



HAL
open science

A Review on some large deviation results

Philippe Barbe

► **To cite this version:**

Philippe Barbe. A Review on some large deviation results. Annales de l'ISUP, 1994, XXXVIII (1), pp.3-24. hal-03659824

HAL Id: hal-03659824

<https://hal.science/hal-03659824>

Submitted on 5 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Pub. Inst. Stat. Univ. Paris
XXXVIII, fasc. 1, 1994, 3 à 24

A Review on some large deviation results

Philippe BARBE
CNRS - Université Paul Sabatier

October 1993

Running title : large deviation results

Abstract : We review some recent results on large deviations related to a sample of independent and identically distributed random variables. It includes the sequential empirical measure, rank statistics, fractional ARIMA processes, linear processes, linear models and various new contraction principles.

Keywords : large deviations, sequential measure, time series, rank statistics, contraction principle.

AMS (1991) classification : 60F10

Correspondance to : Philippe Barbe

Laboratoire de Statistique et Probabilités

Université Paul Sabatier

31062 TOULOUSE

FRANCE

1. Introduction.

The aim of this paper is to review some new techniques and results in the area of large deviations that have been obtained recently by the author and some coauthors.

We first recall what is a large deviation principle (LDP).

Consider a topological space Θ endowed with its Borel σ -field, and let X_λ be a family of Θ -valued random variables (r.v.'s) indexed by $\lambda \in [0, \infty)$.

Definition 1.1. *The family $(X_\lambda)_{\lambda>0}$ obeys a LDP if there exists a function $I(\cdot)$ from Θ into $[0, \infty]$ such that for any closed set $\Omega \in \Theta$,*

$$\limsup_{\lambda \rightarrow \infty} \lambda^{-1} \log P(X_\lambda \in \Omega) \leq -I(\Omega) \quad (1.1)$$

and for any open set $\Omega \in \Theta$,

$$\liminf_{\lambda \rightarrow \infty} \lambda^{-1} \log P(X_\lambda \in \Omega) \geq -I(\Omega) \quad (1.2)$$

where we define for all $\Omega \in \Theta$,

$$I(\Omega) := \inf\{I(x) : x \in \Omega\} \quad (1.3)$$

The function $I(\cdot)$ is called the rate function. It is convenient to introduce the following definition.

Definition 1.2. *The rate function $I(\cdot)$ is a good rate function if it is lower semi-continuous and the level sets $\Gamma_c := \{x \in \Theta : I(x) \leq c\}$ are compact.*

If (1.1) and (1.3) hold, we say that we have a LDP on closed sets while we say that we have a LDP on open sets if (1.2) and (1.3) hold.

Definitions 1.1. and 1.2. are classical. Many large deviations results may be found in the books by Azencott (1980), Varadhan (1984), Stroock (1984), Deuschel and Stroock (1991), Dembo and Zeitouni (1993).

LDP's are important for applications in probability and statistics. Among the most well known applications are the Strassen (1964) theorem which may be

obtained with the LDP of Schilder (1961) on Wiener process (see e.g. Deuschel and Stroock (1991, Theorem 1.4.1)), the investigation of Bahadur (1967, 1971) efficiency (see Groeneboom and Oosterhoff (1977) for a review), the study of the increments of various processes such as the Erdős-Rényi (1970) laws (see Csörgö (1979) for a review). Large deviations are also important in simulation techniques, when one wants to simulate rare events (see e.g. Cottrell, Fort and Malgouyres (1983), Bucklew (1990), Sadowsky and Bucklew (1990)). Finally, LDP's are very useful tools in statistical physics and statistical fields theory to study phase transition phenomenas (see e.g. Ellis (1985)).

To make things clear, let us sketch a statistical example. Consider an i.i.d. sample of real valued r.v.'s X, X_1, X_2, \dots, X_n . Assume that $EX = 0$ and $EX^2 = 1$. Then, the empirical mean \bar{X}_n obeys the central limit theorem, i.e.

$$\lim_{n \rightarrow \infty} P(n^{1/2} \bar{X}_n \geq x) = 1 - \Phi(x)$$

where $\Phi(\cdot)$ denotes the standard normal distribution function. If $(\alpha_n)_{n \geq 1}$ is a sequence of positive numbers which tends to ∞ and $n \rightarrow \infty$, the CLT says nothing about $P(n^{1/2} \bar{X}_n \geq \alpha_n x)$, except that this probability tends to 0 when n tends to ∞ . For the special case where $\alpha_n = n^{1/2}$, we are looking at $P(\bar{X}_n \geq x)$, which is just a large deviation in the sense defined above.

Next, assume that the mean EX is unknown, and that the distribution of X is symmetric around its mean. Then, we can estimate the mean in using \bar{X}_n but also, say, in using the median M_n of X_1, \dots, X_n . One way to compare these two estimators when the true mean is say m , is to look at the probabilities $P(|\bar{X}_n - m| \geq x)$ and $P(|M_n - m| \geq x)$. These probabilities are large deviations type probabilities. They are typically of the form

$$P(|\bar{X}_n - m| \geq x) = \exp(-nh_1(x)(1 + o(1))) \text{ as } n \rightarrow \infty,$$

$$P(|M_n - m| \geq x) = \exp(-nh_2(x)(1 + o(1))) \text{ as } n \rightarrow \infty$$

If $h_1(x) \geq h_2(x)$ for all x , \bar{X}_n is in some sense a better estimator than M_n since it converges to m at a faster rate. Hence, a test based on \bar{X}_n should be better than a test based on M_n . This is very roughly what is the comparison of tests in the Bahadur sense.

The problem of obtaining a LDP has two aspects : to prove that the family of r.v.'s obeys a LDP, and then to identify the rate function. Moreover, for some applications it may also be interesting to study the continuity or the convexity of the rate function.

In the easiest case, we may obtain a LDP from a previously known LDP in using a change of variable formula. For instance, let Θ' be a second topological space.

Theorem 1.1. *Let h be a continuous mapping from Θ into Θ' and assume that $(X_\lambda)_{\lambda>0}$ obeys a LDP with rate $I(\cdot)$. Then, $h(X_\lambda)$ obeys a LDP with rate $I \circ h^{-1}$.*

Theorem 1.1. is called a contraction principle. Its proof may be found in the above mentioned books on large deviations.

There exist many LDP's for various rv.'s such as solutions of stochastic differential equations, sums of r.v.'s, empirical probability measure of a Markov chain, etc. In this review, we concentrate only upon two aspects and their applications. The first one is an extension of classical results on the empirical p.m. of an i.i.d. sequence and on processes with independent increments. The second one is some extension of the contraction principle. It turns out that many processes that are defined in using a sequence of i.i.d. r.v.'s can be handled with the results described below.

