DEMONSTRATIO MATHEMATICA
Vol. XLV No 2 2012

Wojciech Niemiro and Ryszard Zieliński

REMARKS ON UNIFORM CONVERGENCE OF RANDOM VARIABLES AND STATISTICS

Dedicated to Professor Agnieszka Plucińska

Abstract. Convergence in distribution, convergece in probability, and convergence almost surely, *uniform with respect to a family of probability distributions*, is considered. These concepts appeared to be appropriate tools for asymptotic theory of mathematical statistics and many partial results are scattered in the literature of the subject. The aim of this note is to present a unified review of the results in a general and abstract setup. We examine a few rather paradoxical examples which hopefully shed some light on the subtleties of the underlying definitions and the role of asymptotic approximations in statistics. A motivation for considering these problems is provided by their applications.

1. Introduction

The uniform convergence of statistics like $\bar{X}_n$ and random variables like $\sqrt{n}(\bar{X}_n - \mu(\theta))/\sigma(\theta)$ is considered. Definitions of several basic concepts in mathematical statistics involve a quantifier "for all values of parameter θ". This quantifier appears in definitions of an *unbiased* estimator, a *confidence interval*, a *uniformly most powerful test*, or in *robustness framework*, to mention only a few most obvious examples. As long ago as in 1979 Ibragimov and Khas'minskii [6] considered uniformly consistent estimators on subsets of the parameter space. Uniform asymptotics in the robustness framework, with uniformity in a neighborhood ot model distributions, has been considered in Salibian-Barrera et al. [9], Omelka et al. [8] and Berrendero et al. [1]. In some papers the uniformity of asymptotics has been achieved by reducing a parameter space to a suitable subspace of probability distributions, as in the case of asymptotic properties of sample quantiles (Zieliński [10]) or in the case of Dvoretzky–Kiefer–Wolfowitz inequality for smoothed empirical distribution functions [3]. The following example illustrates how much the concept of uniform asymptotics is neglected.

2000 *Mathematics Subject Classification*: 62B99, 62E20, 60F05.

1.1. Example. Encyclopedia of Statistical Sciences, Second Edition, Wiley 2006, includes the following passage:

A textbook standard confidence interval for p is given by $\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}$, *where...*

Of course, a confidence interval, by definition, covers the unknown parameter with probability at least $1-\alpha$ for all values of this parameter. Unfortunately,

$$\mathbb{P}_p(\hat{p} - z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}) \to 0,$$

if $p \to 0$ or $p \to 1$ for every n, just because $\mathbb{P}_p(\hat{p}\hat{q} = 0) \to 0$. The reason of the bad behaviour of this asymptotic confidence interval is well-known: the normal approximation fails if p or q is small. Put differently, for smaller p one needs larger n for the asymptotics to work. Formally speaking, the convergence in distribution $\hat{p}/\sqrt{\hat{p}\hat{q}/n} \to \mathrm{N}(0,1)$ is not *uniform with respect to parameter* p. One may reasonably argue that this fact is not of practical importance in applications where we "do not expect p or q to be too close to 0". However this practically oriented statement is not logically compatible with declaring that the space of parameters is the interval $]0,1[$.

The notion of convergence in distribution, uniform with respect to the family of probability distributions is not new, but it is usually only marginally considered in standard courses of asymptotic statistics. The relevant results are scattered in the literature. In our paper we review some facts from a unified perspective and examine several quite paradoxical examples.

2. Definitions

Consider a statistical space $(\Omega, \mathcal{F}, \{\mathbb{P}_\theta : \theta \in \Theta\})$. Let us say that a *random variable* is a function $Z : \Theta \times \Omega \to \mathbb{R}$ such that for every $\theta \in \Theta$ the mapping $Z(\theta) : \omega \mapsto Z(\theta, \omega)$ is $\mathcal{F}$-measurable. As usual, the argument ω will most often be supressed, while the argument θ will be explicitly written to avoid misunderstanding. Thus we write e.g. $\mathbb{P}_\theta(Z(\theta) \in B)$ instead of $\mathbb{P}_\theta\{\omega : Z(\theta, \omega) \in B\}$. A random variable T which does not depend on θ (i.e. $T : \Omega \to \mathbb{R}$) is called a *statistic*. A random variable which does not depend on ω is called a deterministic function. This terminology might not be quite orthodox but we find it convenient. In what follows, $Z_n(\theta)$, $R_n(\theta)$ etc. denote random variables, while T_n, X_n etc. stand for statistics.

