big[\big|h(X_1-\theta_n)\big|\, \mathbf{1}_{\{X_1\neq\theta_n\}\cap\{|X_1-\theta_n|\leq\alpha\}}\Big] = 0, \tag{2.6}$$

$$\lim_{\beta\to\infty} \sup_{n\geq N_0} \mathbb{E}\Big[\big|h(X_1-\theta_n)\big|\, \mathbf{1}_{\{|X_1-\theta_n|\geq\beta\}}\Big] = 0. \tag{2.7}$$

When $\limsup_{x\to 0}|h(x)| < \infty$, then (2.6) is satisfied because $h(\cdot)$ is uniformly bounded on compact subsets of $\mathbb{R}$ by (H.3). When $\limsup_{x\to 0}|h(x)| = \infty$, then (2.6) follows directly from (E.4), (H.4) and the dominated convergence theorem (DCT).

Assume for the remaining of the proof that

$$n \geq N_0 \quad \text{and} \quad \beta > \beta_0 > \gamma,$$

where γ and β_0 are fixed in (E.5). Separate the expectation in (2.7) in two parts :

$$(a)+(b) \doteq \mathbb{E}\Big[\big|h(X_1-\theta_n)\big|\, \mathbf{1}_{\{|X_1-\theta_n|\geq\beta\}\cap\{|\theta_n-\theta|\leq\gamma\}}\Big] + \mathbb{E}\Big[\big|h(X_1-\theta_n)\big|\, \mathbf{1}_{\{|X_1-\theta_n|\geq\beta\}\cap\{|\theta_n-\theta|>\gamma\}}\Big].$$

By (H.5).2 and the DCT, we have $(a) \to 0$ as $\beta \to \infty$, uniformly in n. For the term (b), condition on the value of $X_1-\theta$, integrate by parts (see (E.6) and (H.5).1) and then use (E.5) and (H.5).5. We obtain

$$\begin{aligned}
(b) &= \int_{\{(u,x)\,:\,|u|\geq\beta,\ |x-u|>\gamma\}} |h(u)|\, \mathbb{P}((X_1-\theta_n, X_1-\theta)\in d(u,x)) \\
&= \int_{-\infty}^{\infty} \left(\int_{\{u\,:\,|u|\geq\beta,\ |x-u|>\gamma\}} |h(u)|\, \mathbb{P}(x-(\theta_n-\theta)\in du \,|\, X_1-\theta=x)\right) \mathbb{P}(X_1-\theta\in dx) \\
&= \int_{-\infty}^{-(\beta+\gamma)} \left\{\begin{array}{l} \Big[-|h(u)|\,\mathbb{P}(\theta_n-\theta\leq x-u\,|\,X_1-\theta=x)\Big]\Big|_{u=x+\gamma}^{-\beta} \\ +\displaystyle\int_{x+\gamma}^{-\beta} \mathrm{sign}(h(u))\,h'(u)\,\mathbb{P}(\theta_n-\theta\leq x-u\,|\,X_1-\theta=x)\,du \end{array}\right\} \mathbb{P}(X_1-\theta\in dx) \\
&\quad + \int_{-\infty}^{\infty} \lim_{t\to\infty} \left\{\begin{array}{l} \Big[-|h(u)|\,\mathbb{P}(\theta_n-\theta\leq x-u\,|\,X_1-\theta=x)\Big]\Big|_{u=(x+\gamma)\vee\beta}^{t} \\ +\displaystyle\int_{(x+\gamma)\vee\beta}^{t} \mathrm{sign}(h(u))\,h'(u)\,\mathbb{P}(\theta_n-\theta\leq x-u\,|\,X_1-\theta=x)\,du \\ +\Big[|h(u)|\,\mathbb{P}(\theta_n-\theta\geq x-u\,|\,X_1-\theta=x)\Big]\Big|_{u=-t}^{(x-\gamma)\wedge(-\beta)} \\ -\displaystyle\int_{-t}^{(x-\gamma)\wedge(-\beta)} \mathrm{sign}(h(u))\,h'(u)\,\mathbb{P}(\theta_n-\theta\geq x-u\,|\,X_1-\theta=x)\,du \end{array}\right\} \mathbb{P}(X_1-\theta\in dx) \\
&\quad + \int_{\beta+\gamma}^{\infty} \left\{\begin{array}{l} \Big[|h(u)|\,\mathbb{P}(\theta_n-\theta\geq x-u\,|\,X_1-\theta=x)\Big]\Big|_{u=\beta}^{x-\gamma} \\ -\displaystyle\int_{\beta}^{x-\gamma} \mathrm{sign}(h(u))\,h'(u)\,\mathbb{P}(\theta_n-\theta\geq x-u\,|\,X_1-\theta=x)\,du \end{array}\right\} \mathbb{P}(X_1-\theta\in dx) \\
&\leq \int_{-\infty}^{-(\beta+\gamma)} \big\{|h(x+\gamma)|+0\big\}\, \mathbb{P}(X_1-\theta\in dx) \\
&\quad + C\int_{-\infty}^{\infty} \left\{\begin{array}{l} |h((x+\gamma)\vee\beta)|\,e^{-|x-((x+\gamma)\vee\beta)|^p} + \displaystyle\int_{\beta}^{\infty} |h'(u)|\,e^{-|x-u|^p}du \\ |h((x-\gamma)\wedge(-\beta))|\,e^{-|x-((x-\gamma)\wedge(-\beta))|^p} + \displaystyle\int_{-\infty}^{-\beta} |h'(u)|\,e^{-|x-u|^p}du \end{array}\right\} \mathbb{P}(X_1-\theta\in dx) \\
&\quad + \int_{\beta+\gamma}^{\infty} \big\{|h(x-\gamma)|+0\big\}\, \mathbb{P}(X_1-\theta\in dx)
\end{aligned}$$

