



A new family of omega-square-type statistics with Bahadur local optimality for the location family of generalized logistic distributions

Jean-Renaud Pycke

► To cite this version:

Jean-Renaud Pycke. A new family of omega-square-type statistics with Bahadur local optimality for the location family of generalized logistic distributions. *Statistics and Probability Letters*, 2021, 170, 10.1016/j.spl.2020.108999 . hal-03008958

HAL Id: hal-03008958

<https://hal.science/hal-03008958>

Submitted on 15 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

A new family of omega-square-type statistics with Bahadur local optimality for the location family of generalized logistic distributions

Jean-Renaud Pycke

University of Evry

Abstract

We introduce a family of weighted omega-square statistics. We show that they provide a family of locally asymptotically optimal goodness-of-fit tests, in the sense of Bahadur, for the location problem associated with (type IV) generalized logistic distributions. Our paper extends and unifies some results already known for the classical Cramér-von Mises and Anderson-Darling statistics which are optimal with respect to the hyperbolic secant and the logistic distributions.

Keywords: Bahadur efficiency, nonparametric tests, integral statistics
2008 MSC: 62F05

1. Introduction

Let, for $x \in \mathbb{R}$, $F_n(x) := n^{-1} \times (\text{number of observations} \leq x)$, denote the empirical distribution function based on a sample $(X_i)_{1 \leq i \leq n}$ of size n drawn from a population with absolutely continuous distribution function $F(x) = P(X_1 \leq x)$, and probability density function (p.d.f.) $f = F'$. In order to test the goodness-of-fit hypothesis $H_0 : f = f_0$, let us introduce the uniform empirical process $\{\mathbf{U}_n(t) : 0 < t < 1\}$ defined by $\mathbf{U}_n(t) := n^{-1/2} \sum_{i=1}^n (1_{\{F_0(X_i) \leq t\}} - t)$ and $\{\mathbf{B}(t) : 0 < t < 1\}$ the standard Brownian bridge, i.e. the centred Gaussian process with covariance function $\mathbb{E}\{\mathbf{B}(s)\mathbf{B}(t)\} = \min(s, t) - st$. An omega-square statistic is defined to be

$$n\omega_{n,q}^2 := \int_0^1 q(t) \mathbf{U}_n^2(t) dt = n \int_{-\infty}^{\infty} q\{F_0(x)\} [F_n(x) - F_0(x)]^2 f_0(x) dx, \quad (1)$$

Preprint submitted to preprint

November 1, 2020

where $q : (0, 1) \rightarrow (0, \infty)$ is a non-negative weight function satisfying

$$\int_0^1 t(1-t)q(t)dt < \infty. \quad (2)$$

The latter condition ensures that, under H_0 , the convergence in law $n\omega_{n,q}^2 \rightarrow \int_0^1 q(t)\mathbf{B}^2(t)dt$ takes place (take $a(n) = b(n) = 0$ and $p = 2$ in relation (5.3.59) of Theorem 3.3 in [6]). Two celebrated cases are the Cramér-von Mises and the Anderson-Darling statistics

$$W_n^2 = \int_0^1 \mathbf{U}_n^2(t)dt, \quad A_n^2 = \int_0^1 \frac{\mathbf{U}_n^2(t)}{t(1-t)}dt \quad (3)$$

which are (1) with $q(t) = 1$ and $q(t) = 1/[t(1-t)]$.

We propose the following generalization of these two statistics. Assume $\nu_1, \nu_2 \in (0, \infty)$ and let $I_{\nu_1, \nu_2} : (0, 1) \rightarrow (0, 1)$ denote the normalized incomplete Beta function $t \mapsto I_{\nu_1, \nu_2}(t) = [B(\nu_1, \nu_2)]^{-1} \int_0^t u^{\nu_1-1}(1-u)^{\nu_2-1}du$, where B is Euler Beta function $B(\nu_1, \nu_2) = \Gamma(\nu_1)\Gamma(\nu_2)/\Gamma(\nu_1+\nu_2)$. Useful properties of special functions related to Γ are provided in Chapter 6 of [1], see in particular formulas 6.1.6, 6.1.8–9, 6.1.15, 6.1.18, 6.1.47, 6.2.2, 6.3.1, 6.4.10, and 6.6.1–3. We will make use repeatedly, for the incomplete Beta function and its inverse, of the elementary symmetry relations $I_{\nu_2, \nu_1}(1-t) = 1 - I_{\nu_1, \nu_2}(t)$ and $I_{\nu_2, \nu_1}^{-1}(1-t) = 1 - I_{\nu_1, \nu_2}^{-1}(t)$, for $0 < t < 1$.

