



Working Paper 08-42
Statistics and Econometrics Series 11
September 2008

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

**ASYMPTOTIC PROPERTIES OF A GOODNESS-OF-FIT TEST
BASED ON MAXIMUM CORRELATIONS**

Aurea Grané⁽¹⁾ and Anna V. Tchirina⁽²⁾

Abstract

We study the efficiency properties of the goodness-of-fit test based on the Q_n statistic introduced in Fortiana and Grané (2003) using the concepts of Bahadur asymptotic relative efficiency and Bahadur asymptotic optimality. We compare the test based on this statistic with those based on the Kolmogorov-Smirnov, the Cramér-von Mises and the Anderson-Darling statistics. We also describe the distribution families for which the test based on Q_n is asymptotically optimal in the Bahadur sense and, as an application, we use this test to detect the presence of hidden periodicities in a stationary time series.

Keywords: Bahadur asymptotic relative efficiency; goodness-of-fit; local asymptotic optimality; L -statistics; maximum correlation.

AMS subject classification: 62G30, 62G20.

(1) A. Grané, Statistics Department, Universidad Carlos III de Madrid, C/ Madrid 126, 28903 Getafe (Madrid), Spain, e-mail: aurea.grane@uc3m.es

(2) A. V. Tchirina, St. Petersburg Electrotechnical University, 5 Professor Popov str., St. Petersburg 197376, Russia, e-mail: antch@mail.ru

Work partially supported by Spanish grants SEJ2007-64500 and MTM2006-09920 (Ministry of Education and Science- FEDER) and by the Russian Fund for Basic Research RFBR 07-01-00159.

Asymptotic properties of a goodness-of-fit test based on maximum correlations

Aurea Grané⁽¹⁾ and Anna V. Tchirina⁽²⁾

⁽¹⁾ *Statistics Department. Universidad Carlos III de Madrid.*

⁽²⁾ *St. Petersburg Electrotechnical University.*

Abstract: We study the efficiency properties of the goodness-of-fit test based on the Q_n statistic introduced in Fortiana and Grané (2003) using the concepts of Bahadur asymptotic relative efficiency and Bahadur asymptotic optimality. We compare the test based on this statistic with those based on the Kolmogorov-Smirnov, the Cramér-von Mises and the Anderson-Darling statistics. We also describe the distribution families for which the test based on Q_n is locally asymptotically optimal in the Bahadur sense and, as an application, we use this test to detect the presence of hidden periodicities in a stationary time series.

Keywords: Bahadur asymptotic relative efficiency; goodness-of-fit; local asymptotic optimality; L-statistics; maximum correlation.

AMS subject classification: 62G30, 62G20.

1 Introduction

Let x_1, \dots, x_n be a sample of independent and identically distributed (iid) random variables with cumulative distribution function (cdf) F and let $x_{(1)}, \dots, x_{(n)}$ be the ordered sample. In Fortiana and Grané (2003) we proposed and studied the following L -statistic

$$Q_n = \frac{6}{n^2} \sum_{i=1}^n (2i - n - 1) x_{(i)} \quad (1)$$

as a general-purpose goodness-of-fit test of uniformity, a test which extends to case 0 in Stephens (1986), testing for a fully specified continuous distribution. The statistic (1) was obtained from $\rho^+(F_n, F)$, the Hoeffding's maximum correlation between F and the empirical distribution function F_n ,

$$Q_n = \frac{s_n}{\sqrt{1/12}} \rho^+(F_n, F),$$

Work supported partially by Spanish grants SEJ2007-64500 and MTM2006-09920 (Ministry of Education and Science-FEDER) and by the Russian Fund for Basic Research RFBR 07-01-00159.

Corresponding author: Aurea Grané. Statistics Department, Universidad Carlos III de Madrid, C/ Madrid 126, 28903-Getafe, Spain.

E-mails: aurea.grane@uc3m.es, antch@mail.ru

Date: September 11, 2008.

where s_n is the sample standard deviation, and it is close related to the well-known Gini mean difference statistic. We found that Q_n has good properties as a goodness-of-fit test: since it is an L -statistic, it is possible to find its exact distribution under the null hypothesis for small samples and, additionally, it is asymptotically normally distributed. The test based on Q_n can advantageously replace those of Kolmogorov-Smirnov, Cramér-von Mises and Anderson-Darling for a wide range of alternatives.

In the present article we study efficiency properties of the goodness-of-fit test based on Q_n using the concepts of Bahadur asymptotic relative efficiency and Bahadur asymptotic optimality. In section 2 we describe a set of five parametric families of probability distributions with support on $[0, 1]$, used in Fortiana and Grané (2003) and Grané and Fortiana (2006) as alternatives to the $[0, 1]$ -uniform distribution and illustrate the power of the test based on Q_n . In section 3 we prove a large deviations result for Q_n , which enables us to compute its Bahadur exact slope, and study the its Bahadur asymptotic relative efficiency with respect to the Kolmogorov-Smirnov, Cramér-von Mises and Anderson-Darling statistics, for the set of alternative distributions introduced in Section 2. In section 4 we find domains where Q_n is locally asymptotically optimal in the Bahadur sense. As an application of the test of uniformity based on Q_n , in Section 5 we consider the problem of detecting the presence of hidden periodicities of unspecified frequency in a time series, through the analysis of its cumulative periodogram. We conclude in Section 6.

2 Parametric alternative distributions

In this Section we describe five parametric families of alternative probability distributions with support on $[0, 1]$. We have chosen them so that either the mean or the variance differs from those of the null hypothesis, the uniform distribution, which in each case is obtained for a particular value of the parameter. We will use them in order to compare the asymptotic efficiency of the test based on Q_n with respect to those based on the Kolmogorov-Smirnov, the Cramér-von Mises and the Anderson-Darling statistics. These parametric families are defined by the following cumulative distribution functions:

- A1. Lehmann alternatives. A family of asymmetric distributions which concentrate around 1, for $\alpha > 1$, and around 0, for $\alpha < 1$.

$$F_\alpha(x) = x^\alpha, \quad 0 \leq x \leq 1, \quad \alpha > 0. \quad (2)$$

- A2. Centered distributions having a U-shaped pdf, for $\beta \in (0, 1)$, or wedge-shaped pdf, for $\beta > 1$.

$$F_\beta(x) = \begin{cases} \frac{1}{2} (2x)^\beta, & 0 \leq x \leq 1/2, \\ 1 - \frac{1}{2} (2(1-x))^\beta, & 1/2 \leq x \leq 1. \end{cases} \quad (3)$$

- A3. Compressed uniform alternatives,

$$F_\gamma(x) = \frac{x - \gamma}{1 - 2\gamma}, \quad \gamma \leq x \leq 1 - \gamma, \quad (4)$$

where $0 \leq \gamma \leq 1/2$.

