IMPORTANCE SAMPLING FOR LARGE AND MODERATE LARGE DEVIATION SIMULATION OF TESTS AND ESTIMATORS

Michael Ermakov

October 7, 1998

1991 MSC: primary 62E25, secondary 60F10, 65C05

Key words: Importance sampling, Large deviations , Empirical measure, Hadamard-type derivative, Statistical simulation.

Abstract

We find the effective importance sampling procedures for the simulation of large and moderate large deviations of tests and estimators. The computational burden of these effective procedures has no exponential rate as in the direct simulation. The results are applied to the simulation of large and moderate large deviations of L,M,R-statistics and omega-square tests.

Introduction. For the analysis of large and moderate large deviations of tests and estimators the Monte-Carlo technique usually is applied in the more subtle form then the direct simulation of such rare events. The direct simulation requires both a large volume of computations and an investigation of a random number quality since the small fluctuations of random number distributions may cause serious deviations in estimation of small probabilities. There exist convenient approaches to the minimization of computational work and usually one of these approaches is applied in a simulation. The most widespread method for the large deviation simulation is the importance sampling. In the importance sampling procedure the data are generated using a probability distribution different from the true underlying distribution. After that the observed events are weighted to reflect their true relative frequency.

On the base of large deviation theory the problem of the choice of optimal weights in the importance sampling has been solved for a wide range of applications (see Siegmund, 1976; Bucklew, 1990; Bucklew, Ney and Sadowsky, 1990; Sadowsky and Bucklew, 1990; Chen, Lu, Sadowsky and Yao (1993); Sadowsky, 1991, 1996; Lehtonen and Nyhrinen, 1992; Barone, Gigli and Piccioni, 1995). The optimal weights were found using the standard approach of the analysis of large deviations of sums of random variables. The most part of statistical procedures has usually only approximately linear or even nonlinear character and, as a consequence, can not be reduced directly to such an approach. Thus, so far, these results in statistics were applied only for the special models (see Siegmund, 1976; Sadowsky and Bucklew, 1990; Barone, Gigli and Piccioni, 1995). The effective simulation of large and moderate large deviations in this area requires the additional investigation of the problem.

As wellknown the statistical functionals usually can be represented as the functionals of empirical probability measures. Using this fact we develop a similar approach of effective importance sampling based on the theorems about the large and moderate large deviations of empirical measures (see Groeneboom, Oosterhoff and Ruymgaart (GOR), 1979; Ermakov, 1995). The results on efficient simulation of large deviations are obtained in an evident form expressed in terms of Kullback–Leibler information numbers and admits the clear interpretation: the effective importance sampling measures are the solutions of extremal problem of minimization of Kullback-Leibler information numbers on the set of large deviations. Although the straightforward calculations of Kullback-Leibler information numbers and the corresponding probability measures for the efficient simulation represent essential difficulties we can make use these results for the obtaining approximate solutions.

The difficulties arising in the efficient simulation of large deviations were the main reason to investigate a similar problem in the moderate large deviation setting. The domains of moderate large deviations of statistics usually admit the approximations by half-spaces or convex sets in the space of all probability measures. As a consequence, in practice, the testing assumptions of theorems about the efficient moderate large deviation simulation does not represent such serious difficulties as for the large ones. Naturally, the moderate large deviation simulation has also an independent interest for the applications. The results, in this problem, are expressed in terms of the Hellinger metric and the functional admitting the interpretation as the Fisher information. The densities of measures for the efficient simulation of the most of widespread statistics, in particular, L,M and R statistics, are given in a direct form based on their influence functions.

2. Importance sampling for large deviations. Let $\Im$ be the σ-field of Borel sets in Hausdorff space S, Λ the space of all probability measures (pms) in $(S, \Im)$ and $X_1, \ldots, X_n$ i.i.d.r.v.'s with pm $P \in \Lambda$. Denote $\hat{P}_n$ the empirical measure of $X_1, \ldots, X_n$. For any $P, Q \in \Lambda$ define the Kullback-Leibler information number

$$K(Q, P) = \int_S q \log q \; dP, \quad q = \frac{dQ}{dP},$$

if Q absolutely continuous w.r.t. P and $K(Q, P) = \infty$ otherwise. Denote $K(\Omega, P) = \inf\{K(Q, P) : Q \in \Omega\}$ for any $P \in \Lambda$ and $\Omega \subset \Lambda$.

Introduce on the space Λ the τ-topology of weak convergence. In τ-topology a sequence of pms $Q_n \in \Lambda$ converges to pm $Q \in \Lambda$ iff

$$\lim_{n \to \infty} \int_S f \, dQ_n = \int_S f \, dQ$$

for each bounded measurable function $f : S \to R^1$. In what follows all topological properties in Λ (convergence, closeness, compactness and so on) will be considered w.r.t. τ-topology. The closure and the interior of a set $\Omega \subset \Lambda$ in the τ -topology will be denoted by cl (Ω) and int (Ω) respectively.

Let $T : \Lambda \to R^1$ be a fixed functional. For any $b > 0$ denote $\Omega(b) = \{Q : T(Q) > b, \, Q \in \Lambda\}$.

Our arguments are based on the following theorems (see Lemma 2.3, Theorems 3.1 and 3.2 in GOR, 1979).

Theorem 2.1. *Let $P \in \Lambda$ and let $\Omega \subset \Lambda$. Suppose that $K(\mathrm{cl}(\Omega),\ P) = K(\mathrm{int}(\Omega), P)$. Then*

$$\lim_{n \to \infty} n^{-1} \log P(\hat{P}_n \in \Omega) = -K(\Omega, P). \tag{2.1}$$

There exists a pm $Q \in$ cl (Ω) such that Q is absolutely continuous w.r.t. P and $K(Q, P) = K(\Omega, P)$.

Theorem 2.2. *Let the functional $T\colon \Lambda \to R^1$ be continuous in τ-topology and let $T(P) \neq b$. Then*

$$\lim_{n \to \infty} n^{-1} \log P(T(\hat{P}_n) > b) = -K(\Omega(b), P). \tag{2.2}$$

There exists a pm $Q_b \in \Omega(b)$ such that $K(\Omega(b), P) = K(Q_b, P)$.

As wellknown (see Bucklew, 1990; Hammersley and Handscomb, 1964; Sadowsky and Bucklew, 1990), the importance sampling approach allows to define easily the trivial procedure for the explicit calulation of the true probability $P(T(\hat{P}_n) > b)$. However this procedure can not be applied directly since its application requires the explicit knowledge of this probability $P(T(\hat{P}_n) > b)$. Moreover, in this procedure the simulated random variables $Y_1, \ldots, Y_n$ are usually essentially dependent. In the paper we shall be considering essentially more narrow class of procedures. We shall find the effective importance sampling procedures in the class of all procedures simulating independent random variables $Y_1, \ldots, Y_n$.