The starting point in section 2 will be a classical LDP on the empirical probability measure (p.m.) of n i.i.d. r.v.'s. We shall show that this result does not give easily LDP's on some classical statistics such as the rank statistics, and that it is natural to study LDP's for the point process of the sample. We shall also show the limitation of the classical contraction principle.

In section 3, we shall describe a technique which enables us to obtain LDP's for the empirical p.m. associated to a rescaled version of the point process of the sample.

In section 4, we explain some ideas which lead to some refinements of the contraction principle.

Finally, section 5 is devoted to review some applications.

At this stage, we mention that the aim of this paper is more to explain the general tools and techniques than to give very precise results. Sometimes, to give

precise results would have length this paper considerably. We hope that it will be more benefit for the reader to have a kind of general exposition, leaving out the too technical details that may be found in the papers quoted in references.

2. LDP for the empirical measure.

We now consider a sequence of i.i.d. r.v.'s X_1, \dots, X_n on a Hausdorff topological space \mathcal{X} . Their empirical p.m. is defined by

$$P_n := n^{-1} \sum_{1 \leq i \leq n} \delta_{X_i}.$$

We endow the space $\mathcal{P}(\mathcal{X})$ of all p.m.'s on \mathcal{X} with the topology of weak convergence and the corresponding Borel σ -field. The p.m. P_n is a random element in $\mathcal{P}(\mathcal{X})$. The interest of P_n is that many quantities of statistical interest may be written as functional of P_n .

For instance, if \mathcal{X} is a vector space, we can define the mean functional on $\mathcal{P}(\mathcal{X})$ by

$$M(\mu) := \int x d\mu(x), \mu \in \mathcal{P}(\mathcal{X}).$$

Then, the empirical mean $n^{-1} \sum_{1 \leq i \leq n} X_i$ is just $M(P_n)$. If $\mathcal{X} = \mathbb{R}$, the empirical distribution function (d.f.) is

$$F_n(x) := n^{-1} \sum_{1 \leq i \leq n} \mathbb{I}(X_i \leq x).$$

It is a random element in the space $\mathcal{D}(\mathbb{R})$ of all càdlàg functions of \mathbb{R} . If we define the mapping F from $\mathcal{P}(\mathbb{R})$ into $\mathcal{D}(\mathbb{R})$ by

$$F(\mu)(x) := \int \mathbb{I}(y \leq x) d\mu(y)$$

we see that $F_n = F(P_n)$. We deduce that the median, says, is also a function of P_n .

Therefore, having in mind the contraction principle, we see that there is a certain interest in having a LDP on P_n . This has been worked out by many authors among them Sanov (1957), Borovkov (1962), Stone (1974), Hoadley (1976), Donsker and Varadhan (1976), Bahadur and Zabell (1979), Groeneboom, Oosterhoff and Ruymgaart (1979) Csizár (1984) and De Acosta (1992). In the sequel,

we shall make use of the following consequence of any one of the papers mentioned above after Donsker and Varadhan (1976).

Theorem 2.1. *The p.m. P_n obeys a LDP in $\mathcal{P}(\mathcal{X})$ endowed with the weak topology, with good rate function the Kullback-Leibler information number*

$$K(Q, P) = \begin{cases} \int \log(dQ/dP)dQ & \text{if } Q \leq P \\ \infty & \text{otherwise} \end{cases}$$

The meaning of Theorem 2.1. is roughly that if Ω is a set of p.m.'s, then

$$P(P_n \in \Omega) \approx \exp(-n \inf_{Q \in \Omega} K(Q, P)).$$

Now, if T_n is a statistics which may be written as a function $T(P_n)$, and if $T(\cdot)$ is continuous, then Theorem 2.1. and the contraction principle lead to a LDP for T_n . Using this approach, Groeneboom, Oosterhoff and Ruymgaart (1979) and Groeneboom and Shorack (1981) obtain LDP's for L-statistics.

Now, we can make clear a first limitation of the contraction principle. The mean functional $M(\cdot)$ defined above is generally not weakly continuous since the function $x \rightarrow x$ is generally unbounded. Therefore, we cannot obtain a LDP on the empirical mean in using Theorem 2.1..

For the functional $F(\cdot)$ ($\mathcal{X} = \mathbb{R}$ in this case) a different problem occurs. For a fixed y , the function $x \rightarrow \mathbb{I}(x \leq y)$ as a function from \mathbb{R} into \mathbb{R} is bounded but not continuous. Consequently, $F(\cdot)$ is not weakly continuous, and we cannot obtain a LDP on the empirical d.f., even at a fixed point, with Theorem 2.1. and the contraction principle.

We now turn to a more sophisticated limitation of Theorem 2.1.. Assume that $\mathcal{X} = \mathbb{R}$ and let $R_{i,n}$ be the rank of X_i in the sample X_1, \dots, X_n . Consider the rank statistics

$$T_n := n^{-1} \sum_{1 \leq i \leq n} a(i/n, R_{i,n}/n)$$

when $a(\cdot, \cdot)$ is a continuous and bounded function (this requirement on $a(\cdot, \cdot)$ may be considerably weakened but we try to avoid extra complications here). The statistics T_n cannot be written as a functional of P_n . However, one may obtain a LDP for T_n in using Theorem 2.1. as follows. Define a new sequence of i.i.d. r.v.'s uniformly distributed over $[0, 1]$, U_1, \dots, U_n , which are independent

of the X_i 's. Define the empirical p.m. $Q_n := n^{-1} \sum_{1 \leq i \leq n} \delta_{(U_i, X_i)}$. Replacing X_i by (U_i, X_i) , we see that Theorem 2.1. leads to a LDP for Q_n . The p.m. Q_n is a random element in the space $\mathcal{P}([0, 1] \times \mathbb{R})$ of all p.m.'s on $[0, 1] \times \mathbb{R}$. Define the mapping $F_i, i = 1, 2$ from $\mathcal{P}([0, 1] \times \mathbb{R})$ into $\mathcal{D}[0, 1]$ if $i = 1$ and $\mathcal{D}(\mathbb{R})$ if $i = 2$ by

$$F_i(\mu)(x) := \int \mathbb{I}(y_i \leq x) d\mu(y_1, y_2) \quad i = 1, 2.$$

Notice that $F_1(Q_n)$ is just the empirical p.m. of U_1, \dots, U_n , while $F_2(Q_n)$ is that of X_1, \dots, X_n . Since the U_i 's are independent of the X_i 's, we have $Q_n \stackrel{d}{=} Q'_n := n^{-1} \sum_{1 \leq i \leq n} \delta_{(U_{i,n}, X_i)}$ where $U_{1,n} \leq \dots \leq U_{n,n}$ is the order statistics of U_1, \dots, U_n . Then, notice that $i/n = F_1(Q'_n)(U_{i,n})$ while $R_{i,n}/n \stackrel{d}{=} F_2(Q'_n)(X_i) \stackrel{d}{=} F_2(Q_n)(X_i)$. Hence, since the X_i 's are i.i.d.,

$$T_n \stackrel{d}{=} \int a(F_1(Q_n)(u), F_2(Q_n)(x)) dQ_n(u, x).$$

Since a LDP deals only with the distribution of r.v.'s, we see that T_n may be represented as a function of Q_n . Hence, there is a hope to obtain a LDP for T_n from Theorem 2.1. and the contraction principle. This has been done by Praagman (1988) in using the above described technique. A direct proof of a LDP for T_n is due to Woodworth (1970).

