Math. Proc. Camb. Phil. Soc. (2000), **128**, 561 561
Printed in the United Kingdom

A large deviations heuristic made precise

BY NEIL O'CONNELL
*BRIMS, Hewlett–Packard Laboratories, Bristol BS*34 8*QZ.*
e-mail: noc@hplb.hpl.hp.com

(*Received* 13 *May* 1998; *revised* 21 *June* 1999)

Abstract

Sanov's Theorem states that the sequence of empirical measures associated with a sequence of i.d.d. random variables satisfies the large deviation principle (LDP) in the weak topology with rate function given by a relative entropy. We present a derivative which allows one to establish LDPs for symmetric functions of many i.d.d. random variables under the condition that (i) a law of large numbers holds whatever the underlying distribution and (ii) the functions are uniformly Lipschitz. The heuristic (of the title) is that the LDP follows from (i) provided the functions are 'sufficiently smooth'. As an application, we obtain large deviations results for the stochastic bin-packing problem.

1. *The heuristic*

We begin with some definitions. Let $\mathscr{X}$ be a Hausdorff topological space with Borel σ-algebra $\mathscr{B}$ and let μ_n be a sequence of probability measures on $(\mathscr{X}, \mathscr{B})$. A *rate function* is a non-negative lower semicontinuous function on $\mathscr{X}$. We say that the sequence μ_n satisfies the *large deviation principle* (LDP) with rate function I, if for all $B \in \mathscr{B}$,

$$-\inf_{x\in B^\circ} I(x) \leqslant \liminf_n \frac{1}{n}\log \mu_n(B) \leqslant \limsup_n \frac{1}{n}\log \mu_n(B) \leqslant -\inf_{x\in \overline{B}} I(x).$$

Here B° and $\overline{B}$ denote the interior and closure of B, respectively. Alternatively, if Z_n is a sequence of realizations of the μ_n, we can say that the sequence Z_n satisfies the LDP. A rate function is *good* is its level sets $\{x: I(x) \leqslant a\}$, $a \geqslant 0$, are compact.

Let $X_1, X_2, \ldots$ be a sequence of independent random variables taking values in a compact Polish space (E, d) with common law μ. Suppose that, for each n, $f_n: E^n \to \mathbb{R}$ is symmetric and measurable and that, for any underlying distribution μ, we have a strong law of large numbers:

$$f_n(X_1, \ldots, X_n) \xrightarrow{\text{a.s.}} f(\mu).$$

Then, provided the f_n are sufficiently smooth, we might hope to deduce the existence of a large deviation principle for the sequence $f_n(X_1, \ldots, X_n)$ with rate function given by

$$J(y) = \inf\{H(\nu|\mu): f(\nu) = y\}, \tag{1}$$

where $H(\cdot|\mu)$ is the relative entropy function:

$$H(\nu|\mu) = \begin{cases} \displaystyle\int_E d\nu \log \frac{d\nu}{d\mu} & \nu \ll \mu \\ \infty & \text{otherwise.} \end{cases}$$

This is a heuristic application of the (extended) contraction principle. It is useful because, put simply, laws of large numbers are easier to prove than large deviation principles. Some readers may find it reassuring to note that Cramér's theorem can be 'derived' from this heuristic, letting the f_ns be the empirical means of real variables.

For a more interesting example, consider the stochastic bin-packing problem. Here E is the unit interval and, for $x \in [0,1]^n$, $nf_n(x_1, \ldots, x_n)$ is the smallest number of unit-sized bins required to pack n objects of respective sizes $x_1, \ldots, x_n$. It is well-known, and easy to prove, that if $X_1, X_2, \ldots$ is a sequence of independent random variables taking values in $[0,1]$ with common law μ, then

$$f_n(X_1, \ldots, X_n) \xrightarrow{\text{a.s.}} c(\mu)$$

for some (finite) $c(\mu)$ called the 'packing constant'. According to the heuristic, the large deviation principle follows with rate function given by

$$J(y) = \inf\{H(\nu|\mu) : c(\nu) = y\}.$$

We shall see later that this statement is (almost) correct.

