JOURNAL OF MULTIVARIATE ANALYSIS 35, 168–185 (1990)

Limiting Values of Large Deviation Probabilities of Quadratic Statistics

G. A. M. JEURNINK AND W. C. M. KALLENBERG

Faculty of Applied Mathematics, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

Communicated by the Editors

Application of exact Bahadur efficiencies in testing theory or exact inaccuracy rates in estimation theory needs evaluation of large deviation probabilities. Because of the complexity of the expressions, frequently a local limit of the nonlocal measure is considered. Local limits of large deviation probabilities of general quadratic statistics are obtained by relating them to large deviation probabilities of sums of k-dimensional random vectors. The results are applied, e.g., to generalized Cramér–von Mises statistics, including the Anderson–Darling statistic, Neyman's smooth tests, and likelihood ratio tests. © 1990 Academic Press, Inc.

1. INTRODUCTION

Application of Bahadur efficiencies in testing theory or inaccuracy rates in estimation theory needs evaluation of large deviation probabilities. Often the available expressions for these probabilities are very complicated. Therefore many authors take a local limit of the nonlocal measure, which is important from a statistical point of view.

In a number of special cases the local limits are derived by relating the minimization of certain Kullback–Leibler "distances" to associated Euler–Lagrange differential equations. Typical examples of this elegant method are Nikitin [19] and Groeneboom and Shorack [5]. For certain general linear rank tests Kremer [13–15] gives an elegant theory of local comparison in the Bahadur sense. For more references see [8].

The first *general* result in this area is due to Wieand [23]. He has shown that the limiting *approximate* Bahadur efficiency equals the limiting Pitman efficiency. Approximate Bahadur efficiency, however, is in itself as a

Received September 19, 1989; revised April 9, 1990.

AMS classifications: primary 62 G 20; secondary 60 F 10.

Keywords and phrases: exact Bahadur efficiency, large deviations, generalized Cramér–von Mises statistics, quadratic statistics, Hilbert–Schmidt operator, eigenvalues, eigenfunctions, Neyman's smooth tests, likelihood ratio tests.

0047-259X/90 $3.00

measure of performance of tests of little value, since monotone transformations of a test statistic may lead to entirely different approximate Bahadur slopes, cf. [3]. Similar results hold for the approximate inaccuracy rate, cf. [9]. We therefore prefer to investigate exact Bahadur efficiency and exact inaccuracy rate.

The first *general* results on *exact* Bahadur efficiencies and *exact* inaccuracy rates are given in [11]. This approach is, however, restricted to statistics, which are, in a local sense, asymptotically normal, due to an approximation by a sum of i.i.d. random variables. Quadratic statistics are beyond the scope of this work. Recently Inglot and Ledwina [7] have developed a new approach using strong approximations. By this method several complicated statistics can be handled in an elegant way. A more detailed discussion is given in [8].

Here the more elementary method of [11] is extended to quadratic statistics, in that way unifying and clarifying isolated earlier results and presenting new results on this type of statistics by putting them in a general theory. This is done by relating the large deviation probabilities of the statistics to large deviation probabilities of sums of k-dimensional i.i.d. random vectors. The local limit of the latter large deviation probabilities can be easily obtained. Furthermore, the present method gives the asymptotically optimal directions for the growth of the power of the considered tests. Those directions are generated by eigenfunctions of an associated Hilbert–Schmidt operator. It is remarkable that this operator is *not* the same as the operator associated with the covariance function, which plays an important role in local theory of generalized Cramér–von Mises statistics, cf. Remark 4.4.

The paper is organized in the following way: in Section 2 finite and infinite weighted sums of squares of linear functionals are considered. General quadratic statistics are treated in Section 3 by relating them to statistics of the infinite dimensional type treated in Section 2. Functional analytic methods play an important role here. Generalized Cramér–von Mises statistics can be written in the form of quadratic statistics. They are treated in Section 4. A particular example in this class is the Anderson–Darling statistic. The general theory of this paper leads to a new proof and an explanation of results in [5]. A lot of other examples are presented: Neyman's smooth tests, Neuhaus' goodness of fit tests, Watson's statistic, and likelihood ratio statistics in a general framework.

2. Weighted sums of Squares

Let $\mathscr{X}$ be a set, $\mathscr{A}$ a σ-field of subsets of $\mathscr{X}$ and $\mathscr{P}$ the collection of all probability measures on $\mathscr{A}$. The estimators and test statistics studied here are of the form $T_n = T(\hat{P}_n)$, where T is a fixed functional and $\hat{P}_n$ is the empirical probability measure based on n observations. For $P, Q \in \mathscr{P}$ the Kullback–Leibler information number $K(Q, P)$ is defined by

$$K(Q, P) = \begin{cases} E_Q \log \dfrac{dQ}{dP} & \text{if } Q \ll P, \\ \infty & \text{otherwise.} \end{cases} \tag{2.1}$$

For $P \in \mathscr{P}$ and $\Omega \subset \mathscr{P}$ (Ω possibly depending on P) we define

$$K(\Omega, P) = \inf_{Q \in \Omega} K(Q, P). \tag{2.2}$$

Nonlocal measures like Bahadur efficiency or inaccuracy rate are often given in terms of Kullback–Leibler information numbers of the form $K(\Omega_\varepsilon, P)$ with

$$\Omega_\varepsilon = \{Q \in \mathscr{P}: T(Q) - T(P) > \varepsilon\} \tag{2.3}$$

or $\Omega_\varepsilon = \{Q \in \mathscr{P}: |T(Q) - T(P)| > \varepsilon\}$, where T is a functional with values in $\bar{\mathbb{R}}$, the extended real line, cf., e.g., [4]. The quantity $K(\Omega_\varepsilon, P)$ is often not very transparent and therefore local expansions are useful to get insight in the performance of the test or estimator based on the functional T. It is the purpose of this paper to expand $K(\Omega_\varepsilon, P)$ as $\varepsilon \to 0$, where Ω_ε is of the form (2.3). Extension of the results to sets Ω_ε, e.g., of the form $\{Q \in \mathscr{P}: |T(Q) - T(P)| > \varepsilon\}$ is straightforward.