Let pms $Q_{n1}, \ldots, Q_{nl} \in \Lambda$ be absolutely continuous w.r.t. pm P and let $p_1, \ldots, p_l$ be nonnegative real numbers such that $p_1 + \ldots + p_m = 1$. Let us, by simulation procedure, we get t-independent samples $Y_1^{(i)}, \ldots, Y_n^{(i)}$, $1 \le i \le t$ of i.i.d.r.v.'s with pm $Q_{n\nu_i} \in \Lambda$ where ν_i is a random index, $P(\nu_i = j) = p_j$, $1 \le j \le l$. Denote $\hat{Q}_n^{(i)}$ the empirical measure of $Y_1^{(i)}, \ldots, Y_n^{(i)}$, $1 \le i \le l$. We shall study the importance sampling estimators of $P(T(\hat{P}_n) > b)$ which is defined as follows

$$\hat{V}_{nt} = t^{-1} \sum_{i=1}^{t} \chi(T(\hat{Q}_n^{(i)}) > b)\, w_{ni}^{-1}. \tag{2.3}$$

where

$$w_{ni} = \sum_{j=1}^{m} p_j u_{nij}, \quad u_{nij} = \prod_{s=1}^{n} q_{nj}(Y_s^{(i)})$$

with $q_{nj} = dQ_{nj}/dP$, $1 \le j \le l$. The using importance sampling simulation based on the mixtures of pms Q_{nj}, $1 \le j \le m$, and the corresponding additional randomization by the random index ν allow to define the effective procedures for the essentially more wide class of statistical problems.

By straightforward calculations we get

$$E_Q[\hat{V}_{nt}] = E_Q[\hat{V}_{n1}] = P(T(\hat{P}_n) > b) \tag{2.4}$$

and

$$Var_Q[\hat{V}_{nt}] = t^{-1}(E_Q[\hat{U}_n] - (E_Q[\hat{V}_{n1}])^2) \tag{2.5}$$

where $\hat{U}_n = \chi(T(\hat{Q}_n^{(1)}) > b)\, w_{n1}^{-2}$.

By Theorem 2.2 and (2.4), we have

$$E_Q[\hat{V}_{n1}] = \exp\{-nK(\Omega(b), P)(1 + o(1))\} \tag{2.6}$$

as $n \to \infty$. Here Q denotes the probability measure of simulation.

Therefore we get the following assertion.

Lemma 2.1. *Let the functional $T : \Lambda \to R^1$ be continuous in τ-topology and let $T(P) \neq a$. Then for any sequences $Q_{n1}, \ldots, Q_{nl} \in \Lambda$ and nonnegative numbers $p_1, \ldots, p_l$, $p_1 + \ldots + p_l = 1$,*

$$\liminf_{n \to \infty} n^{-1} \log E\,[\hat{U}_n] \ge -2K(\Omega(b), P). \tag{2.7}$$

Lemma 2.1 allows to introduce naturally the notion of asymptotic efficiency of importance sampling procedures. We say that the importance sampling procedure $\hat{V}_{nt}$ is asymptotically efficient if

$$\lim_{n\to\infty} n^{-1} \log E\,[\hat{U}_n] = -2K(\Omega(b), P). \tag{2.8}$$

A similar notion of asymptotic efficiency in the other terms has been introduced in Bucklew (1990) and Bucklew and Sadowsky (1990). Roughly speaking, an importance sampling procedure is asymptotically efficient if the computational burden grows less than exponentially fast.

For the problem of large deviation estimation of $P(\hat{P}_n \in \Omega)$ with a given set $\Omega \subset \Lambda$ the importance sampling procedure is defined similarly

$$\hat{V}_{nt} = t^{-1} \sum_{i=1}^{t} \chi(\hat{Q}_n^{(i)} \in \Omega)\, w_{ni}^{-1}. \tag{2.9}$$

Here the analog of Lemma 2.1 holds also in the notation $\hat{U}_n = \chi(\hat{Q}_n^{(1)} \in \Omega) w_{n1}^{-2}$ and $\Omega(b) = \Omega$. Therefore the same definition of asymptotic efficiency can be introduced for this problem as well. Naturally the importance sampling procedures for the estimation of large deviation probabilities $P(T(\hat{P}_n) > b)$ can be considered as a particular case of the more general procedure (2.9). It suffices to put only $\Omega = \Omega(b)$.

Let the pm $R \in \Lambda$ be absolutely continuous w.r.t. pm $P \in \Lambda$ and let $r = dR/dP$. Denote $\Gamma_R = \{Q : \int_S \log r\, dQ > K(\Omega, P), \quad Q \in \Lambda\}$.

Theorem 2.3. *Let $\Omega \subset \Lambda$ and let $P \in \Lambda$. Let $K(\text{ cl }(\Omega), P) = K(\text{ int }(\Omega), P)$. Suppose there exists only a finite number of pms $R_1, \ldots, R_m$ such that $R_i \in$ cl (Ω) and $K(R_i, P) = K(\Omega, P)$ for all $1 \le i \le m$. Denote $r_i = dR_i/dP$ and suppose $E[r_i^{-1}(X_1)] < \infty$ for all $1 \le i \le m$. Suppose also $\Omega \subset \cup_{i=1}^{m} \Gamma_{R_i}$. Then the importance sampling procedures* (2.9) *are asymptotically efficient with given pms $Q_{n1} = Q_1, \ldots, Q_{nl} = Q_l$ iff the set of pms $Q_1, \ldots, Q_l$ contains the set of pms $R_1, \ldots, R_m$ and $p_i \neq 0$ for all i such that $Q_i = R_i$.*

Let the functional $T{:}\Lambda \to R^1$ be continuous in τ-topology. Then $K(\text{ cl }(\Omega(b)),\ P) = K(\text{int }(\Omega(b)),\ P)$. Therefore, if the set $\Omega = \Omega(b)$ satisfies all the other assumptions of the theorem, then the same statement holds also for the importance sampling procedure (2.3).

The proof of Theorem 2.3 will be omitted. Similar arguments are given below in the proof of Theorem 3.3 about the effective importance sampling simulation of moderate large deviations. The proof of these theorems unites the technique for the analysis of large deviations of empirical measures (see GOR, 1979; Ermakov, 1993,1995) with the reasonings utilized in the proof of the efficiency of importance sampling procedures (see Sadowsky and Bucklew, 1990; Sadowsky, 1996 and references in these papers).

