ESAIM: PS 24 (2020) 581–606
https://doi.org/10.1051/ps/2020020

ESAIM: Probability and Statistics
www.esaim-ps.org

THE PROBABILITIES OF LARGE DEVIATIONS FOR A CERTAIN CLASS OF STATISTICS ASSOCIATED WITH MULTINOMIAL DISTRIBUTION

Sherzod M. Mirakhmedov*

Abstract. Let $\eta = (\eta_1, \ldots, \eta_N)$ be a multinomial random vector with parameters $n = \eta_1 + \cdots + \eta_N$ and $p_m > 0$, $m = 1, \ldots, N$, $p_1 + \cdots + p_N = 1$. We assume that $N \to \infty$ and $\max p_m \to 0$ as $n \to \infty$. The probabilities of large deviations for statistics of the form $h_1(\eta_1) + \cdots + h_N(\eta_N)$ are studied, where $h_m(x)$ is a real-valued function of a non-negative integer-valued argument. The new large deviation results for the power-divergence statistics and its most popular special variants, as well as for several count statistics are derived as consequences of the general theorems.

Mathematics Subject Classification. 60F10, 62E20, 62G20.

Received October 8, 2019. Accepted August 3, 2020.

1. Introduction

Let $\eta = (\eta_1, \ldots, \eta_N)$ be a random vector of frequencies of a multinomial model $M(n, N, \mathrm{P})$ on $n \geq 1$ observations classified into $N > 1$ cells (categories) with the cell probabilities $\mathrm{P} = (p_1, \ldots, p_N)$, $p_1 + \cdots + p_N = 1$, all $p_j > 0$. Hence, $n = \eta_1 + \cdots + \eta_N$ and

$$P\{\eta_1 = m_1, \ldots, \eta_N = m_N\} = \frac{n!}{m_1! \cdot \ldots \cdot m_N!} p_1^{m_1} \cdot \ldots \cdot p_N^{m_N},$$

where arbitrary non-negative integer m_js are such that $m_1 + \cdots + m_N = n$. This multinomial distribution is a probabilistic model for a random categorical data: If each of n independent trials can result in any of N possible types of outcome, and the probability that the outcome is of a given type is the same in every trial, the numbers of outcomes of each of the N types have a multinomial joint distribution. It is convenient also to describe this model in terms of a multinomial random allocation of n particles into N cells: n particles are allocated into N cells indexed 1 through N at random, successively and independently of each other , the probability of a particle falls into cell with index l is $p_l > 0$, $l = 1, \ldots, N$, $p_1 + \cdots + p_N = 1$, then η_l is the number of particles in the cell with index l after allocation of all n particles.

The classical tests goodness of fit on the probabilities of cells based on the chi-square type statistics (*i.e.* on the statistics having an asymptotic chi-square distribution), and assume that the number of cells N is fixed.

Keywords and phrases: Chi-square statistic, count statistics, log-likelihood ration statistic, large deviations, multinomial distribution, Poisson distribution, power divergence statistics.

Institute of Mathematics, Academy of Sciences of Uzbekistan, Mirzo Ulugbek str., 81, 100125 Tashkent, Uzbekistan.

* Corresponding author: shmirakhmedov@yahoo.com

The most well-known such statistics are the Pearson's chi-square statistics and the log-likelihood ratio statistics, which are special variants of the power-divergence statistics, introduced by Cressie and Read [3] There is huge literature where interest and results have followed many aspects: the asymptotic distributional and statistical properties and recommendations in applications of power-divergence statistics and its special variants in the case fixed N, see [4, 25, 27] and references within. However, the assumption "N is fixed" becomes restrictive in several contexts. Indeed: Mann and Wald [18] have obtained the relation $N \sim cn^{2/5}$, where $c > 0$ depends on the test size, concerning the optimal choice of the number of groups in the chi-square goodness of fit test. Koehler and Larntz [15] have explored the practical importance of the asymptotic normality results of chi-squared and log-likelihood ratio statistics in case N increases. Consider a problem of testing uniformity distribution on a fixed interval in which the interval partitioned into subintervals of equal length; if it is desired to achieve a specified expected frequency ν(say) for each subinterval, then N subintervals are used for a sample size of n, where N is selected to make n/N close to ν. For another motivation for increasing N associated with "big-data" applications, see [30, 32]. At last, a wide class of random variables of interest in the above described multinomial random allocation of particles into cells assumes that the number of cells N increases together with the number of particles n, see [16, 17].

In this paper we shall focus our attention to the case when $N = N(n) \to \infty$ as $n \to \infty$. It turned out that in the case $N \to \infty$ one can study a very general class of statistics of the form

$$R_N(\eta) = \sum_{l=1}^{N} h_l(\eta_l), \tag{1.1}$$

where $h_1(x), \ldots, h_N(x)$ (may depend on N and n) are real-valued functions defined on the non-negative axis. These functions can be random, then it is assumed that for any $x_1, \ldots, x_N$, random variables (r.v.) $h_1(x_1), \ldots, h_N(x_N)$ are mutually independent and independent of $(\eta_1, \ldots, \eta_N)$. We will pay special attention to the family of power-divergence statistics (PDS), which is a subclass of (1.1) with $h_l(x) = 2x\left((x/np_l)^d - 1\right)/d(d+1)$, although a more general subclass is the family of ϕ-divergences, where $h_l(x) = np_l\phi(x/np_l)$, see [5, 19]. The most notably special versions of PDS are

$$\chi^2_N = \sum_{m=1}^{N} \frac{(\eta_m - np_m)^2}{np_m}, \Lambda_N = 2\sum_{m=1}^{N} \eta_m \ln \frac{\eta_m}{np_m}, T^2_N = 4\sum_{m=1}^{N} (\sqrt{\eta_m} - \sqrt{np_m})^2. \tag{1.2}$$

The statistics (1.2) are known as Pearson's chi-squared (χ^2), the log-likelihood ratio (LR) and the Freeman-Tukey (FT) statistics, respectively. Special variants of statistics (1.1), that are not PDS, are the following count statistics (CS):

$$\mu_r = \sum_{m=1}^{N} \mathrm{I}\{\eta_m = r\}, w_r = \sum_{m=1}^{N} \mathrm{I}\{\eta_m \geq r\}, C_n = \sum_{m=1}^{N} (\eta_m - 1)\mathrm{I}\{\eta_m > 1\}, \tag{1.3}$$

where $r = 0, 1, \ldots, n$ and $\mathrm{I}\{A\}$ is the indicator function of A. In terms of the aforementioned random allocation scheme of n particles into N cells the statistics μ_r, w_r, and C_n count respectively: the number of cells that contain exactly $r \geq 0$ particles, the number of cells that contain at least $r \geq 1$ particles, and the number of collisions (*i.e.* the number of times a particle falls in a cell that already has a particle in it). Consider now a random allocation with random level. Let the cell with index m be assigned a non-negative, integer valued r.v. ν_m, $m = 1, \ldots, N$. The cell with index m is said to be filled up if $\eta_m \geq \nu_m$ after allocation all particles. A special version of (1.1) that of interest is the number of unfilled cells Φ_N, say, viz.

$$\Phi_N = \sum_{m=1}^{N} B_m(\eta_m), \tag{1.4}$$

where $B_m(x)$ is a Bernoulli r.v.: $P\{B_m(x) = 1\} = P\{\nu_m > x\}$, $P\{B_m(x) = 0\} = P\{\nu_m \le x\}$, $m = 1, \ldots, N$.

If $h_m(x) = h(x)$ for all m then (1.1) is said to be symmetric statistics. Obviously χ^2_N, Λ_N and T^2_N are symmetric if $p_1 = ... = p_N = N^{-1}$. All statistics (1.3) are symmetric, whereas statistic Φ_N is symmetric if $\nu_1, \ldots, \nu_N$ are i.i.d. r.v.s. The PDS, specifically statistics (1.2), play a leading role in tests goodness-of-fit on grouped data, whereas the CS (1.3), (1.4) are used in problems associated with occupancy problems, see [14, 16]. We also draw the attention of readers to the article by L'ecuyer *et al.* [17], where the statistics (1.1) and its special versions, including (1.2) and (1.3) are used in construction the serial tests for testing of uniformity and independence of the output sequence of general-purpose uniform random number generators.

The class of statistics (1.1) has been studied in the literature. The Central Limit Theorem among with Berry-Essen kind bound and Edgeworth asymptotic expansion results with different refinements can be found in the papers [6, 8, 13, 19–24, 26, 29].

In the present paper we are interested in $P\left\{\tilde{R}_N(\eta) \ge x_N\right\}$, where $\tilde{R}_N(\eta)$ is standardized version of $R_N(\eta)$ and $x_N \to \infty$, in the situation when $N \to \infty$ and $\max p_m \to 0$, as $n \to \infty$. According to our best knowledge such probabilities of large deviations have been studied earlier by Siragdinov *et al.* [36] and Ivchenko and Mirakhmedov [11] (see Thm. 2.1 of Sect. 2). Their results correspond to the Cramér zone, $x_N = o(\sqrt{N})$, and assume two restrictions: (i) the sparseness condition: $\lim_{n\to\infty}(n/N) = \lambda \in (0,\infty)$, and (ii) the Cramér condition : $\max_{1\le k\le N} E\exp\{H\,|h_k(\xi)|\} < \infty$, $\exists H > 0$, here ξ is a Poisson r.v. with expectation λ. We refer also to [12], who have studied large deviation probabilities of χ^2_N and Λ_N; in particularly, his results corresponding to the case $x_N = o(\sqrt{N})$ propose that $N\min p_m \ge c > 0$, for Λ_N statistic, and in addition, $\sqrt{n}N^{-3/2}x_N \to \infty$ and $N = o(\sqrt{n})$ (*i.e.* the model is so dense that the average of the observations $n/N \gg \sqrt{n}$) for χ^2_N statistic. Remark that Cramér's condition (ii) is fulfilled for Λ_N, but for χ^2_N does not; also the sparseness condition (i) is too restrictive in applications, for instance in a problem of testing goodness-of-fit. Kallenberg's results also do not cover some important situations; see Remarks 4.6 and 4.8 of Section 4. The prime driver of this work was the need for results on the probabilities of large deviations of the χ^2 and LR statistics in order to study the intermediate properties of the tests goodness of fit, see Appendix B. The results of aforementioned authors don't cover the needs. In contrast, while the new results of the present paper mostly corresponds to the deviations of zone $x_N = o(N^{1/6})$, their advantage is that they applicable for the *arbitrary multinomial models* (not necessary sparse) and for statistics that *do not necessarily satisfy the Cramér's condition.* Further, the proof of Theorem 2.2 is based on the technique different from the methods of existing papers. Core part of our technique is that we reduce the cumulants of the statistics $R_N(\eta)$ to the cumulants of the sum of independent r.v.s, next we use a modified variant of Lemma 2.3 of Saulis and Statulevicius (1991) [35] .

The rest of the paper is organized as follows. The main results are presented in Section 2; these theorems are applied for the class of PDS in Section 3, and in Section 4 for the statistics (1.2), (1.3) and (1.4); the proofs are presented in Section 5. For the reader's convenience, the auxiliary Assertions are collected in Appendix A; at last, some results on the asymptotical properties of goodness-of-fit tests on uniformity of a multinomial model, obtained using the results of Section 3, are presented in Appendix B.

Throughout the paper we use c_i to denote absolute constants whose value may differ at each occurrence, all asymptotic statements are considered as $n \to \infty$; $\varsigma \sim F$ stands for "r.v. ς has the distribution$F$", $\mathcal{C}_k(\varsigma)$- the cumulant of order k of the r.v. ς, $Poi(\lambda)$- the Poisson distribution with parameter $\lambda > 0$ and $\Phi(u)$- the standard normal distribution function. Everywhere in the next sections $\xi, \xi_1, \ldots, \xi_N$ are independent r.v.s, where $\xi_m \sim Poi(np_m)$;$\xi \sim Poi(\lambda_n)$,$\lambda_n = n/N$, $\pi_l(\lambda) = \lambda^l e^{-\lambda}/l!$, $p_{\max} = \max_{1\le k\le N} p_k$, $p_{\min} = \min_{1\le k\le N} p_k$.

2. The general results

Set

$$\begin{aligned} g_m(\xi_m) &= h_m(\xi_m) - Eh_m(\xi_m) - \gamma_n(\xi_m - np_m), \\ A_N &= \sum_{m=1}^{N} Eh_m(\xi_m), \gamma_n = n^{-1}\sum_{m=1}^{N} \mathrm{cov}\,(h_m(\xi_m),\xi_m), \\ \widetilde{\sigma}_N^2 &= \sum_{m=1}^{N} Varh_m(\xi_m),\ \sigma_N^2 = \sum_{m=1}^{N} Var\ g_m(\xi_m) = \widetilde{\sigma}_N^2 - n\gamma_n^2. \end{aligned} \tag{2.1}$$

Note that under very general set-up A_N and σ_N^2 becomes the asymptotic value of $ER_N(\eta)$ and $VarR_N(\eta)$, respectively. In particularly, under the conditions of below Theorems 2.1 and 2.2

$$ER_N(\eta) = A_N + o(N)\ and\ VarR_N(\eta) = \sigma_N^2(1 + o(1)). \tag{2.2}$$

The following theorem was proved by Ivchenko and Mirakhmedov [11], see also [36]. We state it for completeness and for ease sake of comparison.

