ON ASYMPTOTIC APPROXIMATIONS TO THE DISTRIBUTIONS OF STATISTICS CONSTRUCTED FROM SAMPLES WITH RANDOM SIZES

Vladimir Bening
Moscow State University,
Moscow, Russia
RUDN University,
Moscow, Russia

Victor Korolev
Moscow State University,
Moscow, Russia
IPI FRC CSC RAS,
Moscow, Russia
Hangzhou Dianzi University,
Hangzhou, China

Alexander Zeifman
Vologda State University,
Vologda, Russia
IPI FRC CSC RAS;
ISEDT RAS
Email: a_zeifman@mail.ru

KEYWORDS

Sample of random size; asymptotic expansions; mixtures of probability laws; Laplace distribution; Student distribution; transfer theorem.

ABSTRACT

Due to the stochastic character of the intensities of information flows in high performance information systems, the size of data available for the statistical analysis can be often regarded as random. In the paper general theorem concerning the asymptotic expansions of the distribution function of the statistics based on the sample of random size was proved. Some examples are presented for the cases where the sample size has the negative binomial or discrete Pareto distributions.

INTRODUCTION

In classical problems of mathematical statistics, the size of the available sample, i. e., the number of available observations, is traditionally assumed to be deterministic. In the asymptotic settings it plays the role of infinitely increasing *known* parameter. At the same time, in practice very often the data to be analyzed is collected or registered during a certain period of time and the flow of informative events each of which brings a next observation forms a random point process. Therefore, the number of available observations is unknown till the end of the process of their registration and also must be treated as a (random) observation. For example, this is so in insurance statistics where during different accounting periods different numbers of insurance events (insurance claims and/or insurance contracts) occur and in high performance information systems where due to the stochastic character of the intensities of information flows, the size of data available for the statistical analysis can be often regarded as random. Say, the statistical algorithms applied in high-frequency financial applications must take into consideration that the number of events in a limit order book during a time unit essentially depends on the intensity of order flows. Moreover, contemporary statistical procedures of insurance and financial mathematics do take this circumstance into consideration as one of possible ways of dealing with heavy tails. However, in other fields such as medical statistics or quality control this approach has not become conventional yet although the number of patients with a certain disease varies from month to month due to seasonal factors or from year to year due to some epidemic reasons and the number of failed items varies from lot to lot. In these cases the number of available observations as well as the observations themselves are unknown beforehand and should be treated as random to avoid underestimation of risks or error probabilities.

In asymptotic settings, statistics constructed from samples with random sizes are special cases of random sequences with random indices. The randomness of indices usually leads to that the limit distributions for the corresponding random sequences are heavy-tailed even in the situations where the distributions of non-randomly indexed random sequences are asymptotically normal see, e. g., [1], [2], [3]. For example, if a statistic which is asymptotically normal in the traditional sense, is constructed on the basis of a sample with random size having negative binomial

Proceedings 31st European Conference on Modelling and
Simulation ©ECMS Zita Zoltay Paprika, Péter Horák, Kata Váradi,
Péter Tamás Zwierczyk, Ágnes Vidovics-Dancs, János Péter Rádics (Editors)
ISBN: 978-0-9932440-4-9/ ISBN: 978-0-9932440-5-6 (CD)

distribution, then instead of the expected normal law, the Student distribution with power-type decreasing heavy tails appears as an asymptotic law for this statistic.

In the present paper, asymptotic expansions (a.e:s) are obtained for the distribution functions (d.f:s) of statistics constructed from samples with random sizes. These results continue the studies started in [1] – [6]. The obtained a.e:s directly depend on the a.e. for the d.f. of the random sample size and on the a.e. for the d.f. of the statistic based on the sample with a non-random size. Such statements are conventionally called transfer theorems. So, we may say that in this paper transfer theorems are presented for the a.e:s of the d.f:s of statistics constructed from samples with random size. Unlike previous works, here we concentrate our attention on the case of non-normalized statistics.

We use conventional notation: $\mathbb{R}$ is the set of real numbers, $\mathbb{N}$ is the set of natural numbers, $\Phi(x)$ and $\varphi(x)$ are the d.f. and the probability density of the standard normal law, respectively.

In Section 2 the main result is presented and proved, Section 3 contain some examples, Section 4 is devoted to the normalized statistics.

