C. R. Acad. Sci. Paris, Ser. I 356 (2018) 692–695

Contents lists available at ScienceDirect

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com

Probability theory

Scaling and non-standard matching theorems

Mise à l'échelle et théorèmes d'appariement non-standard

Michel Talagrand

23, rue Louis-Pouey, 92800 Puteaux, France

ARTICLE INFO

Article history:
Received 11 April 2018
Accepted 13 April 2018

Presented by Gilles Pisier

ABSTRACT

Consider the standard Gaussian measure μ on $\mathbb{R}^2$. Consider independent r.v.s $(X_i)_{i\le N}$ distributed according to μ, and an independent copy $(Y_i)_{i\le N}$ of these r.v.s. We prove that, for some number C and N large, we have

$$\frac{(\log N)^2}{C} \le \mathrm{E} \inf_\pi \sum_{i\le N} d(X_i, Y_{\pi(i)})^2 \le C(\log N)^2 , \tag{1}$$

where the infimum is over all permutations π of $\{1,\dots,N\}$. The striking point of this result is the factor $(\log N)^2$. Indeed, if instead of μ we consider the uniform distribution on the unit square, it is well known that the proper factor is $\log N$. The upper bound was proved by Michel Ledoux (2017) [3].

RÉSUMÉ

Considérons une suite indépendente $(X_i)_{i\le N}$ de variables aléatoires distribuées comme la mesure gaussienne canonique μ sur $\mathbb{R}^2$ et une copie independente $(Y_i)_{i\le N}$ de cette même suite. Pour une certaine constante universelle C et $N \ge 2$, nous avons les inégalités

$$\frac{(\log N)^2}{C} \le \mathrm{E} \inf_\pi \sum_{i\le N} d(X_i, Y_{\pi(i)})^2 \le C(\log N)^2 \tag{1}$$

où l'infimum est pris sur toutes les permutations π de $\{1,\dots,N\}$. La borne supérieure a été prouvée par Michel Ledoux (2017) [3], qui conjecturait que l'inégalité (1) était correcte avec un facteur $\log N$ et non pas $(\log N)^2$. C'est précisement l'apparence de ce facteur $(\log N)^2$ qui est non standard.

E-mail address: michel.talagrand@gmail.com.

https://doi.org/10.1016/j.crma.2018.04.018

1. Introduction

Consider independent r.v.s $(X_i)_{i\le N}$ uniformly distributed on the unit square of $\mathbb{R}^2$ and an independent copy $(Y_i)_{i\le N}$ of these variables. It is well known that, for each $p>1$, one has

$$\frac{N^{1-p/2}(\log N)^{p/2}}{C_p}\le \mathsf{E}\inf_\pi\sum_{i\le N} d(X_i,Y_{\pi(i)})^p\le C_pN^{1-p/2}(\log N)^{p/2} \tag{2}$$

for some number C_p depending on p only. An important idea here is that the lower bound follows from Hölder's inequality and the case $p=1$, and that this lower bound also holds if the distribution of the X_i is not too different from uniform, say its density is constant within a multiplicative factor 2.

One may like, of course, to investigate what happens when the uniform distribution is replaced by an unbounded distribution μ. In the case $p=1$, this was done, in particular, by J. Yukich [5].

In this note, we bring forward a completely elementary scaling property that does not seem to have been previously noticed. Since this property is not specific to dimension 2, we will explain it in its proper setting where it is much clearer. Consider $k\ge 3$ and assume now that the r.v.s X_i and Y_i are uniform over $[0,1]^k$. Then one should replace (1) by the (much easier) inequality

$$\frac{N^{1-p/k}}{C_{k,p}}\le \mathsf{E}\inf_\pi\sum_{i\le N} d(X_i,Y_{\pi(i)})^p\le C_{k,p}N^{1-p/k}\,, \tag{3}$$

where $C_{k,p}$ is independent of N. It is the case $p=k$, which gives rise to interesting scaling effects. To explain the heuristic, we assume now that the X_i have a common distribution μ. Let us assume that in a certain box A of side a, the probability μ is nearly uniform, in a sense that its density ρ with respect to Lebesgue's measure varies on that square by at most a factor 2, say $b\le\rho\le 2b$. Assuming $n=Nba^k\ge N^{1/100}$, there are about n points X_i that belong to A. The typical distance between a point X_i and the closest point Y_j is about $a/n^{1/k}$. We then expect that, for any permutation π, we have

$$\sum_{X_i\in A} d(X_i,Y_{\pi(i)})^k\ge \frac{1}{C_k}n\times a^k/n=\frac{a^k}{C_k}\,. \tag{4}$$