- A4. A bimodal locally uniform distribution, with probability mass concentrated near both extremes, 0 and 1.

$$F_\delta(x) = \begin{cases} x/2\delta, & 0 \leq x \leq \delta, \\ 1/2, & \delta \leq x \leq 1 - \delta, \\ 1 + (x - 1)/2\delta, & 1 - \delta \leq x \leq 1, \end{cases} \quad (5)$$

where $0 \leq \delta \leq 1/2$.

- A5. A locally uniform distribution, mixture of A3 and A4-type distributions.

$$F_{\epsilon,\eta}(x) = 2\eta\epsilon \left[\frac{1}{2\epsilon} \mathbf{1}_{[0,\epsilon] \cup [1-\epsilon,1]}(x) \right] + (1 - 2\eta\epsilon) \left[\frac{1}{1 - 2\epsilon} \mathbf{1}_{[\epsilon,1-\epsilon]}(x) \right], \quad (6)$$

where $0 < \epsilon < \frac{1}{2}$, $0 \leq \eta \leq \frac{1}{2\epsilon}$. The parameter ϵ determines the width of the three intervals (analogous role than γ in (4) and δ in (5)). The parameter $\eta \in (0, \frac{1}{2\epsilon})$ gives the relative proportion of A4, thus when $\eta < 1$ the probability mass is more concentrated around the centre. In particular, this family reduces to A3 for $\eta = 0$, and to A4 for $\eta = \frac{1}{2\epsilon}$.

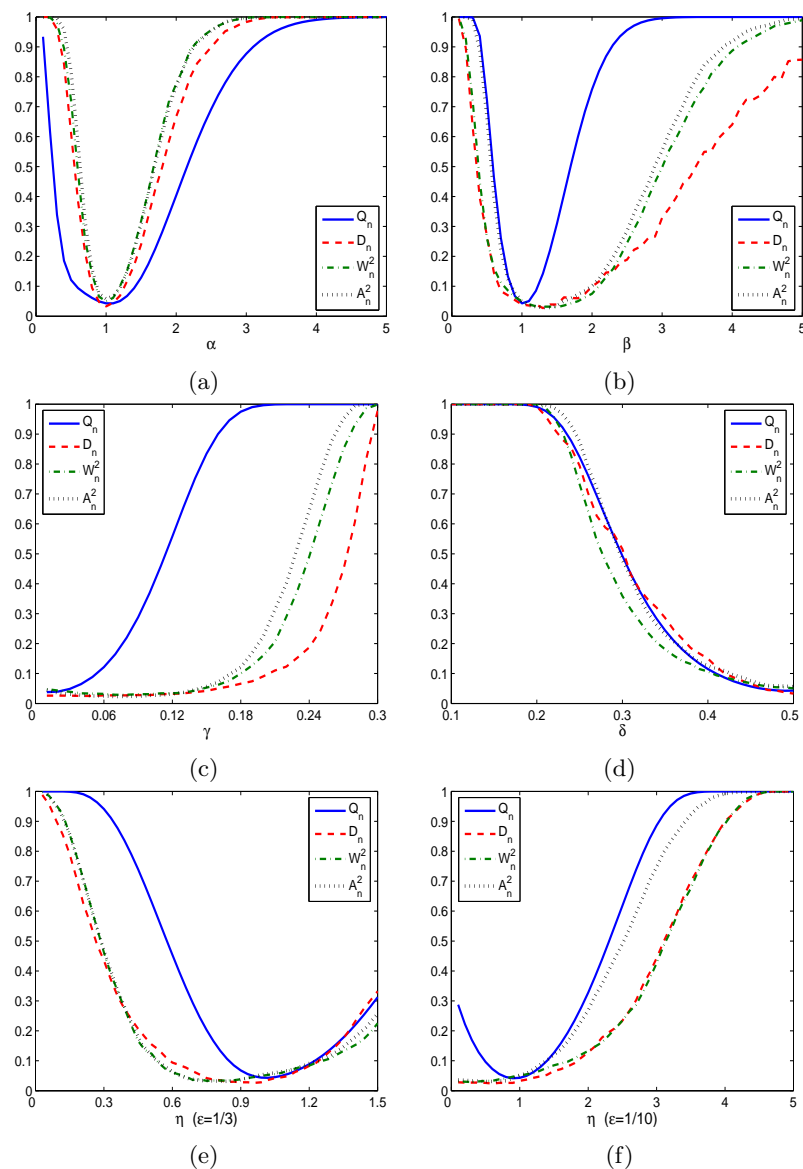
2.1 Power study

As an illustration of the good performance of the test based on Q_n , we depict only the power functions at a 5% significance level and a sample size of $n = 20$. We compare them with those obtained for the tests based on the Kolmogorov-Smirnov statistic, D_n , the Cramér-von Mises statistic, W_n^2 and the Anderson-Darling statistic, A_n^2 . For all of them we have considered the critical regions derived from the corresponding asymptotical distributions that can be found in Kolmogorov (1933) for D_n , Anderson and Darling (1952) for W_n^2 , Anderson and Darling (1954) for A_n^2 and Fortiana and Grané (2003) for Q_n . These powers have been estimated from $N = 1000$ simulated samples of size $n = 20$ as the relative frequency of values of the corresponding statistic in the critical region. For each family we have taken 30 different values of the parameter. Figure 1 shows the power curves. The test based on Q_n is consistent for all the families of alternatives studied.

3 Asymptotic relative efficiency

From the available concepts of asymptotic relative efficiency (ARE), Bahadur's is the most adequate to compare Q_n , the Kolmogorov-Smirnov statistic, D_n , the Cramér-von Mises statistic, W_n^2 , and the Anderson-Darling statistic, A_n^2 , since they have different asymptotic distributions. Thorough treatments of ARE can be found in Serfling (1980) and Nikitin (1995) as well as in the classic Bahadur (1971). For easy reference, we include a summary of the necessary concepts, notation and results.

Figure 1: Power functions at 5% significance level and a sample size $n = 20$ for the tests based on Q_n , D_n , W_n^2 and A_n^2 for (a) A1 family, (b) A2 family, (c) A3 family, (d) A4 family, (e) A5 family with $\epsilon = 1/3$ and (f) A5 family with $\epsilon = 1/10$.