MAIN RESULTS FOR NON-NORMALIZED STATISTICS

Consider random variables (r.v:s) $N_1, N_2, \ldots$ and $X_1, X_2, \ldots$, defined on the same probability space $(\Omega, \mathcal{A}, \mathsf{P})$. The r.v:s $X_1, X_2, \ldots X_n$ will be treated as observations with n being a non-random sample size, whereas the r.v:s N_n will be treated as random sample sizes depending on the parameter $n \in \mathbb{N}$. For example, if the r.v. N_n has the geometric distribution $\mathsf{P}\big(N_n = k\big) = frac1n(1-\frac{1}{n})^{k-1}$, $k \in \mathbb{N}$, then $\mathsf{E}N_n = n$, that is, the r.v. N_n is parameterized by its expectation n.

Assume that for each $n \geq 1$ the r.v. N_n takes only natural values, that is, $N_n \in \mathbb{N}$ and are independent of the sequence $X_1, X_2, \ldots$. Everywhere in what follows consider the r.v:s $X_1, X_2, \ldots$ to be independent and identically distributed (i.i.d) with some common d.f. $F(x)$. By $T_n = T_n(X_1, \ldots, X_n)$ denote a statistic, that is, real measurable function of observations $X_1, \ldots, X_n$. We focus on the situation where the number of available observations is large, that is, $n \to \infty$. Assume that the d.f. of the non-normalized statistic T_n weakly converges to some d.f. $G(x)$, that is,

$$\mathsf{P}\big(T_n < x\big) \longrightarrow G(x), \quad n \to \infty, \qquad (2.1)$$

In every continuity point of $G(x)$. Assume that, as $n \to \infty$, the random sample size N_n tends to infinity in probability, that is, for any $M > 0$

$$\mathsf{P}(N_n \leq M) \longrightarrow 0, \quad n \to \infty. \qquad (2.2)$$

Consider the limit behavior of the d.f. of the statistic constructed from the sample of a random size, that is, of the statistic $T_{N_n}(\omega) \equiv T_{N_n(\omega)}(X_1(\omega), \ldots, X_{N_n(\omega)}(\omega))$, $\omega \in \Omega$. As is shown in the following lemma, under the conditions (2.1) and (2.2) the limit law for T_{N_n} is the same as for T_n.

Lemma 2.1. *Let conditions* (2.1) *and* (2.2) *hold. Then* $\mathsf{P}\big(T_{N_n} < x\big) \longrightarrow G(x)$, $n \to \infty$, *at every continuity point of* $G(x)$.

The **proof** is a simple exercise on the application of the formula of total probability.

Now assume that the d.f. of the *non-normalized* statistic T_n admits an a.e. described by the following condition.

Condition 1. *There exist constants* $l \in \mathbb{N}$, $\alpha > \frac{1}{2}$, $C_1 > 0$, *a differentiable d.f.* $G(x)$ *and differentiable bounded functions* $g_i(x)$, $i = 1, \ldots, l$ *such that*

$$\sup_x \Big|\mathsf{P}\big(T_n < x\big) - G(x) - \sum\nolimits_{i=1}^{l} \frac{g_i(x)}{n^{i/2}}\Big| \leq \frac{C_1}{n^{\alpha}}, \ n \in \mathbb{N}.$$

Also assume that the d.f. of the *normalized* random sample size T_n admits an a.e. described by the following condition.

Condition 2. *There exist constants* $m \in \mathbb{N}$, $\beta > m/2$, $C_2 > 0$, *functions* $0 < v(n) \uparrow \infty$, $(n \to \infty)$, $u(n) \in \mathbb{R}$, *a d.f.* $H(x)$ *with* $H(0+) = 0$ *and functions with bounded variations* $h_j(x)$, $j = 1, \ldots, m$, *such that*

$$\sup_{x \geq 0} \Big|\mathsf{P}\Big(\frac{N_n}{v(n)} - u(n) < x\Big) - H(x) -$$
$$- \sum\nolimits_{j=1}^{m} \frac{h_j(x)}{n^{j/2}}\Big| \leq \frac{C_2}{n^{\beta}}, \quad n \in \mathbb{N}.$$

Everywhere in what follows we denote $y_n = y - u(n)$.

