

RYSZARD ZIELIŃSKI (Warszawa)

## EFFECTIVE WLLN, SLLN AND CLT IN STATISTICAL MODELS

*Abstract.* Weak laws of large numbers (WLLN), strong laws of large numbers (SLLN), and central limit theorems (CLT) in statistical models differ from those in probability theory in that they should hold uniformly in the family of distributions specified by the model. If a limit law states that for every  $\varepsilon > 0$  there exists  $N$  such that for all  $n > N$  the inequalities  $|\xi_n| < \varepsilon$  are satisfied and  $N = N(\varepsilon)$  is explicitly given then we call the law effective. It is trivial to obtain an effective statistical version of WLLN in the Bernoulli scheme, to get SLLN takes a little while, but CLT does not hold uniformly. Other statistical schemes are also considered.

**1. The Bernoulli scheme.** Let  $X, X_1, X_2, \dots, X_n, \dots$  be iid random variables with

$$P_\theta\{X = 1\} = 1 - P_\theta\{X = 0\} = \theta, \quad \theta \in (0, 1),$$

and let  $S_n = \sum_{i=1}^n X_i$ .

WLLN states that, for every fixed  $\theta \in (0, 1)$ ,  $S_n/n \rightarrow \theta$  in probability, which can be written in the form

$$\forall \theta \in (0, 1) \quad \forall \varepsilon > 0 \quad \forall \eta > 0 \quad \exists N \quad \forall n \geq N \quad P_\theta \left\{ \left| \frac{S_n}{n} - \theta \right| \geq \varepsilon \right\} \leq \eta.$$

An appropriate  $N$  is given by the formula  $N = \theta(1 - \theta)/\eta\varepsilon^2$ .

In the related statistical model all we know about  $\theta$  is that  $\theta \in (0, 1)$  so that the above result is of no use: the statistical version may be formulated as follows:

---

2000 *Mathematics Subject Classification*: 62E12, 62E20, 60F05, 60F15.

*Key words and phrases*: weak laws of large numbers, strong laws of large numbers, central limit theorems, statistical models, effective limit laws, uniform limit laws, Bernoulli scheme, exponential distribution, quantiles.

## PROPOSITION 1.

$$(A) \quad \forall \varepsilon > 0 \quad \forall \eta > 0 \quad \exists N \quad \forall n \geq N \quad \forall \theta \in (0, 1) \quad P_\theta \left\{ \left| \frac{S_n}{n} - \theta \right| \geq \varepsilon \right\} \leq \eta;$$

$$(B) \quad \text{an appropriate } N \text{ is } N(\varepsilon, \eta) = -\frac{\frac{1}{2} + \frac{2}{3}\varepsilon}{\varepsilon^2} \log \frac{\eta}{2}.$$

A better known but less exact formula in (B) is  $N(\varepsilon, \eta) = 1/4\eta\varepsilon^2$ .

*Proof.* Remark A(i) in Serfling (1980, Sec. 2.5.4) gives us

$$P_\theta \left\{ \left| \frac{S_n}{n} - \theta \right| \geq \varepsilon \right\} \leq 2\varepsilon \left\{ -\frac{\varepsilon^2}{\frac{1}{2} + \frac{2}{3}\varepsilon} n \right\}$$

for all  $n = 1, 2, \dots$  ■

Here and further on, Part (A) states the uniform convergence and Part (B) makes the law effective.

SLLN states that, for every fixed  $\theta \in (0, 1)$ ,  $S_n/n \rightarrow \theta$  a.s. If we use the fact that  $\xi_n \rightarrow 0$  a.s. iff  $\forall \varepsilon > 0 \lim_{N \rightarrow \infty} P\{\bigcup_{n=N}^{\infty} \{|\xi_n| > \varepsilon\}\} = 0$  iff  $\forall \varepsilon > 0 \quad \forall \eta > 0 \quad \exists N \quad P\{\bigcup_{n=N}^{\infty} \{|\xi_n| > \varepsilon\}\} < \eta$ , an appropriate effective statistical version of the law takes on the following form:

## PROPOSITION 2.

$$(A) \quad \forall \varepsilon > 0 \quad \forall \eta > 0 \quad \exists N \quad \forall \theta \in (0, 1) \quad P_\theta \left\{ \bigcup_{n=N}^{\infty} \left\{ \left| \frac{S_n}{n} - \theta \right| \geq \varepsilon \right\} \right\} \leq \eta;$$