3. Importance sampling for the moderate large deviations. Let $T : \Lambda \to R^1$ be a functional continuous in τ-topology and let P be a limit point of a sequence of sets $\Omega_n \subset \Lambda$. Let $T(P) = 0$ and $b_n \to 0$, $nb_n^2 \to \infty$ as $n \to \infty$. In section we find the effective importance sampling procedures for the estimation of moderate large deviation probabilities $P(T(\hat{P}_n) > b_n)$ and $P(\hat{P}_n \in \Omega_n)$.

For any $P, Q \in \Lambda$ define the Hellinger distance

$$\rho(Q, P) = \left(\int_S \left(\left(\frac{dQ}{dR}\right)^{1/2} - \left(\frac{dP}{dR}\right)^{1/2} \right)^2 dR \right)^{1/2}, \quad R = \frac{1}{2}(P + Q).$$

For any $\Omega \subset \Lambda$ denote $\rho(\Omega,\ P) = \inf\{\rho(Q, P) : Q \in \Omega\}$.

Introduce the space Λ_0 of all charges G on $(S, \Im)$ having the bounded variation and such that $G(S) = 0$. Define the τ-topology in Λ_0 similarly to that on Λ. All topological properties in Λ_0 will be considered w.r.t. τ-topology.

For any $G \in \Lambda_0$ and $P \in \Lambda$ define the functional

$$\rho_0(G :\ P) = \left(\int_S g^2 dP \right)^{1/2}, \quad g = \frac{dG}{dP} \tag{3.1}$$

if G is absolutely continuous w.r.t. P and $\rho_0(G :\ P) = \infty$ otherwise. For any $\Omega_0 \subset \Lambda_0$ denote $\rho_0(\Omega_0 :\ P) = \inf\{\rho_0(G :\ P), G \in \Omega_0\}$. From the viewpoint of statistical applications, the functional ρ_0 can be considered as a natural analog of the Fisher information.

Let P be a limit point of sets Ω_n.

Make the following Assumptions.

A. There exist an open set $\Omega_0 \subset \Lambda_0$ and a function $\omega(t)$, $\omega(t)/t \to 0$ as $t \to 0$ such that

1. for any sequence of charges $G_n \in \Omega_0$ there exists a sequence of pms $Q_n \in \Omega_n$ such that $\rho(Q_n, P + b_n G_n) < \omega(\rho(P, P + b_n G_n))$.

2. for any sequence of pms $Q_n \in \Omega_n$ there exists a sequence of charges $G_n \in \Omega_0$ such that $\rho(Q_n, P + b_n G_n) < \omega(\rho(Q_n, P))$.

Thus the sets $P + b_n\Omega_0$ can be interpreted as the "linear approximations" of the sets Ω_n in the Hellinger metric. It is easily seen that $\rho(\Omega_n, P) = \frac{1}{2} b_n \rho_0(\Omega_0 :\ P)(1 + o(1))$ as $n \to \infty$.

B. There exists a homogeneous functional $T_0 :\ \Lambda_0 \to R^1$ having the order one of homogeneity and a function $\omega :\ \omega(t)/t \to 0$ as $t \to 0$ such that the functional T_0 is continuous in τ-topology and

$$|T(Q) - T(P) - T_0(Q - P)| < \omega(T_0(Q - P)) \tag{3.2}$$

for any $Q \in \Lambda$.

Let B hold. Then, denote $\Omega_0 = \{G :\ T_0(G) > 1, G \in \Lambda_0\}$.

Theorems 3.1 and 3.2 below follow easily from Theorem 3.2 in Borovkov and Mogulskii (1980) and Theorem 3.1 in Ermakov (1995).

Theorem 3.1. *Assume* A. *Let* $b_n \to 0$, $nb_n^2 \to \infty$ *as* $n \to \infty$. *Then*

$$\lim_{n\to\infty} (2n\rho^2(\Omega_n, P))^{-1} \log P(\hat{P}_n \in \Omega_n) = -1. \tag{3.3}$$

There exists a charge $G \in \mathrm{cl}\ (\Omega_0)$ *such that* $\rho(\Omega_n, P) = \rho(P + b_n G, P)(1 + o(1)) = \frac{1}{2} b_n \rho_0(\Omega_0 : P)(1 + o(1)) = \frac{1}{2} b_n \rho_0(G : P)(1 + o(1))$ *as* $n \to \infty$.

Theorem 3.2. *Assume* B. *Let* $b_n \to 0$, $nb_n^2 \to \infty$ *as* $n \to \infty$. *Then*

$$\lim_{n\to\infty} \left(\frac{1}{2} n b_n^2 \rho_0^2(\Omega_0, P)\right)^{-1} \log P(T(\hat{P}_n) > b_n) = -1. \tag{3.4}$$

There exists a charge $G \in \mathrm{cl}\ (\Omega_0)$ *such that* $\rho(\Omega(b_n), P) = \rho(P + b_n G, P)(1 + o(1)) = \frac{1}{2} b_n \rho_0(\Omega_0 : P)(1 + o(1)) = \frac{1}{2} b_n \rho_0(G : P)(1 + o(1))$ *as* $n \to \infty$.

The importance sampling procedures are defined similarly to that in section 2. It suffices only to replace b by b_n in the definition (2.3) and Ω by Ω_n in the corresponding definition (2.9).

Denote $\Omega_n = \Omega(b_n)$ if B holds. Lemma 3.1 below represents a direct analog of Lemma 2.1.

Lemma 3.1. *Assume* A *or* B. *Let* $b_n \to 0$, $nb_n^2 \to \infty$ *as* $n \to \infty$. *Then for any sequence of importance sampling procedures*

$$\liminf_{n\to\infty} (2n\rho^2(\Omega_n, P))^{-1} \log E[\hat{U}_n] \geq -2 \tag{3.5}$$

or, in the other terms,

$$\liminf_{n\to\infty} (nb_n^2)^{-1} \log E[\hat{U}_n] \geq -\rho_0^2(\Omega_0 : P). \tag{3.6}$$

Lemma 3.1 follows immediately from Theorems 3.1 and 3.2.

We say that sequences of pms $Q_{n1}, \ldots, Q_{nm}$ and nonnegative real numbers $p_1, \ldots, p_m$, $p_1 + \ldots + p_m = 1$, generate asymptotically effective procedures of importance sampling if the equality is attained in (3.5) and (3.6).

Make the following additional assumptions.