Theorem 2.1. *Let conditions* 1 *and* 2 *hold. Then*

$$\sup_x \Big|\mathsf{P}\big(T_{N_n} < x\big) - G_n(x)\Big| \leq$$
$$\leq C_1 \mathsf{E} N_n^{-\alpha} + 2\frac{C_2}{n^{\beta}} \sup_x \sum\nolimits_{i=1}^{l} |g_i(x)|,$$

where

$$G_n(x) = G(x) + \sum\nolimits_{i=1}^{l} \frac{g_i(x)}{(v(n))^{i/2}} \int_{1/v(n)}^{\infty} y^{-i/2} \times$$
$$\times d\Big(H(y_n) + \sum\nolimits_{j=1}^{m} n^{-j/2} h_j(y_n)\Big) =$$

$$= G(x)+\sum_{i=1}^{l} g_i(x)\int_1^\infty z^{-i/2} d\Big(H(z/v(n)-u(n))+$$

$$+\sum_{j=1}^{m} n^{-j/2}\; h_j(z/v(n)-u(n))\Big).$$

Proof. Estimate the difference under consideration above in the following way:

$$\sup_x\Big|\mathsf{P}\big(T_{N_n}<x\big)-G_n(x)\Big|\le I_{1n}+I_{2n}, \qquad (2.4)$$

where

$$I_{1n}\equiv\sup_x\Big|\mathsf{P}\big(T_{N_n}<x\big)-\mathsf{E}\Big(G(x)+\sum_{i=1}^{l}\frac{g_i(x)}{N_n^{i/2}}\Big)\Big|,$$

$$I_{2n}\equiv\sup_x\Big|\mathsf{E}\Big(G(x)+\sum_{i=1}^{l}\frac{g_i(x)}{N_n^{i/2}}\Big)-G_n(x)\Big|.$$

Estimate I_{1n} using condition 1 and the formula of total probability. We have

$$I_{1n}=\sup_x\Big|\sum_{k=1}^{\infty}\mathsf{P}\big(N_n=k\big)\Big(\mathsf{P}\big(T_k<x\big)-$$

$$-G(x)-\sum_{i=1}^{l}k^{-i/2}g_i(x)\Big)\Big|\le$$

$$\le\sum_{k=1}^{\infty}\mathsf{P}\big(N_n=k\big)\sup_x\Big|\mathsf{P}\big(T_k<x\big)-$$

$$-G(x)-\sum_{i=1}^{l}k^{-i/2}g_i(x)\Big|\le$$

$$\le C_1\sum_{k=1}^{\infty}\frac{1}{k^\alpha}\mathsf{P}\big(N_n=k\big)=C_1\mathsf{E}N_n^{-\alpha}. \qquad (2.5)$$

To estimate I_{2n}, use condition 2 and integration by parts. We have

$$I_{2n}=\sup_x\Big|\mathsf{E}\Big(G(x)+\sum_{i=1}^{l}N_n^{-i/2}g_i(x)\Big)-G(x)-$$

$$-\sum_{i=1}^{l}\big(v(n)\big)^{-i/2}g_i(x)\int_{1/v(n)}^{\infty}y^{-i/2}d\Big(H(y_n)+$$

$$+\sum_{j=1}^{m}\frac{h_j(y_n)}{n^{j/2}}\Big)\Big|=\sup_x\Big|\sum_{i=1}^{l}\mathsf{E}N_n^{-i/2}g_i(x)-$$

$$-\sum_{i=1}^{l}g_i(x)(v(n))^{-i/2}\int_{1/v(n)}^{\infty}\frac{1}{y^{i/2}}\times$$

$$\times d\Big(H(y_n)+\sum_{j=1}^{m}\frac{h_j(y_n)}{n^{j/2}}\Big)\Big|=$$

$$=\sup_x\Big|\sum_{i=1}^{l}g_i(x)\int_1^\infty y^{-i/2}d\mathsf{P}\big(N_n<y\big)-$$

$$-\sum_{i=1}^{l}g_i(x)\big(v(n)\big)^{-i/2}\int_{1/v(n)}^{\infty}y^{-i/2}\times$$

