OPTIMAL MATCHING AND EMPIRICAL MEASURES

J. E. YUKICH

(Communicated by William D. Sudderth)

ABSTRACT. Using results from optimal matching, we find the exact order of convergence for $\beta(P_n, P)$ and $\rho(P_n, P)$, where β denotes the dual bounded Lipschitz metric, ρ the Prokhorov metric and P_n the n th empirical measure associated to P , the uniform measure on the unit square. The results solve a long-open problem in empirical measures.

1. INTRODUCTION

This article describes some interesting connections between certain optimal matching problems and Glivenko-Cantelli convergence of empirical measures. More precisely, it is observed that average edge length matching and minimax grid matching problems have close connections to empirical processes. Classical duality results from linear programming as well as the deep recent results of Shor [Sh1] and Leighton and Shor [LS] play an important connecting role.

It turns out that tight bounds for these matching problems are essentially equivalent to sharp rates for Glivenko-Cantelli convergence. In this way, some long-open problems in empirical processes are solved in a relatively simple fashion. Also, some new estimates for matching problems are found in the non-i.i.d. case.

Notation. Throughout, $\mathcal{P}(\mathbf{R}^d)$ denotes the set of all Borel probability measures on $\mathbf{R}^d$, $d \geq 1$. Weak-star convergence in $\mathcal{P}(\mathbf{R}^d)$ is metrizable by various metrics; the following definition recalls two of them.

Definition 1.1.

- (i) For all $P, Q \in \mathcal{P}(\mathbf{R}^d)$ define the dual bounded Lipschitz metric

$$\beta(P, Q) := \sup_{f \in BL(1)} \left| \int f(dP - dQ) \right|,$$

where $BL(1) := \{f : \mathbf{R}^d \rightarrow \mathbf{R} \text{ such that } \|f\|_\infty \leq 1 \text{ and } |f(x) - f(y)| \leq \|x - y\|\}$.

Received by the editors September 19, 1988 and, in revised form December 20, 1988.

1980 *Mathematics Subject Classification.* Primary 60B10, 05C70; Secondary 60D05, 68C05.

Key words and phrases. Average edge length, minimax matching length, empirical measures, Prokhorov metric, dual bounded Lipschitz metric.

(ii) Define the Prokhorov metric

$$\rho(P, Q) := \inf\{\varepsilon > 0 : P(A) \leq Q(A^\varepsilon) + \varepsilon \quad \text{for any Borel set } A\},$$

where A^ε is the ε -neighborhood of A , that is

$$A^\varepsilon := \{y \in \mathbf{R}^d : \|y - z\| < \varepsilon \quad \text{for some } z \in A\}.$$

The metric β was apparently first used by Fortet and Mourier [FM]; see Dudley [Du1] for general relations between β and ρ .

Definition 1.2. Let $\mathcal{F}$ be a class of functions on $\mathbf{R}^d$, $d \geq 1$, and let $P \in \mathcal{P}(\mathbf{R}^d)$. The $\mathcal{L}^1(P)$ bracketing number for $\mathcal{F}$ is defined for all $\varepsilon > 0$ as

$$N_{[\cdot]}(\varepsilon, \mathcal{F}, P) := \inf\{n : \exists f_1, \dots, f_n \text{ measurable such that for all } f \in \mathcal{F} \text{ there are } i, j \leq n \text{ with } f_i \leq f \leq f_j \text{ pointwise and } \int |f_j - f_i| dP \leq \varepsilon\}.$$

$\log N_{[\cdot]}(\varepsilon, \mathcal{F}, P)$ is called metric entropy with bracketing and was first introduced by Dudley [Du3].

Definition 1.3. Say that $f(x) = O(g(x))$ iff $\exists C < \infty$ such that $\lim_{x \rightarrow \infty} f(x)/g(x) \leq C$; $f(x) = \Omega(g(x))$ iff $\exists C > 0$ such that $\lim_{x \rightarrow \infty} f(x)/g(x) \geq C$.

Throughout, c and C denote finite, strictly positive constants; their values may change from line to line.

2. AVERAGE EDGE LENGTH AND THE DUAL BOUNDED LIPSCHITZ METRIC β

Given two sets of points $X_1, \dots, X_n$ and $Y_1, \dots, Y_n$ match the X points to the Y points so as to minimize the sum of the edge lengths (equivalently, minimize the average edge length). Thus, find

$$T_n := \min_{\pi} \sum_{i=1}^n d(X_{\pi(i)}, Y_i),$$

where d denotes distance and the min is taken over all permutations π of the integers $1, \dots, n$.