Theorem 2.1. *Assume that* $\boldsymbol{n} \to \boldsymbol{\infty}$, $N \to \infty$ *such that*

$$\lambda_n \to \lambda \in (0,\infty), \tag{2.3}$$

$$Np_{\max} \le c_1, \tag{2.4}$$

$$\lim\inf N^{-1}\sigma_N^2 > 0, \tag{2.5}$$

$$\max_{1\le m\le N} Ee^{H|h_m(\xi_m)|} \le c_2,\ \exists\, H > 0. \tag{2.6}$$

Then uniformly in $x_N \ge 0$, $x_N = o(\sqrt{N})$

$$P\left\{R_N > x_N\sigma_N + A_N\right\} = (1 - \Phi(x_N)) \exp\left\{\frac{x_N^3}{\sqrt{N}} M_N\left(\frac{x_N}{\sqrt{N}}\right)\right\}\left(1 + o\left(\frac{1 + x_N}{\sqrt{N}}\right)\right), \tag{2.7}$$

$$P\left\{R_N < -x_N\sigma_N + A_N\right\} = \Phi(-x_N)\exp\left\{-\frac{x_N^3}{\sqrt{N}} M_N\left(-\frac{x_N}{\sqrt{N}}\right)\right\}\left(1 + o\left(\frac{1 + x_N}{\sqrt{N}}\right)\right), \tag{2.8}$$

where $M_N(u) = \mu_{0N} + \mu_{1N}u + \ldots$ is a series which for sufficiently large N can be majorized by a power series $\mu_0 + \mu_1 u + \ldots$ (*i.e.* $|\mu_{jN}| \le \mu_j, j = 0,1,\ldots$) converging in a neighbourhood of zero. Exact formula for μ_{0N} and μ_{1N} is given by Siragdinov *et al.* [36].

Our main new result is as follows. Set $K_n(a,b) = \left(n^{1-b}p_{\max}^{-b}\right)^{1/(\overline{a}+1)}$, where integers $a \ge 0$, $b \ge 0$ and $\overline{a} = \max(1,a)$.

Theorem 2.2. *Let $n \to \infty$, $N \to \infty$, $p_{\max} \to 0$ and the functions $h_m(\,\cdot\,)$ be non-negative. If (i) For each integer $s \in [3\ ,\ k_n]$ there exist non-negative a_1, a_2 and b_1, b_2 such that*

$$E\left((\xi_m - np_m)^2 h_m^s(\xi_m)\right) = O\left(s^a (np_m)^b E h_m^s(\xi_m)\right), \tag{2.9}$$

where $a = a_1\ ,\ b = b_1$ for all $m \in \mathcal{N} \subseteq (1,2,\ldots,N)$, and $a = a_2\ ,\ b = b_2$ for all $m \in (1,2,\ldots,N)\backslash\mathcal{N}$, and

$$k_n = o\left(\min(p_{\max}^{-1}, K_n(a_1,b_1), K_n(a_2,b_2))\right), \tag{2.10}$$

$$|\mathcal{C}_s(h_m(\xi_m))| \le (s!)^{1+\nu} V_n^{s-2} \omega_m^2\ ,\ m = 1,\ldots,N, \tag{2.11}$$

then for all x_N such that

$$0 \le x_N = o\left(\min\left(W_N^{1/(1+2v)}, k_n^{1/2}\right)\right), \tag{2.12}$$

where $W_N = V_n^{-1}\sigma_N \min\left(1, \sigma_{\mathrm{N}}^2/\widetilde{\omega}_N^2\right)$, $\widetilde{\omega}_N^2 = \omega_1^2 + \cdots + \omega_N^2$, it holds

$$P\left\{R_N > x_N\sigma_N + A_N\right\} = (1 - \Phi(x_N))\left(1 + O\left((x_N+1) W_N^{-1/(1+2v)}\right)\right), \tag{2.13}$$

and

$$P\left\{R_N < -x_N\sigma_N + A_N\right\} = \Phi(-x_N)\left(1 + O\left((x_N+1) W_N^{-1/(1+2v)}\right)\right). \tag{2.14}$$

To compare these two theorems, we make the following comments. The deviation zone $x_N = o(\sqrt{N})$ of Theorem 2.1 is maximally possible for this Cramér type result; application, however, of Theorem 2.1 is limited by the sparse multinomial models only, since conditions (2.3) and (2.4), and by the class of statistics satisfying the Cramér condition (2.6). For instance, the class of PDS with parameter $d > 0$ does not satisfy the condition (2.6). In contrast, Theorem 2.2 is effective mostly in the deviation zone $x_N = o(N^{1/6})$, see Section 3, but its advantage is that it applicable for the arbitrary multinomial models and for statistics that do not necessarily satisfy the Cramér condition. Note that for the sparse models Theorem 2.2 extends Theorem 2.1 for the deviation zone $x_N = o(N^{1/6})$, since it includes also statistics that do not satisfy the Cramér condition.

Remark 2.3. Condition (i) is not necessary, its appearance due to the technique of proof; see proof of Proposition 5.1, the relation (5.9). When applying Theorem 2.2 to statistics (1.2), (1.3) and (1.4), this condition (i) effects through the quantity $K_n(a,b)$, restricting us to the deviation of order $o(N^{1/6})$. Actually, due to this facts we restricted condition (ii) to the assumption $\nu \ge 1$, instead of $\nu \ge 0$, as it should have been to get the Linnik's kind large deviation result, alike to Lemma 2.3 of Saulis and Statulevichius [35].

3. Application to the Power Divergence Statistics (PDS)

The PDS of [3] is defined as

$$CR_N(d) = \frac{2}{\mathrm{d}(d+1)} \sum_{l=1}^{N} \eta_l \left[(\eta_l/np_l)^d - 1\right] = \frac{2}{\mathrm{d}(d+1)} \sum_{l=1}^{N} np_l (\eta_l/np_l)^{d+1} - \frac{2n}{\mathrm{d}(d+1)}.$$

Note that for $d = 0$ the PDS is defined by continuity: $CR_N(0) = \lim_{d\to 0} CR_N(d) = \Lambda_N$. We assume that $d > -1$ in order Theorems 2.1 and 2.2 to be applicable. Further, since we deal with standardized variant of

PDS, we can, whenever it is necessary, consider PDS as a special version of statistics $R_N(\eta)$ with kernel functions $h_l(x) = h_{d,l}(x)$, where

$$h_{d,l}(x) = np_l\,(x/np_l)^{d+1},\ d > -1,\ d \neq 0,\ \text{else}\ h_l\,(x) = h_{o,l}\,(x) = 2x\ln\,(x/np_l) \tag{3.1}$$

The notation (3.1) has the following form: for $d > -1$, $d \neq 0$:

$$A_N(d) = \frac{2n}{\mathrm{d}(d+1)}\left(\sum_{m=1}^{N} p_m E(\xi_m/np_m)^{d+1} - 1\right), \tag{3.2}$$

$$\gamma_n(d) = \frac{2}{\mathrm{d}(d+1)}\sum_{m=1}^{N} p_m E\left((\xi_m/np_m)^{d+1}(\xi_m - np_m)\right), \tag{3.3}$$

$$\widetilde{\sigma}_N^2\,(d) = \frac{4n^2}{\mathrm{d}^2\,(d+1)^2}\sum_{m=1}^{N} p_m^2 \mathrm{var}\,(\xi_\mathrm{m}/np_m)^{d+1} \tag{3.4}$$

and

$$A_N(0) = 2n\sum_{m=1}^{N} p_m E\,[(\xi_m/np_m)\ln(\xi_m/np_m)], \tag{3.5}$$

$$\gamma_n(0) = 2n^{-1}\sum_{m=1}^{N} E\,((\xi_m - np_m)\xi_m\ln(\xi_m/np_m)), \tag{3.6}$$

$$\widetilde{\sigma}_N^2(0) = 4n^2\sum_{m=1}^{N} p_m^2\mathrm{var}\,[(\xi_m/np_m)\ln(\xi_m/np_m)]. \tag{3.7}$$

Again, we remind that $A_N(d)$and

$$\sigma_N^2(d) = \widetilde{\sigma}_N^2(d) - n\gamma_n^2(d) \tag{3.8}$$

are asymptotic expectation and asymptotic variance of $CR_N(d)$, respectively.

3.1. The sparse multinomial model: $\boldsymbol{\lambda_n \to \lambda \in (0,\infty)}$

Theorems 2.1 and 2.2 imply

Theorem 3.1. *Suppose $n \to \infty$, $N \to \infty$ such that*

$$\lambda_n \to \lambda \in (0,\infty)\ \textit{and}\, c_3 \leq Np_m \leq c_4,\ \textit{some}\, c_3 > 0, c_4 > 0. \tag{3.9}$$

(i) Let $-1 < d \leq 0$. Then uniformly in $x_N \geq 0$, $x_N = o(\sqrt{N})$ for the $P\{CR_N(d) > x_N\sigma_N(d) + A_N(d)\}$ and $P\{CR_N(d) < -x_N\sigma_N(d) + A_N(d)\}$ the relations (2.7) and (2.8), respectively, are hold.

(ii) Let an integer $d \geq 1$. *Then uniformly in* $x_N \geq 0, x_N = o(N^{1/2(1+2d)})$ *one has*

$$P\{CR_N(d) > x_N\sigma_N(d) + A_N(d)\} = (1 - \Phi(x_N))(1 + o(1)), \tag{3.10}$$

$$P\{CR_N(d) < -x_N\sigma_N(d) + A_N(d)\} = \Phi(-x_N)(1 + o(1)). \tag{3.11}$$

(iii) Let a non-integer $d > 0$. *Then uniformly in* $x_N \geq 0$,

$x_N = o(\min(N^{1/8}, N^{1/2(1+2\hat{d})}), \ \hat{d} = \max(1,d),$ the relations (3.10) and (3.11) are hold.

3.2. The dense multinomial model: $\lambda_n \to \infty$

In this case, the PDS $CR_N(d)$ for each $d > -1$ behave asymptotically alike to chi-square statistics.

Theorem 3.2. *Let* $p_{\max} = o(1)$ *and* $np_{\min} \to \infty$. *Then for each* $d > -1$ *uniformly in* $x_N \geq 0$, $x_N = o\left(\min(N^{1/6}, p_{\max}^{-1/4})\right)$ *one has*

$$P\left\{CR_N(d) > x_N\sqrt{2N} + N\right\} = (1 - \Phi(x_N))(1 + o(1)), \tag{3.12}$$

$$P\left\{CR_N(d) < -x_N\sqrt{2N} + N\right\} = \Phi(-x_N)(1 + o(1)). \tag{3.13}$$

Remark 3.3. From $np_{\min} \to \infty$ it follows $\lambda_n \to \infty$. The condition $np_{\min} \to \infty$ can be replaced by $\lambda_n \to \infty$ and $Np_{\min} \geq c > 0$, which imply in turn $np_{\min} \to \infty$. It is obvious that if $p_{\max} = O(N^{-2/3})$, then relations (3.12) and (3.13) are hold uniformly in non-negative $x_N = o(N^{1/6})$.

Remark 3.4. Let $p_m = (1-\alpha)/N^{1-\alpha}m^{\alpha}$, $m = 1, \ldots, N$ and $\alpha \in [0, 1)$. This example of the multinomial model has been considered by Rempala and Wesolowski [32], as it is of interest when testing for signal-noise threshold in data with large number of support points [28]. We have $np_{\min} = (1-\alpha)\lambda_n$, $p_{\max} = (1-\alpha)/N^{1-\alpha}$, $Np_m = (1-\alpha)N^{\alpha}m^{-\alpha}$, $1-\alpha \leq Np_m \leq (1-\alpha)N^{\alpha}$. Theorem 3.1 can be applied for the case $\alpha = 0$, *i.e.* for the uniform multinomial model only. Theorem 3.2 for this model states that if $\lambda_n \to \infty$ then the relations (3.12) and (3.13) hold uniformly in $x_N \geq 0$ such that: $x_N = o(N^{1/6})$ if $\alpha \in [0, 1/3]$, and $x_N = o(N^{(1-\alpha)/4})$ if $\alpha \in (1/3, 1)$.

3.3. The very sparse multinomial model :$\lambda_n \to 0$

The $CR_N(d)$, being divergence measure between proposed and observed models, depends on the intended model. As it turns out, this fact plays a role for very sparse models; for instance the asymptotic value of the expectation and variance of statistic Λ_N differ essentially for the non-uniform and for the uniform multinomial model. The standardized version of $CR_N(d)$ coincides with standardized version of $R_N^d = h_{d,1}(\eta_1) + \cdots + h_{d,N}(\eta_N)$, where $h_{d,m}(x)$ is defined in (3.1). Set $\mathrm{P}_{jN}(a) = p_1^{j-a} + \cdots + p_N^{j-a}$. The asymptotic value of expectation and variance of R_N^d, when $d \neq 0$, are equal, respectively to

$$A_N(d) = n^{1-d}\mathrm{P}_{1N}(d),$$

$$\sigma_N^2(d) = 2n^{2(1-d)}\left[(2^{2d} - 1)\mathrm{P}_{2N}(2d) - 2(2^d - 1)\mathrm{P}_{1N}(d)\mathrm{P}_{2N}(d)\right]. \tag{3.14}$$

Consider now $R_N^0 = \Lambda_N$. Let $(p_1,\ldots,p_N) \neq (N^{-1},\ldots,N^{-1})$ and a r.v. Z_N be such that $P\{Z_N = -\ln np_m\} = p_m, m = 1,\ldots,N$. For the asymptotic value of expectation and variance of R_N^0 we have

$$A_N(0) = 2nEZ_N,\ \sigma_N^2(0) = 4n\mathrm{var}Z_N. \tag{3.15}$$

Let, now, $(p_1,\ldots,p_N) = (N^{-1},\ldots,N^{-1})$. Then

$$A_N(d) = n\lambda_n^{-d},\ \sigma_N^2(d) = 2(2^d-1)^2 n\lambda_n^{1-2d}, d > -1, d \neq 0, \tag{3.16}$$

$$A_N(0) = 2\ln 2\ n\lambda_n,\ \sigma_N^2(0) = 8\ln^2 2\ n\lambda_n. \tag{3.17}$$

Set $\tilde{R}_N^d = \eta_1^{1+d} + \cdots + \eta_N^{1+d}$, $d > -1$, $d \neq 0$, and else $\tilde{R}_N^0 = 2\eta_1 \ln\eta_1 + \cdots + 2\eta_N \ln\eta_N$. For the uniform model standardized version of R_N^d coincides with $(\tilde{R}_N^d - \tilde{A}_N(d))\widetilde{\sigma}_N^{-1}(d)$, where $\tilde{A}_N(d) = n$, $\widetilde{\sigma}_N^2(d) = 2(2^d-1)^2 n\lambda_n$, for $d \neq 0$, else $\tilde{A}_N(0)$ and $\widetilde{\sigma}_N^2(0)$ are as in formula (3.17).

