UZBEKISTAN ACADEMY OF SCIENCES
V.I.ROMANOVSKIY INSTITUTE OF MATHEMATICS

UZBEK MATHEMATICAL JOURNAL

Journal was founded in 1957. Until 1991 it was named by "Izv. Akad. Nauk UzSSR, Ser. Fiz.-Mat. Nauk". Since 1991 it is known as "Uzbek Mathematical Journal". It has 4 issues annually.

Volume 65 **Issue 4** **2022**

Uzbek Mathematical Journal is abstracting and indexing by

MathSciNet **Zentralblatt** **Math VINITI**

Starting from 2018 all papers published in Uzbek Mathematical Journal you can find in **EBSCO** and **CrossRef**.

TASHKENT - 2022

Uzbek Mathematical Journal
2022, Volume 65, Issue 4, pp.128-136
DOI: 10.29229/uzmj.2022-4-11

Remarks on the Pitman's Efficiency of Goodness of Fit Tests Based on Grouped Data

Mirakhmedov Sh.M., Bozarov U.A.

Abstract. We consider the problem of testing the goodness of fit of a continuous distribution to a set of observations grouped into equal probability intervals. We are interested in a class of tests based on symmetric statistics, which are defined as the sum of a function of interval-frequencies. The effect of changing of the number of intervals on the Pitman efficiency for a family of contamination alternatives is studied. It is assumed that the number of intervals increases asymptotically as the number of observations grows.

Keywords: Asymptotic efficiency, chi-square statistic, goodness of fit, multinomial distribution, power divergence statistics.

MSC (2010): 62G10, 62G20

1 Introduction

Consider the classical problem of testing the goodness of fit of a sample of size n has come from an absolutely continuous distribution F_0. Through probability integral transformation $z \to F_0(z)$ this problem reduces to testing for uniformity over $[0,1]$, and hence without loss of generality we will assume that F_0 is Uniform $[0,1]$ distribution. So, we consider the null hypothesis $H_0 : f(x) = 1,\ x \in [0,1]$, versus a sequence of contamination alternatives:

$$H_1 : f(x) = 1 + \delta(x) g_n(x), \tag{1.1}$$

where $\delta(n) \to 0$, as $n \to \infty$, $\int_0^1 g_n(x)dx = 0$, $0 < \inf_n ||g_n||_2 \le \sup_n ||g_n||_\infty < \infty$ $||.||_\infty$ denotes the supremum norm, $||.||_2$ is the $L_2[0,1]$ norm.

Remark 1.1. These alternatives converge to H_0 with a rate determined by $\delta(n)$, whereas function $g_n(x)$ defines the path along which one goes from alternative to hypothesis. For the asymptotic efficiency of h-tests the actual direction of approach to the hypothesis is immaterial, but the rate of convergence plays a role. Therefore, without loss of generality and to keep future notation simple we shall assume that $||g_n||_2 = 1$.

Assume that a set of n observations are grouped into N equal probability intervals, and we consider tests based on the symmetric statistics of the form

$$S^h_{n,N} = \sum_{l=1}^{N} h(\eta_l), \tag{1.2}$$

where h is a nonlinear real-valued function defined on the non-negative axis, $\eta_1, \ldots, \eta_N$ is the random vector of frequencies of intervals. Assume that the large values of $S^h_{n,N}$ reject the hypothesis. The test based on $S^h_{n,N}$ is called h-test for brevity.

An important flexible sub-class of statistics (1.2) is the class of Power Divergence Statistics (PDS) of Creese and Read [1], where $h(x) = h_d(x)$ where $\lambda_{n,N} = n/N$,

$$h_d(x) = 2(d(d+1))^{-1}x[(x/\lambda_{n,N})^d - 1], \;\; d > -1, d \neq 0 \text{ else}$$

$$h_0(x) = 2xlog(x/\lambda_{n,N}). \tag{1.3}$$

We emphasis the following important variants of statistics (1.2): the PDSs

$$\chi^2_N = \lambda^{-1}_{n,N} \sum_{m=1}^{N} (\eta_m - \lambda_{n,N})^2, \;\; \Lambda_N = 2 \sum_{m=1}^{N} \eta_m log(\eta_m/\lambda_{n,N}) \text{ and}$$