If P denotes the uniform distribution on $[0, 1]^2$ and if the X and Y points are chosen at random, independently of each other with distribution P , then there exist constants c and C such that with probability $1 - o(1)$

$$(2.1) \quad c(n \log n)^{1/2} < T_n < C(n \log n)^{1/2}.$$

This result is originally due to Ajtai, Komlós and Tusnády [AKT], where T_n denotes transportation cost. Karp [Ka] had apparently proved $T_n = O(n^{1/2} \log n)$ at an earlier date. Later, Shor showed that the lower bound in (2.1) holds with probability $1 - 2^{-n^\varepsilon}$ for any $\varepsilon < 1$ [Sh1, Theorem 2.1.3] and therefore

$$\liminf \frac{T_n}{(n \log n)^{1/2}} \geq c \quad \text{a.s.}$$

Finding the minimum average edge length T_n/n is an example of the bipartite matching problem: given a complete bipartite graph G with weights (e.g. distances) associated to the edges, find a minimum perfect matching. Bipartite matching has a dual problem, with the maximum solution to the dual problem equal to the minimum weight matching. The dual problem is just the linear programming dual of bipartite matching [PS], [Ke].

The dual relationship involving T_n shows that

$$(2.2) \quad T_n := T_n(X, Y) = \sup_{f \in L(1)} \sum_{i=1}^n (f(X_i) - f(Y_i)),$$

where $L(1)$ denotes the set of Lipschitz(1) functions f defined on $[0, 1]^2$, i.e.

$$|f(x) - f(y)| \leq \|x - y\| \quad \forall x, y \in [0, 1]^2.$$

This is just a special case of the Kantorovich-Rubinstein theorem applied to the uniform discrete measures with support on $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$, respectively (see e.g. [Fe]). Without loss of generality $L(1)$ can be replaced by $BL(1)$ in (2.2).

It is important to observe that the stochastic convergence rate for T_n is, modulo constant factors, identical to that for

$$D_n := D_n(X) := \sup_{f \in BL(1)} \left| \sum_{i=1}^n (f(X_i) - Ef(X_i)) \right|, \quad X_i \text{ i.i.d. } P.$$

To see this, observe that $T_n(X, Y) \leq D_n(X) + D_n(Y)$ shows one direction and by now classical symmetrization techniques show the other (see e.g. Lemma 2.5 of [GZ]).

The quantity D_n/n is just the supremum of the empirical process indexed by $BL(1)$ functions. In other words, since the X_i , $i \geq 1$, are i.i.d. P ,

$$D_n/n = \sup_{f \in BL(1)} \left| \int f(dP_n - dP) \right|,$$

where P_n denotes the usual n th empirical measure, i.e. $P_n(\omega) = n^{-1}(\delta_{X_1(\omega)} + \dots + \delta_{X_n(\omega)})$, where δ_x denotes the unit mass at x . For other connections between matching problems and empirical processes, see Rhee and Talagrand [RT].

By definition of β , $\beta(P_n, P) = D_n/n$ and (2.1) consequently implies:

Theorem 2.1. *There exist constants c and C such that*

$$\lim_{n \rightarrow \infty} Pr \{c \leq (n/\log n)^{1/2} \beta(P_n, P) \leq C\} = 1.$$

Modulo only constant factors, this gives the exact rate of convergence for $\beta(P_n, P)$, improving upon the existing estimates of Dudley [Du2], Zuker [Zu], Massart [Ma] and Bakhvalov [Ba]. Whether there is a constant C such that $\lim_{n \rightarrow \infty} (n/\log n)^{1/2} \beta(P_n, P) = C$ a.s. remains an open question.

It should be noted that an a.s. higher dimensional analog of Theorem 2.1 is already known: if P denotes the uniform measure on the unit cube $[0, 1]^d$, $d \geq 3$, then there are constants $c := c(d)$ and $C := C(d)$ such that

$$0 < c \leq \liminf n^{1/d} \beta(P_n, P) \leq \overline{\lim} n^{1/d} \beta(P_n, P) \leq C < \infty \quad \text{a.s.}$$

The lower bound always holds and is implied by Bakhvalov's results [Ba]; for a direct proof, see Dudley [Du4, Theorem 1]. The upper bound follows from standard entropy estimates [KT] for the Lipschitz(1) functions $f: [0, 1]^d \rightarrow \mathbb{R}$ and the a.s. bounds of Alexander [Al, Corollary 2.5 of Correction].