3.1 Some previous concepts and notation

3.1.1 Bahadur exact slope and Bahadur efficiency

We consider a sequence $s = \{x_n\}_{n \in \mathbb{N}}$ of random variables, iid with probability distribution P_θ , where $\theta \in \Theta$ is a parameter. We want to test $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1 = \Theta - \Theta_0$, with the test statistic $\mathbb{T} = \{T_n\}_{n \in \mathbb{N}}$, where for each $n \in \mathbb{N}$, $T_n(s) = T_n(x_1, \dots, x_n)$ is measurable with respect to the natural sigma-algebra of $\mathbb{R}^{\mathbb{N}}$,

with values in some, possibly unbounded, interval $I \subseteq \overline{\mathbb{R}}$. We assume, to simplify, that T_n has the same distribution for each $\theta \in \Theta_0$. Let $C(t) \subset I$, where $t \in I$, be a critical region for this test, for a given n and signification level (e. g., $C(t) = (t, \infty)$ for a one-sided test on $I = \mathbb{R}$), and let

$$p_n(t) = P_\theta(T_n \in C(t)), \quad t \in \mathbb{R}, \quad \theta \in \Theta_0, \quad (7)$$

be the actual size of the test. Bahadur's approach amounts to studying the rate of decrease of p_n when n increases. The p -value, defined as the random variable

$$L_n(s) \equiv L_n(x_1, \dots, x_n) \equiv p_n(T_n), \quad (8)$$

represents the degree to which the test \mathbb{T} rejects H_0 . When $C(t) = (t, +\infty)$ or $C(t) = (-\infty, t)$ and the distribution of T_n is continuous, L_n is uniform in $[0, 1]$ in the null case. For a given $\varepsilon \in (0, 1)$, let $N(\varepsilon)$ be the smallest integer m such that $L_n < \varepsilon$ for all $n \geq m$ and $N(\varepsilon) = \infty$ if no such m exists, i.e., $N(\varepsilon)$ is the smallest sample size required so that the test based on \mathbb{T} becomes and remains significant at level ε . If, for some positive function $c : \Theta_1 \rightarrow \mathbb{R}_+$, the following convergence in probability takes place,

$$\frac{1}{n} \log L_n \xrightarrow[n \rightarrow \infty]{P_\theta} -\frac{1}{2} c(\theta), \quad \theta \in \Theta_1, \quad (9)$$

then c is called the *Bahadur exact slope* of \mathbb{T} . It is related with $N(\varepsilon)$ by

Theorem A (Bahadur 1971, thm. 7.1) If (9) holds and $0 < c(\theta) < \infty$, then

$$\lim_{\varepsilon \rightarrow 0} \frac{N(\varepsilon)}{2 \log(1/\varepsilon)} = \frac{1}{c(\theta)}, \quad \text{a.s. } [P_\theta], \quad \theta \in \Theta_1. \quad (10)$$

If \mathbb{T}_1 and \mathbb{T}_2 are test statistics, and $c_i(\theta)$, $N_i(\varepsilon)$ correspond to \mathbb{T}_i as above,

$$\lim_{\varepsilon \rightarrow 0} \frac{N_2(\varepsilon)}{N_1(\varepsilon)} = \frac{c_1(\theta)}{c_2(\theta)}, \quad \text{a.s. } [P_\theta], \quad \theta \in \Theta_1.$$

This quotient is the *Bahadur efficiency* of the test based on \mathbb{T}_1 relative to the test based on \mathbb{T}_2 , and we will denote it by $\text{ARE}(\theta; \mathbb{T}_1, \mathbb{T}_2)$. \mathbb{T}_1 is more efficient at θ than \mathbb{T}_2 in the sense of Bahadur if $\text{ARE}(\theta; \mathbb{T}_1, \mathbb{T}_2) > 1$. A useful method for computing exact slopes is:

Theorem B (Bahadur 1971, thm. 7.2) If, for some function $b : \Theta_1 \rightarrow I$, the following convergence in probability takes place,

$$T_n \xrightarrow[n \rightarrow \infty]{P_\theta} b(\theta), \quad \text{for each } \theta \in \Theta_1, \quad (11)$$

and, for each $t \in I$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log p_n(t) = f(t), \quad (12)$$

where f is a continuous function on I , then (9) holds with $c(\theta) = -2 f(b(\theta))$.

Usually, $b(\theta)$ is computed from Glivenko-Cantelli theorem, while $f(t)$ depends on the behaviour of the large deviations of \mathbb{T} .

3.2 Results for D_n , W_n^2 and A_n^2

Here we reproduce some results for the Kolmogorov–Smirnov, the Cramér–von Mises and the Anderson–Darling statistics and we compute their Bahadur exact slopes for the parametric alternative distributions A1–A5.

Result 1 The statistic $D_n \equiv \|F_n - F_0\|$, where $\|\cdot\|$ is the supremum norm, verifies hypotheses (11) and (12) of Theorem B with

$$b(\theta) \equiv \|F_\theta - F_0\|$$

and $f(t) = \inf\{g(t, s), 0 \leq s \leq 1\}$, with

$$g(t, s) = \begin{cases} (t+s) \log \frac{t+s}{s} + (1-t-s) \log \frac{1-t-s}{1-s} & 0 \leq s \leq 1-t, \\ \infty & s > 1-t, \end{cases}$$

for $t \in (0, 1)$.

Proof: See Bahadur (1971), Groeneboom and Shorack (1981) and Shorack and Wellner (1986). Our D_n is equivalent to $D_{1,n}$ in Shorack and Wellner’s notation.

A direct application of Glivenko–Cantelli theorem gives the function $b(\theta)$ of (11) and Groeneboom and Shorack (1981) proved the large deviation result. \square

To compute Bahadur’s slope for D_n , we use the approximation $f(t) \approx 2t^2 + O(t^3)$ if $t \rightarrow 0$ and $f(t) = \infty$ if $t \rightarrow 1$ (See lemma 5.1 of Bahadur 1971). Table 1 contains function $c(\theta)$ for D_n for alternatives A1–A5.

Table 1: Bahadur exact slopes of D_n for A1–A5

Alternative	θ	$c(\theta)$
A1	α	$4 \frac{(1/\alpha-1)^2}{\alpha^{2/(\alpha-1)}}$
A2	β	$\frac{(1/\beta-1)^2}{\beta^{2/(\beta-1)}}$
A3	γ	$4\gamma^2$
A4	δ	$4(\frac{1}{2} - \delta)^2$
A5	(ϵ, η)	$4\epsilon^2(\eta - 1)^2$

Result 2 The statistic $W_n^2 = n \int_{-\infty}^{+\infty} (F_n - F_0)^2 dF_0$, verifies hypotheses (11) and (12) of Theorem B with

$$b(\theta) = \int_{-\infty}^{+\infty} (F_\theta - F_0)^2 dF_0$$

and $f(t) = \frac{\pi^2 t}{2} + O(t^2)$.