$$\lesssim \mathbb{E}\Big[|h(X_1-\theta+\gamma)|\mathbf{1}_{\{|X_1-\theta+\gamma|\geq\beta\}}\Big] + \mathbb{E}\Big[|h(\beta)|\,e^{-|X_1-\theta-\beta|^p}\Big] + \int_\beta^\infty \mathbb{E}\Big[|h'(u)|\,e^{-|X_1-\theta-u|^p}\Big]du$$
$$+\ \mathbb{E}\Big[|h(X_1-\theta-\gamma)|\mathbf{1}_{\{|X_1-\theta-\gamma|\geq\beta\}}\Big] + \mathbb{E}\Big[|h(-\beta)|\,e^{-|X_1-\theta+\beta|^p}\Big] + \int_{-\infty}^{-\beta} \mathbb{E}\Big[|h'(u)|\,e^{-|X_1-\theta-u|^p}\Big]du,$$

where $y \lesssim z$ means $y \leq (1 \vee C)z$. As $\beta \to \infty$, the first and fourth terms go to 0 by (H.5).2 and the DCT, the second and fifth terms go to 0 by (H.5).3 and the DCT, the third and sixth terms go to 0 by (H.5).4 and the DCT. None of the terms depended on n, so the convergence is uniform in $n \geq N_0$. □

If $\{\theta_n^\star\}_{n\in\mathbb{N}}$ is a sequence of M-estimators, then the next lemma proposes an easy-to-verify condition on the tail probabilities of $\theta_n^\star$ for (E.5) in Lemma 2.3 to hold uniformly in the set of estimators

$$\mathcal{E}_{n,\theta} \mathrel{\hat{=}} \{\theta + v(\theta_n^\star - \theta)\}_{v\in[0,1]}, \quad \text{for some } \theta \in \mathbb{R}. \tag{2.8}$$

Lemma 2.5. *Let $\theta \in \mathbb{R}$ and let $X_1, X_2, X_3, \ldots$ be a sequence of i.i.d. random variables. Let $\{\theta_n^\star\}_{n\in\mathbb{N}}$ be a sequence of estimators satisfying*

$$\sum_{i=1}^n \psi(X_i - \theta_n^\star) = 0, \tag{2.9}$$

where $\psi : \mathbb{R} \to \mathbb{R}$ is measurable, non-decreasing and $\psi(0) = 0$. Assume that there exist $N \geq 1$ and $C, \gamma, p > 0$ such that

$$\sup_{n\geq N} \mathbb{P}\big(|\theta_n^\star - \theta| \geq |t|\big) \leq Ce^{-|t|^p}, \quad \textit{for all } |t| \geq \gamma. \tag{2.10}$$

Then, condition (E.5) from Lemma 2.3 is satisfied uniformly on $\mathcal{E}_{n,\theta}$, namely :

(E.5.unif) *There exist $N_2 \geq 2$, $C, \gamma, p > 0$ and $\beta_0 > \gamma$ such that, for $\mathbb{P}(X_1 - \theta \in \cdot\,)$-almost-all $x \in \mathbb{R}$, we have*

