JOURNAL OF APPLIED STATISTICS
2020, VOL. 47, NOS. 13–15, 2610–2622
https://doi.org/10.1080/02664763.2020.1761952



# Goodness-of-fit tests for the logistic location family

Ya. Yu. Nikitin[a,b] and I. A. Ragozin[a,b]

[a]Department of Mathematics and Mechanics, Saint-Petersburg State University, Saint-Petersburg, Russia; [b]Department of Mathematics and Business Informatics, National Research University – Higher School of Economics, Saint-Petersburg, Russia

**ABSTRACT**
We construct two *U*-empirical tests for the logistic location family which are based on appropriate characterization of this family using independent exponential shifts. We study the limiting distributions and local Bahadur efficiency of corresponding test statistics under close alternatives. It turns out that the present tests are considerably more efficient than the recently proposed similar tests based on another characterization. The efficiency calculations are accompanied by the simulation of power for new tests together with the previous ones.Both efficiency and power turn out to be very high. Finally we consider the application of our tests to real data example.

**ARTICLE HISTORY**
Received 26 November 2019
Accepted 23 April 2020

**KEYWORDS**
Logistic distribution; *U*-statistics; Bahadur efficiency; characterization; Kolmogorov test

**2010 MATHEMATICS SUBJECT CLASSIFICATIONS**
60F10; 62F03; 62G20; 62G30

## 1. Introduction

The logistic distribution family was apparently introduced by Verhulst in the mid nineteenth century. It has been used in many different areas such as logistic regression, reliability theory, physical models, neural networks, hydrology, public health, and, more recently, in finance. Typically, it appears as a substitute for the normal law because it is also symmetric and bell-shaped but has heavier tails. The importance and significance of logistic distribution is described in detail in the handbook [8].

However there are few goodness-of-fit tests which were designed just for this family. Among the rare tests of this type we can mention the tests developed in [1,4,17,26]. Recently there appeared a number of statistical tests based on characterizations. Such tests often are parameter-free and efficient, they have appealing properties based on some hidden features of characterizations.

The survey of such tests can be found in [21] but there is not a single reference to tests for the logistic family. The point is that the number of known characterizations of the logistic family is surprisingly small in comparison with those for the exponential and normal families. This remark was already done by Kotz in [14]. Nevertheless, even now this state of affairs continues to persist. The only research [18] proposing new tests based on a characterization for the logistic distribution appeared when the paper [21] was completed.

**CONTACT** Ya. Yu. Nikitin y.nikitin@spbu.ru, yanikit47@mail.ru Department of Mathematics and Mechanics, Saint-Petersburg State University, Universitetskaya nab. 7/9, Saint-Petersburg 199034, Russia National Research University – Higher School of Economics, Soyuza Pechatnikov 16, Saint-Petersburg 190008, Russia

It seems that one of the first characterizations of logistic distribution by the property of the sample median with random Laplace shift appeared in [9]. After a long pause the interesting paper of Lin and Hu [16] emerged, containing a number of new characterizations of the standard logistic distribution. They introduced the idea of characterization by the random shift of order statistics. Later some papers expanding and developing this direction of research were published, see, e.g.[2,3,12,27].

Also, it appears that the first tests based on the characterization of the logistic location family has been build in [24]. In particular, the underlying characterization uses the equidistribution of two order statistics with the one-sided random exponential shift. The resulting statistics turn out to have reasonably high local Bahadur efficiency against natural alternatives.

In the present paper we use a similar but different characterization which appeared in [2] and [3]. This characterization also includes the equidistribution of smallest order statistics but with independent *two-sided* exponential shifts. We build and study two statistics based on this characterization: one is of the integral type while another is the Kolmogorov-type statistic. Noteworthy is the fact that the new statistics are location-free under the null hypothesis. Evaluating their local Bahadur efficiency we observe that this efficiency is significantly higher than in case of previous tests from [24]. The explanation for this may be the greater symmetry of the two-sided shift model.

The computation of efficiency is associated with the simulation of power for new tests introduced in the present paper and previous ones. We calculate the power for sample size 20 and 50, standard significance levels and for two natural alternatives. One is the standard normal distribution, the second is the logistic distribution with larger scale parameter.

We finish by considering the real data example concerning the annual rainfall in Los Angeles from [5]. Our tests confidently accept the composite hypothesis that the corresponding density is logistic with arbitrary location parameter. In the same time we reject the null hypothesis when in reality the sample has the standard Cauchy distribution.

The structure of the paper is as follows. We formulate the basic characterizations of the logistic location family and construct the corresponding test statistics in Section 2. They turn out to be $U$-statistics with complicated but bounded kernels. In Section 3, we introduce some close alternatives to the logistic distribution, and calculate their Kullback–Leibler distance from the null hypothesis. In Section 4, we study the asymptotic properties, and, in particular, the local Bahadur efficiency of the integral test. The same analysis is being implemented in Section 5 for the Kolmogorov-type statistic. Section 6 is devoted to the power study of our tests and to the analysis of an example of real data. We collect and discuss the obtained results in Section 7.