$$\times d\Big(H(y_n)-\sum_{j=1}^{m}\frac{h_j(y_n)}{n^{j/2}}\Big)\Big|.$$

Changing the variables in the first integral we obtain $I_{2n}=$

$$=\sup_x\Big|\sum_{i=1}^{l}\frac{g_i(x)}{(v(n))^{i/2}}\int_{\frac{1}{v(n)}}^{\infty}\frac{1}{y^{i/2}}d\Big(\mathsf{P}\Big(\frac{N_n}{v(n)}<y\Big)-$$

$$-H(y_n)-\sum_{j=1}^{m}\frac{h_j(y_n)}{n^{j/2}}\Big)\Big|.$$

Using integration by parts, the boundedness of the functions $g_i(x)$, $i=1,...,l$ and condition 2, we obtain the inequalities

$$I_{2n}\le\frac{C_2}{n^\beta}\sup_x\sum_{i=1}^{l}\big|g_i(x)\big|+$$

$$+\sup_x\Big|\sum_{i=1}^{l}\frac{g_i(x)}{(v(n))^{i/2}}\int_{\frac{1}{v(n)}}^{\infty}\Big(\mathsf{P}\Big(\frac{N_n}{v(n)}-u(n)<$$

$$<y_n\Big)-H(y_n)-\sum_{j=1}^{m}n^{-j/2}h_j(y_n)\Big)dy^{-i/2}\Big)\Big|\le$$

$$\le 2\frac{C_2}{n^\beta}\sup_x\sum_{i=1}^{l}\big|g_i(x)\big|. \qquad (2.6)$$

Now the desired assertion follows from (2.4), (2.5) and (2.6). The theorem is proved.

EXAMPLES

Here we present two examples of the application of theorem 2.1 for statistics constructed from samples with special random sample sizes. We will consider the a.e:s for the d.f. of the sample mean constructed from samples with random sizes. Similar results can be obtained for statistics admitting the Edgeworth-type a.e:s for the d.f. under a non-random sample size.

Let $X_1, X_2, \ldots$ be i.i.d. r.v:s with $\mathsf{E}X_1=\mu$, $0<\mathsf{D}X_1=\sigma^2$, $\mathsf{E}|X_1|^{3+2\delta}<\infty$, $\delta\in(0,\frac{1}{2})$ and $\mathsf{E}\big(X_1-\mu\big)^3=\mu_3$. For $n\in\mathbb{N}$ denote

$$T_n=\frac{X_1+\ldots+X_n-n\mu}{\sigma\sqrt{n}}. \qquad (3.1)$$

Also assume that the r.v. X_1 satisfies the CramEr condition (C)

$$\limsup_{|t|\to\infty}|\mathsf{E}\exp\{itX_1\}|<1.$$

Then with the account of theorem 6.3.2 from [7] we obtain

$$\sup_x\Big|\mathsf{P}\big(T_n<x\big)-\Phi(x)-$$

$$-\frac{\mu_3}{6\sqrt{n}\sigma^3}(1-x^2)\varphi(x)\Big|\le\frac{C_1}{n^{1/2+\delta}} \qquad (3.2)$$

with $C_1>0$, $\delta\in(0,\frac{1}{2})$, $n\in\mathbb{N}$. Thus, statistic (3.1) satisfies condition 1 of theorem 2.1 with $\alpha=\frac{1}{2}+\delta$, $l=1$, $G(x)=\Phi(x)$, $g_1(x)=\frac{\mu_3}{6\sigma^3}(1-x^2)\varphi(x)$. It is easy to see that $\sup_x|g_1(x)|<\infty$.

A. Sample size with the negative binomial distribution

Let the random sample size N_n have the negative binomial distribution with parameters $p = \frac{1}{n}$ and $r > 0$, that is

$$\mathsf{P}\big(N_n = k\big) = \frac{(k+r-2)\cdots r}{(k-1)!}\frac{1}{n^r}\Big(1-\frac{1}{n}\Big)^{k-1}, \quad k \in \mathbb{N}. \tag{3.4}$$