Consider the weight function

$$q_{\nu_1, \nu_2}(t) := [I_{\nu_1, \nu_2}^{-1}(t)]^{1-2\nu_1} [I_{\nu_2, \nu_1}^{-1}(1-t)]^{1-2\nu_2} \quad (0 < t < 1), \quad (4)$$

with I_{ν_1, ν_2}^{-1} denoting the inverse function of I_{ν_1, ν_2} . Elementary computations show that (2) holds for q_{ν_1, ν_2} provided $\nu_1, \nu_2 > 0$. The associated omega-square statistic is given by

$$n\omega_n^2(\nu_1, \nu_2) := \int_0^1 q_{\nu_1, \nu_2}(t)\mathbf{U}_n^2(t)dt \quad (n \geq 1), \quad (5)$$

so that, as is easily checked, $n\omega_n^2(1/2, 1/2)$ is the Cramer-von Mises statistic and $n\omega_n^2(1, 1)$ the Anderson-Darling statistic.

The aim of this paper is to discuss the Bahadur efficiency of the omega-square statistic (5) and show that it provides a locally asymptotically optimal (LAO) family of tests, in the sense of Bahadur, for the location parameter problem

$$H_0 : f(x) = f_0(x) \tag{6}$$

$$\text{against the alternative } H_1 : f(x) = f_0(x + \theta), \theta \neq 0, \tag{7}$$

in the case when f_0 is the generalized logistic distributions f_{ν_1, ν_2} given by (10). Our main result is Theorem 1 in Section 3.

In this way we extend results already known for Cramér - von Mises and Anderson-Darling statistics (3), who provide LAO tests for the hyperbolic secant and the logistic distributions (11) (see [13], Corollary 1 p. 225, and Theorem 6.3.5 p. 227). Note that for the hyperbolic secant distribution a simple linear rank test of location which is asymptotically most powerful is discussed in [11].

The remainder of our paper is organised as follows. In Section 2 we recall some basic definitions and facts about the generalized logistic distributions and compute some quantities that will be required in the determination of Bahadur efficiency.

Section 3 deals with Bahadur efficiency. Recall that this concept was systematically developed by Bahadur in [2], [3], [4] and [5]. Further expositions, results and references are provided by [8], [16], and [13]. The latter will remain our main reference throughout this paper, in particular Chapter 6, since our main result can be seen as a corollary of Theorem 6.2.4 p. 219. Our main contribution is to make explicit the weight functions (4) associated with distributions (10) for reaching Bahadur optimality.

In Section 4 we group some technical results and proofs used in the previous sections.

Let us conclude this introduction by recalling the following facts from Bahadur's theory of efficiency. For further details, see [13], Chapter 1, especially Theorem 1.2.2 p. 6 (restatement of Theorem 7.2 p.27 in [4]), which plays a central role in Bahadur's theory and whose notations will be used in the sequel.

Given the sequence of statistics $T = (T_n)_{n \geq 1}$, used for the test (6)–(7), the *Bahadur exact slope*, denoted $c_T(\theta)$, appears to be a fundamental quantitative characteristic of T . It is a nonrandom variable characterizing the asymptotic (as $n \rightarrow \infty$) efficiency of (T_n) : the bigger the exact slope $c_T(\theta)$ is, the most efficient the associated test will be, in the sense of Bahadur. See examples of computations of some exact slopes in [4], §8.

An important upper bound, an analogue of Cramér-Rao inequality in the problem of estimating, is $c_T(\theta) \leq 2K(\theta)$, K denoting the Kullback-Leibler

divergence (15). Due to this upper bound, it is natural to state that (T_n) is *asymptotically optimal* provided equality holds for all $\theta \neq 0$. A somewhat weaker, but mathematically more amenable property, is to say that (T_n) is *locally* asymptotically optimal provided

$$c_T(\theta) \sim 2K(\theta) \quad (\theta \rightarrow 0). \quad (8)$$

Let us now proceed with the exposition of our results.

2. The generalized logistic distribution

One defines a proper cumulative distribution functions (c.d.f.), by setting

$$F_{\nu_1, \nu_2}(x) := I_{\nu_1, \nu_2}\left(\frac{1 + \tanh x}{2}\right) \quad (-\infty < x < \infty). \quad (9)$$

The associated p.d.f., $f_{\nu_1, \nu_2} := F'_{\nu_1, \nu_2}$, admits the equivalent expressions

$$f_{\nu_1, \nu_2}(x) = \frac{2}{B(\nu_1, \nu_2)} \frac{e^{(\nu_1 - \nu_2)x}}{(e^x + e^{-x})^{\nu_1 + \nu_2}} = \frac{2}{B(\nu_1, \nu_2)} \left[\frac{1 + \tanh(x)}{2} \right]^{\nu_1} \left[\frac{1 - \tanh(x)}{2} \right]^{\nu_2} \quad (10)$$

This p.d.f. was discussed by [15] (see formula (2) p. 762), see also [10]. This distribution is referred to as the type IV generalized logistic distribution in [9], whose Chapter 23 provides basic facts, formulas, and more references about the logistic distribution and its generalizations, including (10). From [9], relations (23.64) – (23.66) p. 142, we easily infer that a random variable $\mathbf{X}_{\nu_1, \nu_2}$ with density f_{ν_1, ν_2} has expectation and variance given by $\mathbb{E}\mathbf{X}_{\nu_1, \nu_2} = [\psi(\nu_1) - \psi(\nu_2)]/2$ and $\mathbb{V}\mathbf{X}_{\nu_1, \nu_2} = [\psi'(\nu_1) + \psi'(\nu_2)]/4$ where in usual notation, $\psi(z) := \Gamma'(z)/\Gamma(z)$.