## 2. The characterization and the construction of test statistics

The characterization used in [24] for the construction of the goodness-of-fit tests for the logistic family was obtained in [12] and is formulated as follows.

**Theorem 2.1:** *Let $X$ and $Y$ be independent identically distributed random variables with the density $f$, and let $Z$ be a random variable independent from $X$ and $Y$ and having the standard* *exponential distribution. Then the equality in distribution*

$$X \stackrel{d}{=} \min(X, Y) + Z \tag{1}$$

*holds if and only if f is the density of the logistic location family, namely* $f(x) = l(x - \theta), \theta \in \mathbb{R}$, *where*

$$l(x) = \frac{e^x}{(1 + e^x)^2}. \tag{2}$$

The authors of [24] have built two tests using characterization (1) and studied their local Bahadur efficiency. As one of the tests was based on the Kolmogorov-type statistic which has a non-normal distribution, the use of this type of efficiency is quite justified. The indicated efficiency turned out to be reasonably high for certain natural alternatives.

Below we use another characterization of the logistic family and construct similar but different tests which give significantly higher efficiency values. This characterization can be found in [2] and is formulated below. Similar but more complicated results can be found in [3].

**Theorem 2.2:** *Let X and Y be as in the previous theorem, and let* $Z_1, Z_2$ *be independent from X and Y random variables with standard exponential distribution. Then the equality in distribution*

$$\min(X, Y) + Z_1 \stackrel{d}{=} \max(X, Y) - Z_2 \tag{3}$$

*holds if and only if f is the density of the logistic location family.*

Equality (3) is more complicated than a similar equation (1) but the tests built on the basis of the equality (3) give a higher efficiency and power than in the case of (2). We will see it later.

Let $X_1, \dots, X_n$ be independent identically distributed observations with density $f$. We are testing the null hypothesis $H_0$ according to which $f(x) = l(x - \theta), \theta \in \mathbb{R}$, where $l$ is the standard logistic density (2) against some close alternatives which will be described in the next section.

Denote by $F_n(t)$ the usual empirical df namely

$$F_n(t) = n^{-1} \sum_{i=1}^{n} I\{X_i < t\}, \quad t \in \mathbb{R}.$$

Consider two $U$–empirical df's:

$$U_{1,n}(t) = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} I\left\{\max(X_i, X_j) < t\right\}, \quad t \in \mathbb{R},$$

and

$$U_{2,n}(t) = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} I\left\{\min(X_i, X_j) < t\right\}, \quad t \in \mathbb{R}.$$

The next step is to consider two shifted $U$–empirical df's in accordance with both parts of equality (3). The $U$-empirical df $U_n^+$ corresponds to the positive exponential shift while

$U_n^-$ is in conformity with negative exponential shift. We have

$$U_n^+(t) = \int_0^\infty U_{1,n}(t+s)e^{-s}\,\mathrm{d}s = \begin{pmatrix} n \\ 2 \end{pmatrix}^{-1} \sum_{1\le i<j\le n} \int_0^\infty I\left\{\max(X_i, X_j) < t+s\right\} e^{-s}\,\mathrm{d}s$$

$$= \binom{n}{2}^{-1} \sum_{1\le i<j\le n} \left(e^{-\max(0,\max(X_i,X_j)-t)}\right),$$

and

$$U_n^-(t) = \int_0^\infty U_{2,n}(t-s)e^{-s}\,\mathrm{d}s = \binom{n}{2}^{-1} \sum_{1\le i<j\le n} \int_0^\infty I\left\{\min(X_i, X_j) < t-s\right\} e^{-s}\,\mathrm{d}s$$

$$= \binom{n}{2}^{-1} \sum_{1\le i<j\le n} \left(1 - e^{(\min(X_i,X_j)-t)}\right) I\left\{\min(X_i, X_j) < t\right\}$$

$$= \binom{n}{2}^{-1} \sum_{1\le i<j\le n} \left(1 - e^{\min(0,\min(X_i,X_j)-t)}\right).$$

To test the null hypothesis $H_0$ we suggest two test statistics based on the characterization from Theorem 2.2: the integral statistic

$$IU_n = \int_{-\infty}^{\infty} \left(U_n^+(t) - U_n^-(t)\right) dF_n(t) \tag{4}$$

and the Kolmogorov-type statistic

$$QU_n = \sup_t |U_n^+(t) - U_n^-(t)|. \tag{5}$$

Our aim is to calculate their local Bahadur efficiencies against the alternatives described in the next section. The outline of Bahadur efficiency can be found in the original papers [6,7], in the monograph [19], and, e.g. in the survey [21, § 3]. The measure of Bahadur efficiency for a sequence of test statistics $\{T_n\}$ is its exact slope $c_T(\theta)$, where $\theta$ is the alternative value of parameter. The calculation of exact slope depends on the large deviation asymptotics of $\{T_n\}$ under the null hypothesis and on its limit almost surely under the alternative. The value of Bahadur efficiency can be defined as

$$\mathit{eff}_T(\theta) = c_T(\theta)/2K(\theta), \tag{6}$$

where $K(\theta)$ is the Kullback–Leibler 'distance' [7, § 4] between the alternative and the composite null hypothesis. Very often, one takes the limit as $\theta \to 0$ in the right-hand side of (6) obtaining, thereby, the local Bahadur efficiency.

