Journal of Statistical Planning and Inference 140 (2010) 3017–3025

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

On Greenwood goodness-of-fit test

Sherzod M. Mirakhmedov

GIK Institute of Engineering Sciences & Technology, Pakistan

ARTICLE INFO

Article history:
Received 23 February 2009
Received in revised form
31 March 2010
Accepted 7 April 2010
Available online 13 April 2010

Keywords:
Greenwood statistic
Goodness-of-fit
Spacings
Asymptotic efficiency
Large deviation probabilities
Gamma distribution

MSC:
62G10
62G20
60F10

ABSTRACTS

The Greenwood test is based on the sum of squares of disjoint spacings. For the goodness-of-fit problem within the class of symmetric tests, the Greenwood test is known to be optimal in terms of the Pitman asymptotic efficiency (AE), whereas it is much inferior to the log-spacings test in terms of the Bahadur AE. In this paper we extend these properties of Greenwood test showing that it remains optimal in terms of the weak intermediate and intermediate AE, but it is much inferior to those tests satisfying the Cramer condition in terms of the strong intermediate AE.

1. Introduction

Let $X_{1,n} \le X_{2,n} \le \ldots \le X_{n-1,n}$ be the order statistics of a sample of size $n-1$ drawn from a continuous distribution concentrated on the interval [0,1]. The disjoint s-spacings are defined by $W_{k,n}^{(s)} = X_{ks,n} - X_{(k-1)s,n}$, $k=1,2,\ldots,N'$, $W_{N'+1,n}^{(s)} = 1 - X_{N's,n}$, with notation $X_{0,n}=0$ and $X_{n,n}=1$ and $N' = \lfloor n/s \rfloor$ is the largest integer that does not exceeds n/s. We assume that s may increase together with n and $s=o(n)$.

The goodness-of-fit problem is to test if the sample drawn from Uniform [0,1] distribution. A large class of the tests of uniformity based on spacings has the general form

$$S_N^f = \sum_{k=1}^{N} f(nW_{k,n}^{(s)}), \tag{1.1}$$

where $N = \lceil n/s \rceil$ is the smallest integer greater than or equal to n/s and $f(u)$ is a real-valued function. A test based on the statistic S_N^f is called, for brevity, f-test. Three most popular cases of S_N^f are

$$G_N^2 = \sum_{k=1}^{N} (nW_{k,n}^{(s)})^2,$$

E-mail address: shmirakhmedov@yahoo.com

0378-3758/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2010.04.002

called Greenwood's statistic

$$L_N = \sum_{k=1}^{N} \log(nW_{k,n}^{(s)}),$$

called log-spacings statistic, and

$$R_N = \sum_{k=1}^{N} \left| nW_{k,n}^{(s)} - s \right|,$$

called Rao's statistic.

Considerable attention has been devoted in literature to statistics of type (1.1). For detailed information, applications, and references, consult, the papers by Greenwood (1946), Darling (1953), Le Cam (1958), Pyke (1965, 1972), Rao (1972), Rao and Sethuraman (1975), Deheuvels (1985), Jammalamadaka et al. (1989), Zhou and Jammalamadaka (1989), Ghosh and Jammalamadaka (2000), Jammalamadaka and Goria (2004), Mirakhmedov (2005, 2006), Gatto and Jammalamadaka (2006), Mirakhmedov and Naeem (2008a, b) and Mirakhmedov et al. (2009).