$$T^2_N = 4 \sum_{m=1}^{N} (\sqrt{\eta_m} - \sqrt{\lambda_{n,N}})^2 \tag{1.4}$$

which are the chi-square statistic, the log-likelihood ratio statistic and the Freeman-Tukey statistic, respectively; and the count statistics (CS)

$$\mu_r = \sum_{m=1}^{N} I\{\eta_m = r\}, r \geq 0, \;\; w_l = \sum_{m=1}^{N} I\{\eta_m \geq l\}, l \geq 1, \;\text{ and}$$

$$C_n = \sum_{m=1}^{N} (\eta_m - 1) I\{\eta_m > 1\}, \tag{1.5}$$

where $I\{\cdot\}$ denotes the indicator function, which are respectively, the number of intervals consisting exactly r and at least l observations, and the number of collisions (that is, the number of observations that we observe in intervals already containing observations). These CS have been considered in the literature in various contexts; see, for instance, L'ecuyer et al [7].

Our objective in this paper is to compare the performance of two h-tests in term of relative Pitman efficiency, the approach, that is probably the most widely used in statistical inference. In the method of grouped data the problem of choice of number of intervals is still of interest. There have been many attempts to resolve the problem. Among many others we refer to Quine and Robinson [9], who investigated the effect of different choice of number of classes in Pitmann's and Bahadur's efficiencies of the chi-square and log-likelihood ratio tests. In this article we extend the results on Pitman efficiency of Quine and Robinson (1985) to the class of h-tests.

2 Results

Our results based on the central limit theorem for statistics (1.3). In turn proof of such limit result is based on well-known fact that the distribution of random vector of frequencies $(\eta_1, \ldots, \eta_N)$ can be represented as the joint conditional distribution of independent random variables $(\xi_1, \ldots, \xi_N)$ given $\xi_1 + \ldots + \xi_N = n$, where ξ_m is a Poisson random variable with parameter np_m,

$$p_m = N^{-1}(1 + \delta(n)\Delta_{n,m}),\ \Delta_{n,m} = N \int_{(m-1)/N}^{m/N} g_n(x)dx, m = 1, \ldots,\ N \tag{2.1}$$

where $\max_m |\Delta_{n,m}| \le c < \infty,\ \Delta_{n,1}+\ldots+\Delta_{n,N} = 0$ and $N^{-1}(\Delta^2_{n,1}+\ldots+\Delta^2_{n,N}) = 1$.

In what follows we adopt the following notation: $\xi \sim Poi(\lambda)$ stands for "a r.v. ξ has Poisson distribution with parameter $\lambda > 0$"; $\xi, \xi_1, \ldots, \xi_N$ are independent r.v.s such that $\xi \sim Poi(\lambda_{n,N})$ and $\xi_m \sim Poi(np_m)$, where $\lambda_{n,N} = n/N$ is the average of observations per intervals; $\Phi(u)$ denotes a standard normal distribution function; c_j is a positive constant, may not the same in each its occurrence; all asymptotic statements are considered as $n \to \infty$. We are concerned with the case where $N = N(n) \to \infty$ as $n \to \infty$.

Let $P_i, E_i S^h_{n,N}$ and $Var_i S^h_{n,N}$ be the probability, expectation and variance of $S^h_{n,N}$, respectively, counted under $H_i, i = 0, 1$. Set

$g(\xi) = h(\xi) - Eh(\xi) - r_n(\xi - \lambda_{n,N}), r_n = \lambda^{-1}_{n,N} cov(h(\xi), \xi)$

$\sigma^2(h) = Var g(\xi) = Var h(\xi)(1 - corr^2(h(\xi), \xi))$.

The following proposal is the basis for studying the Pitman asymptotic efficiency of h-tests.

Proposition 2.1.