Theorem 3.5. *Let $\lambda_n \to 0$, $n\lambda_n \to \infty$, $Np_{\max} \le C$ and $d \neq 0$. Then*

$$P\left\{R_N^d > x_N\sigma_N(d) + A_N(d)\right\} = (1-\Phi(x_N))(1+o(1)),$$

$$P\left\{R_N^d < -x_N\sigma_N(d) + A_N(d)\right\} = \Phi(-x_N)(1+o(1))$$

are hold

(i) uniformly in $x_N \ge 0$, $x_N = o\left(\min(n^{1/6}, W_N^{1/(1+2\hat{d})}\right)$ for integer $d \ge 1$,

(ii) uniformly in $x_N \ge 0$, $x_N = o\left(\min(n^{1/8}, W_N^{1/(1+2\hat{d})}\right)$ for non-integer $d > -1$,

where $A_N(d)$ and $\sigma_N^2(d)$ are defined in (3.14),

$$W_N = \sigma_N(d)\min(1,\sigma_N^2(d)/n^{1-d}\mathrm{P}_{1N}(d)),\ \hat{d} = \max(1,d).$$

Corollary 3.6. *Let $\lambda_n \to 0$, $n\lambda_n^3 \to \infty$, $p_1 = ... = p_N = N^{-1}$ and $d \neq 0$. Then*

$$P\left\{\tilde{R}_N^d > x_N\left|2^d-1\right|\sqrt{2n\lambda_n} + n\right\} = (1-\Phi(x_N))(1+o(1)),$$

$$P\left\{\tilde{R}_N^d < -x_N\left|2^d-1\right|\sqrt{2n\lambda_n} + n\right\} = \Phi(-x_N)(1+o(1)),$$

are hold

(i) uniformly in $x_N \ge 0$, $x_N = o\left((n\lambda_n^3)^{1/2(1+2\hat{d})}\right)$ for integer $d \ge 1$,

(ii) uniformly in $x_N \ge 0$, $x_N = o\left(\min(n^{1/8}, (n\cdot\lambda_n^3)^{1/2(1+2\hat{d})}\right)$ for non-integer $d > -1$.

Theorem 3.7. *Let $\lambda_n \to 0$, $n\lambda_n \to \infty$, $Np_{\max} \le C$, and $(p_1,\ldots,p_N) \neq (N^{-1},\ldots,N^{-1})$. Then uniformly in $x_N \ge 0$, $x_N = o(n^{1/6}\min(1,(VarZ_N)^{1/6}))$ it hold (see (3.15))*

$$P\left\{\Lambda_N > 2x_N\sqrt{nVarZ_N} + 2nEZ_N\right\} = (1-\Phi(x_N))(1+o(1)),$$

$$P\left\{\Lambda_N < -2x_N\sqrt{nVarZ_N} + 2nEZ_N\right\} = \Phi(-x_N)(1+o(1)).$$

Theorem 3.8. *Let $\lambda_n \to 0$, $n\lambda_n \to \infty$ and $(p_1, \dots, p_N) = (N^{-1}, \dots, N^{-1})$. Then uniformly in $x_N \ge 0$,$x_N = o(\min(n^{1/8}, (n\lambda_n^3)^{1/6}))$ it holds*

$$P\left\{\Lambda_N > x_N 2\ln 2 \ \sqrt{2n\lambda_n} + 2\ln 2 \ \ n\lambda_n\right\} = (1 - \Phi(x_N))(1 + o(1)),$$

$$P\left\{\Lambda_N < -x_N 2\ln 2 \ \sqrt{2n\lambda_n} + 2\ln 2 \ n\lambda_n\right\} = \Phi(-x_N)(1 + o(1)).$$

We end Section 3 by noting that the theorems of this section should be regarded as results under the null hypothesis that the cell probabilities are $p_1, \dots, p_N$.

4. Application to the special variants of the statistic $R_N(\eta)$

Consider application of results of Sections 2 and 3 to the statistics (1.2), (1.3) and (1.4). Remind that (1.2) statistics are PDS: $\chi_N^2 = CR_N(1)$, $\Lambda_N = CR_N(0)$ and $T_N^2 = CD_N(-1/2)$, whereas the (1.3) and (1.4) statistics are out of the class of PDS.

4.1. Chi-square statistic

For the χ_N^2 statistic $h_m(u) = (u - np_m)^2 / np_m$,

$$\widetilde{\sigma}_N^2 = \sum_{m=1}^{N} \frac{1}{np_m} + 2N, \ \sigma_N^2 = 2N + \sum_{m=1}^{N} \left(\frac{1}{np_m} - \frac{1}{\lambda_n}\right) = \widetilde{\sigma}_N^2 - N\lambda_n^{-1}. \tag{4.1}$$

Set $\hat{X}_N^2 = (\chi_N^2 - N)/\sigma_N$. Since $\chi_N^2 = CR_N(1)$ as the consequence of Theorem 3.1(ii) and Theorem 3.2 we obtain

Theorem 4.1. *Let $n \to \infty$, $N \to \infty$ such that the condition (3.9) or*

$$\lambda_n \to \infty \ \text{ and } \ c_1 \le Np_{\min} \le Np_{\max} \le c_2 N^{1/3}, \tag{4.2}$$

is fulfilled. Then for all x_N: $0 \le x_N = o(N^{1/6})$ one has

$$P\left\{\hat{X}_N^2 > x_N\right\} = (1 - \Phi(x_N))(1 + o(1)) \ \text{ and } P\left\{\hat{X}_N^2 < -x_N\right\} = \Phi(-x_N)(1 + o(1)). \tag{4.3}$$

Remark 4.2. Let Y be a r.v. such that $P\{Y = p_m^{-1}\} = p_m$, $m = 1, \dots, N$. Then $EY = N$ and $\mathrm{var}Y = \sum_{m=1}^N p_m^{-1} - N^2 = n \sum_{m=1}^N \left((np_m)^{-1} - \lambda_n^{-1}\right) \ge 0$. Hence $\sigma_N^2 \ge 2N$. Under the conditions (4.2) $np_{\min} \to \infty$, and hence $\sigma_N^2 = 2N(1 + o(1))$.

From Theorem 2.2 we derive the relations (4.3) under weaker conditions than Theorem 4.1. Set $\nabla_n = \max\left(1, (np_{\min})^{-1}\right)$.

Theorem 4.3. *Let $p_{\max} = o(1)$. Then the relations (4.3) are hold uniformly in $0 \le x_N = o\left(\min\left(\left(\sigma_N^3 / \widetilde{\sigma}_N^2 \nabla_n\right)^{1/3}, n^{1/6}, p_{\max}^{-1/4}\right)\right)$.*

Corollary 4.4. *Let $np_{\min} \ge c_0 > 0$. Then the relations (4.3) are hold uniformly in $0 \le x_N = o(\min(N^{1/6}, p_{\max}^{-1/4}))$, while if additionally $p_{\max} = O(N^{-2/3})$ then uniformly in $0 \le x_N = o(N^{1/6})$.*

For the example of multinomial model considered in Remark 3.4 Corollary 4.4 says that if $\lambda_n \geq c_0$, some $c_0 > 0$ (the condition weaker than $\lambda_n \to \infty$ as in Rem. 3.4) the relations (4.3) hold uniformly in non-negative $x_N = o(\min(N^{1/6}, N^{(1-\alpha)/4}))$.

The following consequence of Theorem 4.3 is useful in studying the intermediate asymptotic efficiency of the chi-squared test in verifying uniformity of a discrete distribution versus a family of sequences of alternatives approaching the hypothesis, see Appendix B.

Corollary 4.5. *Let* $p_m = N^{-1}(1 + \delta(n)\ell_{m,n})$, $m = 1,\ldots,N$, *where* $\delta(n) \to 0$ *and*

$$\sum_{m=1}^{N} \ell_{m,n} = 0, \ \ \frac{1}{N}\sum_{m=1}^{N} \ell_{m,n}^2 = \ell^2 < \infty.$$

Then for arbitrary λ_n *and* x_N *such that* $0 \leq x_N = o\left(N^{1/6}\min(1,\lambda_n^{2/3})\right)$ *the relations (4.3) hold with* $\sigma_N^2 = 2N$.

Remark 4.6. From (4.3) when $x_N \to \infty$ we obtain $\ln P\left\{\chi_N^2 > x_N\sigma_N + N\right\} = -x_N^2/2(1+o(1))$. From [12], Eq. (2.17)) this relation follows if $\ln N = o(x_N^2)$, $x_N = o(N^{1/6})$ and $N = o(n^{3/8})$ (*i.e.* the multinomial model is so dense that $\lambda_n \gg n^{5/8}$). Last condition excludes, for instance, the case recommended by Mann and Wald [18], who obtained the relation $N = cn^{2/5}$ concerning the optimal choice of the number of groups in chi-square goodness of fit test.

4.2. Likelihood ratio statistic Λ_N

For the statistic Λ_N Theorem 3.1 and 3.2 imply

Theorem 4.7. *(i) Let* $n \to \infty$ *and* $N \to \infty$ *such that the condition (3.9) is fulfilled. Then for the* $P\{\Lambda_N > x_N\sigma_N(0) + A_N(0)\}$ *and* $P\{\Lambda_N < -x_N\sigma_N(0) + A_N(0)\}$ *the relations (2.7) and (2.8), respectively, are hold uniformly in* $x_N \geq 0$, $x_N = o(\sqrt{N})$.
(ii) Let $p_{\max} = O(N^{-2/3})$ *and* $np_{\min} \to \infty$. *Then uniformly in* $0 \leq x_N = o\left(N^{1/6}\right)$ *one has*

$$\begin{aligned} P\left\{\Lambda_N > x_N\sqrt{2N} + N\right\} &= (1-\Phi(x_N))(1+o(1)), \\ P\left\{\Lambda_N < -x_N\sqrt{2N} + N\right\} &= \Phi(-x_N)(1+o(1)). \end{aligned} \tag{4.4}$$

Remark 4.8. Theorem 4.7 (ii) imply $\log P\left\{\Lambda_N > x_N\sqrt{2N} + N\right\} == -x_N^2/2\,(1+o(1))$ if $x_N \to \infty$ and $x_N = o(N^{1/6})$. From ([12], Eq. (2.13)) such relation follows under additionally conditions $N = o\left(n^{3/7}\right)$ and $\ln N = o(x_n^2)$.

4.3. The Freeman–Tukey statistic T_N^2

FT statistic is $CR_N(-1/2)$, therefore directly from Theorems 3.1 and 3.2 it follows

Theorem 4.9. *(i) Let* $n \to \infty$, $N \to \infty$ *such that the condition (3.9) is fulfilled. Then for the*

$$P\left\{T_N^2 > x_N\sigma_N(-1/2) + A_N(-1/2)\right\} \text{ and } P\left\{T_N^2 < -x_N\sigma_N(-1/2) + A_N(-1/2)\right\}$$

the relations (2.7) and (2.8), respectively, are hold uniformly in $0 \leq x_N = o(\sqrt{N})$.
(ii) Let $p_{\max} = O(N^{-2/3})$ *and* $np_{\min} \to \infty$. *Then uniformly in* $0 \leq x_N = o(N^{1/6})$ *we have*

$$P\left\{T_N^2 > x_N\sqrt{2N} + N\right\} = (1-\Phi(x_N))(1+o(1))$$

$$P\left\{T_N^2 < -x_N\sqrt{2N} + N\right\} = \Phi(-x_N)(1+o(1)).$$

4.4. The statistic $CR_N(2/3)$

This statistic has been recommended by Cressie-Read ([3], p. 463)). Theorems 3.1 and 3.2 directly imply

Theorem 4.10. *Let $n \to \infty$ and $N \to \infty$.*

(i) If the condition (3.9) is fulfilled then uniformly in $0 \le x_N = o(N^{1/8})$

$$P\left\{CR_N(2/3) > x_N\sigma_N(2/3) + A_N(2/3)\right\} = (1 - \Phi(x_N))(1 + o(1)),$$

$$P\left\{CR_N(2/3) < -x_N\sigma_N(2/3) + A_N(2/3)\right\} = \Phi(-x_N)(1 + o(1))$$

(ii) if $p_{\max} = O(N^{-2/3})$ and $np_{\min} \to \infty$ then uniformly in $0 \le x_N = o(N^{1/6})$

$$P\left\{CR_N(2/3) > x_N\sqrt{2N} + N\right\} = (1 - \Phi(x_N))(1 + o(1)),$$

$$P\left\{CR_N(2/3) < -x_N\sqrt{2N} + N\right\} = \Phi(-x_N)(1 + o(1)).$$

4.5. The count statistics

4.5.1. First we Consider the CS (1.3)

Remind for the μ_r statistic $h_m(x) = \mathrm{I}\{x = r\}$, and set $\pi_s(\lambda) = \lambda^s e^{-\lambda}/s!$. The notation (2.1) in this case has the following form:

$$A_N = A_{rN} = \sum_{m=1}^{N} \pi_r(np_m), \;\; \gamma_N = \gamma_{rN} = n^{-1}\sum_{m=1}^{N}(r - np_m)\pi_r(np_m),$$

$$\sigma_N^2 = \sigma_{rN}^2 = \sum_{m=1}^{N} \pi_r(np_m)(1 - \pi_r(np_m)) - n\gamma_{rN}^2.$$

Theorem 4.11. *For fixed $r \ge 0$ uniformly in*

$$0 \le x_N = o\left(\min\left(\sigma_{rN}^{1/3}, (N^{-1}\sigma_{rN}^3)^{1/3}, n^{1/4}, (np_{\max}^2)^{-1/4}\right)\right)$$

one has

$$P\left\{\mu_r > x_N\sigma_{rN} + A_{rN}\right\} = (1 - \Phi(x_N))(1 + o(1)), \tag{4.5}$$

$$P\left\{\mu_r < -x_N\sigma_{rN} + A_{rN}\right\} = \Phi(-x_N)(1 + o(1)). \tag{4.6}$$

Let there exist $c_1 > 0$ and $c_2 > 0$ such that

$$c_1 \le Np_{\min} \le Np_{\max} \le c_2. \tag{4.7}$$

The condition $\sigma_{rN} \to \infty$, that is var$\mu_r \to \infty$, is necessary condition in order above kind large deviation results could be valid for μ_r. In turn, this condition is fulfilled for the following cases of behavior of the parameter $\lambda_n = n/N$:

(i) $\lambda_n \to \lambda \in (0,\infty)$.

