Weierstraß-Institut
für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Preprint ISSN 0946 – 8633

On moderate deviation probabilities of empirical bootstrap measure

Mikhail Ermakov

submitted: 22nd November 2004

Mechanical Engineering Problems Institute
Russian Academy of Sciences
Bolshoy pr. VO 61
199178 St. Petersburg
Russia
e-mail: ermakov@random.ipme.ru

No. 983
Berlin 2004

2000 *Mathematics Subject Classification.* 60F10.

Key words and phrases. large deviations, moderate deviations, bootstrap, empirical measure.

The paper was written partially during the author staying in WIAS Institute.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Mohrenstraße 39
10117 Berlin
Germany

Fax: + 49 30 2044975
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Abstract

We establish the moderate deviation principle for the common distribution of empirical measure and empirical bootstrap measure (empirical measure obtaining by the bootstrap procedure). For the most widespread statistical functionals depending on empirical measure (in particular differentiable and homogeneous functionals) we compare their asymptotic of moderate deviation probabilities with the asymptotic given by the bootstrap procedure.

1. Introduction. Let S be a Hausdorff space, $\Im$ the σ-field of Borel sets in S and Λ the space of all probability measures (pms) on $(S, \Im)$. Let $X_1, \ldots, X_n$ be i.i.d.r.v.'s taking values in S according to a pm $P \in \Lambda$ and let $\hat{P}_n$ be the empirical probability measure of $X_1, \ldots, X_n$. The distributions of statistics depending on the sample $X_1, \ldots, X_n$ are often analyzed on the base of the bootstrap procedure (see Hall (1992), Mammen (1992), Efron and Tibshirany(1993) and references therein). For given statistics $V(X_1, \ldots, X_n)$, we simulate independent samples $X^*_{1i}, \ldots, X^*_{ni}, 1 \le i \le k$ having the probability measure $\hat{P}_n$ and treat the empirical distribution of $V(X^*_{1i}, \ldots, X^*_{ni}), 1 \le i \le k$ as the estimator of the distribution of $V(X_1, \ldots, X_n)$. What is of special interest, are the estimates of large and moderate deviation probabilities of $V(X_1, \ldots, X_n)$. Such problems constantly emerge in confidence estimation and hypothesis testing. The significant levels in the confidence estimation and the p-values in the hypothesis testing have usually small values and can be often correctly analyzed using the theorems on large and moderate deviations. From this viewpoint it is natural to compare the probabilities of large and moderate deviations of $V(X_1, \ldots, X_n)$ and $V(X^*_1, \ldots, X^*_n)$. In paper we carry out such a comparison for the moderate deviation probabilities in a slightly different setting. The statistics $V(X_1, \ldots, X_n)$ can be usually represented as a functional $T(\hat{P}_n)$ of the empirical measure $\hat{P}_n$, that is, $V(X_1, \ldots, X_n) = T(\hat{P}_n)$. Similarly, $V(X^*_1, \ldots, X^*_n) = T(P^*_n)$, where P^*_n is the empirical probability measure of $X^*_1, \ldots, X^*_n$. Thus, we reduce the problem to the study of moderate deviation probabilities of $T(\hat{P}_n) - T(P)$ and $T(P^*_n) - T(\hat{P}_n)$ on the base of moderate deviation principle.

The problems related to large and moderate deviation probabilities of empirical measures have been treated in many papers (see Sanov, 1957; Groeneboom, Oosterhoff, Ruymgaart, 1979 (GOR); Borovkov and Mogulskii, 1980; Dembo and Zeitouni, 1993; Ermakov, 1995; Eichelsbacher and Schmock, 2002; Arcones, 2003 and references therein). These papers contain complete results proved under rather general assumptions. Our goal is to develop similar techniques for the moderate deviation probabilities of $(P^*_n - \hat{P}_n) \times (\hat{P}_n - P)$ and to make use of these techniques to compare the probabilities of deviations $T(\hat{P}_n) - T(P)$ and $T(P_n^*) - T(\hat{P}_n)$. Thus, we intend to study the asymptotic of the probabilities $P(P_n^* \times \hat{P}_n \in \bar{\Omega}_n)$ with $\bar{P} = P \times P$ as a limiting point of $\bar{\Omega}_n \subset \Lambda^2$. Hereafter we make use of the standard notation. We denote $Q_2 \times Q_1$ the Cartesian product of pms $Q_2, Q_1 \in \Lambda$ and $\Lambda^2 = \Lambda \times \Lambda$ the set of all product measures $Q_2 \times Q_1$ with $Q_2, Q_1 \in \Lambda$.

The large deviation probabilities of empirical bootstrap measure have been studied earlier in Chaganty (1997)and Chaganty, Karandikar (1996). These results were established in terms of topology of weak convergence. In paper we consider the moderate deviation setting for the τ_Φ-topology allowing to study moderate deviations for functionals having unbounded influence functions. Our approach make use of new Arcones (2002) results. The results on large deviations probabilities of $P_n^* \times \hat{P}_n$ are far from being "computable", except for some special cases (see Chaganty (1997)). At the same time the moderate deviation principle allows to find easily the asymptotic and to compare the probabilities of moderate deviations of $T(\hat{P}_n) - T(P)$ and $T(P_n^*) - T(\hat{P}_n)$ for the majority of widespread statistics.

In paper we make use of the following notation. We denote C, c arbitrary positive constants which can have different values even on the same line, $\chi(A)$ the indicator of an event A, and $[t]$ the integral part of a real number t. The integration domain in almost all integrals is the set S. Thus it will be convenient to omit the subscript S and to write such integrals as $\int$ instead of $\int_S$.

2. Main Results We begin with the definition of τ_Φ-topology. Fix a sequence b_n such that $b_n \to 0, nb_n^2 \to \infty, b_{n+1}/b_n \to 1$ as $n \to \infty$. Suppose there are given the set Φ of measurable functions f satisfying the following

$$\lim_{n\to\infty} (nb_n^2)^{-1} \log(nP(|f(X)| > b_n^{-1})) = -\infty \tag{2.1}$$

Define the set Λ_Φ of pms $P \in \Lambda$ such that $\int |f(X)| dP < \infty$ for all $f \in \Phi$. The τ_Φ-topology is the coarsest topology in Λ_Φ that makes continuous for all $f \in \Phi$ the map $\Lambda_\Phi \ni P \to \int f dP$. From now on, all topological notion will be related to the τ_Φ-topology. For any set $\Omega \subset \Lambda_\Phi$ denote $\mathrm{cl}(\Omega)$ and $\mathrm{int}(\Omega)$ the closure and the interior of Ω respectively. Define the τ_Φ-topology in Λ_Φ^2 as the corresponding product topology. If Φ is the set Φ_0 of all bounded measurable functions, the τ_Φ-topology coincides with the τ-topology (see GOR (1979), Eichelsbacher and Schmock (1996)). In what follows, we suppose $\Phi_0 \subset \Phi$.

Define the linear spaces Λ_0 and $\Lambda_{0\Phi}$ induced by all differences $P - Q$ with $P, Q \in \Lambda$ and $P, Q \in \Lambda_\Phi$ respectively. Define the τ_Φ-topologies in $\Lambda_{0\Phi}$ and $\Lambda_{0\Phi}^2$ similarly to that in Λ_Φ and Λ_Φ^2 respectively. For any set $\bar{\Omega}_0 \subset \Lambda_{0\Phi}^2$ denote $\mathrm{cl}(\bar{\Omega}_0)$ and $\mathrm{int}(\bar{\Omega}_0)$ the closure and the interior of $\bar{\Omega}_0$ respectively.

For any $G \in \Lambda_0$ define the rate function

$$\rho_0^2(G : P) = \frac{1}{2} \int \left(\frac{dG}{dP} \right)^2 dP$$

if G is absolutely continuous w.r.t. P and $\rho_0(G : P) = \infty$ otherwise. In statistics the functional $2\rho_0^2$ has the interpretation as the Fisher information. The rate function ρ_0^2 naturally arises in the study of moderate deviation probabilities of empirical measures $\hat{P}_n$ (see Borovkov and Mogulskii (1980); Ermakov (1995) and Arcones (2003)). In the bootstrap setting the rate function ρ_{0b}^2 has slightly more cumbersome definition.

For any $\bar{G} = G_2 \times G_1 \in \Lambda_{0\Phi}^2$ denote

$$\rho_{0b}^2(\bar{G} : P) = \rho_0^2(G_2 : P) + \rho_0^2(G_1 : P).$$

Similarly to the proof of Lemma 2.2 in GOR (1979) it is easy to show that the functions $G \to \rho_0(G : P)$, $\bar{G} \to \rho_{0b}(\bar{G} : P)$ with $G \in \Lambda_{0\Phi}$, $\bar{G} \in \Lambda_{0\Phi}^2$ respectively are τ_Φ lower semicontinuous.