With $r = 1$ we obtain the geometric distribution. In [8] (relation (6.112) on p. 233) the following bound was presented for the rate of convergence of the normalized sample size N_n to the gamma-distribution:

$$\sup_{x\geq 0}\Big|\mathsf{P}\Big(\frac{N_n}{\mathsf{E}N_n} < x\Big) - H_r(x)\Big| \leq \begin{cases} \dfrac{C_r}{n}, & r \geq 1, \\ \dfrac{C_r}{n^r}, & r \in (0,1), \end{cases} \tag{3.5}$$

where $C_r > 0$, $n \in \mathbb{N}$ and

$$H_r(x) = \frac{r^r}{\Gamma(r)}\int_0^x e^{-ry}y^{r-1}dy, \quad x \geq 0, \tag{3.6}$$

is the gamma-d.f. with shape parameter $r > 0$ coinciding with scale parameter. In this case

$$\mathsf{E}N_n = r(n-1) + 1. \tag{3.7}$$

So, from (3.5) – (3.7) it follows that the random sample size N_n satisfies condition 2 with $v(n) = r(n-1)+1$, $H(x) = H_r(x)$, $m = 1$, $h_1(x) \equiv 0$, $C_2 = C_r > 0$, $u(n) = 0$, $n \in \mathbb{N}$,

$$\beta = \begin{cases} 1, & r \geq 1, \\ r, & r \in (1/2, 1). \end{cases}$$

Further, using the equality $\big(1+x\big)^\gamma = \sum_{k=0}^\infty \gamma(\gamma-1)\cdots(\gamma-k+1)x^k/k!$, $|x| < 1$, $\gamma \in \mathbb{R}$, it is easy to obtain that for $r > 0$, $r \neq 1$, $n \in \mathbb{N}$

$$\mathsf{E}N_n^{-1} = [(n-1)(1-r)]^{-1}(n^{1-r} - 1) = O(n^{-r}). \tag{3.8}$$

For $r = 1$ we have $\mathsf{E}N_n^{-1} = (n-1)^{-1}\log n$, $n > 1$. Now using the Hölder inequality, we obtain

$$\mathsf{E}N_n^{-\alpha} \leq \big(\mathsf{E}N_n^{-1}\big)^\alpha (\alpha \leq 1),$$

$$\mathsf{E}N_n^{-\alpha} = O(n^{-r(1/2+\delta)}), \ r > 0, \ r \neq 1,$$

$$\mathsf{E}N_n^{-\alpha} = O\Big(\Big(\frac{\log n}{n}\Big)^{1/2+\delta}\Big), \quad r = 1, \ n \in \mathbb{N}. \tag{3.9}$$

So, using theorem 2.1, relations (3.2), (3.3), (3.5)–(3.9) and the equality

$$\int_{(r(n-1)+1)^{-1}}^\infty \sqrt{y}dH_r(y) = \int_0^\infty \sqrt{y}dH_r(y) + O\big(1/n\big) =$$

$$= \int_0^\infty \sqrt{y}\frac{r^r}{\Gamma(r)}e^{-ry}y^{r-1}dy + O\big(1/n\big) =$$

$$= \frac{r^r}{\Gamma(r)}\int_0^\infty e^{-y}\frac{y^{r-1/2}}{r^{r+1/2}}dy + O\big(1/n\big) =$$

$$= \frac{\Gamma(r+1/2)}{\Gamma(r)\sqrt{r}} + O\big(1/n\big),$$

we obtain the following assertion.

Theorem 3.1. *Let a statistic T_n have the form (3.1), where $X_1, X_2, \ldots$ are i.i.d. r.v:s with $\mathsf{E}X_1 = \mu$, $0 < \mathsf{D}X_1 = \sigma^2$, $\mathsf{E}|X_1|^{3+2\delta} < \infty$, $\delta \in (0, \frac{1}{2})$ and $\mathsf{E}(X_1 - \mu)^3 = \mu_3$. Also assume that the r.v. X_1 satisfies the Cramér condition (C). Assume that the r.v. N_n has the negative binomial distribution (3.4) with some $r > 0$. Then for $r > (1+2\delta)^{-1}$, as $n \to \infty$, the d.f. of T_{N_n} admits the a.e.*

$$\sup_x\Big|\mathsf{P}\big(T_{N_n} < x\big) - \Phi(x) -$$

$$-\frac{\mu_3\Gamma(r+1/2)}{6\sigma^3\Gamma(r)\sqrt{r^2(n-1)+r}}(1-x^2)\varphi(x)\Big| =$$

$$= \begin{cases} O\big((n^{-1}\log n)\big)^{1/2+\delta}, & r = 1, \\ O\big(n^{-\min\{1, r(1/2+\delta)\}}\big), & r > 1, \\ O\big(n^{-r(1/2+\delta)}\big), & (1+2\delta)^{-1} < r < 1. \end{cases}$$

