ASYMPTOTIC EFFICIENCY AND LOCAL OPTIMALITY OF DISTRIBUTION – FREE TESTS BASED ON U- AND V-STATISTICS

by

Yakov Nikitin

St.Petersburg University, Russia and ICER

July 2001

Abstract. Many well-known goodness-of-fit and symmetry tests are based on U- et V-statistics. Recent progress in the analysis of their large deviations enables to find new expressions for their local Bahadur efficiency in case of bounded kernels and to formulate the conditions of local optimality. For some known statistics and such alternatives as location, scale and skew models this leads to original characterizations of distributions.

Key words : U-statistics, large deviations, Bahadur efficiency, local optimality.

Acknowledgement: Parts of this paper were written when the author was visiting ICER in Turin. ICER is gratefully acknowledged for providing financial support and excellent working conditions. The research was also supported in part by grants of RFBR No.99-01-0724 and 01-01-0245.

1. INTRODUCTION

Numerous distribution-free goodness-of-fit and symmetry tests are based on U- and V-statistics. The examples of them are the Wilcoxon one-sample test, the Maesono test [14], the Hollander-Proshan test of exponentiality [12], the Cramér – von Mises and Watson tests, and others.

To compare the statistical tests in large samples one commonly uses the concept of asymptotic efficiency. The efficiency of many concrete tests including those based on U- and V-statistics was intensively studied during last 40 years and most of these results are taken together in [16]. U- and V-statistics with non-degenerate kernels are asymptotically normal under null-hypothesis and alternative and their efficiency is calculated using the well-known Pitman approach. However many important statistics like the Cramér – von Mises or Watson statistics have degenerate kernels with non-normal limiting distributions and Pitman efficiency (PE) is not appropriate. Therefore one uses for such statistics the exact or the approximate Bahadur efficiency (BE).

The concepts of PE and BE are rather close. On one hand [4, App.2] for asymptotically normal statistics PE usually coincides with the local approximate BE as well as with local exact BE. On the other hand [25],[9] even for statistics with non-normal limiting distribution the local approximate BE is equal under mild conditions to the limiting PE. The development of large deviation theory which is the key tool for the calculation of exact BE ensured the use of this kind of efficiency for the comparison of tests and now PE and BE of most known statistics are known [16].

Next step in the investigation of BE and PE is the study of local asymptotic optimality (LAO) conditions . We are interested for which distributions of observations under the null-hypothesis and alternative the given sequence of statistics attains the maximum of efficiency. First remarks on the importance of this "inverse" problem are found in [5] and [23] but the detailed study was initiated in [15] and exposed later in [16]. The basic result consists in the description of the domain of LAO using the leading functions which are proper for every sequence of statistics.

All these results were obtained for many particular statistics but not for large classes of them like U- and V-statistics with more or less general kernels. The cause of that was the lack of information on Chernoff type large deviations for U- and V-statistics. General large deviation principle obtained in [8] is not sufficient for statistical applications. For some time it was believed the problem has been solved in [7] but later some mistakes were found in the proofs and even the result of [7] appeared to be false.

In recent papers [17], [18] the result of [7] was corrected for bounded non-degenerate and weakly degenerate (in other terms, of rank 1 and 2) kernels. Note that just bounded kernels are typical for distribution-free tests based on U- and V-statistics. For reader's convenience main results of [17] are quoted in Section 2. Using them we find new formulas for local exact BE (in Section 3) and describe the LAO domains (in Section 4) of tests based on U- and V-statistics with almost arbitrary non-degenerate or weakly degenerate m-variate bounded kernels. This generalizes and unifies sparse issues known so far only for particular kernels. We present also some examples of calculations for concrete statistics and alternatives.

For asymptotically normal statistics this program can be carried out also for PE using [21], the results are similar differing only in regularity conditions.

