Transnational Journal of Mathematical Analysis and Applications
Vol. 8, Issue 1, 2020, Pages 1-15
ISSN 2347-9086
Published Online on December 28, 2020

http://jyotiacademicpress.org

RANDOM SIZE SAMPLES, ASYMPTOTIC EXPANSIONS AND THE DEFICIENCY CONCEPT IN STATISTICS

V. E. BENING

Faculty of Computational Mathematics and Cybernetics
Lomonosov Moscow State University
Russia
e-mail: bening@yandex.ru

Abstract

Due to a 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 deficiencies of some estimators and 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.

1. 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

2020 Mathematics Subject Classification: Primary 60F05; Secondary 62E20, 91B30, 91B70.
Keywords and phrases: sample of random size, asymptotic expansions, asymptotic deficiency, statistical estimator, risk function, asymptotic relative efficiency.
Research supported by the Moscow Center for Fundamental and Applied Mathematics.
Received October 9, 2020; Revised November 23, 2020

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 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 the 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]-[3], [11]. 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 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 and deficiencies of some statistical estimators are calculated. These results continue the studies started in [1]-[6], [11]. 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.

2. 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

$$\mathsf{P}(N_n = k) = \frac{1}{n}\left(1 - \frac{1}{n}\right)^{k-1}, \; k \in N,$$

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,

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

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 $M > 0$,

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

Consider the limit behaviour 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*

$$\mathsf{P}(T_{N_n} < x) \to G(x), \quad n \to \infty,$$

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 $l \in \mathbb{N}$, $\alpha > l/2$, $C_1 > 0$, a differentiable d.f. $G(x)$ and differentiable bounded functions $g_i(x)$, $i = 1, \dots, l$ such that

$$\sup_x \left| \mathsf{P}(T_n < x) - G(x) - \sum_{i=1}^{l} n^{-i/2} g_i(x) \right| \le \frac{C_1}{n^{\alpha}}, \quad n \in \mathbb{N}.$$

Also assume that the d.f. of the *normalized* random sample size N_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 $h_j(x)$, $j = 1, \dots, m$, such that

$$\sup_{x \ge 0} \left| \mathsf{P}\left(\frac{N_n}{v(n)} - u(n) < x \right) - H(x) - \sum_{j=1}^{m} n^{-j/2} h_j(x) \right| \le \frac{C_2}{n^{\beta}}, \quad n \in \mathbb{N}.$$

Theorem 2.2. *Let Conditions* 1 *and* 2 *hold. Then*

$$\sup_x \left| \mathsf{P}(T_{N_n} < x) - G_n(x) \right| \le C_1 \,\mathsf{E}\, N_n^{-\alpha} + 2 \frac{C_2}{n^{\beta}} \sup_x \sum_{i=1}^{l} |g_i(x)|,$$

where

$$G_n(x) = G(x) + \sum_{i=1}^{l} (v(n))^{-i/2} g_i(x)$$

$$\times \int_{1/v(n)}^{\infty} y^{-i/2} d\left(H(y - u(n)) + \sum_{j=1}^{m} n^{-j/2} h_j(y - u(n)) \right)$$

$$= G(x) + \sum_{i=1}^{l} g_i(x)$$

$$\times \int_{1}^{\infty} z^{-i/2} d\left(H(z / v(n) - u(n)) + \sum_{j=1}^{m} n^{-j/2} h_j(z / v(n) - u(n)) \right).$$

Proof. It follows from Theorem 3.1 in [6] in case of non-normalized statistics T_{N_n}.

3. Examples

Here we present two examples of the application of Theorem 2.2 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 $\mathsf{E}(X_1 - \mu)^3 = \mu_3$. For $n \in \mathbb{N}$ denote

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

Also assume that the r.v. X_1 satisfies the Cramer condition (C). Then with the account of Theorem 6.3.2 from [7] (see also [12]), we obtain

$$\sup_x \left| \mathsf{P}(T_n < x) - \Phi(x) - \frac{\mu_3}{6\sqrt{n}\sigma^3}(1 - x^2)\varphi(x) \right| \le \frac{C_1}{n^{1/2+\delta}} \tag{3.2}$$

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 $\sup_x |g_1(x)| < \infty$.