For any set $A \in \Im$ and any charge $G \in \Lambda_0$ denote $|G|(A) = \sup\{G(B) - G(D) : B \subset A, D \subset A\}$. Thus the measure $|G|$ is the variation of charge G.

Let the charges $H, H_n \in \Lambda_{0\Phi}$ satisfy the following assumptions.

A. There hold $P_n = P + b_n H_n \in \Lambda_\Phi, P + b_n H \in \Lambda_\Phi$ and $H_n \to H$ as $n \to \infty$ in τ_Φ-topology.

A1. For any $f \in \Phi$

$$\sup_n \int f^2 dH_n < C < \infty.$$

B1. For any $f \in \Phi$

$$\lim_{n\to\infty} (nb_n^2)^{-1} \log \left(nb_n \int \chi(|f(x)| > b_n^{-1}) d|H_n| \right) = -\infty.$$

Define the charge $O \in \Lambda_{0\Phi}$ such that $O(A) = 0$ for all measurable sets $A \in \Im$. For each $G \in \Lambda_{0\Phi}$ denote $\tilde{G} = O \times G$.

Theorem 2.1. *Assume* A,A1 *and* B1. *Let* $\bar{\Omega}_0 \subset \Lambda_{0\Phi}^2$. *Then the following Moderate Deviation Principle (MDP) holds*

$$\liminf_{n\to\infty} (nb_n^2)^{-1} \log P_n((P_n^* - \hat{P}_n) \times (\hat{P}_n - P_0) \in b_n \bar{\Omega}_0) \ge -\rho_{0b}^2(\mathrm{int}(\bar{\Omega}_0 - \tilde{H}), P) \quad (2.2)$$

and

$$\limsup_{n\to\infty} (nb_n^2)^{-1} \log P_n((P_n^* - \hat{P}_n) \times (\hat{P}_n - P) \in b_n \bar{\Omega}_0) \le -\rho_{0b}^2(\mathrm{cl}(\bar{\Omega}_0 - \tilde{H}), P). \quad (2.3)$$

Remark 2.1. A similar version of theorem on moderate deviation probabilities of empirical measures has been proved in Borovkov and Mogulskii (1980) in the case of τ-topology with $H_n = H = O$.

Remark 2.2. In hypothesis testing the tests behaviour are often analyzed for the alternatives P_n converging to the hypothesis P. Such a setting is considered in

Theorem 2.1. Naturally if we suppose that the charges H_n, H are absent, we get the usual form of moderate deviation theorem. The techniques of moderate deviation theorems with the sequences of pms P_n converging to pm P can be implemented also in the proofs of importance sampling theorems studying the problem of simulation of moderate deviation probabilities.

The analogy of Theorem 2.1 is also valid for the moderate deviations of $P_k^* \times \hat{P}_n$, where P_k^* is the empirical measure of independent sample $X_1^*, \ldots, X_k^*$ distributed with the pm $\hat{P}_n$ and $k = k(n)$, $k/n \to \nu > 0$ as $n \to \infty$.

For any $\bar{G} = G_2 \times G_1 \in \Lambda_0^2$ denote the rate function

$$\rho_{0\nu}^2(\bar{G} : P) = \nu\, \rho_0^2(G_2 : P) + \rho_0^2(G_1 : P).$$

For any $\bar{\Omega}_0 \subset \Lambda_0^2$ we set $\rho_{0\nu}(\bar{\Omega}_0 : P) = \inf\{\rho_{0\nu}(\bar{G} : P) : \bar{G} \in \bar{\Omega}_0\}$.

Theorem 2.2. *Assume* A,A1 *and* B1. *Then the following Moderate Deviation Principle (MDP) holds*

$$\liminf_{n\to\infty}(nb_n^2)^{-1} \log P_n((P_k^* - \hat{P}_n) \times (\hat{P}_n - P_0) \in b_n\bar{\Omega}_0) \ge -\rho_{0\nu}^2(\text{int}(\bar{\Omega}_0 - \tilde{H}), P) \quad (2.4)$$

and

$$\limsup_{n\to\infty}(nb_n^2)^{-1} \log P_n((P_k^* - \hat{P}_n) \times (\hat{P}_n - P) \in b_n\bar{\Omega}_0) \le -\rho_{0\nu}^2(\text{cl}(\bar{\Omega}_0 - \tilde{H}), P). \quad (2.5)$$

The proof of Theorem 2.2 is akin to that of Theorem 2.1 and is omitted. From now on, we assume $k = n$.

The moderate deviation principle for empirical measures holds for the wider zones of moderate deviations. In this setting a version of Theorem 2.1 is valid for the sets Ψ of functions f such that

$$\lim_{n\to\infty} (nd_n^2)^{-1} \log(nP(|f(X)| > nd_n)) = -\infty \quad (2.6)$$

where $d_n \to 0, nd_n^2 \to \infty, d_{n+1}/d_n \to 1$ as $n \to \infty$.

Assume the following.

B2. For any $f \in \Psi$

$$\lim_{n\to\infty} (nb_n^2)^{-1} \sup_{m\ge n} \log\left(nb_n \int \chi(|f(x)| > nb_n) d|H_m|\right) = -\infty.$$

Using the reasoning of Lemma 2.5 in Eichelsbacher and Lowe (2003) we get that B2 implies A1.

Theorem 2.3. *Assume* A *with* $\Phi = \Psi$ *and* B2. *Let* $\Omega_0 \subset \Lambda_{0\Psi}$. *Then, the Moderate Deviation Principle holds*

$$\liminf_{n\to\infty}(nd_n^2)^{-1} \log P_n(\hat{P}_n \in P + d_n\Omega_0) \ge -\rho_0^2(\text{int}(\Omega_0 - H), P_0) \quad (2.7)$$

and

$$\limsup_{n\to\infty}(nd_n^2)^{-1}\log P_n(\hat{P}_n \in P + d_n\Omega_0) \le -\rho_0^2(\mathrm{cl}(\Omega_0 - H), P_0) \tag{2.8}$$

The Proposition 2.4 given below shows that moderate deviation principle often does not hold for the empirical bootstrap measure if (2.1) is replaced by (2.6).

Theorem 2.4.. *Let random variable $Y = f(X), EY = 0$ satisfies (2.6). Let*

$$\lim_{n\to\infty} nP(|Y| > d_n^{-1}) = 0, \tag{2.9}$$

Let a sequence $r_n, d_n^{-1} < r_n < nd_n, r_nd_n \to \infty, nd_n/r_n \to \infty$ as $n \to \infty$ be such that

$$\lim_{n\to\infty} (nd_n^2)^{-1}\log\left(nP\left(r_n < Y < r_{1n}\right)\right) = 0, \tag{2.10}$$

and

$$\lim_{n\to\infty} (r_nd_n)^{-1}\log\left|\frac{\log P(r_n < Y < r_{1n})}{r_nd_n}\right| = 0 \tag{2.11}$$

where $r_n < r_{1n} < d_n, r_{1n}/(nd_n) \to 0$ as $n \to \infty$.

Let $Y_1, \dots, Y_n$ be independent copies of Y and let $Y_1^*, \dots, Y_n^*$ be obtained from $Y_1, \dots, Y_n$ using the bootstrap procedure. Then

$$\lim_{n\to\infty} (nd_n^2)^{-1}\log P\left(\sum_{i=1}^n Y_i^* > nd_n\right) = 0. \tag{2.12}$$

Remark 2.3. Denote $v_n = nd_n/r_n$. Then (2.10) holds if (2.9) fullfilled and $v_n \log v_n = o(nd_n^2)$. Suppose that c.d.f. $F(x) = P(Y < x)$ is continuous strictly monotone function. Define a_n the equation $na_n^2 = 1 - F(na_n)$. It is easy ti verify that if $na_n^2 = n^\gamma, 0 < \gamma < 1$ then one can take $r_n = n^{1/2-\gamma+\epsilon}, 0 < \epsilon < 2\gamma$. Putting $ka_k = r_n$ we get

$$|\log nP(Y > r_n)| = O(ka_k^2) = O(n^{\frac{\gamma(1-\gamma+2\epsilon)}{1+\gamma}}).$$

Thus (2.9) is satisfied.

Arguing similarly one can show that the same statement holds for any sequence $a_n, na_n^2 = \phi(n)$ such that $\phi'(x) = \frac{d\phi}{dx}(x)$ is monotone decreasing function and $\phi'(x) > cx^{\gamma-1}$ with $0 < \gamma < 1$.

Theorem 2.5 given below shows that the moderate deviation principle holds for the empirical bootstrap measures with high probability even if (2.1) or (2.6) does not hold.

Theorem 2.5. *Let $d_n \to 0, nd_n^2 \to \infty$ as $n \to \infty$ and let Ψ be a set of functions f such that*

$$P(|f(X)| > d_n^{-1}) < h(d_n) \tag{2.13}$$

where $nh(d_n) \to 0$ as $n \to \infty$.