2.1. Definition. Let $Z_1(\theta), \ldots, Z_n(\theta), \ldots$ be a sequence of random variables. Let F be a continuous cumulative distribution function on $\mathbb{R}$. The sequence $Z_n(\theta)$ converges to F in distribution uniformly in $\theta \in \Theta$ if

$$\sup_{\theta\in\Theta} \sup_{-\infty<x<\infty} |\mathbb{P}_\theta(Z_n(\theta) \le x) - F(x)| \to 0 \qquad (n \to \infty).$$

We will then write

$$Z_n(\theta) \rightrightarrows_{\mathrm{d}} F.$$

More explicitly, Definition 2.1 stipulates that

$$\forall_\varepsilon \exists_{n_0} \forall_{n \geq n_0} \forall_\theta \forall_x \quad |\mathbb{P}_\theta \{\omega : Z_n(\theta, \omega) \leq x\} - F(x)| < \varepsilon.$$

Let us emphasize that Definition 2.1 assumes that F *does not depend on* θ *and it is continuous.*

2.2. Definition. A sequence $Z_1(\theta), \ldots, Z_n(\theta), \ldots$ of random variables converges to 0 in probability uniformly in $\theta \in \Theta$ if

$$\sup_{\theta \in \Theta} \mathbb{P}_\theta (|Z_n(\theta)| > \varepsilon) \to 0 \qquad (n \to \infty),$$

for every $\varepsilon > 0$. We will then write

$$Z_n(\theta) \rightrightarrows_{\mathrm{pr}} 0 \quad \text{or} \quad Z_n(\theta) = o_{\mathrm{up}}(1).$$

Explicitly,

$$\forall_\varepsilon \forall_\eta \exists_{n_0} \forall_{n \geq n_0} \forall_\theta \quad \mathbb{P}_\theta \{\omega : |Z_n(\theta, \omega)| > \varepsilon\} < \eta.$$

Definition 2.2 is *not* a special case of 2.1, because the probability distribution concentrated at 0 has discontinuous c.d.f. However, a standard definition of uniform convergence generalizes both Definitions 2.1 and 2.2. We defer a discussion on this to Appendix B.

2.3. Definition. A sequence $Z_1(\theta), \ldots, Z_n(\theta), \ldots$ of random variables converges to 0 almost surely uniformly in $\theta \in \Theta$ if

$$\sup_{\theta \in \Theta} \mathbb{P}_\theta (\sup_{k \geq n} |Z_k(\theta)| > \varepsilon) \to 0 \qquad (n \to \infty),$$

for every $\varepsilon > 0$. We will then write

$$Z_n(\theta) \rightrightarrows_{\mathrm{a.s.}} 0 \quad \text{or} \quad Z_n(\theta) = o_{\mathrm{uas}}(1).$$

Explicitly,

$$\forall_\varepsilon \forall_\eta \exists_{n_0} \forall_\theta \quad \mathbb{P}_\theta \{\omega : \exists_{n \geq n_0} |Z_n(\theta, \omega)| > \varepsilon\} < \eta.$$

2.4. Definition. A sequence $Z_1(\theta), \ldots, Z_n(\theta), \ldots$ of random variables is uniformly bounded in probability if

$$\limsup_{n \to \infty} \sup_{\theta \in \Theta} \mathbb{P}_\theta (|Z_n(\theta)| > m) \to 0 \qquad (m \to \infty).$$

We will then write

$$Z_n(\theta) = O_{\mathrm{up}}(1).$$

Uniform boundedness in probability is equivalent to

$$\forall_\varepsilon \exists_m \exists_{n_0} \forall_{n \geq n_0} \forall_\theta \quad \mathbb{P}_\theta \{\omega : |Z_n(\theta, \omega)| > m\} < \varepsilon.$$

We can now proceed to uniform versions of two fundamental statistical concepts, consistency and asymptotic normality. Consider a function $g : \Theta \to \mathbb{R}$ and a sequence $T_1, \ldots, T_n, \ldots$ of statistics ($T_n : \Omega \to \mathbb{R}$ is regarded as an estimator of $g(\theta)$).