The fundamental fact here is that this quantity is independent of b, so that in a sense the points in A do not overall contribute like $\mu(A)$, but rather like $\lambda_k(A)$, the k-dimensional volume of A. For this reason, we should heuristically have

$$\sum_{i\le N} d(X_i,Y_{\pi(i)})^k\ge\frac{U}{C_k} \tag{5}$$

where U is the area of the union of the squares A as described above. In the case where μ is the standard Gaussian measure on $\mathbb{R}^k$, it turns out that the union of such squares contains a sphere of radius $\sqrt{\log N}/C$, which is volume $(\log N)^{k/2}/C$, and we have completed the heuristic proof of the following.

Proposition 1.1. *Assume $k\ge 3$. If the sequence $(X_i)_{i\le N}$ is independently distributed according to the standard Gaussian measure μ and $(Y_i)_{i\le N}$ is an independent copy of this sequence, then*

$$\mathsf{E}\inf_\pi\sum_{i\le N} d(X_i,Y_{\pi(i)})^k\ge\frac{(\log N)^{k/2}}{C_k}\,. \tag{6}$$

It is extremely easy to turn the heuristic argument into a rigorous proof. There is every reason to believe that the bound (6) can be reversed, and we outline below a variation of the standard "transportation method" that should prove that, but we have not performed the computations to check that it works.

The situation is more subtle when $k=2$. There is another type of more global fluctuations in a random sample. These fluctuations create the extra factor $\log N$ in (2). These fluctuations are also present inside each box A, and created an extra factor $\log n\simeq\log N$. The lower bound in (5) has to be replaced by $(U\log N)/C$, and this is how one reaches the lower bound in (1).

In the next section, we make the previous ideas precise in order to prove the lower bound in (1). The upper bound in (1) was proved by Michel Ledoux [3] (by adapting the methods of [2]), who introduced the problem. We will give a completely elementary proof of this upper bound, using only (2).

It should be noted that our proofs depend heavily on the fact that the tails of the Gaussian distribution decrease fast. More precisely, calling $\rho(x)$ the density of this distribution with respect to Lebesgue's measure, the regions where $\rho(x)\ge 1/N$ and $\rho(x)\ge N^{-1/100}$ have areas of the same order. It seems certain that our result can be extended to the case of a distribution μ with density proportional to $\exp(-d(0,x)^\alpha)$ with respect to Lebesgue's measure (where $\alpha>0$) (replacing of course the factor $(\log N)^2$ by $(\log N)^{1+2/\alpha}$). But what happens is the case where μ has a density proportional to $(1+d(0,x))^{-\alpha}$ is far less clear.

2. Lower bounds

From now on, μ denotes the standard Gaussian measure on $\mathbb{R}^2$ and X_i and Y_i are as in the introduction. The first lemma goes back to the paper [1]. The reader may also find the argument in Section 3.6 of [4].

Lemma 2.1. *Consider a square A of side a in $\mathbb{R}^2$. Assume that $N\mu(A) \geq N^{1/100}$ and that the density $\rho(x) = (2\pi)^{-1}\exp(-d(0,x)^2/2)$ of μ with respect to Lebesgue's measure varies by at most a factor 2 over A. Then with probability close to 1, there exists a Lipschitz function f on $\mathbb{R}^2$ which is zero outside A and is such*

$$\left|\sum_{i\leq N} f(X_i) - \sum_{j\leq N} f(Y_j)\right| \geq \frac{1}{C} a\sqrt{N\mu(A)}(\log N)^{1/2}\,. \tag{7}$$

Here a is just a scaling factor, and A contains about $N\mu(A)$ points X_i and Y_j. The condition $N\mu(A) \geq N^{1/100}$ ensures that $\log(N\mu(A))$ is of order $\log N$.