Proof: See Nikitin (1976, 1979, 1980), Mogul'skii (1977).

Taking $T_n = W_n^2/n$ and applying Glivenko–Cantelli theorem, we get the expression of $b(\theta)$. Mogul'skii (1977) proved that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P(W_n^2 > nt) \approx -\frac{\pi^2 t}{2},$$

where $t = t(n) = o(1)$. Hence, the exact slope of W_n^2 is given by

$$c(\theta) \approx \pi^2 \int_{-\infty}^{+\infty} (F_\theta - F_0)^2 dF_0.$$

□

Table 2 contains function $c(\theta)$ for W_n^2 for alternatives A1–A5.

Table 2: Bahadur exact slopes of W_n^2 for A1–A5

Alternative	θ	$c(\theta)$
A1	α	$\frac{2\pi^2}{3} \frac{(\alpha-1)^2}{(\alpha+2)(2\alpha+1)}$
A2	β	$\frac{\pi^2}{6} \frac{(\beta-1)^2}{(\beta+2)(2\beta+1)}$
A3	γ	$\frac{\pi^2}{3} \gamma^2$
A4	δ	$\frac{\pi^2}{12} (2\delta - 1)^2$
A5	(ϵ, η)	$\frac{\pi^2}{3} \epsilon^2 (\eta - 1)^2$

Result 3 The statistic $A_n^2 = n \int_{-\infty}^{+\infty} \frac{(F_n - F_0)^2}{F_0(1-F_0)} dF_0$, verifies hypotheses (11) and (12) of Theorem B with

$$b(\theta) = \int_{-\infty}^{+\infty} \frac{(F_\theta - F_0)^2}{F_0(1-F_0)} dF_0,$$

and $f(t) = t + O(t)$, $t \rightarrow 0$.

Proof: See Nikitin (1976, 1979, 1980).

Taking $T_n = A_n^2/n$ and applying Glivenko–Cantelli theorem, we get the expression of $b(\theta)$. Nikitin (1979, 1980) proved that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P(A_n^2 > nt) \approx -t, \quad t \rightarrow 0.$$

Hence, the exact slope of A_n^2 is given by

$$c(\theta) \approx 2 \int_{-\infty}^{+\infty} \frac{(F_\theta - F_0)^2}{F_0(1-F_0)} dF_0.$$

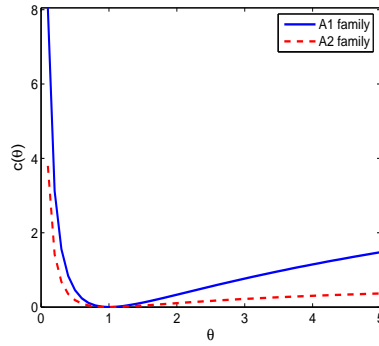
□

Table 3 contains function $c(\theta)$ for A_n^2 for alternatives A3–A5, while for families A1 and A2, it has been numerically computed. Figure 2 contains a graphical representation of $c(\theta)$ for A1 and A2 families.

Table 3: Bahadur exact slopes of A_n^2 for A3–A5

Alternative	θ	$c(\theta)$
A3	γ	$\frac{4}{(2\gamma-1)^2} [\gamma(-1 + 2\gamma - \gamma \log \gamma) + (-1 + 4\gamma - 3\gamma^2) \log(1 - \gamma)]$
A4	δ	$\frac{1}{\delta^2} [\delta(-1 + 2\delta - \delta \log \delta) + (-1 + 4\delta - 3\delta^2) \log(1 - \delta)]$
A5	(ϵ, η)	$\frac{4(\eta-1)^2}{(2\epsilon-1)^2} [\epsilon(-1 + 2\epsilon - \epsilon \log \epsilon) + (-1 + 4\epsilon - 3\epsilon^2) \log(1 - \epsilon)]$

Figure 2: Bahadur exact slope of A_n^2 for A1 and A2 alternatives.



3.3 A large deviations result for the Q_n statistic

Since Q_n is an L -statistic, in order to find its large deviations function under the hypothesis of uniformity, we can use the well-known connection between the uniform distribution and the exponential one. This method is less universal than the Sanov theorem for L -statistics (see Groeneboom and Shorack 1981), but it permits us to obtain a more explicit form of the deviation function. We start with a technical lemma:

Lemma 3.1 *Under the null hypothesis of uniformity, the Q_n statistic defined in (1) can be written as follows:*

$$Q_n \stackrel{d}{=} \sum_{j=1}^n w_{j,n} y_j,$$

where $w_{j,n} = 6n^{-2}(n-j+1)(j-1)$, $y_j = z_j / \sum_{i=1}^{n+1} z_i$ and z_1, \dots, z_{n+1} are iid $\sim \text{Exp}(1)$ random variables.

Proof: Let x_1, \dots, x_n be iid $\sim U[0, 1]$ random variables and let $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ be the ordered sample. For $i = 1, \dots, n+1$, consider the uniform spacings $\delta_i = x_{(i)} - x_{(i-1)}$, where, by convention, $x_{(0)} = 0$ and $x_{(n+1)} = 1$.

Noticing that $x_{(i)} = \sum_{j=1}^i \delta_j$, for $i = 1, \dots, n$, and after regrouping summands Q_n can be written as:

$$Q_n = \frac{6}{n^2} \sum_{i=1}^n (2i - n - 1)x_{(i)} = \frac{6}{n^2} \sum_{j=1}^n (n - j + 1)(j - 1) \delta_j \stackrel{d}{=} \sum_{j=1}^n w_{j,n} y_j,$$

where $w_{j,n} = 6n^{-2}(n - j + 1)(j - 1)$. The last equality is obtained applying Lemma 1 in Dwass (1961) (see also, for example, Azlarov and Volodin 1986 or Chap. 21 in Shorack and Wellner 1986), which asserts that the random vectors (y_1, \dots, y_{n+1}) , where $y_j = z_j / \sum_{i=1}^{n+1} z_i$ with z_1, \dots, z_{n+1} being iid $\sim \text{Exp}(1)$ random variables, and $(\delta_1, \dots, \delta_{n+1})$ have the same distribution. \square

In the following theorem we compute the rough large deviation asymptotics of the centered statistic $|Q_n - 1|$ under uniformity:

THEOREM 1 *Consider the function $W(x) = 6x(1 - x) - 1$, for $0 < x < 1$. Under the hypothesis of uniformity the large deviation functions of $Q_n - 1$ and $1 - Q_n$ may be written in the form*

$$f_R(a) \equiv \lim_{n \rightarrow \infty} n^{-1} \log P_\theta(Q_n - 1 > a) = \inf_{t \geq 0} \int_0^1 \log \frac{1}{1 - t(W(x) - a)} dx,$$

and

$$f_L(a) \equiv \lim_{n \rightarrow \infty} n^{-1} \log P_\theta(1 - Q_n > a) = \inf_{t \geq 0} \int_0^1 \log \frac{1}{1 - t(-W(x) - a)} dx,$$

$f_R(a)$ and $f_L(a)$ being continuous for $0 < a < 0.5$ and $0 < a < 1$, respectively, and having the following asymptotics as $a \rightarrow 0+$:

$$f_{R,L}(a) = - \left(2 \int_0^1 W^2(x) dx \right)^{-1} a^2(1 + o(1)) = -\frac{5}{2} a^2(1 + o(1)).$$

Finally,

$$f(a) \equiv \lim_{n \rightarrow \infty} n^{-1} \log P_\theta(|Q_n - 1| > a) = \max(f_R(a), f_L(a)) = -\frac{5}{2} a^2(1 + o(1)).$$

Proof: Since the Q_n statistic has the same distribution as a quotient of linear combinations of iid Exponential random variables (see Lemma 3.1), the proof follows by arguing as in the proof of the Theorem 2 of Tchirina (2007) and applying the Plachky–Steinebach theorem (Plachky and Steinebach 1975, see also Shorack and Wellner 1986). \square

The next step is to obtain the Bahadur exact slope of Q_n . This is stated in the following proposition:

Proposition 3.1 *The centered statistic $|Q_n - 1|$ verifies hypotheses (11) and (12) of Theorem B with*

$$b(\theta) = \left| \int_0^1 J(t) F_\theta^-(t) dt - 1 \right| \quad \text{and} \quad f(t) = -\frac{5}{2} t^2(1 + o(1)), \quad 0 \leq t \leq 3/2,$$

where $J(t) = 6(2t - 1)$ and F_θ^- is the pseudo-inverse of the distribution function F_θ . Then the Bahadur exact slope of $|Q_n - 1|$ is

$$c(\theta) \approx 5 \left(\int_0^1 J(t) F_\theta^-(t) dt - 1 \right)^2.$$

Proof: Let T be a functional defined on the set of distribution functions $\{F_\theta, \theta \in \Theta\}$, $T(F_\theta) = \int_0^1 J(t) F_\theta^-(t) dt$, where $J(t) = 6(2t - 1)$ and F_θ^- is the pseudo-inverse of F_θ . Then we can write Q_n as the result of applying T to the empirical distribution function,

$$Q_n \equiv T(F_n) = \int_0^1 J(t) F_n^-(t) dt,$$

whenever the integral exists. Since J is a function of bounded variation in $(0, 1)$, we can apply Theorem (19.1.3) in Shorack and Wellner (1986), which asserts the almost sure convergence of $|Q_n - 1|$ to $b(\theta) \equiv |T(F_\theta) - 1|$, when $n \rightarrow \infty$, for $\theta \in \Theta_1$, which proves (11). Function $f(t)$ of (12) is given by Theorem 1 above and finally, applying Theorem B, the Bahadur exact slope of $|Q_n - 1|$ is $c(\theta) \approx 5(T(F_\theta) - 1)^2$. \square

Function $c(\theta)$ for $|Q_n - 1|$ for alternatives A1–A5 is given in Table 4.

Table 4: Bahadur exact slopes of $|Q_n - 1|$ for A1–A5

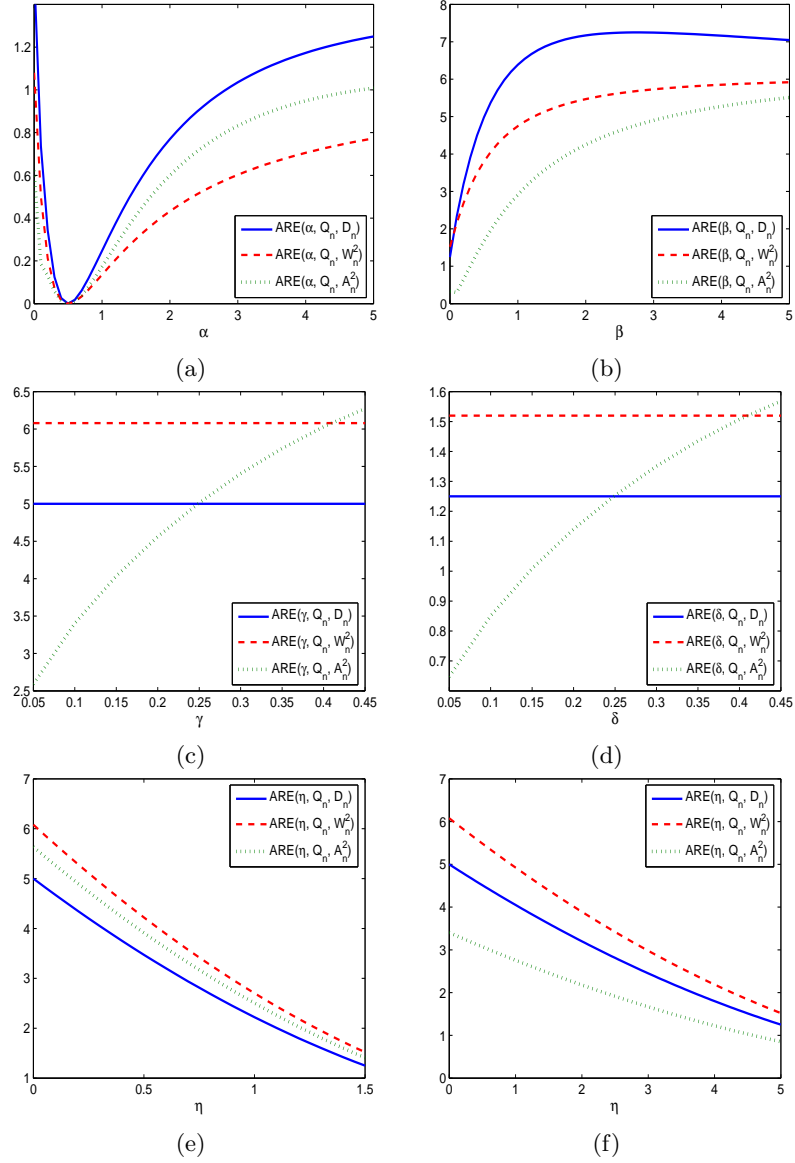
Alternative	θ	$c(\theta)$
A1	α	$5 \left[\frac{6\alpha}{(1+\alpha)(1+2\alpha)} - 1 \right]^2$
A2	β	$5 \left[\frac{3}{2} \frac{3\beta+1}{(1+\beta)(1+2\beta)} - 1 \right]^2$
A3	γ	$20\gamma^2$
A4	δ	$5\left(\frac{1}{2} - \delta\right)^2$
A5	(ϵ, η)	$20\epsilon^2(\eta - 1)^2(1 - \epsilon\eta)^2$