B. Sample size with the discrete Pareto distribution

In [4], an example related to the theory of records was considered of a sequence of r.v:s $N_n(s)$ depending on a natural parameter $s \in \mathbb{N}$ such that

$$\mathsf{P}\big(N(s) \geq k\big) = \frac{s}{s+k-1}, \quad k \geq 1 \tag{3.10}$$

(also see [9], [10]). Let now $N^{(1)}(s), N^{(2)}(s), \ldots$ be i.i.d. r.v:s with distribution (3.10). Define the r.v:s $N_n(s) = \max_{1\leq j\leq n} N^{(j)}(s)$. Then, as was shown in [4],

$$\lim_{n\to\infty}\mathsf{P}\Big(\frac{N_n(s)}{n} < x\Big) = e^{-s/x}, \quad x > 0. \tag{3.11}$$

In [5] the following bound of the rate of convergence in (3.10) was obtained: there exists a constant $C_s \in (0, \infty)$ such that

$$\sup_{x\geq 0}\Big|\mathsf{P}\Big(\frac{N_n(s)}{n} < x\Big) - e^{-s/x}\Big| \leq \frac{C_s}{n}, \quad n \in \mathbb{N}. \tag{3.12}$$

So, from (3.12) it follows that the r.v. $N_n(s)$ satisfies condition 2 of theorem 2.1 with

$$v(n) = n, \quad H(x) = e^{-s/x}, \quad m = 1, \quad h_1(x) \equiv 0,$$

$$C_2 = C_s > 0, \quad u(n) = 0, \quad \beta = 1. \tag{3.13}$$

Consider $\mathsf{E}N_n^{-1}(s)$ in more detail. From the definition of $N_n(s)$ and (3.10) we have

$$\mathsf{P}\big(N_n(s)=k\big)=\Big(\frac{k}{s+k}\Big)^n-\Big(\frac{k-1}{s+k-1}\Big)^n= \\ =sn\int_{k-1}^{k}\frac{x^{n-1}}{(s+x)^{n+1}}dx.$$

Therefore,

$$\mathsf{E}N_n^{-1}(s)=\sum\nolimits_{k=1}^{\infty}\frac{1}{k}\mathsf{P}\big(N_n(s)=k\big)= \\ =sn\sum\nolimits_{k=1}^{\infty}\frac{1}{k}\int_{k-1}^{k}\frac{x^{n-1}}{(s+x)^{n+1}}dx\leq \qquad (3.14) \\ \leq sn\sum\nolimits_{k=1}^{\infty}\int_{k-1}^{k}\frac{x^{n-2}}{(s+x)^{n+1}}dx= \\ =sn\int_{0}^{\infty}\frac{x^{n-2}}{(s+x)^{n+1}}dx=\frac{1}{s(n-1)}=O(n^{-1}),$$

see [11], formula 856.12. Now using the Hölder inequality we obtain

$$\mathsf{E}N_n^{-\alpha}\leq\big(\mathsf{E}N_n^{-1}\big)^{\alpha},\ \alpha\leq 1, \\ \mathsf{E}N_n^{-\alpha}=O(n^{-(1/2+\delta)}),\ \alpha=1/2+\delta. \qquad (3.15)$$

Now from theorem 2.1, relations (3.12)–(3.15) and the equality

$$\int_{n^{-1}}^{\infty}\sqrt{y}de^{-s/y}=\sqrt{s}\int_{0}^{\infty}y^{-1/2}e^{-y}dy+O(n^{-1})= \\ =\sqrt{s}\Gamma(1/2)+O(n^{-1})=\sqrt{\pi s}+O(n^{-1})$$

we directly obtain the following theorem.