2.5. Definition. Statistic T_n is a uniformly consistent estimator of $g(\theta)$ if

$$T_n - g(\theta) = o_{\mathrm{up}}(1).$$

2.6. Definition. Statistic T_n is a strongly uniformly consistent estimator of $g(\theta)$ if

$$T_n - g(\theta) = o_{\mathrm{uas}}(1).$$

2.7. Definition. Statistic T_n is a uniformly $\sqrt{n}$-consistent estimator of $g(\theta)$ if

$$\sqrt{n}\,[T_n - g(\theta)] = O_{\mathrm{up}}(1).$$

2.8. Definition. Statistic T_n is a uniformly asymptotically normal (UAN) estimator of $g(\theta)$ if there exists a function $\sigma : \Theta \to \mathbb{R}$ such that

$$\frac{\sqrt{n}}{\sigma(\theta)}[T_n - g(\theta)] \rightrightarrows_{\mathrm{d}} \Phi,$$

where Φ is the c.d.f. of the standard normal distribution.

3. Properties

Some well-known properties of the o_{p}, O_{p} and the convergence in distribution concepts are clearly inherited by their uniform analogues, o_{up}, O_{up} and $\rightrightarrows_{\mathrm{d}}$, respectively. However, we have to be aware of some nuances as it will be seen below.

3.1. Lemma. *If* $Z_n(\theta) \rightrightarrows_{\mathrm{pr}} 0$ *and* $\varrho : \mathbb{R} \to \mathbb{R}$ *is a Borel measurable function such that* $\lim_{z\to 0} \varrho(z) = 0$ *then* $\varrho\,(Z_n(\theta)) \rightrightarrows_{\mathrm{pr}} 0$.

Proof. For every $\varepsilon > 0$ there is a $\delta > 0$ such that $|z| \le \delta$ implies $|\varrho(z)| \le \varepsilon$. Hence

$$\mathbb{P}_\theta(|\varrho(Z_n(\theta))| > \varepsilon) \le \mathbb{P}_\theta((|Z_n(\theta)| > \delta) \le \sup_{\theta\in\Theta} \mathbb{P}_\theta\,(|Z_n(\theta)| > \delta)$$

and the RHS tends to 0. ∎

3.2. Lemma. *If* $X_n(\theta) = O_{\mathrm{up}}(1)$ *and* $R_n(\theta) = o_{\mathrm{up}}(1)$ *then* $X_n(\theta)R_n(\theta) = o_{\mathrm{up}}(1)$.

Proof. Fix $\varepsilon, \eta > 0$. Take suitable n_0 and m such that $\sup_\theta \mathbb{P}_\theta(|X_n(\theta)| > m) < \eta$ for $n \ge n_0$. Then choose n_1 such that $\sup_\theta \mathbb{P}_\theta(|R_n(\theta)| > \varepsilon/m) < \eta$ for $n \ge n_1$. For $n \ge \max(n_0, n_1)$ we thus have

$$\mathbb{P}_\theta(|X_n(\theta)R_n(\theta)| > \varepsilon) \le \mathbb{P}_\theta(|X_n(\theta)| > m) + \mathbb{P}_\theta(|R_n(\theta)| > \varepsilon/m) < 2\eta,$$

for all θ. ∎

An important special case obtains if $R_n(\theta) = r_n(\theta)$ are deterministic functions. Then $R_n(\theta) \rightrightarrows_{\rm pr} 0$ reduces to ordinary uniform convergence $r_n(\theta) \rightrightarrows 0$.