Let us then explain how to prove the lower bound in (1). We say that a square B in $\mathbb{R}^2$ is k-dyadic if its side has length 2^{-k} and its vertices have coordinates that are integers multiple of 2^{-k}. Considering constants $C_1, \dots$ large enough, we cover the disc centered at the origin of diameter $\sqrt{\log N}/C_1$ by disjoint k-dyadic squares where k is such that $2^{-k} \simeq 1/(C_2\sqrt{\log N})$. Then ρ varies by a factor at most 2 on each such square. Let $\mathcal{B}$ be this family of squares. Given such a square B, we denote by A_B the square with the same center but one-forth of the area. For N large (which we assume in the rest of the proof) for each square $B \in \mathcal{B}$, the square A_B satisfies $\mu(A_B)N \geq N^{1/100}$. What happens in different squares are basically independent events, so with probability close to 1 the subset $\mathcal{B}'$ of $\mathcal{B}$ consisting of the squares B for which the square A_B satisfies the conclusion of Lemma 2.1 (the existence of a Lipschitz function f as in (7)) has a cardinality at least 1/2 of the cardinality of $\mathcal{B}$.

Assuming that this is the case, given a permutation π of $\{1,\dots,N\}$ we bound from below the quantity

$$D = \sum_{i\leq N} d(X_i, Y_{\pi(i)})^2\,. \tag{8}$$

Consider $B \in \mathcal{B}'$, and $I_B = \{i \leq N\,;\, X_i \in A\}$. Consider then the Lipschitz function $f = f_B$ as in Lemma 2.1 applied to the square A_B. Then using the Cauchy–Schwarz inequality,

$$\left|\sum_{i\in I_B}(f(X_i) - f(Y_{\pi(i)}))\right|^2 \leq \operatorname{card} I_B \sum_{i\in I_B} d(X_i, Y_{\pi(i)})^2\,. \tag{9}$$

The plan is now to find a lower bound on the left-hand side. First, $\sum_{i\in I_B} f(X_i) = \sum_{i\leq N} f(X_i)$ since f is zero outside A_B. Next,

$$\sum_{i\in I_B} f(Y_{\pi(i)}) = \sum_{j\leq N} f(Y_j) - \sum_{i\notin I_B} f(Y_{\pi(i)})\,.$$

Consequently,

$$\left|\sum_{i\in I_B}(f(X_i) - f(Y_{\pi(i)}))\right| \geq \left|\sum_{i\leq N} f(X_i) - \sum_{j\leq N} f(Y_j)\right| - \sum_{i\notin I_B} |f(Y_{\pi(i)})|$$

The first term on the right is $\geq 2^{-k}(N\mu(A_B))^{1/2}\sqrt{\log N}/C$. Assuming $D \leq (\log N)^3$ (for otherwise there is nothing to prove), the second term is of much lower order because, when $f(Y_{\pi(i)}) \neq 0$, then $Y_{\pi(i)} \in A_B$, so that if, moreover, $i \notin I_B$, then $X_i \notin B$, and then $d(X_i, Y_{\pi(i)}) \geq 2^{-k-2}$. The number of terms in the sum is then polylogarithmic in N, and each term is $\leq 2^{-k}$. Consequently,

$$\left|\sum_{i\in I_B}(f(X_i) - f(Y_{\pi(i)}))\right|^2 \geq \frac{1}{C} 2^{-2k} N\mu(A_B)\log N\,.$$

Since $\operatorname{card} I_B$ if of order $N\mu(B)$, going back to (9) gives $\sum_{i\in I_B} d(X_i, Y_{\pi(i)})^2 \geq 2^{-2k}(\log N)/C$, exactly as in the heuristic. Summation over $B \in B'_{\mathcal{B}}$ concludes the proof since 2^{-2k} is just the area of B.