## 3. Evaluation of Kullback–Leibler information

We begin with the description of densities $f_i(x,\theta)$ and (when necessary) of their df's $F_i(x,\theta), x \in \mathbb{R}, i = 1, 2, 3$, which will be considered in this paper as alternatives to the logistic location family.

(1) Scale alternative with the density

$$f_1(x,\theta) = \frac{e^{\theta + xe^{\theta}}}{(1+e^{xe^{\theta}})^2}.$$

(2) Generalized hyperbolic cosine alternative with the density

$$f_2(x,\theta) = \frac{\Gamma(\theta+2)}{\Gamma^2(\frac{\theta}{2}+1)} \frac{e^{(x+\theta x/2)}}{(1+e^x)^{\theta+2}}.$$

(An explanation of this formula is as follows. Instead of the traditional form of the logistic density $(4\cosh^2(x/2))^{-1}$, $x \in \mathbb{R}$, we consider $(4\cosh^{2+\theta}(x/2))^{-1}$, $x \in \mathbb{R}$, and make the appropriate normalization.)

(3) Sine-alternative in the spirit of the paper [15] with the df for small $\theta$ $F_3(x,\theta) = L(x) - \theta \sin(2\pi L(x))$, and the density

$$f_3(x,\theta) = l(x) - 2\pi\theta \cos(2\pi L(x)) l(x).$$

Here $L(x) = (1+\exp(-x))^{-1}$, $x \in \mathbb{R}$ is the df of the standard logistic distribution. All these alternatives at $\theta = 0$ go over to the logistic distribution.

Let $K(\theta)$ be the infimum of the Kullback–Leibler ‘distance’ between the alternative and the family of logistic density with unknown location. The alternative densities $f_i$, $i = 1, 2, 3$, are regular enough to satisfy the assertion, see [7, Section 4]

$$2K(\theta) \sim I_f(0) \cdot \theta^2, \quad \text{as} \quad \theta \to 0.$$

Here $I_f(0)$ is the Fisher information at zero for the density $f(x,\theta)$ which is equal to

$$I_f(0) = \int_{-\infty}^{\infty} \frac{|f'_\theta(x,0)|^2}{f(x,0)}\, \mathrm{d}x.$$

Then the formula (6) for the local Bahadur efficiency takes the form [6,19]:

$$\mathit{eff}_T = \lim_{\theta \to 0} \frac{c_T(\theta)}{I_f(0)\theta^2}. \tag{7}$$

Now let us find the value of $I_j(0)$, $j = 1, 2, 3$ for our three alternatives. Using the tables of integrals [10], we obtain:

$$I_1(0) = \int_{-\infty}^{\infty} \frac{e^x(e^x + x + 1 - xe^x)^2}{(1+e^x)^4}\, \mathrm{d}x \approx 1.430. \tag{8}$$

In the same way we find the Fisher information $I_2(0)$ and $I_3(0)$ and put them all together in the Table 1.

**Table 1.** Fisher information for three alternatives.

| | Alternatives | | |
|---|---|---|---|
| Fisher information | $f_1$ | $f_2$ | $f_3$ |
| $I_f(0)$ | 1.430 | 0.355 | 19.739 |

## 4. Integral type statistic

Consider the auxiliary function

$$g(x,y;z) = e^{-\max(0,\max(x,y)-z)} - (1 - e^{\min(0,\min(x,y)-z)}),\ x,y,z \in \mathbb{R}.$$

Performing the integration in (4), we see that the integral statistic $IU_n$ is asymptotically equivalent to the $U$-statistic of degree 3 with the centered kernel

$$\Phi(x,y,z) = \frac{1}{3}\left(g(x,y;z) + g(y,z;x) + g(x,z;y)\right).$$

Let us calculate the projection of this kernel:

$$\Psi(t) = \mathbb{E}\left(\Phi(X,Y,Z)|Z=t\right) = \mathbb{E}\left(\frac{1}{3}g(X,Y;t) + \frac{2}{3}g(X,t;Y)\right).$$

It is easy to show that the characterization we use implies $\mathbb{E}(g(X,Y;t)) = 0$ for any $t$, and it remains to calculate the second summand. We have

$$\mathbb{E}(g(X,t;Y)) = \mathbb{E}\left(e^{-\max(0,\max(X,Y)-t)}\right) - \mathbb{E}\left((1 - e^{\min(0,\min(X,Y)-t)})\right) := I_1(t) - I_2(t).$$