Let there exist $t > 2$ and increasing positive function $q(x) > x^t$ such that

$$Eq(f^2(X) - Ef^2(X)) < \infty \tag{2.14}$$

for all $f \in \Psi$.

Then for any $\Omega_0 \subset \Lambda_{0\Psi}$ for any $\epsilon > 0$ and $n > n_0(\epsilon)$ there hold

$$(nd_n^2)^{-1} \log \hat{P}_n(P_n^* \in \hat{P}_n + d_n\Omega_0) \geq -\rho(\text{int}(\Omega_0), P) - \epsilon \tag{2.15}$$

and

$$(nd_n^2)^{-1} \log \hat{P}_n(P_n^* \in \hat{P}_n + d_n\Omega_0) \leq -\rho(\text{cl}(\Omega_0), P) + \epsilon \tag{2.16}$$

with probability

$$\kappa_n = \kappa_n(\epsilon, \Omega_0) = 1 - C(\epsilon, \Omega_0)\left[nh(d_n) + \inf_y \{nq(y)/y + \right]$$

$$\exp\{-\frac{t}{t+2}\frac{n\epsilon}{y}\log\left(\frac{n\epsilon y^{t-1}}{C(\Omega_0)} + 1\right)\Big\} + \exp\{-\delta n \log(\delta h^{-1}(d_n/\delta)) + n\delta\}]$$

where $\delta = \delta(\epsilon, \Omega_0) > 0$.

Remark 2.4. We do not suppose that that the set Ψ contains all functions f satisfying (2.13).

Remark 2.5. The proof utilizes the estimate of rate of convergence $\frac{1}{n}\sum_1^n f^2(X_s)$ to $Ef^2(X)$ for all $f \in \Psi$. To get such an estimate we suppose (2.14) that causes the additional term $\inf_y\{nq(y)/y + \exp\{-\frac{t}{t+2}\ \frac{n\epsilon}{y}\log(\frac{n\epsilon y^{t-1}}{C(\Omega_0)} + 1)\}\}$ in the probability κ_n.

The proofs of Theorems 2.1 and 2.3, Proposition 2.4 and Theorem 2.5 will be given in sections 4,5 and 6 respectively.

In Lemma 2.6 we show that, if (2.6) holds, then (2.6) holds for any sequence $r_n = o(d_n), nr_n^2 \to \infty$ as $n \to \infty$.

Lemma 2.6. *Let (2.6) holds. Then for any sequence $r_n, r_{n+1}/r_n \to 1, r_n/d_n \to 0, nr_n^2 \to \infty$ as $n \to \infty$ it holds*

$$\lim_{n\to\infty} (nr_n^2)^{-1} \log nP(|f(X)| > nr_n) = -\infty. \tag{2.17}$$

Proof. Define the sequence d_{k_n} such that $k_n d_{k_n} \leq nr_n < (k_n + 1)d_{k_n+1}$. We have

$$(nr_n^2)^{-1} \log nP(|f(X)| > nr_n) = \frac{n}{k_n}((k_n{+}1)d_{k_n+1}^2)^{-1} \log nP(|f(X_1)| > (k_n{+}1)d_{k_n+1}) =$$

$$\frac{n}{k_n}((k_n{+}1)d_{k_n+1}^2)^{-1} \log \frac{n}{k_n} + \frac{n}{k_n}((k_n{+}1)d_{k_n+1}^2)^{-1} \log k_n P(|f(X_1)| > (k_n{+}1)d_{k_n+1})) \doteq$$

$$I_{1n} + \frac{n}{k_n} I_{2n}.$$

By (2.6), we have $I_{2n} \to -\infty$ as $n \to \infty$. Thus, if (2.17) does not hold, $\frac{k_n}{n} I_{1n} \to \infty$ as $n \to \infty$. Hence for any $C > 0$ for all $n > n_0(C)$ it holds

$$\left(\log \frac{n}{k_n}\right)^{1/2} > C(k_n+1)^{1/2} d_{k_n+1} > C\frac{n}{(k_n+1)^{1/2}} n^{1/2} r_n.$$

Therefore

$$\left(\frac{k_n}{n} \log \frac{n}{k_n}\right)^{1/2} > C n^{1/2} r_n.$$

Since $n^{1/2} r_n \to \infty$ as $n \to \infty$ we get the contradiction.

3.Examples. In section we establish the asymptotics of moderate deviation probabilities for the differentiable and homogeneous functionals depending on empirical measure $\hat{P}_n$ and empirical bootstrap measure P_n^*. The functionals of such types often emerge in statistics.

Example 3.1. Differentiable statistical functionals. We suppose that the functional $T : \Lambda \to R^1$ admits a linear approximation of the following type.

C. There exist a real function $r \in \Phi$, $\int r dP = 0$, and a seminorm N in Λ_0 continuous in τ_Φ-topology in $\Lambda_{0\Phi}$ satisfying the following. For any $Q \in \Lambda$,

$$\left| T(Q) - T(P) - \int r\, dQ \right| < \omega(N(Q-P)).$$

Hereafter $\omega(t)$ is an increasing function such that $\omega(t)/t \to 0$ as $t \to 0$.

Thus we suppose that the functional $T(Q)$ has the Gato derivative h and such a linear approximation admites the uniform estimate expressed in terms of seminorm N. This assumption is not unnatural. For example, if the fuctional $T(Q)$ has the bounded second Gato derivatives, this assumtion holds. The assumptions of differentiability are the standard tool for the proof of asymptotic normality of statistics $T(\hat{P}_n)$ (see Serfling (1980)) and in implicit form were also used for the study of moderate deviation probabilities (see Jureckova, Kallenberg and Veraverbeke (1988); Inglot, Kallenberg and Ledwina (1990),(1992); and Ermakov (1994)).

If C holds, then, as it follows easily from Theorem 2.1, for any sequence P_n converging to P_0 and satisfying A, A1, B1 we have

$$\lim_{n\to\infty} (nb_n^2)^{-1} \log P_n(T(P_n^*) - T(\hat{P}_n) > b_n) =$$

$$\lim_{n\to\infty} (nb_n^2)^{-1} \log P_n \left(\int r\, d(P_n^* - \hat{P}_n) > b_n \right) =$$

$$-\frac{1}{2} \inf \left\{ \int (g_2^2 + g_1^2)\, dP : \int g_2 r\, dP > 1,\ g_1, g_2 \in L_2(P) \right\} =$$

$$-\frac{1}{2} \left(\int r^2\, dP \right)^{-1}, \qquad (3.1)$$

$$\lim_{n\to\infty}(nb_n^2)^{-1}\log P_n(T(\hat{P}_n)-T(P_n)>b_n)=$$

$$\lim_{n\to\infty}(nb_n^2)^{-1}\log P_n\left(\int r\,d(\hat{P}_n-P_n)>b_n\right)=$$

$$-\frac{1}{2}\inf\left\{\int g^2\,dP:\int gr\,dP>1,\;\; g\in L_2(P)\right\}=-\frac{1}{2}\left(\int r^2 dP\right)^{-1}. \qquad (3.2)$$

Thus, the asymptotics of moderate deviations probabilities of $T(P_n^*)-T(\hat{P}_n)$ and $T(\hat{P}_n)-T(P)$ coincide. At the same time

$$\lim_{n\to\infty}(nb_n^2)^{-1}\log P_n(T(P_n^*)-T(P_n)>b_n)=$$

$$\lim_{n\to\infty}(nb_n^2)^{-1}\log P_n\left(\int rd(P_n^*-P_n)>b_n\right)=$$

$$-\frac{1}{2}\inf\left\{\int(g_2^2+g_1^2)\,dP:\int(g_2-g_1)r\,dP>1,\;\; g_1,g_2\in L_2(P)\right\}=$$

$$-\frac{1}{4}\left(\int r^2 dP\right)^{-1}. \qquad (3.3)$$

The proof of first equality in (3.1) is very easy and (3.2),(3.3) are obtained by a similar technique. Define a sequence C_n such that $C_n\to\infty$, $\omega(C_nb_n)/b_n\to 0$ as $n\to\infty$. By Theorem 3.2 and C, we have

$$P_n\left(\int r\,d(P_n^*-\hat{P}_n)>b_n+\omega(C_nb_n)\right)-P_n(N(P_n^*-\hat{P}_n)>C_nb_n)<$$

$$P_n(T(P_n^*)-T(\hat{P}_n)>b_n)$$

$$<P_n\left(\int r\,d(P_n^*-\hat{P}_n)>b_n-\omega(C_nb_n)\right)+P_n(N(P_n^*-\hat{P}_n)>C_nb_n) \qquad (3.4)$$

and

$$P_n(N(P_n^*-\hat{P}_n)>C_nb_n)<\exp\{-CnC_n^2b_n^2\}.$$

The asymptotic of $P_n(\int r\,d(P_n^*-\hat{P}_n)>b_n)$, given in (3.1), follows directly from Theorem 2.1.