Theorem 3.2. *Let a statistic* T_n *have the form* (3.1), *where* $X_1, X_2, \ldots$ *are i.i.d. r.v:s with* $\mathsf{E}X_1=\mu$, $0<\mathsf{D}X_1=\sigma^2$, $\mathsf{E}|X_1|^{3+2\delta}<\infty$, $\delta\in(0,\frac{1}{2})$ *and* $\mathsf{E}(X_1-\mu)^3=\mu_3$. *Also assume that the r.v.* X_1 *satisfies the Cramér condition* (C). *Assume that the r.v.* N_n *has the discrete Pareto distribution* (3.10). *Then, as* $n\to\infty$, *the d.f. of* $T_{N_n(s)}$ *admits the a.e.*

$$\sup_x\Big|\mathsf{P}\big(T_{N_n(s)}<x\big)-\Phi(x)- \\ -\frac{\mu_3\sqrt{\pi s}}{6\sigma^3\sqrt{n}}(1-x^2)\varphi(x)\Big|=O\Big(\frac{1}{n^{1/2+\delta}}\Big).$$

CONCLUSION

Due to the stochastic character of the intensities of information flows in high performance information systems, the size of data available for the statistical analysis can be often regarded as random. In the paper general theorem concerning the asymptotic expansions of the distribution function of the statistics based on the sample of random size was proved. Some examples are presented for the cases where the sample size has the negative binomial or discrete Pareto distributions.

ACKNOWLEDGEMENT

The publication was financially supported by the Ministry of Education and Science of the Russian Federation (the Agreement number 02.a03.21.0008) and by Russian Foundation for Basic Research, project 15-07-02652.

REFERENCES

[1] *Gnedenko B. V., Korolev V. Yu.* Random Summation. Limit Theorems and Applications. – Boca Raton: CRC Press, 1996.

[2] *Bening V. E., Korolev V. Yu.* Generalized Poisson Models and Their Applications in Insurance and Finance. – Utrecht: VSP, 2002.

[3] *Bening V. E., Korolev V. Yu.* On an application of the Student distribution in the theory of probability and mathematical statistics // Theory of Probability and Its Applications, 2005. Vol. 49. No. 3. P. 377–391.

[4] *Bening V. E., Korolev V. Yu.* Some statistical problems related to the Laplace distribution // Informatics and Its Applications, 2008. Vol. 2. No. 2. P. 19 – 34.

[5] *Lyamin O. O.* On the rate of convergence of the distributions of some statistics to the Laplace and Student distributions // Bulletin of Moscow University. Ser. 15 Computational Mathematics and Cybernetics, 2011. No. 1. P. 39 – 47.

[6] *Bening V. E., Galieva N. K, Korolev V. Yu.* Asymptotic expansions for the distributiond functions of statistics constructed from samples with random sizes // Informatics and Its Applications, 2013. Vol. 7. No. 2. P. 75 – 83.

[7] *Petrov V. V.* Sums of Independent Random Variables. – Berlin–New York: Springer, 1975.

[8] *Bening V. E., Korolev V. Yu., Sokolov I. A., Shorgin S. Ya.* Randomized Models and Methods Of the Theory of Reliability of Information and Technical Systems. – Moscow: Torus Press, 2007 (in Russian).

[9] *Wilks S. S.* Recurrence of extreme observations // Journal of American Mathematical Society, 1959. V. 1, No. 1, P. 106 – 112.

[10] *Nevzorov V. B.* Records: Mathematical Theory. – Providence, RI: American Mathematical Society, 2000. [15]

[11] *Dwight H. B.* Tables of Integrals and Other Mathematical Data. – New York: The MacMillan Company, 1961.

AUTHOR BIOGRAPHIES

VLADIMIR BENING is Doctor of Science in physics and mathematics, professor, Faculty of Computational Mathematics and Cybernetics, M.V. Lomonosov Moscow State University; senior researcher, Peoples' Friendship University of Russia (RUDN University), Moscow. His email is `bening@yandex.ru`.

VICTOR KOROLEV is Doctor of Science in physics and mathematics, professor, Head of Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M.V. Lomonosov Moscow State University; leading scientist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences; professor, Hangzhou Dianzi University, Hangzhou, China. His email is victoryukorolev@yandex.ru.

ALEXANDER ZEIFMAN is Doctor of Science in physics and mathematics; professor, Heard of Department of Applied Mathematics, Vologda State University; senior scientist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences; principal scientist, Institute of Socio-Economic Development of Territories, Russian Academy of Sciences. His email is a_zeifman@mail.ru.