The reason why it is somehow so complicated to handle rank statistics is essentially that the empirical p.m. P_n does not keep track of the order in which the sample X_1, \dots, X_n came out. The p.m. P_n is the sample up to a permutation. The reason why the Pragman (1988) technique works is that

$$\mathbb{P}_n := n^{-1} \sum_{1 \leq i \leq n} \delta_{(i/n, X_i)} \stackrel{d}{=} n^{-1} \int \delta_{(F_1(Q_n)(u), x)} dQ_n(u, x), \quad (2.1)$$

and that \mathbb{P}_n keeps track of the order in which the sample came out. The p.m. \mathbb{P}_n is called the sequential empirical measure (s.e.m.). It is clear that representation (2.1) and Theorem 2.1. and the contraction principle lead to a LDP for \mathbb{P}_n . However, one can do much better.

Instead of considering a sequence X_1, \dots, X_n in \mathbb{R} , consider an i.i.d. random field $(X_i)_{i \in \mathbb{N}^d}$ in a general Hausdorff topological space \mathcal{X} , and define the s.e.m.

$$\mathbb{P}_{n,d} := n^{-d} \sum_{i \in \mathbb{n}[0,1]^d} \delta_{(i/n, X_i)}.$$

This p.m. is a random element in the space $\mathcal{P}([0, 1]^d \times \mathcal{X})$ of all p.m.'s on $[0, 1]^d \times \mathcal{X}$, endowed with the topology of weak convergence. Clearly, $\mathbb{P}_n = \mathbb{P}_{n,1}$. Now, if $d > 1$, representation (2.1) is useless. The case $d > 1$ is of interest in statistical physics and probably in other fields such as image processing. Barbe and Broniatowski (1993) developed a new technique which leads to a LDP for $\mathbb{P}_{n,d}$ and is explained here in section 3.

To make perfectly clear the interest of \mathbb{P}_n , consider the weighted V-statistics

$$V_n := n^{-2} \sum_{1 \leq i, j \leq n} w(i/n, j/n) a(X_i, X_j)$$

where $a(\cdot, \cdot)$ is, say, some function from $\mathcal{X} \times \mathcal{X}$ into \mathbb{R} . The limiting distribution of V_n has been obtained very recently by O'Neil and Redner (1993) for V-statistics of degree 2, and Major (1992) for an arbitrary degree. Clearly, $V_n = V(\mathbb{P}_n)$ if we define for any $\mu \in \mathcal{M}([0, 1] \times \mathcal{X})$,

$$V(\mu) := \int \int w(u, v) a(x, y) d\mu(u, x) d\mu(v, y).$$

LDP on \mathbb{P}_n leads also some older classical results. If \mathcal{X} is a vector space, consider the partial sum process

$$S_n(t) := n^{-1} \sum_{1 \leq i \leq nt} X_i$$

which is a random element in $\mathcal{D}[0, 1]$. Define the mapping $S(\cdot)$ from $\mathcal{M}([0, 1] \times \mathcal{X})$ into $\mathcal{D}[0, 1]$ by

$$S(\mu)(t) := \int \mathbb{I}(u \leq t) x d\mu(u, x).$$

Then, $S_n(\cdot) = S(\mathbb{P}_n)(\cdot)$ so that a LDP on \mathbb{P}_n will lead to one for $S_n(\cdot)$.

The limiting behaviour of the sem \mathbb{P}_n is as follows. If A is a Borel subset of $[0, 1]^d$, and B a Borel subset of \mathcal{X} , it is easy to check that $\mathbb{P}_n(A \times B)$ converges a.s. to $|A| \times P(B)$ where $|\cdot|$ is the Lebesgue measure on $[0, 1]^d$ and P the p.m. of X_1 . Since a measure is completely defined by its values on rectangles, \mathbb{P}_n converges a.s. to $\mathbb{IP} := |\cdot| \times P$. The weak convergence of the process $\nu_n := n^{1/2}(\mathbb{P}_n - \mathbb{IP})$ as a process indexed by a class of functions follows from Theorem (10.6) in Pollard (1990).

3. Large deviation principle for the sequential empirical measure.

We denote $\mathcal{B}([0,1]^d)$ the Borel σ -field of $[0,1]^d$ for the usual topology. If $A \in \mathcal{B}([0,1]^d)$, we define

$$\mathbb{P}_{d,n} \langle A \rangle := n^{-1} \sum_{i/n \in A} \delta_{X_i}.$$

We may see $\mathbb{P}_{d,n} \langle \cdot \rangle$ in two ways :

i) it is a process indexed by sets $A \subset \mathcal{B}([0,1]^d)$ and taking its values in the space $\mathcal{M}_1(\mathcal{X})$ of all measures on \mathcal{X} with mass less than 1. Notice that if $A \cap B = \emptyset$ then $\mathbb{P}_{d,n} \langle A \rangle$ is independent of $\mathbb{P}_{d,n} \langle B \rangle$. Hence, as a process indexed by sets, $\mathbb{P}_{d,n} \langle \cdot \rangle$ has independent increments. Barbe and Broniatowski (1993a) extended the proof of McBride (1974) and Lynch and Sethuraman (1989) to obtain a fairly general deviation theory for abstract set indexed processes with independent increments. This leads to a LDP for the process $\mathbb{P}_{d,n} \langle \cdot \rangle$ (see below).

ii) If B is a Borel subset of \mathcal{X} , $\mathbb{P}_{d,n} \langle A \rangle (B) = \mathbb{P}_{d,n}(A \times B)$. Hence we may view $\mathbb{P}_{d,n} \langle \cdot \rangle (\cdot)$ as a measure on the product space $[0,1]^d \times \mathcal{X}$. This measure may be written as a function of $\mathbb{P}_{d,n} \langle \cdot \rangle$,

$$\Psi(\mathbb{P}_{d,n} \langle \cdot \rangle)(A \times B) = \mathbb{P}_{d,n} \langle A \rangle (B).$$

Since a measure is completely determined by its values on rectangles, it is easy to see that $\Psi(\cdot)$ is one-to-one. Provided we endow the space where $\mathbb{P}_{d,n} \langle \cdot \rangle$ lives with a suitable topology, the mapping $\Psi(\cdot)$ is continuous when $\mathcal{P}([0,1]^d \times \mathcal{X})$ is endowed with the topology of weak convergence. Then, using the contraction principle, we obtain a LDP on $\mathbb{P}_{d,n}$ from that on $\mathbb{P}_{d,n} \langle \cdot \rangle$.