Example 3.2. Variance. Let $T(P)=Var_P[X]=E_P[X^2]-(E_P[X])^2$ and let $S=R^1$. The functional $T(P)$ has the influence function $r(x)=x^2-2xE[X]-E[X^2]+2[EX]^2$ and

$$T(Q)-T(P)-\int r\,dQ=-\left(\int x\,d(Q-P)\right)^2. \qquad (3.5)$$

Thus, if $r\in\Phi$ and $f(x)=x\in\Phi$, we have

$$\lim_{n\to\infty}(nb_n^2)^{-1}\log P(T(P_n^*)-T(\hat{P}_n)>b_n)=$$

$$\lim_{n\to\infty}(nb_n^2)^{-1}\log P(T(\hat{P}_n)-T(P)>b_n)=-\frac{1}{2}(\mathrm{Var}[X^2-2X\,E[X]])^{-1}. \qquad (3.6)$$

Example 3.3. Homogeneous functionals. It is easily seen that the analogues of (3.1)-(3.3) hold also in the case of an arbitrary norm $N:\Lambda_0\to R^1$ such that N is continuous in τ_Φ-topology in $\Lambda_{0\Phi}$

$$\lim_{n\to\infty}(nb_n^2)^{-1}\log P(N(\hat{P}_n-P)>b_n)=-\frac{1}{2}\rho_0^2(\Omega_0:P). \qquad (3.7)$$

$$\lim_{n\to\infty}(nb_n^2)^{-1}\log P(N(P_n^*-\hat{P}_n)>b_n)=$$

$$-\frac{1}{2}\inf\left\{\int g_2^2+g_1^2dP:\ N(G_2)\ge 1;\, g_1=\frac{dG_1}{dP},g_2=\frac{dG_2}{dP};\, G_2,G_1\in\Lambda_0\right\}=$$

$$-\frac{1}{2}\rho_0^2(\Omega_0:P). \qquad (3.8)$$

and

$$\lim_{n\to\infty}(nb_n^2)^{-1}\log P(N(P_n^*-P)>b_n)=-\frac{1}{4}\rho_0^2(\Omega_0:P). \qquad (3.9)$$

Here $\Omega_0=\{G:N(G)>1,G\in\Lambda_0\}$.

In particular, the statements (3.7) and (3.9) are valid for the functional N corresponding to the test statistics of Kolmogorov and omega-square types

$$N(Q-P,P)=\max\{|F(x)-F_0(x)|q(F_0(x)):x\in S\} \qquad (3.10)$$

and

$$N(Q-P,P)=\left(\int_S(F(x)-F_0(x))^2q(F_0(x))dF_0(x)\right)^{1/2} \qquad (3.11)$$

respectively. Here q is a bounded weight function, $S=R^1$, P_0 and F,F_0 are the distribution functions of Q,P respectively. These norms depend on the probability measure P additionally. Thus the statement (3.8) holds only in the case of Kolmogorov test statistic.

Example 3.4. Now we show that the presence of weight function q does not influence seriously on the asymptotic (3.8). Assume the following.

C1. There exists function $\omega(t),\omega(t)/t\to 0$ as $t\to 0$ such that, for all $P,Q,R\in\Lambda_\Phi$

$$|N(Q-P,P)-N(Q-P,R)|\le\omega(\sup_x|\bar{F}(x)-F_0(x)|)$$

where $\bar{F}$ stands for the distribution function of R.

The functionals $N(Q-P,P)$ defined by (3.10),(3.11) satisfy C1 if the function q is continuous in $[0,1]$.

Let $\hat{F}_n$ be the distribution function of $\hat{P}_n$. Then, by Theorem 2.3,

$$P(\omega(\sup_x|\hat{F}_n(x)-F_0(x)|)>cb_n)\le\exp\{-CnC_nb_n^2\}$$

where $C_n \to \infty$ as $n \to \infty$.

Hence, estimating similarly to (3.4), we get

$$\lim_{n\to\infty} (nb_n^2)^{-1} \log P(N(P_n^* - \hat{P}_n, \hat{P}_n) > b_n) =$$

$$\lim_{n\to\infty} (nb_n^2)^{-1} \log P_0(N(P_n^* - \hat{P}_n, P) > b_n) = -\frac{1}{2}\rho_0^2(\Omega_0 : P). \tag{3.12}$$

Example 3.5. Let us find the asymptotic

$$J \doteq \lim_{n\to\infty} (nb_n^2)^{-1} \log P_0(N^\gamma(P_n^* - P) - N^\gamma(\hat{P}_n - P) > b_n)$$

with $\gamma > 0$.

By Theorem 2.1, we get

$$J = \inf \left\{ \int (r^2 + g^2) dP_0 : N^\gamma(G+R) - N^\gamma(G) \geq 1; \right.$$

$$\left. g = \frac{dG}{dP_0}, r = \frac{dR}{dP_0}; G, R \in \Lambda_0 \right\} \doteq \inf V(G,R). \tag{3.13}$$

Since $N(G+R) \leq N(G) + N(R)$, we get

$$J \geq \inf \left\{ \int (r^2 + g^2) dP : (N(G) + N(R))^\gamma - N^\gamma(G) \geq 1; \right.$$

$$\left. g = \frac{dG}{dP}, r = \frac{dR}{dP}; G, R \in \Lambda_0 \right\} \doteq \inf U(G,R). \tag{3.14}$$

Define the charge $H \in \mathrm{cl}(\Omega_0)$ such that $\rho_0(H : P) = \left(\int h^2\, dP\right)^{1/2} = \rho_0(\Omega_0 : P)$ with $h = \frac{dH}{dP}$. Here the set Ω_0 is the same as in example 3.3. It is easy to see that for the fixed G

$$\arg\inf_R U(G,R) = \lambda H \tag{3.15}$$

with constant $\lambda \in R^1$.

Let $r = \lambda h$ and let us consider the problem of minimization of $U(G, \lambda H)$ with respect to G. We begin with the dual problem. Let $N(R) = d$ =const and one need to find

$$\sup \left\{ (N(G) + d)^\gamma - N^\gamma(G) : \int g^2 dP = 1 \right\}$$

Let $\gamma \geq 1$. Since the function $(x+d)^\gamma - x^\gamma$ is convex the supremum is attained on the charge $G_0 = c\tilde{G}$ where $\tilde{G} = \arg\sup\{N(G) : \int g^2 dP_0\}$ and $\tilde{g} = \frac{d\tilde{G}}{dP} = h/\rho_0$. Therefore $\inf\{U(G,R) : G, R \in \Lambda_0\}$ is attained on the charges G, R having the densities $g = ah$, $r = dh$ with $a, d \in R^1$. However $V(aH, dH) = U(aH, dH)$. Hence we get

$$J = \inf\{d^2 + a^2 : (d+a)^\gamma - a^\gamma > 1\} \int h^2(s) ds (1 + o(1)). \tag{3.16}$$

In particular, if $\gamma = 1$, $J = \rho_0^2(\Omega_0, P)$.

If $\gamma < 1$, then $\arg\sup\{(x+d)^\gamma - x^\gamma : x \geq 0\} = 0$. Therefore $U = d^\gamma$ and

$$J = \inf\left\{\int r^2 dP : N^\gamma(R) \geq 1, r = \frac{dR}{dP}\right\} = \rho_0^2(\Omega_0, P).$$

4. Proofs of Theorem 2.1 and 2.3. For each $r > 0$ define the sets $\Gamma_{0r} = \{G \in \Lambda_0 : \rho_0(G : P) \leq r\}$ and $\Gamma_r = \{\bar{G} \in \Lambda_0^2 : \rho_{0b}(\bar{G} : P) \leq r\}$.

Lemma 4.1. Let (2.6) hold. Then

i. $\Gamma_r \subset \Lambda_{0\Psi}^2$,

ii. the set Γ_r is τ_Ψ-compact and sequentially τ_Ψ-compact set in $\Lambda_{0\Phi}^2$.

Proof. Let $\phi \in \Psi$. Then, by Lemma 2.5, in Eichelsbacher and Lowe (2003), there holds

$$\int \phi^2(x) dP < \infty. \tag{4.1}$$

For any charge $\bar{G} = G_1 \times G_2 \in \Lambda_0^2$ and any measurable set $A \subset S$ we have

$$\int_A |\phi_1| d|G_1| + \int_A |\phi_2| d|G_2| \leq$$

$$\alpha\left(\int_A \phi_1^2 dP + \int_A \phi_2^2 dP\right) + \alpha^{-1}\left(\int_A \left(\frac{dG_1}{dP}\right)^2 dP + \int_A \left(\frac{dG_2}{dP}\right)^2 dP\right) \tag{4.2}$$

for all $\alpha > 0$. This implies i if $A = S$.