Average edge length in the nonuniform and non-i.i.d. case. Conversely, it is worth noting that rates for $\beta(P_n, P)$ yield rates for T_n ; this is so in the nonuniform and non-i.i.d. case. This is of interest in optimization theory. For example, if the X and Y points are i.i.d. P , P any law on $[0, 1]^d$, $d \geq 3$, then

$$\overline{\lim} n^{-1+1/d} T_n \leq \overline{\lim} 2n^{1/d} \beta(P_n, P) \leq C \quad \text{a.s.},$$

which again follows from entropy estimates and the a.s. bounds of [Al]. In fact, the estimate

$$ET_n \leq 2En\beta(P_n, P) \leq C(n^{1-1/d})$$

holds when the X and Y points are weakly dependent [Ga], which is intuitively expected.

These remarks also apply to the non-i.i.d. case. For example, if $\mathcal{L}(X_i) = \mathcal{L}(Y_i) = P_{(i)} \in \mathcal{P}([0, 1]^2)$, $i \geq 1$, then the inequality

$$T_n := T_n(X, Y) \leq D_n(X) + D_n(Y)$$

implies that upper bounds on convergence rates for D_n yield the same upper bound for T_n . Combining the entropy bound [KT]

$$\log N_{[\cdot]} \left(\varepsilon, BL(1), n^{-1} \sum_{i=1}^n P_{(i)} \right) = O(\varepsilon^{-2})$$

with Corollary 2.5 of [Al] shows that in the non-i.i.d. case

$$\overline{\lim} \frac{T_n}{n^{1/2} \log n} \leq C \quad \text{a.s.}$$

The $\log n$ term could probably be sharpened to $\log^{1/2} n$; however, this remains open.

As a final extension, suppose that the X and Y points have values in an arbitrary compact metric space S . Bounds on the transportation cost follow from rates for $\beta(P_n, P)$, where $\mathcal{L}(X) = \mathcal{L}(Y) = P \in \mathcal{P}(S)$; see e.g. [Du2].

3. MINIMAX GRID MATCHING AND THE PROKHOROV METRIC ρ

Letting n be a perfect square, consider the unit square in the plane that contains the set G of n grid points arranged in a regularly spaced $n^{-1/2} \times n^{-1/2}$

array and n random points, $X_i(\omega)$, $1 \leq i \leq n$, located independently and randomly according to the uniform distribution P on the square. From now on, $P_{(n)}$ will denote the collection of random points $\{X_1(\omega), \dots, X_n(\omega)\}$.

Given $P_{(n)}$, let $L(P_{(n)}, G)$ denote the minimum length such that there exists a perfect matching of the (random) points in $P_{(n)}$ to the grid points G for which the distance between every pair of matched points is at most $L(P_{(n)}, G)$. In other words, $L(P_{(n)}, G)$ is the minimum over all perfect matchings of the maximum distance between any pair of matched points; $L(P_{(n)}, G)$ is thus called the *minimax matching length* for $P_{(n)}$ [LS]. In their remarkable paper, Leighton and Shor [LS] show that

$$(3.1) \quad c \log^{3/4} n \leq n^{1/2} L(P_{(n)}, G) \leq C \log^{3/4} n;$$

moreover (3.1) holds with “very high probability”, i.e., with probability exceeding $1 - n^{-\alpha}$, where $\alpha = \Omega(\log^{1/2} n)$. The somewhat unexpected result (3.1) is based upon a dual discrepancy result (the discrepancy of a region R of the square is the absolute value of the difference between the expected number of points contained in R and the actual number of points in R). Let Γ be a partition of $[0, 1]^2$ into $n/\log^{3/2} n$ subsquares, each with side length $\log^{3/4} n/n^{1/2}$. Leighton and Shor [LS] show that every simply connected region whose boundary lies along the edges of Γ has discrepancy at most $O(p \log^{3/4} n/n^{1/2})$, where p is the perimeter of the region.

Considering this dual discrepancy result it is not surprising that the Prokhorov distance $\rho(P_n, P)$ between P and the empirical measure P_n is related to $L(P_{(n)}, G)$. In fact, it will be shown that the upper bound in (3.1) implies the upper bound of the following result.