Note that f_{ν_1, ν_2} is positively skewed if $\nu_1 > \nu_2$, negatively skewed if $\nu_1 < \nu_2$. It is symmetric about zero if and only if $\nu_1 = \nu_2$ and in this case can be rewritten as $f_{\nu, \nu}(x) = \Gamma(\nu + 1/2)/[\sqrt{\pi}\Gamma(\nu) \cosh^{2\nu} x]$. The Laplace and the normal distributions can be seen as limit cases of $f_{\nu, \nu}$, in view of relations $\lim_{\nu \rightarrow 0^+} (2\nu)^{-1} f_{\nu, \nu}(\frac{x}{2\nu}) = 2^{-1} e^{-|x|}$ and $\lim_{\nu \rightarrow \infty} (\sqrt{2\nu})^{-1} f_{\nu, \nu}(\frac{x}{\sqrt{2\nu}}) = (2\pi)^{-1/2} e^{-\frac{x^2}{2}}$. This aspect is addressed by Theorem 3 in [7]. Two important particular cases are $(\nu_1, \nu_2) = (1/2, 1/2)$ and $(\nu_1, \nu_2) = (1, 1)$, giving rise to the hyperbolic secant and the logistic distributions

$$f_{\frac{1}{2}, \frac{1}{2}}(x) = \frac{1}{\pi \cosh(x)}, \quad f_{1, 1}(x) = \frac{1}{2 \cosh^2(x)} \quad (x \in \mathbb{R}), \quad (11)$$

(see [9], formulas (23.74) – (23.75) p. 147, and (23.6) – (23.9) p. 115-116).

Let us now introduce some analytical features related to these distributions in the context of Bahadur efficiency.

A key role is played by the density-quantile function

$$t \in (0, 1) \mapsto w(t) := f \circ F^{-1}(t) \quad (12)$$

associated with a distribution F with density f (see formula (33.70) p. 678 in [9], §3 in [14], formulas (1) and (10) in Chapter 18 of [17] for definitions of the density-quantile function, and then equalities (6.1.4) p. 210, (6.2.3) p. 213 in [13] for its use in our context).

Proposition 1. The following properties hold:

- (i) The density-quantile function associated with distribution F_{ν_1, ν_2} , is given by

$$w_{\nu_1, \nu_2}(t) = f_{\nu_1, \nu_2} \circ F_{\nu_1, \nu_2}^{-1}(t) = \frac{2I_{\nu_1, \nu_2}^{-1}(t)^{\nu_1} I_{\nu_2, \nu_1}^{-1}(1-t)^{\nu_2}}{B(\nu_1, \nu_2)} \quad (0 < t < 1). \quad (13)$$

- (ii) An absolutely continuous distribution F has a density-quantile function $w = f \circ F^{-1}$ of the form

$$w = Cw_{\nu_1, \nu_2} \quad (14)$$

with $C \in (0, \infty)$, if and only if $F(x) = F_{\nu_1, \nu_2}(Cx + \theta)$ for some $\theta \in \mathbb{R}$.

Proof. Let us set $F_{\nu_1, \nu_2}^{-1}(t) = x$, so that

$$t = I_{\nu_1, \nu_2}\left(\frac{1 + \tanh x}{2}\right), \quad I_{\nu_1, \nu_2}^{-1}(t) = \frac{1 + \tanh x}{2}, \quad I_{\nu_2, \nu_1}^{-1}(1-t) = \frac{1 - \tanh x}{2}.$$

Thus in these notations (13) becomes a reformulation of (10).

Now put $Q = F^{-1}$, $q = Q'$ and let m be the median of a distribution with density function f . Differentiation of $F \circ Q(t) = t$ gives $q(t)f[Q(t)] = q(t)w(t) = 1$ so that the quantile function associated with F is, in the general case,

$$Q(t) = m + \int_{\frac{1}{2}}^t q(u) du = m + \int_{\frac{1}{2}}^t \frac{du}{w(u)} \quad (0 < t < 1).$$

The quantile function of distribution function $x \mapsto F(Cx + \theta)$ is

$$\frac{Q(t) - \theta}{C} = \frac{m - \theta}{C} + \int_{\frac{1}{2}}^t \frac{du}{Cw(u)},$$

and this formula, in the case of $F_{\nu_1, \nu_2}(Cx + \theta)$, becomes

$$\frac{Q_{\nu_1, \nu_2}(t) - \theta}{C} = \frac{m_{\nu_1, \nu_2} - \theta}{C} + \int_{\frac{1}{2}}^t \frac{du}{Cw_{\nu_1, \nu_2}(u)}$$

so that (14) holds for $F_{\nu_1, \nu_2}(Cx + \theta)$.

