EFFECTIVE WLLN, SLLN, AND CLT IN STATISTICAL MODELS

Ryszard Zieliński

Inst. Math. Polish Acad. Sc.
P.O.Box 21, 00-956 Warszawa 10, Poland
e-mail: rziel@impan.gov.pl

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 than we call the law effective. It is trivial to obtain the 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.

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

1. THE BERNOULLI SCHEME

Let $X, X_1, X_2, \ldots, X_n, \ldots$ be iid random variables with

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

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

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

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

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

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

PROPOSITION 1.

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

$$(B) \qquad \textit{the appropriate } N = N(\varepsilon, \eta) = \frac{1}{4\eta\varepsilon^2}.$$

□

The formula is useful for example for constructing the confidence interval for an unknown θ, with an a priori postulated accuracy and confidence level. Here and further on Part (A) states the uniform convergence and Part (B) makes the law effective. Part (B) may be improved by the argument used in the proof of Proposition 2 below (Bernstein inequality)

SLLN states that, under every fixed $\theta \in (0, 1)$, $S_n/n \to \theta$ *a.s.* Using the fact that $\xi_n \to 0$ *a.s.* iff $\forall \varepsilon > 0\ lim_{N\to\infty} P\{\bigcup_{n=N}^{\infty}\{|X_n| > \varepsilon\}\} = 0$ iff $\forall \varepsilon > 0\, \forall \eta > 0\, \exists N\ P\{\bigcup_{n=N}^{\infty}\{|X_n| > \varepsilon\}\} < \eta$, an appropriate effective statistical version of the law takes on the form

PROPOSITION 2.

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

(B)

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

PROOF.

By a rather crude estimation one obtains

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

and then by the Bernstein inequality for the Bernoulli scheme (Serfling 1980, Jakubowski et al. 2001) in the form

$$P_\theta \{ |\frac{S_n}{n} - \theta| > \varepsilon \} \leq 2e^{-n\varepsilon^2/4}$$

the following estimation holds

$$P_\theta \left\{ \bigcup_{n=N}^{\infty} \{ |\frac{S_n}{n} - \theta| > \varepsilon \} \right\} < \frac{2e^{-N\varepsilon^2/4}}{1 - e^{-\varepsilon^2/4}}$$

which enables us to explicitly fix N as any integer such that

$$N > -\frac{4}{\varepsilon^2} \, log \left(\frac{\eta}{2} \left(1 - e^{-\varepsilon^2/4} \right) \right).$$

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

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

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

The maximal inequality for inverse martingales gives us

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

and in consequence

$$P_\theta\left\{\bigcup_{n=N}^{\infty}\{|\frac{S_n}{n}-\theta|>\varepsilon\}\right\} = P\left\{\sup_{k\geq N}|T_k|\geq\varepsilon\right\}$$
$$= \lim_{m\to\infty} P\left\{\max_{N\leq k\leq m}|T_k|\geq\varepsilon\right\}$$
$$\leq \lim_{m\to\infty}\frac{1}{4N\varepsilon^2} = \frac{1}{4N\varepsilon^2}$$

Now PROPOSITION 2(A) holds for any $N \geq \dfrac{1}{4\varepsilon^2\eta}$ (second line in Table 1).

□

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

ε	η		
	0.1	0.01	0.001
0.1	3,596	4,517	5,438
	250	2,500	25,000
0.01	543,695	635,799	727,902
	25,000	250,000	2,500,000
0.001	$7.28 * 10^7$	$8.20 * 10^7$	$9.12 * 10^7$
	$0.25 * 10^7$	$2.50 * 10^7$	$25 * 10^7$

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

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

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 is

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

or even in a stronger form: "uniformly in x".