(B) *an appropriate*  $N$  *is*

$$N(\varepsilon, \eta) = \min \left\{ -\frac{1}{\chi} \log \left( \frac{\eta}{2} (1 - e^{-\chi}) \right), \frac{1}{4\eta\varepsilon^2} \right\},$$

where

$$\chi = \frac{\varepsilon^2}{\frac{1}{2} + \frac{2}{3}\varepsilon}.$$

*Proof.* By a rather crude estimation one obtains

$$P_\theta \left\{ \bigcup_{n=N}^{\infty} \left\{ \left| \frac{S_n}{n} - \theta \right| \geq \varepsilon \right\} \right\} \leq \sum_{n=N}^{\infty} P_\theta \left\{ \left| \frac{S_n}{n} - \theta \right| \geq \varepsilon \right\}$$

By the inequality given in the proof of Proposition 1 the following estimate holds:

$$P_\theta \left\{ \bigcup_{n=N}^{\infty} \left\{ \left| \frac{S_n}{n} - \theta \right| \geq \varepsilon \right\} \right\} \leq 2 \frac{e^{-N\chi}}{1 - e^{-\chi}},$$

and

$$2 \frac{e^{-N\chi}}{1 - e^{-\chi}} \leq \eta$$

amounts to

$$N \geq -\frac{1}{\chi} \log\left(\frac{\eta}{2} (1 - e^{-\chi})\right).$$

Table 1 (first line) exhibits  $N = N(\varepsilon, \eta)$  for some  $\varepsilon$  and  $\eta$ .

Another formula for  $N$  may be constructed as follows (Wesołowski 2002). Define  $Y_i = X_i - \theta$ ,  $T_k = k^{-1} \sum_{i=1}^k Y_i$ , and  $\mathcal{G}_k = \sigma(T_k, T_{k+1}, \dots)$ . Then  $(T_k, \mathcal{G}_k)_{k=1,2,\dots}$  is an inverse martingale:

$$E(T_k | \mathcal{G}_{k+1}) = E(T_k | T_{k+1}) = \frac{1}{k} \sum_{i=1}^k E(Y_i | T_{k+1}) = T_{k+1}, \quad k = 1, 2, \dots$$

The maximal inequality for inverse martingales gives us

$$P\left\{\max_{N \leq k \leq m} |T_k| \geq a\right\} \leq \frac{\text{Var}(T_N)}{a^2} = \frac{\theta(1 - \theta)}{Na^2} \leq \frac{1}{4Na^2}$$

and in consequence

$$\begin{aligned} P_\theta \left\{ \bigcup_{n=N}^{\infty} \left\{ \left| \frac{S_n}{n} - \theta \right| \geq \varepsilon \right\} \right\} &= P\left\{ \sup_{k \geq N} |T_k| \geq \varepsilon \right\} \\ &= \lim_{m \rightarrow \infty} P\left\{ \max_{N \leq k \leq m} |T_k| \geq \varepsilon \right\} \\ &\leq \lim_{m \rightarrow \infty} \frac{1}{4N\varepsilon^2} = \frac{1}{4N\varepsilon^2}. \end{aligned}$$

Now Proposition 2(A) holds for any  $N \geq 1/4\varepsilon^2\eta$  (second line in Table 1). ■

**Table 1.**  $N(\varepsilon, \eta)$

$\varepsilon$	$\eta$		
	0.1	0.01	0.001
0.1	400	530	660
	250	2500	25000
0.01	58400	70067	81733
	25000	250000	2500000
0.001	$0.81 \cdot 10^7$	$0.93 \cdot 10^7$	$1.04 \cdot 10^7$
	$0.25 \cdot 10^7$	$2.50 \cdot 10^7$	$25 \cdot 10^7$

CLT for the Bernoulli scheme holds for every  $\theta \in (0, 1)$  separately, even in a stronger version (“uniformly in  $x$ ”):

$$\forall \theta \in (0, 1) \quad \sup_x \left| P_\theta\{S_n \leq x\} - \Phi\left(\frac{x - n\theta}{\sqrt{n\theta(1 - \theta)}}\right) \right| \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

The classical CLT for the Bernoulli scheme may be written in the form

$$\forall \theta \forall x \forall \varepsilon \exists N = N(\theta, x, \varepsilon) \forall n \geq N \quad \left| P_\theta\left\{ \frac{S_n - n\theta}{\sqrt{n\theta(1 - \theta)}} \leq x \right\} - \Phi(x) \right| \leq \varepsilon.$$

What statisticians need, however, is

$$\forall x \forall \varepsilon \exists N = N(x, \varepsilon) \forall n \geq N \forall \theta \quad \left| P_\theta \left\{ \frac{S_n - n\theta}{\sqrt{n\theta(1-\theta)}} \leq x \right\} - \Phi(x) \right| \leq \varepsilon$$

or even in a stronger form: “uniformly in  $x$ ”.