Conversely, if for the distribution F , (14) holds, then its quantile function can be written, if we set $\theta = m_{\nu_1, \nu_2} - Cm$, as

$$Q(t) = \frac{m_{\nu_1, \nu_2} - \theta}{C} + \int_{\frac{1}{2}}^t \frac{du}{Cw_{\nu_1, \nu_2}(u)} = \frac{Q_{\nu_1, \nu_2}(t) - \theta}{C}.$$

Taking the inverse functions we obtain $F(x) = F_{\nu_1, \nu_2}(Cx + \theta)$. \square

Two other analytic properties that prove useful in computing the local asymptotic Bahadur efficiency are the following. The first is the regularity condition (6.2.1) p. 212 in [13], given below by the first relation in (18). It relates local behaviour of the Kullback-Leibler divergence

$$K(\theta) := \int_{-\infty}^{\infty} f_0(x + \theta) \log \left\{ \frac{f_0(x + \theta)}{f_0(x)} \right\} dx \quad (15)$$

with respect to the Fisher information at the origin (see in [17], §5 of Chapter 4, Definition 1, formula (1) p. 181, and [13] p. 223),

$$I_0(f_0) := \int_{-\infty}^{\infty} \frac{f_0'(x)^2}{f_0(x)} dx = \int_0^1 w_0'(t)^2 dt. \quad (16)$$

The second analytic property is related to the functional

$$b_{q, f_0}(\theta) := \int_{-\infty}^{\infty} \{F_0(x + \theta) - F_0(x)\}^2 q[F_0(x)] f_0(x) dx, \quad (17)$$

easily seen to be the limit of $\omega_{n, q}^2$ defined by (1) under the alternative hypothesis (7) (in [13], see (1.2.6) p. 6 and the formula for $\omega_{n, q}^2$ in Table 1 p. 75, where the power $1/k$ should be read $2/k$). For computational convenience

we decide, as is customary in practice, to deal with the wide class of densities f_0 satisfying, for these two quantities, as $\theta \rightarrow 0$,

$$K(\theta) \sim \frac{\theta^2}{2} I_0(f_0), \quad b_{q,f_0}(\theta) \sim \theta^2 \int_{-\infty}^{\infty} q[F_0(x)] f_0^3(x) dx = \theta^2 \int_0^1 q(t) w_0^2(t) dt, \quad (18)$$

with $w_0 = f_0 \circ F_0^{-1}$. Let us now state some regularity conditions sufficient for (18) to hold. They are adapted from conditions (R1) – (R5) p. 211-212 in [13]. As usual $f^{(k)}$ will denote the k -th order derivative of f , for $k \geq 1$, with $f^{(1)} = f'$ and $f^{(2)} = f''$.

Lemma 1. If the probability density function $f_0 = F_0'$ satisfies the following conditions (C1)–(C3), then the location parameter family $\{f_0(x+\theta), \theta \in \Theta\}$, satisfies the asymptotic equalities (18).

(C1): f_0 is positive, bounded, three times differentiable with $\lim_{\pm\infty} f_0 = \lim_{\pm\infty} f_0' = 0$.

(C2): There exist integrable functions h_i , $i = 1, 2$, such that: $\forall \theta \in \Theta, \forall x \in \mathbb{R} : |f_0^{(i)}(x+\theta)| < h_i(x)$, $\int_{\mathbb{R}} h_i(y) dy < \infty$.

(C3): There exists an integrable function h_3 and a constant M such that: $\forall \theta \in \Theta, \forall x \in \mathbb{R} : |\{\log[f_0(x+\theta)]\}^{(3)}| < h_3(x)$, $\int_{\mathbb{R}} f_0(y+\theta) h_3(y) dy \leq M$.

Proof. In the case of the location parameter model it is easily seen that conditions (R1) – (R4) are implied by (C1). Next, (R5) is the first relation in (18), and (C1) – (C3) imply conditions 1-3 p. 26-27 in [12], which are sufficient for it to hold. \square

From now on the set of functions satisfying conditions (C1) – (C3) will be denoted \mathfrak{G} .

Proposition 2. Assume $\nu_1, \nu_2 > 0$. Then $f_{\nu_1, \nu_2} \in \mathfrak{G}$ and consequently (18) holds with $f_0 = f_{\nu_1, \nu_2}$.

