Goodness-of-fit tests for the uniformity based on Ahsanullah characterization

M. S. Karakulov, Ya. Yu. Nikitin [1], K. Yu. Volkova [2]

The uniform distribution is one of the widely used distribution in statistical modeling, computer science. Moreover, goodness-of-fit testing problem for a given distribution can usually be adjust to test uniformity on $[0, 1]$.

The first test of the uniformity based on a characterization was proposed by Hashimoto and Shirahata in [2]. We use the different way of test constructing based on Ahsanullah characterization [1] of the uniform law: *Let* $X_1, \dots, X_k, k \geq 2$ *be i.i.d. observations having an absolutely continuous (with respect to Lebesgue measure) d.f.* F *concentrated on* $[0, 1]$. *Then the sample has the uniform law* $U[0, 1]$ *if and only if the ratio of order statistics*

$$X_{1,k}/X_{2,k} \sim U[0, 1].$$

Let $X_1, \dots, X_n$ be i.i.d. observations with the continuous d.f. F. Using above characterization we construct $U-$empirical d.f. related to the characterization

$$G_n(t) = \binom{n}{k}^{-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbf{1}\{X_{1,\{i_1,\dots,i_k\}}/X_{2,\{i_1,\dots,i_k\}} < t\}, \quad t \in [0, 1].$$

In order to test the hypothesis H_0 that $F(x) = x$, $x \in [0, 1]$ we propose two statistics:

$$I_n = \int_0^1 (t - G_n(t))dt \quad \text{and} \quad D_n := \sup_{0 \leq t \leq 1} |t - G_n(t)|.$$

The statistic I_n is asymptotically equivalent to the non-degenerate U-statistic, therefore it is asymptotically normal. The statistic D_n has the non-normal limiting distribution, hence we use the Bahadur approach as a method of calculation of the asymptotic efficiency, while the classical Pitman approach to the efficiency is not applicable. However, it is known that the local Bahadur efficiency and the limiting Pitman efficiency usually coincide, see [4].

We find the logarithmic large deviation asymptotic of both sequences of statistics under H_0. This allows us to calculate their local Bahadur efficiency under numerous local alternatives concentrated on $[0, 1]$. It is seen that the integral statistic I_n more efficient than the Kolmogorov statistic D_n as usually happens in goodness-of-fit testing, see [3].

[1] Saint-Petersburg State University, Russia, E-mail: y.nikitin@spbu.ru. The research was supported by the grant RFBR 16-01-00258.

[2] Saint-Petersburg State University, Russia, E-mail: k.volkova@spbu.ru

However, the local efficiencies for both sequences of statistics are reasonably high. For instance, when we construct statistics under the characterization with fix $k = 2$ then for the so-called "parabolic" alternative with density $f(x, \theta) = 1 - \theta(x - 1)(x - \frac{1}{3}), \theta > 0$ the first statistic has the efficiency 0.957 while the second has the efficiency 0.877. So our tests can be recommended for testing. Finally we describe the conditions of local asymptotic optimality (most favorable alternatives) for our statistics.

References

[1] Ahsanullah M. *On characterizations of the uniform distribution based on functions of order statistics*// Aligarh J. Statist., 1989, v. 9, p. 1–6.

[2] Hashimoto T., Shirahata S. *A goodness-of-fit test based on a characterization of uniform distribution*// J. Japan Stat. Soc., 1993, v. 23, N 2, p. 123–130.

[3] Nikitin Y. *Asymptotic efficiency of nonparametric tests*// Cambridge University Press, NY, 1995.

[4] Wieand H. S. *A condition under which the Pitman and Bahadur approaches to efficiency coincide*// The Ann. of Math. Stat., 1976, v. 4, p. 1003–1011.