To make clear how we can obtain a LDP on $\mathbb{P}_n \langle \cdot \rangle$, we need to interpret Theorem 2.1. as $P(P_n \approx Q) \approx \exp(-nK(Q, P))$. The \approx signs are not well defined but one should understand $P(P_n \approx Q)$ as the probability that P_n is very close to Q . The second sign \approx means somehow a kind of asymptotic relation which holds when n gets large. Let Ω be a set of functions from $\mathcal{B}([0,1]^d)$ into $\mathcal{M}_1(\mathcal{X})$. Call a finite family $A = (A_1, \dots, A_k)$ a partition if each A_i is a rectangle in $[0,1]^d$ and the A_i 's are disjoint. Let $\Omega \langle A_i \rangle := \{Q \langle A_i \rangle : Q \in \Omega\}$. For any partition A , notice the following facts :

$$n \mathbb{P}_n \langle A_i \rangle / \#(nA_i \cap \mathbb{N}) \stackrel{d}{=} P_{\#(nA_i \cap \mathbb{N})}$$

. If $i \neq j$, $\mathbb{P}_n \langle A_i \rangle$ is independent of $\mathbb{P}_n \langle A_j \rangle$.

. $\#(nA_i \cap \mathbb{N})/n \approx |A_i|$ for n large.

Hence, using Theorem 2.1., we see that

$$P(\mathbb{P}_n \approx \mathbb{Q}) \leq P(n\mathbb{P}_n \langle A_i \rangle / \#(nA_i \cap \mathbb{N}) \approx n\mathbb{Q}_n \langle A_i \rangle / \#(nA_i \cap \mathbb{N}) : 1 \leq i \leq k) \quad (3.1)$$

$$\begin{aligned} &\approx \prod_{1 \leq i \leq k} P(P_{\#(nA_i \cap \mathbb{N})} \approx \mathbb{Q}_n \langle A_i \rangle / |A_i|) \\ &\approx \prod_{1 \leq i \leq k} \exp(-\#(nA_i \cap \mathbb{N})K(\mathbb{Q} \langle A_i \rangle / |A_i|, P)) \\ &\approx \exp(-n \sum_{1 \leq i \leq k} |A_i|K(\mathbb{Q} \langle A_i \rangle / |A_i|, P)). \end{aligned} \quad (3.2)$$

Now, view \mathbb{Q} as a p.m. on the product space $[0, 1]^d \times \mathcal{X}$ and denote $\mathbb{Q}^1(u)$ ($0 \leq u \leq 1$) the conditional p.m. of \mathbb{Q} given that its first marginal at u . Then, if A_i shrinks around some u , $\mathbb{Q} \langle A_i \rangle / |A_i|$ tends to $\mathbb{Q}^1(u)$ and we see that when each element of the partition A gets smaller and smaller, the expression (3.2) should tend to

$$\exp(-n \int K(\mathbb{Q}^1(u), P) du),$$

which is just $\exp(-nJ(\mathbb{Q}))$ with $J(\mathbb{Q}) := K(\mathbb{Q}, P)$. Now, if we take an other function in Ω , say \mathbb{Q}' , we obtain that $P(\mathbb{P}_n \approx \mathbb{Q}') \approx \exp(-nJ(\mathbb{Q}'))$. Assume for instance that $J(\mathbb{Q}') > J(\mathbb{Q})$. Then,

$$\begin{aligned} &P(\mathbb{P}_n \approx \mathbb{Q}' \text{ or } \mathbb{P}_n \approx \mathbb{Q}) \\ &\leq P(\mathbb{P}_n \approx \mathbb{Q}') + P(\mathbb{P}_n \approx \mathbb{Q}) \\ &\approx \exp(-nJ(\mathbb{Q}')) + \exp(-nJ(\mathbb{Q})) \\ &\approx \exp(-n \inf(J(\mathbb{Q}), J(\mathbb{Q}'))) \text{ as } n \rightarrow \infty. \end{aligned}$$

This explains that

$$P(\mathbb{P}_n \in \Omega) \leq \exp(-n \inf_{\mathbb{Q} \in \Omega} J(\mathbb{Q})(1 + o(1))) \text{ as } n \rightarrow \infty.$$

For the lower bound on open sets, notice that if Ω is open and $\mathbb{Q} \in \Omega$, there exists a neighborhood $\{\mathbb{Q}\}^\epsilon$ of \mathbb{Q} in Ω . If the partition A contains only some very

small sets, any p.m. \mathbb{Q}' such that $\mathbb{Q}' < A_i > \approx \mathbb{Q} < A_i >$ for any $A_i \subset A$ belongs to $\{\mathbb{Q}\}^\epsilon$. Hence,

$$P(\mathbb{P}_n \in \Omega) \geq P(\mathbb{P}_n < A_i > \approx \mathbb{Q} < A_i > : A_i \in A)$$

and the same argument as above yields the lower bound on open sets. The final result is as follows.

Theorem 3.1. *The sem $\mathbb{P}_{d,n}$ obeys a LDP on $\mathcal{P}([0,1]^d \times \mathcal{X})$ endowed with the topology of weak convergence, with good and convex rate function*

$$J(\mathbb{Q}) := \begin{cases} K(\mathbb{Q}, \mathbb{P}) & \text{if } \mathbb{Q}(\cdot \times \mathcal{X}) = |\cdot| \\ \infty & \text{otherwise} \end{cases}$$

Notice that since $\mathbb{P}_n(\cdot \times \mathcal{X}) = n^{-1} \sum_{i/n \in [0,1]^d} \delta_{i/n}$ converges weakly to $|\cdot|$ in a purely deterministic way, it is not surprising that the rate functional $J(\cdot)$ is infinite on measures \mathbb{Q} which do not have a uniform first marginal on $[0,1]^d$.

In fact, this construction may be used for general processes indexed by sets and with independent increments. For instance, Barbe and Broniatowski (1993a) consider an extremal process. Assume that $\mathcal{X} = \mathbb{R}$ and let

$$M_n(t) := \max_{1 \leq i \leq nt} X_i.$$

To obtain a LDP for $M_n(\cdot)$, we embed it into a process indexed by sets. For $A \in [0,1]$, define

$$M_n < A > := \max_{i/n \in A} X_i.$$

Then, $M_n < \cdot >$ is a process indexed by sets, and with independent increments. We can obtain a LDP for $M_n < \cdot >$ just applying the general result of Barbe and Broniatowski (1993a). Next, we map $M_n < \cdot >$ to $M_n(\cdot)$ in considering its restriction to the sets of the form $[0,t]$. Since $M_n(t) = M_n < [0,t] >$, the LDP for $M_n < \cdot >$ and the contraction principle lead to a LDP for $M_n(\cdot)$. Notice that $M_n < \cdot >$ has independent increments, while $M_n(\cdot)$ has not.