Proof. A simple look at (10) shows that for f_{ν_1, ν_2} the only non-trivial assertion in (C1) might be $\lim_{\pm\infty} f_{\nu_1, \nu_2}' = 0$, but the latter is a straightforward consequence of Lemma 4.

With the change of variables $x \rightarrow x + \theta$, Lemma 4 also shows that for $k = 1, 2$, functions $f_{\nu_1, \nu_2}^{(k)}(x + \theta)$ can be written as $f_k(x + \theta)$ for some $f_k \in \mathcal{L}$.

Then Lemma 3, combined with the triangle inequality, shows that f_k satisfies $|f_k(x + \theta)| < h_k(x)$ for some $h_k \in \mathcal{L}$, $k = 1, 2$.

As for h_3 , elementary calculus leads to

$$|\{\log[f_{\nu_1, \nu_2}(x + \theta)]\}^{(3)}| = 2(\nu_1 + \nu_2)|\tanh(x + \theta)(1 - \tanh^2(x + \theta))| < 2(\nu_1 + \nu_2)$$

so that the constant function $h_3(x) = M = 2(\nu_1 + \nu_2)$ is suitable. \square

Being now assured that (18) will hold, let us explicit the Fisher information appearing in this relation.

Proposition 3. The generalized logistic density function f_{ν_1, ν_2} has a Fisher information (16) given by

$$I_0(\nu_1, \nu_2) = \frac{4\nu_1\nu_2}{1 + \nu_1 + \nu_2} \quad (\nu_1, \nu_2 > 0). \quad (19)$$

Proof. By integrating over the real line relation (25) from Lemma 5, we get

$$\frac{1 + \nu_1 + \nu_2}{4\nu_1\nu_2} \int_{-\infty}^{\infty} \frac{f'_{\nu_1, \nu_2}(x)^2}{f_{\nu_1, \nu_2}(x)} dx = \frac{\nu_1(1 + \nu_2)}{\nu_1 + \nu_2} - \frac{2\nu_1\nu_2}{\nu_1 + \nu_2} + \frac{\nu_2(1 + \nu_1)}{\nu_1 + \nu_2} = 1$$

and (19) is established. \square

3. Bahadur efficiency of Cramer-von Mises statistics

Let us start by computing the integral appearing on the right-hand side of (18).

Proposition 4. The generalized logistic distribution with density function f_{ν_1, ν_2} satisfies

$$\int_{-1}^1 q_{\nu_1, \nu_2}(t) w_{\nu_1, \nu_2}^2(t) dt = \frac{4\nu_1\nu_2}{(1 + \nu_1 + \nu_2)(\nu_1 + \nu_2)B^2(\nu_1, \nu_2)}. \quad (20)$$

Proof. In view of (4) and (13) the integral on the left side of (20) equals

$$\begin{aligned} \frac{4}{B^2(\nu_1, \nu_2)} \int_0^1 I_{\nu_1, \nu_2}^{-1}(t) I_{\nu_2, \nu_1}^{-1}(1 - t) dt &= \frac{4}{B^2(\nu_1, \nu_2)} \int_0^1 u(1 - u) \frac{u^{\nu_1-1}(1 - u)^{\nu_2-1} du}{B(\nu_1, \nu_2)} \\ &= \frac{4B(\nu_1 + 1, \nu_2 + 1)}{B^3(\nu_1, \nu_2)} \end{aligned}$$

where we have used the change of variable $t = I_{\nu_1, \nu_2}(u)$. Then the result follows from the identity $B(a + 1, b + 1) = abB(a, b)/[(1 + a + b)(a + b)]$. \square

Now recall that we wish LAO condition (8) to hold. The Kullback-Leibler divergence on the right-hand side of this relation satisfies asymptotic equality in (18). As for the slope on the left-hand side, a fundamental role is played by the eigenvalue problem

$$y'' = -\lambda qy, \quad y(0) = y(1) = 0 \quad (21)$$

(see [13], (2.4.5) p. 57 and (6.2.12) p. 219 with $k = 2$).

In the case of omega-square statistics all eigenvalues λ are simple and positive (see [13] p. 57-58). The principal eigenvalue (i.e. the smallest) is denoted $\lambda_0(q)$, the eigenspace \mathbf{V} associated with λ_0 is called the *leading set*.

The leading set characterizes the efficiency a statistic. It consists of all possible density-quantile functions w_0 of distributions for which our statistic will be locally asymptotically optimal for the location problem (6) – (7). Let us summarize these properties, as a corollary of Theorem 6.2.4 p. 219 in [13].