Fix $\epsilon > 0$. Let $\alpha = r/\epsilon$ and $n = n(\epsilon)$ is such that

$$\frac{r}{\epsilon}\left(\int_{|\phi_1|>n} \phi_1^2 dP + \int_{|\phi_2|>n} \phi_2^2 dP\right) < \epsilon$$

Then, by (4.2), we get

$$\int |\phi_1| d|G_1| + \int |\phi_2| d|G_2| - \int_{|\phi_1|<n} |\phi_1| d|G_1| + \int_{|\phi_2|<n} |\phi_2| d|G_2| < 2\epsilon$$

Hence the map $\Gamma_r \ni \bar{G} = G_1 \times G_2 \to \int |\phi_1| d|G_1| + \int |\phi_2| d|G_2|$ is τ_Ψ-continuous as the uniform limit of functions

$$\int_{|\phi_1|<n} \phi_1 dG_1 + \int_{|\phi_2|<n} \phi_2 dG_2.$$

This implies that the τ and τ_Ψ-topologies coincide in Γ_r. Since the sets Γ_{0r} and $\Gamma_r \subset \Gamma_{or}^2$ are τ-compact and sequentially τ-compact these sets are τ_Ψ-compact and sequentially τ_Ψ-compact as well. This completes the proof of Lemma 4.1.

Note that τ_Ψ continuity implies τ_Φ continuity. Hence the sets Γ_r and Γ_{0r} are τ_Φ compacts as well.

For any $u, v \in R^k$ denote $u'v$ the inner product of u and v. For any $f \in \Phi$ and any charge $G \in \Lambda_{0\Phi}$ denote $< f, G >= \int f dG$.

Let $f_1, \ldots, f_{k_1}, g_1, \ldots, g_{k_2} \in \Phi$ and $G \in \Lambda_{0\Phi}$ Let $Ef_i(X) = 0, Eg_j(X) = 0, 1 \le i \le k_1, 1 \le j \le k_2$. Define the covariance matrices $R_f = \{E[f_i(X)f_j(X)]\}_{i,j=1}^{k_1}$ and $R_g = \{E[g_i(X)g_j(X)]\}_{i,j=1}^{k_2}$.

Denote $f = \{f_i\}_{i=1}^{k_1}$ and $g = \{g_i\}_{i=1}^{k}$.

By Dawson-Gartner Theorem (see Dembo and Zeitouni (1993)), Theorem 2.1 follows from Lemma 4.2 given bellow.

Lemma 4.2. *Assume (2.1) and* A,A1,B1. *Then, for the random vectors* $U_n(\bar{X}) = \left(\frac{1}{n}\sum_{i=1}^n f_1(X_i), \ldots, \frac{1}{n}\sum_{i=1}^n f_{k_1}(X_i),\ \frac{1}{n}\sum_{i=1}^n g_1(X_i^*), \ldots, \frac{1}{n}\sum_{i=1}^n g_{k_2}(X_i^*)\right)$ *the MDP holds, that is, for any* $\Omega \subset R^{k_1+k_2}$

$$\liminf_{n\to\infty}(nb_n^2)^{-1}\log P_n(U_n(\bar{X}) \in b_n\Omega) \ge -\inf_{x\in \mathrm{int}(\Omega)} x'I_{f,g}x \tag{4.3}$$

and

$$\limsup_{n\to\infty}(nb_n^2)^{-1}\log P_n(U_n(\bar{X}) \in b_n\Omega) \le -\inf_{x\in \mathrm{cl}(\Omega)} x'I_{f,g}x \tag{4.4}$$

where for any $x = (y, z) \in R^{k_1+k_2}, y \in R^{k_1}, z \in R^{k_2}$

$$x'I_{f,g}x = \sup_{t\in R^{k_1}, s\in R^{k_2}} \left(t'y + s'z - < t'f, H > -\frac{1}{2}t'R_f t - \frac{1}{2}s'R_g s\right).$$

Note that, if there exist R_f^{-1} *and* R_g^{-1}, *then*

$$x'I_{fg}x = \frac{1}{2}((y- < f, H >)'R_f^{-1}(y- < f, H >) + \frac{1}{2}z'R_g^{-1}z.$$

Lemma 4.2 follows from Lemmas 4.3 and 4.4 given below.

Lemma 4.3. *Assume (2.1). Then for any* $C > 0$

$$\lim_{n\to\infty}(nb_n^2)^{-1}\log P_n(\max_{1\le i\le k_1}\max_{1\le j\le n}|f_i(X_j)| > cb_n^{-1}) = -\infty, \tag{4.5}$$

$$\lim_{n\to\infty}(nb_n^2)^{-1}\log P_n(\max_{1\le i\le k_2}\max_{1\le j\le n}|g_i(X_j^*)| > cb_n^{-1}) = -\infty. \tag{4.6}$$

Proof. We have

$$P_n(\max_{1\le i\le k_1}\max_{1\le j\le n}|f_i(X_j)| > cb_n^{-1}) \le n\sum_{i=1}^{k_1} P_n(|f(X_1)| > cb_n^{-1})$$

By (2.1) and B1, this implies (4.5).

Hence we have

$$P_n(\max_{1\le i\le k_2}\max_{1\le j\le n}|g_i(X_j)|>cb_n^{-1})=O(\exp\{-Cnb_n^2\})$$

for any $C>0$. This implies (4.6).

For any $h\in\Phi$ denote $h_n(x)=h(x)\chi(|h(x)|<b_n^{-1})$.

Lemma 4.4. *Let $f_1,\dots,f_{k_1},g_1,\dots,g_{k_2}\in\Phi$. Then, for the random vectors $\tilde U_n(\bar X)=\left(\frac{1}{n}\sum_{i=1}^n f_{1n}(X_i),\dots,\frac{1}{n}\sum_{i=1}^n f_{k_1n}(X_i),\frac{1}{n}\sum_{i=1}^n g_{1n}(X_i^*),\dots,\frac{1}{n}\sum_{i=1}^n g_{k_2n}(X_i^*)\right)$ the MDP holds, that is, (4.3) and (4.4) are valid with $U_n(\bar X)=\tilde U_n(\bar X)$.*

By Gartner-Ellis Theorem (see Dembo and Zeitouni (1993)) Lemma 4.4 follows from Lemma 4.5 given below.

Lemma 4.5. *Let $f_i\in\Phi, g_j\in\Phi$ for all $1\le i\le k_1, 1\le j\le k_2$. Then*

$$\lim_{n\to\infty}(nb_n^2)^{-1}\log E_n\exp\left\{b_n\sum_{i=1}^n t'f_n(X_i)+b_n\sum_{i=1}^n s'(g_n(X_i^*)-\bar g_n)\right\}=$$

$$<t'f,H>-\frac{1}{2}t'R_ft-\frac{1}{2}s'R_gs$$

where $\bar g_n=\left(\frac{1}{n}\sum_{s=1}^n g_{1n}(X_s),\dots,\frac{1}{n}\sum_{s=1}^n g_{k_2n}(X_s)\right)$, $f_n=\{f_{in}\}_1^{k_1}, g_n=\{g_{jn}\}_1^{k_2}$.

Proof. We have

$$E_n\exp\left\{b_n\sum_{i=1}^n t'f_n(X_i)+b_n\sum_{i=1}^n s'(g_n(X_i^*)-\bar g_n)\right\}=$$

$$E_n\exp\left\{b_n\sum_{i=1}^n t'f_n(X_i)\right\}\prod_{i=1}^n(1+b_n(s'(g_n(X_i^*)-\bar g_n))+$$

$$\frac{b_n^2}{2}(s'(g_n(X_i^*)-\bar g_n))^2+O\left(\frac{b_n^3}{6}(s'(g_n(X_i^*)-\bar g_n))^3\right)\Bigg)=$$

$$E_n\left[\exp\left\{b_n\sum_{i=1}^n t'f_n(X_i)\right\}\left(1+\frac{b_n^2}{2n}\sum_{i=1}^n(s'(g_n(X_i)-\bar g_n))^2+\right.\right.$$

$$\left.\left.O\left(\frac{b_n^3}{n}\sum_{i=1}^n(s'(g_n(X_i)-\bar g_n))^3\right)\right)^n\right]\doteq I_n. \tag{4.7}$$

By straightforward calculations, we get

$$E_{\hat P_n}(s'(g_n(X_i^*)-\bar g_n))^2=\frac{1}{n}\sum_{i=1}^n(s'(g_n(X_i)-\bar g_n))^2=$$

$$\frac{1}{n}\sum_{i=1}^n(s'(g_n(X_i)-E_n[g_n(X_1)]))^2-(s'\bar g_n-E_n[s'g_n(X_1)])^2. \tag{4.8}$$