The technique described in this section may be used to obtain LDP for random walks process in a set endowed with an associative operation (the extremal process is in fact an exemple of such a random walk).

4. Contraction principles.

In this section we use the general formalism of large deviations introduced in the first section, and we are going to discuss the contraction principle.

Varadhan (1984) gives the following extension of the contraction principle, where we use notation of Definition 1.1..

Theorem 4.1. *If h is a uniform limit over compact sets of a family of continuous function h_λ , then $h_\lambda(X_\lambda)$ obeys a LDP with rate $I \circ h^{-1}$.*

Having in mind the example given in the introduction for the functionals $M(\cdot)$ and $F(\cdot)$, one can see that this theorem still does not give a LDP for $M(P_n)$ and $F(P_n)$ from Theorem 2.1.. However, it suggests the possibility that we can approximate h by some nice functions. For instance, going back to the function $M(\cdot)$ in the introduction, it suggests to define a truncated version of $M(\cdot)$. In the example of the mean, we assume Θ to be a separable Banach space.

For any $\tau > 0$, define $T_\tau(x) := x$ if $\|x\| \leq \tau$ and $T_\tau(x) := \tau x / \|x\|$ otherwise. Define the truncated version M_τ of $M(\cdot)$, from $\mathcal{P}(\Theta)$ into Θ by

$$M_\tau(\mu) := \int T_\tau(x) d\mu(x).$$

Since $T_\tau(\cdot)$ is continuous and bounded, $M_\tau(\cdot)$ is weakly continuous. Therefore, Theorems 2.1. and 1.1. give a LDP for $M_\tau(P_n)$ for any $\tau > 0$. Moreover, for a fixed μ , $M_\tau(\mu)$ tends to $M(\mu)$ as $\tau \rightarrow \infty$ so that we can reasonably try to take the limit as $\tau \rightarrow \infty$ in the LDP for $M_\tau(P_n)$ to obtain a LDP for $M(P_n)$. To proceed further, notice that $\|M_\tau(P_n) - M(P_n)\| \leq \Delta_n(\tau)$ where $\Delta_n(\tau) := n^{-1} \sum_{1 \leq i \leq n} \|X_i\| \mathbb{I}(\|X_i\| \geq \tau)$.

Assume that

$$\phi(t) := E \exp(t \|X_1\|) < \infty \text{ for all } t > 0. \quad (4.1)$$

Then, using standard Cramér estimates, we see that for all $t > 0$,

$$P(\Delta_n(\tau) \geq \epsilon) = P(\exp(t \sum_{1 \leq i \leq n} \|X_i\| \mathbb{I}(\|X_i\| > \tau)) \geq \exp(nt\epsilon))$$

$$\leq \exp(-nt\epsilon) (E \exp(t \| X_i \| \mathbb{I}(\| X_i \| > \tau)))^n \quad (4.2)$$

Since $\phi(t)$ is finite for all $t > 0$, we can take t arbitrary large in (4.2), and since

$$\lim_{\tau \rightarrow \infty} \log E \exp(t \| X_i \| \mathbb{I}(\| X_i \| > \tau)) = 0 \quad (4.3)$$

we see that (4.1) and (4.2) imply for any $\epsilon > 0$,

$$\lim_{\tau \rightarrow \infty} \limsup_{n \rightarrow \infty} n^{-1} \log P(\| M(P_n) - M_\tau(P_n) \| > \epsilon) = -\infty.$$

In particular, there is a sequence τ_k such that

$$\limsup_{n \rightarrow \infty} n^{-1} \log P(\| M(P_n) - M_{\tau_k}(P_n) \| \geq 1/k^2) \leq -k$$

Hence, if we define

$$\mathcal{F}_k := \{ \mu \in \mathcal{P}(\mathcal{X}) : \| M(\mu) - M_{\tau_k}(\mu) \| \leq 1/k^2 \}, \quad k \geq 1$$

we see that for large n and large k , it is very unlikely (even in the sense of large deviations) that P_n does not belong to \mathcal{F}_k . It means that 'asymptotically', P_n should be in the set $\bar{\mathcal{F}}_m := \bigcap_{k \geq m} \mathcal{F}_k$. This suggests that the rate function for P_n should be infinite outside of $\bar{\mathcal{F}}_m$. Now, we also see that $\bar{\mathcal{F}}_m$ is a set of p.m.'s for which $x \rightarrow x$ is uniformly integrable. And so the functional $M(\cdot)$ will be continuous on the set $\bar{\mathcal{F}}_m$. It turns out that Donsker and Varadhan (1976) shown that $x \rightarrow x$ is uniformly integrable on the level sets of $K(\cdot, P)$ provided (4.1) holds.

Before pointing out the general underlying mechanism, let us mention a flaw of this construction. When we define the set \mathcal{F}_k , we assume that $M(\mu)$ is finite. Since $M(\cdot)$ is not continuous, we cannot infer anything on the closeness property of \mathcal{F}_k . But Cauchy's convergence criterion suggests to define

$$\mathcal{G}_k := \{ \mu \in \mathcal{P}(\mathcal{X}) : \| M_{\tau_{k+1}}(\mu) - M_{\tau_k}(\mu) \| \leq 2/k^2 \}.$$

Clearly, $\mathcal{F}_k \cap \mathcal{F}_{k+1} \subset \mathcal{G}_k$. Hence, $P_n \in \mathcal{G}_k$ with high probability and the preceding arguments may be carried over in replacing \mathcal{F}_k by \mathcal{G}_k . Now, since the M_{τ_k} 's are weakly continuous, \mathcal{G}_k is weakly closed. Moreover, if $\mu \in \bigcap_{k \geq m} \mathcal{G}_k$ then $M_{\tau_k}(\mu)$ is a Cauchy sequence, and therefore, $M(\mu)$ exists.

With notation of Definition 1.1., the general mechanism is as follows (Barbe (1993a)).

Theorem 4.2. Assume that there exists a sequence of sets $\mathcal{G}_k \in \Theta$ such that

$$\limsup_{\lambda \rightarrow \infty} \lambda^{-1} \log P(X_\lambda \in \mathcal{G}_k^c) \leq -k \text{ and}$$

$$\text{for any } c > 0, \text{cl}(A_k) \cap \Gamma_c = \mathcal{G}_k \cap \Gamma_c.$$

Let $\mathcal{H}_k = \bigcap_{i \geq k} \mathcal{G}_i$. Assume that for any k large enough, $h|_{\mathcal{H}_k}$ is continuous at each point of \mathcal{H}_k . Then $I \circ h^{-1}$ is a good rate function provided I is also a good rate function, and $h(X_\lambda)$ obeys a LDP on closed sets.