Assume

$$L_N(h) = E|g(\xi)|^3 \sigma^3(h) sqrt N \to 0 \tag{2.2}$$

and sequences of alternatives H_{1n} as in (1.1). Then

$$P_i\{S^h_{n,N} < u\sigma_{i,n}(h)\sqrt{N} + NA_{i,n}(h)\} = \Phi(u) + o(1),\ i = 0, 1, \ldots \tag{2.3}$$

and if additionally $\delta(n) = o(\lambda^{-1/2}_{n,N})$ then

$$x_{n,N}(h) := \sqrt{N}(A_{1,n}(h) - A_{0,n}(h))/\sigma_{0,n}(h) =$$

$$= \sqrt{n\lambda_{n,N}/2}\delta^2(n)\rho(S^h_{n,N}, \lambda_{n,N})(1 + j(1)), \tag{2.4}$$

where

$$A_{i,n}(h) = N^{-1} \sum_{m=1}^{N} E_i h(\xi_m),\ \sigma^2_{i,n}(h) = N^{-1} \sum_{m=1}^{N} Var_i g(\xi_m),$$

$$\sigma^2_{1,n}(h) = \sigma^2_{0,n}(h)(1 + o(1)) = \sigma^2(h)(1 + o(1)),$$

$$\rho(S^h_{n,N}, \lambda_{n,N}) = corr(h(\xi) - r_n\xi, \xi^2 - (2\lambda_n, N + 1)\xi). \tag{2.5}$$

The asymptotical normality result (2.3) follows from Theorem 2 of Mirakhmedov [8]. Equality (2.4) can be derived by applying the Maclaurin expansion of $(1 + x)^k e^{\lambda_{n,N}x}$ in

$A_{1,n}(h) = \frac{1}{N}\sum_{m=1}^{N}\sum_{k=0}^{\infty}\pi_k(\lambda_{n,N})(1 + \varepsilon_m)^k e^{-\lambda_{n,N}\varepsilon_m}$,

where $\pi_k(\lambda) = \lambda^k e^{-\lambda}/k!$, $\varepsilon_m = \delta_m(n)\Delta_n, m$.

In particular, from Proposition 2.1 and well-known theorem on convergence of moments it follows that $A_{i,n}(h)$ and $\sigma^2_{i,n}(h)$ are the asymptotic value of $N^{-1}E_iS^h_{n,N}$ and $N^{-1}Var_iS^h_{n,N}$, respectively.

Remark 2.1. In fact, by the grouping data the original problem testing uniformity against alternatives (1.1) is reduced to the problem of testing of uniformity of a multinomial distribution against sequences of alternatives (2.1). In context of this we emphasis that the condition (2.2) is fulfilled for the sparse multinomial distributions, i.e. $\lambda_{n,N} \to \lambda \in (0, \infty)$ if $E|h(\xi)|^3 < \infty$, and for the very sparse multinomial distributions, i.e., when $\lambda_{n,N} \to 0$ if $\Delta^2h(0) \neq 0$, where $\Delta h(x) = h(x + 1) - h(x)$. For instance, arbitrary PDS and the CS c_n, μ_r, $r = 1, 2, \ldots$ and w_l, $l = 1, 2$ satisfy this condition. But for the dense multinomial distributions, i.e., when $\lambda_{n,N} \to \infty$, the condition (2.2) may impose an additional condition to $\lambda_{n,N}$. For instance, (2.2) is fulfilled in this case for every PDS, while, for example, for CS c_n, μ_r, $r \geq 0$ and c_n the (2.2) imposes condition $\lambda_{n,N} - \ln N - \ln\ln N \to -\infty$ and $\lambda_{n,N} - \ln N \to -\infty$, respectively.

Next, if $\lambda_{n,N} \to 0$ and statistics such that $\Delta^2h(0) \neq 0$ then

$$\rho(S^h_{n,N}, \lambda_{n,N}) = 1 - \frac{\lambda_{n,N}}{6}\left(\frac{\Delta^3h(0)}{\Delta^2h(0)}\right)^2 + O(\lambda^2_{n,N}), \tag{2.6}$$

and if $\lambda_{n,N} \to \infty$ then for PDS with parameter $d > -1$, see(1.3),

$$\rho(S^h_{n,N}, \lambda_{n,N}) = 1 - \frac{(d - 1)^2}{6\lambda_{n,N}} + O(\lambda^{-2}_{n,N}). \tag{2.7}$$

But for the CS (1.5) $\rho(S^h_{n,N}, \lambda_{n,N}) = o(1)$ if $\lambda_{n,N} \to \infty$.