The integrals $I_1(t), I_2(t)$ have been calculated already in [24], hence we have:

$$I_1(t) = Li_2(-e^t) + \frac{1}{2}\ln^2(1+e^t) - \frac{1}{1+e^t} + \frac{\pi^2}{6},\ t \in \mathbb{R},$$

and

$$I_2(t) = Li_2(-e^t) + t\ln(e^t+1) - \frac{1}{2}\ln^2(1+e^t) + \frac{2e^t+1}{1+e^t},\ t \in \mathbb{R}.$$

Here the Euler's dilogarithm $Li_2$ is given by the formula

$$Li_2(z) = -\int_0^z \frac{\ln(1-t)}{t}\,dt,\ z \in \mathbb{C}.$$

Collecting the calculations together, we see that the dilogarithm disappears, and the projection is an even function

$$\Psi(t) = \frac{2}{3}\left(\ln^2(e^t+1) - t\ln(e^t+1) + \frac{\pi^2}{6} - 2\right),\ t \in \mathbb{R}.$$

Let us calculate the variance of this projection. Using numerical integration, we obtain

$$\Delta^2 = \mathbb{E}\Psi^2(X) \approx 0.00697.$$

Hence the kernel $\Phi$ is non-degenerate, and by Hoeffding's theorem [13] one has:

$$\sqrt{n}IU_n \xrightarrow{d} \mathcal{N}\left(0, 9\Delta^2\right),$$

as $n \to \infty$. Using this limit, we can easily calculate the asymptotic critical values of any prescribed level for the test based on $IU_n$.

The kernel $\Phi$ is non-degenerate, centered and bounded, therefore we can describe the logarithmic large deviations of $U$-statistics with such kernels using [23, Theor.2.1]. For the reader's convenience, we state the corresponding theorem.

**Theorem 4.1:** *Let the sequence of U-statistics $V_n$ with centered kernel $\Phi$ of degree $m \geq 1$ be bounded. We suppose that it is non-degenerate with the positive variance $\sigma^2$ of its projection $\psi(s) = E(\Phi(X_1, \dots, X_m)|X_1 = s)$. Then for any real sequence $\{\gamma_n\}$ such that $\gamma_n \to 0$ one has*

$$\lim_{n\to\infty} n^{-1} \ln \Pr\{V_n \geqslant a + \gamma_n\} = \sum_{j=2}^{\infty} b_j a^j,$$

*where the series on the right-hand side converges for sufficiently small $a > 0$, moreover, $b_2 = -1/(2m^2\sigma^2)$.*

Applying this result to our sequence of test statistics $IU_n$, we get the following statement.

**Theorem 4.2:** *For any $t > 0$ $\lim_{n\to\infty} n^{-1} \ln \mathbb{P}(IU_n > t) = h(t)$, where h is some continuous function such that $h(t) \sim -t^2/18\Delta^2$ as $t \to 0$.*

Now we are able to calculate the local Bahadur exact slope $c_{IU}(\theta)$ of the sequence of statistics $IU_n$ from (7). According to Theorems 1 and 2 from [22], and using Theorem 4.2 we get:

$$c_{IU}(\theta, f) \sim \Delta^{-2} \left( \int_{-\infty}^{\infty} \Psi(x) f'_\theta(x, 0)\, \mathrm{d}x \right)^2 \theta^2, \quad \text{as} \quad \theta \to 0. \tag{9}$$

Next, we proceed to the calculation of local efficiencies for all three alternatives listed above. In case of scale alternative with the density $f_1$, we get by formula (9) the following relation for the local Bahadur slope:

$$c_{IU}(\theta, f_1) \sim 1.339 \cdots \cdot \theta^2, \quad \text{as} \quad \theta \to 0.$$

Hence, by formulae (7) and (8), the local Bahadur efficiency in this case is equal to

$$\mathit{eff}_{IU}(f_1) = \lim_{\theta\to 0} \frac{c_{IU}(\theta, f_1)}{2K_1(\theta)} = \frac{1.339\cdots}{1.430\cdots} \approx 0.937.$$

The calculations for the alternatives with the densities $f_2$ and $f_3$ are quite analogous, and we get

$$\mathit{eff}_{IU}(f_2) = \lim_{\theta\to 0} \frac{c_{IU}(\theta, f_2)}{2K_2(\theta)} \approx 0.864; \quad \mathit{eff}_{IU}(f_3) = \lim_{\theta\to 0} \frac{c_{IU}(\theta, f_3)}{2K_3(\theta)} \approx 0.849.$$

Note that all local efficiencies are collected below in the Table 2.