C1. There exists only a finite number of charges $G_1, \ldots, G_m \in \mathrm{cl}\ (\Omega_0)$ such that $\rho_0(G_j : P) = \rho_0(\Omega_0 : P)$, $1 \leq j \leq m$.

Denote $g_j = dG_j/dP$, $1 \leq j \leq m$. Define the sets $\Psi_j = \Psi_{G_j} = \{G : \rho_0(G + G_j : P) < 2\rho_0(\Omega_0 : P),\ G \in \Lambda_0\}$ for all $1 \leq j \leq m$. The set Ψ_j is the set of all charges G with the densities from the ball in $L_2(P)$ having the center $-g_j$ and the radius $2\rho_0(\Omega_0 : P)$. We put $\Psi = \cap_{j=1}^m \Psi_j$.

C2. $\Omega_0 \cap \Psi = \emptyset$.

For each j, $1 \leq j \leq m$, define the set $\bar{\Gamma}_{G_j} = \{H : \int_S g_j dH < \rho_0^2(\Omega_0 : P),\ H \in \Lambda_0\}$. It is clear that $\bar{\Gamma}_{G_j} \supset \Psi_j$. Thus C2 can be replaced by the stronger assumption.

C3. $\Omega_0 \subset \Lambda_0 \setminus \mathrm{cl}\ (\cap_{j=1}^m \bar{\Gamma}_{G_j})$.

The assumption of a type C3 was made in Theorem 2.3 and is a traditional in the problem of large deviation simulation (see Sadowsy and Bucklew, 1990). This assumption does not fulfilled for the sets Ω_0 generated by nonlinear statistical functionals, in particular omega-square test statistics. The assumption C2, in some extent, allows to avoid this difficulty.

Theorem 3.3. *Assume* A, C1 *and* C2 *or* B, C1 *and* C2. *Let* $b_n \to 0$, $nb_n^2 \to \infty$ *as* $n \to \infty$. *Then the sequence of pms* Q_{nj}, $1 \le j \le l$, *having the densities*

$$q_{nj} = \lambda_{nj} + b_n h_j \chi(h_j > -c_n b_n^{-1}) \tag{3.7}$$

with $c_n b_n^{-1} \to \infty$ *as* $n \to \infty$, $0 < c_n < c < 1$ *and* $\lambda_{nj} \to 1$ *as* $n \to \infty$ *generates asymptotically efficient importance sampling procedures* $\hat{V}_{nt}$ *(see* (2.3),(2.9)*) iff the set of all* h_j, $1 \le j \le l$, *contains the set of all densities* g_i, $1 \le i \le m$ *and* $p_j \ne 0$ *for all* j *such that* $h_j = g_i$.

The same statement is also valid for the sequences of pms $Q_{nj}^{(1)}$ *having the densities*

$$q_{nj}^{(1)} = c(b_n) \exp\{b_n h_j\} \chi(h_j > -c_n b_n^{-1}) \tag{3.8}$$

with $c_n b_n^{-1} \to \infty$ *as* $n \to \infty$, $0 < c_n < c < 1$. *Here* $c(b_n)$ *is a normalizing constant.*

The proof of Theorem 3.3 will be given in section 5.

Remark 3.1. If the functional T is nonlinear, it can turn out that C2 does not hold. In this case the following modification of the procedure can be usefull. Suppose there exists a finite number of charges $G_{m+1}, \ldots, G_{m+a} \in \text{cl}\ (\Omega_0)$ such that $\cap_{j=1}^{m+a} \Psi_{G_j} \cap \Omega_0 = \emptyset$. Consider the importance sampling procedure for the pm Q_{nj} having the densities q_{nj} or $q_{nj}^{(1)}$ with $h_j = g_j = \frac{dG_j}{dP}$, $1 \le j \le m+a$. Then the analysis of the proof of Theorem 3.3 shows that such a procedure is asymptotically efficient. The corresponding likelihood ratios u_{nij}, $j > m$, can have the same or the larger order than the likelihood ratios u_{nij}, $j \le m$, with the essentially smaller probability. Thus u_{nij}, $j > m$, can be considered as the regularization addendums.

Example. Let $T(\hat{P}_n)$ be the test statistic of omega-square type, with the functional

$$T(Q) = \int_0^1 (F(x) - x)^2 r(x) dx$$

where $F(x)$ stands for the distribution function of Q, $S = [0,1]$ and r a weight function continuous in $[0,1]$. Naturally we suppose that P is the uniform distribution in $[0,1]$.

The set Ω_0 equals

$$\Omega_0 = \left\{ G : \int_0^1 H^2(x) r(x)\ dx \ge 1,\ H(x) = G((0,x]),\ G \in \Lambda_0 \right\}.$$

The charges G satisfying $\rho_0(G : P) = \rho_0(\Omega_0 : P)$, $G \in \text{cl}\ \Omega_0$ are set by the equation (see Anderson and Darling, 1952)

$$H'' + \lambda_1 r H = 0, \qquad H(0) = H(1) = 0 \tag{3.9}$$

where $H(x) = G(0,x)$, $x \in (0,1)$ and λ_1 is the largest eigenvalue of (3.9).

Let $\lambda_1 > \lambda_2 > \dots$ be the eigenvalues of (3.9) and let $k = \max\{i : 4\lambda_i \geq \lambda_1\}$. Suppose that for each λ_j, $1 \leq j \leq k$, there exists the unique eigenfunction H_j. Then the charges G_j, $G_j((0,x)) = H_j(x)$ and $G_{k+j}((0,x)) = -H_j(x)$, $1 \leq j \leq k$, satisfy the assumptions of Theorem 3.3. The charges $G_2, \dots, G_k, G_{k+2}, \dots, G_{2k}$, here, play the same part as in Remark 3.1.

4. Importance sampling for the moderate large deviations. The functionals admitting the linear approximation. In section we shall be assuming that the functional $T : \Lambda \to R^1$ is approximately linear, that is, satisfies the following assumptions.

D1. There exist a function $g : S \to R^1$ and a function ω, $\omega(t)/t \to 0$ as $t \to 0$, such that for any $Q \in \Lambda$

$$\left| T(Q) - T(P) - \int_S g \, d(Q-P) \right| < \omega(N(Q-P)). \tag{4.1}$$

Here $N : \Lambda_0 \to R^1$ stands for a norm in Λ_0 continuous in τ- topology.

By Theorem 3.2, if the norm N is continuous in τ- topology, then for any sequence d_n, $d_n \to 0$, $nd_n^2 \to \infty$ as $n \to \infty$ it holds

$$P(N(\hat{P}_n - P) > d_n) \leq \exp\{-cnd_n^2(1 + o(1))\}. \tag{4.2}$$

D2. There exists $c > 0$ such that $E[\exp\{c\, g(X_1)\}] < \infty$.