We have

$$\left|\sum_{i=1}^{n}(s'(g_n(X_i)-\bar g_n))^3\right| \le 8\sum_{i=1}^{n}|s'(g_n(X_i)-E_n[g_n(X_1)])|^3+$$

$$8n|s'(\bar g_n - E_n[g_n(X_1)])|^3 \doteq 8V_1+8nV_2. \qquad (4.9)$$

We have

$$b_n^3|V_1| = b_n^3\sum_{i=1}^{n}|s'(g_n(X_i)-E_n[g_n(X_1)])|^3\chi(|s'(g_n(X_i)-E_n[g_n(X_1)])| \le \epsilon b_n^{-1})+$$

$$b_n^3\sum_{i=1}^{n}|s'(g_n(X_i)-E_n[g_n(X_1)])|^3\chi(|s'(g_n(X_i)-E_n[g_n(X_1)])| \ge \epsilon b_n^{-1}) \le$$

$$2\epsilon|s|b_n^2\sum_{i=1}^{n}|s'(g_n(X_i)-E_n[g_n(X_1)])|^2+$$

$$8\epsilon^3|s|^3\sum_{i=1}^{n}\chi(|s'(g_n(X_i)-E_n[g_n(X_1)])| \ge \epsilon b_n^{-1}) \qquad (4.10)$$

and

$$b_n^3V_2 \le 4|s|^2 b_n|s'(\bar g - E_n[g_n(X_1)])|. \qquad (4.11)$$

By (4.8)-(4.10), we get

$$I_n \le E_n\exp\left\{b_n\sum_{i=1}^{n}t'f_n(X_i)+\frac{b_n^2}{2}\sum_{i=1}^{n}(s'g_n(X_i)-\right.$$

$$E_n[s'g_n(X)])^2(1+\epsilon_n)+\frac{nb_n^2}{2}(s'\bar g_n - E_n[s'g_n(X)])^2+$$

$$\left.+O\left(\epsilon^3|s|^3\sum_{i=1}^{n}\chi(|s'(g_n(X_i)-E_n[g_n(X_1)])| \ge \epsilon b_n^{-1})\right)+O(nb_n^3V_2)\right\}$$

$$\doteq E_n[W_{1n}\exp\{O(nb_n^3V_n)\}] \doteq E_n[W_{1n}W_{2n}] \doteq \tilde I_n \qquad (4.12)$$

where $\epsilon = \epsilon_n \to 0$ as $n \to \infty$.

Define the events $A_n = \{X_1,\dots,X_n : s'\bar g_n - E_n[s'g_n(X_1)] < rb_n\}$ and the complement of this event $\bar A_n$.

We can write

$$\tilde I_n = E_n[W_{1n}W_{2n}\chi(A_n)] + E_n[W_{1n}W_{2n}\chi(\bar A_n)] \doteq U_{1n}+U_{2n}. \qquad (4.13)$$

We have

$$\log[U_{1n}] \le n\log E_n\left[\exp\left\{b_nt'f_n(X_1)+\frac{b_n^2}{2}(s'g(X_1)-E_n[s'g(X_1)])^2(1+\epsilon_n)+\right.\right.$$

$$\chi(|s'(g_n(X_i) - E_n[g_n(X_1)])| \geq \epsilon b_n^{-1}) + O(r^3 b_n^6)\}] \leq$$

$$n \log E_n[1 + b_n(t' f_n(X_1)) + \frac{b_n^2}{2}((t' f_n(X_1))^2 +$$

$$(s' g_n(X_1) - E_n[s' g_n(X_1)])^2 (1 + \delta_n)) + O(\omega_n) + O(r^3 b_n^6)]$$

where $\omega_n = \omega_{1n} + \omega_{2n} + \omega_{3n} + \omega_{4n} + \omega_{5n}$ with

$$\omega_{1n} = \frac{b_n^3}{6}(t' f(X_1))^3, \quad \omega_{2n} = 3\frac{b_n^3}{2}(t' f(X_1))(s' g(X_1) - E_n(s' g(X_1))^2,$$

$$\omega_{3n} = \frac{b_n^4}{8} E_n(s' g(X_1) - E_n(s' g(X_1))^4, \quad \omega_{4n} = \frac{b_n^4}{12}(t' f(X_1))^2 (s' g(X_1) - E_n(s' g(X_1))^2,$$

$$\omega_{5n} = \chi(|s'(g_n(X_1) - E_n[g_n(X_1)])| \geq \epsilon b_n^{-1})$$

By (2.1), we get

$$E_n[\omega_{1n}] \leq \frac{\epsilon |t| b_n^2}{6} E_n(t' f(X_1))^2, \quad E_n[\omega_{2n}] \leq \frac{\epsilon |t| b_n^2}{6} E_n(s' g(X_1) - E_n(s' g(X_1)))^2,$$

$$E_n[\omega_{3n}] \leq \frac{\epsilon^2 |s|^2 b_n^2}{6} E_n(s' g(X_1) - E_n(s' g(X_1)))^2,$$

$$E_n[\omega_{4n}] \leq \frac{\epsilon^2 |t|^2 b_n^2}{24} E_n(s' g(X_1) - E_n(s' g(X_1)))^2$$

and

$$E_n[\omega_{5n}] \leq \epsilon^{-2} b_n^2 E_n[(s'(g_n(X_1) - E_n[g_n(X_1)]))^2 \times$$

$$\chi(|s'(g_n(X_i) - E_n[g_n(X_1)])| \geq \epsilon b_n^{-1})] = o(\epsilon^{-2} b_n^2)$$

where the last equality holds by A and (4.1).

Hence, we get

$$\log(U_{1n}) \leq -\frac{n b_n^2}{2} \left(2 < t' f, H > -t' R_f t - s' R_g s\right) (1 + O(1)) \doteq v_n \tag{4.14}$$

Note that, by Theorem 2.4 in Arcones (2002), we have

$$P_n(s' \bar{g} - E_n[s' g(X_1)] > r b_n) \leq \exp\{-c r^2 b_n^2\} \tag{4.15}$$

for each $r > 0$.

For the proof of (4.15) it suffices to note that (2.1) implies (2.6) and (2.6) implies

$$\lim_{n \to \infty} (n r^2 b_n^2)^{-1} \log(n P(|f(X)| > r n b_n)) = -\infty$$

for $r > 1$. The case $r < 1$ follows from Lemma 2.6.

By the Hoelder inequality, we get

$$U_{2n} \leq (E_n[W_n^{1+\delta}])^{\frac{1}{1+\delta}} (E_n[\exp\{\delta n b_n^3 V_2\} \chi(\bar{A}_n)])^{\frac{1}{\delta}} \leq$$

$$(E_n[W_n^{1+\delta}])^{\frac{1}{1+\delta}}(E_n[\exp\{2\delta nb_n^3V_2\}])^{\frac{1}{2\delta}}(P_n(\bar{A}_n))^{\frac{1}{2\delta}}. \quad (4.16)$$

We have

$$E_n[\exp\{2\delta nb_n^3V_2\}] \le E_n\exp\{2\delta nb_n\sum_{i=1}^n s'(g_n(X_i)-E_n[g_n(X_1)])\}+$$

$$E_n\exp\{-2\delta nb_n\sum_{i=1} ns'(g_n(X_i)-E_n[g_n(X_1)])\} \doteq U_{21n}+U_{22n}. \quad (4.17)$$

We have

$$\log U_{21n} = n\log E_n\Big[1+\frac{\delta^2b_n^2}{2}(s'g_n(X_1)-E_n[s'g_n(X_1)])^2+$$

$$O(\delta^3b_n^3(s'g_n(X_1)-E_n[s'g_n(X_1)])^3\Big] \le$$

$$n\log E_n[1+C(1+|s|)\delta^2b_n^2(s'g_n(X_1)-E_n[s'g_n(X_1)])^2] \le Cn\delta^2b_n^2s'R_gs. \quad (4.18)$$

Estimating similarly we get

$$\log U_{21n} \le Cn\delta^2b_n^2s'R_gs \quad (4.19)$$

Estimating $E_n[W_n^{1+\delta}]$ similarly to U_{1n} we get

$$E_n[W_n^{1+\delta}] \le \exp\{(1+\delta)^2v_n(1+o(1))\}. \quad (4.20)$$

By (4.16)-(4.20), we get

$$U_{2n} \le \exp\{(1+\delta)|v_n|+Cnb_n^2(1+|s|)s'R_gs-(2\delta^{-1})nr^2b_n^2(1+o(1))\}. \quad (4.21)$$

Since the choice of r is arbitrary we get $U_{2n}=o(U_{1n})$ for sufficiently large r.

The proof of lower bound is based on the inequality $I_n \ge E_n[W_{1n}\exp\{-nb_n^3V_2\}]$. The further estimates are similar to the proof of upper bound and are omitted. This completes the proof of Lemma 4.5.

The proof of Theorem 2.3 follows the same arguments and ulilizes the reasoning of Lemma 4.4 together with Lemma 4.6 given bellow.

For any $h \in \Psi$ denote $\tilde{h}_n(x) = h(x)\chi(b_n^{-1} < |h(x)| < nb_n)$.