2. LARGE DEVIATIONS OF U- AND V-STATISTICS

Let $X_1, X_2, ...$ be a sequence of i.i.d. r.v.'s with known continuous distribution function (d.f.) F. Let $\Phi(x_1, ..., x_m)$ be a real symmetric kernel of degree m and denote by

$$U_n = \binom{n}{m}^{-1} \sum_{1 \le i_1 < ... < i_m \le n} \Phi(X_{i_1}, ..., X_{i_m})$$

the Hoeffding's U- statistic with the kernel Φ. The von Mises functional or V-statistic V_n with the kernel Φ is defined in a similar way, namely

$$V_n = n^{-m} \sum_{1 \le i_1, ..., i_m \le n} \Phi(X_{i_1}, ..., X_{i_m}).$$

See [13] for main definitions and classical results in the theory of U- and V-statistics. We are interested in the rough large deviation asymptotics of the statistics U_n and V_n.

Over the whole paper we suppose that the kernel Φ is bounded and that $\mathbf{E}\Phi = 0$. For most distribution-free statistics the kernel is a combination of indicators and hence bounded. We can always replace any r.v. of the form $\Phi(X_1, ..., X_m)$ by the r.v.

$$\tilde{\Phi}(U_1, ..., U_m) = \Phi(F^{-1}(U_1), ..., F^{-1}(U_m)), \tag{1}$$

where $U_1, U_2, ...$ are independent uniform r.v.'s on [0,1] and $\tilde{\Phi}$ is the "reduced" kernel. Hence if necessary we can assume the initial r.v.'s to be uniform on [0,1] owing to the complication of the kernel.

The transition from the kernel Φ to the "reduced" form $\tilde{\Phi}$ by (1) is quite natural for distribution-free statistics and is similar to the transition to a copula from a multidimensional d.f. In the sequel we will use both forms of the kernel depending on convenience.

Denote by I^m the m-dimensional unit cube and denote the projections of the kernels according to formulas

$$\tilde{\psi}(u_1) = \int_{I^{m-1}} \tilde{\Phi}(u_1, ..., u_m)\, du_2 ... du_m,$$
$$\psi(x_1) = \int_{R^{m-1}} \Phi(x_1, ..., x_m)\, dF(x_2) ... dF(x_m),$$

so that $\psi(x) = \tilde{\psi}(F(x))$. An important constant to be used later is

$$\sigma^2 = \int_0^1 \tilde{\psi}^2(x) dx = \int_{R^1} \psi^2(x) dF(x).$$

The first of main results of [17] is as follows.

Theorem 1. *Let the kernel $\tilde{\Phi}$ be bounded and has rank 1, that means that $\sigma^2 > 0$. Then for any real sequence $\{\gamma_n\}$, such that $\gamma_n \to 0$, and sufficiently small $a > 0$ it is true that*

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

where in the right-hand side there is a convergent series with numerical coefficients b_j, moreover $b_2 = -(2m^2\sigma^2)^{-1}$.The same formulation is valid for V-statistics with kernel $\tilde{\Phi}$.

For weakly degenerate statistics with the kernel of rank 2 we have a different asymptotics [17]. Weak degeneracy means [13] that a.e.

$$\int_{I^{m-1}} \tilde{\Phi}(s_1, s_2, ..., s_m) ds_2 ... ds_m = 0\,,$$

while (for $m > 2$)

$$\int_{I^{m-2}} \tilde{\Phi}(s_1, s_2, ..., s_m) ds_3 ... ds_m \neq 0.$$

For any kernel Φ of degree m put

$$\Phi^*(s_1, s_2) = \int_{R^{m-2}} \Phi(s_1, ..., s_m) dF(s_3) ... dF(s_m), \quad \text{if} \quad m > 2,$$

and $\Phi^*(s_1, s_2) = \Phi(s_1, s_2), \quad$ if $\quad m = 2$. The kernel $\tilde{\Phi}^*$ is defined likewise.