(ii) $\lambda_n \to 0$ and $N\lambda_n^{\hat{r}} \to \infty$, where $\hat{r} = \max(2,r)$.

(iii) $\lambda_n \to \infty$ but $\lambda_n \leq \delta c_2^{-1}\ln N + r\ln\ln N$, some delta $\in (0,1)$, c_2 from the condition (4.7).

For all these three cases $\sigma_{rN}^2 \sim \text{var}\mu_r \sim E\mu_r \sim A_{rN} \to \infty$. Moreover, $\sigma_{rN}^2 \sim cN$ for (i) case, $\sigma_{rN}^2 \sim cN\lambda_n^{\hat{r}}$ for (ii) case, and $\sigma_{rN}^2 \sim cN^{1-\delta}$ for (iii) case; here a constant $c > 0$. These (i), (ii) and (iii) cases correspond to the central, left intermediate and right intermediate zones, respectively, defined by Kolchin *et al.* [16]. Applying Theorem 2.1 for (i) case and Theorem 4.6. for (ii) and (iii) cases we obtain the following.

Corollary 4.12. *Let the condition (4.7) is fulfilled. Then for the statistic μ_r, fixed $r \geq 0$, in case (i): the relations (2.7) and (2.8), where $R_N(\eta) = \mu_r$, $A_N = A_{rN}$ and $\sigma_N^2 = \sigma_{rN}^2$, are hold uniformly in $0 \leq x_N = o(N^{1/2})$; in case (ii): if additionally $N = o(n^{3\hat{r}/(3\hat{r}-1)})$, then the relations (4.5) and (4.6) are hold uniformly in $0 \leq x_N = o(N^{1/6}\lambda_n^{\hat{r}/2})$; in case (iii): the relations (4.5) and (4.6) are hold uniformly in $0 \leq x_N = o(N^{(1-\delta)/6})$.*

Remark 4.13. We have $w_1 = N - \mu_0$ and $C_n = \mu_0 - (n - N)$. So

$$P\left\{C_n - EC_n > x\sqrt{VarC_n}\right\} = P\left\{\mu_0 - E\mu_0 > x\sqrt{Var\mu_0}\right\}$$

and

$$P\left\{w_1 - Ew_1 > x\sqrt{Varw_1}\right\} = P\left\{\mu_0 - E\mu_0 < -x\sqrt{Var\mu_0}\right\}.$$

Hence, assertions of Corollary 4.12 corresponding to the case $r = 0$ can be reformulated for w_1 and C_n statistics as well.

4.5.2. The Statistic (1.4) Φ_N

We will consider the case when $p_1 = ... = p_N = N^{-1}$, the classical random allocation scheme, and assume that the levels $\nu_1, \dots, \nu_N$ are independent copies of a non-negative integer-valued r.v. ν such that $P\{\nu = 0\} = \beta_0 < 1$. Set

$$\vartheta = \min\{l : l > 0, P\{\nu = l\}\},\ \ \alpha(\vartheta) = P\{\nu = \vartheta\}/\vartheta!,\ \ \tau(\lambda) = \sum_{l=0}^{\infty} \pi_l(\lambda) P\{\nu > l\}$$

Then

$$A_N = N\tau(\lambda_n),\ \ \gamma_n = \tau'(\lambda_n),\ \ \sigma_N^2 = N\left[\tau(\lambda_n)(1-\tau(\lambda_n)) - \lambda_n(\tau'(\lambda_n))^2\right]. \tag{4.8}$$

Here $\tau'(\lambda)$ is the derivative of $\tau(\lambda)$. It is easy to see that $\tau'(\lambda) = -\sum_{l=0}^{\infty}\pi_l(\lambda)P\{\nu = l+1\}$, and hence the function $\tau(\lambda)$ is strongly decreasing if $\beta_0 < 1$. Notice that, if $\lambda_n \to 0$, then

$$\tau(\lambda_n) = 1 - \beta_0 - \beta(\vartheta)\lambda_n^{\vartheta}(1 + O(\lambda_n)). \tag{4.9}$$

In case $\lambda_n \to \infty$ the following two variants of $\tau(\lambda_n)$'s behavior are

$$\tau(\lambda_n) = \rho\lambda_n^{\chi}e^{-\mu\lambda_n}\left(1 + O(\lambda_n^{-\delta})\right),\ \rho,\ \mu,\ \delta\ > 0 \text{ and } \chi \geq 0, \tag{4.10}$$

$$\tau(\lambda_n) = g\lambda_n^{-\varepsilon}\left(1 + O(\lambda_n^m e^{-\lambda_n})\right) \text{ where } g,\ m \quad > 0 \text{ and } \varepsilon \geq 1. \tag{4.11}$$

From Theorems 2.1 and 2.2 it follows

Theorem 4.14. *Suppose $p_1 = ... = p_N = N^{-1}$ and the levels $\nu_1, \ldots, \nu_N$ are i.i.d. r.v.s..*
(i) Let $\lambda_n \to \lambda \in (0,\infty)$. The relations (2.7) and (2.8), where $R_N(\eta) = \Phi_N$, A_N and σ_N^2 are defined in (4.8), are hold uniformly in $0 \leq x_N = o(\sqrt{N})$.
(ii) Let $\lambda_n \to 0$. The relations (2.13) and (2.14), where $R_N(\eta) = \Phi_N$, A_N and σ_N^2 are defined in (4.8), are valid uniformly in $0 \leq x_N = o\left(N^{1/6}\lambda_n^{\vartheta/2}\right)$.
(iii) Let $\lambda_n \to \infty$ and or (4.9) or (4.10) is fulfilled. The relations (2.13) and (2.14), where $R_N(\eta) = \Phi_N$, A_N and σ_N^2 are defined in (4.8), are valid uniformly: in $0 \leq x_N = o\left(N^{1/6}\lambda_n^{\chi/2}e^{-\mu\lambda_n/2}\right)$ if (4.10) is fulfilled , and in $0 \leq x_N = o\left(N^{1/6}\lambda_n^{-\varepsilon/2}\right)$ if (4.11) is fulfilled.

5. Proofs

We still use the notation of Section 2. Additionally let $\mathcal{C}_k(\varsigma)$ stands for the cumulant of kth order of the r.v.ς and $T_N = h_1(\xi_1) + \cdots + h_N(\xi_N)$, $S_N = (\xi_1 - np_1) + \cdots + (\xi_N - np_N)$; also to keep notation simple we will write R_N instead of $R_N(\eta)$.

Proposition 5.1. *Let the conditions (i) and (ii) of Theorem 2.2 be fulfilled. Then for each integer $s \in [3\ ,\ k_n]$, where k_n satisfies the condition (2.10), it holds*

$$\mathcal{C}_s(R_N) = \mathcal{C}_s(T_N)\left(1 + o\left(1\right)\right). \tag{5.1}$$

Proof. We start from the following

Lemma 5.2. *For any non-negative s it holds*

$$ER_N^s = \upsilon_n \int_{-\pi\sqrt{n}}^{\pi\sqrt{n}} ET_N^s \exp\left\{i\tau\frac{S_N}{\sqrt{n}}\right\} \mathrm{d}\tau,$$

where

$$\upsilon_n \overset{def}{=} \left(2\pi\sqrt{n}P\{S_N = 0\}\right)^{-1} = \frac{n!e^n}{2\pi n^n\sqrt{n}} = \frac{1}{\sqrt{2\pi}}\left(1 + o\left(\frac{1}{n}\right)\right). \tag{5.2}$$

Proof. It is well known that $L((\eta_1, \ldots, \eta_N)) = L((\xi_1, \ldots, \xi_N)/S_N = 0)$, where $L(X)$ stands for the distribution of a random vector X. Hence $ER_N^k = E\left(T_N^k\,|S_N = 0\right)$. On the other hand, $E\left(T_N^k e^{i\tau S_N}\right) = E\left\{e^{i\tau S_N}E\left(T_N^k\,|S_N\right)\right\}$. Now Lemma 5.2 follows by Fourier inversion. The equation (5.2) follows due to fact that $S_N + n \sim Poi(n)$ and Stirling's formula.

Let integer $s \in [3,k_n]$. Use Lemma 5.2 to write

$$ER_N^s = \upsilon_n \sum_{l=1}^{s} \sum\nolimits'_{l,s} \sum\nolimits''_{l} \int_{-\pi\sqrt{n}}^{\pi\sqrt{n}} E\left((h_{j_1}(\xi_{j_1}))^{s_1} \cdot \ldots \cdot (h_{j_l}(\xi_{j_l}))^{s_l} \exp\left\{i\tau\frac{S_N}{\sqrt{n}}\right\}\right) \mathrm{d}\tau, \tag{5.3}$$

where $\sum'_{l,s}$ is the summation over all l-tuples $(s_1, \ldots, s_l)$ with non-negative integer components such that $s_1 + \cdots + s_l = s$; $\sum''_{l}$ is the summation over all l-tuples $(j_1, \ldots, j_l)$ such that $j_i \neq j_r$ for $i \neq r$ and $j_m = 1, 2, \ldots, N$; $m = 1, 2, \ldots, l$.

Set $S_{l,N}=\sum_{i=1}^{l}(\xi_{j_i}-np_{j_i}),\quad d_{l,N}=\sum_{i=1}^{l}p_{j_i},\ l=1,\dots,s$ and write

$$\begin{aligned}
&\int_{-\pi\sqrt{n}}^{\pi\sqrt{n}}E\left((h_{j_1}(\xi_{j_1}))^{s_1}\cdot\ldots\cdot(h_{j_l}(\xi_{j_l}))^{s_l}\exp\left\{i\tau\frac{S_N}{\sqrt{n}}\right\}\right)\mathrm{d}\tau\\
&=\int_{-\pi\sqrt{n}}^{\pi\sqrt{n}}E\left((h_{j_1}(\xi_{j_1}))^{s_1}\cdot\ldots\cdot(h_{j_l}(\xi_{j_l}))^{s_l}\exp\left\{i\tau\frac{S_{l,N}}{\sqrt{n}}\right\}\right)E\exp\left\{i\tau\frac{S_N-S_{l,N}}{\sqrt{n}}\right\}\mathrm{d}\tau\\
&=\int_{-\pi\sqrt{n}}^{\pi\sqrt{n}}E\left((h_{j_1}(\xi_{j_1}))^{s_1}\cdot\ldots\cdot(h_{j_l}(\xi_{j_l}))^{s_l}\exp\left\{i\tau\frac{S_{l,N}}{\sqrt{n}}\right\}\right)\\
&\quad\cdot\left(E\exp\left\{i\tau\frac{S_N-S_{l,N}}{\sqrt{n}}\right\}-\exp\left\{-\frac{\tau^2}{2}(1-d_{l,N})\right\}\right)\mathrm{d}\tau\\
&\quad+\int_{-\pi\sqrt{n}}^{\pi\sqrt{n}}\exp\left\{-\frac{\tau^2}{2}(1-d_{l,N})\right\}E\left((h_{j_1}(\xi_{j_1}))^{s_1}\cdot\ldots\cdot(h_{j_l}(\xi_{j_l}))^{s_l}\left(\exp\left\{i\tau\frac{S_{l,N}}{\sqrt{n}}\right\}-1\right)\right)\mathrm{d}\tau\\
&\quad+E\left[(h_{j_1}(\xi_{j_1}))^{s_1}\cdot\ldots\cdot(h_{j_l}(\xi_{j_l}))^{s_l}\right]\int_{-\pi\sqrt{n}}^{\pi\sqrt{n}}\exp\left\{-\frac{\tau^2}{2}(1-d_{l,N})\right\}\mathrm{d}\tau\overset{def}{=}J_1+J_2+J_3.
\end{aligned}\tag{5.4}$$

We have

$$\begin{aligned}
E\exp\left\{\frac{i\tau(\xi_m-np_m)}{\sqrt{n}}\right\}&=\exp\left\{np_m\left(e^{i\tau/\sqrt{n}}-1-\frac{i\tau}{\sqrt{n}}\right)\right\}\\
&=\exp\left\{-\frac{\tau^2}{2}p_m+\frac{\theta\tau^3}{6\sqrt{n}}p_m\right\},
\end{aligned}\tag{5.5}$$

$$\left|E\exp\left\{\frac{i\tau(\xi_m-np_m)}{\sqrt{n}}\right\}\right|=\exp\left\{-2np_m\sin^2\frac{\tau}{2\sqrt{n}}\right\}\le\exp\left\{-\frac{2p_m}{\pi^2}\tau^2\right\},\tag{5.6}$$

because $\sin^2 u/2\ge u^2/\pi^2$, $|u|\le\pi$, and

$$d_{lN}\le k_np_{\max}=o(1),l=1,\dots,s,\tag{5.7}$$

since (2.10).