3.3. Lemma. *If $X_n(\theta) \rightrightarrows_{\rm d} F$ for some c.d.f. F then $X_n(\theta) = O_{\rm up}(1)$.*

Proof. Fix an $\varepsilon > 0$ and choose m such that $1 - F(m) + F(-m) < \varepsilon$. For sufficiently large n, say $n \geq n_0$ we have $\sup_\theta |\mathbb{P}_\theta(X_n(\theta) \leq x) - F(x)| < \varepsilon$ for all x. Therefore for $n \geq n_0$,

$$\begin{aligned}\mathbb{P}_\theta(|X_n| > m) &\leq \mathbb{P}_\theta(X_n(\theta) \leq -m) + 1 - \mathbb{P}_\theta(X_n(\theta) \leq m)\\ &\leq |\mathbb{P}_\theta(X_n(\theta) \leq -m) - F(-m)| + F(-m)\\ &\qquad + 1 - F(m) + |F(m) - \mathbb{P}_\theta(X_n(\theta) \leq m)|\\ &< \varepsilon + F(-m) + 1 - \Phi(m) + \varepsilon < 3\varepsilon,\end{aligned}$$

for all θ, which proves our assertion. ∎

3.4. Corollary. *Let r_n be a sequence of deterministic functions and assume that $Z_n(\theta) \rightrightarrows_{\rm d} F$. If $r_n(\theta)$ are uniformly bounded then $r_n(\theta)Z_n(\theta) = O_{\rm up}(1)$. If $r_n(\theta) \rightrightarrows 0$ then $r_n(\theta)Z_n(\theta) = o_{\rm up}(1)$.*

Note that the condition $r_n(\theta) \rightrightarrows 0$ is essential. The following example illustrates the situation.

3.5. Example. Suppose $T_n \sim \mathrm{N}(\theta, \theta^2/n)$ under $\mathbb{P}_\theta$, with $\theta \in \Theta = \mathbb{R}$. Then T_n is clearly UAN, because $(\sqrt{n}/\theta)[T_n - \theta] \sim \mathrm{N}(0,1)$. However, T_n is *not* uniformly consistent. The reason is that $\theta/\sqrt{n} \to 0$ pointwise but not uniformly, $\theta/\sqrt{n} \not\rightrightarrows 0$.

3.6. Lemma (A uniform version of Slucki's Theorem). *If $X_n(\theta) \rightrightarrows_{\rm d} F$ and $R_n(\theta) \rightrightarrows_{\rm pr} 0$ then $X_n(\theta) + R_n(\theta) \rightrightarrows_{\rm d} F$.*

Proof. Let us begin with the following self-evident inequalities:

$$\begin{aligned}\mathbb{P}_\theta\left(X_n + R_n \leq x\right) &\leq \mathbb{P}_\theta\left(X_n \leq x + \delta\right) + \mathbb{P}_\theta\left(R_n < -\delta\right)\\ \mathbb{P}_\theta\left(X_n + R_n \leq x\right) &\geq \mathbb{P}_\theta\left(X_n \leq x - \delta\right) - \mathbb{P}_\theta\left(R_n > \delta\right).\end{aligned}$$

It follows that

$$\begin{aligned}\left|\mathbb{P}_\theta\left(X_n + R_n \leq x\right) - F(x)\right| \leq \sup_x \left|\mathbb{P}_\theta\left(X_n \leq x + \delta\right) - F(x+\delta)\right|\\ + \sup_x |F(x+\delta) - F(x)|\\ + \mathbb{P}_\theta\left(|R_n| > \delta\right).\end{aligned}$$

The contribution of the middle term on the RHS can be made arbitrarily small in view of the uniform continuity of F. The first term goes uniformly to 0 because $X_n(\theta) \rightrightarrows_{\rm d} F$ and the third term – because $R_n(\theta) \rightrightarrows_{\rm pr} 0$. ∎

3.7. Lemma (A uniform version of the δ-method). *Let $h : \mathbb{R} \to \mathbb{R}$ be a Borel function differentiable at μ. Assume that h and μ do not depend on θ. If*

$$\frac{\sqrt{n}}{\sigma(\theta)}[Z_n(\theta) - \mu] \rightrightarrows_{\mathrm{d}} \Phi,$$

$h'(\mu) \neq 0$ and $\sigma(\theta) \leq b < \infty$ for some $b > 0$ and for all $\theta \in \Theta$ then

$$\frac{\sqrt{n}}{\sigma(\theta)h'(\mu)}[h(Z_n(\theta)) - h(\mu)] \rightrightarrows_{\mathrm{d}} \Phi.$$