- *For all $u \geq (x+\gamma) \vee \beta_0$ and for all $n \geq N_2$,*
$$\sup_{\theta_n \in \mathcal{E}_{n,\theta}} \mathbb{P}(\theta_n - \theta \leq x - u \,|\, X_1 - \theta = x) \leq Ce^{-|x-u|^p}.$$
- *For all $u \leq (x-\gamma) \wedge (-\beta_0)$ and for all $n \geq N_2$,*
$$\sup_{\theta_n \in \mathcal{E}_{n,\theta}} \mathbb{P}(\theta_n - \theta \geq x - u \,|\, X_1 - \theta = x) \leq Ce^{-|x-u|^p}.$$

Proof. For all $n \geq 2$, let $\theta_{2:n}^\star \mathrel{\hat{=}} \theta_{2:n}^\star(X_2, X_3, \ldots, X_n)$ be an estimator that satisfies

$$\sum_{i=2}^n \psi(X_i - \theta_{2:n}^\star) = 0 \quad \text{and} \quad \theta_{2:n}^\star \overset{\text{law}}{=} \theta_{n-1}^\star. \tag{2.11}$$

Since ψ is non-decreasing and $\psi(0) = 0$,

$$\bullet \quad \theta_n^\star \leq X_1 \implies \psi(X_1 - \theta_n^\star) \geq 0 \overset{(2.9)}{\implies} \sum_{i=2}^n \psi(X_i - \theta_n^\star) \leq 0 \overset{(2.11)}{\implies} \theta_{2:n}^\star \leq \theta_n^\star \leq X_1, \tag{2.12}$$

$$\bullet \quad \theta_n^\star \geq X_1 \implies \psi(X_1 - \theta_n^\star) \leq 0 \overset{(2.9)}{\implies} \sum_{i=2}^n \psi(X_i - \theta_n^\star) \geq 0 \overset{(2.11)}{\implies} \theta_{2:n}^\star \geq \theta_n^\star \geq X_1. \tag{2.13}$$

Let $\theta_n \in \mathcal{E}_{n,\theta}$ for all $n \in \mathbb{N}$. In order to prove (2.14) (respectively (2.15)), we use the following facts in succession : $\theta_n - \theta \leq 0 \implies \theta_n^\star - \theta \leq \theta_n - \theta$ (respectively $\theta_n - \theta \geq 0 \implies \theta_n^\star - \theta \geq \theta_n - \theta$), (2.12) (respectively (2.13)), the independence between X_1 and $\theta_{2:n}^\star$, (2.11), and (2.10).

- For all $u \geq (x+\gamma) \vee \beta_0 > 0$ (note that $x - u \leq -\gamma < 0$) and for all $n \geq N+1$, we have

$$\begin{aligned}
\mathbb{P}(\theta_n - \theta \leq x - u \,|\, X_1 - \theta = x) &\leq \mathbb{P}(\theta_n^\star - \theta \leq x - u \,|\, X_1 - \theta = x) \\
&\leq \mathbb{P}(\theta_{2:n}^\star - \theta \leq x - u) \\
&= \mathbb{P}(\theta_{n-1}^\star - \theta \leq x - u) \\
&\leq Ce^{-|x-u|^p}.
\end{aligned} \tag{2.14}$$

- For all $u \le (x-\gamma) \wedge (-\beta_0) < 0$ (note that $x - u \ge \gamma > 0$) and for all $n \ge N+1$, we have

$$\begin{aligned}\mathbb{P}(\theta_n - \theta \ge x - u \,|\, X_1 - \theta = x) &\le \mathbb{P}(\theta_n^\star - \theta \ge x - u \,|\, X_1 - \theta = x)\\ &\le \mathbb{P}(\theta_{2:n}^\star - \theta \ge x - u)\\ &= \mathbb{P}(\theta_{n-1}^\star - \theta \ge x - u)\\ &\le Ce^{-|x-u|^p}.\end{aligned} \tag{2.15}$$

Simply choose $N_2 \triangleq N+1$ in (E.5.unif). This ends the proof. □

We can now state the main result. The structure of the assumptions is illustrated in Fig. 2.1.

Theorem 2.6. *Let $\theta \in \mathbb{R}$ and let $X_1, X_2, X_3, \ldots$ be a sequence of i.i.d. random variables satisfying*

(X.1) $\mathbb{P}(X_1 = \theta) = 0$.