In this section we consider the class of statistics $T_n = T(\hat{P}_n)$, where

$$T(Q) = \begin{cases} \sum_{i=1}^{\infty} \lambda_i \left\{ \int \phi_i(x)\, dQ(x) \right\}^2 & \text{if } \int |\phi_i(x)|\, dQ(x) < \infty, i = 1, 2, ..., \\ \infty & \text{otherwise,} \end{cases} \tag{2.4}$$

with

$\phi = (\phi_1, \phi_2, ...)$ a measurable function from $(\mathscr{X}, \mathscr{A})$ to $(\mathbb{R}^\infty, \mathscr{B}^\infty)$ with $\mathscr{B}^\infty$ the Borel σ-algebra on $\mathbb{R}^\infty$, $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant 0$, $\lambda_1 > 0$, $E_P \phi_i = 0$, $E_P \phi_i^2 = 1$, $E_P \phi_i \phi_j = 0$, $i \neq j$, $i, j = 1, 2, \ldots$. (2.5)

Condition (2.5) implies that $T(P) = 0$. Since under P the statistics ϕ_i are standardized and uncorrelated, the statistics $T_n = T(\hat{P}_n)$ with T given by (2.4) and (2.5) are under P typically asymptotically distributed as an infinite weighted sum of independent χ^2s. The inner product in $\mathbb{R}^k$ is denoted by $\langle a, b \rangle = \sum_{i=1}^k a_i b_i$, $a, b \in \mathbb{R}^k$, and the Euclidean norm by $\|a\|^2 = \langle a, a \rangle$, $a \in \mathbb{R}^k$.

THEOREM 2.1. *Let T be defined by* (2.4) *and* (2.5). *If*

$$\lim_{k\to\infty} \sup_x \sum_{i=k}^{\infty} \lambda_i \phi_i^2(x) = 0 \tag{2.6}$$

and for each $k>0$ there exists $\delta_k>0$ such that

$$\int \exp\left\{\sum_{i=1}^{k} t_i \lambda_i^{1/2} \phi_i(x)\right\} dP(x) < \infty \qquad \text{for all } (t_1, t_2, ...) \text{ with } \sum_{i=1}^{k} t_i^2 < \delta_k, \tag{2.7}$$

then

$$\lim_{\varepsilon\to 0} \varepsilon^{-1} K(\Omega_\varepsilon, P) = (2\lambda_1)^{-1}.$$

Proof. First consider the finite dimensional case, where $\lambda_{k+1}=0$ for some $k\geqslant 1$ (and hence $\lambda_i=0$ for all $i\geqslant k+1$). Define $\chi_k=(\phi_1, ..., \phi_k)$ and denote the statistical functional by T_k in this case. Let $y=(y_1, ..., y_k)$ and let Λ be the set of all probability measures on $(\mathbb{R}^k, \mathscr{B}^k)$ with $\mathscr{B}^k$ the Borel σ-algebra on $\mathbb{R}^k$. Define $\mu(B)=P(\chi_k^{-1}(B))$, $B\in\mathscr{B}^k$, and for $\nu\in\Lambda$,

$$\tilde{T}_k(\nu) = \begin{cases} \sum_{i=1}^{k} \lambda_i \left\{\int y_i \, d\nu(y)\right\}^2 & \text{if } \sum_{i=1}^{k} \int |y_i| \, d\nu(y) < \infty \\ \infty & \text{otherwise} \end{cases} \tag{2.8}$$

with

$$\lambda_1, ..., \lambda_k \text{ given by (2.5)}, \qquad \int y_i \, d\mu(y) = 0, \qquad \int y_i^2 \, d\mu(y) = 1,$$
$$\int y_i y_j \, d\mu(y) = 0, \qquad i\neq j,\ i, j = 1, ..., k. \tag{2.9}$$

Further, define

$$\tilde{K}(\lambda, \mu) = \begin{cases} E_\lambda \log \dfrac{d\lambda}{d\mu} & \text{if } \lambda \ll \mu, \\ \infty & \text{otherwise;} \end{cases}$$

$$\tilde{\Omega}_\varepsilon = \{\lambda\in\Lambda: \tilde{T}_k(\lambda) > \varepsilon\} \qquad \text{and} \qquad \tilde{K}(\tilde{\Omega}_\varepsilon, \mu) = \inf\{\tilde{K}(\lambda, \mu): \lambda\in\tilde{\Omega}_\varepsilon\}.$$

For each $Q\in\mathscr{P}$ define $\nu(B)=Q(\chi_k^{-1}(B))$, $B\in\mathscr{B}^k$. Then $T_k(Q)=\tilde{T}_k(\nu)$ and, in particular, $T_k(P)=\tilde{T}_k(\mu)=0$ (cf. also (2.5) and (2.9)). Moreover, by Theorem 4.1 in Chapter 2 of [16], $K(Q, P)\geqslant \tilde{K}(\nu, \mu)$ and, hence,

$$K(\Omega_\varepsilon, P) \geqslant \tilde{K}(\tilde{\Omega}_\varepsilon, \mu). \tag{2.10}$$

On the other hand, define for $\nu^* \in \Lambda$, $\nu^* \ll \mu$, the measure Q^* by

$$dQ^*(x) = \frac{d\nu^*}{d\mu}(\chi_k(x))\, dP(x).$$

It is easily seen that $\tilde{T}_k(\nu^*) = T_k(Q^*)$ and $\tilde{K}(\nu^*, \mu) = K(Q^*, P)$. If ν^* is not absolutely continuous w.r.t. μ, we have $\tilde{K}(\nu^*, \mu) = \infty$. Therefore,

$$\tilde{K}(\tilde{\Omega}_\varepsilon, \mu) \geqslant K(\Omega_\varepsilon, P). \tag{2.11}$$

Combination of (2.10) and (2.11) yields

$$K(\Omega_\varepsilon, P) = \tilde{K}(\tilde{\Omega}_\varepsilon, \mu). \tag{2.12}$$