In this short note we present a sufficient condition on the f_n which is both simple to check and justifies the above heuristic. We make no attempt to prove the best possible result; the aim of this note is to give the reader an understanding of where the heuristic comes from and a feeling for the 'type' of condition under which we can expect it to hold.

2. *The main result*

Let (E, d) be a compact Polish space. Denote by $M_1(E)$ the space of Borel probability measures on E, endowed with the weak topology, and by $M_1^\mu(E)$ the subspace of probability measures which are absolutely continuous with respect to μ. Suppose that, for each n, $f_n : E^n \to \mathbb{R}$ is symmetric and Borel measurable and satisfies the Lipschitz condition

$$|f_n(x) - f_n(y)| \leqslant \frac{K}{n} \sum_{i=1}^{n} d(x_i, y_i) \tag{2}$$

for all $x, y \in E^n$, where K is a fixed constant (independent of n). We prove the following. (For $\nu \in M_1(E)$, $\nu^{\otimes n}$ denotes the n-fold product measure on E^n.)

THEOREM 1. *In the above context: if there exists a mapping $f : M_1^\mu(E) \to \mathbb{R}$ such that for each $\nu \in M_1^\mu(E)$,*

$$\lim_{n\to\infty} \int_{\mathbb{R}} x(\nu^{\otimes n} \circ f_n^{-1})(dx) = f(\nu),$$

then f is continuous (relative to the weak topology on $M_1^\mu(E)$) and the sequence $\mu^{\otimes n} \circ f_n^{-1}$ satisfies the LDP in $\mathbb{R}$ with good rate function given by

$$J(y) = \inf\{H(\nu|\mu) : f(\nu) = y\}.$$

Note that in the statement of the theorem we only need to assume mean convergence of the probability measures $\nu^{\otimes n} \circ f_n^{-1}$; under the smoothness condition (2), this is in fact equivalent to the almost sure convergence discussed in the previous section. We remark also that, under the hypotheses of the theorem, the function f is in fact ρ-Lipschitz, where ρ is the MKO metric defined by (3) below.

We will begin by presenting the main ingredients of the proof. In what follows, δ_x denotes the atomic measure with unit mass at the point x, supp μ denotes the support of a measure μ and $\xrightarrow{w}$ denotes weak convergence of probability measures.

Sanov's Theorem. *If X_n are independent random variables taking values in a Polish space (E, d), with common law μ, and we set*

$$L_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i},$$

then the sequence L_n satisfies the LDP *in $M_1(E)$ with good convex rate function $H(\cdot|\mu)$.*

The extended contraction principle. Let $\mathscr{X}$ be a Hausdorff topological space, equipped with its Borel σ-algebra, and let μ_n be a sequence of probability measures on $\mathscr{X}$. Let $\mathscr{Y}$ be another Hausdorff topological space. The usual contraction principle states that, if the sequence μ_n satisfies the LDP in $\mathscr{X}$ with good rate function $I: \mathscr{X} \to \mathbb{R}_+$ and $f: \mathscr{X} \to \mathscr{Y}$ is a continuous mapping, then the sequence $\mu_n \circ f^{-1}$ satisfies the LDP in $\mathscr{Y}$ with good rate function given by

$$J(y) = \inf \{I(x): f(x) = y\}.$$

The extended contraction principle applies to the case where we have, for each n, a mapping $f_n: \mathscr{X} \to \mathscr{Y}$ and wish to obtain an LDP for the sequence $\mu_n \circ f_n^{-1}$. There are a number of statements in the literature, dating back to the seminal paper of Varadhan [**15**] (see also [**2**] and [**12**]) which are roughly equivalent to the following. For completeness, we have included a short proof in the appendix.