D1 can be considered as a version of the condition of Hadamard differentiability of functional T. Such a type of assumptions usually is utilized for the proof of asymptotic normality of L, M and R statistics (see Serfling, 1980; Denker, 1985) and, in implicit form, the same technique was applied also for the study of their large deviations (see Jureckowa, Kallenberg and Veraverbeke, 1988; Inglot, Kallenberg and Ledwina,1992; Ermakov, 1994). Note that the τ - continuiuty of the norm N can be replaced by the weaker assumption (4.2) (see Inglot, Kallenberg and Ledwina, 1992). Thus D1 and D2 allow to investigate the problem of the moderate large deviation simulation for the statistical functionals having the Hadamard derivative, in particular, L, M and R statistics.

For any function $h \in L_2(P)$ denote $\sigma_h^2 = E[h^2(X_1)]$. We put $\sigma^2 = \sigma_g^2$.

Lemma 4.1. *Assume* D1 *and* D2. *Let $b_n \to 0, nb_n^2 \to \infty$ as $n \to \infty$. Then for any sequence of importance sampling procedures*

$$\liminf_{n\to\infty} (nb_n^2)^{-1} \log E[\hat{U}_n] \geq -\sigma^{-2}. \tag{4.3}$$

Therefore, if D1,D2 hold, one can get the lower bound for the asymptotic efficiency of importance sampling procedures in the more evident form.

Theorem 4.1. *Assume* D1 *and* D2. *Let $b_n \to 0$, $nb_n^2 \to \infty$ as $n \to \infty$ and let $h \in L_2(P)$. Then the sequences of pms Q_n having the densities*

$$q_n = \lambda_{nj} + b_n \, h \, \chi(h > -c_n b_n^{-1}) \tag{4.4}$$

and pms $Q_n^{(1)}$ having the densities

$$q_n^{(1)} = c(b_n)\exp\{b_n h\}\chi(h > -c_n b_n^{-1}) \tag{4.5}$$

with $c_n b_n^{-1} \to \infty$ as $n \to \infty$, $0 < c_n < c < 1$ generate asymptotically efficient importance sampling procedures $\hat{V}_{nt}$ iff $h = \sigma^{-2} g$.

Proof of Lemma 4.1. Without loss of generality we shall assume that $\omega(t)$ is strictly monotone function. Define the inverse function ϕ for the function ω such that $\phi(s) = t$ implies $\omega(t) = s$. By the Cramer Theorem (see Saulis and Statulevichius, 1989) and (4.2), we have

$$P(T(\hat{P}_n) - T(P) > b_n) \geq P\left(\sum_{s=1}^{n} g(X_s) > nb_n(1-\delta_n)\right) - P(N(\hat{P}_n - P) > \phi(b_n\delta_n)) \geq$$

$$\exp\left\{-\frac{1}{2\sigma^2} nb_n^2(1-\delta_n)^2(1+o(1))\right\} - \exp\{-cn\phi^2(b_n\delta_n)\} \tag{4.6}$$

for any sequence $\delta_n > 0$, $\delta_n \to 0$ as $n \to \infty$ and, in particular, a sequence δ_n such that $\phi(b_n\delta_n)/b_n \to \infty$ as $n \to \infty$. Therefore (4.6) implies (4.3).

The proof of Theorem 4.1 makes use the asymptotic of moderate large deviation probabilities of empirical measures for the more general setting than in Theorems 3.1 and 3.2. This proof is based on the asymptotic of $\log P_n(\hat{P}_n \in \Omega_n)$ with a sequence of pms P_n converging to P. Such an asymptotic was obtained in Ermakov (1995), Theorem 3.1.

Make the following Assumption.

E. There exists a sequence of charges $H_n \in \Lambda_0$ such that H_n are absolutely continuous w.r.t. P, $\rho(P_n, P + b_n H_n) = o(\rho(P_n, P))$ as $n \to \infty$ and

$$\lim_{n\to\infty} \int_S \left(\frac{dH_n}{dP}\right)^2 \chi\left(\left|\frac{dH_n}{dP}\right| > C_n\right) dP = 0 \tag{4.7}$$

for any sequence $C_n \to \infty$ as $n \to \infty$.

Theorem 4.2. *Let P_n converge to P in the τ - topology, let $b_n \to 0$, $nb_n^2 \to \infty$ as $n \to \infty$ and let* A *and* E *hold . Then*

$$\lim_{n\to\infty} (2n\rho^2(\Omega_n, P_n))^{-1} \log P_n(\hat{P}_n \in \Omega_n) = -1. \tag{4.8}$$

Proof of Theorem 4.1. The reasoning will be given for the sequences of pms Q_n. The case of pms $Q_n^{(1)}$ is similar.

Define the sequence of pms R_n having the densities

$$r_n(x) = \frac{dR_n}{dP}(x) = c^{-1}(b_n)(\lambda_n + b_n h(x))^{-1}\chi(h(x) > -c_n b_n^{-1}\sigma_h^{-1}).$$

Here

$$c(b_n) = E[(\lambda_n + b_n h(X_1))^{-1}\chi(h(X_1) > -c_n b_n^{-1}\sigma_h^{-1})] = 1 + b_n^2\sigma_h^2(1+o(1)). \tag{4.9}$$

We have

$$E[\hat{U}_n] = c^n(b_n) P_{R_n}(T(\hat{P}_n) - T(P) > b_n) \le I_{n1} + I_{n2} \tag{4.10}$$

where

$$I_{n1} = c^n(b_n) P_{R_n}\left(\int_S g\, d(\hat{P}_n - R_n) + T(R_n) - T(P) > (1-\delta_n) b_n\right),$$

$$I_{n2} = c^n(b_n) P_{R_n}\left(T(\hat{P}_n) - T(R_n) - \int_S g\, d(\hat{P}_n - R_n) \ge b_n \delta_n\right)$$

with $b_n\delta_n/\omega(b_n) \to \infty$ as $n \to \infty$. Hereafter $P_{R_n}(A)$ denotes the probability of event A with respect to pm R_n.