By condition (2.7) and Lemma 4.1 in [11] it is seen that for $\nu \in \Lambda$, $\sum_{i=1}^k \int |y_i|\, d\nu(y) = \infty$ implies $\tilde{K}(\nu, \mu) = \infty$. Hence we may restrict attention to measures ν for which its expectation finitely exists. It now easily follows from Theorem 3 of Bartfai [2] (cf. also (6.4) in [1]) that

$$\tilde{K}(\tilde{\Omega}_\varepsilon, \mu) = -\lim_{n \to \infty} n^{-1} \log \Pr\left(\sum_{i=1}^k \lambda_i (\bar{Y}_i)^2 > \varepsilon \right),$$

where $(\bar{Y}_1, ..., \bar{Y}_k)$ is the mean of n i.i.d. random vectors each distributed according to μ. Note that (2.9) implies that under μ the distribution of these random vectors is not concentrated on a hyperplane in $\mathbb{R}^k$. Application of Theorem 3.2 and Remark 3.3 of Steinebach [22] yields

$$\begin{aligned} \tilde{K}(\tilde{\Omega}_\varepsilon, \mu) &= \inf\{-c(t) + \langle t, a \rangle : a \in \partial A,\, a = c'(t)\} \\ &= -c(t_\varepsilon) + \langle t_\varepsilon, a_\varepsilon \rangle \end{aligned} \tag{2.13}$$

with

$$c(t) = \log\left\{ \int e^{\langle t, y \rangle}\, d\mu(y) \right\} \qquad \text{and} \qquad c' \text{ its derivative}$$

$$A = \left\{ (a_1, ..., a_k) : \sum_{i=1}^k \lambda_i a_i^2 \leqslant \varepsilon \right\}, \qquad t_\varepsilon = (c')^{-1}(a_\varepsilon), \qquad a_\varepsilon \in \partial A.$$

By Taylor's expansion we obtain (cf. Lemma 3.1 in [10])

$$\{-c(t_\varepsilon) + \langle t_\varepsilon, a_\varepsilon \rangle\} \|a_\varepsilon\|^{-2} \to \tfrac{1}{2} \qquad \text{as} \quad \varepsilon \to 0. \tag{2.14}$$

Combination of (2.12), (2.13), and (2.14) yields

$$\liminf_{\varepsilon \to 0} \varepsilon^{-1} K(\Omega_\varepsilon, P) \geqslant \liminf_{\varepsilon \to 0} \tfrac{1}{2} \varepsilon^{-1} \|a_\varepsilon\|^2 \geqslant (2\lambda_1)^{-1}. \tag{2.15}$$

On the other hand, we have

$$\tilde{K}(\tilde{\Omega}_\varepsilon, \mu) \leqslant \inf\left\{\tilde{K}(v, \mu): \left|\int y_1\, dv(y)\right| > (\varepsilon\lambda_1^{-1})^{1/2}\right\}$$
$$= (2\lambda_1)^{-1}\varepsilon + o(\varepsilon) \qquad \text{as} \quad \varepsilon \to 0$$

by Corollary 2.3 in [11], and hence, cf. (2.12),

$$\limsup_{\varepsilon \to 0} \varepsilon^{-1} K(\Omega_\varepsilon, P) \leqslant (2\lambda_1)^{-1}. \tag{2.16}$$

Combination of (2.15) and (2.16) yields the result in the finite dimensional case.

Next consider the general case. Write T_k for T with λ_i replaced by 0, $i \geqslant k+1$, $k = 1, 2, \ldots$. Since $K(\Omega_\varepsilon, P) \leqslant \inf\{K(Q, P): Q \in \mathscr{P}, \ T_1(Q) > \varepsilon\}$, application of the finite dimensional result with $k = 1$ yields

$$\limsup_{\varepsilon \to 0} \varepsilon^{-1} K(\Omega_\varepsilon, P) \leqslant (2\lambda_1)^{-1} \tag{2.17}$$

and hence there exists $\varepsilon_0 > 0$ such that

$$K(\Omega_\varepsilon, P) = \inf\{K(Q, P): Q \in \mathscr{P}, T(Q) > \varepsilon, K(Q, P) \leqslant \varepsilon\lambda_1^{-1}\} \tag{2.18}$$

for all $0 < \varepsilon < \varepsilon_0$. Take $Q \in \mathscr{P}$ with $T(Q) > \varepsilon$, $K(Q, P) \leqslant \varepsilon\lambda_1^{-1}$, and take $\eta > 0$. In view of (2.6) there exists $k = k(\eta)$ such that $\lambda_k > 0$ and

$$\sup_x \sum_{i=k+1}^{\infty} \lambda_i \phi_i^2(x) < \tfrac{1}{2}\eta\lambda_1.$$

Hence we have, writing $|Q - P|$ for the total variation of $Q - P$,

$$\sum_{i=k+1}^{\infty} \lambda_i \left\{\int \phi_i(x)\, dQ(x)\right\}^2$$
$$= \sum_{i=k+1}^{\infty} \lambda_i \left\{\int \phi_i(x)\, d(Q-P)(x)\right\}^2$$
$$\leqslant \sup_{x, z} \sum_{i=k+1}^{\infty} \lambda_i |\phi_i(x)\, \phi_i(z)| \{|Q-P|(\mathscr{X})\}^2$$
$$\leqslant \sup_{x, z} \left\{\sum_{i=k+1}^{\infty} \lambda_i \phi_i^2(x) \sum_{i=k+1}^{\infty} \lambda_i \phi_i^2(z)\right\}^{1/2} 2K(Q, P)$$
$$= 2\varepsilon\lambda_1^{-1} \sup_x \sum_{i=k+1}^{\infty} \lambda_i\, \phi_i^2(x) < \varepsilon\eta,$$

where we have used the inequality $|Q-P|(\mathscr{X})=\{2K(Q,P)\}^{1/2}$, cf., e.g., [12]. Therefore, since $T(Q)>\varepsilon$, we obtain

$$T_k(Q)=T(Q)-\sum_{i=k+1}^{\infty}\lambda_i\left\{\int\phi_i(x)\,dQ(x)\right\}^2>\varepsilon-\varepsilon\eta=\varepsilon(1-\eta),$$

implying, cf. (2.18), for all $0<\varepsilon<\varepsilon_0$ and $\eta>0$,

$$K(\Omega_\varepsilon, P)\geqslant\inf\{K(Q,P)\colon Q\in\mathscr{P},\ T_k(Q)>\varepsilon(1-\eta)\}.$$