We are interested here in asymptotic efficiencies (AE) of *f*-tests, giving most attention to the Greenwood test. There are several concepts of AE, which differ by conditions imposed on the asymptotic behavior of the size α_n, the power β_n and the alternative H_1. The most celebrated is Pitman concept, in which we deal with a sequence of alternatives $H_{1,n}$ converging to H_0 at a rate necessary to keep $\alpha_n \to \alpha$, $\beta_n \to \beta$, $0<\alpha<\beta<1$, as $n\to\infty$. This is a family of Pitman alternatives (we denote it as P_{alt}). Under the Bahadur concept, a family of fixed alternatives (more precisely, the alternatives that do not approach the hypothesis) B_{alt} is considered; it is supposed that the asymptotic power $\beta<1$ and, as consequence, AE of a test is characterized by the *exponential* rate of decrease of the size α_n. Thus, the Pitman and Bahadur concepts apply one of the following principles: move the alternatives steadily closer to the hypothesis (the Pitman concept) or decrease the size of a test to zero (the Bahadur concept). The intermediate AE concept, as defined by Kallenberg (1983), (see also Ivchenko and Mirakhmedov, 1995; Inglot, 1999), applies both principles simultaneously: it is assumed that $\beta_n \to \beta$, $0<\beta<1$, while simultaneously $H_{1,n}\to H_0$ (more slowly than in the Pitman case) and$\alpha_n \to 0$ (more slowly than in the Bahadur approach). This intermediate approach is somewhere between of Pitman and Bahadur approaches of efficiency.

AE of *f*-tests was studied by several authors. It has been proved that the Pitman AE of *f*-test depends on an asymptotic correlation $\mu(s,f)$(see below, relation (2.6)) of the statistic S_N^f with Greenwood statistic (see Del Pino, 1979; Rao and Kuo, 1984; Jammalamadaka and Tiwari,1987; Jammalamadaka et al., 1989; Mirakhmedov and Naeem, 2008b) and hence, Greenwood test is optimal in terms of Pitman AE. Zhou and Jammalamadaka (1989) have shown that Greenwood test is much inferior to the log-spacings test in terms of Bahadur AE. Finally, Mirakhmedov (2006) and Mirakhmedov and Naeem (2008b) show that the intermediate AE of *f*-tests satisfying Cramér condition $E\exp\{H|f(Z)|\}<\infty$, $\exists H>0$, where Z is a random variable with p.d.f. $\gamma_s(u)$=$u^{s-1}e^{-u}/\Gamma(s)$, also depend on $\mu(s,f)$. Note that the log-spacings and Rao's statistics satisfy the Cramér condition, whereas Greenwood statistic does not. Thus, the intermediate AE of Greenwood test has been an open problem. In the present paper we address that problem.

The concept of intermediate AE is based on moderate and large deviation theorems for the test statistics. Such theorems for the statistics S_N^f are presented in Section 3, whereas the main results on AE are discussed in Section 2. For the reader's convenience, the auxiliary assertions are collected in Appendix.

In what follows, C_i is a positive universal constant, and all asymptotic relations are considered as $n\to\infty$.

2. Asymptotic efficiency

Consider testing the hypothesis H_0 of uniformity of the $n-1$ independent observations on [0,1]. Under the alternative hypothesis H_1, we specify the p.d.f. to be

$$p_n(x) = 1 + dl(x)\delta(n), \quad 0 \le x \le 1, \tag{2.1}$$

where $\delta(n)\to 0$ as $n\to\infty$, $d>0$ is a constant and $l(x)$is a function such that

$$\int_0^1 l(x)dx = 0, \quad \int_0^1 l^2(x)dx = 1.$$

Assume that the large values of the statistics of type (1.1) reject H_0. Let P_i, E_i, Var_i and $corr_i$ be the probability, expectation , variance and correlation coefficient, respectively, under H_i; $A_{i,N}$ stand for the asymptotic value of $N^{-1}E_iR_N^f$, i=0,1, and Z stands for a random variable (r.v.) with p.d.f. $\gamma_s(u)$=$u^{s-1}e^{-u}/\Gamma(s)$.

Under the alternative (2.1) (see, for example, Mirakhmedov and Naeem, 2008b)

$$Var_1 S_N^f = Var_0 S_N^f(1+o(1)) = N\sigma^2(f)(1+o(1)),$$

$$x_N(f) \stackrel{def}{=} \sqrt{N}(A_{1N}(f) - A_{0N}(f))/\sigma(f) = \sqrt{\frac{n(s+1)}{2}}\delta^2(n)\mu(s,f)d^2(1+o(1)), \tag{2.2}$$