Remark 2.2. Functional $\rho(S^h_{n,N}, \lambda_{n,N})$ plays an important role in determining the asymptotic properties of h-tests satisfying (2.1). Its sense is clarified by the fact that if (2.2) is fulfilled, then

$$\rho(S^h_{n,N}, \lambda_{n,N}) = corr_0(S^h_{n,N}, \chi^2_N)(1 + o(1))$$

See, Lemma of Ivchenko and Mirakhmedov [6]. In what follows we shall consider statistics $S^h_{n,N}$ which satisfy Proposition 2.1 and $|\rho(S^h_{n,N}, \lambda_{n,N})|$ is bounded away from zero.

Let's turn to the problem of comparison of two tests in term of Pitman's asymptotic relative efficiency. Let $\{T_{1n}\}$ and $\{T_{2n}\}$ be two sequences of test statistics for testing a hypothesis H_0. Assume that the sequence of alternatives H_{1n} converges to H_0 at such a rate that the power of the test of size $\alpha > 0$ using the statistic T_{1n} based on sample size n tends to $\beta \in (\alpha, 1)$ as $n \to \infty$. Let n' be the sample size needed for the power of the size α test based on statistics $T_{2n'}$ under H_{1n} also tend to β as $n' \to \infty$. Assume that $lim(n'/n) = e$ exist and do not depend on particular choice of n'. Then this limit is the Pitman efficiency of T_{1n} wrt T_{2n}, viz.,

$$PE(T_{1n}, T_{2n}) = lim(n'/n).$$

Define efficacy of the test based on statistic T_n as $e(T_n) = \mu_r^2/\sigma_T^2$, where μ_T and σ_r^2 are the mean and variance of the limiting normal distribution under the alternatives when the test statistic T_n have been normalized to have limiting standard normal distribution under the hypothesis. Then due to Fraser [2]

$$PE(T_{1n}, T_{2n}) = e(T_{1n})/e(T_{2n}) \tag{2.8}$$

Assume that the statistics $S^h_{n,N}$ and $S^f_{n,N}$ satisfy condition (2.2). Let's consider Pitman efficiency of h-test wrt f-test. Let $\alpha_{n,N}(h)$ and $\beta_{n,N}(h)$ denote the size and the power of h-test, respectively.

$$\alpha_{n,N}(h) = P_0\{S^h_{n,N} > u_\alpha \sigma_{0,n}(h)\sqrt{N} + NA_{0,n}(h)\} =$$

$$= \Phi(-u_\alpha) + o(1) = \alpha + o(1), \tag{2.9}$$

where $\alpha > 0$, $u_\alpha = \Phi^{-1}(1-\alpha)$. Next, if alternatives (1.1) is such that

$$\delta(n) = c_0(n\lambda_{n,N})^{-1/4}, \tag{2.10}$$

then

$$\beta_{n,N}(h) = P_1\{S^h_{n,N} > u_\alpha \sigma_{0,n}(h)\sqrt{N} + NA_{0,n}(h)\} =$$

$$= P_1\{S^h_{n,N} > \frac{\sigma_{0,n}(h)}{\sigma_{1,n}(h)}(u_\alpha + x_{n,N}(h)) =$$

$$= \Phi(\sqrt{n\lambda_{n,N}/2}\delta^2(n)|\rho(S^h_{n,N}, \lambda_{n,N})| - u_\alpha)(1 + o(1)) =$$

$$= \Phi(c_0|\rho(S^h_{n,N}, \lambda_{n,N})|/\sqrt{2} - u_\alpha)(1 + o(h)) = \beta \in (\alpha, 1) \tag{2.11}$$

Gvanceladze and Chibisov [3] have shown for a similar scheme to ours that the power of the tests symmetrically depending on interval-frequencies tends to the significance level as $n \to \infty$ whenever $N \to \infty$ for $\delta(n) = n^{-1/2}$. A refinement of their result follows from the second line of equation (2.11), which implies that the power of the h-tests tend to the significance level as $n \to \infty$ whenever $N \to \infty$ for $\delta_n = o((n^2/N)^{-1/4})$.