Let $\{\theta_n^\star\}_{n\in\mathbb{N}}$ be a sequence of estimators satisfying **(E.5.unif)** *directly or the conditions in Lemma 2.5. Denote*

$$\mathcal{E}_{n,\theta} \triangleq \{\theta + v(\theta_n^\star - \theta)\}_{v\in[0,1]},$$

and assume that

(E.1.unif) $\theta_n^\star \xrightarrow{\mathbb{P}} \theta$;

(E.2.unif) *For all $n \in \mathbb{N}$, all $i \in \{1,2,\ldots,n\}$ and all $\theta_n \in \mathcal{E}_{n,\theta}$, $(X_i - \theta_n, X_i - \theta) \overset{law}{=} (X_1 - \theta_n, X_1 - \theta)$;*

(E.4.unif) *If $\limsup_{x\to 0}|h(x)| < \infty$, we impose no condition. Otherwise, assume that there exist $N_1 \in \mathbb{N}$, $\alpha_0 > 0$ and a constant $C_{\alpha_0} > 0$ such that*

$$\sup_{n\ge N_1} \sup_{\theta_n\in\mathcal{E}_{n,\theta}} \sup_{A\in\mathcal{B}_{>0}([-\alpha_0,\alpha_0])} \frac{\mathbb{P}(X_1 - \theta_n \in A)}{Lebesgue(A)} \le C_{\alpha_0} < \infty.$$

(E.6.unif) *There exists $N_3 \in \mathbb{N}$ such that for all $n \ge N_3$ and for all $\theta_n \in \mathcal{E}_{n,\theta}$, there exists $A_{n,\theta_n} \in \mathcal{B}(\mathbb{R})$ such that $\mathbb{P}(X_1 - \theta \in A_{n,\theta_n}) = 1$ and, for all $x \in A_{n,\theta_n}$, the conditional measure $\mathbb{P}(x - (\theta_n - \theta) \in \cdot \,|\, X_1 - \theta = x)$, when restricted to $\{u \in \mathbb{R} : |u| \ge \beta_0, |x-u| > \gamma\}$, is absolutely continuous with respect to the Lebesgue measure.*

Finally, assume

(H.1), **(H.2)** *from Proposition 2.1,*

(H.3), **(H.4)**, **(H.5)** *from Lemma 2.3.*

Then, the conclusion in Proposition 2.1 holds uniformly for $\theta_n \in \mathcal{E}_{n,\theta}$, namely

$$\lim_{n\to\infty} \sup_{\theta_n\in\mathcal{E}_{n,\theta}} \mathbb{E}\left|\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i\neq\theta_n\}} h(X_i - \theta_n) - \mathbb{E}\big[h(X_1 - \theta)\big]\right| = 0. \tag{2.16}$$

Proof. We know that (E.5.unif) holds, either directly or via the conditions in Lemma 2.5. By combining (E.4.unif) to (E.6.unif) and (H.3) to (H.5), a proof along the lines of Lemma 2.3 shows

(E.3.unif)

$$\lim_{K\to\infty} \sup_{n\ge N_0} \sup_{\theta_n\in\mathcal{E}_{n,\theta}} \mathbb{E}\left[\big|h(X_1 - \theta_n)\big|\, \mathbf{1}_{\{X_1\neq\theta_n\}\cap\{|h(X_1-\theta_n)|\ge K\}}\right] = 0.$$

By (E.3.unif), (H.2) and the identity $|U_n + V_n|\mathbf{1}_{\{|U_n+V_n|\ge 2K\}} \le 2|U_n|\mathbf{1}_{\{|U_n|\ge K\}} + 2|V_n|\mathbf{1}_{\{|V_n|\ge K\}}$, we deduce

$$\lim_{K\to\infty} \sup_{n\ge N_0} \sup_{\theta_n\in\mathcal{E}_{n,\theta}} \mathbb{E}\left[\big|h(X_1 - \theta_n) - h(X_1 - \theta)\big| \mathbf{1}_{\{X_1\neq\theta_n\}\cap\{|h(X_1-\theta_n)-h(X_1-\theta)|\ge K\}}\right] = 0. \tag{2.17}$$

To conclude, we rerun the proof of Proposition 2.1 with our new assumptions. By (X.1), (H.2), (E.1.unif) and (E.2.unif), the convergence in (2.2) is valid for $\sup_{\theta_n\in\mathcal{E}_{n,\theta}}$ of the expectation. This implies that the convergence in (2.4) is also valid for $\sup_{\theta_n\in\mathcal{E}_{n,\theta}}$ of the expectation. Furthermore, by (H.1), (E.1.unif) and the continuous mapping theorem, we have, for all $\varepsilon > 0$,

$$\lim_{n\to\infty} \sup_{\theta_n\in\mathcal{E}_{n,\theta}} \mathbb{P}\Big(\mathbf{1}_{\{X_1\neq\theta_n\}}\big|h(X_1 - \theta_n) - h(X_1 - \theta)\big| > \varepsilon\Big) = 0. \tag{2.18}$$