Lemma 2. Let X_1, \dots, X_n be a sample drawn from $f_0 = F'_0 \in \mathfrak{G}$ with density-quantile function w_0 . Let $q : (0, 1) \rightarrow (0, \infty)$ be a weight function satisfying $\int_0^1 q(t)dt < \infty$. Let $\lambda_0(q)$ be the principal eigenvalue of the eigenvalue problem (21). Then in the goodness-of-fit problem (6) – (7), the exact slope of the statistic (1) satisfies $c_{q,f_0}(\theta) \sim \theta^2 \lambda_0(q) \int_0^1 q(t)w_0^2(t)dt$. Thus the sequence of statistics $(\omega_{n,q}^2)_{n \geq 1}$ is LAO in the sense of Bahadur for the goodness-of-fit problem (6) – (7) if and only if

$$\lambda_0 \int_0^1 q(t)w_0^2(t)dt = I_0(f_0) \text{ with } w_0(0) = w_0(1) = 0, \quad (22)$$

or equivalently $w_0'' = -\lambda_0 w_0 q$ with $w_0(0) = w_0(1) = 0$.

Note in passing for further use that the fundamental assumption $\int_0^1 q(t)dt < \infty$ of the preceding Lemma holds for $q = q_{\nu_1, \nu_2}$ provided $\nu_1, \nu_2 \in (0, 1)$. Let us now solve the eigenvalue problem for the generalized logistic distributions in the general case $\nu_1, \nu_2 > 0$.

Proposition 5. The principal eigenvalue of problem (21) with $q = q_{\nu_1, \nu_2}$ is

$$\lambda_0(\nu_1, \nu_2) = (\nu_1 + \nu_2)B(\nu_1, \nu_2)^2 \quad (23)$$

and every eigenfunction associated with λ_0 belongs to the leading set

$$\mathbf{V}(\nu_1, \nu_2) = \{Cw_{\nu_1, \nu_2}, C \in \mathbb{R}^*\} = \{C I_{\nu_1, \nu_2}^{-1}(t)^{\nu_1} I_{\nu_2, \nu_1}^{-1}(1-t)^{\nu_2}, C \in \mathbb{R}^*\}. \quad (24)$$

Proof. Let us first show that $y = w_{\nu_1, \nu_2}$ is a solution of (21) for $q = q_{\nu_1, \nu_2}$ and $\lambda = (\nu_1 + \nu_2)B(\nu_1, \nu_2)^2$. To this end we set $t = (1 + \tanh x)/2$ so that we successively obtain

$$\begin{aligned} y(I_{\nu_1, \nu_2}(t)) &= f_{\nu_1, \nu_2}(x) = \frac{2t^{\nu_1}(1-t)^{\nu_2}}{B(\nu_1, \nu_2)}, \\ \frac{1}{B(\nu_1, \nu_2)} t^{\nu_1-1}(1-t)^{\nu_2-1} y'(I_{\nu_1, \nu_2}(t)) &= \frac{2t^{\nu_1-1}(1-t)^{\nu_2-1}}{B(\nu_1, \nu_2)} [\nu_1(1-t) - \nu_2 t] \\ y'(I_{\nu_1, \nu_2}(t)) &= 2[\nu_1(1-t) - \nu_2 t] \\ \frac{1}{B(\nu_1, \nu_2)} t^{\nu_1-1}(1-t)^{\nu_2-1} y''(I_{\nu_1, \nu_2}(t)) &= -2(\nu_1 + \nu_2) \\ y''(I_{\nu_1, \nu_2}(t)) &= -\frac{2(\nu_1 + \nu_2)B(\nu_1, \nu_2)}{t^{\nu_1-1}(1-t)^{\nu_2-1}}. \end{aligned}$$

In view of (4) and (13), $[q_{\nu_1, \nu_2} w_{\nu_1, \nu_2}](I_{\nu_1, \nu_2}(t)) = 2/[B(\nu_1, \nu_2)t^{\nu_1-1}(1-t)^{\nu_2-1}]$, and by comparing these last two equalities the differential equation in (24) is satisfied. Limit conditions are obviously fulfilled.

Now this solution $y = w_{\nu_1, \nu_2}$ to problem (24) has no zero over $(0, 1)$, and a classical argument from the theory of Sturm-Liouville operators shows that $\lambda = \lambda_0$ (see [13] p. 58). \square

Remark 1. It is already known (see [13] p. 220 and Theorem 6.2.5 p. 221), that the principal eigenvalues and the leading sets for the Cramér-von Mises statistic $\omega_n^2(1/2, 1/2)$, and the Anderson-Darling's statistic $\omega_n^2(1, 1)$ are given by $\lambda_0(1/2, 1/2) = \pi^2$, $\lambda_0(1, 1) = 2$, $\mathbf{V}(\frac{1}{2}, \frac{1}{2}) = \{C \sin(\pi t), C \neq 0\}$ and $\mathbf{V}(1, 1) = \{Ct(1-t), C \neq 0\}$. Formulas (23) – (24) unify and extend these two cases.

In view of Proposition 5, and remark following Lemma 2, a corollary is

Theorem 1. Assume $\nu_1, \nu_2 \in (0, 1)$. The omega-square statistic $n\omega_n^2(\nu_1, \nu_2)$ defined by (4) – (5) is locally asymptotically optimal, in the sense of Bahadur, for testing $H_0 : f = f_0(\cdot)$, versus $H_1 : f = f_0(\theta + \cdot)$, $\theta \neq 0$, if and only if $f_0 = f_{\nu_1, \nu_2}$.