The latter, however, is not true. To see this one should prove that

$$\exists x \exists \varepsilon \forall N \exists n \geq N \exists \theta \quad \left| P_\theta \left\{ \frac{S_n - n\theta}{\sqrt{n\theta(1-\theta)}} \leq x \right\} - \Phi(x) \right| > \varepsilon.$$

It is sufficient to prove that

$$\exists x \exists \varepsilon \forall n \exists \theta \quad \left| P_\theta \left\{ \frac{S_n - n\theta}{\sqrt{n\theta(1-\theta)}} \leq x \right\} - \Phi(x) \right| > \varepsilon.$$

To this end take  $x = 0$  and  $\varepsilon = 1/4$ . Then LHS =  $|P_\theta\{S_n \leq n\theta\} - 1/2|$ . If for any fixed  $n$  one takes  $\theta$  such that  $n\theta < 1$  and  $(1-\theta)^n > 3/4$ , then  $P_\theta\{S_n \leq n\theta\} = P_\theta\{S_n = 0\} = (1-\theta)^n > 3/4$  and LHS  $> \varepsilon$ . It follows that CLT does not hold uniformly in the statistical model with  $\theta \in (0, 1)$ .

It is interesting to observe that a similar result holds in the inverse Binomial scheme (negative Binomial distribution). Let  $Y$  be the number of experiments needed to observe a first success:

$$P_\theta\{Y = y\} = (1-\theta)^{y-1}\theta, \quad E_\theta Y = \frac{1}{\theta}, \quad \text{Var}_\theta Y = \frac{1-\theta}{\theta^2}.$$

If  $Y, Y_1, Y_2, \dots$  are iid and  $T_n = \sum_{i=1}^n Y_i$  then

$$\begin{aligned} P_\theta \left\{ \frac{T_n - n/\theta}{\sqrt{n(1-\theta)/\theta^2}} \leq x \right\} \Big|_{x=0} &= P_\theta\{T_n \leq n/\theta\} - \Phi(0) \\ &> P_\theta\{T_n \leq n\} - \Phi(0) \\ &= \theta^n - 1/2, \end{aligned}$$

which tends to  $1/2$  as  $\theta \rightarrow 1$ .

One may conclude that typical difficulties in constructing confidence intervals for  $\theta$  (e.g. Brown *et al.* 2001), based on normal approximation, arise from the fact that CLT does not hold uniformly.

**2. Exponential distribution.** If  $X_1, X_2, \dots$  are iid random variables with probability density function  $\lambda^{-1}e^{-x/\lambda}$ ,  $x > 0$ ,  $\lambda > 0$ , and  $S_n = \sum_{i=1}^n X_i$ , then the SLLN

$$S_n/n \rightarrow \lambda \quad \text{a.s.}$$

does not hold uniformly in  $\lambda > 0$ , and the CLT

$$\forall x \quad \left| P_\lambda \left\{ \frac{S_n/n - \lambda}{\lambda} \sqrt{n} \leq x \right\} - \Phi(x) \right| \rightarrow 0$$

holds uniformly.

To prove the former it is enough to observe that for some fixed  $\varepsilon > 0$ ,  $\eta > 0$ , and for each  $n$ , one can find  $\lambda > 0$  such that

$$P_\lambda \left\{ \left| \frac{S_n}{n} - \lambda \right| < \varepsilon \right\} < \eta,$$

which, by the fact that  $S_n/n$  has gamma distribution  $\Gamma(n, \lambda/n)$  with the shape parameter  $n$  and scale parameter  $\lambda/n$ , easily follows from the estimate

$$P_\lambda \left\{ \left| \frac{S_n}{n} - \lambda \right| < \varepsilon \right\} = \frac{1}{\Gamma(n)} \int_{n(1-\varepsilon/\lambda)}^{n(1+\varepsilon/\lambda)} t^{n-1} e^{-t} dt < \frac{2n\varepsilon}{\lambda} \frac{1}{\sqrt{2\pi n}}.$$