By combining (2.17) and (2.18), the $\sup_{\theta_n\in\mathcal{E}_{n,\theta}}$ of the expectation on the right-hand side of (2.5) converges to 0. In summary, we have shown that $\sup_{\theta_n\in\mathcal{E}_{n,\theta}}$ of the expectations in (2.2), (2.4) and (2.5) all converge (respectively) to 0. Hence, the conclusion of Proposition 2.1 holds for $\sup_{\theta_n\in\mathcal{E}_{n,\theta}}$ of the expectation, which is exactly the claim made in (2.16). □

Remark 2.7. By following the proof of Theorem 2.6, we see that (X.1), (H.1), (H.2), (E.1.unif), (E.2.unif) and (E.3.unif) alone imply the conclusion in (2.16). The other assumptions in the statement of the theorem are simply there to give a more practical way to verify (E.3.unif).

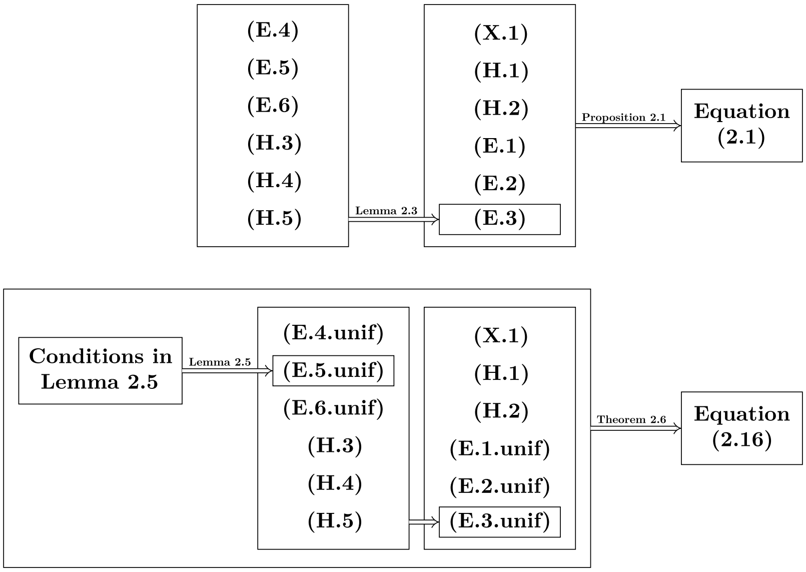

Fig. 2.1. Logical structure of the assumptions and their implications.

3. Example

We now give an application of the previous theorem. The context of the problem is described at the end of Section 1.

Lemma 3.1. *Let $X_1, X_2, X_3, \dots$ be a sequence of i.i.d. random variables with density function*

$$f_{X_1}(x) \triangleq \frac{1}{4\sigma} e^{-\frac{1}{2}\left|\frac{x-\mu}{\sigma}\right|}, \quad x \in \mathbb{R},$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$. Define $h : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ by

$$h(y) \triangleq sign(y) \log|y|.$$

Let

$$\mu_n^\star \triangleq median(X_1, X_2, \dots, X_n) \triangleq \begin{cases} X_{((n+1)/2)}, & \text{if } n \text{ is odd}, \\ \frac{1}{2}(X_{(n/2)} + X_{(n/2+1)}), & \text{if } n \text{ is even}. \end{cases} \tag{3.1}$$

For $v \in [0, 1]$, define $\mu_{n,v}^\star \triangleq \mu + v(\mu_n^\star - \mu)$, and let $\mathcal{E}_{n,\mu} \triangleq \{\mu_{n,v}^\star\}_{v\in[0,1]}$. Then,

$$\lim_{n\to\infty} \sup_{v\in[0,1]} \mathbb{E}\left|\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{X_i \neq \mu_{n,v}^\star\}} h(X_i - \mu_{n,v}^\star) - \mathbb{E}\left[h(X_1 - \mu)\right]\right| = 0. \tag{3.2}$$

Proof. Without loss of generality, assume that $\mu = 0$. Below, we verify the conditions of Theorem 2.6.