Furthermore under this assumption the LAO condition (22) holds with explicit values given by (19), (20) and (23).

The following table completes Table 3 of §2.3 p. 80 in [13]. It gives the local index (see [13] p. 79 for details)

$$\lambda_0(\nu_1, \nu_2) \int_{-\infty}^{\infty} q_{\nu_1, \nu_2} \{F_0(x)\} f_0(x)^3 dx$$

of our statistics in the goodness-of-fit problem 6 – 7 for various densities $f_0 = f_{\nu_1, \nu_2}$.

Statistics	$f_{1/4, 3/4}$	$f_{1/4, 1/4}$	$f_{1/2, 1/2}$	$f_{3/4, 3/4}$	$f_{1, 1}$	Gaussian	Cauchy
$\omega_{1, 1}^2 = A_n^2$	0.323	0.155	0.494	0.899	<i>0.333</i>	<i>0.961</i>	<i>0.335</i>
$\omega_{3/4, 3/4}^2$	0.320	0.159	0.498	0.900	0.330	0.942	0.350
$\omega_{1/2, 1/2}^2 = W_n^2$	0.311	0.163	0.500	0.896	<i>0.328</i>	<i>0.907</i>	<i>0.375</i>
$\omega_{1/4, 1/4}^2$	0.286	0.167	0.491	0.865	0.314	0.831	0.416
$\omega_{1/4, 3/4}^2$	0.375	0.130	0.414	0.754	0.277	0.807	0.277

Table 1: Bahadur local indices for location alternatives

The italicized values can also be found in Table 3 of §2.3 of [13]. The framed values are the maximum of the columns and their row gives the most efficient statistic in the case of the corresponding distribution. The first five columns confirm the optimality of ω_{ν_1, ν_2}^2 provided $f_0 = f_{\nu_1, \nu_2}$. For the Gaussian distribution the efficiency increases with (ν_1, ν_2) and A_n^2 remains the most efficient, even when compared with other usual statistics in [13]. Interestingly, it appears that for the Cauchy distribution the efficiency decreases with (ν_1, ν_2) so that $\omega_{1/4, 1/4}^2$ has the best local index. The asymmetric weight function associated with $\omega_{1/4, 3/4}^2$ performs poorly in the case of symmetric distributions.

4. Useful technical results

Lemma 3. Assume $\nu_1, \nu_2, \theta_0 > 0$. There exists a constant $A = A(\nu_1, \nu_2, \theta_0)$ such that: $\forall \theta \in (-\theta_0, \theta_0), \forall x \in \mathbb{R} : f_{\nu_1, \nu_2}(x + \theta) \leq A(\nu_1, \nu_2, \theta_0) f_{\nu_1, \nu_2}(x)$.

Proof. The equality $\tanh(x + \theta) = (\tanh x + \tanh \theta) / (1 + \tanh x \tanh \theta)$ yields relations $[1 \pm \tanh(x + \theta)] / [1 \pm \tanh x] = [1 \pm \tanh \theta] / [1 + \tanh x \tanh \theta] \leq 2 / (1 - \tanh \theta_0)$ and the result follows from (10). \square

Lemma 4. Let \mathcal{L} be the real vector space generated by the set of functions $\{f_{\nu_1, \nu_2} : \nu_1, \nu_2 > 0\}$. If $\nu_1, \nu_2 > 0$, then $f'_{\nu_1, \nu_2}, f''_{\nu_1, \nu_2} \in \mathcal{L}$.

Proof. Differentiation gives $f'_{\nu_1, \nu_2}(x) = 2f_{\nu_1, \nu_2}(x)[\nu_1 e^{-x} - \nu_2 e^x] / [e^x + e^{-x}]$ and it is easily seen from (10) that the latter function is a linear combination of f_{ν_1, ν_2+1} and f_{ν_1+1, ν_2} . Thus $f'_{\nu_1, \nu_2} \in \mathcal{L}$. This first result valid for any $\nu_1, \nu_2 > 0$ implies in turn $f''_{\nu_1, \nu_2} \in \mathcal{L}$. \square

Lemma 5. One has

$$\frac{1 + \nu_1 + \nu_2}{4\nu_1\nu_2} \frac{[f'_{\nu_1, \nu_2}]^2}{f_{\nu_1, \nu_2}} = \frac{\nu_1(1 + \nu_2)f_{\nu_1, \nu_2+2} - 2\nu_1\nu_2f_{\nu_1+1, \nu_2+1} + \nu_2(1 + \nu_1)f_{\nu_1+2, \nu_2}}{\nu_1 + \nu_2}. \quad (25)$$