Application of the finite dimensional result yields for each $\eta>0$,

$$\liminf_{\varepsilon\to 0}\varepsilon^{-1}K(\Omega_\varepsilon, P)\geqslant(2\lambda_1)^{-1}(1-\eta). \tag{2.19}$$

Sending $\eta\to 0$ in (2.19) and combining the result with (2.17) the proof of the theorem is complete. ∎

Remark 2.1. If ϕ_i is bounded for each i, then (2.7) is satisfied by the Schwarz inequality, even if we replace δ_k by ∞.

As a first application we give an example of statistical functionals of the finite dimensional type.

EXAMPLE 2.1. Consider Neyman's smooth tests for uniformity. The proposed statistical functionals are given by

$$T_k(Q)=\sum_{i=1}^{k}\left\{\int_{(0,1)}\pi_i(x)\,dQ(x)\right\}^2,$$

where $\{\pi_i(x)\}$ are the normalized Legendre polynomials. The conditions of Theorem 2.1 are satisfied and hence $K(\Omega_\varepsilon, P)=\frac{1}{2}\varepsilon+o(\varepsilon)$ as $\varepsilon\to 0$. A simple generalization is obtained by replacing $\{\pi_i\}$ by another complete orthonormal system in $L^2(P)$, say $\{b_i\}$ with $b_0(x)=1$. The statistical functional is now given by

$$T_k(Q)=\sum_{i=1}^{k}\left\{\int_{(0,1)}b_i(x)\,dQ(x)\right\}^2.$$

Such statistics are investigated, e.g. , in [6]. If (2.7) holds, which is, in particular, the case if the functions b_i are bounded on $(0,1)$, then $K(\Omega_\varepsilon, P)=\frac{1}{2}\varepsilon+o(\varepsilon)$ as $\varepsilon\to 0$.

The second application concerns statistical functionals of the infinite dimensional type.

EXAMPLE 2.2. Consider the goodness of fit tests introduced by Neuhaus [17]. Theorem 2.1 can be directly applied, since the ϕ_i's are uniformly bounded and $\sum_{i=1}^{\infty} \lambda_i < \infty$ for those tests.

3. Quadratic Statistics

An important class of statistical functionals is of the form

$$T(Q) = \begin{cases} \iint \psi(s,t)\, dQ(s)\, dQ(t) & \text{if } \iint |\psi(s,t)|\, dQ(s)\, dQ(t) < \infty \\ \infty & \text{otherwise} \end{cases} \tag{3.1}$$

for a suitable chosen function ψ satisfying

$$\int \psi(s,t)\, dP(s) = 0, \qquad P\text{-a.e.} \tag{3.2}$$

Such quadratic statistical functionals appear as second-order differentials of statistical functionals, cf., e.g., Serfling [20, Chap. 6]. Generalized Cramér–von Mises statistics are of this form, as is seen in the following example.

EXAMPLE 3.1. Let $\mathscr{X} = (0, 1)$, $\mathscr{A}$ the Borel σ-algebra on $(0, 1)$ and P the Lebesgue measure on $(0, 1)$. Let w be a nonnegative and Lebesgue-measurable function satisfying

$$\int_{(0,1)} w(u)\, u(1-u)\, du < \infty. \tag{3.3}$$

Define

$$\psi(s,t) = \int_{(0,1)} (1_{[s,1)}(u) - u)(1_{[t,1)}(u) - u)\, w(u)\, du, \quad 0 < s, t < 1.$$

and denote the distribution function of Q by G. In view of (3.3) it can be shown (cf. [8]) that

$$\int_{(0,1)} \int_{(0,1)} \psi(s,t)\, dQ(s)\, dQ(t) = \int_{(0,1)} w(u)\{G(u) - u\}^2\, du,$$

which is the usual form of a generalized Cramér–von Mises statistic.

From now on we assume that

$$\iint e^{\delta\psi(s,t)} + e^{-\delta\psi(s,t)}\, dP(s)\, dP(t) < \infty \qquad \text{for some } \delta > 0, \tag{3.4}$$

It follows by Corollary 4.2 in [11] that

$$\iint |\psi(s, t)| \, dQ(s) \, dQ(t) = \infty \Rightarrow K(Q, P) = \infty. \tag{3.5}$$

Hence, writing $q(s) = dQ/dP(s)$ if $Q \ll P$,

$$K(\Omega_\varepsilon, P) = \inf \left\{ K(Q, P) \colon Q \ll P, \iint |\psi(s, t)| \, dQ(s) \, dQ(t) < \infty, \right.$$
$$\left. \iint \psi(s, t) \, q(s) \, q(t) \, dP(s) \, dP(t) > \varepsilon \right\}.$$

We even can restrict attention to $q \in L^2(P)$.

PROPOSITION 3.1. *Assume that* (3.2) *and* (3.4) *hold. Then*

$$K(\Omega_\varepsilon, P) = \inf \left\{ K(Q, P) \colon Q \ll P, q = dQ/dP \in L^2(P), \right.$$
$$\iint |\psi(s, t)| \, dQ(s) \, dQ(t) < \infty,$$
$$\left. \iint \psi(s, t) \, q(s) \, q(t) \, dP(s) \, dP(t) > \varepsilon \right\}. \tag{3.6}$$

Proof. Let $Q \ll P$ satisfy $K(Q, P) < \infty$, $\iint |\psi(s, t)| \, dQ(s) \, dQ(t) < \infty$, and $\iint \psi(s, t) \, dQ(s) \, dQ(t) > \varepsilon$. Writing $q = dQ/dP$ and

$$q_n = q 1_{q \leqslant n} \left(\int_{q \leqslant n} dQ \right)^{-1},$$

it follows by the dominated convergence theorem that $\int_{q \leqslant n} dQ \to 1$ and

$$\int q_n \log q_n \, dP \to \int q \log q \, dP$$

$$\iint \psi(s, t) \, q_n(s) \, q_n(t) \, dP(s) \, dP(t) \to \iint \psi(s, t) \, dQ(s) \, dQ(t),$$

and hence the desired result is obtained. ∎

Remark 3.1. Note that, in fact, we have proved that we can restrict attention to $q \in L^\infty(P)$, and hence $q \in L^p(P)$ for each $1 \leqslant p \leqslant \infty$.