(X.1) $\mathbb{P}(X_1 = 0) = 0$. This is obvious.

(Conditions in Lemma 2.5) We show that the conditions are satisfied with $\psi(y) \triangleq \text{sign}(y)$ and $\psi(0) \triangleq 0$. Indeed, by (3.1), we know that $\sum_{i=1}^n \psi(X_i - \mu_n^\star) = 0$. Furthermore, for $N \in \mathbb{N}$ and $\gamma > 0$ both large enough (depending on σ), we have, for all $n \geq N$ and all $t \geq \gamma$,

$$\begin{aligned} \mathbb{P}(\mu_n^\star \geq t) &\leq \sum_{k=\lceil n/2\rceil}^n \binom{n}{k} \mathbb{P}(X_1 \geq t)^k \, \mathbb{P}(X_1 \leq t)^{n-k} \\ &\leq (n - \lceil n/2\rceil) \cdot \binom{n}{\lceil n/2\rceil} \cdot \mathbb{P}(X_1 \geq t)^{\lceil n/2\rceil} \\ &\leq \lfloor n/2\rfloor \cdot 2\frac{2^n}{\sqrt{n}} \cdot \left(\frac{1}{2}e^{-\frac{t}{2\sigma}}\right)^{\lceil n/2\rceil} \leq \frac{\sqrt{n}}{2} 2^n e^{-\frac{nt}{8\sigma}} \cdot e^{-\frac{nt}{8\sigma}} \leq \frac{1}{2}e^{-t}. \end{aligned} \tag{3.3}$$

To obtain the third inequality, we use Stirling's formula and assume that N is large enough. To obtain the last inequality, assume that $N \geq 8\sigma$ and $\gamma \geq 8\sigma$. This proves (2.10) with $C = 1$ and $p = 1$.

(E.1.unif) $\mu_n^\star \xrightarrow{\mathbb{P}} 0$. This is explained in Example 5.11 of van der Vaart (1998).

(E.2.unif) For any $v \in [0, 1]$, the estimator $\mu_{n,v}^\star = v\mu_n^\star$ is symmetric with respect to its n variables because the median, $\mu_n^\star$, is symmetric with respect to its n variables. Since the X_i's are i.i.d., the condition is satisfied.

(E.4.unif) We have $\limsup_{x\to 0}|h(x)| = \infty$, so we need to verify the condition. For any $n \geq 2$ and any $v \in [0, 1]$, note that $X_1 - v\mu_n^\star$ has a density function. It suffices to show that the densities are bounded, uniformly in n and v, by a positive constant. Since the density $u \mapsto f_{X_1 - v\mu_n^\star}(u)$ is symmetric around 0, we will assume, without loss of generality, that $u > 0$. For $v \in (0, 1]$, denote $z \triangleq (x - u)/v$ and notice that $z < x$.
When $v \in (0, 1]$ and $n \geq 3$ is odd, we have

$$\begin{aligned} f_{X_1 - v\mu_n^\star}(u) &= \int_{-\infty}^{\infty} f_{X_1 - v\mu_n^\star | X_1}(u \,|\, x) f_{X_1}(x) dx = \int_{-\infty}^{\infty} \frac{1}{v} f_{\mu_n^\star | X_1}(z \,|\, x) f_{X_1}(x) dx \\ &= \int_{-\infty}^{\infty} \frac{1}{v} \binom{n}{\lfloor n/2 \rfloor} (F_{X_1}(z))^{\lfloor n/2 \rfloor} f_{X_1}(z) (1 - F_{X_1}(z))^{\lfloor n/2 - 1 \rfloor} f_{X_1}(x) dx \\ &\leq C\, \|f_{X_1}\|_\infty \underbrace{\int_{-\infty}^{\infty} \frac{1}{v} f_{\mu_{n-2}^\star}(z) dx}_{=1} \\ &= C\, \|f_{X_1}\|_\infty < \infty. \end{aligned}$$

In the inequality above, we took $C \triangleq \sup_{n\geq 3} \binom{n}{\lfloor n/2 \rfloor} / \binom{n-2}{\lfloor (n-2)/2 \rfloor}$, which is finite by Stirling's formula. When $v \in (0, 1]$ and $n \geq 4$ is even, we can apply a similar argument and also obtain a uniform bound. Finally, when $v = 0$ and $n \in \mathbb{N}$, $f_{X_1 - v\mu_n^\star}(u) = f_{X_1}(u) \leq 1/(4\sigma)$.
In summary, $f_{X_1 - v\mu_n^\star}(u)$ is uniformly bounded in $u \in \mathbb{R}$, $n \geq 3$ and $v \in [0, 1]$, which proves (E.4.unif) with any $\alpha_0 > 0$ and any $N_1 \geq 3$.