3.1. Sample size with the negative binomial distribution

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

$$\mathsf{P}(N_n = k) = \frac{(k + r - 2)\cdots r}{(k-1)!}\frac{1}{n^r}\left(1 - \frac{1}{n}\right)^{k-1}, \quad k \in \mathbb{N}. \tag{3.3}$$

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:

$$\sup_{x \geq 0} \left| \mathsf{P}\left(\frac{N_n}{\mathsf{E} N_n} < x \right) - H_r(x) \right| = O(n^{-\min(1, r)}),$$

where $C_r > 0$, $n \in \mathbb{N}$ and $H_r(x)$ is the gamma-d.f. with shape parameter $r > 0$ coinciding with scale parameter.

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 Cramer condition* (C). *Assume that the r.v.* N_n *has the negative binomial distribution* (3.3) *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.*

$$\sup_x \left| \mathsf{P}(T_{N_n} < x) - \Phi(x) - \frac{\mu_3 r^{1/2} \Gamma(r - 1/2)}{6\sigma^3 \Gamma(r) \sqrt{r(n-1) + 1}} (1 - x^2) \varphi(x) \right|$$

$$= O(n^{-\min(1, r(1/2 + \delta))}).$$

3.2. Sample size with the discrete Pareto distribution

In [9], 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

$$\mathsf{P}(N(s) \geq k) = \frac{s}{s + k - 1}, \quad k \geq 1 \tag{3.4}$$

(also see [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],

$$\lim_{n \to \infty} \mathsf{P}\left(\frac{N_n(s)}{n} < x \right) = e^{-s/x}, \quad x > 0. \tag{3.5}$$

In [5], the following bound of the rate of convergence in (3.5) was obtained

$$\sup_{x \geq 0} \left| \mathsf{P}\left(\frac{N_n(s)}{n} < x \right) - e^{-s/x} \right| = O(n^{-1}), \quad n \in \mathbb{N}. \tag{3.6}$$

Now from Theorem 2.2, relations (3.4)-(3.6) 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 Cramer condition* (C). *Assume that the r.v.* N_n *has the discrete Pareto distribution* (3.4). *Then, as* $n \to \infty$, *the d.f. of* $T_{N_n(s)}$ *admits the a.e.*

$$\sup_x \left| \mathsf{P}(T_{N_n(s)} < x) - \Phi(x) - \frac{\mu_3 \sqrt{\pi}}{12\sigma^3 \sqrt{ns}} (1 - x^2)\varphi(x) \right| = O(n^{-1/2-\delta}).$$

4. The Concept of Deficiency

Before turning to the general case of statistics constructed from samples with random size, that is the main aim of the present paper, let us recall the notion of a deficiency of a statistical estimator for the traditional case where the sample size is non-random.

Suppose that $T_n^*(X_1, \ldots, X_n)$ and $T_n(X_1, \ldots, X_n)$ are two competing estimators of $g(\theta)$, $\theta \in \Theta$ based on n observations $X_1, \ldots, X_n$ and let their expected squared errors (risk functions) be denoted by $R_n^*(\theta)$ and $R_n(\theta)$, respectively. An interesting quantitative comparison can be obtained by taking a viewpoint similar to that of the asymptotic relative efficiency (ARE) of estimators, and asking for the number $m(n)$ of observations needed by estimator $T_{m(n)}(X_1, \ldots, X_{m(n)})$ to match the performance of $T_n^*(X_1, \ldots, X_n)$ (based on n observations). Asymptotic

$(n \to \infty)$ comparison of the two estimators involves the comparison of $m(n)$ with n, and this can be carried out in various ways. Although the difference $m(n) - n$ seems to be a very natural quantity to examine, historically the ratio $n / m(n)$ was preffered by almost all authors in view of its simpler behaviour. The first general investigation of $m(n) - n$ was carried out by Hodges and Lehmann ([14]). They name $m(n) - n$ the deficiency of T_n with respect to T_n^* and denote it as

$$d_n = m(n) - n. \tag{4.1}$$

Suppose that for $n \to \infty$, the ratio $n / m(n)$ tends to a limit b, the asymptotic relative efficiency of $T_n(X_1, \dots, X_n)$ with respect to $T_n^*(X_1, \dots, X_n)$. If $0 < b < 1$, we have $d_n \sim (b^{-1} - 1)n$ and further asymptotic information about d_n is not particularly revealing. On the other hand, if $b = 1$, the asymptotic behaviour of d_n, which may now be anything from $o(1)$ to $o(n)$, does provide important additional information.