A stronger version of the second statement may be formulated as the following effective Proposition 3. Define

$$R(x, n) = \begin{cases} \frac{1}{\Gamma(n)} \int_0^{n+x\sqrt{n}} t^{n-1} e^{-t} dt - \Phi(x) & \text{if } x > -\sqrt{n}, \\ 0 & \text{elsewhere.} \end{cases}$$

**PROPOSITION 3.** *If  $X_1, X_2, \dots$  are iid random variables with probability density function  $\lambda^{-1}e^{-x/\lambda}$  and  $S_n = \sum_{i=1}^n X_i$  then*

(A)  $\forall \varepsilon > 0 \exists N = N(\varepsilon) \forall n \geq N \forall \lambda > 0$

$$\sup_x \left| P_\lambda \left\{ \frac{S_n/n - \lambda}{\lambda} \sqrt{n} \leq x \right\} - \Phi(x) \right| < \varepsilon;$$

(B) *an appropriate  $N = N(\varepsilon)$  is given numerically as one such that*

$$\max_x |R(x, N)| \leq \varepsilon.$$

*Proof.* To prove (A) it is enough to observe that

$$P_\lambda \left\{ \frac{S_n/n - \lambda}{\lambda} \sqrt{n} \leq x \right\} = P_\lambda \left\{ \frac{1}{\lambda} \frac{S_n}{n} \leq 1 + \frac{x}{\sqrt{n}} \right\},$$

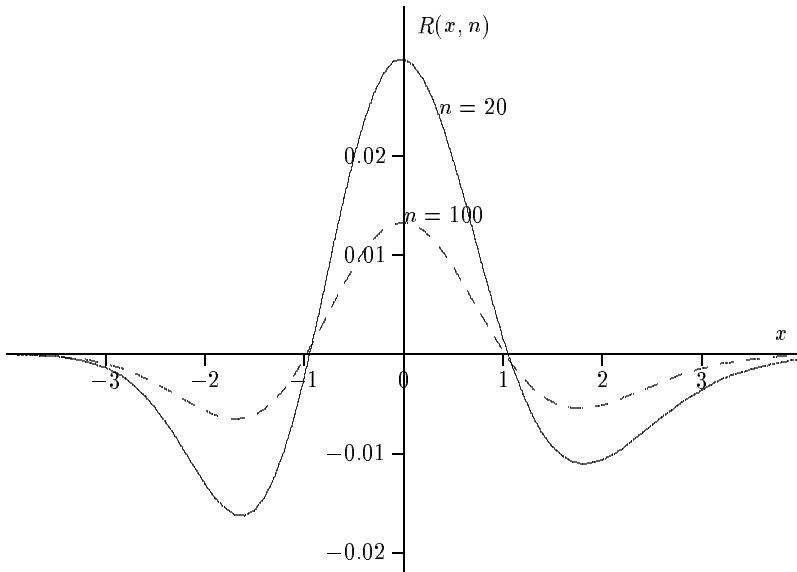
which, due to the fact that  $(1/\lambda)(S_n/n)$  is distributed as  $\Gamma(n, 1/n)$ , does not depend on  $\lambda$ :

$$P_\lambda \left\{ \frac{S_n/n - \lambda}{\lambda} \sqrt{n} \leq x \right\} = \frac{1}{\Gamma(n)} \int_0^{n+x\sqrt{n}} t^{n-1} e^{-t} dt.$$

To prove (B) observe that

$$R(x, n) = P_\lambda \left\{ \frac{S_n/n - \lambda}{\lambda} \sqrt{n} \leq x \right\} - \Phi(x).$$

The function  $R(x, n)$  is continuous and bounded; two examples are exhibited in Fig. 1.

Fig. 1. The function  $R(x, n)$ 

Some values of  $R(x, n)$  presented in Table 2 below enable us to choose a proper  $N$  for typical values of  $\varepsilon$ ; here  $x_n = \arg \max_x R(x, n)$ .

**Table 2.**  $R(x, n)$ 

$n$	$R(x_n, n)$	$x_n$
7	0.050363	-0.931299
8	0.047100	-0.029303
176	0.010025	-0.006282
177	0.009996	-0.006265
707	0.005001	-0.003135
708	0.004998	-0.003132
17 683	0.001000	-0.000540
70 735	0.000500	-0.000224

Explicit formulas neither for  $x_n$  nor for  $R(x_n, n)$  are known to the author. ■