**Table 2.** Comparative local Bahadur efficiencies for test statistics.

| | Test statistic | | | |
|---|---|---|---|---|
| Alternative | $LU_n$ | $IU_n$ | $KU_n$ | $QU_n$ |
| $f_1$ | 0.837 | 0.937 | 0.353 | 0.667 |
| $f_2$ | 0.770 | 0.864 | 0.288 | 0.566 |
| $f_3$ | 0.759 | 0.849 | 0.800 | 0.975 |

## 5. Kolmogorov-type statistic and the table of efficiencies

We return to the statistic $QU_n$ introduced in (5). Its limiting behavior is unknown but it is possible to calculate the approximate critical values by simulation.

This statistic can be considered as the supremum by $t$ of the family of absolute values for $U$ – statistics with the kernels

$$\Phi_1(X,Y;t) = e^{-\max(0,\max(X,Y)-t)} - (1 - e^{\min(0,\min(X,Y)-t)}).$$

These kernels are centered, non-degenerate and bounded. To apply the theorem on large deviations for such statistics from [20], let us calculate the family of projections of these kernels. After extensive computations, we have:

$$\begin{aligned}\Psi_1(s,t) &= \mathbb{E}\left(\Phi_1(X,Y;t)|Y=s\right) = \mathbb{E}\left\{e^{-\max(0,\max(X,s)-t)} - (1 - e^{\min(0,\min(X,s)-t)})\right\}\\ &= e^t\left(\ln(1+e^{\max(s,t)}) - \frac{1}{1+e^{\max(s,t)}} - \max(s,t)\right)\\ &\quad - e^{-t}\left(\frac{(1+e^t)e^{\min(s,t)}}{1+e^{\min(s,t)}} - \ln(1+e^{\min(s,t)})\right)\\ &\quad + \frac{e^{3t}+e^{2t}+I\{s<t\}(e^{2t}+e^t+1)(e^s-e^t)}{e^t(1+e^t)(1+e^s)}.\end{aligned}$$

Now, we should calculate the family of variance functions $\Delta_1^2(t) := \mathbb{E}_X\Psi_1^2(X,t)$ as functions of $t$. After some calculations, we obtain

$$\begin{aligned}\Delta_1^2(t) := \mathbb{E}_X\Psi_1^2(X,t) &= e^{-t}\left(2te^{3t} - 2(t-1)e^{2t} - 3e^t + 2\right)\\ &\quad - 2e^{-2t}\left(e^{4t} - e^{3t} - te^{2t} - e^t + 1\right)\ln(1+e^t) - 2\ln^2(1+e^t).\end{aligned}$$

The supremum of this function is attained for $t = 0$ and equals $1 - 2\ln^2(2)$. Consequently, the key parameter for large deviations is equal to

$$\Delta_1^2 = \sup_{t\in\mathbb{R}} \Delta_1^2(t) = 1 - 2\ln^2(2) \approx 0.00393.$$

From [20] we get the following logarithmic large deviation asymptotics under $H_0$:

**Theorem 5.1:** *For any $z > 0\, \lim_{n\to\infty} n^{-1}\ln\mathbb{P}\{QU_n > z\} = h(z)$, where $h$ is some continuous function, such that $h(z) \sim -z^2/8\Delta_1^2$ as $z \to 0$.*

To this end, we need the following function, see [7, § 7]:

$$b_{QU}(\theta;t) = \mathbb{E}_\theta \Phi_1(X,Y;t) = \mathbb{E}_\theta \left\{ e^{-\max(0,\max(X,Y)-t)} - (1 - e^{\min(0,\min(X,Y)-t)}) \right\},$$

where we used the Glivenko–Cantelli theorem for $U$-empirical df's [11] while the index $\theta$ signifies that the sample has the alternative distribution. Then, the local Bahadur exact slope assumes the representation $c_{QU}(\theta) = 2h(b_{QU}(\theta))$, where $b_{QU}(\theta) = \sup_{t\in\mathbb{R}} |b_{QU}(\theta;t)|$.

Similarly to the expression for the local exact Bahadur slope of the integral statistic, we get the following formula for the Kolmogorov-type statistic, see similar calculations in [25]:

$$c_{QU}(\theta,f) \sim \Delta_1^{-2} \sup_{t\in\mathbb{R}} \left( \int_{-\infty}^{\infty} \Psi_1(x;t) f'_\theta(x,0)dx \right)^2 \cdot \theta^2,\ \theta \to 0.$$

Now we will calculate in details the local Bahadur slope and local Bahadur efficiency of the sequence of statistics $QU_n$. The derivative of the density $f_1(x,\theta)$ with respect to $\theta$ in the point $\theta = 0$ is equal to $f'_{1,\theta}(x,0) = \frac{e^x(e^x+x+1-xe^x)}{(e^x+1)^3}$. First, we calculate the integral:

$$\int_{-\infty}^{\infty} \Psi_1(x,t) f'_{1,\theta}(x,0)dx = \frac{(e^{2t}+1)\ln(1+e^t)}{2e^t} - \frac{te^t+1}{2}.$$