Theorem 2. *Assume that $\mathscr{X}$ is a metric space. Suppose that for each n,* supp $\mu_n \subset \mathscr{X}_n \subset \mathscr{X}$*, $f_n: \mathscr{X}_n \to \mathscr{Y}$ is continuous and the sequence μ_n satisfies the* LDP *in $\mathscr{X}$ with good rate function I with effective domain contained in $\mathscr{X}_\infty \subset \mathscr{X}$. Suppose also that there exists a continuous mapping $f: \mathscr{X}_\infty \to \mathscr{Y}$ such that whenever $x_n \in \mathscr{X}_n$ and $x_n \to x \in \mathscr{X}_\infty$, we have $f_n(x_n) \to f(x)$. Then the sequence $\mu_n \circ f^{-1}$ satisfies the* LDP *in $\mathscr{Y}$ with good rate function given by*

$$J(y) = \inf \{I(x): f(x) = y\}.$$

Combining the extended contraction principle with Sanov's Theorem we obtain the following corollary. Suppose that, for each n, $f_n: E^n \to \mathscr{Y}$ is symmetric and Borel measurable and there exists a continuous mapping $f: M_1^\mu(E) \to \mathscr{Y}$ such that whenever, for $x \in E^\infty$,

$$\frac{1}{n} \sum_{i=1}^{n} \delta_{x_i} \xrightarrow{w} \nu \in M_1^\mu(E)$$

we have $f_n(x_1, \ldots, x_n) \to f(\nu)$. Then the sequence $\mu^{\otimes n} \circ f_n^{-1}$ satisfies the LDP in $\mathscr{Y}$ with good rate function given by

$$J(y) = \inf\{H(\nu|\mu): f(\nu) = y\}.$$

The Monge–Kantorovich–Ornstein (MKO) distance. Let (E, d) be a metric space. For $\pi \in M_1(E^2)$, denote by π_1 and π_2 the respective marginals of π in $M_1(E)$. The MKO distance between two probability measures $\mu, \nu \in M_1(E)$ is defined by

$$\rho(\mu, \nu) = \inf\left\{\int_{E^2} d(x, y)\pi(dx, dy): \pi \in M_1(E^2),\ \pi_1 = \mu,\ \pi_2 = \nu\right\}. \tag{3}$$

This measure of distance was first introduced in 1781 by Monge [**10**] in studying the most efficient way of transporting soil. It was later developed in a measure-theoretic context by Kantorovich [**7**] and Ornstein [**11**], among others (see [**13**] for an extensive survey). For our purposes it is sufficient to note that, if (E, d) is compact, then ρ metrises the weak topology on $M_1(E)$.

THE BIRKOFF–VON NEUMANN THEOREM.[1] *Recall that a non-negative matrix is* doubly stochastic *if each of its rows and columns sum to one. A* permutation matrix *of order n is a matrix of the form* $1\{\sigma(i) = j\}$*, for some* $\sigma \in S_n$*. The Birkoff–von Neumann theorem states that any doubly stochastic matrix can be written as a convex combination of permutation matrices (see, for example* [**1**, theorem 2·1·6]). *Using this remarkable fact, we immediately obtain the following formula for the* MKO *distance between two empirical measures of the same order. For* $x \in E^n$*, set*

$$l_n(x) = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}$$

and denote by S_n *the set of permutations on n elements.*

LEMMA 3. *For* $x, y \in E^n$,

$$\rho(l_n(x), l_n(y)) = \inf_{\sigma \in S_n} \frac{1}{n}\sum_{i=1}^{n} d(x_i, y_{\sigma(i)}).$$

Proof. For $\sigma \in S_n$, set

$$\pi^\sigma = \frac{1}{n}\sum_{i=1}^{n} \delta_{(x_i, y_{\sigma(i)})}.$$

We will argue that for any $\pi \in M_1(E^2)$, with $\pi_1 = l_n(x)$ and $\pi_2 = l_n(y)$, we have

$$\pi = \sum_{\sigma \in S_n} \beta_\sigma \pi^\sigma,$$

for some collection of non-negative numbers β_σ with

$$\sum_{\sigma \in S_n} \beta_\sigma = 1.$$

[1] The Birkoff–von Neumann theorem has an interesting history: the theorem was stated and proved in a different but equivalent formulation by König some 40 years before it was rediscovered by Birkoff (1946) and von Neumann (1953), after whom it was named. References and related history can be found in the preface of [**8**].