Further, from Proposition 2.1 one can easily observe that the efficacy of the h-test under alternatives H_{1n} where $\delta(n)$ is defined as in (2.10) is equal to the limiting value of $x_{n,N}(h)$, see (2.4), that is $e(S^h_{n,N}) = c_0^4 lim\rho^2(S^h_{n,n}, \lambda_{n,N})/2$. Thus, this fact together with (2.8) gives the following.

Theorem 2.1. Let for statistics $S^h_{n,N}$ and $S^f_{n,N}$, which based on the same number of intervals, the condition (2.2) be fulfilled and alternatives H_{1n} (1.1) converge to H_0 at the rate defined as (2.10). Then

$$PE(S^h_{n,N}, S^f_{n,N}) = lim \frac{\rho^2(S^h_{n,N}, \lambda_{n,N})}{\rho^2(S^f_{n,N}, \lambda_{n,N})}$$

It follows from (2.5) that $|\rho(\chi^2_{n,N}, \lambda_{n,N})| = 1$. Hence

$$PE(S^h_{n,N}, \chi^2_{n,N}) = lim\rho^2(S^h_{n,N}, \lambda_{n,N}) \leq 1$$

That is, Pitman efficiency of arbitrary h-test satisfying (2.2) wrt chi-square test less than or equal to 1. Specifically, $PE(S^h_{n,N}, \chi^2_{n,N}) < 1$, if $\lambda_{n,N}$ is bounded away from zero and infinity. Nevertheless, computations shows that even in this situation maximum Pitman efficiency of the PDS with parameter $d \in [0, 2]$ are very close to 1 (see below Table 1). This fact extends corresponding results of Holst [4] and Ivchenko and Medvedev [5]. Further, consider class of PDS, i.e., h_d-tests, where h_d is defined as (1.3), then $PE(S^{h_d}_{n,N}, \chi^2_{n,N}) = 1$ if $\lambda_{n,N} \to \infty$, since (2.7). This is an extension of the statement (1.3) of Quine and Robinson [9] to the class of PDS, where such a fact is presented for the case $S^{h_d}_{n,N} = \Lambda_N$, log-likelihood ratio statistic. At last, for the class of h-tests such that $\Delta^2 h(0) \neq 0$ we have $PE(S^h_{n,N}, \chi^2_{n,N}) = 1$ if $\Delta^2 h(0) \neq 0$ and $\lambda_{n,N} \to 0$, since (2.6). This case is of interest for the testing of uniformity of a multinomial distribution against alternatives (2.1) satisfying (2.10), see, for instance, L'ecuyer et al [7] .

Consider now again h-test of size α, but based on statistic $S^h_{n,N'}$, where N' is another number of intervals. In this case we assume that the number of intervals $N = N(x)$, taken as a function of the continuous variable x, is regularly varying with index $q \in (0, 2)$, i.e. $N(ax)/N(x) \to a^q$ as $x \to \infty$, for all $a > 0$.

Theorem 2.2. Let $L_N(h) \to 0$ (see (2.2)) then

$$PE(S^h_{n,N}, S^h_{n,N'}) = c^{1/(2-q)} \tag{2.12}$$

if $N'(n)/N(n) \to c \in (0, \infty)$, and

$$PE(S^h_{n,N}, S^h_{n,N'}) = \infty \tag{2.13}$$

if $N'(n)/N(n) \to \infty$.

Proof. According to Pitman's approach $\delta(n)$, the rate of convergence of alternatives H_{1n} to the hypothesis, must be chosen so that the power for h-test of size $\alpha > 0$ has a limit $\beta \in (\alpha, 1)$. So, for statistic $S^h_{n,N}$ the relations (2.9), (2.10) and (2.11) are valid. Let n' is a sample size such that the power of the size α test based on statistic $S^h_{n',N'}$ under H_{1n} (1.1) also tends to β as $n' \to \infty$. That is for the statistic $S^h_{n',N'}$ the equations (2.9), (2.10) and (2.11) still hold when n and N are replaced by $n'$and N'. It is clear that $n' = n'(n)$. Due to (2.10) in order for the test based on $S^h_{n',N'}$ to have the same asymptotic power as in (2.11) under H_{1n} (1.1) satisfying (2.10) we must choose n' so that

$$N(n)/n^2 \sim N'(n')/n'^2 \tag{2.14}$$

since $\rho^2(S^h_{n,N}, \lambda_{n>N})$ and $\rho^2(S^h_{n',N'}, \lambda_{n'>N'})$ bounded away from zero, and, on the other hand, the rate of convergence of alternatives H_{1n} remaining unchanged as in (2.10).