Under condition (3.4) we have $\psi \in L^2(P \times P)$ and hence ψ induces a Hilbert–Schmidt operator A from $L^2(P)$ into $L^2(P)$, defined by

$$Aq(t) = \int q(s)\, \psi(s, t)\, dP(s). \tag{3.7}$$

We further assume that the kernel

$$\psi \text{ is symmetric and } \iint \psi^2(s, t)\, dP(s)\, dP(t) > 0. \tag{3.8}$$

Let $\lambda_0, \lambda_1, \lambda_2, \ldots$ denote the sequence of eigenvalues of A and $\phi_0, \phi_1, \phi_2, \ldots$ be a corresponding sequence of orthonormal eigenfunctions, each eigenvalue $\neq 0$ repeated according to its multiplicity. So for all i and j

$$A\phi_i = \lambda_i \phi_i, \qquad P\text{-a.e.}, \qquad \int \phi_i \phi_j\, dP = 0 \quad \text{if} \quad i \neq j, \qquad \int \phi_i^2\, dP = 1. \tag{3.9}$$

In view of (3.2) we may let $\phi_0 \equiv 1$ correspond to the eigenvalue $\lambda_0 = 0$. Without loss of generality , let

$$A\phi_i \equiv \lambda_i \phi_i \qquad \text{if} \quad \lambda_i \neq 0. \tag{3.10}$$

Further, we have

$$\psi(s, t) = \sum_{i=1}^{\infty} \lambda_i \phi_i(s)\, \phi_i(t) \qquad \text{in the } L^2\text{-sense;} \tag{3.11a}$$

i.e.,

$$\lim_{n \to \infty} \iint \left\{ \psi(s, t) - \sum_{i=1}^{n} \lambda_i \phi_i(s)\, \phi_i(t) \right\}^2 dP(s)\, dP(t) = 0 \tag{3.11b}$$

and, hence,

$$0 < \sum_{i=1}^{\infty} \lambda_i^2 = \iint \psi^2(s, t)\, dP(s)\, dP(t) < \infty. \tag{3.12}$$

We therefore may write, under the conditions of Proposition 3.1 and condition (3.8),

$$K(\Omega_\varepsilon, P) = \inf \left\{ K(Q, P) \colon Q \ll P,\, q = dQ/dP \in L^2(P), \right.$$
$$\left. \iint |\psi(s, t)|\, dQ(s)\, dQ(t) < \infty,\, \sum_{i=1}^{\infty} \lambda_i \left(\int q\phi_i\, dP \right)^2 > \varepsilon \right\}. \tag{3.13}$$

Our next aim is to approximate the infinite sum in (3.13) by a finite sum in the sense of condition (2.6) in order to apply Theorem 2.1. We assume in the sequel that

$\mathscr{X}$ is a metric space with metric d, $\mathscr{A}$ is the Borel σ-algebra on $\mathscr{X}$ and the support S of P is compact. (3.14)

We put the following conditions on ψ:

$$\lim_{\substack{d(s,s_0)\to 0\\ s\in S}} \int \left\{\psi(s,t)-\psi(s_0,t)\right\}^2 dP(t)=0 \qquad \text{for each } s_0\in S \tag{3.15}$$

ψ is continuous at every point $(s,s)\in S\times S$ (w.r.t. the relative topology) (3.16)

ψ is a positive semidefinite kernel, which means $\iint \psi(s,t)\,q(s)\,q(t)\,dP(s)\,dP(t)\geq 0$ for all $q\in L^2(P)$. (3.17)

Remark 3.2. If ψ is continuous on $S\times S$ (w.r.t. the relative topology), then ψ is uniformly continuous on $S\times S$ in view of (3.14) and hence (3.4), (3.15), and (3.16) are fulfilled.

The conditions (3.2), (3.4), (3.8), (3.14), (3.15), (3.16), and (3.17) together are called *condition* $\mathscr{C}$.

PROPOSITION 3.2. *Assume that condition $\mathscr{C}$ holds, then*

(i) $\lambda_i\geq 0$ *for all i*, $\sup_i\lambda_i>0$

(ii) ϕ_i *is continuous on S if* $\lambda_i\neq 0$,

(iii) $\int\psi^2(s,t)\,dP(s)=\sum_{i=1}^{\infty}\lambda_i^2\phi_i^2(t)$ *for all* $t\in S$,

(iv) $\lim_{k\to\infty}\sup_{s\in S}\sum_{i=k}^{\infty}\lambda_i\phi_i^2(s)=0$.

Proof. (i) Take $q=\phi_i$ in (3.17) and apply (3.9), yielding $\lambda_i\geq 0$ for all i. By (3.12) we have $0<\sum_{i=1}^{\infty}\lambda_i^2<\infty$ and hence $\sup_i\lambda_i=\max_i\lambda_i>0$.

(ii) Using (3.8) and (3.15) it follows that Aq is continuous on S for each $q\in L^2(P)$. If $\lambda_i\neq 0$ we have $A\phi_i=\lambda_i\phi_i$ (cf. (3.10)) and hence ϕ_i is continuous on S if $\lambda_i\neq 0$.

For a proof of (iii) and (iv) see [8]. ∎

Combination of Proposition 3.1, 3.2, and Theorem 2.1 yields the following result for quadratic statistics:

THEOREM 3.3. *Let T be defined by* (3.1). *Assume that condition $\mathscr{C}$ holds. Then*

$$\lim_{\varepsilon \to 0} \varepsilon^{-1} K(\Omega_\varepsilon, P) = (2\lambda_1)^{-1}, \tag{3.18}$$

where λ_1 is the largest eigenvalue of the operator A, cf. (3.7).