Proof. By the definition of derivative, $h(z) - h(\mu) = h'(\mu)(z-\mu) + \varrho(z)(z-\mu)$, where $\varrho(z) \to 0$ as $z \to \mu$. We can write

$$\begin{aligned} \frac{\sqrt{n}}{\sigma(\theta)h'(\mu)}[h(Z_n(\theta)) - h(\mu)] &= \frac{\sqrt{n}}{\sigma(\theta)}[Z_n(\theta) - \mu] \\ &\quad + \frac{r(Z_n(\theta))}{h'(\mu)} \frac{\sqrt{n}}{\sigma(\theta)}[Z_n(\theta) - \mu] \\ &:= V_n(\theta) + R_n(\theta)V_n(\theta). \end{aligned}$$

By assumption, $V_n(\theta) \rightrightarrows_{\mathrm{d}} \Phi$. Corollary 3.4 implies that $Z_n(\theta) - \mu \rightrightarrows_{\mathrm{pr}} 0$ (note that $\sigma(\theta)/\sqrt{n} \rightrightarrows 0$ because $\sigma(\theta)$ is bounded). Then it follows from Lemma 3.1 that $R_n(\theta) \rightrightarrows_{\mathrm{pr}} 0$. The conclusion now follows from Lemma 3.2 and Lemma 3.6. ∎

A. Appendix: a uniform CLT

In this appendix, we follow Borovkov [2] (Appendix IV, par. 4, Th. 5). However, in contrast with Borovkov, we consider only a fixed limit law $\mathrm{N}(0,1)$. Borovkov does not mention that his sufficient condition for UAN for i.i.d. summands (Condition A.2 below) is also necessary.

We consider a sequence of random variables $X_1(\theta), \ldots, X_n(\theta), \ldots$ defined on a statistical space $(\Omega, \mathcal{F}, \{\mathbb{P}_\theta : \theta \in \Theta\})$. Let Φ be the c.d.f. of $\mathrm{N}(0,1)$.

A.1. Theorem. *Let us assume that for every θ, random variables $X_1(\theta), \ldots, X_n(\theta), \ldots$ are i.i.d. with $\mathbb{E}_\theta X_i(\theta) = \mu(\theta)$ and finite variance $\mathrm{Var}_\theta X_i(\theta) = \sigma^2(\theta)$. Let $S_n(\theta) = \sum_{i=1}^n X_i(\theta)$. Write $X(\theta) = X_1(\theta)$ and*

$$\tilde{X}(\theta) = \frac{X(\theta) - \mu(\theta)}{\sigma(\theta)}$$

for the standardized single variable. Then

$$\sup_\theta \mathbb{E}_\theta \tilde{X}(\theta)^2 \mathbb{I}(|\tilde{X}(\theta)| > a) \to 0 \quad (a \to \infty) \tag{A.2}$$

is a necessary and sufficient condition for

$$\frac{S_n(\theta) - n\mu(\theta)}{\sigma(\theta)\sqrt{n}} \rightrightarrows_{\mathrm{d}} \Phi. \tag{A.3}$$

Proof. The crucial point is to notice that the uniform convergence (A.3), i.e.

$$\sup_{\theta} \sup_{-\infty<x<\infty} \left| \mathbb{P}_\theta \left(\frac{S_n(\theta) - n\mu(\theta)}{\sigma(\theta)\sqrt{n}} \le x \right) - \Phi(x) \right| \to 0$$

is equivalent to the following statement: for every sequence θ_n of elements of Θ we have

$$\sup_{-\infty<x<\infty} \left| \mathbb{P}_{\theta_n} \left(\frac{S_n(\theta_n) - n\mu(\theta_n)}{\sigma(\theta_n)\sqrt{n}} \le x \right) - \Phi(x) \right| \to 0. \tag{A.4}$$

