Statistics and Probability Letters 117 (2016) 136–143

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Local efficiency of integrated goodness-of-fit tests under skew alternatives

A. Durio [a], Ya.Yu. Nikitin [b,c,*]

[a] Department of Economics and Statistics "S. Cognetti de Martiis", University of Turin, Lungo Dora Siena 100/A, 10153, Torino, Italy

[b] Department of Mathematics and Mechanics, Saint Petersburg State University, 7/9 Universitetskaya nab., St. Petersburg, 199034, Russia

[c] National Research University — Higher School of Economics, Souza Pechatnikov, 16, St.Petersburg 190008, Russia

A R T I C L E I N F O

Article history:
Received 22 February 2016
Received in revised form 21 May 2016
Accepted 21 May 2016
Available online 27 May 2016

MSC:
62G10
62G20
62G30

Keywords:
Integrated empirical process
Bahadur efficiency
Skew alternative

A B S T R A C T

The efficiency of distribution-free *integrated* goodness-of-fit tests was studied by Henze and Nikitin (2000, 2002) under location alternatives. We calculate local Bahadur efficiencies of these tests under more realistic generalized skew alternatives. They turn out to be unexpectedly high.

1. Introduction

Goodness-of-fit testing is one of the most important problems in Statistics. If the hypothetical distribution is continuous, one can apply distribution-free tests based on functionals of the empirical process. Most known tests of such type are the Kolmogorov and Cramér–von Mises tests and their variants, see, e.g., Shorack and Wellner (1986) and Nikitin (1995).

In search of new distribution-free tests with possibly better efficiency properties, Henze and Nikitin (2000, 2002) proposed new test statistics based on the *integrated* empirical process. They found their limiting distributions and calculated local Bahadur efficiencies for location alternatives. These efficiencies are comparable with the efficiencies of usual distribution-free tests, but there exist also some interesting distinctions in favor of these new tests. Gradually statistical inference using integrated empirical processes becomes quite popular, see, e.g., Alvarez-Andrade et al. (2015), Bouzebda and El Faouzi (2012), Jing and Wang (2006), and Kuriki and Hwang (2013).

However, the location alternative is a simplest alternative which is not very realistic in practice, particularly because it preserves the symmetry of the underlying distribution. In many situations it is more reasonable to assume asymmetric alternative models. The most interesting and simple example of such alternative models in the case of normal distribution

* Corresponding author at: Department of Mathematics and Mechanics, Saint Petersburg State University, 7/9 Universitetskaya nab., St. Petersburg, 199034, Russia.

E-mail address: yanikit47@mail.ru (Y.Y. Nikitin).

http://dx.doi.org/10.1016/j.spl.2016.05.016
0167-7152/

Table 1
Local Bahadur efficiencies under skew alternatives.

Statistic	Distribution				
	Gauss	Logistic	Arcsine	Uniform	Student-5
D_n	0.637	0.584	0.810	0.750	0.540
ω_n^1	0.955	0.912	0.985	1	0.862
ω_n^2	0.907	0.855	1	0.987	0.802
U_n^2	0.486	0.420	0.662	0.658	0.373
$\bar{D}_n$	0.955	0.912	0.985	1	0.862
$\bar{\omega}_n^1$	0.895	0.855	0.924	0.938	0.808
$\bar{\omega}_n^2$	0.912	0.866	0.963	0.968	0.816
$\bar{U}_n^2$	0.900	0.846	1	0.986	0.792

was introduced in Azzalini (1985). Let Φ and φ denote the distribution function and the density of the standard normal law. Azzalini proposed the skew-normal distribution depending on the real parameter θ and having the density

$$g(x,\theta) = 2\varphi(x)\Phi(\theta x), \quad x \in \mathbb{R},\ \theta \geq 0.$$

It is evident that for any θ the function $g(x,\theta)$ is a density and that for $\theta = 0$ we get the standard normal density. Later the properties of Azzalini's skew-normal model and its generalizations were considered in numerous papers. Finally they were described and collected in Azzalini (2014).

For any symmetric distribution function F with the density f and any symmetric distribution function G with the density g we can consider the *generalized skew distribution* with the density

$$h(x,\theta) = 2f(x)G(\theta x), \quad x \in \mathbb{R},\ \theta \geq 0. \tag{1}$$

Note that this model is more general than that considered in Durio and Nikitin (2002, 2003) in view of the emergence of almost arbitrary distribution function G instead of initial distribution function F. This model is described and advocated in Azzalini (2014).

It is quite interesting to calculate the efficiencies of integrated distribution-free tests mentioned above under the generalized skew alternative (1). We select the Bahadur efficiency as it is well-adapted for such calculations while other types of efficiencies such as Pitman, Chernoff or Hodges–Lehmann are not applicable or do not discriminate between two-sided tests. See Nikitin (1995) for details concerning the calculation of efficiencies and their interrelations.

The calculation of local Bahadur efficiency of common distribution-free tests under skew alternatives was performed in Durio and Nikitin (2002, 2003). In the present paper we calculate the efficiencies of the *integrated* tests under the more general alternative (1).

General expressions for local Bahadur efficiencies in case of one-parameter families of alternatives can be found in Nikitin (1995). However we cannot apply them as the alternative (1) requires some additional analysis. This analysis was partially done in Durio and Nikitin (2002, 2003). We use corresponding results in Sections 2 and 3 when calculating the efficiencies for five examples of symmetric distributions with different tail behaviors. These efficiencies are taken together in Table 1 of Section 4. They demonstrate that the efficiencies of integrated tests are appreciably higher than of usual tests. Section 5 is devoted to the analysis of local optimality of tests under consideration.

2. Tests based on integrated empirical process

Let $X_1, \dots, X_n$ be a random sample from the density $h(x,\theta)$ given by (1) and depending on the known symmetric density f and symmetric distribution function G, and a real parameter $\theta \geq 0$. Let

$$H(x,\theta) = 2\int_{-\infty}^{x} f(u)G(\theta u)du, \quad x \in \mathbb{R},\ \theta \geq 0, \tag{2}$$

be the distribution function corresponding to this density. We want to test the goodness-of-fit hypothesis $H_0 : \theta = 0$ against the alternative $H_1 : \theta > 0$. Let F_n be the empirical distribution function based on the sample $X_1, \dots, X_n$.

Some well-known goodness-of-fit tests are based on the Kolmogorov statistic

$$D_n = \sqrt{n}\sup_t |F_n(t) - F(t)|,$$

on the Chapman–Moses statistic

$$\omega_n^1 = \sqrt{n}\int_{\mathbb{R}} (F_n(t) - F(t))dF(t),$$

on the Cramér–von Mises statistic

$$\omega_n^2 = n\int_{\mathbb{R}} (F_n(t) - F(t))^2\, dF(t),$$