Then we would have, noting that any π of the form $\sum_{\sigma\in S_n}\beta_\sigma\pi^\sigma$ has the correct marginals $l_n(x)$ and $l_n(y)$,

$$\begin{aligned}\rho(l_n(x),l_n(y)) &= \inf\left\{\int_{E^2} d(x,y)\pi(dx,dy)\colon \pi\in M_1(E^2),\ \pi_1=l_n(x),\ \pi_2=l_n(y)\right\}\\ &= \inf\left\{\sum_{\sigma\in S_n}\beta_\sigma\int_{E^2} d(x,y)\pi^\sigma(dx,dy)\colon \beta_\sigma>0,\ \sum_{\sigma\in S_n}\beta_\sigma=1\right\}\\ &= \inf_{\sigma\in S_n}\int_{E^2} d(x,y)\pi^\sigma(dx,dy)\\ &= \inf_{\sigma\in S_n}\frac{1}{n}\sum_{i=1}^n d(x_i,y_{\sigma(i)})\end{aligned}$$

as required. To see that our claim is justified, note that if $\pi\in M_1(E^2)$, with $\pi_1=l_n(x)$ and $\pi_2=l_n(y)$, we can write

$$\pi=\frac{1}{n}\sum_{1\leqslant i,j\leqslant n} a_{ij}\delta_{(x_i,y_j)},$$

where $A=(a_{ij})$ is a doubly stochastic matrix. By the Birkoff–von Neumann Theorem, there exist non-negative constants β_σ with $\sum_{\sigma\in S_n}\beta_\sigma=1$, such that

$$a_{ij}=\sum_{\sigma\in S_n}\beta_\sigma\mathbf{1}_{\sigma(i)=j}$$

for all $1\leqslant i,j\leqslant n$. It follows that

$$\pi=\sum_{\sigma\in S_n}\beta_\sigma\pi^\sigma,$$

as claimed.

A concentration inequality. The final ingredient in the proof is the following elementary concentration inequality, which is an immediate consequence of the Azuma–Hoeffding inequality for martingale differences (see, for example [**9**]).

LEMMA 4. *Let (E,d) be a bounded metric space and $f_n\colon E^n\to\mathbb{R}$ a symmetric (Borel measurable) function which satisfies the Lipschitz condition*

$$|f_n(x)-f_n(y)|\leqslant\frac{K}{n}\sum_{i=1}^n d(x_i,y_i)$$

for all $x,y\in E^n$, where K is a constant independent of n. Let $X_1,\dots,X_n$ be independent random variables in E and write $X^n=(X_1,\dots,X_n)\in E^n$. Then:

$$P(|f_n(X^n)-\mathbb{E}f_n(X^n)|>t)\leqslant 2\exp(-At^2n),$$

where $A=1/2K^2\delta^2$ and $\delta=\sup_{x,y\in E^n} d(x,y)$.