Theorem 2. *Let the kernel $\tilde{\Phi}$ be bounded and weakly degenerate. Let λ_0 be the smallest of numbers λ satisfying the integral equation*

$$x(s_1) = \lambda \int_0^1 \tilde{\Phi}^*(s_1, s_2)\, x(s_2)\, ds_2 \tag{2}$$

and suppose that λ_0 is a simple characteristic number of the linear integral operator with the kernel $\tilde{\Phi}^*$ acting from $L^2[0,1]$ into $L^2[0,1]$. Then for any real sequence $\{\gamma_n\}, \quad \gamma_n \to 0$, and sufficiently small $a > 0$ it is true that

$$\lim_{n\to\infty} n^{-1} \ln \Pr\{U_n \geq a + \gamma_n\} = \sum_{j=2}^{\infty} c_j a^{j/2},$$

where in the right-hand side there is a convergent series with numerical coefficients c_j , moreover $c_2 = -\lambda_0/m(m-1)$. The same relation holds for V-statistics with kernel $\tilde{\Phi}$.

These results enable us to calculate the local Bahadur efficiency of tests based on U- and V-statistics.

3. BAHADUR LOCAL EXACT SLOPES OF TESTS BASED ON U- AND V-STATISTICS

Let $X_1, ..., X_n$ be a sample of i.i.d. univariate observations with density $f(x, \theta)$ depending on real parameter $\theta \in [0, \theta^*]$, $\theta^* > 0$. Denote by $F(x, \theta)$ the d.f. corresponding to this density and put for simplicity $F(x, 0) = F(x)$, $F^{-1}(x, 0) = F^{-1}(x)$.

We are testing the goodness-of-fit hypothesis $H_0 : \theta = 0$ against the alternative $H_1 : \theta > 0$. Let the test be based on test statistic U_n and large values of it are significant.

Denote by $\mathbf{E}_\theta$ the expectation with respect to d.f. $F(x, \theta)$. We suppose that $b_\Phi(\theta) = \mathbf{E}_\theta \Phi(X_1, ..., X_m) > 0$ for all θ. This ensures the consistency of the test based on U_n. By the strong law of large numbers for U- and V-statistics [13, Ch.3] we have a.s. under H_1 that

$$U_n \to b_\Phi(\theta), \quad V_n \to b_\Phi(\theta).$$

In the same time we assume that $b_\Phi(\theta) \to 0$ as $\theta \to 0$.

The measure of BE of the sequence $\{U_n\}$ is the exact slope $c_U(\theta)$ describing the rate of exponential decrease of the attained level. According to the main theorem of Bahadur theory [6, Th. 7.2], [16, Th. 1.2.2] we can write out the main parts of exact slopes of test statistics $\{U_n\}$ with the kernel Φ in the non-degenerate and weakly degenerate case. In the first case we have by Theorem 1 as $\theta \to 0$

$$c_U(\theta) \sim \frac{b_\Phi^2(\theta)}{m^2\sigma^2}.$$

In the second case Theorem 2 yields as $\theta \to 0$

$$c_U(\theta) \sim \frac{2\lambda_0 b_\Phi(\theta)}{m(m-1)}.$$

Under some regularity conditions we will simplify the expression for $b_\Phi(\theta)$ when $\theta \to 0$. First we need the behavior of the likelihood function

$$L(\theta) = L(x_1, ..., x_m; \theta) = \Pi_{j=1}^m f(x_j, \theta)$$

as a function of θ. Denote for brevity h_θ the derivative h'_θ. Using the Taylor expansion in Lagrange form we get for some $y \in (0, \theta)$

$$\begin{aligned} L(\theta) &= \Pi_{j=1}^m f(x_j, 0) + \theta \textstyle\sum_{i=1}^m f_\theta(x_i, 0)\Pi_{j\neq i} f(x_j, 0) \\ &+\tfrac{1}{2}\theta^2 L_{\theta\theta}(0) + \tfrac{1}{6}\theta^3 L_{\theta\theta\theta}(y). \end{aligned}$$

The exact formula for $L_{\theta\theta}(0)$ is as follows

$$L_{\theta\theta}(0) = \sum_{i=1}^m f_{\theta\theta}(x_i, 0)\Pi_{j\neq i} f(x_j, 0) + \sum_{i\neq j}^m f_\theta(x_i, 0) f_\theta(x_j, 0)\Pi_{k\neq i,j} f(x_k, 0).$$