In order to get an upper bound for the $|J_1|$ we first write the integral J_1 as a sum of two integrals, say J'_1 and J''_1, over intervals $|\tau|\le\pi\sqrt{n}/2$ and $\pi\sqrt{n}/2\le|\tau|\le\pi\sqrt{n}$, respectively. Next, noting that $E\exp\{i\tau(S_N-S_{l,N})/\sqrt{n}\}=\prod_m E\exp\{i\tau(\xi_m-np_m)/\sqrt{n}\}$, where the product runs over $(1,\dots,N)$ except $(j_1,\dots,j_l)$, we use relations (5.5) and (5.6), respectively in J'_1 and J''_2. Then quite clear algebra gives

$$\begin{aligned}
|J_1|&\le E|h_{j_1}(\xi_{j_1})|^{s_1}\cdot\ldots\cdot E|h_{j_l}(\xi_{j_l})|^{s_l}\\
&\quad\cdot\left[\int_{-\pi\sqrt{n}/2}^{\pi\sqrt{n}/2}\exp\left\{-\frac{\tau^2}{2}(1-d_{l,N})\right\}\left|\exp\left\{\frac{\theta\tau^3}{6\sqrt{n}}(1-d_{l,N})\right\}-1\right|\mathrm{d}\tau\right.\\
&\quad\left.+\int_{\frac{\pi\sqrt{n}}{2}\le|\tau|\le\pi\sqrt{n}}\left(\left|E\exp\left\{i\tau\frac{S_N-S_{l,N}}{\sqrt{n}}\right\}\right|+\exp\left\{-\frac{\tau^2}{2}(1-d_{l,N})\right\}\right)\mathrm{d}\tau\right]\\
&\le E|h_{j_1}(\xi_{j_1})|^{s_1}\cdot\ldots\cdot E|h_{j_l}(\xi_{j_l})|^{s_l}\left[\int_{-\pi\sqrt{n}/2}^{\pi\sqrt{n}/2}\exp\left\{-\frac{6-\pi}{12}(1-d_{l,N})\tau^2\right\}\frac{|\tau|^3}{6\sqrt{n}}(1-d_{l,N})\mathrm{d}\tau\right.\\
&\quad\left.+C_5\exp\left\{-\frac{n}{4}(1-d_{l,N})\right\}\right].
\end{aligned}$$

Hence,

$$J_1 = O\left(\frac{1}{\sqrt{n}}\right) E\left((h_{j_1}(\xi_{j_1}))^{s_1} \cdot \ldots \cdot (h_{j_l}(\xi_{j_l}))^{s_l}\right). \tag{5.8}$$

since (5.7) and the functions $h_m(\,\cdot\,) \geq 0$. Next apply inequality $\left|e^{it} - 1 - it\right| \leq t^2/2$ to get

$$\begin{aligned}
J_2 &= \frac{C_6}{n} E\left(|h_{j_1}(\xi_{j_1})|^{s_1} \cdot \ldots \cdot |h_{j_l}(\xi_{j_l})|^{s_l} \left(\sum_{i=1}^{l} (\xi_{j_i} - np_{j_i})\right)^2\right) \\
&\leq \frac{C_7}{n} l \sum_{i=1}^{l} E\left(|h_{j_i}(\xi_{j_i})|^{s_i} (\xi_{j_i} - np_{j_i})^2\right) \prod_{m \neq i, m=1}^{l} E|h_{j_m}(\xi_{j_m})|^{s_m} \\
&\leq C_8 E\left((h_{j_1}(\xi_{j_1}))^{s_1} \cdot \ldots \cdot (h_{j_l}(\xi_{j_l}))^{s_l}\right) \frac{s}{n} \sum_{i=1}^{l} \left(s_i^{a_1} (np_{j_i})^{b_1} + s_i^{a_2} (np_{j_i})^{b_2}\right) \\
&\leq C_9 E\left((h_{j_1}(\xi_{j_1}))^{s_1} \cdot \ldots \cdot (h_{j_l}(\xi_{j_l}))^{s_l}\right) \frac{s}{n} \left((np_{\max})^{b_1} s^{\max(1,a_1)} + (np_{\max})^{b_2} s^{\max(1,a_2)}\right) \\
&= o\,(1)\, E\left((h_{j_1}(\xi_{j_1}))^{s_1} \cdot \ldots \cdot (h_{j_l}(\xi_{j_l}))^{s_l}\right),
\end{aligned} \tag{5.9}$$

since the fact that $s_1^a + \cdots + s_l^a = s^{\max(1,a)}$ (because $s_1 + \cdots + s_l = s$), $s \leq k_n$ and conditions (2.9), (2.10).

By a simple algebra we obtain

$$J_3 = \sqrt{2\pi} E\left((h_{j_1}(\xi_{j_1}))^{s_1} \cdot \ldots \cdot (h_{j_l}(\xi_{j_l}))^{s_l}\right) (1 + o\,(1)). \tag{5.10}$$

since (5.7) Now apply (5.8), (5.9) and (5.10) in the (5.4) to get

$$\begin{aligned}
&\int_{-\infty}^{\infty} E\left((h_{j_1}(\xi_{j_1}))^{s_1} \cdot \ldots \cdot (h_{j_l}(\xi_{j_l}))^{s_l} \exp\left\{i\tau \frac{S_N}{\sqrt{n}}\right\}\right) \mathrm{d}\tau \\
&\qquad = \sqrt{2\pi}\,(1 + o\,(1))\, E\left((h_{j_1}(\xi_{j_1}))^{s_1} \cdot \ldots \cdot (h_{j_l}(\xi_{j_l}))^{s_l}\right).
\end{aligned} \tag{5.11}$$

Note that in (5.9) and (5.10), and hence in (5.11), the $o(1)$ is uniform in all l-tuples $(s_{j_1}, \ldots, s_{j_l})$ and $s \leq k_n$. For every integer $s \leq k_n$ and k_n satisfying the condition (2.10) the relations (5.2), (5.3) and (5.11) imply $ER_N^s = ET_N^s\,(1 + o\,(1))$. Proposition 5.1 follows from this equality and Assertion A.5.

Let $\lfloor x \rfloor$ stands for the largest integer which less than or equal to x, $\xi \sim Poi(\lambda)$, $\lambda > 0$, $\mu_v(\lambda) = E(\xi - \lambda)^v$ and D_λ denote differentiation w.r.t. λ.

Lemma 5.3. *For any integer $v \geq 2$ one has*

$$\mu_v(\lambda) = v! \sum_{l=1}^{\lfloor v/2 \rfloor} c_{l,v} \lambda^l, \tag{5.12}$$

where

$$0 < c_{l,v} < 1/l!, l = 1, 2, \ldots, \lfloor v/2 \rfloor, \tag{5.13}$$

and

$$(v+1)c_{l,v+1} = lc_{l,v} + c_{l-1,v-1}, \tag{5.14}$$

here $l = 1,2,\dots,\lfloor (v+1)/2 \rfloor$ *if* v *is an even, and* $l = 1,2,\dots,\lfloor (v+1)/2\rfloor - 1$, $(v+1)c_{\lfloor (v+1)/2\rfloor, v+1} = c_{\lfloor (v-1)/2\rfloor, v-1}$ *if* v *is an odd; here we put* $c_{0,v-1} = 0$.

Proof. Apply the Bruno's formula for $Ee^{i\tau(\xi-\lambda)} = \exp\left\{\lambda\left(e^{i\tau} - 1 - i\tau\right)\right\}$ to get

$$\mu_v(\lambda) = v! \sum \lambda^{s_2+\cdots+s_\nu} \prod_{m=2}^{v} \frac{1}{s_m!(m!)^{s_m}} = v! \sum_{l=1}^{\lfloor v/2 \rfloor} c_{l,v}\lambda^l,$$

where $\sum$ is the summation over all non-negative $s_2,\dots,s_v$ such that $2s_2 + \cdots + vs_v = v$ and $l = s_2 + \cdots + s_v$, $c_{l,v} = \sum \prod_{m=2}^{v} \frac{1}{s_m!(m!)^{s_m}}$, and hence $0 < c_{l,v} < 1/l!$. On the other hand, by Riordan ([33], Eq. (4.8))

$$\mu_{v+1}(\lambda) = v\lambda\mu_{v-1}(\lambda) + \lambda D_\lambda \mu_v(\lambda). \tag{5.15}$$

Just to keep notation simple we put $\kappa = \lfloor v/2 \rfloor$ and $\hat{c}_{i,v} = v!c_{l,v}$. Note that $\hat{c}_{1,v} = 1$ for any integer $\nu \geq 0$. Let ν is even, then $\lfloor (v+1)/2 \rfloor = \kappa$ and $\lfloor (v-1)/2\rfloor = \lfloor (v-2)/2 \rfloor = \kappa - 1$. Therefore (5.12) and (5.15) imply

$$\sum_{l=1}^{\kappa} \hat{c}_{l,v+1}\lambda^l = \sum_{l=1}^{\kappa-1} v\hat{c}_{l,v-1}\lambda^{l+1} + \sum_{l=1}^{\kappa} l\hat{c}_{l,v}\lambda^l = \sum_{l=2}^{\kappa}\left(v\hat{c}_{l-1,v-1} + l\hat{c}_{l,v}\right)\lambda^l + \lambda.$$

For the even v the property (5.14) follows. Let now v is odd, then $\lfloor (v-1)/2 \rfloor = \kappa$ and $\lfloor (v+1)/2 \rfloor = \kappa + 1$. From (5.12) and (5.15) obtain

$$\sum_{l=1}^{\kappa+1} \hat{c}_{l,v+1}\lambda^l = \sum_{l=1}^{\kappa} v\hat{c}_{l,v-1}\lambda^{l+1} + \sum_{l=1}^{\kappa} l\hat{c}_{l,v}\lambda^l = \sum_{l=2}^{\kappa+1} v\hat{c}_{l-1,v-1}\lambda^l + \sum_{l=1}^{\kappa} l\hat{c}_{l,v}\lambda^l$$

$$= \sum_{l=2}^{\kappa}\left(v\hat{c}_{l-1,v-1} + l\hat{c}_{l,v}\right)\lambda^l + v\hat{c}_{\kappa,v-1}\lambda^{\kappa+1} + \lambda.$$

This equality proves (5.14) for the odd v. Lemma 5.3 is proved completely.

Using (5.12) we obtain $\mu_v(\lambda) < \lambda D_\lambda \mu_v(\lambda) \leq \lfloor v/2 \rfloor \mu_v(\lambda)$. This together with (5.15) imply

$$\mu_v(\lambda) < \mu_{v+1}(\lambda) \leq \left(\nu\lambda + \lfloor v/2 \rfloor\right)\mu_v(\lambda). \tag{5.16}$$

Use formula (5.15) for $\mu_{v+2}(\lambda)$, next again apply (5.15) to the derivative of $\mu_{v+1}(\lambda)$, after this use the inequalities $\lambda D_\lambda \mu_{v-1}(\lambda) \leq \mu_v(\lambda)$ (because $l\hat{c}_{l,v-1} \leq \hat{c}_{l,v}$, see (5.14)), $\lambda D_\lambda \mu_v(\lambda) \leq \lfloor v/2 \rfloor \mu_v(\lambda)$ and $\lambda^2 D_\lambda^2 \mu_v(\lambda) \leq \lfloor v/2 \rfloor^2 \mu_v(\lambda)$. Then one can observe that for any integer $\nu \geq 0$

$$(\nu+1)\lambda\mu_v(\lambda) < \mu_{v+2}(\lambda) \leq 2v(\lambda + v)\mu_v(\lambda). \tag{5.17}$$

Proof of Theorem 2.2. Due to Proposition 5.1 and condition (ii) we have for large enough N

$$\begin{aligned}\left|\mathcal{C}_s(\sigma_N^{-1}R_N)\right| &\le 2^{-1}\sigma_N^{-s}\sum_{m=1}^{N}\left|\mathcal{C}_s\left(h_m(\xi_m)\right)\right| \le 2^{-1}(s!)^2\widetilde{\omega}_N^2 V_n^{s-2}\sigma_N^{-s}\\ &\le (s!)^2(2\sigma_N V_n^{-1})^{-(s-2)}(\sigma_N^2/\widetilde{\omega}_N^2)^{-1}\\ &\le (s!)^2(2\sigma_N V_n^{-1}(\min(1,\sigma_N^2/\widetilde{\omega}_N^2))^{1/(s-2)})^{-(s-2)} \le (s!)^2 W_N^{-(s-2)},\end{aligned}$$

for all s: $3\le s\le k_n$, see (2.10). Theorem 2.2 follows now from Assertion A.4 with $\Delta = W_N$ and $\tilde{k}=k_n$.
Proof of Theorem 3.1. From (3.3), (3.4) in the case $d\neq 0$, and in the case $d=0$ from (3.6), (3.7), and the fact that $h_{0,m}(\xi_m) = 2np_m\left(1+\widehat{\xi}_m\right)\ln\left(1+\widehat{\xi}_m\right)$, where $\widehat{\xi}_m = (\xi_m - np_m)/np_m$, using inequalities $x \le (1+x)\ln(1+x) \le x + x^2/2$ one can observe that under the condition (3.9) for large enough N

$$\sigma_N^2(d) \ge cN. \tag{5.18}$$

Case (i) follows from Theorem 2.1, (5.18) and that (see (3.1)) $E\exp\left\{H\left|h_{d,m}(\xi_m)\right|\right\} \le E\exp\left\{2H(np_m)^{-d}\xi_m\right\} \le \exp\left\{np_m\left(\exp\left\{2H(np_m)^{-d}\right\}-1\right)\right\} \le C(d,H,c_3,c_4)$, the constant depending on arguments only, since condition (3.9).
Proof of the cases (ii) and (iii) consist in verifying of the conditions of Theorem 2.2. Due to Riordan ([33], Eq. (3.8)) $E\xi^{s+1} = \lambda(E\xi^s + D_\lambda E\xi^s)$, $\xi\sim Poi(\lambda)$and integer $s\ge 0$. Using this and raw inequality $\lambda^i D_\lambda^i E\xi^s < s^i E\xi^s$ it is not hard to see that $E\xi^{s+2}\le 4\max(s^2,\lambda^2)E\xi^s$ for integer $s\ge 1$, and $E\xi^{s+2}\le 4\max(s^3,\lambda^3)E\xi^s$ for non-integer $s>0$. Hence, the condition (2.9) is fulfilled with $a_1=2$, $b_1=0$, for $s>\lambda$, and $a_2=0$, $b_2=2$, for $s\le\lambda$, if s is an integer, and similarly with $a_1=3$, $b_1=0$ and $a_2=0$, $b_2=3$ for the non-integer s. Further,

$$E\exp\left\{H\left(h_{d,l}(\xi_l)/\sqrt{Var h_{d,l}(\xi_l)}\right)^{1/(d+1)}\right\} = E\exp\left\{H_1\xi_l\right\} \le C_0(d,c_4,\lambda),$$

where $H_1 = H/(Var\xi_l^{d+1})^{1/2(d+1)} \ge H/C_1(d,c_4,\lambda) > 0$, since $np_j \le c_4\lambda$ and $Var\xi_j^{d+1} < E\xi_j^{2(d+1)} \le C_1(d,c_4,\lambda)$. Therefore, due to Assertion A.3 we have

$$\left|\mathcal{C}_s(h_{d,l}(\xi_l))\right| = (Var\ h_{d,l}(\xi_l))^{s/2}\mathcal{C}_s\left(h_{d,l}(\xi_l)/\sqrt{Var\ h_{d,l}(\xi_l)}\right)$$

$$\le (s!)^{1+d}\left(\Delta(\max_{1\le j\le N} Var\ h_{d,l}(\xi_l))^{1/2}\right)^{s-2} Var\ h_{d,l}(\xi_l),$$

where $\Delta = \left(4C_0(d,c_4,\lambda)e^{1+d}((1+d)/\max(1,H_1))^{3(1+d)}\right)^{-1}$. Thus, condition (2.11) is fulfilled with $v = \max(1,d)$, $V_n = \Delta(\max_{1\le l\le N} Var\ h_{d,l}(\xi_l))^{1/2}$ and $\omega_l^2 = Var\ h_{d,l}(\xi_l)$, and hence there exists a $c>0$ such that $W_N = c\sqrt{N}$, since (3.9). Cases (ii) and (iii) of Theorem 3.1 follows.