Proof of Theorem 1. By hypothesis, we have that for all $x,y\in E^n$,

$$|f_n(x)-f_n(y)|\leqslant\frac{K}{n}\sum_{i=1}^n d(x_i,y_i).$$

Recall that for $x \in E^n$,

$$l_n(x) = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}.$$

By the symmetry of the f_n, and Lemma 3, it follows that

$$\begin{aligned}|f_n(x) - f_n(y)| &\leqslant \inf_{\sigma \in S_n} \frac{K}{n}\sum_{i=1}^{n} d(x_i, y_{\sigma(i)}) \\ &= K\rho(l_n(x), l_n(y)).\end{aligned}$$

If Y_i are i.d.d. with common law $\nu \in M_1^\mu(E)$ then, again by hypothesis, $Ef_n(Y^n) \to f(\nu)$. It follows, applying the concentration inequality (and Borel–Cantelli) that $f_n(Y^n) \xrightarrow{\text{a.s.}} f(\nu)$. We also have, by Sanov's Theorem (and Borel–Cantelli), that $l_n(Y^n) \xrightarrow{\text{a.s.}} \nu$ (in the weak topology). In particular, there exists a sequence $y \in E^\infty$ such that $f_n(y^n) \to f(\nu)$ and $l_n(y^n) \xrightarrow{w} \nu$. It follows that, for any $x \in E^\infty$ with $l_n(x^n) \xrightarrow{w} \nu$ we have, by the continuity of ρ,

$$|f_n(x^n) - f_n(y^n)| \leqslant K\rho(l_n(x^n), l_n(y^n)) \to 0,$$

and so $f_n(x^n) \to f(\nu)$, as required. In order to apply the extended contraction principle, it remains to check that f is weakly continuous. For $\mu, \nu \in M_1^\mu(E)$ we can find sequences $x, y \in E^\infty$ such that $l_n(x^n) \xrightarrow{w} \mu$ and $l_n(y^n) \xrightarrow{w} \nu$. From the above,

$$\begin{aligned}|f(\mu) - f(\nu)| &= \lim_{n\to\infty} |f_n(x^n) - f_n(y^n)| \\ &\leqslant \limsup_{n\to\infty} K\rho(l_n(x^n), l_n(y^n)) \\ &= K\rho(\mu, \nu),\end{aligned}$$

as required. Note that we have established more, namely that f is ρ-Lipschitz. This completes the proof of the theorem.

3. *Application to the stochastic bin-packing problem*

The standard reference on bin-packing is the book of Coffman and Lueker [**3**]. For $x \in [0,1]^n$, let $nf_n(x_1, \dots, x_n)$ be the smallest number of unit-sized bins required to pack n objects of respective sizes $x_1, \dots, x_n$. It is well-known, and easy to prove, that if $X_1, X_2, \dots$ is a sequence of independent random variables taking values in $[0,1]$ with common law μ, then

$$f_n(X_1, \dots, X_n) \xrightarrow{\text{a.s.}} c(\mu),$$

for some (finite) $c(\mu)$ called the 'packing constant'. Mean convergence, which is all we need to apply the theorem, follows from a simple subadditivity argument.

Concentration inequalities for this problem, using the Azuma–Hoeffding inequality, have been obtained by MacDiarmid [**9**]. Grimmett [**6**] obtains partial results on the existence of an LDP, using subadditivity arguments. Here we obtain, using Theorem 1, more refined large deviations results and a variational formula for the rate function.

Suppose for the moment that μ is supported on a finite subset E of $[0,1]$ and let $\epsilon > 0$ be the minimal separation distance between elements of E. Then the f_ns satisfy the Lipschitz condition (2) on E^n with $K = 1/\epsilon$ and we can apply the theorem to get that the sequence $f_n(X^n)$ satisfies the LDP in $[0,1]$ with good rate function given by

$$J(y) = \inf\{H(\nu|\mu): c(\nu) = y\}.$$

For general μ, we need to do some extra work. Let F denote the distribution function associated with μ and for each positive integer m, set

$$\mu_m^+ = \frac{1}{m}\sum_{j=1}^{m} \delta_{F^{-1}(j/m)}$$

and

$$\mu_m^- = \frac{1}{m}\sum_{j=0}^{m-1} \delta_{F^{-1}(j/m)}.$$

Then, as Coffman and Lueker observe,

$$c(\mu) - 1/m \leqslant c(\mu_m^-) \leqslant c(\mu) \leqslant c(\mu_m^+) \leqslant c(\mu) + 1/m.$$