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

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

It is sufficient to prove that

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

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 θ 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 similar result holds in the inverse Binomial scheme (negative Binomial distribution). Let Y be the number of experiments needed to observe first success:

$$P_\theta\{Y = y\} = (1-\theta)^{y-1}\theta, \qquad E_\theta Y = \frac{1}{\theta}, \qquad Var_\theta Y = \frac{1-\theta}{\theta^2}.$$

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

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

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

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

2. EXPONENTIAL DISTRIBUTION

If $X_1, X_2, \ldots$ 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*

$$S_n/n \to \lambda \quad a.s.$$

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

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

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 $\Gamma(n, \frac{\lambda}{n})$ with the shape parameter n and the scale parameter λ/n, easily follows from the following estimation

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

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

Before stating the theorem let us define

$$R(x,n) = \begin{cases} \frac{1}{\Gamma(n)} \int\limits_0^{n+x\sqrt{n}} t^{n-1}e^{-t}\,dt - \Phi(x) & \text{if } x > -\sqrt{n} \\ 0 & \text{elsewhere} \end{cases}$$

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

$$(A) \quad \forall \varepsilon > 0 \; \exists N = N(\varepsilon) \; \forall \lambda > 0 \; \sup_x \left| P_\lambda \left\{ \frac{\frac{S_n}{n} - \lambda}{\lambda} \sqrt{n} \le x \right\} - \Phi(x) \right| < \varepsilon$$

and

$$(B) \quad \begin{array}{c} \text{an appropriate } N = N(\varepsilon) \text{ is given numerically as an } N \\ \text{such that } \max_x |R(x,n)| \le \varepsilon \end{array}$$

PROOF. To prove part (A) of the proposition it is enough to observe that

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

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

$$P_\lambda \left\{ \frac{\frac{S_n}{n} - \lambda}{\lambda} \sqrt{n} \le x \right\} = \frac{1}{\Gamma(n)} \int\limits_0^{n+x\sqrt{n}} t^{n-1}e^{-t}\,dt.$$

To prove part (B) observe that

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

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

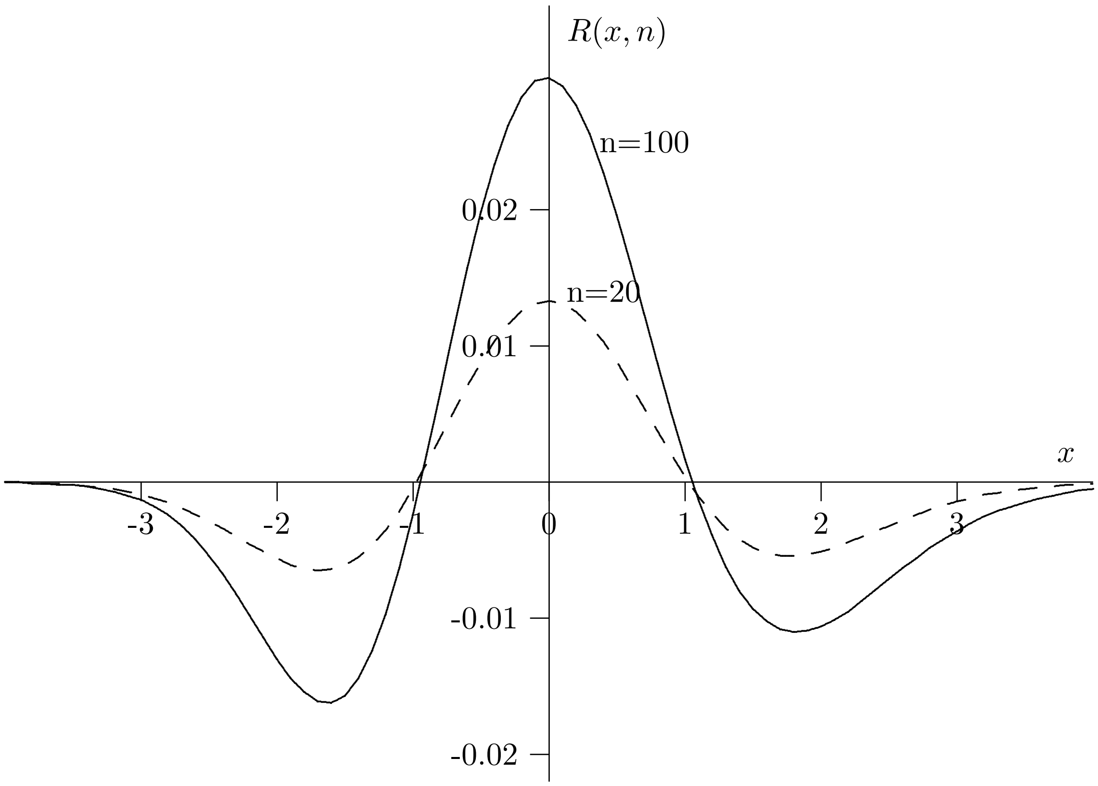

Fig.1. Function R(x,n)

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

Tab. 2. $R(x, n)$

n	$R(x_n, n)$	x_n
7	0.050,363	-0.931,299
8	0.047,100	-0.029,303
176	0.010,025	-0.006,282
177	0.009,996	-0.006,265
707	0.005.001	-0.003,135
708	0.004,998	-0.003,132
17,683	0.001,000	-0.000,540
70,735	0.000,500	-0.000,224

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

$\square$.