Proof of Theorem 3.2. The PDS can be rewritten as

$$CR_N(d) = \sum_{l=1}^{N} np_j\hat{h}_{j,d}(\eta_j),$$

where

$$\begin{aligned}\hat{h}_{j,d}(x) &= \frac{2}{\mathrm{d}(d+1)}\left[(x/np_l)^{d+1} - (d+1)(x/np_l) + d\right] \\ &= \frac{2}{\mathrm{d}(d+1)}\left[\left(1+\frac{x-np_j}{np_l}\right)^{d+1} - (d+1)\left(1+\frac{x-np_j}{np_l}\right) + d\right], d \neq 0, \\ \hat{h}_{j,0}(x) &= 2\left[(x/np_l)\log(x/np_l) - (x-np_j)/np_j\right] \\ &= 2\left[\left(1+\frac{x-np_j}{np_j}\right)\log\left(1+\frac{x-np_j}{np_j}\right) - \left(1+\frac{x-np_j}{np_j}\right)\right]. \end{aligned} \tag{5.19}$$

Set $\widehat{\xi}_m = (\xi_m - np_m)/np_m$. Recall that $np_{\min} \to \infty$, so the r.v. $\sqrt{np_m}\widehat{\xi}_m$ has asymptotically normal distribution. Hence

$$\widehat{\xi}_m = O_p\left((np_m)^{-1/2}\right). \tag{5.20}$$

Now, we apply Taylor expansion formula in (5.19) to get

$$\hat{h}_{m,d}(\xi_m) = \widehat{\xi}_m^2 + O_p\left(\widehat{\xi}_m^3\right) = \widehat{\xi}_m^2 + O_p\left((np_m)^{-3/2}\right), d > -1. \tag{5.21}$$

By virtue of this fact and that $E\widehat{\xi}_m^2 = (np_m)^{-1}$, $EO_p\left((np_m)^{-3/2}\right) = O\left((np_m)^{-3/2}\right)$ we obtain under the condition (3.12) for all $d > -1$: $A_N(d) = N + o(N)$ and

$$\sigma_N^2(d) = \sum_{m=1}^{N}\frac{1}{np_m} + (2-\lambda_n^{-1})N + \sum_{m=1}^{N} O\left((np_m)^{-5/2}\right) = 2N(1+o(1)). \tag{5.22}$$

Here we used the fact that $\lambda_n \to \infty$ as $np_{\min} \to \infty$, since $1 = p_1 + \cdots + p_N \geq Np_{\min}$ and hence $np_{\min} \leq \lambda_n$. Further proof consists in verifying the conditions of Theorem 2.2. Note that $\hat{h}_{j,d}(x) \geq 0$ for all $d \in (-\infty, \infty)$. Next, by (5.20) and (5.21) we have.

$$E\hat{h}_{m,d}^s(\xi_m) = E\widehat{\xi}_m^{2s}\left(1 + O_p\left((np_m)^{-1/2}\right)\right) = E\widehat{\xi}_m^{2s} + O\left((np_m)^{-(2s+1)/2}\right), \tag{5.23}$$

since $EO_p\left((np_m)^{-(2s+1)/2}\right) = O\left((np_m)^{-(2s+1)/2}\right)$ and by Lemma 5.3 $E\widehat{\xi}_m^{2s} > c(s)(np_m)^{-s}$, where $1 \leq c(s) < s$. Similarly, using (5.20), (5.21) and (5.17) we get

$$\begin{aligned} E\hat{h}_{m,d}^s(\xi_m)(\xi_m - np_m)^2 &= E\widehat{\xi}_m^{2s}(\xi_m - np_m)^2 + O\left((np_m)^{-(2s-1)/2}\right) \\ &\leq 4s(np_m + 2s)E\widehat{\xi}_m^{2s} + O\left((np_m)^{-(2s-1)/2}\right). \end{aligned} \tag{5.24}$$

So the condition (2.9) is fulfilled with $b = 1$ and $a = 1$ if $np_m \geq 2s$, and with $b = 0$ and $a = 2$ if$np_m < 2s$. Next, putting $\tilde{h}_{m,d}(\xi_m) = np_m\hat{h}_{m,d}(\xi_m)$ by virtue (5.21) we obtain

$$E\tilde{h}_{m,d}(\xi_m) = 1 + O\left((np_m)^{-1/2}\right), Var\tilde{h}_{m,d}(\xi_m) = 2 + O\left((np_m)^{-1/2}\right). \tag{5.25}$$

Set $h^*_{m,d}(\xi_m) = \left(\tilde{h}_{m,d}(\xi_m) - E\tilde{h}_{m,d}(\xi_m)\right)/\sqrt{Var\tilde{h}_{m,d}(\xi_m)}$. Using (5.21), (5.25) and Lemma 3.2 (relations (5.12), (5.13)) we obtain

$$E\left|h^*_{m,d}(\xi_m)\right|^s \le 2^s\left((np_m)^{-s}E(\xi_m - np_m)^{2s} + O\left((np_m)^{-s/2}\right)\right)/(Var\tilde{h}_{m,d}(\xi_m))^{s/2}$$
$$\le (2s)!2^{s/2} + O\left((np_m)^{-1}\right) \le (s!)^2 2^{(5s-1)/2}e/\pi\sqrt{s} \le (s!)^2 2^{6(s-2)},$$

since $(2s)! \le (s!)^2 2^{2s} e/\pi\sqrt{2s}$, due to inequalities $\sqrt{2\pi m}\ m^m e^{-m} \le m! \le e\sqrt{m}\ m^m e^{-m}$. Therefore, from Assertion A.2 it follows that

$$\left|\mathcal{C}_s\left(\tilde{h}_{m,d}(\xi_m)\right)\right| \le (s!)^2 2^{6(s-2)}2^{s/2} < (s!)^2 2^{7(s-2)}.$$

Hence, condition (2.11) is fulfilled with $v = 1$, $V_n = 2^7$ and $\omega^2_m = 1$. Thus, for the quantities from (2.10) we have $\widetilde{\omega}^2_N = N,\ \sigma^2_N = 2N(1 + o(1))$, $W_N = 2^{-7}\sqrt{N}$, $K_n(a_1, b_1) = K_n(2, 0) = n^{1/3}$ and $K_n(a_2, b_2) = K_n(1, 1) = p_{\max}^{-1/2}$. Now Theorem 3.2 follows from Theorem 2.2.

Proof of Theorems 3.5, 3.7 and 3.8 will be concluded by applying Theorem 2.2. Formulas (3.14), (3.15), (3.16) and (3.17) are obtained by direct calculation using not hard algebra, taking into account the fact that $np_{\max} \to 0$. For the kernel function $h_{d,m}(x)$given in (3.1) reasoning alike to proof of the case (ii) of Theorem 3.1 it can be observed that the condition (2.9) is fulfilled with $a = 2$,$b = 0$, if $d \ge 1$ is an integer, and with $a = 3$,$b = 0$ for the non-integer $d > -1$and $d = 0$, since $np_{\max} \to 0$. Using Bruno's formula for derivatives of composite function and Stirling's formula we obtain for $d > -1$, $d \ne 0$

$$E(h_{d,m}(\xi_m))^s \le (np_m)^{-d}E\xi_m^{s(\hat{d}+1)}$$

$$\le (np_m)^{-d}E\xi_m^{\left\lfloor s(\hat{d}+1)\right\rfloor+1} \le 2\left(\left\lfloor s(\hat{d}+1)\right\rfloor + 1\right)!(np_m)^{1-d}$$

$$\le 2\sqrt{2\pi(\left\lfloor s(\hat{d}+1)\right\rfloor + 1)}(\left\lfloor s(\hat{d}+1)\right\rfloor + 1)^{(\left\lfloor s(\hat{d}+1)\right\rfloor+1)}\exp\left\{-(\left\lfloor s(\hat{d}+1)\right\rfloor + 1)\right\}(np_m)^{1-d}$$

$$\le (s!)^{1+\hat{d}}[C(d)]^{s-2}(np_m)^{1-d},$$

where $\hat{d} = \max(0,d)$, $C(d) = 2^{4(1+\hat{d})}(1 + \hat{d})^{3(1+\hat{d})}$. From this and Lemma 3.1 of [35], it follows that

$$|\mathcal{C}_s(h_{d,m}(\xi_m))| \le (s!)^{\hat{d}+1}(2C(d))^{(s-2)}(np_m)^{1-d}. \tag{5.26}$$

For the case $d = 0$ we have $Eh^s_{0,m}(\xi_m) = E\xi^s_m\log^s(\xi_m/np_m) \le (np_m)^{-1}E\xi^{s+1}_m$. Therefore, inequality (5.26) with $d = 0$ still true for this case also. Thus the condition (ii) of Theorem 2.2 is fulfilled with $\nu = \max(1,d)$, $V_n = 2C(d)$ and $\omega^2_m = (np_m)^{1-d}$, $d > -1$. Theorem 3.5, 3.7 and 3.8 follows from Theorem 2.2 by using appropriate formula given in (3.14)–(3.17).

Proof of Theorem 4.1 follows in immediate manner from Theorem 3.1 (ii), case $d = 1$, and Theorem 3.2.

Proof of Theorem 4.3. For the chi-square statistic $h_m(x) = (x - np_m)^2/np_m$, hence $A_N = N$, σ^2_N and $\widetilde{\sigma}^2_N$ as in (4.1). Set $\widetilde{\xi}_m = (\xi_m - np_m)/\sqrt{np_m}$. Then $h_m(\xi_m) = \widetilde{\xi}^2_m$ and $E\widetilde{\xi}^{2s}_m(\xi_m - np_m)^2 \le 4s(np_m + 2s)E\widetilde{\xi}^{2s}_m$, since (5.17), hence the condition (2.9) is fulfilled with $b = 1$ and $a = 1$ if $np_m \ge 2s$, and with $b = 0$ and $a = 2$ if

$np_m < 2s$. We would remind the notation $\nabla_n = \max\left(1,(np_{\min})^{-1}\right)$. Set $\varsigma_m = (\widetilde{\xi}_m^2 - E\widetilde{\xi}_m^2)/\sqrt{Var\widetilde{\xi}_m^2}$. Using Lemma 5.3 we obtain $|E\varsigma_m^s| \le 2^s(2s)!(Var\widetilde{\xi}_m^2)^{-s/2}\sum_{l=0}^{s-1}(np_m)^{-l} \le (s!)^2\left(2^7\nabla_n/\sqrt{Var\widetilde{\xi}_m^2}\right)^{s-2}$ for all $s \ge 3$, since $(2s)! \le (s!)^2 2^{2s}e/\pi\sqrt{2s}$, $\sum_{l=0}^{s-1}(np_m)^{-l} \le s\nabla_n^{s-2}$ and $Var\widetilde{\xi}_m^2 = (np_m)^{-1} + 2$. Hence, by Assertion A.2 $|\mathcal{C}_s(\varsigma_m)| \le (s!)^2\left(2^8\nabla_n/\sqrt{Var\widetilde{\xi}_m^2}\right)^{s-2}$. That is $\left|\mathcal{C}_s\left(\widetilde{\xi}_m^2\right)\right| \le (s!)^2\left(2^8\nabla_n\right)^{s-2}Var\widetilde{\xi}_m^2$, $s \ge 3$. Thus condition (ii) of Theorem 4.3 is fulfilled with $V_n = 2^8\nabla_n$, $\omega_m = Var\widetilde{\xi}_m^2$ and $3 \le s \le k_n = o\left(\min(n^{1/3}, p_{\max}^{-1/2}\right)$, because $K_n(1,1) = p_{\max}^{-1/2}$ and $K_n(2,0) = n^{1/3}$. Note that here $\sigma_N^2 \le \widetilde{\sigma}_N^2 = \widetilde{\omega}_N^2 = Var\widetilde{\xi}_1^2 + \cdots + Var\widetilde{\xi}_N^2$. Theorem 4.3 follows from Theorem 2.2 and Remark 3.3.

Proof of Corollary 4.4. We have $\widetilde{\sigma}_N^2 \le N(2 + c_0^{-1})$, $\nabla_n \le \max(1,c_0^{-1})$ and $n \ge c_0N$, since $np_{\min} \ge c_0$. Yet $\sigma_N^2 \ge 2N$, see Remark 4.2. Corollary 4.4 follows.

Proof of Corollary 4.5 is straightforward, since in this case $\nabla_n = \max(1,\lambda_n^{-1})$, $\sigma_N^2 = N\left(2 + d^2\Delta^2(n)\lambda_n^{-1}(1+o(1)\right)$ and $\widetilde{\sigma}_N^2 = 2N\left(1 + (2\lambda_n)^{-1}(1+o(1))\right) \le 2N\nabla_n$.