**3. Quantiles.** It is well known (e.g. Serfling 1980) that if  $x_q = x_q(F)$  is the unique quantile of order  $q$  of the distribution  $F$  and  $k(n)/n \rightarrow q$ , then  $X_{k(n):n} \rightarrow x_q$  a.s. Here  $X_{k:n}$  is the  $k$ th order statistic from the sample  $X_1, \dots, X_n$ . The convergence, however, is not uniform: for each  $\varepsilon$ , each  $\eta$ , and every  $n$  one can find a distribution  $F$  with the unique quantile  $x_q$  such that

$$P_F\{|X_{k(n):n} - x_q| > \varepsilon\} \geq 1 - \eta.$$

A necessary and sufficient condition for uniform convergence has been given

in Zieliński (1998). An effective uniform asymptotic theorem for a smaller class of model distributions may be stated as follows. For a fixed  $q \in (0, 1)$ , consider the class  $\mathcal{F}(q, \vartheta)$  of all locally (at the  $q$ th quantile  $x_q$ ) continuous and strictly increasing distributions  $F$  such that the densities  $f$  at the  $q$ th quantile  $x_q$  satisfy  $f(x_q) \geq \vartheta > 0$ .

PROPOSITION 4.

(A)  $\forall \varepsilon > 0 \forall \eta > 0 \exists N = N(\varepsilon, \eta) \forall F \in \mathcal{F}(q, \vartheta)$

$$P_F\left\{\sup_{n \geq N} |X_{k(n):n} - x_q| > \varepsilon\right\} < \eta;$$

(B)  $N(\vartheta, \varepsilon, \eta) \geq -\frac{8 \log\left(\frac{1}{2}(1 - \exp\{-\frac{1}{8}\vartheta^2\varepsilon^2\})\right)\eta}{\vartheta^2\varepsilon^2}.$

*Proof.* If

$$\delta = \inf_{F \in \mathcal{F}} \min\{q - F(x_q - \varepsilon), F(x_q + \varepsilon) - q\}$$

for a class  $\mathcal{F}$  of distributions, then for every  $F \in \mathcal{F}$ ,

$$P_F\left\{\sup_{n \geq N} |X_{k(n):n} - x_q| > \varepsilon\right\} < \frac{2\tau^N}{1 - \tau}$$

with  $\tau = \exp\{-\delta^2/2\}$  (Serfling 1980). In the class  $\mathcal{F}(q, \vartheta)$  we have

$$\lim_{0 < t \rightarrow 0} \frac{F(x_q + t) - q}{t} = \lim_{0 < t \rightarrow 0} \frac{q - F(x_q - t)}{t} \geq \vartheta$$

so that there exists  $t_0 > 0$  such that for all  $t < t_0$ ,

$$F(x_q + t) - q \geq \frac{1}{2}\vartheta t, \quad q - F(x_q - t) \geq \frac{1}{2}\vartheta t,$$

and in consequence, for all sufficiently small  $\varepsilon$  (for  $\varepsilon < t_0$ ),

$$\delta = \min\{q - F(x_q - \varepsilon), F(x_q + \varepsilon) - q\} \geq \frac{1}{2}\vartheta\varepsilon.$$

Now

$$\tau = \exp\{-\delta^2/2\} \leq \exp\{-\frac{1}{8}\vartheta^2\varepsilon^2\}.$$

Solving, with respect to  $N$ , the equation

$$\frac{2\tau^N}{1 - \tau} = \eta$$

we obtain the result. ■

Table 3 below gives us an insight into how large samples are needed to get the prescribed accuracy of the asymptotics.

**Table 3.**  $N(\vartheta, \varepsilon, \eta)$ 

$\eta$	$\vartheta$	$\varepsilon$	
		0.05	0.10
0.1	0.5	159 398	35 414
	1.0	35 414	7 745
	2.0	7 745	1 660
0.01	0.5	188 871	42 782
	1.0	42 782	9 587
	2.0	9 587	2 120

**4. Some ineffective uniform asymptotic results.** Consider the problem of the previous section. As an ineffective asymptotic theorem we have the following corollary (Zieliński 1998): *if  $F$  is a continuous and strictly increasing distribution function and  $k(n)/n \rightarrow q$  then  $X_{k(n):n} \rightarrow x_q$  a.s. uniformly in the family of distributions  $\{F_\theta(x) = F(x - \theta), -\infty < \theta < \infty\}$ .*