The term $L_{\theta\theta\theta}(y)$ has more complicated structure but contains only the terms like $f_{\theta\theta\theta}(x_1, y)\Pi_{j\neq 1} f(x_j, y)$, $f_{\theta\theta}(x_1, y) f_\theta(x_2, y)\Pi_{j\neq 1,2} f(x_j, y)$ and $\Pi_{i=1}^3 f_\theta(x_i, y)\Pi_{j\neq 1,2,3} f(x_j, y)$ if $m \geq 3$ with $y \in [0, \theta]$. The asymptotics of $b_\Phi(\theta)$

as $\theta \to 0$ is different in the non-degenerate and weakly degenerate case. In the first case let impose the following conditions.
Conditions ND: there exists such $\delta > 0$ that for any $0 \le \theta \le \delta$

$$\int_{R^m} \Phi(x_1, ..., x_m) f_{\theta\theta}(x_1, \theta) \Pi_{j \neq 1} f(x_j, \theta) dx_1 ... dx_m < \infty,$$

$$\int_{R^m} \Phi(x_1, ..., x_m) f_{\theta}(x_1, \theta) f_{\theta}(x_2, \theta) \Pi_{k \neq 1,2} f(x_k, \theta) dx_1 ... dx_m < \infty.$$

Using the Taylor expansion under this condition we get easily as $\theta \to 0$

$$b_\Phi(\theta) \sim m \int_{R^1} \psi(x) f_\theta(x, 0) dx \cdot \theta. \tag{3}$$

As Φ is bounded, the conditions ND are true if for $0 \le \theta \le \delta$

$$\int_{R^1} |f_\theta(x, \theta)| dx < \infty, \quad \int_{R^1} |f_{\theta\theta}(x, \theta)| dx < \infty. \tag{4}$$

Conditions (4) are valid for many families of densities.

In the weakly degenerate case the integral in (3) vanishes and the main term is of order $O(\theta^2)$. To get the required asymptotics we impose another set of conditions.
Conditions WD: for $0 \le \theta \le \delta$

$$\int_{R^m} \Phi(x_1, ... x_m) f_{\theta\theta\theta}(x_1, \theta) \Pi_{j \neq 1} f(x_j, \theta) dx_1 ... dx_m < \infty,$$

$$\int_{R^m} \Phi(x_1, ..., x_m) f_{\theta\theta}(x_1, \theta) f_\theta(x_2, \theta)\, \Pi_{j \neq 1,2} f(x_j, \theta) dx_1 ... dx_m < \infty,$$

$$\int_{R^m} \Phi(x_1, ... x_m) \Pi_{i=1}^3 f_\theta(x_i, \theta) \Pi_{j \neq 1,2,3} f(x_j, \theta) dx_1 ... dx_m < \infty,\ m \ge 3.$$

Under these conditions we get similarly that as $\theta \to 0$

$$b_\Phi(\theta) \sim \frac{m(m-1)}{2} \int_{R^2} \Phi^*(x_1, x_2) f_\theta(x_1, 0) f_\theta(x_2, 0) dx_1 dx_2 \cdot \theta^2. \tag{5}$$

Clearly for the validity of conditions WD it is sufficient to verify the assumptions (4) plus the condition: $\int_{R^1} |f_{\theta\theta\theta}(x, \theta)| dx < \infty$ for $0 \le \theta \le \delta$ which often takes place, too.

The asymptotics (3) and (5) imply the following formulas for the local exact slopes $c_U(\theta)$.

THEOREM 3. *Suppose that $\{U_n\}$ is a sequence of non-degenerate U- or V-statistics with bounded kernels and conditions ND are true. Then as $\theta \to 0$*

$$c_U(\theta) \sim \frac{1}{\sigma^2}\Big(\int_{R^1} \psi(x) f_\theta(x,0)dx\Big)^2 \cdot \theta^2. \tag{6}$$

THEOREM 4. *Suppose that $\{U_n\}$ is a sequence of weakly degenerate U- or V-statistics satisfying conditions of Theorem 2 and conditions WD. Then as $\theta \to 0$*

$$c_U(\theta) \sim \lambda_0 \int_{R^2} \Phi^*(x_1,x_2) f_\theta(x_1,0) f_\theta(x_2,0) dx_1 dx_2 \cdot \theta^2. \tag{7}$$