Proof of Theorem 4.7, 4.9 and 4.10 follows straightforwardly from Theorems 3.1 and 3.2 by putting $d = 0$, $d = -1/2$ and $d = 2/3$ respectively.

Proof of Theorem 4.11 follows from Theorem 2.2. Indeed. In this case $h_m(x) = \mathrm{I}\{x = r\}$ and it is easy to see that the condition (i) of Theorem 2.2 is fulfilled with $a_1 = 0$, $b_1 = 2$ and $a_2 = 0$, $b_2 = 0$, where $\mathcal{N} = \{j{:}np_j > r\}$. Set $\hat{h}_m(\xi_m) = \mathrm{I}\{\xi_m = r\} - \pi_r(np_m)$. We have $\left|E\exp\left\{z\hat{h}_m(\xi_m)\right\}\right| \le e$, within disk $|z| \le 1$. Hence, by well-known Cauchy inequality for coefficients of a power series we have $\left|E\hat{h}_m^s(\xi_m)\right| \le s!e^{s-2}$. Therefore, rewriting line by line the proof of Lemma 3.1 of [35] for the r.v. $\hat{h}_m(\xi_m)$, taking into account that $E\hat{h}_m^2(\xi_m) \le 1$, we find that $|\mathcal{C}_s(h_m(\xi_m))| \le s!(2e)^{s-2}$. Hence the condition (2.11) is fulfilled with $v = 0$(hence with $v = 1$ as well), $V_n = 2e$ and $\omega_m = 1$, $m = 1,\ldots,N$. Thus, we can apply Theorem 2.2 with $K_n(a_1,b_1) = (np_{\max}^2)^{-1/2}$, $K_n(a_2,b_2) = n^{1/2}$, $W_N = \sigma_{rN}\min(1,N^{-1}\sigma_{rN}^2)$. Theorem 4.11 follows.

Proof of Corollary 4.12 follows straightforwardly from Theorem 4.11.

Proof of Theorem 4.14. Remind that for the statistic Φ_N the kernel function $h_m(x) = \mathrm{I}\{v_m > x\}$. If $\lambda_n \to \lambda \in (0,\infty)$, then it is not hard to verify that all conditions of Theorem 2.1 are fulfilled, and hence the case (i) follows.

The cases (ii) and (iii) we will derive from Theorem 2.2. Observe that $Eh_m^s(\xi_m) = \tau(\lambda_n)$ for any $s \ge 1$. Using this fact we see that $Eh_m^s(\xi_m)(\xi_m - \lambda_n)^2 \le \lambda_n(3\lambda_n + 2)Eh_m^s(\xi_m)$. Hence the condition (i) of Theorem 2.2 is fulfilled with

$$a = 0, b = 0 \;\; if \; \lambda_n \le 1, \tag{5.27}$$

$$a = 0, b = 2 \;\; if \; \lambda_n > 1. \tag{5.28}$$

Next, alike to the above proof of Theorem 4.11 it can be shown that $|\mathcal{C}_s(h_m(\xi_m))| \le s!(2e)^{s-2}$, that is in this case also the condition (2.11) is fulfilled with $v = 0$(hence with $v = 1$as well),

$$V_n = 2e \;\; and \;\; \omega_m = 1, \; m = 1,\ldots,N. \tag{5.29}$$

Let $\lambda_n \to 0$. Then from (4.8) and (4.9) by simple algebra we obtain $A_N = N(1-\beta_0)(1+o(1))$, $\sigma_N^2 = N\alpha(\vartheta)\lambda_n^{\vartheta}(1+O(\lambda_n))$. The case (ii) follows from Theorem 2.2, (5.27) and (5.29).

Case (iii). In this case from (4.8) we have $\sigma_N^2 = N\rho\lambda_n^{\chi}e^{-\mu\lambda_n}(1+o(1))$ if (4.10) is fulfilled, and $\sigma_N^2 = Ng\lambda_n^{-\varepsilon}(1+o(1))$ if (4.11) is fulfilled. The proof of the case (iii) concludes from Theorem 2.2, (5.28) and (5.29).

Appendix A.

Below ς is a r.v. with $E\varsigma = 0$, $Var\varsigma = 1$, $\mathcal{C}_k(\varsigma)$ and $\alpha_k(\varsigma)$ cumulant and moment of kth order, respectively of the r.v. ς. The following conditions and assertions are taken from the book by Saulis and Statulevicius [35] and paper by Amosova [1].

Bernstein condition (B_ν): there exist constants $\nu \geq 0$ and $B > 0$ such that $|\alpha_k(\varsigma)| \leq (k!)^{\nu+1}B^{k-2}$, for all $k = 3,4,\ldots$.

Statulevicius condition(S_ν): there exist constants $\nu \geq 0$ and $\Delta > 0$ such that $|\mathcal{C}_k(\varsigma)| \leq (k!)^{\nu+1}\Delta^{-(k-2)}$, for all $k = 3,4,\ldots$.

Linnik condition (L_γ): there exist constants $\gamma \in (0,1]$ and $H > 0$ such that

$$E\exp\{H|\varsigma|^{\gamma}\} \leq C(\gamma).$$

Remark A.1. The Linnik condition is written here in a simplified and convenient for us form; actually it was introduced by Linnik with $\gamma = 4\alpha/(1+2\alpha)$, where $0 < \alpha < 1/2$. Condition $(L_1)$is the Cramer condition.

Assertion A.2. ([35], Lem. 3.3)). If ς satisfy condition (B_ν)then it also satisfy condition (S_ν) with $\Delta = (2\max(1,B))^{-1}$.

Assertion A.3. ([1], pp. 2032–2033)**.** If ς satisfy condition (L_γ)then it also satisfy condition (S_ν) with $v = \gamma^{-1}-1$ and $\Delta = \left(4C(\gamma)e^{1/\gamma}(\gamma H)^{-3/\gamma}\right)^{-1}$.

Assertion A.4. Let a r.v. ς depending on a parameter Δ satisfy condition (S_v) for some $\nu \geq 1$and all k such that $3 \leq k \leq \tilde{k}$. Then there exist constants $c_1 > 0$ and $c_2 > 0$such that for all x: $0 \leq x \leq c_1\min\left(\sqrt{\tilde{k}}, \Delta^{1/(1+2v)}\right)$ one has

$$P(\varsigma > x) = (1-\Phi(x))\left(1 + c\frac{(x+1)}{\Delta^{1/(1+2v)}}\right),$$

$$P(\varsigma < -x) = \Phi(-x)\left(1 + c\frac{(x+1)}{\Delta^{1/(1+2v)}}\right). \tag{A.1}$$

Assertion A.3 can be proved by reasoning alike to those presented by [35] at page in 35, in course of proof of Lemma 2.3, see relations (2.62), (2.65), (2.66) and Lemma 2.2 of the book. Actually it can be proved more general result: assume condition $(S_v)$for some $\nu \geq 0$, and get relation alike (2.6) of Lemma 2.3 of [35], instead of (A.1).

The following relation is well-known (see [35], Eq. (1.34))

Assertion A.5. One has

$$\mathcal{C}_k(\varsigma) = k!\sum(-1)^{m_1+m_2+\cdots+m_k-1}(m_1+m_2+\cdots+m_k-1)!\prod_{l=1}^{k}\frac{1}{m_l!}\left(\frac{\alpha_l(\varsigma)}{l!}\right)^{m_l}$$

here Σ is summation over all non-negative integer $m_1, m_2, \ldots, m_k$ such that $m_1 + 2m_2 + \cdots + km_k = k$.

Appendix B.

We still use notation of the previous sections. Let $(\eta_1, \ldots, \eta_N)$ be a random vector of frequencies of multinomial model $M(n, N, \mathrm{P})$, see Introduction. Consider problem of verifying the hypothesis H_0: $p_m = 1/N$, $m = 1, \ldots, N$, against the sequences of alternatives

$$H_1 = H_{1n} : p_m = \frac{1}{N}\left(1 + \delta(n)\Delta_{m,n}\right), m = 1, \ldots, N, \tag{B.1}$$

where the sequence $\delta(n) \to 0$ as $n \to \infty$, and $\Delta_{m,n}$s such that

$$\max_{1 \le m \le N} |\Delta_{m,n}| \le C, \ \sum_{m=1}^{N} \Delta_{m,n} = 0\,, \ \ \frac{1}{N}\sum_{m=1}^{N} \Delta_{m,n}^2 = 1. \tag{B.2}$$

We are interested in tests based on symmetric variant of statistics (1.1):

$$S_N^h = \sum_{m=1}^{N} h(\eta_m),$$

where h is a non-linear function defined on non-negative axis. Test based on S_N^h statistic is called h-test for brevity. Assume that large values of the statistic S_N^h reject the hypothesis, and $n\lambda_n \to \infty$, as $n \to \infty$, where $\lambda_n = n/N$, the average of observations in the cells.

Remark B.1. One can consider tests based on class of non-symmetric variants of statistics (1.1), but such tests will not be goodness-of-fit, since their asymptotic power will depend on Δ_ms, *i.e.* on alternatives. For details see [6, 9].

Asymptotic properties of h-tests have been studied in [6, 9, 10, 20, 34]. We refer also to [12, 30, 31] and [27], where various asymptotic properties of particular variants of h-test were treated.

The objective of this section is to demonstrate the advantages of Theorem 2.2. To this end, we briefly present some results about "intermediate properties" of h-tests, derived by applying theorems about PDS, the consequences of Theorem 2.2. The detailed analysis of the asymptotic efficiencies (AE) of h-tests is subject of another paper.

In what follows $\alpha_n(h)$ and $\beta_n(h)$ denote the size and power of h-test, respectively; P_i, $E_i S_N^h$ and $Var_i S_N^h$ stand for the probability, expectation and variance of S_N^h , respectively counted under H_i, $i = 0, 1$.

First we make some comments to clarify the meaning of "intermediate properties" of h-tests. There are several approaches to the asymptotic comparison of tests which differ by the conditions imposed on the asymptotic behaviour of the size, the power and the sequence of alternatives; *i.e.* of $\alpha_n(h)$, $\beta_n(h)$ and alternatives H_{1n}, by a condition on the $\delta(n)$, in our case. The conditions imposed on two of them provide a condition for the third. The most common is Pitman's approach, which assumes that $\alpha_n(h) \to \alpha > 0$ and a sequence of alternatives converge to the hypothesis at the rate necessary to $\beta_n(h) \to \beta \in (\alpha, 1)$, In our task this Pitman sub-family of alternatives is H_{1n} (B.1) where $\delta(n) = (n\lambda_n)^{-1/4}$ (note that there is no power of h-tests for the alternatives (B.1) with $\delta(n) = o\left((n\lambda_n)^{-1/4}\right)$). Within class of h-tests (satisfying the asymptotic normality condition of S_N^h) the χ^2 test has most Pitman AE, see [6, 9, 10]. Another "extreme" setting assumes that $\beta_n(h) \to \beta \in (0, 1)$ and the alternatives do not approach the H_0, then the significance level $\alpha_n(h)$ decreases in exponential rate. For our problem this means $\delta(n)$ is a constant. This approach is at the basis of the concept of Bahadur AE; it was developed by Ronzhin (1984), who have considered the sparse multinomial models ($\lambda_n \to \lambda \in (0,\infty)$) and a subclass of h-tests satisfying the Cramér condition (2.6). Again we remind that class of PDS with parameter $d > 0$, which includes the chi-squared statistic, does not satisfy the Cramér condition. We refer to Quine and Robinson (1985), who showed, in particularly, that for the dense models ($\lambda_n \to \infty$) the chi-square test is inferior to the log-likelihood ratio test in terms of the Bahadur AE, in contrast to that these two tests have the same Pitman AE. The same verdict was presented by Kallenberg [12], who have considered as dense models as $\lambda_n/\sqrt{n} \to \infty$. Further, the intermediate (between Pitman and Bahadur settings) approach assumes $\beta_n(h) \to \beta \in (0,1)$and sequences of alternatives approach the hypothesis at the rate slower than in Pitman setting. These conditions provide $\alpha_n(h) \to 0$, but slower than in Bahadur setting. The essence of this setting is that the significance level goes to 0 as n increases, while the asymptotic power, under the underlying sequence of local alternatives, should be non-degenerate. Such requirements call for a balancing of the rates at which the significance levels tend to 0 and the alternatives approach the null model. In our task the sequences of alternatives (B.1) such that

$$\delta(n) \to 0 \ \textit{and} \ (n\lambda_n)^{1/4}\delta(n) \to \infty, \tag{B.3}$$

specifies $\Im_{alt}$-the family of sequences of intermediate alternatives.

The intermediate properties of the h-tests have been studied for the first time by Ivchenko and Mirakhmedov [11]. Let $A_i(h)$ and $\sigma_i^2(h)$ be asymptotic value of $N^{-1}E_iS_N^h$ and $N^{-1}Var_iS_N^h$, $i=0,1$, respectively. Assume that$A_1(h) - A_0(h) > 0$. As a measure of performance of h-test [11] proposed so called "α-intermediate AE" (α-IAE), *viz.* the asymptotic value of

$$\begin{aligned} e_N^\alpha(h) &= -\log P_0\left\{S_N^h > NA_1(h)\right\} \\ &= -\log P_0\left\{S_N^h - NA_0(h) > x_N(h)\sqrt{N}\sigma_0(h)\right\}, \end{aligned} \tag{B.4}$$

where $x_N(h) = \sqrt{N}(A_1(h) - A_0(h))/\sigma_0(h)$.