If $\lim_{n\to\infty} d_n$ exists, it is called the asymptotic deficiency of T_n with respect to T_n^* and denote as d. At points where no confusion is likely, we shall simply call d the deficiency of T_n with respect to T_n^*.

The deficiency of T_n relative to T_n^* will then indicate how many observations one loses by insisting on T_n, and thereby provides a basis for deciding whether or not the price is too high. If the risk functions of these two estimators are

$$R_n(\theta) = \mathsf{E}_\theta\ (T_n - g(\theta))^2, \quad R_n^*(\theta) = \mathsf{E}_\theta\ (T_n^* - g(\theta))^2,$$

then by definition, $d_n(\theta) \equiv d_n = m(n) - n$, for each n, may be found from

$$R_n^*(\theta) = R_{m(n)}(\theta). \tag{4.2}$$

In order to solve (4.2), $m(n)$ has to be treated as a continuous variable. This can be done in a satisfactory manner by defining $R_{m(n)}(\theta)$ for non-integral $m(n)$ as

$$R_{m(n)}(\theta) = (1 - m(n) + [m(n)])R_{[m(n)]}(\theta) + (m(n) - [m(n)])\, R_{[m(n)]+1}(\theta)$$

(cf. [14]).

Generally $R_n^*(\theta)$ and $R_n(\theta)$ are not known exactly and we have to use approximations. Here these are obtained by observing that $R_n^*(\theta)$ and $R_n(\theta)$ will typically satisfy asymptotic expansions of the form

$$R_n^*(\theta) = \frac{a(\theta)}{n^r} + \frac{b(\theta)}{n^{r+s}} + o(n^{-(r+s)}), \tag{4.3}$$

$$R_n(\theta) = \frac{a(\theta)}{n^r} + \frac{c(\theta)}{n^{r+s}} + o(n^{-(r+s)}), \tag{4.4}$$

for certain $a(\theta)$, $b(\theta)$ and $c(\theta)$ not depending on n and certain constants $r > 0$, $s > 0$. The leading term in both expansions is the same in view of the fact that ARE is equal to one. From (4.1)-(4.4) is now easily follows that (see [14])

$$d(\theta) \equiv d = \begin{cases} \pm\infty, & 0 < s < 1, \\ \dfrac{c(\theta) - b(\theta)}{ra(\theta)}, & s = 1, \\ 0, & s > 1. \end{cases} \tag{4.5}$$

The situation where $s = 1$ seems to be the most interesting one. Hodges and Lehmann ([14]) demonstrate the use of deficiency in a number of simple examples for which this is the case (see also [12]). The present paper consists of a number of applications of the deficiency concept in problems of point estimation in the case when number of observations is random.

5. Estimators Based on the Sample with Random Size

Everywhere in what follows it will be assumed that $\mathsf{E}N_n = n,$ that is, the expected sample size equals the sample size for the case where it is non-random.

Theorem 5.1. *Suppose that numbers* $a(\theta)$, $b(\theta)$ *and* $C(\theta) > 0$, $\alpha > 0$, $r > 0$, $s > 0$ *exist such that*

$$\left| R_n^*(\theta) - \frac{a(\theta)}{n^r} - \frac{b(\theta)}{n^{r+s}} \right| \le \frac{C(\theta)}{n^{r+s+\alpha}} ,$$

then

$$\left| R_{N_n}(\theta) - a(\theta)\, \mathsf{E}\, N_n^{-r} - b(\theta)\, \mathsf{E}\, N_n^{-r-s} \right| \le C(\theta)\, \mathsf{E}\, N_n^{-r-s-\alpha},$$

$$R_{N_n}(\theta) = \mathsf{E}_\theta (T_{N_n} - g(\theta))^2.$$

The proof directly follows from the total probability formula.