The problem with Theorem 4.2. is that it gives only the upper bound on closed sets. The lower bound on open sets may be handled with the following result which is an extension of Lemma 2.1.4 in Deuschel and Stroock (1989) and is proved in Barbe (1993a).

Theorem 4.3. Assume that (Θ', ρ) is a metric space. If h_n is a sequence of function from Θ into Θ' such that for any $\epsilon \geq 0$,

$$ii) \lim_{n \rightarrow \infty} \limsup_{\lambda \rightarrow \infty} \lambda^{-1} \log P(\rho(h_n(X_\lambda), h(X_\lambda)) \geq \epsilon) = -\infty \text{ where}$$

$$i) h_n \text{ is continuous at each point } x \in \Theta \text{ such that } I(x) < \infty$$

Then $h(X_\lambda)$ obeys a LDP on open sets with rate function $I \circ h^{-1}$.

Generally speaking, Theorems 4.1. and 4.1. are the tools to deal with functional which are not continuous due to a lack of boundedness. For functionals of the empirical p.m. or of the sem of the form $\int \psi(x) d\mu(x)$, they are the tool to handle unbounded function $\psi(\cdot)$.

Now, the other example we pointed out in the introduction was the empirical d.f. at a fixed point x which is not a continuous functional of P_n . Recall that P_n obeys a LDP with rate $K(\cdot, P)$ and that if $K(Q, P)$ is finite, then $Q \ll P$. Now, assume that x is a continuity point of P . If $K(Q, P) < \infty$, x is also a continuity point of Q . This suggests that for the the contraction principle to hold, we just need the function h to be continuous on the set where the rate function $I(\cdot)$ is finite. It turns out that this is true (see Barbe (1993b)) and explains assumption ii) in Theorem 4.3.. This is the tool to handle the non-continuity that occurs in the rank statistics example (see e.g. Barbe (1993b), Barbe and Hallin (1993)).

To conclude this section, we would like to mention that the contraction principle holds under rather weak conditions. There is however a point which is not

clear : what is the minimum requirement to obtain a contraction principle? Having in mind the example of the mean functional $M(\cdot)$, it should lead to a LDP on $M(P_n)$ from that on P_n assuming only that the moment generating function $\phi(t)$ in (4.1) is finite in a neighborhood of 0. One can obtain LDP on $M(P_n)$ from that on P_n using some convexity arguments (see e.g. Groeneboom, Oosterhoff and Ruyngaert (1978)) but it would be very interesting to avoid convexity techniques.

5. Some applications.

In this section we review some applications of the above mentioned results.

i) Application to sequential statistics

Many sequential statistics may be written as function of the sequence $(\mathbb{P}_n)_{n \geq 1}$. Assume that we have to consider \mathbb{P}_n for $1 \leq n \leq m$ (this typically arises when one wants to obtain an evaluation of the tail probability of a stopping time). Then for some $t \in (0, 1]$, $\mathbb{P}_n = \mathbb{P}_{[nt]}$. But

$$\mathbb{P}_{[nt]} := (m/[nt]) \int \delta_{(mu/[nt], x)} \mathbb{I}(u \leq t) d\mathbb{P}_m(u, x).$$

For m large, $m/[nt] \approx 1/t$ and roughly

$$\mathbb{P}_{[nt]} \approx t^{-1} \int \delta_{(u/t, x)} \mathbb{I}(u \leq t) d\mathbb{P}_m(u, x).$$

Hence, we see that some sequential statistics may be approximated by functional of the sem \mathbb{P}_n . This approach has been developed in Barbe and Broniatowski (1993b) to obtain a LDP on sequential rank statistics.

ii) Application to fractional ARIMA process.

Consider a fractional ARIMA process

$$(1 - B)^d \phi(B) Y_t = \theta(B) X_t$$

where ϕ, θ are two polynomials with root outside the unit circle and B is the lag operator $BX_t = X_{t-1}$. Akonom and Gouriéroux (1987) shown that approximately when n is large and $0 \leq t \leq 1$,

$$Y_{[nt]} \approx c \sum_{1 \leq i \leq [nt]} (nt - i)^d X_i,$$



for some constant c which depends on d, ϕ and θ . We see that this representation leads to

$$\begin{aligned} Y_{[nt]} &\approx cn^d \sum_{1 \leq i \leq [nt]} (t - i/n)^d X_i \\ &= cn^{d-1} \int \mathbb{I}(u \leq t)(t - u)^d x d\mathbb{P}_n(u, x) \end{aligned}$$

Using this approximation, Barbe and Broniatowski (1993c) shown that $n^{-d+1}Y_{[n]}$ obeys a LDP in $\mathcal{D}[0, 1]$. The rate function is given by

$$I(f) := \int h(f^{(d)}(x)\theta(1)/\phi(1))dx$$

where $h(x) := \sup_{t>0}(tx - \log E \exp(tX_1))$, and $f^{(d)}$ is the fractional derivative of order d of f (for fractional calculus, see e.g. Oldam and Spanier (1974), Ross '1974)). When $d = 1$ and $\phi = \theta = 1$, $Y_{[nt]}$ is the classical partial sum process. Thus, we have a generalisation of the result of Varadhan (1966) on partial sum to fractional ARIMA process.

iii) Conditional s.e.m. and linear model

Assume that $X_i = (X_{1,i}, X_{2,i})$ is a pair of r.v.'s (we do not assume that $X_{1,i}$ is independent of $X_{2,i}$). One can use the construction developed in section 3 to obtain a LDP for the \mathbb{P}_n conditioned on $X_{2,1}, \dots, X_{2,n}$. This has been done in Barbe (1993a) in order to handle linear models. Conditional LDP for the usual empirical p.m. has been obtained by Comets (1989) by a slightly different technique. Indeed, if one consider the linear model

$$Z_i = aX_{2,i} + X_{1,i}$$

one often consider such model where $X_{2,i}$ is either deterministic, either random. In the random case, one often consider the model conditioned on the $X_{2,i}$'s. All these cases may be handled. It turns out that many statistics are functionals of \mathbb{P}_n conditioned on $X_{2,1}, \dots, X_{2,n}$. To make the idea clear, a LDP for \mathbb{P}_n conditioned on $X_{2,1}, \dots, X_{2,n}$ is obtained in considering the process $\mathbb{P} < . >$ as in section 3 and reworking all the construction conditionally on $X_{2,1}, \dots, X_{2,n}$.