Theorem 3.1. *There exist constants c and C such that*

$$(3.2) \quad \lim_{n \rightarrow \infty} Pr \left\{ c \leq \frac{n^{1/2}}{\log^{3/4} n} \rho(P_n, P) \leq C \right\} = 1.$$

The bounds in (3.2) thus resolve the long-open problem of finding the exact order of convergence of $\rho(P_n, P)$. This problem was originally raised by Dudley [Du2] and since then there has been substantial progress, notably by Zuker [Zu] and Massart [Ma]. Massart actually determines the exact order of convergence of $\rho(P_n, P)$ up to a power of $\log n$; (3.2) supplies the missing power.

Proof. To prove the upper bound in (3.2) it suffices to show it for n a perfect square, since for $0 \leq j \leq 2n^{1/2} + 1$, it is easily verified that $\rho(P_{n+j}, P) \leq \rho(P_{n+j}, P_n) + \rho(P_n, P) = O(n^{-1/2}) + \rho(P_n, P) = O(\log^{3/4}(n+j)/(n+j)^{1/2}) + \rho(P_n, P)$. Thus, with n a perfect square, let G_n be the probability measure which places mass n^{-1} at each of the n grid points. By definition of ρ , it is clear that $\rho(P_n, G_n) \leq L(P_{(n)}, G)$. Moreover, $\rho(G_n, P) = O(n^{-1/2})$, since if A contains exactly k of the n grid points in G , then $k/n = G_n(A) \leq P(A^{n^{-1/2}})$. Since ρ satisfies the triangle inequality, the upper bound in (3.2) is established.

To establish the lower bound in (3.2), we first need a lemma involving the discrepancy of lower layers in the unit square. Recall that a subset B of $[0, 1]^2$ is called a lower layer if it is closed and it contains (x, y) whenever it contains some point (x', y') with $x' \geq x$, $y' \geq y$.

Lemma 3.2. *With probability $1 - o(1)$ there exists a lower layer A such that*

$$(3.3) \quad P_n(A) - P(A) \geq c \log^{3/4} n / n^{1/2}.$$

Proof. This estimate is part of the main theorem of Rhee and Talagrand [RT], who refer to an earlier result of Shor [Sh2, Theorem 3 and its proof] involving the dual problem of upright matching. (Leighton and Shor [LS] also state [without proof] that Shor's theorem implies (3.3) with very high probability.) Shor's result is nonetheless one about expectations and so (3.3) is not completely obvious. To see that (3.3) does hold, make the following additions to the proof of Shor's Theorem 3.

Given n , recall [Sh2] that there are $\Theta(\log n)$ stages involved in constructing the desired lower layer A . At the i th stage, there are 2^i "central quadrilaterals" which are candidates for inclusion in the lower region. Each quadrilateral has area $1/(6 \cdot 4^i \log^{1/2} n)$; by appropriate choice of the constant implicit in $i = O(\log n)$, we may assume that $n/4^i \log^{1/2} n \rightarrow \infty$ as $n \rightarrow \infty$. For the j th such quadrilateral Q_{ij} , $1 \leq j \leq 2^i$, define the corresponding random variable D_{ij} to be the number of $+$ points minus the number of $-$ points in Q_{ij} (take D_{ij} as zero if this difference is negative). Notice that $D_{ij} \leq |2K_{ij} - m_{ij}|$, where K_{ij} is a binomial random variable with parameters m_{ij} and $1/2$, $0 \leq m_{ij} \leq n$. The parameter m_{ij} is random and has expected value equal to $n \cdot (\text{area of } Q_{ij})$, the expected number of points in Q_{ij} . D_{ij} represents the discrepancy for Q_{ij} and $\sum_{j \leq 2^i} D_{ij}$ the total added discrepancy resulting from stage i . Note that the D_{ij} are independent.

Following Shor [Sh2, p. 192], there exists $c > 0$ such that for n large and all i, j , $1 \leq j \leq 2^i$,

$$ED_{ij} \geq cn^{1/2} / (2^i \log^{1/4} n).$$

Since $ED_{ij}^2 = O(n/4^i \log^{1/2} n)$, independence implies the upper bound

$$\text{Var} \sum_{j \leq 2^i} D_{ij} = O(n/2^i \log^{1/2} n)$$

for the variance of the sum. For $i = \Theta(\log n)$, Chebyshev's inequality implies that the lower bound

$$(3.4) \quad \sum_{j \leq 2^i} D_{ij} \geq \frac{c}{2} (n^{1/2} / \log^{1/4} n)$$

holds with probability at least $1 - n^{-\alpha}$, where α depends only on the constant implicit in the lower estimate $i = \Omega(\log n)$.