We will now use arguments similar to those of Quine and Robinson [9]. Note that if $N'(n)/N(n) \to c \in (0, \infty)$ then $N'(n')/N(n') \to c$ also. By (2.14) and that $N(x)$ is regularly varying function with index q we have

$$N(n)/n^2 \sim cN(an')/n'^2 \sim N(an')/(an')^2, \text{ where } a = c^{-1/(2-q)}. \tag{2.15}$$

The function $R(x) = N(x)/x^2$ is regularly varying with index $q - 2 < 0$, therefore $R(r_x x)/R(x) \to r^{q-2}$ as $x \to \infty$, if $r_x \to r \in [0, \infty]$, where $0 < r_x < \infty$. Let a sequence of n'/n have a sub-sequence for which $n'/n \to b \in [0, \infty]$, then for this subsequence

$$R(an')/R(n) = R((an'/n)n)/R(n) \to (ab)^{q-2}. \tag{2.16}$$

Since (2.15) $(ab)^{q-2} = 1$, that is $b = c^{1/(2-q)}$, do not depend on particular sub-sequence. Thus for whole sequence $n'/n \to c^{1/(2-q)}$, and hence (2.12) follows. Let now $N'(n)/N(n) \to \infty$. Then for arbitrary large $C > 0$ there exists $n' = n'C$ such that $N'(n') \geq CN(n')$, therefore from (2.11) we obtain $N(n)/n^2 \sim N'(n')/n'^2 > CN(n')/n'^2 \sim N(An')/(An')^2$, where $A = C^{-1/(2-q)}$. Thus there exists a subsequence of n'/n for which (2.16) hold with a replaced by A. But now $(Ab)^{q-2} > 1$ and hence $b > A^{-1} = C^{1/(2-q)}$. This mean that asymptotically $n' > C^{1/(2-q)}n$, giving (2.13). Proof of Theorem 2.2 completed.

Theorem 2.2 extends statements (1.1) and (1.2) of Quine and Robinson [9], where equalities (2.12) and (2.13) were derived for χ^2_N and Λ_N statistics (see (1.3)) to the class of h-tests satisfying condition (2.2). So, all conclusions of Quine and Robinson ([9] on effect of changing the number of intervals to Pitman efficiency of chi-square test still hold for the class of h-tests. In particular, if we deal with the contamination alternatives the number of intervals should not be too large.

3 Some computational results

It is seen in fact that the Pitman efficiency of h-test depend on the asymptotic behavior of the parameter $\lambda_{n,N}$ and $|\rho(S^h_{n,N}, \lambda_{n,N})|$, the asymptotical correlation coefficient under the hypothesis between the test statistic $S^h_{n,N}$ and the chi-square statistic; so a statistic that is more correlated with the chi-square statistic should be considered preferable. For the PDS in the following Table 1 the values of $|\rho(h_d, \lambda)|$ are presented for various λ and $d > -1$.

Table 1. The value of $|\rho(S^{h_d}_{n,N}, \lambda)|$ for different d and λ.