Proof. For $Q \ll P$, $K(Q, P)$ and $T(Q)$ remain unchanged if we replace $\mathscr{X}$ by S. Therefore without loss of generality we assume $\mathscr{X} = S$. In view of Proposition 3.2, ϕ_i is continuous on S. By (3.14) S is compact and hence ϕ_i is bounded for each i. Moreover, by Proposition 3.2(iv),

$$\sum_{i=n}^{\infty} |\lambda_i \phi_i(s)\, \phi_i(t)| \leqslant \left\{\sum_{i=n}^{\infty} \lambda_i \phi_i^2(s)\right\}^{1/2} \left\{\sum_{i=n}^{\infty} \lambda_i \phi_i^2(t)\right\}^{1/2} \to 0$$

uniformly on $S \times S$ as $n \to \infty$, implying that $\tilde{\psi}(s, t) = \sum_{i=1}^{\infty} \lambda_i \phi_i(s)\, \phi_i(t)$ is well defined on $S \times S$. Let

$$\tilde{T}(Q) = \iint \tilde{\psi}(s, t)\, dQ(s)\, dQ(t) = \sum_{i=1}^{\infty} \lambda_i \left\{\int \phi_i(x)\, dQ(x)\right\}^2,$$

then $\tilde{T}$ satisfies (2.4) and (2.5). (Note that $\phi_0 \equiv 1$ and hence $E_P \phi_i = 0$, $i = 1, 2, ...,$ by (3.9).) It is ensured by Proposition 3.2(iv) and Remark 2.1 that conditions (2.6) and (2.7) are satisfied and therefore by Theorem 2.1,

$$\lim_{\varepsilon \to 0} \varepsilon^{-1} \inf\{K(Q, P): Q \in \mathscr{P}, \tilde{T}(Q) > \varepsilon\} = (2\lambda_1)^{-1}. \tag{3.19}$$

By (3.11), $T(Q) = \tilde{T}(Q)$ on the set

$$\left\{Q \in \mathscr{P}: Q \ll P, q = dQ/dP \in L^2(P), \iint |\psi(s, t)|\, dQ(s)\, dQ(t) < \infty\right\}.$$

Note that $\iint |\tilde{\psi}(s, t)|\, dQ(s)\, dQ(t) < \infty$ for all $Q \in \mathscr{P}$, since $\tilde{\psi}$ is bounded on $S \times S$. In view of Proposition 3.1, (3.5), and (3.13) we therefore have

$$\begin{aligned}
&\inf\{K(Q, P): Q \in \mathscr{P}, \tilde{T}(Q) > \varepsilon\} \\
&\quad = \inf\left\{K(Q, P): Q \ll P, q = dQ/dP \in L^2(P), \tilde{T}(Q) > \varepsilon\right\} \\
&\quad = \inf\Big\{K(Q, P): Q \ll P, q = dQ/dP \in L^2(P), \\
&\qquad\quad \iint |\psi(s, t)|\, dQ(s)\, dQ(t) < \infty, \tilde{T}(Q) > \varepsilon\Big\} = K(\Omega_\varepsilon, P).
\end{aligned}$$

In combination with (3.19) this establishes (3.18). ∎

EXAMPLE 3.2. Consider the Watson statistic, given by

$$T(Q)=\int_{[0,1]}\left[G(u)-u-\int_{[0,1]}\{G(v)-v\}\,dv\right]^2 du, \tag{3.20}$$

where G is the distribution function corresponding to Q. If $Q \ll P$ with P the Lebesgue measure on $[0, 1]$, then

$$T(Q)=\int_{[0,1]}\int_{[0,1]}\psi(s,t)\,dQ(s)\,dQ(t)$$

with continuous

$$\psi(s,t)=\min(s,t)-\tfrac{1}{2}(s+t)+\tfrac{1}{2}(s-t)^2+\tfrac{1}{12}.$$

It is easily seen, cf. Remark 3.2, that condition C of Section 3 is satisfied and that for $j=1, 2, ...,$

$$\lambda_{2j-1}=\lambda_{2j}=1/(4\pi^2 j^2)$$

(and $\phi_{2j-1}(t)=\sqrt{2}\sin(2j\pi t)$. $\phi_{2j}(t)=\sqrt{2}\cos(2j\pi t)$). Hence,

$$K(\Omega_\varepsilon, P)=2\pi^2\varepsilon+o(\varepsilon) \qquad \text{as} \quad \varepsilon\to 0.$$

4. GENERALIZED CRAMÉR–VON MISES STATISTICS

In this section local limits of large deviation probabilities of generalized Cramér–von Mises statistics are obtained. As is seen in Example 3.1 generalized Cramér–von Mises statistics can be written in the form of quadratic statistics. First, we consider statistical functionals

$$T(Q)=\int_{(0,1)} w(u)\{G(u)-u\}^2\,du, \tag{4.1}$$

where G is the distribution function of Q and w is a non-negative, Lebesgue-measurable function, which is *integrable* on $(0, 1)$.

To apply the theory of Section 3 we slightly change the framework of Example 3.1. Let $\mathscr{X}=[0,1]$, $\mathscr{A}$ the Borel σ-algebra on $[0, 1]$ and P the Lebesgue measure on $[0, 1]$. Since $K(Q, P)=\infty$ if Q is not absolutely continuous w.r.t. P, we restrict attention to $Q\in\mathscr{P}$ with $Q\ll P$ and hence $Q(\{0\})=Q(\{1\})=0$. For those Q's, $T(Q)$ is of the form (3.1) with

$$\psi(s,t)=\int_{[0,1]}(1_{[s,1]}(u)-u)(1_{[t,1]}(u)-u)\,w(u)\,du, \quad 0\leqslant s, t\leqslant 1, \tag{4.2}$$

by Example 3.1. Note that since w is integrable, ψ is well defined for all $s, t \in [0, 1]$. Moreover ,it follows by dominated convergence that ψ is continuous on $\mathscr{X} \times \mathscr{X}$. It is easily seen that condition C holds with $S = \mathscr{X}$ in (3.14). Hence application of Theorem 3.3 yields the following result.