We also have the related fact that $\mu^{\otimes n} \circ f_n^{-1}$ is majorized (respectively minorized) by $(\mu_m^+)^{\otimes n} \circ f_n^{-1}$ (respectively $(\mu_m^-)^{\otimes n} \circ f_n^{-1}$). Combining these observations, we obtain the following 'large deviation principle' (the following statement can be reformulated as an LDP with respect to a coarser topology on the unit interval) for the sequence $\mu^{\otimes n} \circ f_n^{-1}$: for $c(\mu) < q < 1$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu^{\otimes n}(f_n^{-1}[q,1]) \leqslant -\limsup_{m\to\infty} \inf\{H(\nu|\mu): c(\nu) \geqslant q - 1/m\}$$

and

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu^{\otimes n}(f_n^{-1}[q,1]) \geqslant -\liminf_{m\to\infty} \inf\{H(\nu|\mu): c(\nu) > q + 1/m\};$$

similar bounds hold for deviations to the left of $c(\mu)$. We give a step-by-step proof of the first inequality; the others can be obtained similarly. First note that, since $\mu^{\otimes n} \circ f_n^{-1}$ is majorized by $(\mu_m^+)^{\otimes n} \circ f_n^{-1}$, and the latter satisfies the LDP for any fixed m,

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n} \log \mu^{\otimes n}(f_n^{-1}[q,1]) &\leqslant \limsup_{n\to\infty} \frac{1}{n} \log (\mu_m^+)^{\otimes n}(f_n^{-1}[q,1]) \\ &\leqslant -\inf\{H(\nu|\mu_m^+): c(\nu) \geqslant q\} \\ &= -\inf\{H(\nu_m^+|\mu_m^+): c(\nu_m^+) \geqslant q\} \\ &= -\inf\{H(\nu|\mu): c(\nu_m^+) \geqslant q\} \\ &\leqslant -\inf\{H(\nu|\mu): c(\nu) + 1/m \geqslant q\}.\end{aligned}$$

The first *equality* follows from the fact that $H(\nu|\mu_m^+) = +\infty$ for $\nu \neq \nu_m^+$; the second follows from the observation, due to Pinsker (see, for example [**5**, lemma 6·5·16]), that $H(\nu|\mu) \geqslant H(\nu_m^+|\mu_m^+)$ for any ν and μ; the final inequality follows from the fact that $c(\nu_m^+) \leqslant c(\nu) + 1/m$.

Note that, if J is continuous and increasing (respectively decreasing) to the right (respectively to the left) of $c(\mu)$, the LDP holds with rate function J. To say more than that requires a careful analysis of the variational problem and is beyond the scope of this paper.

4. *Concluding remarks*

The main result of this paper can be extended in many directions. In the case of unbounded d, Sanov's Theorem can be extended to hold in the corresponding MKO topology under a moment condition such as

$$\mathbb{E}\exp\left[\delta d(X_1, x)\right] < \infty,$$

for some $\delta > 0$ and $x \in E$; in this case Theorem 1 can be extended by assuming $f_n(X^n)$ converges almost surely to $f(\mu)$ in the hypothesis. There are concentration inequalities for unbounded random variables which one can appeal to. Extension to functions taking values in a more general space is also possible.

An appealing feature of Theorem 1 is that we deduce the LDP under the type of condition which is usually associated with concentration inequalities. The particular Lipschitz condition we assume is, in a sense, the most naive in this class; it would be interesting to see if the LDP can be obtained under more sophisticated (milder!) conditions that have been developed in that area (see, for example [**14**]).

Finally, a word of caution. In many problems of combinatorial optimization, such as the travelling salesman and longest increasing subsequence problems, the functionals of interest are highly discontinuous and the heuristic discussed in this paper breaks down. This is because such functionals depend on much finer properties of the empirical measure than those which are asymptotically captured in the weak topology. There is a recent paper by Deuschel and Zeitouni [**5**] which beautifully illustrates this point for the longest increasing subsequence problem.