This fact is just a special case of „sequential definition of uniform convergence": given a sequence ψ_n of functions, $\sup_\theta |\psi_n(\theta)| \to 0$ holds true if and only if $\psi_n(\theta_n) \to 0$ for every sequence θ_n. Therefore if we fix a sequence θ_n and

$$X_{nk} = \frac{X_k(\theta_n) - \mu(\theta_n)}{\sigma(\theta_n)\sqrt{n}}, \qquad (k = 1, \ldots, n),$$

we can use the classical Lindeberg–Feller theorem for triangular arrays (e.g. Borovkov [2] or Dudley [4]). It should be emphasized that theorems for triangular arrays allow the rows to be defined on different probability spaces. Clearly, we have $\sum_{k=1}^n X_{nk} = S_n(\theta_n)/(\sigma(\theta_n)\sqrt{n})$, $\mathbb{E}_{\theta_n} X_{nk} = 0$, $\sum_{k=1}^n \mathbb{E}_{\theta_n} X_{nk}^2 = 1$ and $\max_{k=1}^n \mathbb{E}_{\theta_n} X_{nk}^2 = 1/n \to 0$. It remains to check the Lindeberg condition. If (A.2) holds then

$$L_n := \sum_{k=1}^n \mathbb{E}_{\theta_n} X_{nk}^2 \mathbb{I}(|X_{nk}| > \varepsilon) = \mathbb{E}_{\theta_n} \tilde{X}(\theta_n)^2 \mathbb{I}(|\tilde{X}(\theta_n)| > \varepsilon\sqrt{n}) \to 0,$$

so the Lindeberg condition is fulfilled and (A.4) follows. Conversly, if (A.2) does not hold then for some sequence (θ_n) we have $L_n \not\to 0$. The Feller's theorem (e.g. [4], note to par. 9.4) implies that (A.4) is not true. ∎

A.5. Remark. The condition (A.2) follows from the following "Lyapunov type" condition

$$\sup_\theta \mathbb{E}_\theta |\tilde{X}(\theta)|^{2+\delta} < \infty \quad \text{for some } \delta > 0.$$

Indeed, $\mathbb{E}_\theta \tilde{X}(\theta)^2 \mathbb{I}(|\tilde{X}(\theta)| > a) \le a^{-\delta} \mathbb{E}_\theta |\tilde{X}(\theta)|^{2+\delta}$.

A.6. Example. (CLT for the Bernoulli scheme, [7]) Let $X = X_1, \ldots, X_n$, ... be i.i.d. with $\mathbb{P}_\theta(X = 1) = \theta = 1 - \mathbb{P}_\theta(X = 0)$. The parameter space is $\Theta =]0, 1[$. We have

$$\tilde{X}(\theta) = \frac{X - \theta}{\sqrt{\theta(1-\theta)}}.$$

It is easy to see that for θ sufficiently close to 0,

$$\mathbb{E}_\theta \tilde{X}(\theta)^2 \mathbb{I}(|\tilde{X}(\theta)| > a) \geq \mathbb{E}_\theta \left(\frac{X-\theta}{\sqrt{\theta(1-\theta)}} \right)^2 \mathbb{I}(X=1) = 1-\theta,$$

so the condition (A.2) is not satisfied. Therefore,

$$\frac{\sum_{i=1}^n X_i - \theta}{\sqrt{\theta(1-\theta)}} \not\rightrightarrows_{\mathrm{d}} \Phi \qquad (0<\theta<1).$$

Thus the CLT for the Bernoulli scheme (de Moivre–Laplace Theorem) is not uniform.

However, if we restrict the parameter space to a compact subset $I \subset]0,1[$, it is easy to see that the CLT becomes uniform. Indeed,

$$\mathbb{E}_\theta \tilde{X}(\theta)^4 = \frac{1+2\theta^2-3\theta^4}{\theta^2(1-\theta)^2}.$$

Theorem A.1 combined with Remark A.5 yields immediately a uniform CLT:

$$\frac{\sum_{i=1}^n X_i - n\theta}{\sqrt{n\theta(1-\theta)}} \rightrightarrows_{\mathrm{d}} \Phi, \qquad \delta \in I.$$

A.7. Example. (CLT for the Negative Binomial scheme, [7]) Suppose $Y = Y_1, \ldots, Y_n, \ldots$ are i.i.d. and have the geometric distribution, $\mathbb{P}_\theta(Y=k) = \theta(1-\theta)^{k-1}$ for $k = 1, 2, \ldots$.