Under certain condition of the Liyapunov type, see ([10], Eq. (14)), the statistic $\hat{S}_N^h = (S_N^h - NA_0(h))/\sqrt{N}\sigma_0(h)$ has an asymptotically normal distribution, and

$$\begin{aligned} x_N(h) &= \rho(h,\lambda_n)\sqrt{n\lambda_n/2}\ \delta^2(n)(1+o(1)), \\ \rho(h,\lambda_n) &= corr\left(h(\xi) - \tau_n\xi, \xi^2 - (2\lambda_n+1)\xi\right), \end{aligned} \tag{B.5}$$

By virtue of these facts, it can be shown that the asymptotical power of h-tests with critical region $\left\{\hat{S}_N^h > \rho(h,\lambda_n)\sqrt{n\lambda_n/2}\ \delta^2(n) + c\right\}$, for arbitrary c, is bounded away from 0 and 1, while $e_N^\alpha(h)$ asymptotically coincides with $-\log\alpha_n(h)$. Hence, the h-test having the largest asymptotic value of $e_N^\alpha(h)$ is optimal within class of h-tests. Further, for intermediate alternatives (B.3) $x_N(h) \to \infty$. Therefore, asymptotic analysis of $e_N^\alpha(h)$ is based on the results on the probabilities of large deviations counted under the hypothesis; the range of large deviations depend on the rate of $\sqrt{n\lambda_n}\delta^2(n) \to \infty$. Thus, the "distance" $\delta(n)$ determines the pertaining range of large deviation for the statistic S_N^h; this effect is common for all h-tests, and hence the $\alpha-$IAE of various h-tests differ through the functional $|\rho(h,\lambda_n)|$.

The statement of Ivchenko and Mirakhmedov [11] in our notation reads as follows: Let $\lambda_n \to \lambda \in (0,\infty)$, sequences of alternatives H_{1n} specified by (B.3). Then

$$\frac{e_N^\alpha(h)}{n\lambda_n\delta^4(n)} = \frac{1}{4}\rho^2(h,\lambda_n)(1+o(1)), \tag{B.6}$$

provided either (i) $E|h(\xi)|^{2+\varepsilon} < \infty$, some $\varepsilon > 0$, and $\delta(n) = O(N^{-1/4}\log^{1/4}N)$, or (ii) the Cramér condition (C) is fulfilled.

Note that for the chi-squared test part (i) can be applied only, since χ_N^2 statistic does not satisfy the Cramér condition. On the other hand $|\rho(h,\lambda_n)| = 1$ iff $h(x) = x^2$, *i.e.* for χ_N^2 statistic only. Therefore, from the result of Ivchenko and Mirakhmedov we may conclude that within class of h-tests the chi-squared test is optimal in terms of α-IAE for intermediate alternatives (B.3) with $\delta(n) = O(N^{-1/4}\log^{1/4}N)$, but for more distant alternatives, when $\delta(n)N^{1/4}\log^{-1/4}N \to \infty$, α-IAE of the chi-squared test their result leaves to be open. Another remaining open problem is α-IAE of h-tests for the dense models. Theorem 2.2 allows addressing these problems.

We consider h-tests, where $h(x) = h_d(x) = x^{d+1}$, $d > -1, d \neq 0$, else $h_0(x) = 2x\ln x$. The statistics of these tests coincide with the class of standardized versions of PDS $CR_N(d)$ under the uniformity of null hypothesis.

Theorem B.2. *Let $\lambda_n \to \lambda > 0$, then the equality (B.6), where $h(x) = h_d(x)$, is hold:*

1. *for every sequence of alternatives of the family $\Im_{alt}$ (B.3), if $-1 < d \leq 0$;*
2. *for every sequence of alternatives of a sub-family of the family $\Im_{alt}$ such that $\delta(n) = o\left(N^{-d/2(1+2d)}\right)$, if integer $d > 0$;*
3. *for every sequence of alternatives of a sub-family of the family $\Im_{alt}$ such that $\delta(n) = o\left(\min(N^{-3/16}, N^{-d/2(1+2d)})\right)$, if non- integer $d > 0$.*

Theorem B.3. *Let $\lambda_n \to \infty$, then relation (B.6), where $h(x) = h_d(x)$,$d > -1$, is hold for every sequence of alternatives of a sub-family of the family $\Im_{alt}$ such that $\delta(n) = o\left(\left(n\lambda_n^2\right)^{-1/6}\right)$.*

Proofs. By Theorems 3.1 and 3.2 one has under the conditions of the theorem that $e_N^\alpha(h) = -\ln\Phi(-x_N(h)) + o(1)) = x_N^2(h)/2(1+o(1))$, since $\ln\Phi(-x_N) = -x_N^2/2(1+o(1))$ as $x_N \to \infty$. Now, Theorems B.2 and B.3 follow from (B.5), and simple algebra to determine conditions for the $\delta(n)$ (*i.e.* to determine a sub-family of alternatives), which follows from corresponding conditions on $x_N = x_N(h)$ of Theorems B.2 and B.3, respectively.

The functional $\rho(h, \lambda_n)$ play key role in task of comparison of h-tests in terms of α-IAE, as it is in terms of Pitman AE. Its sense is clarified by the fact that $\rho(h, \lambda_n) = corr_0(S_N^h, \chi_N^2)(1+o(1))$, see Lemma 1 in [11]. We emphasis also that if $\lambda_n \to \infty$, then $\rho(h_d, \lambda_n) = 1 + o(1)$ for all $d > -1$. (But for the CS (1.6) $\rho(h, \lambda_n) = o(1)$ if $\lambda_n \to \infty$, that is way we have restricted ourselves by the class of PDS). As consequence of Theorems B.2 and B.3 we conclude the followings.

1. Within class of h-tests the chi-squared test is optimal in term of $\alpha-$IAE for the intermediate alternatives (B.3) with $\delta\left(n\right) = o\left(\left(n\lambda_n^2\right)^{-1/6}\right)$.
2. The chi-squared test is unique optimal test for the sparse models and a sub-family of intermediate alternatives (B.3) with $\delta(n) = o\left(N^{-1/6}\right)$. In [11] emphasized that for the chi-squared test $e_N^\alpha(h) = o(n\lambda_n\delta^4(n))$ if alternatives (B.3) at distant $\delta(n)N^{1/6}\log^{-1/3}N \to \infty$, and hence the chi-squared test is much inferior to h-tests satisfying the Cramér condition (2.6), in particularly to tests based on Λ_N and T_N^2 statistics.
3. If we consider dense models and a sub-family of intermediate alternatives (B.3) with $\delta(n) = o\left(\left(n\lambda_n^2\right)^{-1/6}\right)$, then h-tests based on PDS $CD_N(d)$ have the same $\alpha-$IAE for all $d > -1$.

Remark B.4. In the upcoming article, in particular for dense models, the following results are proved.

1. For the chi-square test one has $e_N^\alpha(h) = o(n\lambda_n\delta^4(n))$ if either (a) $\sqrt{n} = o(N)$ and a sub-family of intermediate alternative (B.3) such that $\delta(n)\left(n\lambda_n\right)^{1/6}\ln^{-1/3}(N^2/n) \to \infty$, or (b) $N = o(\sqrt{n})$ and a sub-family of intermediate alternative (B.3) such that $\delta(n)\left(n\lambda_n^2\right)^{1/8} \to \infty$.
2. For the log-likelihood ratio test one has $e_N^\alpha(h) = 4^{-1}n\lambda_n\delta^4(n)(1+o(1))$, for every sequences of intermediate alternatives (B.3).

Thus, in situations (a) or (b) the chi-square test is inferior to the log-likelihood test in terms of α-IAE . Next, Inglot [7] pointed out that "comparison of tests using α-IAE can be interpreted only in terms of comparing intermediate slopes of test statistics". This mean that α-IAE of h-tests coincides with asymptotic intermediate efficiency in weak sense due to Inglot [7], but not of comparing the sample sizes as it is in the definition of asymptotic relative efficiency notion. Nevertheless, Theorems B.2 and B.3 together with the theorem on asymptotic normality of statistics S_N^h made possible to study the asymptotic relative efficiency of two h-tests, which is based on the limit of the ratio of sample sizes guaranteeing for both tests the same significance level (may tending to zero) and asymptotically non-degenerate power.

Acknowledgements. The author thanks anonymous reviewer for his valuable constructive comments that lead to an improved version of the paper.

References

[1] N.N. Amosova, Necessity of the Cramer, Linnik and Statulevicius conditions for Large deviation probabilities. *J. Math. Sci.* **109** (2002) 2031–2036.
[2] S.A. Bykov and V.A. Ivanov, The conditions of asymptotic normality of multidimensional randomized decomposable statistics. *Discrete Math. Appl.* **1** (1991) 219–227.
[3] N. Cressie and T.R.C. Read, Multinomial goodness-of-Fit tests. *J. R. Statist. Soc. Ser. B* **46** (1984) 440–464.
[4] N. Cressie and T.R.C. Read, Pearson's X^2 and the Log-likelihood Ratio Statistic G^2: a comparative review. *Int. Statist. Rev.* **57** (1989) 19–43.
[5] I. Csiszár, Information measures: a critical survey, in *Transactions of the 7th Prague Conference on Information Theory, Statistical Decision Functions, Random Processes* (1997) 73–86.
[6] L. Holst, Asymptotic normality and efficiency for certain goodness-of-fit tests. *Biometrica* **59** (1072) 137–145.
[7] T. Inglot, Generalized intermediate efficiency of goodness-of-fit tests. *Math. Methods Statist.* **8** (1999) 487–509.
[8] V.A. Ivanov, On a generalization of separable statistics. *Theory Probab. Appl.* **30** (1985) 835–840.
[9] G.I. Ivchenko and Y.I. Medvedev, Decomposable statistics and a hypothesis testing. The small sample case. *Theor. Probabl. Appl.* **23** (1978) 796–806.
[10] G.I. Ivchenko and S.A. Mirakhmedov, On limit theorems for the decomposable statistics and efficiency of the corresponding statistical tests. *Discrete Math. Appl.* **2** (1992) 547–562.
[11] G.I. Ivchenko and S.A. Mirakhmedov, Large deviations and intermediate efficiency of the decomposable statistics in multinomial scheme. *Math. Methods Statist.* **4** (1995) 294–311.
[12] W.C.M. Kallenberg, On moderate and large deviations in multinomial distributions. *Ann. Statist.* **13** (1985) 1554–1580.
[13] T.V. Khmaladze, Martingale limit theorems for divisible statistics. *Theor. Probabl. Appl.* **28** (1984) 530–548.
[14] T.V. Khmaladze, Convergence properties in certain occupancy problems including the Karlin–Rauoult law. *J. Appl. Probab.* **48** (2011) 1095–1113.
[15] K.J. Koehler and K. Larntz, An empirical investigation of goodness-of-fit statistics for spars multinomial. *J. Am. Stat. Assoc., Sec. Theory Methods* **75** (1980) 336–344.
[16] V.F. Kolchin, B.A. Sevostyanov and V.P. Chistyakov, Random Allocations. Halstead Press (Wiley), New York (1978).
[17] P. L'ecuyer, R. Simard and S. Wegenkittl, Sparse serial tests of uniformity for random number generators. *SIAM J. Sci. Comput.* **24** (2002) 652–668.
[18] H.B. Mann and A. Wald, On the choice of the number of intervals in the application of the chi-square test. *Ann. Math. Statist.* **13** (1942) 306–317.
[19] D. Morales, L. Pardo and I. Vajda, Asymptotic divergence of estimates of discrete distributions. *J. Stat. Plann. Inference* **48** (1995) 347–369.
[20] S.A. Mirakhmedov, Approximation of the distribution of multi-dimensional randomized divisible statistics by normal distribution (multinomial scheme). *Theory Probabl. Appl.* **32** (1987) 696–707.
[21] S.A. Mirakhmedov Randomized decomposable statistics in the scheme of independent allocation of particles into boxes. *Discrete Math. Appl.* **2** (1992) 91–108.
[22] S.M. Mirakhmedov,[1] Asymptotic normality associated with generalized occupancy problems. *Stat. Probabl. Lett.* **77** (2007) 1549–1558.
[23] S.M. Mirakhmedov, S.R. Jammalamadaka and M. Ibrahim, On Edgeworth expansion in generalized urn models. *J. Theor. Probab.* **27** (2014) 725–753.
[24] S.A. Mrakhmedov, Estimates of closeness of the distribution of a randomized separable statistic to the normal law in a multinomial scheme. *Theory Probabl. Appl.* **30** (1986) 192–196.
[25] D.S. Moore, Tests of Chi-squared type, in Goodness-of-Fit Techniques, edited by R.B. D'Agostino and M.A. Stephens. Marcel Dekker, New York (1986) 63–95.
[26] C. Morris, Central limit theorems of multinomial sums. *Ann. Statist.* **3** (1975) 165–188.
[27] L. Pardo, Statistical Inference based on Divergence Measures. Chapman & Hall–CRC, Boca Raton (2006).
[28] M. Pietrzak, G.A. Rempala, M. Seweryn and J. Wesołowski, Limit theorems for empirical Rényi entropy and divergence with applications to molecular diversity analysis. *TEST* **25** (2016) 654–673.
[29] M.P. Quine and J. Robinson, Normal approximations to sums of scores based on occupancy numbers. *Ann. Probab.* **12** (1984) 794–804.
[30] M.P. Quine and J. Robinson, Efficiencies of chi-square and likelihood ratio goodness-of-fit tests. *Ann. Statist.* **13** (1985) 727–742.
[31] T.R.C. Read and N.A.C. Cressie, Goodness-of-Fit Statistics for Discrete Multivariate Data. Springer, New York (1988).
[32] G.A. Rempala and J. Wesolowski, Double asymptotics for the chi-square statistic. *Stat. Probabl. Lett.* **119** (2016) 317–325.
[33] J. Riordan, Moment recurrence relations for binomial, Poisson and hypergeometric frequency distributions. *Ann. Math. Stat.* **8** (1937) 103–111.

[1]S.M. Mirakhmedov is the former Mirakhmedov S.A.

[34] A.F. Ronzgin, A theorem on the probabilities of large deviations for the decomposable statistics and its application in statistics. *Mate. Notes* **36** (1984) 603–615.

[35] L. Saulis and V. Statulevicius, Limit Theorems for Large Deviations., Vol. 232. Kluwer Academic Publishers, Dordrecht, Boston, London (1991).

[36] S.Kh. Siragdinov, S.A. Mirakhmedov and S.A. Ismatullaev, Large deviation probabilities for decomposable statistics in a polynomial scheme. *Theory Probab. Appl.* **34** (1989) 706–719.