This has an interesting application in econometrics. It is well known that when one has to test some hypothesis on a coefficient in a linear model, one may use at least three different tests : the Wald test, the Lagrange multiplier test, the likelihood ratio test. It turns out that these three statistics have the same limiting

distribution under the null hypothesis and under a sequence of local alternative. In Barbe and Bertail (1993), we obtain the exact Bahadur slopes of these test statistics which show that they are not equivalent.

iv) Sequential empirical measure in a space of sequences

We can generalize the result on \mathbb{P}_n in the following direction. Let $X := (X_i)_{i \in \mathbb{Z}}$ be a sequence of i.i.d. r.v.'s. We consider the space \mathcal{X}^∞ of all sequences $(x_i)_{i \leq 0}$. This space is endowed with the norm $\| (x_i)_{i \leq 0} \|_a := \sum_{i \leq 0} a_i |x_i|$ for some summable positive sequence $a = (a_i)_{i \leq 0}$. We define $\theta^j X := (X_{i+j})_{i \leq 0}$, and consider the sem

$$\mathbb{P}_n^\infty := n^{-1} \sum_{1 \leq i \leq n} \delta_{(\theta^i X, i/n)}.$$

This is a p.m. on the infinite product space $\mathcal{X}^\infty \times [0, 1]$. Barbe (1993) obtained a LDP for \mathbb{P}_n^∞ in using the following technique. We first define

$$\mathbb{P}_n^k := n^{-1} \sum_{1 \leq i \leq n} \delta_{(\theta_k^i X, i/n)}$$

where $\theta_k^i X = (X_{i+j})_{-k \leq j \leq 0}$. Then, \mathbb{P}_n^k is a p.m. on $\mathcal{X}^{k+1} \times [0, 1]$. For fixed k , the $\theta_k^i X$'s are k -dependent. Hence, the process $\mathbb{P}_n^k \langle A \rangle := n^{-1} \sum_{i/n \in A} \delta_{(\theta_k^i X, i/n)}$ has not independent increments. However, define

$$\overline{\mathbb{P}}_n^k \langle A \rangle := n^{-1} \sum \delta_{(\theta_k^i X, i/n)}$$

where the sum is taken over all the integers i such that i/n is in A and $(i + \alpha)/n$ is in A for any $-k < \alpha < k$. Then, the process $\overline{\mathbb{P}}_n^k \langle . \rangle$ has independent increments. The construction of section 3 may be carried over to obtain a LDP for the process $\overline{\mathbb{P}}_n^k \langle . \rangle$. Now, notice that if A is an interval $[a, b]$, then

$$\sup_B |\overline{\mathbb{P}}_n^k \langle A \rangle (B) - \mathbb{P}_n^k \langle A \rangle (B)| \leq 2k/n.$$

Hence, in a deterministic way, $\overline{\mathbb{P}}_n^k \langle . \rangle$ is a good approximation of $\mathbb{P}_n^k \langle . \rangle$. It enables us to carry over $\mathbb{P}_n^k \langle . \rangle$ the LDP for $\overline{\mathbb{P}}_n^k \langle . \rangle$, and then obtain a LDP for \mathbb{P}_n^k .

In fact, as mentioned in Barbe (1993c), we can obtain LDP for the sem of hypermixing Markov process in this way in using a remark of Dembo and Zajic (1993).

Now, to obtain the LDP for \mathbb{P}_n^∞ , we have to make $k \rightarrow \infty$ in the LDP on \mathbb{P}_n^k . The problem is that \mathbb{P}_n^∞ and \mathbb{P}_n^k are not defined on the same space. For this, we define a mapping ψ_k from $\mathcal{P}(\mathcal{X}^k \times [0, 1])$ into $\mathcal{P}(\mathcal{X}^\infty \times [0, 1])$,

$$\psi_k(\mu) := \delta_{\underline{0}} \times \mu$$

where $\delta_{\underline{0}}$ is the Dirac measure on the constant sequence $\underline{0}$. The mapping ψ_k is continuous for the weak topologies. Thus, $\psi_k(\mathbb{P}_n^k)$ obeys a LDP. Now, we go on the other way, from \mathbb{P}_n^∞ to \mathbb{P}_n^k , in defining a mapping T_k from \mathcal{X}^∞ into \mathcal{X}^k by $(T_k(x))_i := x_i$ if $-k \leq i \leq 0$ and $(T_k(x))_i := 0$ if $i < -k$. Then

$$\mathbb{P}_n^k = \int \delta_{(u, T_k x)} d\mathbb{P}_n^\infty =: \tau_k(\mathbb{P}_n^\infty).$$

The mapping $\tau_k(\cdot)$ is weakly continuous. Using the ideas of the contraction principle given in Theorems 4.1. and 4.2., we can check that with very high probability the sequence $(\tau_k(\mathbb{P}_n^\infty))_{k>0}$ is a Cauchy sequence. This enables us to make $k \rightarrow \infty$ and obtain a LDP for \mathbb{P}_n^∞ . In fact this approach is very roughly the same as the classical one using projective spaces (see Deuschel and Stroock (1989) and Dembo and Zeitouni (1993)). However, it leads to a LDP under a better topology.

Indeed, to see the interest of a LDP on \mathbb{P}_n^∞ , consider a linear process $Y_i := \sum_{j \geq 0} a_j X_{i-j}$. One can reasonably be interested in the trajectory of this linear process. But the trajectory is in one-to-one mapping with the s.e.m.

$$\Pi_n := n^{-1} \sum_{1 \leq i \leq n} \delta_{(Y_i, i/n)}.$$

Define the mapping π from \mathcal{X}^∞ into \mathcal{X} by $\pi(x) := \sum_{j \geq 0} a_j X_{i-j}$. We see that

$$\Pi_n = \int \delta_{(\pi x, u)} d\mathbb{P}_n^\infty(u, x) =: \Pi(\mathbb{P}_n^\infty).$$

Thanks to the norm $\|\cdot\|_a$ on \mathcal{X}^∞ , the mapping $\Pi(\cdot)$ is weakly continuous. Using the usual contraction principle, Barbe (1993c) turns the LDP for \mathbb{P}_n^∞ into a LDP for Π_n .

6. Conclusion.

We would like to conclude in mentioning that we passed over a major challenge. The contraction principles given in the Theorem 2.1. and in section 4 lead to rate functions $I \circ h^{-1}$. First, it is nice when $I(\cdot)$ is known explicitly. Then, the mapping $h(\cdot)$ can be rather complicated. To actually calculate $I \circ h^{-1}$ may be extremely difficult. It is beyond the scope of this paper to describe some methods, but that may be first a good deal of variational calculus in some rather complicated space (such as $\mathcal{P}(\mathcal{X}^\infty)$). Then, if one wants a numerical answer, that can be a good deal to obtain it with a computer.