Lemma 4.6. *Let $f \in \Psi$ Then, for any $\delta > 0$,*

$$\lim_{n\to\infty}(nb_n^2)^{-1}\log P_n\left(\frac{1}{n}\sum_{i=1}^n\tilde{f}_n(X_i) > \delta b_n\right) = -\infty.$$

Lemma 4.6 was proved in Arcones (2003) in the case of $P_n = P$ (see (2.8) in Arcones (2003)). The presence of $\sup_{m>n}$ in B2 and A with $\Phi = \Psi$ allows to repeat the arguments of the proof of (2.8) in Arcones in the setting Lemma 4.6.

5. Proof of Theorem 2.4. Define the events $A_{ni} = U_{ni} \cup V_{ni}, 1 \le i \le n$ with $U_{ni} = \{y_i : |y_i| < d_n^{-1}\}$ and $V_{ni} = \{y_i : r_n < y_i < r_{1n}\}$. Denote $A_n = \cap_{i=1}^n A_{ni}$. By (2.9), we get

$$P(A_n) > 1 - P(\max_{1 \le i \le n} |Y_i| > d_n^{-1}) >$$

$$1 - nP(|Y_1| > d_n^{-1}) = 1 + o(1). \tag{5.1}$$

Denote P_{cn} the conditional probability measure Y_1 under the condition $Y_1 \in A_{n1}$. By (5.1), it suffices to prove (2.12) if pm P is replaced by pm P_{cn}. Denote $p_n = P_{cn}(r_n < Y_i < r_{1n})$. Define the events $W_n(k_n) = \{Y_1, \dots, Y_n : n - k_n$ random variables $Y_1, \dots, Y_n$ belong $(-d_n^{-1}, d_n^{-1})$ and k_n random variables belong $(r_n, r_{1n})\}$. Suppose that $k = k_n \to \infty, \frac{k_n}{n} \to 0$ as $n \to \infty$. We have

$$v_n \doteq P_{cn}(W_n(k)) = \frac{n!}{(n-k)!k!} p_n^k (1-p_n)^{n-k}(1+o(1)) =$$

$$\exp\{n \log n - (n-k)\log(n-k) - k \log k(1+o(1)) + k \log p_n + (n-k)\log(1-p_n)\} =$$

$$\exp\left\{-(n-k)\log\frac{n-k}{n(1-p_n)} - k \log \frac{k}{np_n}(1+o(1))\right\} =$$

$$\exp\{-n(1-k/n)(-k/n+p_n)(1+o(1)) + k\log[k/(np_n)](1+o(1))\} =$$

$$\exp\{(k_n - np_n - k_n\log(k_n/(np_n))(1+o(1))\}. \tag{5.2}$$

It follows from (2.10),(5.2) that we can choose $k_n \to \infty, nk_np_n \to 0$ as $n \to \infty$ such that

$$o(nd_n^2) = |\log v_n| = k_n|\log(np_n)|(1+o(1)). \tag{5.3}$$

Let us consider the asymptotic of $\sum_{i=1}^n Y_i^*$ if $W_n(k_n)$ holds and l_n random variables $Y_i^*, 1 \le i \le n$ belong $(\frac{1}{2}r_n, 2r_n)$. By Lemma 2.3 in Arcones (2003),

$$P_{nc}\left(\sum_{i=1}^{n-l_n} Y_i^* < cnd_n ||Y_i^*| < d_n^{-1}, 1 \le i \le n - l_n\right) = 1 - o(1). \tag{5.4}$$

By (5.2), we get

$$P(l_n > u_n) = \exp\left\{-u_n \log\frac{u_n}{k_n}(1+o(1))\right\}. \tag{5.5}$$

Thus, if $u_n = c\frac{nd_n}{r_n} = c\frac{nd_n^2}{r_nd_n}$, then, by (5.3),

$$\log\frac{u_n}{k_n} \le \log\frac{c\log(np_n)}{r_nd_n}$$

Hence, by (5.5),(2.11), we get

$$P\left(\sum_{i=1}^{l_n} Y_i^* > cnd_n|Y_i \in V_{ni}, n - l_n \le i \le n\right) > P(l_n > u_n) =$$

$$\exp\left\{-\frac{nd_n^2}{r_nd_n}\log\frac{c\log(np_n)}{r_nd_n}\right\} = \exp\{-o(nd_n^2)\}. \tag{5.6}$$

Now (2.12) follows from (5.1),(5.4),(5.6).

6.Proof of Theorem 2.5. We begin with the proof of upper bound (2.16). Denote $\eta = \rho_0^2(\mathrm{cl}(\Omega_0), P)$ and fix $\delta, 0 < \delta < \eta$. For any $f_1, \ldots, f_l \in \Phi$, $G \in \Gamma_{0,\eta-\delta}$ and $\gamma > 0$ denote

$$U(f_1, \ldots, f_l, G, \gamma) = \left\{R : \left|\int f_i d(R-G)\right| < \gamma, R \in \Lambda_{0\Phi}, 1 \le i \le l\right\}.$$

Since $\Lambda_{0\Phi}$ is Hausdorf space, the space $\Lambda_{0\Phi}$ is regular space (Theorem B2 in Dembo and Zeitouni (1993)). Thus for each $G \in \Gamma_{0,\eta-\delta}$ there exists $U(f_1, \ldots, f_l, G, \gamma) \subset \Lambda_{0\Phi} \setminus \mathrm{cl}(\Omega_0)$. The set $\Gamma_{0,\eta-\delta}$ is compact. Therefore there exists finite covering $\Gamma_{0,\eta-\delta}$ by the sets $U_1 = U(f_{11}, \ldots, f_{1l_1}, G_1, \gamma_1), \ldots, U_k = U(f_{k1}, \ldots, f_{kl_k}, G_k, \gamma_k)$.

Hence the set $\Lambda_{0\Phi}\setminus\Gamma_{0\eta}$ can be covered a finite number of sets $\tilde{U}_i = \tilde{U}(h_{1i}, \ldots, h_{m_i i}, G_i, \gamma_{1i}, \ldots, \gamma_{m_i i})$

$$\tilde{U}_i = \left\{R : \int h_{ji} d(R - G_i) > \gamma_{ji}, R \in \Lambda_{0\Phi}, 1 \le j \le m_i\right\}.$$

with $1 \le i \le t$.

Thus it remains to show that

$$(nd_n^2)^{-1}\log \hat{P}_n\left(\int f d(P_n^* - \hat{P}_n - d_nG) > -\gamma d_n\right) \le -\frac{(\gamma + \int f dG)^2}{\text{Var f(Y)}} - \epsilon \tag{6.1}$$

with probability $1 - c\kappa_n(\epsilon, U(f, G, \gamma))$ for all $f \in \Phi$ and $n > n_0(\epsilon, f)$.

By (2.13), it suffices to prove (6.1) if the condition

$$\max_{1\le s\le n} |f(X_s)| < d_n^{-1} \tag{6.2}$$

holds.

Denote $s_n^2 = \frac{1}{n}\sum_{i=1}^n f^2(X_i) - \left(\frac{1}{n}\sum_{i=1}^n f(X_s)\right)^2$.

By Corollary 2 in Fuc and Nagaev (1971) we get

$$P(|s_n^2 - \mathrm{Var} f(X)| > \epsilon) < \inf_y \left[nq(y)/y + \exp\left\{-\frac{t}{t+2}\frac{n\epsilon}{y}\log\left(\frac{n\epsilon y^{t-1}}{C_q} + 1\right)\right\}\right].$$

Thus to prove (6.1) we can suppose that

$$|s_n^2 - \mathrm{Var}[f(Y)]| < \epsilon. \tag{6.3}$$

Denote $d_nt_n = b_n + \gamma d_n \int f dG$.

Let (6.2),(6.3) hold. To prove (6.1) we apply the following Theorem (see Ermakov (1999)).

Theorem 6.1. *Let $Y_1, \ldots, Y_n$ be i.i.d.r.v.'s. Let $EY = 0$ and $EY^2 = \sigma^2$. Let*

$$E[\exp\{d_n|Y|(1+\epsilon)\}] < C_1 < \infty \tag{6.4}$$

with $\epsilon > 0$ and

$$E|Y|^3 < C_2 d_n^{-1}\omega(d_n) \tag{6.5}$$

where $\omega(x) \to 0$ as $x \to 0$.

Then

$$(nd_n^2)^{-1}\log P\left(\sum_{i=1}^n Y_i > nd_n\right) = -\frac{1}{2}\sigma^{-2}(1 + O(\omega(d_n))) \tag{6.6}$$

where the remainder term in (6.6) is uniform w.r.t. pms P satisfying (6.6),(6.7) with the same constants C_1, C_2.

It follows from (6.2) that (6.6) holds.

By Lemma 2.3 in Arcones (2003) and (2.13),(2.14)

$$P(|\bar f| > r_n|\ |f(X_i)| < d_n^{-1}, 1 \le i \le n) \le \exp\left\{-\frac{nd_n^2}{2\sigma_f^2}(1 + o(1))\right\} \tag{6.7}$$

where $\sigma_f^2 = Ef^2(X)$.