THEOREM 4.1. *Let T be defined by* (4.1) *with w a non-negative, Lebesgue measurable function, which is integrable on* (0, 1). *Then*

$$\lim_{\varepsilon \to 0} \varepsilon^{-1} K(\Omega_\varepsilon, P) = (2\lambda_1)^{-1}, \tag{4.3}$$

where λ_1 is the largest eigenvalue of the operator defined in (3.7) *with ψ given by* (4.2).

Remark 4.1. For other proofs of this result see [7], especially Remark 4.

Next consider generalized Cramér–von Mises statistics with w *not necessarily integrable* on (0, 1). An important example is $w(u) = \{u(1-u)\}^{-1}$ leading to the Anderson–Darling statistic. We return to the framework of Example 3.1. So $\mathscr{X} = (0, 1)$, A the Borel σ-algebra on (0, 1) and P the Lebesgue measure on (0, 1). We assume that w satisfies

Condition D. w is a nonnegative, Lebesgue-measurable function on (0, 1)

w is integrable on $[\eta, 1-\eta]$ for each $\eta > 0$

$$\int_{(0,s)} u^2 w(u)\, du \leqslant -c_1 \log(1-s) \qquad 0 < s < 1 \text{ for some } c_1 > 0$$

$$\int_{(t,1)} (1-u)^2 w(u)\, du \leqslant -c_2 \log t \qquad 0 < t < 1 \text{ for some } c_2 > 0.$$

We have the following generalization of Theorem 4.1.

THEOREM 4.2. *Let T be defined by* (4.1) *with w satisfying Condition* D. *Then*

$$\lim_{\varepsilon \to 0} \varepsilon^{-1} K(\Omega_\varepsilon, P) = (2\lambda_1)^{-1}, \tag{4.4}$$

where λ_1 is the largest eigenvalue of the operator defined in (3.7) *with ψ given by* (4.5).

We give only a sketch of the proof. Technical details can be found in [8].

Sketch of Proof. If w satisfies Condition D, then $\int_0^1 w(u)\, u(1-u)\, du < \infty$. Define the function ψ as in Example 3.1 by

$$\psi(s, t) = \int_{(0,1)} (1_{[s,1)}(u) - u)(1_{[t,1)}(u) - u)\, w(u)\, du, \qquad 0 < s, t < 1. \tag{4.5}$$

In view of (3.13),

$$K(\Omega_\varepsilon, P) = \inf\left\{K(Q, P)\colon Q \ll P, q = dQ/dP \in L^2(P),\right.$$
$$\left.\iint |\psi(s, t)|\, dQ(s)\, dQ(t) < \infty, \sum_{i=1}^{\infty} \lambda_i \left(\int q\phi_i\, dP\right)^2 > \varepsilon\right\}.$$

By Condition D there exists $\delta > 0$ such that

$$\int \exp\{t\lambda_1^{1/2}\phi_1(x)\}\, dP(x) < \infty \qquad \text{for all} \quad |t| < \delta.$$

Moreover, $K(\Omega_\varepsilon, P) \leqslant \inf\{K(Q, P)\colon Q \in \mathscr{P}, \lambda_1(\int \phi_1\, dQ)^2 > \varepsilon\}$ and hence, by Theorem 2.1,

$$\limsup_{\varepsilon \to 0} \varepsilon^{-1} K(\Omega_\varepsilon, P) \leqslant (2\lambda_1)^{-1}. \tag{4.6}$$

It remains to show that

$$\liminf_{\varepsilon \to 0} \varepsilon^{-1} K(\Omega_\varepsilon, P) \geqslant (2\lambda_1)^{-1}. \tag{4.7}$$

The idea is to approximate T by a truncated version T_η, where the integration is over $[\eta, 1-\eta]$. Theorem 4.1 is applied on T_η. Finally it is shown, using Proposition 1(iii) in [5], that for Q with $K(Q, P) \leqslant (2\lambda_1)^{-1}\varepsilon$ and $T(Q) > \varepsilon$, $\varepsilon^{-1}T_\eta(Q)$ is close to $\varepsilon^{-1}T(Q)$ (uniformly in ε) and the largest eigenvalue of the operator associated with T_η is close to λ_1 if η is small. ∎

Remark 4.2. The class of functions w satisfying Condition D includes $w(u) = \{u(1-u)\}^{-1}$, leading to the Anderson–Darling statistic. In that case $\lambda_1 = \frac{1}{2}$ and hence,

$$K(\Omega_\varepsilon, P) = \varepsilon + o(\varepsilon) \qquad \text{as} \quad \varepsilon \to 0,$$

which is (2.5) in Groeneboom and Shorack [5]. The proof of Theorem 4.2 is a new proof of this result.

Remark 4.3. (Likelihood ratio test in a rather general situation). Let $X_1, X_2, \ldots$ be i.i.d. random variables on $(\mathscr{X}, \mathscr{A})$ with density $p_\theta(x)$, $\theta \in \Theta \subset \mathbb{R}^k$, with respect to some measure μ. Consider the testing problem

$H_0: \theta = \theta_0$ against $H_1: \theta \neq \theta_0$ for some $\theta_0 \in \operatorname{int} \Theta$. We write p for p_{θ_0} and P for the corresponding probability measure. Let M denote the functional generating the maximum likelihood estimator, i.e., $M(\hat{P}_n)$ is the maximum likelihood estimator of θ based on $X_1, ..., X_n$. Define

$$T(Q) = \int \log p_{M(Q)}(x)\, dQ(x) - \int \log p(x)\, dQ(x),$$

yielding $T(\hat{P}_n)$ as minus the log likelihood ratio statistic for testing H_0 against H_1, thus rejecting for large values of $T(\hat{P}_n)$. Under suitable regularity conditions (which are satisfied, e.g., for exponential families) it can be shown that

$$K(\Omega_\varepsilon, P) = \varepsilon + o(\varepsilon) \qquad \text{as} \quad \varepsilon \to 0.$$

The main idea is that T is "close" to a quadratic statistic. We do not present details here.