3.4 Bahadur efficiency of Q_n relative to D_n , W_n^2 and A_n^2

In order to compare the efficiency of Q_n with respect to D_n , W_n^2 and A_n^2 we will use the concept of asymptotic relative efficiency (ARE), introduced in Section 3. We will say that Q_n is more efficient (in the Bahadur sense) than the corresponding statistic being considered if the quotient of their exact slopes is greater than 1. Since these quotients can be easily obtained from Tables 1–4, here we prefer to illustrate the behaviour of Q_n through Figure 3.

It turns out that, except for the A1 family, the Q_n statistic is more efficient in the Bahadur sense than D_n , W_n^2 and A_n^2 for nearly all the values of the parameters. It is interesting to mention that for A3 and A4 families the ratios between Q_n and D_n or W_n^2 do not depend on the parameters. Note also that for the A3 and A5 families the ratios are always greater than 1.

Figure 3: Bahadur efficiency of $|Q_n - 1|$ relative to D_n , W_n^2 and A_n^2 for (a) A1 family, (b) A2 family, (c) A3 family, (d) A4 family, (e) A5 family with $\epsilon = 1/3$ and (f) A5 family with $\epsilon = 1/10$.



4 Local Bahadur asymptotic optimality

In this section we will describe the domains of local Bahadur asymptotic optimality for the Q_n statistic. We start introducing some general concepts and notation.

4.1 Some previous concepts

Let x_1, \dots, x_n be a sequence of iid random variables with distribution function F_θ having support $(0, 1)$, depending on a parameter θ that takes values in an interval Θ , and let f_θ to be the corresponding density function. Also suppose that $F_\theta(x) = x$ at $(0, 1)$ if and only if $\theta = 0$.

Consider the problem of testing $H_0 : \theta = 0$ vs. $H_1 : \theta \neq 0$ with the aid of a sequence of statistics \mathbb{T} , large values of \mathbb{T} being significant, and let $c_{\mathbb{T}}(\theta)$ be the Bahadur exact slope of \mathbb{T} introduced in Section 3.

Here we are interested in comparing the exact slope of \mathbb{T} with its upper bound in terms of the Kullback-Leibler information numbers given by the Bahadur-Raghavachari inequality (Raghavachari 1970 and Bahadur 1971):

$$c_{\mathbb{T}}(\theta) \leq 2K(\theta) = 2 \int_0^1 f_\theta(x) \log f_\theta(x) dx.$$

If equality takes place for all $\theta \neq 0$, the sequence \mathbb{T} is said to be *asymptotically optimal* in the Bahadur sense. Moreover, the *exact Bahadur efficiency* of \mathbb{T} is defined by the ratio

$$e_{\mathbb{T}}^B(\theta) = \frac{c_{\mathbb{T}}(\theta)}{2K(\theta)}. \quad (13)$$

Notice that $e_{\mathbb{T}}^B(\theta) = 1$ in the case of asymptotical optimality. A significantly less restrictive condition, consists in studying when the right-hand side of (13) tends to 1 as $\theta \rightarrow 0$. This limit yields the *local Bahadur asymptotic efficiency* of \mathbb{T} :

$$e_{\mathbb{T}}^{B,loc}(\theta) = \lim_{\theta \rightarrow 0} \frac{c_{\mathbb{T}}(\theta)}{2K(\theta)}$$

and the class of those distribution families for which $e_{\mathbb{T}}^{B,loc}(\theta) = 1$ will describe the *local Bahadur asymptotic optimality (LAO) domain* for \mathbb{T} .

4.2 LAO domain for Q_n

In order to obtain the local Bahadur asymptotic optimality domain for Q_n we introduce functions

$$H(x) = \frac{\partial}{\partial \theta} F_\theta(x) \Big|_{\theta=0} \quad \text{and} \quad h(x) = \frac{\partial}{\partial \theta} f_\theta(x) \Big|_{\theta=0}$$

and assume the following regularity conditions:

- R1. $h(x) = H'(x)$ for almost all $x \in (0, 1)$,
- R2. $\int_0^1 h^2(x) dx < +\infty$,
- R3. It is possible to change the order of the operations of differentiating in θ at zero and integrating in x .

Lemma 4.1 *For the alternative distributions satisfying conditions R1–R3, the Kullback-Leibler information has the following asymptotical expansion at $\theta = 0$:*

$$K(\theta) = \frac{\theta^2}{2} \int_0^1 h^2(x) dx + o(\theta^2), \quad (14)$$

Proof: To prove this lemma we expand $K(\theta)$ in a Taylor series around $\theta = 0$. The proof is straightforward noticing that $K(0) = 0$, and taking into account condition R3 in differentiating with respect to θ . \square

Remark 1 Note that the parametric families A3, A4 and A5 of Section 2 do not satisfy condition (14). Their Kullback-Leibler numbers are of the order 1 in θ instead of order 2. One can explain this fact by the dependence of their supports on θ . The Bahadur efficiency of our test tends to 0 as θ tends to 0 as well as the efficiencies of the tests based on D_n , A_n^2 and W_n^2 . At the same time, the local comparison of their exact Bahadur slopes still makes sense.

Proposition 4.1 *Under the regularity conditions R1–R3 the local Bahadur asymptotic efficiency of Q_n takes the form*

$$e_{Q_n}^{B,loc} = \frac{\left(\int_0^1 W(x) h(x) dx \right)^2}{\int_0^1 W^2(x) dx \int_0^1 h^2(x) dx}, \quad (15)$$

where $W(x) = 6x(1-x) - 1$.

Proof:

Function $b(\theta)$ of Proposition 3.1 can be written as $b(\theta) = \max\{b_R(\theta), b_L(\theta)\}$, where $b_R(\theta) = \int_0^1 J(t) F_\theta^-(t) dt - 1$ and $b_L(\theta) = 1 - \int_0^1 J(t) F_\theta^-(t) dt$, with $J(t) = 6(2t - 1)$.