Using Wolfram Mathematica package, we found that the supremum of the square of this function is attained for $t = 0$ and equals $(\ln(2) - \frac{1}{2})^2$. Hence the local exact Bahadur slope has the following representation:

$$\begin{aligned} c_{QU}(\theta,f_1) &\sim \Delta_1^{-2} \sup_{t\in\mathbb{R}} \left( \frac{(e^{2t}+1)\ln(1+e^t)}{2e^t} - \frac{te^t+1}{2} \right)^2 \cdot \theta^2 \\ &= \frac{(\ln(2)-\frac{1}{2})^2}{1-2\ln^2(2)} \cdot \theta^2 \approx 0.954 \cdot \theta^2, \end{aligned}$$

so that the local Bahadur efficiency is equal to

$$\mathit{eff}_{QU}(f_1) = \lim_{\theta\to 0} \frac{c_{QU}(\theta,f_1)}{2K_1(\theta)} = \frac{0.954\cdots}{1.430\cdots} \approx 0.667.$$

The calculations for two other alternatives are quite similar, so we get after some lengthy computations

$$\mathit{eff}_{QU}(f_2) \approx 0.566; \quad \mathit{eff}_{QU}(f_3) \approx 0.975.$$

Let us collect all efficiencies we found as well as the corresponding efficiencies for the tests from [24] in the separate table. There $LU_n$ and $KU_n$ are, respectively, the integral statistic and the Kolmogorov-type statistic from the paper [24] while $IU_n$ and $QU_n$ are new statistics from the present paper.

We see that in all cases the new test statistics are noticeably more efficient than the previous ones from [24]. We believe that it happens due to greater symmetry of the characterization based on (3) with respect to (1). At the same time, the integral statistic is more efficient than the Kolmogorov one. It is a common situation which takes place for most Bahadur efficiency comparisons, see [19].

**Table 3.** Critical values for the statistic $\sqrt{n}IU_n$ and $\sqrt{n}QU_n$.

| | $\sqrt{n}IU_n$ | | | | $\sqrt{n}QU_n$ | |
|---|---|---|---|---|---|---|
| | 5% | | 1% | | 5% | 1% |
| n | Lower | Upper | Lower | Upper | Upper | Upper |
| 20 | −0.106 | 0.234 | −0.142 | 0.312 | 0.935 | 1.158 |
| 50 | −0.125 | 0.206 | −0.170 | 0.276 | 0.940 | 1.160 |

**Table 4.** Power for the statistics $IU_n$,$LU_n$,$QU_n$ and $KU_n$.

| | $IU_n$ | | $LU_n$ | | $QU_n$ | | $KU_n$ | |
|---|---|---|---|---|---|---|---|---|
| $n$ | 5% | 1% | 5% | 1% | 5% | 1% | 5% | 1% |
| 20 | 0.758 | 0.415 | 0.602 | 0.319 | 0.749 | 0.461 | 0.589 | 0.294 |
| 50 | 0.997 | 0.974 | 0.995 | 0.935 | 0.992 | 0.949 | 0.960 | 0.813 |

Very high efficiencies of new statistics should be highlighted. This motivates us to undertake the further study, namely the simulation of their power against common alternatives.

## 6. Power study of new statistics

It is interesting to simulate the powers of new statistics for small sample sizes (as usually, we take $n = 20$ and $n = 50$) and to compare them with the integral statistic $LU_n$ and Kolmogorov statistic $KU_n$ from [24].

We begin by simulating critical values of new statistics for two customary significance level $\alpha = 0.05$ and $\alpha = 0.01$. In the case of integral statistics we provide two tails, the upper and lower. We have simulated 10,000 values of considered statistics and got the following critical values (Table 3).

Next step is to obtain via simulation the powers of all four statistics for a couple of standard alternatives. In this capacity, we take the standard normal distribution and the logistic distribution with nonunit scale parameter $\theta = 0.75$, 0.5 and 0.25, see the alternative $f_1$ in previous sections.

### 6.1. Norm(0,1)

Below we consider the power of our statistics $IU_n, QU_n$ against standard normal distribution for $n = 20$ and $n = 50$, comparing them with the previous statistics $LU_n$, $KU_n$ from [24] (Table 4).

### 6.2. Logistic scale alternative

Now calculate power of our statistics $IU_n, QU_n$ against scale alternative with $\theta = 0.75; 0.5; 0.25$ for n = 20;50 and compare it again with the statistics $LU_n$, $KU_n$ from [24] (Table 5).