To compare these expressions with their potential upper bounds we recall the Bahadur-Raghavachari inequality [5], [6] according to which for all θ

$$c_U(\theta) \le 2K(\theta), \tag{8}$$

where $K(\theta)$ is the Kullback-Leibler information

$$K(\theta) = \int_{R^1} \ln \frac{f(x,\theta)}{f(x,0)} f(x,\theta)dx.$$

It is well-known that in typical cases one has the asymptotics [6], [16]:

$$2K(\theta) \sim I(f) \cdot \theta^2, \text{ as } \quad \theta \to 0, \tag{9}$$

where $I(f) = \int_{R^1}[f_\theta^2(x,0)/f(x,0)]dx$ is the Fisher information. Hence the local (absolute) BE $e_U^B(f)$ can be computed as follows: under conditions of Theorem 3 and (9)

$$e_U^B(f) = \Big(\int_{R^1} \psi(x) f_\theta(x,0)dx\Big)^2/\sigma^2 I(f) \le 1,$$

under conditions of Theorem 4 and (9)

$$e_U^B(f) = \lambda_0 \int_{R^2} \Phi^*(x_1,x_2) f_\theta(x_1,0) f_\theta(x_2,0) dx_1 dx_2 / I(f) \le 1.$$

These expressions generalize numerous formulas for efficiencies of tests based on U- and V-statistics known in particular cases [16].

For testing of symmetry we need some modifications. Let the null hypothesis $H_0^{'}$ be the hypothesis of symmetry with respect to 0 of d.f. F and the parametric alternative $H_1^{'}$ consists in that this d.f. is $F(x,\theta)$ with the density $f(x,\theta)$ which is symmetric with respect to zero only for $\theta = 0$. If a test of symmetry is based on U- or V-statistic with kernel Φ then the general formulas (6) and (7) are still valid. But as the null hypothesis is composite, the inequality (8) looks differently, see [11], [16], namely for all θ

$$c_U(\theta) \leq 2K_1(\theta) = 2\int_{R^1} f(x,\theta) \ln \frac{2f(x,\theta)}{f(x,\theta)+f(-x,\theta)} dx. \tag{10}$$

In typical cases the function $K_1(\theta)$ in (10) has as $\theta \to 0$ the asymptotics

$$K_1(\theta) \sim \frac{1}{8} I_1(f) \cdot \theta^2,$$

where

$$I_1(f) = \int_{R^1} \frac{(f_\theta(x,0) - f_\theta(-x,0))^2}{f(x,0)} dx.$$

Hence the efficiency of symmetry tests based on $\{U_n\}$ can be calculated according to formulas (6) and (7) if we replace $I(f)$ by $\frac{1}{4}I_1(f)$.

We give now two simple examples. Let the observations be normal with shift alternative so that $F(x,\theta) = N(x-\theta)$, where N is the standard normal d.f. with the density n. Consider the Wilcoxon one-sample statistic W_n for testing of symmetry which is a U- statistic with the kernel $\Phi_1(s,t) = \mathbf{1}\{s+t>0\} - \frac{1}{2}$. This kernel is non-degenerate with $\psi(s) = \frac{1}{2} - F(s)$ and $\sigma^2 = \frac{1}{12}$, and all imposed regularity conditions are true. By Theorem 3 we obtain integrating by parts as $\theta \to 0$

$$c_W(\theta) \sim 12(\int_{R^1} (N(x) - 1/2) n'(x) dx)^2 \cdot \theta^2 = \frac{3}{\pi} \cdot \theta^2.$$

As $\frac{1}{4}I_1(n) = 1$, we get $e_W^B(n) = \frac{3}{\pi} \approx 0.955$, a well-known classical result ascending to Pitman who considered of course the PE.