Therefore we can suppose that the addendums with $\bar f$ are negligible in $\sum_{i=1}^n (f(X_i) - \bar f)^3$. Thus it suffices to estimate

$$\sum_{i=1}^n f^3(X_i) = \sum_{i=1}^n f^3(X_i)\chi(|f(X_i)| < \delta d_n^{-1}) +$$

$$\sum_{i=1}^n f^3(X_i)\chi(\delta d_n^{-1} < f(X_i) < d_n^{-1}) \doteq I_1 + I_2. \tag{6.8}$$

We have

$$I_1 \le \delta d_n^{-1} s_n^2. \tag{6.9}$$

Denote $k_n = \sum_{i=1}^n \chi(\delta d_n^{-1} < f(X_i) < d_n^{-1})$.

We have

$$P(k_n > \delta n) \le \exp\{-t\delta n\}(1 + p_n e^t)^n \le \exp\{-t\delta n + np_n e^t\} =$$

$$\exp\{-\delta n \log(\delta/p_n) + \delta n\}.$$

where $p_n = P(f(X_1) > \delta d_n^{-1}) \le h(d_n/\delta)$ and $t = \log(\delta/p_n)$.

Hence

$$I_2 < \delta d_n^{-1}\sum_{i=1}^n f^2(X_i) \tag{6.10}$$

with probability $\kappa_n(\epsilon, U(f, G, \gamma))$.

By (6.9),(6.10), we get

$$\sum_{i=1}^{n} f^3(X_i) < \delta d_n^{-1} \sum_{i=1}^{n} f^2(X_i).$$

with probability $\kappa_n(\epsilon, U(f, G, \gamma))$..

Hence (6.4),(6.5) holds with $P = \hat{P}_n$ that completes the proof of (6.1).

Since

$$\inf_{G,G_0} \left\{ \frac{(\int f dG_0 - \gamma)^2}{\sigma_f^2} : \int \left(\frac{d(G+G_0)}{dP} \right)^2 dP > \eta, \int f dG = \gamma \right\} =$$

$$\inf_{G} \left\{ \frac{(\int f dG)^2}{\sigma_f^2} : \int \left(\frac{dG}{dP} \right)^2 dP > \eta \right\} = \eta$$

is attained with $\frac{dG}{dP} = \eta^{1/2} \sigma_f^{-1} f$ then (6.1) implies the upper bound.

The proof of lower bound (2.15) is based on standard arguments (see GOR (1979)). For each $\delta > 0$ there exists open set $U = U(f_1, \dots, f_l, G, \gamma)$ such that $U \subset \mathrm{int}(\Omega_0)$ and $\rho_0^2(U, P) < \eta + \epsilon$. Hence it suffices to find the lower bound of

$$(nd_n^2)^{-1} \log \hat{P}_n(P_n^* \in \hat{P}_n + d_n U)$$

if (6.2) and (6.3) hold.

By (6.7), for any $\epsilon_1 > 0$ we get

$$P_c(\hat{P}_n \in P + d_n U(f_1, \dots, f_l, O, \epsilon_1)) = 1 - \exp\{-cnd_n^2(1 + o(1))\}$$

where P_c is the conditional distribution of $X_1, \dots, X_n$ if

$$\max_{1 \le i \le l} \max_{1 \le s \le n} |f_i(X_s)| < d_n^{-1}$$

holds.

Thus, in what follows, we can suppose that $\hat{P}_n \in P + d_n U(f_1, \dots, f_l, O, \epsilon_1)$.

Denote $U_1 = U(f_1, \dots, f_l, G, \gamma - \epsilon)$. Then

$$\hat{P}_n(P_n^* \in \hat{P}_n + d_n U) > \hat{P}_n(P_n^* \in \hat{P}_n + d_n U_1).$$

Suppose that

$$\lambda(f_1) \doteq \sigma_{f_1}^{-2} \left(\int f_1 dG + \gamma + \epsilon_1 \right)^2 < \sigma_{f_i}^{-2} \left(\int f_i dG + \gamma + \epsilon_1 \right)^2 \qquad (6.11)$$

for all $2 \le i \le l$.

If the equality in (6.11) is attained for some $i, 2 \le i \le l$ we can replace the set U_1 another set $U_2 = U(f_1, \dots, f_{i-1}, (1+\delta) f_i, f_{i+1}, \dots, f_l, G, \gamma - \epsilon)$ with $\delta > 0$.

The probability $\hat{P}_n(P_n^* \in \hat{P}_n + d_n U_1)$ can be represented as linear combination of probabilities

$$\hat{P}_n\left(\int f_i d(P_n^* - \hat{P}_n - d_n G) > -(\gamma - \epsilon) d_n\right)$$

with $1 \le i \le l$ and

$$\hat{P}_n\left(\int f_i d(P_n^* - \hat{P}_n - d_n G) > (\gamma - \epsilon) d_n\right)$$

with $1 \le i \le l$.

By (6.1), all these probabilities with $f_i, 2 \le i \le l$ have the smaller order then $\exp\{-n d_n^2 (\lambda(f_1) - \epsilon)(1 + o(1))\}$.

Thus it suffices to show

$$\hat{P}_n\left(\int f_1 d(P_n^* - \hat{P}_n - d_n G) > -(\gamma - \epsilon) d_n\right) \ge \exp\{-n d_n^2 (\lambda(f_1) - \epsilon)(1 + o(1))\}$$

and this statement follows from Theorem 6.1 using the same reasoning as in the proof of upper bound.

References

1. Arcones M.A. (2002) Moderate deviations of empirical processes. In *Stochastic Inequalities and Applications* 189-212 ed.E.Gine, C.Houdre and D.Nualart. Birkhauser Boston

2. Bickel, P.J. and Freedman, D.A. (1981) Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196-1217.

3. Borovkov,A.A. and Mogulskii,A.A. (1980) On probabilities of large deviations in topological spaces. II. *Sibirsk. Math.Zh.* **21**(5) 12-26.

4. Chaganty,N.R. (1997) Large deviations for joint distributions and statistical applications. *Sankhya,A.*, **59** 147-166.

5. Chaganty,N.R. and Karandikar,R.L. (1996) Some properties of the Kullback-Leibler number. *Sankhya,A.*, **58** 69-80.

6. Dembo,A. and Zeitouni,O. (1993) *Large Deviations Techniques and Applications*. Jones and Bartlett, Boston.

7. Efron,B. and Tibshirany,R.J. (1993) *An Introduction to the Bootstrap*. Chapman and Hall, London, New York.

8. Eichelsbacher,P. and Schmock,U. (2002) Large deviations of products of empirical measures in strong topologies and applications. *Annals de L'Institute Henri Poincare. Probability and Statistics* **38** 779-797.

9. Eichelsbacher P. and Lowe M. (2003) Moderate deviations for i.i.d. random variables. *ESAIM: Probability and Statistics* **7** 207-216.

10. Ermakov,M.S. (1995) Asymptotic minimaxity of tests of Kolmogorov and omega-square type. *Theory Probab. Appl.* **40** 54-67.

11. Ermakov,M.S. (1998) On lower bounds of moderate deviation probabilities of tests and estimators. *J.Math.Sci.* **88** 36-42.

12. Ermakov, M.S. (1999) On lower bound for sharp asymptotic of large deviations of statistical estimates. Probl.Inf.Transm. 35(3) 46-60.

13. Fuc D.H. and Nagaev S.V. (1971) Probability inequalities for sums of independent random variables. *Theory Probab.Appl* **16** 660-675.

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

15. Hall,P. (1992) *The Bootstrap and Edgeworth Expansion.* Springer Verlag, New York.

16. Inglot,T.,Kallenberg,W.C.M. and Ledwina,T. (1992) Strong moderate deviation theorems. *Ann. Probab.* **20** 987-1003.

17. Inglot.T. and Ledwina,T. (1990) On probabilities of excessive deviations for Kolmogorov-Smirnov, Cramer-von Mises and chi-square statistics. *Ann. Statist.* **18** 1491-1495.

18. Jureckova,J., Kallenberg,W.C.M. and Veraverbeke,N. (1988) Moderate and Cramer type large deviation theorems for M-estimators. *Statist. Probab. Lett.* **6** 191-199.

19. Mammen,E. (1992) *When Does Bootstrap Work?* Springer Verlag, New York.

20. Sanov,I. (1957) On the probability of large deviations of random variables. *Math. Sbornik*, **42** (In Russian) English translation: Selected Translations in Mathematical Statistics and Probability, 1961, 1 213-244.

21. Serfling R. J. (1980) *Approximation theorems of mathematical statistics*, J. Wiley New York.

22. Stone,M. (1974) Large deviations of empirical probability measures. *Ann. Statist.*, **2** 362-366.