If $\mathcal{J}_n$ denotes the collection of those indices i for which (3.4) holds, then $\text{card } \mathcal{J}_n = \Omega(\log n)$. It follows that with probability $1 - o(1)$

$$\sum_{i \in \mathcal{J}_n} \sum_{j \leq 2^i} D_{ij} = \Omega(n^{1/2} \log^{3/4} n)$$

and therefore (3.3) holds provided that the boundary of the lower region has a slope between -1 and $+1$. This last condition may be satisfied without affecting the analysis [Sh2]. Q.E.D.

Now return to the proof of Theorem 3.1. Since

$$\sup_{B \in \mathcal{L}\mathcal{L}'} |P(B) - G_n(B)| = O(n^{-1/2}),$$

where $\mathcal{L}\mathcal{L}$ denotes the class of all lower layers, Lemma 3.2 implies that with probability $1 - o(1)$

$$(3.5) \quad P_n(A) - G_n(A) \geq c \log^{3/4} n / n^{1/2}.$$

By the defining property of a lower layer it is clear that for all $\varepsilon > 0$

$$G_n(A^\varepsilon) - G_n(A) \leq 2(\varepsilon + n^{-1/2}),$$

i.e.,

$$(3.6) \quad G_n(A) - G_n(A^\varepsilon) \geq -2(\varepsilon + n^{-1/2}).$$

Combining (3.5) and (3.6) yields

$$P_n(A) - G_n(A^\varepsilon) \geq c \log^{3/4} n / n^{1/2} - 2(\varepsilon + n^{-1/2}).$$

Setting $\varepsilon := \rho := \rho(P_n, G_n)$ and using the definition of ρ it follows that with probability $1 - o(1)$

$$c \log^{3/4} n / n^{1/2} - 2(\rho + n^{-1/2}) \leq P_n(A) - G_n(A^\rho) \leq \rho,$$

i.e., $3\rho \geq c \log^{3/4} n / n^{1/2} - 2n^{-1/2}$. Thus, $\rho(P_n, G_n) \geq c \log^{3/4} n / n^{1/2}$ and since $\rho(P, G_n) = O(n^{-1/2})$, the triangle inequality implies that the first inequality in (3.2) holds. Q.E.D.

Notice that the above proof actually shows $\overline{\lim}(n^{1/2} / \log^{3/4} n) \rho(P_n, P) \leq C$ a.s.; whether $\lim_{n \rightarrow \infty} (n^{1/2} / \log^{3/4} n) \rho(P_n, P) = C$ a.s. remains an open question, however. Also, concerning the possible extensions of Theorem 3.1 to general probability measures P , see Dudley [Du2]. Among other things he showed that for Cantor-type measures P on $[0, 1]^d$, rates for $\rho(P_n, P)$ can be about as slow as $n^{-1/(k+2)}$, where $k = d \log 2 / \log 3$. Later, Zuker [Zu, §4] showed that these rates can actually be about as slow as $n^{-1/k}$, $k < d$.

Extensions to higher dimension. Let P denote the uniform measure on the unit cube $[0, 1]^d$, $d \geq 3$, and $\rho^{(d)}(P_n, P)$ the Prokhorov distance between P and

P_n . If G denotes a $n^{-1/d} \times \cdots \times n^{-1/d}$ grid on $[0, 1]^d$ and G_n the probability measure which places mass n^{-1} at each of the n grid points, then

$$(3.7) \quad \rho^{(d)}(G_n, P) = O(n^{-1/d}).$$

Massart [Ma] has shown that there is a positive constant $c := c(d)$ such that

$$\liminf \frac{n^{1/d} \rho^{(d)}(P_n, P)}{\log^{1/d} n} \geq c \quad \text{a.s.}$$

and thus (3.7) and the inequality

$$|\rho^{(d)}(P_n, P) - \rho^{(d)}(P, G_n)| \leq \rho^{(d)}(P_n, G_n) \leq L(P_n, G)$$

imply the a.s. lower bound $L(P_n, G) = \Omega((\log n/n)^{1/d})$; (cf. [LS]).

Actually it is possible to show that this lower bound is tight. In a related article [SY] it is shown that there is a finite constant $C := C(d)$ such that the d -dimensional minimax grid matching length $L(P_n, G)$ satisfies ($d \geq 3$)

$$(3.8) \quad \overline{\lim} L(P_n, G)(n/\log n)^{1/d} \leq C \quad \text{a.s.}$$

While (3.8) is of interest in multi-dimensional bin packing and allocation problems [LS and KLM], it shows that the a.s. exact order rate of convergence for $\rho^{(d)}(P_n, P)$ is $(\log n/n)^{1/d}$, $d \geq 3$.