Appendix A. Proof of Theorem 2

Here we present a proof of Theorem 2.

Denote by $\mathbb{N}^*$ the extended natural numbers and equip $\mathbb{N}^*$ with the metric

$$h(n, m) = \left|\frac{1}{n} - \frac{1}{m}\right|,$$

with the convention that $1/\infty = 0$. Then (trivially) the sequence (μ_n, n) satisfies the LDP in $\mathscr{X} \otimes \mathbb{N}^*$, equipped with the product topology, with good rate function given by

$$I_e(x, n) = \begin{cases} I(x) & n = \infty \\ \infty & \text{otherwise.} \end{cases}$$

We can restrict this LDP to the (measurable) subspace

$$\bigcup_{n\in\mathbb{N}^*} \mathscr{X}_n \otimes \{n\},$$

as the effective domain of I_e lies in this subspace (see, for example [**4**, lemma 4·1·5]). The statement of the theorem now follows by applying the usual contraction principle to the mapping

$$F\colon \bigcup_{n\in\mathbb{N}^*} \mathscr{X}_n \otimes \{n\} \to \mathscr{Y},$$

defined by

$$F(x, n) = \begin{cases} f_n(x) & x \in \mathscr{X}_n,\ n < \infty \\ f(x) & n = \infty. \end{cases}$$

Note that we have used the fact that since $\mathscr{X}$ is a metric space, we can check continuity of F using sequences.

Acknowledgement. I would like to express my thanks to the anonymous referee for careful reading of the manuscript and many helpful suggestions.

REFERENCES

[**1**] R. B. Bapat and T. E. S. Raghavan. *Non-negative matrices and applications* (Cambridge University Press, 1997).

[**2**] N. R. Chaganty. Large deviations for joint distributions and statistical applications. Technical report TR93-2, Department of Mathematics and Statistics, Old Dominion University, Norfolk, VA (1993).

[**3**] E. G. Coffman Jr. and George S. Lueker. *Probabilistic analysis of packing and partitioning algorithms* (Wiley, 1991).

[**4**] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications* (Jones and Bartlett, 1992).

[**5**] J.-D. Deuschel and O. Zeitouni. On increasing subsequences of i.i.d. samples. *Combinatorics, Probability and Computing*, to appear.

[**6**] G. R. Grimmett. Large deviations in subadditive processes and first-passage percolation. *Contemporary Mathematics* **41** (1985), 175–194.

[**7**] L. V. Kantorovich. On the transfer of masses. *Dokl. Akad. Nauk. SSSR* **37** (7–8) (1942), 227–229.

[**8**] L. Lovász and M. D. Plummer. *Matching theory* (Elsevier, 1986).

[**9**] C. J. H. MacDiarmid. On the method of bounded differences. In: *Surveys in Combinatorics*, London Mathematical Society Lecture Notes **141**, 148–188 (Cambridge University Press, 1989).

[**10**] G. Monge. Mémoire sur la théorie des déblais et des remblais. Histoire de l'Académie des sciences de Paris, avec les Mémoires de mathématique et de physique pour la même année (1781), 666–704.

[**11**] D. Ornstein. An application of ergodic theory to probability theory. *Ann. Prob.* **1** (1) (1973), 43–65.

[**12**] A. Puhalskii. Large deviation analysis of the single server queue. *Queueing Systems* **21** (1995), 5–66.

[**13**] S. T. Rachev. The Monge–Kantorovich mass transference problem and its stochastic applications. *Theor. Probab. Appl.* **29** (4) (1984), 647–676.

[**14**] M. Talagrand. A new look at independence. *Ann. Prob.* **24** (1) (1995), 1–34.

[**15**] S. R. S. Varadhan. Asymptotic probabilities and differential equations. *Comm. Pure Appl. Math.* **19** (1966), 261–286.