Proof. From $f'_{\nu_1, \nu_2}(x) = 2f_{\nu_1, \nu_2}(x)[\nu_1e^{-x} - \nu_2e^x]/[e^x + e^{-x}]$, we infer

$$\begin{aligned} \frac{f'_{\nu_1, \nu_2}(x)^2}{f_{\nu_1, \nu_2}(x)} &= \frac{8}{B(\nu_1, \nu_2)} \frac{(\nu_1e^{-x} - \nu_2e^x)^2 e^{(\nu_1 - \nu_2)x}}{(e^x + e^{-x})^{\nu_1 + \nu_2 + 2}} \\ &= \frac{8}{B(\nu_1, \nu_2)} \frac{\nu_1^2 e^{(\nu_1 - \nu_2 - 2)x} + \nu_2^2 e^{(\nu_1 - \nu_2 + 2)x}}{(e^x + e^{-x})^{\nu_1 + \nu_2 + 2}}, \\ \frac{B(\nu_1, \nu_2)}{8} \frac{f'_{\nu_1, \nu_2}(x)^2}{f_{\nu_1, \nu_2}(x)} &= \frac{\nu_1^2 B(\nu_1, \nu_2 + 2)}{2} f_{\nu_1, \nu_2+2}(x) + \frac{\nu_2^2 B(\nu_1 + 2, \nu_2)}{2} f_{\nu_1+2, \nu_2}(x) \\ &\quad - \nu_1\nu_2 B(\nu_1 + 1, \nu_2 + 1) f_{\nu_1+1, \nu_2+1}(x). \end{aligned}$$

Taking into account the identities $B(a+1, b+1) = abB(a, b)/[(a+b)(a+b+1)]$, and $B(a+1, b) = aB(a, b)/(a+b)$, this relation becomes

$$\begin{aligned} \frac{f'_{\nu_1, \nu_2}(x)^2}{4f_{\nu_1, \nu_2}(x)} &= \frac{\nu_1^2\nu_2(1 + \nu_2)f_{\nu_1, \nu_2+2}(x) + \nu_2^2\nu_1(1 + \nu_1)f_{\nu_1+2, \nu_2}(x)}{(\nu_1 + \nu_2)(1 + \nu_1 + \nu_2)} \\ &\quad - \frac{2(\nu_1\nu_2)^2 f_{\nu_1+1, \nu_2+1}(x)}{(\nu_1 + \nu_2)(1 + \nu_1 + \nu_2)}, \end{aligned}$$

and (25) readily follows. \square

References

- [1] Abramowitz, M., Stegun, I. A., and Romer, R. H. (1988). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. American Association of Physics Teachers.
- [2] Bahadur, R. R. (1960). On the asymptotic efficiency of tests and estimates. *Sankhyā: The Indian Journal of Statistics*, pages 229–252.
- [3] Bahadur, R. R. (1967). Rates of convergence of estimates and test statistics. *The Annals of Mathematical Statistics*, 38(2):303–324.

- [4] Bahadur, R. R. (1971). *Some limit theorems in statistics*, volume 4. siam.
- [5] Bahadur, R. R. et al. (1960). Stochastic comparison of tests. *Annals of Mathematical Statistics*, 31(2):276–295.
- [6] Csörgö, M. and Horváth, L. (1993). *Weighted approximations in probability and statistics*. J. Wiley & Sons.
- [7] Faliva, M. and Zoia, M. G. (2017). A distribution family bridging the gaussian and the laplace laws, gram-charlier expansions, kurtosis behaviour, and entropy features. *Entropy*, 19(4):149.
- [8] Groeneboom, P. and Oosterhoff, J. (1977). Bahadur efficiency and probabilities of large deviations. *Statistica neerlandica*, 31(1):1–24.
- [9] Johnson, N., Kotz, S., and Balakrishnan, N. (1994). Continuous univariate distributions (volume 2).
- [10] Kalbfleisch, J. D. and Prentice, R. L. (2011). *The statistical analysis of failure time data*, volume 360. John Wiley & Sons.
- [11] Kravchuk, O. (2005). Rank test of location optimal for hyperbolic secant distribution. *Communications in Statistics-Theory and Methods*, 34(7):1617–1630.
- [12] Kullback, S. (1959). *Information theory and statistics*. John Wiley & Sons.
- [13] Nikitin, Y. (1995). *Asymptotic efficiency of nonparametric tests*. Cambridge University Press.
- [14] Parzen, E. (1980). Quantile functions, convergence in quantile, and extreme value distribution theory. Technical report.
- [15] Prentice, R. L. (1976). A generalization of the probit and logit methods for dose response curves. *Biometrics*, pages 761–768.
- [16] Savage, I. R. (1969). Nonparametric statistics: a personal review. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 107–144.
- [17] Shorack, G. R. and Wellner, J. A. (2009). *Empirical processes with applications to statistics*. SIAM.