(E.6.unif) In our case, this is trivial because the conditional density $f_{X_1 - v\mu_n^\star | X_1}(\cdot \,|\, x)$ exists for all $x \in \mathbb{R}$, all $n \geq 2$ and all $v \in (0, 1]$.

(H.1) The function h is continuous on $\mathbb{R} \setminus \{0\}$, so $\mathcal{D}_h = \emptyset$ and thus $\mathbb{P}(X_1 \in \mathcal{D}_h) = 0$.

(H.2) $\mathbb{E}\big|h(X_1)\big| \leq \int_{|x|\leq 1} |\log|x|\,| \frac{1}{4\sigma} dx + \int_{|x|\geq 1} |x| f_{X_1}(x) dx \leq \frac{2}{4\sigma} + 2\sigma < \infty$.

(H.3) For all $x_0 \in \mathbb{R} \setminus \{0\}$, $\limsup_{x\to x_0} |h(x)| < \infty$. This is obvious.

(H.4) $\int_{|u|\leq \alpha_0} |\log|u|\,| du < \infty$ is true for any $\alpha_0 > 0$ since $\int_{|u|\leq 1} |\log|u|\,| du = 2$.

(H.5)
1. This is obviously true for any $\beta_0 > 0$ (use the fundamental theorem of calculus).
2. For any $\gamma > 0$ and any $\beta_0 > \gamma$, the supremum $\sup_{|t|\leq\gamma} |h(X_1 - t)| \mathbf{1}_{\{|X_1 - t| \geq \beta_0\}}$ is attained at the boundary with probability 1 (not necessarily the same end of the boundary for different ω's). Therefore, take $M = |h(X_1 - \gamma)| \mathbf{1}_{\{|X_1 - \gamma| \geq \beta_0\}} + |h(X_1 + \gamma)| \mathbf{1}_{\{|X_1 + \gamma| \geq \beta_0\}}$. It is easy to show that $\mathbb{E}[M] < \infty$ because $|\log|x|\,| \leq |x|$ for $|x| \geq 1$ and $\int_{|x|\geq(1\vee\beta_0)} |x| f_{X_1 \pm \gamma}(x) dx < \infty$.
3. We need to verify this condition for $p = 1$ since this is the p that we used above to verify the conditions of Lemma 2.5. First, $\lim_{|\beta|\to\infty} |h(\beta)| e^{-|x-\beta|^p} = 0$ is true for all $x \in \mathbb{R}$ and all $p > 0$ (true in particular for $p = 1$). For the second part, assume that $\beta \geq 1$. We have

$$\begin{aligned} \mathbb{E}[e^{-|X_1 - \beta|}] &= \int_{(-\infty,0)\cup(0,\beta)\cup(\beta,\infty)} e^{-|x-\beta|} \cdot \frac{1}{4\sigma} e^{-\frac{1}{2\sigma}|x|} dx \\ &\leq \frac{1}{2} e^{-|\beta|} \underbrace{\int_{-\infty}^{0} \frac{1}{2\sigma} e^{-\frac{1}{2\sigma}|x|} dx}_{=1} + \frac{|\beta|}{4\sigma} e^{-(1\wedge\frac{1}{2\sigma})|\beta|} + \frac{1}{4\sigma} e^{-\frac{1}{2\sigma}|\beta|} \underbrace{\int_{\beta}^{\infty} e^{-|x-\beta|} dx}_{=1} \\ &\leq \frac{|\beta|}{2} \left(1 \vee \frac{1}{2\sigma}\right) e^{-(1\wedge\frac{1}{2\sigma})|\beta|}. \end{aligned} \tag{3.4}$$

By the symmetry of f_{X_1}, we also have (3.4) for $\beta \leq -1$. Hence, for any $\beta_0 \geq 1$,

$$\sup_{|\beta|\geq\beta_0} \mathbb{E}\left[\left(|h(\beta)| e^{-|X_1-\beta|}\right)^2\right] < \infty,$$

which is a well-known sufficient condition for the uniform integrability of $\{|h(\beta)| e^{-|X_1-\beta|}\}_{|\beta|\geq\beta_0}$, see e.g. Klenke (2014, Corollary 6.21).
4. Take any $\beta_0 \geq 1$, then (3.4) implies

$$\int_{|u|\geq\beta_0} \mathbb{E}\big[|h'(u)| e^{-|X_1-u|}\big] du \leq \frac{1}{\beta_0} \int_{|u|\geq\beta_0} \mathbb{E}\big[e^{-|X_1-u|}\big] du < \infty.$$