Let observations $X_1, \dots, X_n$ have expectation $g(\theta)$ and variance $\sigma^2(\theta)$. The customary estimator for $g(\theta)$ based on n observation is

$$T_n = \frac{1}{n} \sum_{i=1}^{n} X_i. \tag{5.1}$$

This estimator is unbiased and consistent, and its variance is

$$R_n^*(\theta) = \mathsf{D}_\theta T_n = \frac{\sigma^2(\theta)}{n}.$$

If this estimator T_{N_n} based on the sample with random size we have

$$R_{N_n}(\theta) = \mathsf{E}_\theta (T_{N_n} - g(\theta))^2 = \sigma^2(\theta)\, \mathsf{E}\, N_n^{-1}.$$

If $g(\theta)$ is given, we consider the estimator for $\sigma^2(\theta)$ in the form

$$\overline{T}_n = \frac{1}{n} \sum_{i=1}^{n} (X_i - g(\theta))^2. \tag{5.2}$$

This estimator is unbiased and consistent, and its variance is

$$\overline{R}_n^*(\theta) = \mathsf{D}_\theta \overline{T}_n = \frac{\mu_4(\theta) - \sigma^4(\theta)}{n}. \tag{5.3}$$

For this estimator $\overline{T}_{N_n}$ with random size one have

$$\overline{R}_{N_n}(\theta) = \mathsf{E}_\theta(\overline{T}_{N_n} - g(\theta))^2 = (\mu_4(\theta) - \sigma^4(\theta))\, \mathsf{E}\, N_n^{-1}. \tag{5.4}$$

In the proceeding example, suppose that $g(\theta)$ is unknown but that instead of (5.2) we are willing to consider any estimator of the form

$$\widetilde{T}_n^{(\gamma)} \equiv \widetilde{T}_n = \frac{1}{n+\gamma} \sum_{i=1}^{n} (X_i - T_n)^2, \quad \gamma \in \mathbb{R}. \tag{5.5}$$

If $\gamma \neq -1$, this will not be unbiased but may have a smaller expected squared error that the unbiased estimator with $\gamma = -1$.

One can easily find see formula (3.6) in [13] and formula (27.4.2) in [12]

$$\widetilde{R}_n^*(\theta) = \mathsf{E}_\theta(\widetilde{T}_n - g(\theta))^2$$

$$= \frac{\sigma^4(\theta)}{n(n+\gamma)^2}\Big((n-1)\left((\mu_4(\theta)/\sigma^4(\theta) - 1)(n-1) + 2\right) + n(\gamma+1)^2\Big), \tag{5.6}$$

and hence

$$\widetilde{R}_n^*(\theta) = \sigma^4(\theta)\left(\frac{\mu_4(\theta)/\sigma^4(\theta) - 1}{n}\right.$$

$$\left. + \frac{(\gamma+1)^2 - 2(\mu_4(\theta)/\sigma^4(\theta) - 1) + 2 - 2\gamma(\mu_4(\theta)/\sigma^4(\theta) - 1)}{n^2}\right)$$

$$+ O(n^{-3}). \tag{5.7}$$

Using Theorem 5.1, we have

$$\tilde{R}_{N_n}(\theta) = \mathsf{E}_\theta(\tilde{T}_{N_n}(X_1, \dots, X_{N_n}) - g(\theta))^2$$

$$= \sigma^4(\theta)\Big((\mu_4(\theta)/\sigma^4(\theta) - 1)\ \mathsf{E}\ N_n^{-1} + ((\gamma+1)^2 - 2(\mu_4(\theta)/\sigma^4(\theta) - 1)$$

$$+ 2 - 2\gamma(\mu_4(\theta)/\sigma^4(\theta) - 1))\ \mathsf{E}\ N_n^{-2}\Big) + O(\mathsf{E}\ N_n^{-3}). \tag{5.8}$$

6. Deficiencies of Some Estimators Based on the Samples with Random Size

When the deficiencies of statistical estimators constructed from samples of random size $N_{m(n)}$ and the corresponding estimators constructed from samples of non-random size n (under the condition $\mathsf{E}\ N_n = n$) are evaluated, we actually compare the expected size $m(n)$ of a random sample with n by virtue of the quantity $d_n = m(n) - n$ and its limit value (see also [11]).