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 $k(n)/n \to q$, then $X_{k(n):n} \to x_q$ *a.s.* Here $X_{k:n}$ is the k-th order statistic from the sample $X_1, X_2, \ldots, X_n$. The convergence is however not uniform: for each ε, for each η, and for every n one can find a distribution F with the unique quantile x_q such that

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

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 distributions F such that the densities f at the qth quantile x_q exist and they satisfy $f(x_q) \geq \vartheta > 0$.

PROPOSITION 4.

$$(A) \qquad \forall\varepsilon > 0\ \forall\eta > 0\ \exists N = N(\varepsilon,\eta)\ \forall F \in \mathcal{F}(q,\vartheta) \quad P_F\left\{\sup_{n\geq N}\left|X_{k(n):n} - x_q\right| > \varepsilon\right\} < \eta$$

and

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

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 $F \in \mathcal{F}$

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

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

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

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

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

and in consequence, for all sufficiently small ε (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 in how large samples are needed to get the prescribed accuracy of the asymptotic.

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

		ε	
η	ϑ	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 NON EFFECTIVE UNIFORM ASYMPTOTIC RESULTS

Consider the problem as in the previous Section. As an non effective asymptotic theorem we have the following Corollary (Zieliński 1998): *if F is a continuous and strictly increasing distribution function and $k(n)/n \to q$ then $X_{k(n):n} \to x_q$ 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 $a(X, \theta) = \big(a_1(X, \theta), \ldots, a_m(X, \theta)\big)$ 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 in Θ uniformly with respect to θ if

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

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

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

uniformly in θ.

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 $\int a_j^2(x, \theta) P_\theta(dx)$, $j = 1, \ldots, m$, converge uniformly in θ, then $s_n(\theta)/\sqrt{n}$ converges to a normal random variable $N\Big(0, \sigma^2(\theta)\Big)$ uniformly with respect to θ, where $\sigma^2(\theta) = E_\theta\big(a^T(X, \theta) a(X, \theta)\big)$.*

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 (Borovkow 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 $SLLN$, or CLT) holds. Then formulate the statistical theorem: if the condition C is satisfied uniformly in a given statistical model then $WLLN$ (or, respectively, $SLLN$, or 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 Kolmogov 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 $P\{D_n > \varepsilon\} \to 0$ for every $\varepsilon > 0$ holds uniformly. That means that for every $\varepsilon > 0$ and for 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 reads that $N(0.15, 0.1) = 65$ and $N(0.05, 0.01) = 1,060$. The values were obtained numerically and no explicite formula for $N(\varepsilon, \eta)$ is known.

REFERENCES

Birnbaum, Z.W. (1952): *Numerical tabulation of the distribution of Kolmogorov's statistic for finite sample size*, JASA 47, 425–441

Boratyńska, A. and Zieliński, R. (1997): *Asymptotic behavior of sample median in a parametric model*, Annales UMCS, Sectio A, Vol. LI.1,2, 11–14

Borovkov, A.A. (1998): *Mathematical statistics*, Gordon and Breach

Brown, L.D., Cai, T.T. and DasGupta, A. (2001): *Interval estimation for a Binomial Proportion*, Statistical Science 16, 2, 101–133

Ibragimov, I.A. and Has'minskii (1981): *Statistical estimation. Asymptotic theory*. Springer

Jakubowski, J. and Sztencel, R. (2001): *Wstęp do teorii prawdopodobieństwa*. Wyd. II. Script, Warszawa

Serfling, R.J. (1980): *Approximation theorems of mathematical statistics*. Wiley

Wesolowski, J. (2002): Private communication

Zieliński, R. (1998): *Uniform strong consistency of sample quantiles*, Statist. Probab. Lett. 37, 115-119