5. Take any $\beta_0 \geq 1$, then, for all $|u| \geq \beta_0$,

$$-\operatorname{sign}(u) \cdot \operatorname{sign}(h(u)) \cdot h'(u) = -\operatorname{sign}(u) \cdot \operatorname{sign}(u) \cdot \frac{1}{|u|} \leq 0.$$

This ends the proof. □

Acknowledgments

We would like to thank the anonymous reviewer for helpful comments that contributed to improving the final version of the paper. F. Ouimet is supported by a NSERC Doctoral Program Alexander Graham Bell scholarship (CGS D3).

References

Blum, J.R., 1955. On the convergence of empiric distribution functions. Ann. Math. Stat. 26, 527–529, MR0070871.

de Wet, T., Randles, R.H., 1987. On the effect of substituting parameter estimators in limiting χ^2 U and V statistics. Ann. Statist. 15 (1), 398–412, MR885745.

Dehardt, J., 1971. Generalizations of the Glivenko-Cantelli theorem. Ann. Math. Stat. 42, 2050–2055, MR0297000.

Desgagné, A., Lafaye de Micheaux, P., 2018. A powerful and interpretable alternative to the Jarque-Bera test of normality based on 2nd-power skewness and kurtosis, using the Rao's score test on the APD family (in press), J. Appl. Stat. http://dx.doi.org/10.1080/02664763.2017.1415311.

Desgagné, A., Lafaye de Micheaux, P., Leblanc, A., 2013. Test of normality against generalized exponential power alternatives. Comm. Statist. Theory Methods 42 (1), 164–190, MR3004652.

Ferguson, T.S., 1996. A Course in Large Sample Theory. Texts in Statistical Science Series. Chapman & Hall, London, p. x+245, MR1699953.

Fernández, C., Osiewalski, J., Steel, M.F.J., 1995. Modeling and inference with v-spherical distributions. J. Amer. Statist. Assoc. 90 (432), 1331–1340, MR1379475.

Giné, E., Zinn, J., 1984. Some limit theorems for empirical processes. Ann. Probab. 12 (4), 929–998, MR757767.

Klenke, A., 2014. Probability Theory, second ed. In: Universitext, Springer, London, p. xii+638, MR3112259.

Komunjer, I., 2007. Asymmetric power distribution: theory and applications to risk measurement. J. Appl. Econometrics 22 (5), 891–921, MR2395888.

LeCam, L., 1953. On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates. Univ. Calif. Publ. Statist. 1, 277–329, MR0054913.

Pollard, D., 1984. Convergence of Stochastic Processes. In: Springer Series in Statistics, Springer-Verlag, New York, p. xiv+215, MR762984.

Rubin, H., 1956. Uniform convergence of random functions with applications to statistics. Ann. Math. Stat. 27, 200–203, MR0076246.

Stroock, D.W., 2011. Probability Theory, second ed. Cambridge University Press, Cambridge, p. xxii+527, MR2760872.

Talagrand, M., 1987. The Glivenko-Cantelli problem. Ann. Probab. 15 (3), 837–870, MR893902.

van de Geer, S.A., 2000. Applications of Empirical Process Theory. In: Cambridge Series in Statistical and Probabilistic Mathematics, vol. 6, Cambridge University Press, Cambridge, p. xii+286, MR1739079.

van der Vaart, A.W., 1998. Asymptotic Statistics. In: Cambridge Series in Statistical and Probabilistic Mathematics, vol. 3, Cambridge University Press, Cambridge, p. xvi+443, MR1652247.

van der Vaart, A.W., Wellner, J.A., 1996. Weak Convergence and Empirical Processes. In: Springer Series in Statistics, Springer-Verlag, New York, p. xvi+508, MR1385671.

Vapnik, V.N., Červonenkis, A. Ya., 1971. The uniform convergence of frequencies of the appearance of events to their probabilities. Teor. Veroyatnost. i Primenen. 16, 264–279, MR0288823.

Vapnik, V.N., Červonenkis, A. Ya., 1981. Necessary and sufficient conditions for the uniform convergence of empirical means to their true values. Teor. Veroyatn. Primen. 26 (3), 543–563, MR627861.