Remark 4.4. It follows from Remark 4.1 in [11] that the infimum $K(\Omega_\varepsilon, P)$ in Theorem 2.1 is "approximately" attained by taking

$$dQ_\varepsilon(x) = \{1 + (\varepsilon \lambda_1^{-1})^{1/2} \phi_1(x)\}\, dP(x)$$

with ϕ_1 and λ_1 given in (2.5). A similar statement holds for $K(\Omega_\varepsilon, P)$ in Theorems 3.3, 4.1, and 4.2. This implies that the direction given by Q_ε is approximately the optimal direction for the growth of the power of the test based on T, cf. also Neuhaus [18, Sect. 4]. In particular, for the Anderson–Darling statistic the largest eigenvalue equals $\lambda_1 = \frac{1}{2}$ and the corresponding eigenfunction is given by $\phi_1(s) = \sqrt{3}(1 - 2s)$. Therefore, the infimum $K(\Omega_\varepsilon, P)$ is "approximately" attained by taking $dQ_\varepsilon(s) = \{1 + (2\varepsilon)^{1/2} \sqrt{3}(1 - 2s)\}\, ds$ with distribution function $G_\varepsilon(t) = t + t(1-t)\sqrt{6\varepsilon}$. This explains Remark 10 in [5]. In local theory of generalized Cramér–von Mises statistics the covariance function

$$H(s, t) = \{\min(s, t) - st\}\{w(s)\, w(t)\}^{1/2}$$

plays an important role, cf., e.g., [21, p. 223]. The function ψ defined in (4.5) differs from H. For the operators A_ψ and A_H associated with ψ and H, respectively, the eigenvalues unequal to zero are the same. However, the corresponding eigenfunctions are not the same. They are related by $\phi_j = \sqrt{\lambda_j}(f_j w^{-1/2})'$, with ϕ_j and f_j eigenfunctions of A_ψ and A_H, respectively. (Note that, in their notation, $f_j w^{-1/2}$ satisfies the differential equation on page 223 of [21] and not f_j itself.) Because of the statistical interpretation of the eigenfunctions, it seems to be more appropriate to consider

ψ than H in this context. For the likelihood ratio statistic discussed in Remark 4.3 the involved eigenvalues unequal zero are all the same, which corresponds with its omnibus character of no preference in a special direction.

Acknowledgment

We thank the referees for their careful reading and their helpful comments which have led to a much more concise presentation of the results.

References

[1] Bahadur, R. R., and Zabell, S. L. (1979). Large deviations of the sample mean in general vector spaces. *Ann. Probab.* **7** 587–621.

[2] Bartfai, P. (1978). *Large Deviations of the Sample Mean in Euclidean Spaces.* Mimeograph Series No. 78-13, Statistics Department, Purdue University.

[3] Gleser, L. J. (1964). On a measure of test efficiency proposed by R. R. Bahadur. *Ann. Math. Statist.* **35** 1537–1544.

[4] Groeneboom, P., Oosterhoff, J., and Ruymgaart, F. H. (1979). Large deviation theorems for empirical probability measures. *Ann. Probab.* **7** 553–586.

[5] Groeneboom, P., and Shorack, G. R. (1981). Large deviations of goodness of fit statistics and linear combinations of order statistics. *Ann. Probab.* **9** 971–987.

[6] Inglot, T., Jurlewicz, T., and Ledwina, T. (1987). On some smooth tests of fit. Preprint.

[7] Inglot, T., and Ledwina, T. (1990). On probabilities of excessive deviations for Kolmogorov–Smirnov, Cramér–von Mises, and chi-square statistics. *Ann Statist.* **18**. To appear.

[8] Jeurnink, G. A. M., and Kallenberg, W. C. M. (1988). *Limiting Exact Efficiency of Quadratic Statistics. TW*-memorandum No. 737. University of Twente.

[9] Jurečková, J., and Kallenberg, W. C. M. (1987). On local inaccuracy rates and asymptotic variances. *Statist. Decisions* **5** 139–158.

[10] Kallenberg, W. C. M. (1981). Bahadur deficiency of likelihood ratio tests in exponential families, *J. Multivariate Anal.* **11** 506–531.

[11] Kallenberg, W. C. M., and Ledwina, T. (1987). On local and nonlocal measures of efficiency. *Ann. Statist.* **15** 1401–1420.

[12] Kemperman, J. H. B. (1969). On the optimum rate of transmitting information. In *Probability and Information Theory.* Lecture Notes in Math., Vol. 89, pp. 126–169. Springer-Verlag, Berlin.

[13] Kremer, E. (1979). Approximate and local Bahadur efficiency of linear rank tests in the two-sample problem. *Ann. Statist.* **7** 1246–1255.

[14] Kremer, E. (1981). Local Bahadur efficiency of rank tests for the independence problem, *J. Multivariate Anal.* **11** 532–543.

[15] Kremer, E. (1982). Local comparison of linear rank tests in the Bahadur sense. *Metrika* **29** 159–173.

[16] Kullback, S. (1959). *Information Theory and Statistics.* Wiley, New York.

[17] Neuhaus, G. (1986). *A Class of Quadratic Goodness of Fit Tests.* Preprint No. 11, Institut für Mathematische Stochastik, Universität Hamburg.

[18] NEUHAUS, G. (1987). Local asymptotics for linear rank statistics with estimated score functions. *Ann. Statist.* **15** 491–512.

[19] NIKITIN, YA. YU. (1980). Bahadur efficiency of ω^2-type criteria in the several sample case. *Zap. Nauchn. Sem. Leningrad Otdel Mat. Inst. Steklov (LOMI)* **98** 140–148.

[20] SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

[21] SHORACK, G. R., AND WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

[22] STEINEBACH, J. (1978). Convergence rates of large deviation probabilities in the multi-dimensional case. *Ann. Probab.* **6** 751–759.

[23] WIEAND, H. S. (1976). A condition under which the Pitman and Bahadur approaches to efficiency coincide. *Ann. Statist.* **4** 1003–1011.