Another example is the ω^2 goodness-of-fit statistic which is a V-statistic of degree 2 with the weakly degenerate kernel [13, Sect.7.5]

$$\tilde{\Phi}_2(s,t) = \frac{1}{2}(s^2+t^2) - \max(s,t) + \frac{1}{3}. \tag{11}$$

The integral equation (2) with the kernel $\tilde{\Phi}_2$ can be reduced to the Sturm-Liouville boundary problem for $y(t) = \int_0^t x(s)ds$ of the form

$$y''(t) + \lambda y(t) = 0, \quad y(0) = y(1) = 0.$$

Characteristic numbers are $\lambda_k = k^2\pi^2, k = 1, 2, ...$ and the corresponding eigenfunctions are $C \sin k\pi t, k = 1, 2, ...$ The first characteristic value π^2 is simple and hence by (7) after some simplifications we get as $\theta \to 0$

$$c_{\omega^2}(\theta) \sim \pi^2 \int_{R^1} n^3(x)dx \cdot \theta^2 \sim \frac{\pi}{2\sqrt{3}} \cdot \theta^2.$$

It is known that $I(n) = 1$, hence $e^B_{\omega^2}(n) \approx 0.907$, which is also a well-known result [16, Sect. 2.6].

4. CONDITIONS OF LOCAL ASYMPTOTIC OPTIMALITY FOR TESTS BASED ON U - AND V - STATISTICS

As is known [5],[15] ,[16, Ch.6] the local asymptotic optimality (LAO) in Bahadur sense of a sequence of statistics means the asymptotic equivalence of local exact slope and $2K(\theta)$ as $\theta \to 0$. Denote for a given kernel Φ by $\mathcal{F}(\Phi)$ the set of densities $f(x, \theta)$ with d.f.'s $F(x, \theta)$ for which are satisfied the following regularity conditions. For non-degenerate kernels these are conditions (ND) and (9), for weakly degenerate case we assume conditions of Theorem 4 and (9). We suppose also that $\lim_{x\to-\infty} F_\theta(x, 0) = 0$ and that for almost all x one has

$$F_{x\theta}(x, 0) = F_{\theta x}(x, 0). \tag{12}$$

For such densities the LAO condition means that for any sequence U_n of U- or V-statistics with bounded kernel one has

$$(\int_{R^1} \psi(x) f_\theta(x, 0)dx)^2 = \sigma^2 I(f) \tag{13}$$

in the non-degenerate case, and

$$\lambda_0 \int_{R^2} \Phi^*(x_1, x_2) f_\theta(x_1, 0) f_\theta(x_2, 0)dx_1dx_2 = I(f) \tag{14}$$

in the weakly degenerate case.

We are interested for which densities $f(x,\theta)$ from $\mathcal{F}(\Phi)$ this is true; such densities form the domain of LAO in $\mathcal{F}(\Phi)$. A useful tool is the leading function of the sequence of statistics $\{T_n\}$ [16, Ch.6]. It is such real function $v_T(x)$ on [0,1], $v_T(0) = v_T(1) = 0$ that the condition $f(x,\theta) \in \mathcal{F}(\Phi)$ is equivalent to $F_\theta(F^{-1}(x),0) = Cv_T(x)$ for some constant C and almost all x.

The analysis of (13) is quite easy. By Cauchy-Schwarz inequality

$$(\int_{R^1} \psi(x) f_\theta(x,0)dx)^2 \le \int_{R^1} \psi^2(x) f(x,0)dx \cdot I(f) = \sigma^2 I(f)$$

with equality iff for some constant C_1 and almost all x one has

$$f_\theta(x,0) = C_1\psi(x)f(x,0), \tag{15}$$

or equivalently

$$\frac{d\ln f(x,\theta)}{d\theta}|_{\theta=0} = C_1\psi(x) = C_1\tilde{\psi}(F(x)). \tag{16}$$

The equation (15) may be also rewritten in the form

$$F_\theta(x,0) = C_1 \int_{-\infty}^{x} \psi(x)dF(x)$$

which is equivalent to

$$v_U(x) := F_\theta(F^{-1}(x),0) = C_1 \int_0^x \psi(F^{-1}(u))du = C_1 \int_0^x \tilde{\psi}(u)du =: C_1\tilde{\Psi}(x).$$

Hence the leading function of the sequence of statistics $\{U_n\}$ is equal to $\tilde{\Psi}(x) = \int_0^x \tilde{\psi}(u)du$. We can formulate the obtained result as follows.