	λ									
d	0.05	0.1	0.5	1.0	1.5	2.0	3.0	10	20	50
-2/3	0.9933	0.9838	0.9400	0.8768	0.8314	0.7811	0.7266	0.9257	0.9740	0.9900
-1/2	0.9942	0.9838	0.9402	0.8909	0.8545	0.8321	0.8001	0.9480	0.9803	0.9920
-1/3	0.9950	0.9839	0.9620	0.9192	0.89891	0.8743	0.8573	0.9615	0.9834	0.9940
0	0.9970	0.9940	0.9720	0.9525	0.9400	0.9350	0.9369	0.9793	0.9897	0.9960
1/3	0.9983	0.9840	0.9845	0.9758	0.9699	0.9714	0.9797	0.9928	0.9961	0.9980
1/2	0.9989	0.9979	0.9900	0.9898	0.9815	0.9791	0.9879	0.9972	0.9993	0.9985
2/3	0.9999	0.9924	0.9901	0.9900	0.9930	0.9945	0.9961	0.9977	0.9996	0.9990
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3/2	0.9984	0.9844	0.9900	0.9901	0.9930	0.9925	0.9879	0.9977	0.9997	0.9989
2	0.9917	0.9843	0.9618	0.9617	0.9583	0.9632	0.9716	0.9883	0.9929	0.9960
5/2	0.9759	0.9519	0.9220	0.9192	0.9237	0.9323	0.9389	0.9704	0.9835	0.9920
3	0.9449	0.9391	0.8631	0.8627	0.8876	0.8933	0.8981	0.9526	0.9708	0.9880
4	0.7917	0.8049	0.7443	0.7495	0.7736	0.7921	0.8164	0.8989	0.9392	0.9720
5	0.6323	0.6708	0.6047	0.6225	0.6582	0.6741	0.7103	0.8363	0.9012	0.9520

Table shows that the PDS with $d \leq 5/2$ are preferable than that of $d > 5/2$ for all range of λ. While this property of PDS more pronounced for the very sparse and dense models. It is surprise that for the moderate λ the PDS with parameter $d \in [1/3, 2]$ appears to be asymptotically more correlated with chi-square statistic than the log-likelihood ratio statistic, where $d = 0$. But log-likelihood ratio statistic exhibit high limiting correlation with chi-square-statistic than the PDS with d<0, i.e. satisfying Crame'r condition, $0.9335 \leq \rho(h_0, \lambda) \leq 1$ and $\arg\min \rho(h_0, \lambda) = 2.3750$. The PDS $CR_N(2/3)$ exhibit highest limiting correlation with chi-square-statistic for all range of $\lambda : 0.9900 \leq \rho(h_2/3, \lambda) \leq 1$. This confirms recommendation of Cressie and Read [1].

References

1. Cressie N.A.C. and Read T.R. C. (1984). Multinomial goodness-of-fit tests, J. Roy. Statist. Soc. Ser B. 46, 440-464.

2. Fraser, D.A.S. (1957). Nonparametric Methods in Statistics. John Wiley. New York.

3. Gvanceladze L.G. and Chibisov D.M. (1979). On tests of fit based on grouped data. In Contribution to Statistics, J.Hajek Mamorial Valume. J.Jurechkova. ed. 79-89. Academia, Prague.
4. Holst L. (1972). Asymptotic normality and efficiency for certain goodness-of-fit tests. Biometrica, 59, p.137-145.
5. Ivchenko G.I. and Medvedev Y.I.(1978). Decomposable statistics and verifying of tests. Small sample case. Theory of Probability Appl., 23, 796-806.
6. Ivchenko G.I. and Mirakhmedov S.A. (1995). Large deviations and intermediate efficiency of the decomposable statistics in multinomial scheme. Math. Methods in Statist., 4, p.294-311.
7. L'ecuyer P., Simard R. and Wegenkittl S. (2002). Sparse serial tests of uniformity for random number generators. SIAM J. Sci. Comput. 24, 2.
8. Mirakhmedov S.A.(1992), Randomized decomposable statistics in the scheme of independent allocating particles into boxes. Discrete Math. Appl. 2, p.91-108.
9. Quine M.P. and Robinson J. (1985). Efficiencies of chi-square and likelihood ratio goodness-of-fit tests, The Annals of Statistics. 13.

Sherzod M. Mirakhmedov
V.I.Romanovskiy Institute of Mathematics, Uzbekistan Academy of Sciences, 9, University str., Tashkent 100174, Uzbekistan, e-mail: shmirakhmedov@yahoo.com
Ulugbek A. Bozarov
Dep. Theory of probability and Mathematical Statistics. National University of Uzbekistan. Tashkent-l00174. e-mail: Bazarov.u.a@gmail.com

Contents