Since N is continuous in the τ- topology and the τ- topology is weaker then the topology of convergence on variation then we have (see Ermakov, 1995)

$$N(R_n - P) < C \int_S |r_n - 1|\, dP < C b_n.$$

Hence

$$|T(R_n) - T(P) - \int_S g\, d(R_n - P)| < \omega(N(R_n - P)) < \omega(C b_n) \tag{4.11}$$

By D2 and Theorem 4.2, we have

$$I_{n2} = c^n(b_n) P_{R_n}\left(T(\hat{P}_n) - T(R_n) - \int_S g\, d(\hat{P}_n - R_n) > b_n\delta_n\right) <$$

$$c^n(b_n) P_{R_n}(N(\hat{P}_n - R_n) > \phi(b_n\delta_n)) \le \exp\{n b_n^2 \sigma_h^2 (1 + o(1)) - c_1 n \phi^2(b_n \delta_n)\}. \tag{4.12}$$

Since $0 < C_1 < r_n(x) < C_2 < \infty$, the assumptions of Theorem 3.2 in Saulis and Statulevichius (1989) fulfilled, and, using (4.11), we have

$$I_{n1} \le c^n(b_n) P_{R_n}\left(\int_S g\, d(\hat{P}_n - R_n) > (1 - C\delta_n) b_n - \int_S g\, d(R_n - P)\right) =$$

$$c^n(b_n) P_{R_n}\left(\int_S g\, d(\hat{P}_n - R_n) > (1 - C\delta_n) b_n + b_n \int_S gh\, dP + o(b_n)\right) =$$

$$\exp\left\{n b_n^2 \sigma_h^2 - \frac{1}{2} n b_n^2 \sigma^{-2} \left(1 - C\delta_n + \int_S gh\, dP\right)^2 + o(n b_n^2)\right\}. \tag{4.13}$$

It is easy to see that the infimum of the right-hand side of (4.13) is attained if $h = \sigma^{-2} g$. Hence

$$I_{n1} < \exp\{-n b_n^2 \sigma^{-2} (1 + o(1))\}.$$

This completes the proof of Theorem 4.1.

5. Proof of Theorem 3.3.. The reasoning is based on a standard technique for the analysis of large deviations of empirical measures (see GOR, 1979) and its modification on the case of moderate large deviations (see Ermakov 1993,1995). Naturally, the preceding ideas (see Sadowsky and Buklew (1990), Buklew, Ney and Sadowsky (1990)) developed for effective simulation by importance sampling procedure play the essential part as well.

We begin with the proof of sufficiency of the theorem statement.

Denote $\Pi = \Pi_k = \{S_i\}_1^k$ a partition of S consisting of a finite number of Borel sets S_i, $1 \le i \le k$.

For any $Q \in \Lambda$, $G \in \Lambda_0$ and a partition $\Pi = \{S_i\}_1^k$ of S denote

$$\rho^2(Q, P|\Pi) = \sum_{i=1}^{k} (P^{1/2}(S_i) - Q^{1/2}(S_i))^2,$$

$$\rho_0(G, P|\Pi) = \sum_{i=1}^{k} \frac{G^2(S_i)}{P(S_i)}.$$

Here we suppose that $P(S_i) \neq 0$ for all $1 \le i \le k$.

It is known that (see, for example, Borovkov and Mogulskii, 1980; Ermakov, 1993,1995)

$$\rho(Q, P) = \sup_{\Pi} \rho(Q, P|\,\Pi), \quad \rho_0(G, P) = \sup_{\Pi} \rho_0(G, P|\,\Pi) \tag{5.1}$$

where the supremum is taken over all partitions Π of S.

For any $\delta > 0$, $C_1 > 0$ define a partition $\Pi = \Pi_{C_1,\delta} = \{S_i\}_1^k$ such that $p_i = P(S_i) > 0$ for all $1 \le i \le k$,

$$\min_{y \in S_k} \max_{1 \le j \le m} |g_j(y)| > C_1$$

and for each $1 \le j \le m$, $1 \le i \le k-1$ for all $y \in S_i$

$$\max_{x \in S_i} g_j(x) \le g_j(y) \le \min_{x \in S_i} g_j(x) + \delta$$

Lemma 5.1. *Let the assumptions of Theorem 3.3 be satisfied. Then*

$$\lim_{\substack{\delta \to 0 \\ C_1 \to \infty}} \min_{1 \le j \le k} (2\rho(\Omega_n, P))^{-1} \rho(\Omega_n : P - b_n G_j | \Pi_{C_1 \delta}) =$$

$$\lim_{\substack{\delta \to 0 \\ C_1 \to \infty}} \min_{1 \le j \le k} (2\rho_0(\Omega_0, P))^{-1} \rho_0(\Omega_0 + G_j : P | \Pi_{C_1 \delta}) = 1.$$

The proof of Lemma 5.1 follows the line of that of Lemma 2.4 in GOR (1979) (see also Ermakov (1995)) and will be omitted.

For all $1 \le j \le m$ denote $\xi_{nj} = E[g_j(X)\chi(g_j(X) < -c_n b_n^{-1})]$. Since $c_n b_n^{-1} \to \infty$ as $n \to \infty$, ee get $\xi_{nj} = o(1)$ as $n \to \infty$. Therefore the events $\cup_{s=1}^n \cup_{j=1}^m \{X_s : g_j(X_s) < -c_n b_n^{-1}\}$ will influence in the proof only on the remainder terms of estimates.

For all i, $1 \le i \le k$, and j, $1 \le j \le m$ denote $q_{nji} = Q_{nj}(S_i)$, $g_{ji} = G_j(S_i)$. We put $\alpha_{nj} = 1 + b_n \int_S g_j \chi(g_j > -c_n b_n^{-1})\, dP$. It is clear that $\alpha_{nj} = 1 - b_n \xi_{nj} = 1 + o(b_n)$.

In what follows, in order to simplify the estimates, we consider the case $m = 1$. The case of arbitrary m will be later reduced to this one. For all i, $1 \le i \le k$, we put $r_i = \int_{S_i} g_1^2 dP$, $d_{ni} = \int_{S_i} q_{n1}^{-1} dP$. Denote $\gamma_n^2 = \frac{1}{4} b_n^2 \sum_{i=1}^k g_{1i}^2 / p_i$.