Two more general theorems concerning the convergence of

$$s_n(\theta) = \sum a(X_i, \theta),$$

where  $a(X, \theta) = (a_1(X, \theta), \dots, a_m(X, \theta))$  is a given vector-valued function, are taken from Borovkov (1998). To state the theorems recall that an integral  $\int \psi(x, \theta) P_\theta(dx)$  is said to be *convergent uniformly in  $\theta \in \Theta$*  if

$$\sup_{\theta \in \Theta} \int_{|\psi(x, \theta)| > N} |\psi(x, \theta)| P_\theta(dx) \rightarrow 0 \quad \text{as } N \rightarrow \infty.$$

**THEOREM 1 (Uniform law of large numbers).** *If  $a(\theta) = \int a(x, \theta) P_\theta(dx)$  converges uniformly in  $\theta \in \Theta$ , then*

$$P_\theta \left\{ \left| \frac{s_n(\theta)}{n} - a(\theta) \right| > \varepsilon \right\} \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

*uniformly in  $\theta$ .*

To state the central limit theorem assume that  $a(\theta) = 0$  (or take  $a'(X, \theta) = a(X, \theta) - a(\theta)$  instead of  $a(X, \theta)$ ).

**THEOREM 2 (Uniform central limit theorem).** *If  $\int a_j^2(x, \theta) P_\theta(dx)$ ,  $j = 1, \dots, m$ , converges uniformly in  $\theta$ , then  $s_n(\theta)/\sqrt{n}$  converges to a normal random variable  $N(0, \sigma^2(\theta))$  uniformly with respect to  $\theta$ , where  $\sigma^2(\theta) = E_\theta(a^T(X, \theta)a(X, \theta))$ .*

**5. Comments.** Though of great importance for statistical inference, the literature on uniform asymptotic theorems in statistical models, and especially on effective limit laws, is extremely scarce. Perhaps the only two examples of specific theorems for statistical models are the above result on sample quantiles and a general result on uniform consistency of maximum likelihood estimators (Borovkov 1998, Ibragimov *et al.* 1981). Other



uniform versions of asymptotic theorems are mostly constructed as follows: take a probability asymptotic theorem which states that if a distribution under consideration satisfies a condition  $C$  then WLLN (or SLLN, CLT) holds. Then formulate the statistical theorem: if the condition  $C$  is satisfied uniformly in a given statistical model then WLLN (or, respectively, SLLN, CLT) holds uniformly (Ibragimov *et al.* 1981).

If a distribution-free statistic in a model under consideration is available, the problem of uniform limit laws is automatically solved, but constructing an effective limit law may be difficult. As an example consider the Kolmogorov statistic  $D_n = \sup_x |F_n(x) - F(x)|$  in a statistical model with  $F$  continuous; here  $F_n(x)$  is the empirical distribution function. It is well known that the distribution of  $D_n$  does not depend on the specific distribution  $F$  so that the stochastic convergence  $P\{D_n > \varepsilon\} \rightarrow 0$  for every  $\varepsilon > 0$  holds uniformly. That means that for every  $\varepsilon > 0$  and every  $\eta > 0$  there exists  $N = N(\varepsilon, \eta)$  such that for all  $F$  continuous and for all  $n > N$ ,  $P\{D_n > \varepsilon\} < \eta$ . In Birnbaum (1952) one finds that  $N(0.15, 0.1) = 65$  and  $N(0.05, 0.01) = 1,060$ . The values were obtained numerically and no explicit formula for  $N(\varepsilon, \eta)$  is known.

**Acknowledgments.** I wish to thank Dr. Zbigniew Szkutnik for helpful comments and suggested improvements.

## References

- Z. W. Birnbaum (1952), *Numerical tabulation of the distribution of Kolmogorov's statistic for finite sample size*, J. Amer. Statist. Assoc. 47, 425–441.
- A. A. Borovkov (1998), *Mathematical Statistics*, Gordon and Breach.
- L. D. Brown, T. T. Cai and A. DasGupta (2001), *Interval estimation for a Binomial Proportion*, Statist. Sci. 16, 101–133.
- I. A. Ibragimov and R. Z. Has'minskiĭ (1981), *Statistical Estimation. Asymptotic Theory*, Springer.
- R. J. Serfling (1980; new edition 2002), *Approximation Theorems of Mathematical Statistics*, Wiley.
- J. Wesołowski (2002), private communication.
- R. Zieliński (1998), *Uniform strong consistency of sample quantiles*, Statist. Probab. Lett. 37, 115–119.

Institute of Mathematics  
 Polish Academy of Sciences  
 P.O. Box 21  
 00-956 Warszawa 10, Poland  
 E-mail: rziel@impan.gov.pl

Received on 13.5.2003;  
 revised version on 9.9.2003

(1690)