Acknowledgements : All the work presented in this paper to which I contributed has been done between June 1992 and July 1993 when I was essentially unemployed. I would like to express my deep gratitude to David M. Mason and Marc Hallin who invited me respectively in the fall and spring semesters and provided me a regular research environment as well as financial support at the CNWS-University of Delaware and, on funds of the FNRS, at the Université Libre de Bruxelles.

References

- Akonom, J., Gouieroux, C. (1987) : A functional limit theorem for fractional process, Tech. Report, CEPREMAP, Paris.
- Azencott, R. (1980) : *Grandes Déviations et Applications*, in *Ecole d'été de Probabilités de Saint-Flour*, VIII-1978, pp.1-176, edited by P.L. Hennequin, Lecture Notes in Mathematics, 774, Springer, Berlin.
- Bahadur, R.R. (1967) : Rate of convergence of estimates and tests statistics, *Ann. Math. Statist.*, 38, 303-324.
- Bahadur, R.R. (1971) : *Some Limit Theorems in Statistics*, SIAM, Philadelphia.
- Bahadur, R.R., Zabell, S.L. (1979) : Large deviations of the sample mean in general vector spaces, *Ann. Probab.*, 7, 587-621.
- Barbe, Ph. (1993a) : Large deviation principle in linear models, preprint.

Barbe, Ph. (1993b) : Note on a weak form of the contraction principle, preprint.

Barbe, Ph. (1993c) : Large deviation principle for Banach-valued linear processes, preprint.

Barbe, Ph., Bertail, P. (1993) : The LR, Wald and Lagrange tests are not equivalent!, in preparation.

Barbe, Ph., Broniatowski, M. (1993a) : Deviation principle for set indexed processes with independent increments, preprint.

Barbe, Ph., Broniatowski, M. (1993b) : Large deviations principle for sequential rank processes, preprint.

Barbe, Ph., Broniatowski, M. (1993c) : Large deviations principle for fractional ARIMA processes, preprint.

Barbe, Ph., Hallin, M. (1993) : Bahadur efficiency for serial rank tests, in preparation.

Borovkov, A.A. (1962) : New limit theorem in boundary-value problem for sum of independent terms (in Russian), *Sibirsk. Math. Z.*, 3, 645-695 (English translation in *Selected Transl. Math. Statist. Probab.* (1965), 5, 315-372).

Bucklew, J.A. (1990) : *Large Deviation Techniques in Decision, Simulation and Estimation*, Wiley, New York.

Comets, F. (1989) : Large deviation estimates for a conditional probability distribution. Application to random interaction Gibbs measures, *Probab. Theor. Rel. Fields*, 80, 407-432.

Cottrell, M., Fort, J.C., Malgouyres, G. (1983) : Large deviations and rare events in the study of stochastic algorithms, *IEEE Trans. Automat. Control*, vol AC-28 (9), 907-920.

Csizár, I. (1984) : Sanov property, generalized I-projection and a conditional limit theorem, *Ann. Probab.*, 12, 768-793.

Csörgő, S. (1979) : Erdős-Rényi law, *Ann. Statist.*, 7, 772-787.

De Acosta, A. (1992) : On large deviations of the empirical measures in the τ -topology, preprint.

Dembo, A., Zajic, T. (1993) : Large deviations : from empirical mean and measure to partial sum processes, preprint.

Dembo, A., Zeitouni, O. (1993) : *Large Deviations Techniques and Applica-*

tions, Jones and Bartlett, Boston.

Deuschel, J.D., Stroock, D. (1989) : *Large Deviations*, Academic Press, Boston.

Donsker, M., Varadhan, S.R.S. (1976) : Asymptotic evaluation of certain Markov process expectations for large time, III, *Comm. Pure Appl. Math.*, 29, 389-461.

Ellis, R.S. (1985) : *Entropy, Large Deviations and Statistical Mechanics*, Springer, Berlin.

Erdős, P., Rényi, A. (1970) : On a new law of large numbers, *J. Analyse Math.*, 13, 103-111.

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

Groeneboom, P., Oosterhoff, J. (1977) : Bahadur efficiency and probability of large deviations, *Statist. Neerl.*, 31, 1-24.

Groeneboom, P. Schorack, G.R. (1981) : Large deviations of goodness of fit statistics and linear combinations of order statistics, *Ann. Probab.*, 9, 971-987.

Hoadley, A.B. (1976) : On the probability of large deviations of functional of several empirical cdf's, *Ann. Math. Statist.*, 38, 360-381.

Lynch, J., Sethuraman, J. (1987) : Large deviation for processes with independent increments, *Ann. Probab.*, 15, 610-627.

Major, P. (1992) : Asymptotic distributions for weighted U-statistics, preprint.

McBride, J. (1974) : *Functional Analogue of Iterated logarithm Type Laws for Empirical Distribution Function whose Arguments Tend to 0 at an Intermediate Rate*, Ph.D. Thesis, Chicago, Illinois.

Oldam, K.B., Spanier, K. (1974) : *Fractional Calculus*, Academic Press, New York.

O'Neil, K.A., Redner, R.A. (1993) : Asymptotic distributions of weighted U-statistics of degree 2, *Ann. Probab.*, 21, 1159-1169.

Pollard, D. (1990) : *Empirical Processes : Theory and Applications*, IMS, Hayward, CA.

Pragman, J. (1988) : Bahadur efficiency of rank tests for the change-point, *Ann. Statist.*, 16, 198-217.

Ross, B. (ed) (1974) : *Fractional Calculus and its Applications*, Lecture Notes in Mathematics, 457, Springer, Berlin.

- Sadowsky, J.S., Bucklew, J.A. (1990) : On large deviation theory and asymptotic efficient Monte Carlo estimation, *IEEE, Trans. Inf. Theor.*, vol. 36, 579-588.
- Sanov, I.N. (1957) : On the probability of large deviations of random variables (in Russian), *Math. Sb.*, 5b, 42, 11-14 (English transl. in *Selected Transl. Maht. Statist. Probab.*, (1961), 1, 213-344.
- Schilder, M. (1966) : Some asymptotic formula for the Wiener integrals, *Trans. Amer. Math. Soc.*, 125, 63-85.
- Stone, M. (1974) : Large deviations of the empirical probability measures, *Ann. Statist.*, 2, 362-366.
- Strassen, V. (1964) : An invariance principle for the law of the iterated logarithm, *Z. Wahrsch. verw. Geb.*, 3, 211-226.
- Stroock, D. (1984) *An introduction to the Theorie of Large Deviations*, Springer, Berlin.
- Varadhan, S.R.S. (1966) : Asymptotic probabilities and differential equations, *Comm. Pure Appl. Math.*, 19, 261-286.
- Varadhan, S.R.S. (1984) : *Large Deviations and Applications*, SIAM, Philadelphia.
- Woodworth, G.G. (1970) : Large deviations and Bahadur efficiency on linear rank statistics, *Ann. Math. Statist.*, 41, 251-283.