We will use the following elementary facts about the geometric distribution (see `mathworld.wolfram.com` for example):

$$\mu(\theta) = \mathbb{E}_\theta(Y) = \frac{1}{\theta}, \qquad \sigma^2(\theta) = \mathrm{Var}_\theta(Y) = \frac{1-\theta}{\theta^2},$$
$$m_4(\theta) = \mathbb{E}_\theta(Y-\mu(\theta))^4 = \frac{(1-\theta)(\theta^2-9\theta+9)}{\theta^4}.$$

Just as in the previous example we can show that

$$\frac{\sum_{i=1}^n \theta Y_i - n}{\sqrt{n(1-\theta)}} \not\rightrightarrows_{\mathrm{d}} \Phi \qquad (0<\theta<1),$$

because the uniform convergence fails for θ close to 1.

If the parameter space is $\Theta =]0, 1-\delta]$ with $\delta > 0$ then a uniform CLT follows again from Theorem A.1 and Remark A.5. Now we have

$$\tilde{Y}(\theta) = \frac{\theta Y - 1}{\sqrt{1-\theta}} \quad \text{and} \quad \mathbb{E}_\theta \tilde{Y}(\theta)^4 = \frac{\theta^2}{1-\theta} + 9.$$

Consequently,

$$\frac{\sum_{i=1}^n \theta Y_i - n}{\sqrt{n(1-\theta)}} \rightrightarrows_{\mathrm{d}} \Phi \qquad (0<\theta<1-\delta).$$

It is perhaps interesting that uniformity in CLT does not imply uniformity in LLN.

A.8. Example. (CLT is uniform and LLN is not, [11]) Consider the family of exponential distributions $\{\mathbb{P}_\theta = \mathrm{Ex}(\theta), \theta \in]0, \infty[\}$, where by $\mathrm{Ex}(\theta)$ we mean the distribution with density $f_\theta(x) = \theta \mathrm{e}^{-\theta x}$. It is easy to verify condition (A.2). Since $\mu(\theta) = 1/\theta$ and $\sigma(\theta) = 1/\theta$, we have $\theta X \sim \mathrm{Ex}(1)$. Therefore random variable $\tilde{X}(\theta) = \theta X - 1$ has a probability distribution independent of θ and finite variance. It follows that condition (A.2) is fulfilled and by Theorem A.1 the uniform CLT holds. On the other hand, $\theta \bar{X}_n = \theta \sum_{i=1}^n /n \sim \mathrm{Gamma}(n, n)$ has also a distribution free of θ, so $\sup_{\theta>0} \mathbb{P}_\theta(|\bar{X}_n - 1/\theta| > \varepsilon) = \sup_{\theta>0} \mathbb{P}_\theta(|\theta\bar{X}_n - 1| > \theta\varepsilon) = 1$. It is therefore not true that $\bar{X}_n \rightrightarrows_{\mathrm{pr}} 1/\theta$: the Weak LNN is not uniform.

On the other hand, there exist models which admit uniform LLN but not a uniform CLT.

A.9. Example. (SLLN is uniform and CLT is not, [11]) The uniform CLT fails for the Bernoulli model, as shown in Example A.6. On the other hand, not only the Weak LLN but also the Strong LLN hold uniformly in $\theta \in]0, 1[$. For a proof we refer to [11].

B. Appendix: a general definition of uniform convergence in distribution

Definition 2.1 can be generalized in the following way (e.g. Borovkov [2], Chapter II, par. 37, Def. 2). Let $Z_1(\theta), \dots, Z_n(\theta), \dots$ be a sequence of random variables defined on a statistical space $(\Omega, \mathcal{F}, \{\mathbb{P}_\theta : \theta \in \Theta\})$. Let $\{F_\theta : \theta \in \Theta\}$ be a family of probability distributions.