**Table 5.** Power for the statistics $IU_n$, $QU_n$, $LU_n$, $KU_n$.

| | | $IU_n$ | | $QU_n$ | | $LU_n$ | | $KU_n$ | |
|---|---|---|---|---|---|---|---|---|---|
| $\theta$ | n | 5% | 1% | 5% | 1% | 5% | 1% | 5% | 1% |
| $\theta = 0.75$ | 20 | 0.963 | 0.839 | 0.953 | 0.806 | 0.911 | 0.696 | 0.883 | 0.668 |
| | 50 | 1 | 1 | 1 | 0.999 | 1 | 0.999 | 1 | 0.993 |
| $\theta = 0.5$ | 20 | 0.621 | 0.286 | 0.673 | 0.400 | 0.506 | 0.264 | 0.542 | 0.277 |
| | 50 | 0.980 | 0.917 | 0.966 | 0.892 | 0.962 | 0.84 | 0.911 | 0.706 |
| $\theta = 0.25$ | 20 | 0.203 | 0.09 | 0.24 | 0.089 | 0.156 | 0.055 | 0.202 | 0.067 |
| | 50 | 0.484 | 0.252 | 0.463 | 0.244 | 0.423 | 0.188 | 0.372 | 0.149 |

### 6.3. *Testing real data sets*

Finally, we consider a sample of real data which is the annual rainfall (in inches) during March recorded at Los Angeles Civic Center from 1973 to 2006:

2.70 3.78 4.83 1.81 1.89 8.02 5.85 4.79 4.10 3.54 8.37 0.28 1.29 5.27 0.95 0.26 0.81 0.17 5.92 7.12 2.74 1.86 6.98 2.16 0.00 4.06 1.24 2.82 1.17 0.32 4.31 1.17 2.14 2.87

The Los Angeles rainfall data have been used earlier by some authors, see [5] where we found this example. The authors of [5] analyzed the above rainfall data assuming it has the logistic distribution with known location $\mu = 2.905$ and scale $\sigma = 1.367$. It was observed that the Kolmogorov-Smirnov test and the corresponding $p$-value showed quite well fit to the above data.

We want to apply our tests $IU_n$ and $QU_n$ to this data. Before that, we need to divide the data by 1.367 to obtain a unit scale. The location is not important here as our tests are location-free. We calculated the values of our statistics and their $p$-values (Table 6). It is seen that the above data confidently support the hypothesis $H_0$, and this corresponds to the conclusion of [5]. The result is somewhat unexpected as the annual rainfall could rather be assumed normal or truncated normal, and not logistic.

Next, for contrast, we examined the second data set consisting of 50 values generated from standard Cauchy distribution which resembles much the logistic distribution. Two densities are similar at-a-glance and not so easy distinguished.

−0.189 −17.680 0.015 −3.059 −1.543 1.877 3.624 0.653 −0.845 0.747 −0.151 −0.061 −1.373 0.130 −14.752 0.595 3.670 0.434 −64.448 −3.843 −14.150 2.179 −2.730 0.266 0.543 −0.773 1.202 0.585 −0.349 −2.405 −0.683 0.277 −18.711 −0.171 1.682 −0.987 0.131 −0.380 3.662 0.937 −1.767 0.141 −0.544 4.343 10.563 −1.040 −0.413 −0.285 −0.837 0.318

We calculated the same test statistics and their $p$-values which are put also into Table 6. Our tests surely reject the null hypothesis for the second data set, and the $p$-values are convincing enough. This result was certainly expected.

**Table 6.** Values of statistics $IU_n$, $QU_n$ for the data sets.

| | The first data set | | The second data set | |
|---|---|---|---|---|
| Test | Test statistic value | $p$-Value | Test statistic value | $p$-Value |
| IU | 0.002 | $p = 0.864$ | −0.028 | $p = 0.002$ |
| QU | 0.119 | $p = 0.209$ | 0.146 | $< 0.002$ |

## 7. Conclusion and discussion

We have built two new location-free tests for the logistic distribution based on a recent characterization. They have the $U$-empirical structure, and we undertook their asymptotic analysis, especially the study of their Bahadur efficiency. The tests are not difficult for implementation and have remarkably high local Bahadur efficiencies at least against common alternatives to logistic family. These efficiencies are appreciably higher than the efficiencies of previous tests recently proposed by the authors.

The empirical power of new tests which was simulated for natural alternatives and small samples with sizes $n = 20$ and $n = 50$ confirmed the ordering of tests by their Bahadur efficiency and turned out to be rather high too. We also examined a set of real data related to annual rainfall in Los Angeles during 34 years and confirmed the (unexpected) conclusion of other researchers that it obeys the logistic distribution.

The simplicity, high efficiency and power of the new tests make them attractive. These tests significantly replenish the meager set of goodness-of-fit tests for the logistic family available to statisticians, and this put on the agenda their subsequent realization in the $R$ language.

## Acknowledgments

We would like to thank the Associate Editor and the Referees for their valuable remarks.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This work was supported by the Russian Fund of Basic Research [grant number 20-51-12004].