THEOREM 5. *The domain of LAO in $\mathcal{F}(\Phi)$ for non-degenerate U- and V-statistics satisfying the conditions of Theorem 1 is defined by (15) or (16), and the leading function is $\tilde{\Psi}(x)$.*

In the weakly degenerate case put $g(x) = f_\theta(F^{-1}(x),0)/f(F^{-1}(x),0)$. Then we may rewrite (14) in the form

$$\lambda_0 \int_{I^2} \tilde{\Phi}^*(x_1,x_2)g(x_1)g(x_2)dx_1dx_2 = \int_0^1 g^2(x)dx. \tag{17}$$

The integral operator with the kernel $\tilde{\Phi}^*$ in (17) is the symmetric compact operator from $L^2[0,1])$ into $L^2[0,1]$. Denote it by $\mathcal{K}$ and by $(\cdot\,,\cdot)$ the scalar product in $L^2[0,1]$. The left-hand side of (17) is the quadratic form $\lambda_0(\mathcal{K}g, g)$. According to the Rayleigh principle its maximum under condition $(g,g) = 1$ is equal to 1 and is attained on the eigenfunction $\tilde{\varphi}_0$ corresponding to λ_0 [22, Sect. 93]. Hence the condition of LAO takes the form

$$f_\theta(x,0) = C_2\tilde{\varphi}_0(F(x))f(x,0), \tag{18}$$

or, equivalently, by (12)

$$F_\theta(x,0) = C_2 \int_0^{F(x)} \tilde{\varphi}_0(u)du =: C_2\tilde{\Phi}_0(F(x)).$$

Thus the leading function of the sequence of weakly degenerate statistics $\{U_n\}$ is $\tilde{\Phi}_0(x) = \int_0^{F(x)} \tilde{\varphi}_0(u)du$. This result is in conformity with [9] where approximate BE was used instead of exact BE.

THEOREM 6. *The domain of LAO in $\mathcal{F}(\Phi)$ for weakly degenerate U- and V-statistics satisfying the conditions of Theorem 2 is defined by (18), and the leading function is $\tilde{\Phi}_0(x)$.*

As an example consider again the classical ω^2 - test with the kernel (11). We found the eigenfunction $C\cos\pi t$ corresponding to the first characteristic value π^2. Hence the leading function is $C\sin\pi t$ that agrees with [15], [16].

If the initial family of densities is the location family so that $f(x,\theta) = f(x-\theta)$ then the condition of LAO (18) reads $F'(x) = C\sin\pi F(x)$. The solution of this equation is [16, Ch.6] the hyperbolic cosine distribution with the density $(\pi\cosh x)^{-1}, x \in R^1$ to the scale factor. Hence the domain of LAO consists of such distributions. If we replace the location family by scale family on R^+ we get [16] the right-sided Cauchy distribution.

Another interesting one-parameter family of alternatives is the skew family which was introduced and studied in [1], [2], [3], [10], and [20], among others. This alternative is asymmetric that can be more realistic in practical studies. The skew family of densities for any symmetric density h with d.f. H is given by the formula $f(x,\theta) = 2h(x)H(\theta x)$, $\theta \geq 0$. Suppose that the density h is such that all regularity conditions stated above are fulfilled. Careful analysis shows that it is sufficient to assume that h is positive within its support, has bounded second derivative and finite absolute moment of any order larger than 2. In that case the equation determining the LAO domain reads

$$\int_{-\infty}^{x} uf(u)du = -C_3\sin(\pi F(x)), C_3 > 0$$

which is equivalent to

$$xf(x) = -C_4 f(x) \cos(\pi F(x)), C_4 > 0. \qquad (19)$$

The solution of (19) on the set where $f \neq 0$ is

$$F(x) = \pi^{-1} \arcsin(x/C_4) + 1/2, \quad -C_4 \leq x \leq C_4,$$

with the density

$$f(x) = \left(\pi\sqrt{C_4^2 - x^2}\right)^{-1} \mathbf{1}\{-C_4 \leq x \leq C_4\}.$$

It can be shown that this arcsine density satisfies all required regularity conditions. Note that we got a new characterization in the class $\mathcal{F}(\tilde{\Phi}_2)$ of the symmetric arcsine density by the LAO property of ω^2 - statistic under skew alternative. The characterizations of this density exist [19], [24] but are very rare.