We now apply the results of Section 5 to the four examples given in this section. Let M_n be the Poisson r.v. with parameter $n-1,\ n \geq 2$. Define the random size as $N_n = M_n + 1$, then $\mathsf{E}\ N_n = n$ and

$$\mathsf{E}\ N_n^{-1} = \frac{1}{n} + \frac{1}{n^2} + o(n^{-2}). \tag{6.1}$$

The deficiency of T_{N_n} relative to T_n (see (5.1)) is given by (5.1), (6.1) and (4.5) with

$$r = s = 1, \quad a(\theta) = \sigma^2(\theta), \quad b(\theta) = 0, \quad c(\theta) = \sigma^2(\theta),$$

and hence is equal to

$$d = 1.$$

Similarly, the deficiency of $\overline{T}_{N_n}$ relative to $\overline{T}_n$ (see (5.2)) is given by (5.3), (5.4), (6.1) and (4.5) with

$$r = s = 1, \quad a(\theta) = c(\theta) = \mu_4(\theta) - \sigma^4(\theta), \quad b(\theta) = 0,$$

and hence is equal to

$$\overline{d} = 1.$$

Consider now third example (see (5.5)). We have

$$\mathsf{E}\, N_n^{-2} \sim \frac{1}{n^2}, \quad n \to \infty.$$

Now the deficiency of $\widetilde{T}_{N_n}$ relative to $\widetilde{T}_n$ (see (5.5)) is given by (5.7), (5.8) and (6.1) with $r = s = 1$ and hence is equal to $\widetilde{d} = 1$ and the deficiency of $\widetilde{T}_{N_n}^{(\gamma_1)}$ relative to $\widetilde{T}_{N_n}^{(\gamma_2)}$ (see (5.5), (5.8)) is given by (6.1) and (4.5) with $r = s = 1$ and hence is equal to

$$\widetilde{d}_{\gamma_1, \gamma_2} = (\gamma_1 - \gamma_2)\left(\frac{\gamma_1 + \gamma_2 + 2}{\mu_4(\theta) / \sigma^4(\theta) - 1} - 2\right).$$

References

[1] V. E. Bening and V. Yu. Korolev, Generalized Poisson Models and their Applications in Insurance and Finance, VSP, Utrecht, 2002.

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

[3] V. E. Bening and V. Yu. Korolev, On an application of the Student distribution in the theory of probability and mathematical statistics, Theory of Probability and its Applications 49(3) (2005), 377-391.

DOI: https://doi.org/10.1137/S0040585X97981159

[4] V. E. Bening and V. Yu. Korolev, Some statistical problems related to the Laplace distribution, Informatika i Ee Primeneniya 2(2) (2008), 19-34.

[5] O. O. Lyamin, On the rate of convergence of the distributions of certain statistics to the Laplace and Student distributions, Moscow University Computational Mathematics and Cybernetics 35(1) (2011), 37-46.

DOI: https://doi.org/10.3103/S0278641911010055

[6] V. E. Bening, N. K. Galieva and V. Yu. Korolev, Asymptotic expansions for the distribution functions of statistics constructed from samples with random sizes, Informatika i Ee Primeneniya 7(2) (2013), 75-83.

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

DOI: https://doi.org/10.1007/978-3-642-65809-9

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

[9] S. S. Wilks, Recurrence of extreme observations, Journal of the Australian Mathematical Society 1(1) (1959), 106-112.

DOI: https://doi.org/10.1017/S144678870002512X

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

[11] V. E. Bening, On deficiencies of some estimators based on samples of random size, Vestnik TVGU, Seriya: Prikladnaya Matematica [Herald of Tver State University, Series: Applied Mathematics] 1 (2015), 5-14 (in Russian).

[12] V. E. Bening, Asymptotic Theory of Testing Statistical Hypotheses: Efficient Statistics, Optimality, Power Loss, and Deficiency, Berlin: Walter de Gruyter, 2011, 277 pp. ISBN 978 - 3 - 11 - 093599 -8.

[13] H. Cramér, Mathematical Methods of Statistics, Princeton University Press, Princeton, 1946, 631 pp.

[14] J. L. Hodges and E. L. Lehmann, Deficiency, Annals of Mathematical Statistics 41(3) (1970), 783-801.

DOI: https://doi.org/10.1214/aoms/1177696959

■