ACKNOWLEDGMENT

This paper originates in conversations with Michel Talagrand, who pointed out to the author the articles [Sh1] and [LS] and their connections to empirical processes. In particular, Talagrand was well aware that the determination of the rate of Glivenko-Cantelli convergence was a simple consequence of the Leighton-Shor result; he generously left it to the author to supply the small details and write the present article. The author also thanks P. Shor for several enlightening discussions and a referee for valuable comments leading to the present improved and corrected version.

REFERENCES

- [AKT] M. Ajtai, J. Komlós and G. Tusnády, *On Optimal Matchings*, *Combinatorica* **4** (1984), 259-264.
- [Al] K. S. Alexander, *Probability inequalities for empirical processes and a law of the iterated logarithm*, *Ann. Prob.* **12** (1984), 1041-1067; Correction **15** (1987), 428-430.
- [Ba] N. S. Bakhvalov, *On approximate calculation of multiple integrals* (in Russian), *Vestnik Mosk. Ser. Mat. Mekh. Astron. Fiz. Khim* **4** (1959), 3-18.
- [Du1] R. M. Dudley, *Distances of probability measures and random variables*, *Ann. Math. Stat.* **39** (1968), 1563-1572.
- [Du2] ———, *The speed of mean Glivenko-Cantelli convergence*, *Ann. Math. Stat.* **40** (1969), 40-50.
- [Du3] ———, *Central limit theorems for empirical measures*, *Ann. Prob.* **6** (1978), 899-929; Correction **7**(1979), 909-911.

- [Du4] ———, *Empirical and Poisson processes on classes of sets or functions too large for central limit theorems*, Z. Wahrsch. verw. Geb. **61** (1982), 355-368.
- [Fe] X. Fernique, *Sur le théorème de Kantorovitch-Rubinstein dans les espaces polonais*, Lecture Notes in Math. **850** (1981), 6-10.
- [FM] R. Fortet and E. Mourier, *Convergence de la répartition empirique vers la répartition théorique*, Ann. Scient. École Norm. Sup. **70** (1953), 266-285.
- [Ga] P. Gaenssler, *A note on a result of Dudley on the speed of Glivenko-Cantelli convergence*, Ann. Math. Stat. **41** (1970), 1339-1343.
- [GZ] E. Giné and J. Zinn, *Some limit theorems for empirical processes*, Ann. Prob. **12** (1984), 929-989.
- [Ka] R.M. Karp, Private communication cited in [AKT] (1982).
- [Ke] J. H. B. Kemperman, *On the role of duality in the theory of moments*, Lecture notes in Economic and Mathematical Systems **215** (1983), 63-92, Springer, Berlin.
- [KLM] R. M. Karp, M. Luby and A. Marchetti-Spaccamela, *Probabilistic analysis of multidimensional bin packing problems*, Proc. 16th ACM Symp. on Theory of Computing (1984), 289-298.
- [KT] A. N. Kolmogorov and V. M. Tikhomirov, *ϵ -entropy and ϵ -capacity of sets in function spaces*, Am. Math. Soc. Transl. (Ser. 2.) **17** (1961), 277-364; = *Uspekhi Mat. Nauk.* 14, vyp.2(86) (1959), 3-86.
- [LS] T. Leighton, and P. W. Shor, *Tight bounds for minimax grid matching, with applications to the average case analysis of algorithms*, Proc. 18th ACM Symp. on Theory of Computing (1986), 91-103.
- [Ma] P. Massart, *About the Prokhorov distance between the uniform distribution over the unit cube in $\mathbf{R}^d$ and its empirical measure*, Prob. Theory and Related Fields **79**(3) (1988), 431-450.
- [PS] C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization, Algorithms and Complexity*, Prentice Hall, NJ, 1982.
- [RT] W. S. Rhee and M. Talagrand, *Exact bounds for the stochastic upward matching problem*, Trans. Amer. Math. Soc. **307** (1988), 109-125.
- [Sh1] P. W. Shor, *Random planar matching and bin packing*, Ph.D. thesis, M.I.T. (1985).
- [Sh2] ———, *The average-case analysis of some on-line algorithms for bin packing*, Combinatorica **6** (1986), 179-200.
- [SY] P. W. Shor and J. E. Yukich, *Minimax grid matching and empirical measures*, preprint.
- [Zu] M. Zuker, *Speeds of convergence of random probability measures*, Ph.D. thesis, M.I.T. (1974).