References

[1] Azzalini A. *A class of distributions which includes the normal ones*// Scand. J. Stat., 1985, v. 12, p. 171-178.

[2] Azzalini A. *Further results on a class of distributions which includes the normal ones*// Statistica, 1986, v. 46, p. 199-208.

[3] Azzalini A., Capitanio A. *Statistical applications of multivariate skew distribution*// J. Roy. Stat. Soc., 1999, v. B61, p. 579 - 602.

[4] Bahadur R. R *Stochastic comparison of tests*// Ann. Math. Stat., 1960, v. 31, p. 276–295.

[5] Bahadur R. R. *Rates of convergence of estimates and test statistics*//Ann.Math.Stat., 1967, v.38, p.303 - 324.

[6] Bahadur R. R. *Some limit theorems in statistics*// SIAM, Philadelphia, 1971.

[7] Dasgupta R. *On large deviation probabilities of U- statistics in non-i.i.d. case* // Sankhyā, 1984, v. A46, p.110–116.

[8] Eichelsbacher P., Löwe M. *A Large Deviation Principle for m-variate von Mises-statistics and U- Statistics* // J.Theoret. Prob., 1995. v.8, N 4, p.807–823.

[9] Gregory G. G. *On efficiency and optimality of quadratic tests.*// Ann.Statist., 1980, v.8, N 1, p.116 – 131.

[10] Henze N. *A probabilistic representation of the 'skew-normal' distribution*// Scand.J.Stat., 1986, v.13, p. 271 – 275.

[11] Ho Nguen Van. *Asymptotic efficiency in the Bahadur sense for the signed rank tests*// In: Proc. Prague Symp. on Asympt. Statistics, Charles Univer. Press, Prague, 1974, v. II, p. 127–156.

[12] Hollander M., Proschan F. *Testing whether new is better than used* // Ann. Math. Stat., 1972, v. 43, p. 1136–1146.

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

[14] Maesono Y. *Competitors of the Wilcoxon signed rank test*// Ann. Inst.Stat.Math., 1987, v.39, Pt.A, p.363 – 375.

[15] Nikitin Ya. Yu. *Local asymptotic Bahadur optimality and characterization problems*// Probab. Theory Appl., 1984, v.29, p. 79–92.

[16] Nikitin Ya. *Asymptotic efficiency of nonparametric tests*// Cambridge University Press, 1995.

[17] Nikitin Ya. Yu., Ponikarov E. V. *Large deviations of Chernoff type for U- statistics and von Mises functionals*// Proc. of St.Petersburg Math. Soc., 1999, v.7, p.124 – 167, in Russian. To be translated by Amer. Math. Soc.

[18] Nikitin Ya. Yu., Ponikarov E. V. *Large deviations of Chernoff type for U- and V-statistics* // Doklady RAN, 1999, v. 369, p. 13 – 16.

[19] Norton R. M. *Moment properties and the arc-sine law*// Sankhyā, 1978, v. A40, N 2, p.192 – 198.

[20] Pewsey A. *Problems of inference for Azzalini's skew-normal distribution*// J. of Appl. Stat., 2000, v. 27, p. 859 - 870.

[21] Rao C. R. *Criteria of estimation in large samples*// Sankhyā, 1963, v. A25, p.189–206.

[22] Riesz F., Sz.-Nagy B. *Leçons d'analyse fonctionnelle*// Académiai Kiado, Budapest, 6th edition, 1972.

[23] Savage I. R. *Nonparametric statistics: a personal review* // Sankhyā, 1969, v.A 31, p. 107–144.

[24] Shantaram R. *A characterization of the arcsine law*// Sankhyā, 1978, v. A40, N 2, p. 199 – 207.

[25] Wieand H. S. *A condition under which the Pitman and Bahadur approaches to efficiency coincide*// Ann.Statist., 1976, v. 4, p. 1003–1011.