Defining function $W(t)$ such that $J(t) = -W'(t)$ and $\int_0^1 W(t) dt = 0$, we obtain $W(t) = 6t(1-t) - 1$. After considering regularity conditions R1 and R3, the Taylor expansions of $b_R(\theta)$ and $b_L(\theta)$ at $\theta = 0$ are:

$$b_R(\theta) = -\theta \int_0^1 W(x) h(x) dx + o(\theta), \quad b_L(\theta) = \theta \int_0^1 W(x) h(x) dx + o(\theta).$$

From Theorem 1 and Lemma 4.1 we have that

$$\frac{c_{Q_n}(\theta)}{2K(\theta)} = \frac{\left(2 \int_0^1 W^2(x) dx \right)^{-1} b^2(\theta)(1 + o(1))}{\frac{1}{2} \theta^2 \int_0^1 h^2(x) dx + o(\theta^2)}. \quad (16)$$

Finally, since $b^2(\theta) = \theta^2 \left(\int_0^1 W(x) h(x) dx \right)^2 + o(\theta^2)$, one gets expression (15), after taking the limit as $\theta \rightarrow 0$ in expression (16). \square

The class of distribution functions following regularity conditions R1–R3 for which equation (15) equals 1 describe the LAO domain of Q_n . From the Cauchy-Bunyakowsky inequality we see that the maximum efficiency is attained if and only if $h(x) = \theta W(x)$, for $\theta \in \mathbb{R}$.

The most natural example of such a family of densities is given by

Ex1. $f_{\theta}(x) = 1 + \theta(6x(1-x) - 1)$, $0 < x < 1$, $-2 \leq \theta \leq 1$.

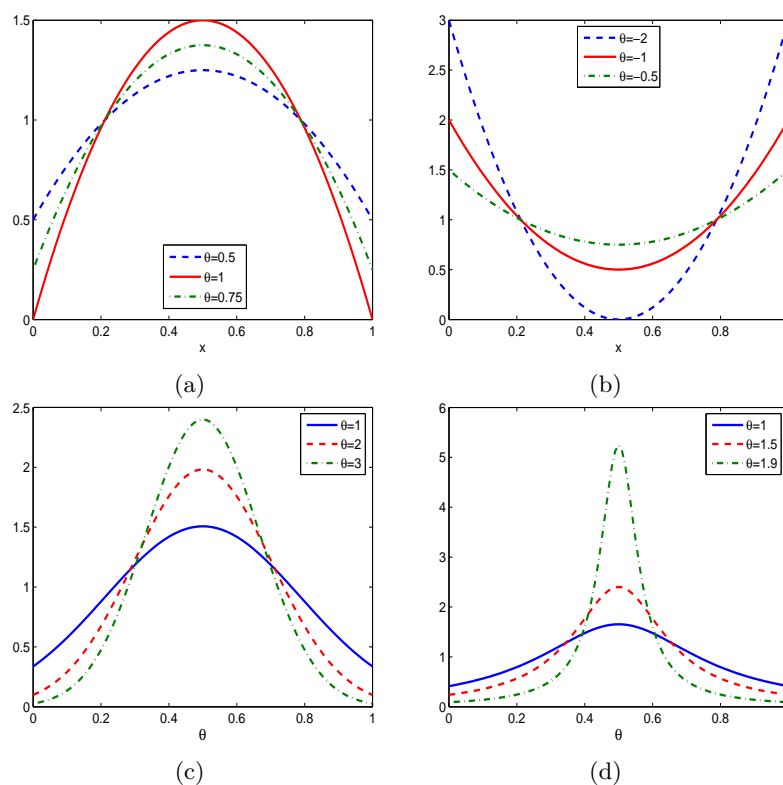
Other examples of families of densities for which Q_n is the most locally efficient statistic in the Bahadur sense are:

Ex2. $f_{\theta}(x) = \frac{\sqrt{6\theta}}{\sqrt{\pi} e^{\theta/2} \operatorname{erf}\left(\sqrt{\frac{3\theta}{2}}\right)} \exp\{\theta(6x(1-x) - 1)\}$, $0 < x < 1$, $\theta > 0$,
 where $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ is the error function,

Ex3. $f_{\theta}(x) = \frac{\sqrt{3(2-\theta)\theta}}{2 \arctan \sqrt{\frac{3\theta}{2-\theta}}} (1 - \theta(6x(1-x) - 1))^{-1}$, $0 < x < 1$, $0 < \theta < 2$.

In Figure 4 we depict those families for different values of the parameter θ .

Figure 4: Examples of families of densities of the LAO domain for Q_n : (a)-(b) Ex1, (c) Ex2 and (d) Ex3.



5 An application to time series analysis

In this section we illustrate with an example a possible application of the test of uniformity based on Q_n in the context of statistical inference for a time series based on the frequency domain properties of the series.

We consider the problem of detecting the presence of hidden periodicities of unspecified frequency in a time series, through the analysis of its cumulative periodogram. In the remaining of this section we will use the notation of Brockwell and Davis (1990).

Let $\{X_t\} \equiv \{X_t, 1 \leq t \leq n\}$ be a real valued stationary time series and $I(w_j)$, with $w_j = 2\pi j/n$, the corresponding periodogram. Proposition 10.2.1 of Brockwell and Davis (1990) asserts that testing that $\{X_t\}$ is Gaussian white noise is equivalent to test that the random variables

$$Y_i = \frac{\sum_{k=1}^i I(w_k)}{\sum_{k=1}^q I(w_k)}, \quad i = 1, \dots, q-1, \quad (17)$$

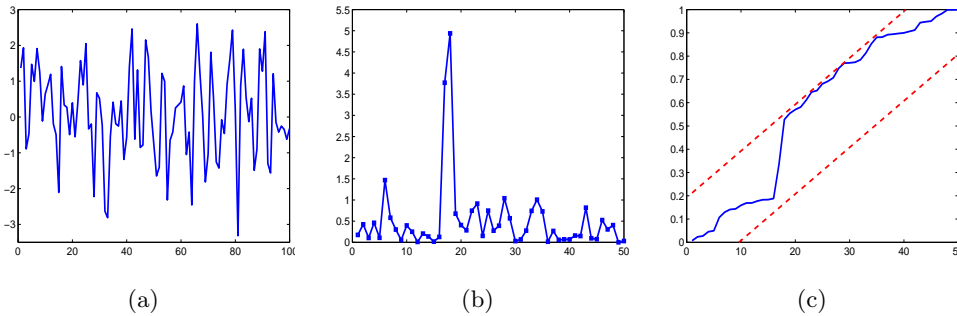
for $q = \lfloor (n-1)/2 \rfloor$, are distributed as the order statistics of $q-1$ iid $\sim U(0,1)$ random variables.