B.1. Definition. Uniform convergence in distribution $Z_n(\theta) \rightrightarrows_{\mathrm{d}} F_\theta$ holds if for every continuous and bounded function h,

$$\sup_\theta \left| \mathbb{E}_\theta h(Z_n(\theta)) - \int h \mathrm{d}F_\theta \right| \to 0.$$

If we take $F_\theta = \Phi$, we reduce Definition 2.1 to a special case of B.1. Moreover, if we take $F_\theta = \delta_0 = \mathbb{I}_{[1,\infty[}$, i.e. the c.d.f. of a probability concentrated at zero, then $Z_n(\theta) \rightrightarrows_{\mathrm{d}} \delta_0$ is equivalent to $Z_n(\theta) \rightrightarrows_{\mathrm{pr}} 0$, as defined by 2.2. However, some caution is necessary. There are some nuances related to the uniform convergence to laws which *depend on* θ. The apparent analogue of 2.1, i.e.

$$\sup_{\theta\in\Theta} \sup_{-\infty<x<\infty} |\mathbb{P}_\theta (Z_n(\theta) \le x) - F_\theta(x)| \to 0 \qquad (n \to \infty),$$

is *not* equivalent to $Z_n(\theta) \rightrightarrows_{\mathrm{d}} F_\theta$.

We freely identify probability laws with their c.d.f.'s—thus writing $\rightrightarrows_{\mathrm{d}} \mathrm{N}(0,1)$ instead of $\rightrightarrows_{\mathrm{d}} \Phi$ and so on.

B.2. Example. Consider the Bernoulli scheme, just as in Example A.6. Let $X = X_1, \ldots, X_n, \ldots$ be i.i.d. with $\mathbb{P}_\theta(X=1) = \theta = 1 - \mathbb{P}_\theta(X=0)$. The parameter space is $\Theta =]0,1[$. Let $\bar{X}_n = \sum_{i=1}^n X_i/n$. On the one hand we know that

$$\frac{\sqrt{n}}{\sqrt{\theta(1-\theta)}}[\bar{X}_n - \theta] \not\rightrightarrows_{\mathrm{d}} \mathrm{N}(0,1),$$

see also [11]. On the other hand Theorem 2 in par. 37, Chapter II in [2] implies that

$$\sqrt{n}[\bar{X}_n - \theta] \rightrightarrows_{\mathrm{d}} \mathrm{N}\left(0, \theta(1-\theta)\right).$$

References

[1] J. R. Berrendero, R. H. Zamar, *A note on the uniform asymptotic normality of location M-estimates*, Metrika 63 (2006), 55–69.

[2] A. A. Borovkov, *Mathematical Statistics*, Gordon and Breach (1998).

[3] Z. Ciesielski, R. Zieliński, *Polynomial and spline estimators of the distribution function with prescribed accuracy*, Applicationes Mathematicae 36(1) (2009), 1–12.

[4] R. M. Dudley, *Real Analysis and Probability*, Wadsworth & Brooks (1989).

[5] W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. II, Wiley (1966).

[6] I. A. Ibragimov, R. Z. Khas'minskii, *Asymptotic theory of estimation* (in Russian), „Nauka", Moscow, 1979 [MR545339, 81h:62004].

[7] W Niemiro, R. Zieliński, *Uniform asymptotic normality for the Bernoulli scheme*, Appl. Math. 34(2) (2007), 215–221.

[8] M Omelka, M. Salibian-Barrera, *Uniform asymptotics for S- and MM-regression estimators*, Ann. Inst. Stat. Math. 62 (2010), 897–927.

[9] M. Salibian-Barrera, R. H. Zamar, *Uniform asymptotics for robust location estimates when the scale is unknown*, Ann. Statist. 32(4) (2004), 1434–1447.

[10] R. Zieliński, *Uniform strong consistency of sample quantiles*, Statist. Probab. Lett. 37 (1998), 115–119.

[11] R. Zieliński, *Effective WLLN, SLLN and CLT in statistical models*, Appl. Math. 31(1) (2004), 117–125.

Wojciech Niemiro
UMK TORUŃ AND UW WARSZAWA
E-mail: wniemiro@gmail.com

Ryszard Zieliński
IMPAN, WARSZAWA

Received December 12, 2011.