Expanding in the Taylor series, we get

$$q_{nji} = \alpha_{nj} p_i + b_n g_{1i} + o(b_n) = p_i + b_n g_{1i} + o(b_n), \tag{5.2}$$

$$d_{ni} = \alpha_{n1} p_i - b_n g_{1i} + b_n^2 r_i (1 + o(1)), \quad 1 \le i \le k. \tag{5.3}$$

By Lemma 5.1, there exists $\beta = \beta(\delta, C_1)$, $\beta(\delta, C_1) \to 1$ as $\delta \to 0$, $C_1 \to \infty$ such that

$$E[\hat{U}_n] = E[w_{n1}^{-1} \chi(\hat{P}_n \in \Omega_n)] \le$$
$$E[w_{n1}^{-1} \chi(\rho^2(\hat{P}_n, P - b_n G_1 | \Pi_{\delta, C_1}) > 4\beta\gamma_n^2) \doteq \bar{I}_n(\Pi_{\delta, C_1}). \tag{5.4}$$

Applying the Stirling formula, we get

$$\bar{I}_n(\Pi_{\delta, C_1}) < \sum{}' \frac{n!}{(nz_{n1})! \dots (nz_{nk})!} \prod_{i=1}^k d_{ni}^{nz_{ni}} \le$$
$$C \sum{}' \exp\left\{ \left(\frac{1}{2} - \frac{k}{2} \right) \log n - \frac{1}{2} \sum_{i=1}^k \log z_{ni} - \right.$$
$$\left. n \sum_{i=1}^k z_{ni} \log \frac{z_{ni}}{p_i} + n \sum_{i=1}^k z_{ni} \log \frac{d_{ni}}{p_i} \right\} \doteq I(Z_{n1}). \tag{5.5}$$

Here the summation $\sum'$ is taken over the set Z_{n1} of all $z = (z_{n1}, \dots, z_{nk})$ such that $nz_{n1}, \dots, nz_{nk}$ are nonnegative whole numbers, $z_{n1} + \dots + z_{nk} = 1$ and

$$J_{n1}(z) = \frac{1}{4} \sum_{i=1}^k (z_{ni}^{1/2} - (p_i - b_n g_{1i})^{1/2})^2 > \beta\gamma_n^2. \tag{5.6}$$

Introduce also the set Z_{n2} of all $z = (z_{n1}, \dots, z_{nk})$ such that $nz_{n1}, \dots, nz_{nk}$ are nonnegative whole numbers, $z_{n1} + \dots + z_{nk} = 1$ and

$$J_{n2}(z) = \sum_{i=1}^k (z_{ni}^{1/2} - p_i^{1/2})^2 > \beta\gamma_n^2. \tag{5.7}$$

It is clear that $Z_{n1} \subset Z_{n2}$ for all $n > n_0(\delta, C_1)$ and therefore $I(Z_{n1}) \le I(Z_{n2})$.

Fix $\epsilon > 0$, and, for $s = 0, 1, 2, \dots$ define the sets $Z_{ns2} = \{z : (1 + \epsilon s)\gamma_n^2 < J_{n2} \le (1 + \epsilon(s+1))\gamma_n^2, z \in Z_{n2}\}$. The asymptotic of number of elements Z_{ns2} presents the asymptotic of the number of elements between two ellipsoids and has the following expression

$$C\epsilon \left(\Gamma \left(\frac{k}{2} \right) \right)^{-1} (2\pi)^{\zeta - 1} (n\gamma_n)^{2\zeta} (1 + \epsilon s)^{\zeta - 1} \prod_{i=1}^k p_i^{1/2} (1 + o(1)) \tag{5.8}$$

as $n \to \infty$. Here $\zeta = (k-1)/2$ and $\Gamma(\frac{k}{2})$ stands for the value of gamma function at the point $\frac{k}{2}$.

Expanding $z_{ni} \log \frac{z_{ni}}{p_i}$ and $z_{ni} \log \frac{d_{ni}}{p_i}$ in the Taylor series by the powers $(z_{ni}^{1/2} - p_i^{1/2}) p_i^{-1/2}$, we get

$$- \sum_{i=1}^k z_{ni} \log \frac{z_{ni}}{p_i} = -2 \sum_{i=1}^k (z_{ni}^{1/2} - p_i^{1/2})^2 (1 + o(1)), \tag{5.9}$$

$$\sum_{i=1}^k z_{ni}\log\frac{d_{ni}}{p_i} = \sum_{i=1}^k p_i(1+2(z_{ni}^{1/2}-p_i^{1/2})p_i^{-1/2}+(z_{ni}^{1/2}-p_i^{1/2})^2p_i^{-1})\times$$

$$\log\left(1+(\alpha_n-1)-b_n\frac{g_{1i}}{p_i}+b_n^2\frac{r_i}{p_i}(1+o(1))\right) =$$

$$-2b_n\sum_{i=1}^k g_{1i}(z_{ni}^{1/2}-p_i^{1/2})p_i^{-1/2}-\frac{1}{2}b_n^2\sum_{i=1}^k\frac{g_{1i}^2}{p_i}+b_n^2\sum_{i=1}^k\frac{r_i}{p_i}+o(b_n^2) =$$

$$-4\sum_{i=1}^k(z_{ni}^{1/2}-p_i^{1/2})(q_{n1i}^{1/2}-p_i^{1/2})+2\sum_{i=1}^k(q_{n1i}^{1/2}-p_i^{1/2})^2+O(\epsilon b_n^2)+o(b_n^2). \qquad (5.10)$$

Hence, by straightforward calculations, we get

$$-\sum_{i=1}^k z_{ni}\log\frac{z_{ni}}{p_i}+\sum_{i=1}^k z_{ni}\log\frac{d_{ni}}{p_i} =$$

$$-4\sum_{i=1}^k(z_{ni}^{1/2}-p_i^{1/2})^2+2\sum_{i=1}^k(z_{ni}^{1/2}-q_{n1i}^{1/2})^2+O(\epsilon b_n^2)+o(b_n^2) =$$

$$-2\sum_{i=1}^k(z_{ni}^{1/2}+q_{n1i}^{1/2}-2p_i^{1/2})^2+4\sum_{i=1}^k(q_{n1i}^{1/2}-p_i^{1/2})^2+O(\epsilon b_n^2)+o(b_n^2). \qquad (5.11)$$

Now, (5.4) - (5.8),(5.11), C2 together imply

$$I(Z_{n1}) \le \sum_{s=0}^\infty I(Z_{ns2}\cap Z_{n1}) \le C\sum_{s=0}^\infty(1+\epsilon s)^{\zeta-1}\epsilon(n\gamma_n^2)^\zeta\times$$

$$\exp\{-4n(1+2\epsilon s)\gamma_n^2(1+o(1))\}. \qquad (5.12)$$

Choose a sequence $\epsilon=\epsilon_n=o(1)$ such that $n\epsilon_n\gamma_n^2\to\infty$ as $n\to\infty$. Then, (5.12) implies

$$I_n(Z_n) < \exp\{-4n\gamma_n^2(1+o(1))\} \qquad (5.13)$$

as $n\to\infty$.