We will use Example 10.2.1 of Brockwell and Davis (1990) to illustrate our purpose. In that example, the data were generated from the following process

$$X_t = \cos(\pi t/3) + Z_t, \quad t = 1, \dots, 100, \quad (18)$$

where $\{Z_t\}$ was Gaussian white noise with variance 1. Now, instead of testing the null hypothesis that $\{X_1, \dots, X_{100}\}$ is generated by a Gaussian white noise sequence, we will consider the test of uniformity based on Q_n for the random variables Y_1, \dots, Y_{49} obtained from equation (17). A realization of this series is depicted in Figure 5. It is clear that the series contains an added deterministic periodic component of unspecified frequency and, hence, the null hypothesis should be rejected. The value of our statistic is $Q_n = 1.1539$, with a p-value of 0.0060, whereas the value of the Kolmogorov-Smirnov statistic is $D_n = 0.2099$. Figure 5-(c) shows the cumulative periodogram and the Kolmogorov-Smirnov boundaries for this series at a 5% significance level, from where we can see that using the test based on D_n we do not reject the null hypothesis even at a 5% significance level. However, the test based on Q_n rejects the null hypothesis at a 1% significance level.

Figure 5: (a) The series $\{X_1, \dots, X_{100}\}$ generated from equation (18), (b) the corresponding periodogram ordinates $I(2\pi j/100)$, $j = 1, \dots, 50$, and (c) the corresponding standardized cumulative periodogram and the Kolmogorov-Smirnov confidential bounds at a 5% significance level



6 Concluding remarks

For the Q_n statistic introduced in Fortiana and Grané (2003) we have proved a large deviations result under the null hypothesis of uniformity and obtained its Bahadur exact slopes for five parametric families A1–A5. We have compared the test of uniformity based on our statistic with those based on the Kolmogorov-Smirnov, D_n , Cramér-von Mises, W_n^2 and Anderson-Darling, A_n^2 , in terms of Bahadur efficiency. We can conclude that, first, the Q_n statistic is more efficient (in the Bahadur sense) than D_n , W_n^2 and A_n^2 for A2, A3 and A5 families. Second, for the A4 family, it is always more efficient than D_n and W_n^2 , and it is also more efficient than A_n^2 for a wide range of values of parameter δ . Using the concepts of Bahadur asymptotic optimality, we have found the class of distribution functions for which the Q_n statistic is locally asymptotically optimal. Finally, as an application, we have seen that the test of uniformity based on Q_n can be used to detect the presence of hidden periodicities of unspecified frequency in a stationary time series.

References

- Anderson, T. W. and D. A. Darling (1952). Asymptotic theory of certain “Goodness of fit” criteria based on stochastic processes. *The Annals of Mathematical Statistics* 23, 193–212.
- Anderson, T. W. and D. A. Darling (1954). A test of goodness of fit. *Journal of the American Statistical Association* 49, 765–769.
- Azlarov, T. A. and N. A. Volodin (1986). *Characterization problems associated to the exponential distribution*. New York: Springer-Verlag.
- Bahadur, R. (1971). *Some limit theorems in statistics*. Philadelphia, PA: SIAM.
- Brockwell, P. and R. Davis (1990). *Time Series: Theory and Methods*. New York: Springer-Verlag.
- Dwass, M. (1961). The distribution of linear combinations of random divisions of an interval. *Trabajos de Estadística e Investigación Operativa* 12, 11–17.
- Fortiana, J. and A. Grané (2003). Goodness-of-fit tests based on maximum correlations and their orthogonal decompositions. *Journal of the Royal Statistical Society B* 65, 115–126.
- Grané, A. and J. Fortiana (2006). An adaptive goodness-of-fit test. *Communications in Statistics A. Theory and Methods* 35, 1141–1155.
- Groeneboom, P. and G. Shorack (1981). Large deviations of goodness-of-fit statistics and linear combinations of order statistics. *The Annals of Probability* 9, 971–987.
- Kolmogorov, A. (1933). On the empirical determination of a distribution. In S. Kotz and N. Johnson (Eds.), *Breakthroughs in Statistics. Volume II. Methodology and Distribution*, pp. 106–113. New York: Springer-Verlag.
- Mogul’skii, A. A. (1977). Remarks on large deviations for the ω^2 statistic. *Theory of Probability and its Applications* 12, 166–171.

- Nikitin, Y. (1995). *Asymptotic efficiency of nonparametric tests*. New York: Cambridge University Press.
- Nikitin, Y. Y. (1976). Bahadur relative asymptotic efficiency of statistics based on the empirical distribution function (in Russian). *Soviet Math. Dokl.* 17, 1645–1649. transl.: Soviet Math. Reports 17: 1645–1649.
- Nikitin, Y. Y. (1979). Large deviations and asymptotic efficiency of integral-type statistics I (in Russian). *Zap. Nauch. Sem. Leningrad Otdel. Math. Inst. Steklov LOMI* 85, 175–187. transl.: J. Soviet Math. 20 (1980): 2224–2231.
- Nikitin, Y. Y. (1980). Large deviations and asymptotic efficiency of integral-type statistics II (in Russian). *Zap. Nauch. Sem. Leningrad Otdel. Math. Inst. Steklov LOMI* 97, 151–175. transl.: J. Soviet Math. 24 (1984): 585–603.
- Planchky, D. and J. Steinebach (1975). A theorem about probabilities of large deviations with an application to queuing theory. *Periodica Mathematica Hungarica* 6, 343–345.
- Raghavachari, M. (1970). On a theorem of bahadur on the rate of convergence of test statistics. *The Annals of Mathematical Statistics* 41, 1695–1699.
- Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. New York: John Wiley & Sons.
- Shorack, G. R. and J. A. Wellner (1986). *Empirical processes with applications to statistics*. New York: John Wiley & Sons.
- Stephens, M. A. (1986). Tests based on EDF statistics. In R. B. D’Agostino and M. A. Stephens (Eds.), *Goodness-of-fit Techniques*, pp. 97–193. New York: Marcel Dekker, Inc.
- Tchirina, A. V. (2007). Asymptotic properties of exponentiality tests based on L-statistics. *Acta Applicandae Mathematicae* 97, 297–309.