## References

[1] N. Aguirre and M. Nikulin, *Chi-squared goodness-of-fit test for the family of logistic distributions*, Kybernetika 30 (1994), pp. 214–222.
[2] M. Ahsanullah, V.B. Nevzorov, and G.P. Yanev, *Characterizations of distributions via order statistics with random exponential shifts*, J. Appl. Stat. Sci. 18 (2010), pp. 297–305.
[3] M. Ahsanullah, G.P. Yanev, and C. Onica, *Characterizations of logistic distribution through order statistics with independent exponential shifts*, Econ. Qual. Control 27 (2011), pp. 85–96.
[4] E.E.A.A. Aly and M.A. Shayib, *On some goodness-of-fit tests for the normal, logistic and extreme-value distributions*, Commun. Stat. Theory Methods 21 (1992), pp. 1297–1308.
[5] A. Asgharzadeh, R. Valiollahi, and M. Abdi, *Point and interval estimation for the logistic distribution based on record data*, SORT-Stat. Oper. Res. Trans. 1 (2016), pp. 89–112.
[6] R.R. Bahadur, *Stochastic comparison of tests*, Ann. Math. Stat. 31 (1960), pp. 276–295.
[7] R.R. Bahadur, *Some Limit Theorems in Statistics*, SIAM, Philadelphia, 1971.
[8] N. Balakrishnan, *Handbook of the Logistic Distribution*, CRC Press, Boca Raton, 1991.
[9] E.O. George and G.S. Mudholkar, *A characterization of the logistic distribution by a sample median*, Ann. Inst. Stat. Math. 33 (1981), pp. 125–129.
[10] I.S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series, and Products*, 4th ed., A. Jeffrey and D. Zwillinger, eds., Academic Press, New York, 1994.

[11] R. Helmers, P. Janssen, and R. Serfling, *Glivenko-Cantelli properties of some generalized empirical DF's and strong convergence of generalized L-statistics*, Probab. Theor. Relat. Fields 79 (1988), pp. 75–93.

[12] C.Y. Hua and G.D. Lin, *Characterizations of the logistic and related distributions*, J. Math. Anal. Appl. 463 (2018), pp. 79–92.

[13] V.S. Korolyuk and Yu.V. Borovskikh, *Theory of U-Statistics*, Kluwer, Dordrecht, 1994.

[14] S. Kotz, *Characterizations of statistical distributions: a supplement to recent surveys*, Int. Stat. Rev. 42 (1974), pp. 39–65.

[15] C. Ley and D. Paindaveine, *Le Cam optimal tests for symmetry against Ferreira and Steel's general skewed distributions*, J. Nonpar. Stat. 21 (2009), pp. 943–967.

[16] G.D. Lin and C.Y. Hu, *On characterizations of the logistic distribution*, J. Stat. Plan. Infer. 138 (2008), pp. 1147–1156.

[17] S.G. Meintanis, *Goodness-of-fit tests for the logistic distribution based on empirical transforms*, Sankhya: Ind. J. Stat. 66 (2004), pp. 306–326.

[18] B. Milošević and M. Obradović, *Two-dimensional Kolmogorov-type goodness-of-fit tests based on characterisations and their asymptotic efficiencies*, J. Nonpar. Stat. 28 (2016), pp. 413–427.

[19] Ya. Nikitin, *Asymptotic Efficiency of Nonparametric Tests*, Cambridge University Press, New York, 1995.

[20] Ya.Yu. Nikitin, *Large deviations of U-empirical Kolmogorov-Smirnov tests, and their efficiency*, J. Nonpar. Stat. 22 (2010), pp. 649–668.

[21] Y.Y. Nikitin, *Tests based on characterizations, and their efficiencies: a survey*, Acta Comment. Univ. Tartuens. Math. 21 (2017), pp. 3–24.

[22] Ya.Yu. Nikitin and I. Peaucelle, *Efficiency and local optimality of distribution-free tests based on U- and V- statistics*, Metron LXII (2004), pp. 185–200.

[23] Ya.Yu. Nikitin and E.V. Ponikarov, *Rough large deviation asymptotics of Chernoff type for von Mises functionals and U-statistics*, Proc. St. Petersburg Math. Soc. 7 (1999), pp. 124–167. Engl. Transl. in AMS Transl., Ser. 2 203 (2001), pp. 107–146.

[24] Y.Y. Nikitin and I.A. Ragozin, *Goodness-of-fit tests based on a characterization of logistic distribution*, Vestnik St. Petersb. Univ. Math. 52 (2019), pp. 169–177.

[25] Ya.Yu. Nikitin and K.Yu. Volkova, *Efficiency of exponentiality tests based on a special property of exponential distribution*, Math. Meth. Stat. 25 (2016), pp. 54–66.

[26] M.A. Stephens, *Tests of fit for the logistic distribution based on the empirical distribution function*, Biometrika 66 (1979), pp. 591–595.

[27] V.O. Zykov and V.B. Nevzorov, *Some characterizations of families of distributions, including logistic and exponential ones, by properties of order statistics*, J. Math. Sci. 176 (2011), pp. 203–206.