Suppose that m is arbitrary. Then, we have

$$E[\hat U_n] = \int_S\left(\sum_{j=1}^m p_j u_{nj}\right)^{-1}\chi(\hat P_n\in\Omega_n)\,dP \le$$

$$\sum_{t=1}^m\int_S\left(\sum_{j=1}^m p_j u_{nj}\right)^{-1}\chi(\min_t\rho(\hat P_n,P-b_nG_t|\Pi)\ge 2\beta b_n\rho(\Omega_0+G_t:P)) \le$$

$$\sum_{j=1}^m p_j^{-1}\int_S u_{nj}^{-1}\chi(\rho(\hat P_n,P-b_nG_j|\Pi)\ge 2\beta b_n\rho(\Omega_0+G_j:P)). \qquad (5.14)$$

Therefore the problem was reduced to the case of $m=1$ considered above.

In the proof of necessity we follow to Sadowsky and Bucklew (1990). Suppose $p_1=0$ (if $p_j=0$ for any j, $2\le j\le m$, the arguments are similar). Then for any

κ, $0 < \kappa < \kappa_0$ there exist κ_1, $0 < \kappa < \kappa_1$, and $\delta_0 > 0$, $C_1 > 0$ such that the set $W_\kappa = W_\kappa(\Pi_{C\delta}) = \{G : \rho_0(G - (1+\kappa)G_1 \, : P|\Pi_{C\delta}) < \kappa\gamma_n, G \in \Lambda_0\}$ is contained in Ω_0 for any $C > C_1$ and any $0 < \delta < \delta(C_1) < \delta_0$. Thus we have

$$E[\hat{U}_n] \ge E\left[\left(\sum_{j=2}^{k} p_j u_{nj}\right)^{-1} \chi(\hat{P}_n \in P + b_n W_\kappa)\right] \ge$$

$$k^{-1} \min_{2 \le j \le k} E[u_{nj}^{-1} \chi(\hat{P}_n \in P + b_n W_\kappa)] \ge C I(D_n) \tag{5.15}$$

where D_n is the set of all $z = (z_{n1}, \ldots, z_{nk})$ such that $nz_{n1}, \ldots, nz_{nk}$ are nonnegative whole numbers, $z_{n1} + \ldots + z_{nm} = 1$ and

$$J_{n3}(z) = \sum_{i=1}^{k} (z_{ni}^{1/2} - (p_i + (1+\kappa) b_n g_{n1i})^{1/2})^2 < \kappa_1^2 \gamma_n^2. \tag{5.16}$$

Then the number of elements D_n does not exceed $C(n\gamma_n\kappa)^{2\zeta} \prod_{i=1}^{k} p_i^{1/2}$. Hence, by (5.14),(5.15), arguing similarly to (5.4) - (5.12), we get

$$I(D_n) > C(n\gamma_n^2)^{\zeta} \min_{2 \le j \le k} \exp\left\{-2(1+\kappa)^2 \sum_{i=1}^{k} (q_{n1i}^{1/2} + q_{nji}^{1/2} - 2p_i^{1/2})^2 + 4\sum_{i=1}^{k} (q_{n1i}^{1/2} - p_i^{1/2})^2\right\} =$$

$$C(n\gamma_n^2)^{\zeta} \min_{2 \le j \le k} \exp\left\{-\frac{1}{2} n b_n^2 \int_S (g_1 + g_j)^2 \, dP (1 + o(1)) + \right.$$

$$\left. n b_n^2 \int_S g_1^2 dP + o(n b_n^2)\right\} \tag{5.17}$$

under the corresponding choice of $\kappa = \kappa(n) \to 0$, $\delta = \delta(n) \to 0$, $C_1 = C_1(n) \to \infty$ as $n \to \infty$. This completes the proof of Theorem 3.3.

References.

1. Anderson, T.W. and Darling, D.A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *Ann.Math.Statist.* **23**, 193-213.

2. Barone, P., Gigli, A. and Piccioni, M. (1995). Optimal importance sampling for some quadratic forms of ARMA process. *IEEE Trans. on Inform. Th.* **45**, 1834 - 1844.

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

4. Bucklew, J.A. (1990). *Large deviations techniques in decision, simulation and estimation.* John Wiley and Sons NY.

5. Bucklew J.A, Ney, P. and Sadowsky, J.S. (1990). Monte-Carlo simulation and large deviation theory for uniformly recurrent Markov chains. *J.Appl.Probab.* **27** 44-59.

6. Chen, J-C, Lu, D., Sadowsky, J.S. and Yao, K. (1993). On importance sampling in digital communications. Part I: Fundamentals; Part II: Trelis coded modulation. *IEEE Journal on Selected Areas in Communications.* **11.** 289-308.

7. Denker, M. (1985). *Asymptotic distribution theory in nonparametric statistics.* Vieweg Advanced Lectures in Mathematics. Vieweg Braunschweig, Wiesbaden.

8. Ermakov, M.S. (1993). Large deviations of empirical measures and hypothesis testing. Ẑapiski Nauchn. Semin. LOMI RAN **207**, 37-60.

9. Ermakov, M.S. (1994). On lower bounds of probabilities of moderate large deviations of tests and estimators. *Zapiski Nauchn. Semin. POMI RAN* **216**, 52-61.

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

11. Groeneboom, P., Oosterhoff, J. and Ruymgaart, F.H. (1979). Large Deviation Theorems for Empirical Probability Measures. *Ann.Probab.* **7**, 553-586.

12. Hammersley, J.M. and Handscomb, D.C. (1964). *Monte - Carlo Methods.* NY Chapman and Hall.

13. Inglot, T., Kallenberg, W.C.M. and Ledwina, T. (1992). Strong moderate deviation theorem. *Ann.Probab.* **20**, 987-1004.

14. 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.

15. Lehtonen,T. and Nyhrinen,H. (1992). On asymptotically efficient simulation of ruin probabilities in a Markovian environment. *Scand. Actuarial J.* **1.** 60-75.

16. Sadowsky, J.S. (1991). Large deviation theory and efficient simulation of excessive backlogs in a GI/GI/m queue. *IEEE Trans. Automat. Control.* **36** 1383-1394.

17. Sadowsky, J.S. (1996) On Monte-Carlo estiomation of large deviation probabilities. The Annals of Applied Probability. **6** 399-422.

18. Sadowsky, J.S. and Bucklew, J.A. (1990). On Large Deviation Theory and Asymptotically Efficient Monte-Carlo Estimation. *IEEE Trans. on Inf.Theory.* **36**, 579-588.

19. Saulis, L. and Statulevichius, V. (1989). *Limit theorems for large deviations.* Vilnius Mokslas.

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

21. Siegmund, D. (1976). Importance sampling in the Monte-Carlo study of sequential tests. *Ann.Statist.* **4**, 673-684.