3. Upper bounds

Even though the upper bound is proved by Michel Ledoux [3], it is of interest to provide a much more direct argument. Let us first explain one of the basic idea of our approach. It $(Z_i)_{i\le N}$ are uniformly distributed on $[0,1]^2$, a typical realization of this sequence is such that one can find a permutation π of $\{1,\dots,N\}$ such that $|Z_{\pi(i)} - i/N| \le C\sqrt{N}$. Consequently, considering independent r.v.s $X_i = (X_i^1, X_i^2)$ that are uniformly distributed over $[0,1]$, for the type of problem we can replace this sequence by the sequence $(i/N, X_i^2)$. Assuming now $N = 2k$ for simplicity, we can also replace the sequence X_i by the sequence $T_i = (U_i, X_i^2)$, where $(U_i)_{i\le k}$ is independent uniformly distributed over $[0,1/2]$ and $(U_i)_{k<i\le N}$ is independent uniformly distributed over $[1/2,1]$. In fact, there is no reason to split the interval $[0,1]$ in just two, we may split it in q pieces for $q \le N$. The transportation cost incurred by this transformation is bounded. Each vertical strip of the type $[a,b]\times[0,1]$ can then be split in many pieces, by the same procedure, but the cost of the procedure is of order 1 for *each* such strip. The method will be used with a logarithmic number of strips.

We define $r_0 = 0$ and for $k \ge 1$ we set $r_k = \sqrt{k}$. For $k \ge 1$, consider the set D_k of points whose distance to the origin is between r_{k-1} and r_k, so that for $k \ge 2$, D_k is an annulus of width about $1/\sqrt{k}$. The sequence r_k is chosen so that the density ρ is "constant" on each D_k. We fix N. For each k, we consider an integer N_k with $|N_k - N\mu(D_k)| \le 2$, insuring that $N_k = 0$ for $k \ge 10\log N$. The first step of the proof is to show that, instead of the sequence $(X_i)_{i\le N}$, we can consider a sequence made as follows. For each k we have an independent sequence $(X_k)_{i\le N_k}$ that is distributed in D_k according to the probability μ_k given by $\mu_k(C) = \mu_k(C\cap D_k)/\mu(D_k)$. The points $(X_i)_{i\le N}$ are then replaced by the points $(X_i^k)_{k\ge 1, i\le N_k}$. This is done by working in polar coordinates, in which case $X_i = (r_i,\theta_i)$ where (r_i) and (θ_i) are independent of each other, and using the procedure described above, together with the following tedious one-dimensional result.

Lemma 3.1. *Consider on $\mathbb{R}^+$ the probability measure ν of density $x\exp(-x^2/2)$. Consider an integer N and for $1 \le k \le N$ define a_k by $\nu([a_k,\infty)) = k/N$. Consider independent r.v.s $(Z_i)_{i\le n}$ distributed according to ν. Then $\mathsf{E}\inf_\pi \sum_{i\le N}(a_i - Z_{\pi(i)})^2 \le C\log N$ where as usual the infimum is over all permutations.*

We then consider independent copies $(Y_i^k)_{i\le N_k}$ of the X_i^k. The matching will be constructed by matching the points X_i^k with the points Y_i^k. The procedure differs according to whether $N_k \ge \sqrt{k}r_k$ (which occurs for the small values of k) or whether $N_k \le \sqrt{k}r_k$ (which occurs for the larger values of k since the sequence N_k decreases). Let us first examine the second case. Then the width of the annulus D_k is smaller than the typical distance between a point X_i^k and the closest point Y_i^k (which is about r_k/N_k). We are then facing a trivial one-dimensional problem. The cost of the matching in this annulus will be of order $r_k^2 \le C\log N$ because $N_k = 0$ for the larger values of k.

Let us then examine the case of the smaller values of k, where $N_k \ge \sqrt{k}r_k$. Then we decompose the annulus into $n_k \simeq \sqrt{k}r_k$ equal sectors, each of which looks roughly like a square, and on which ρ is nearly constant. Playing now with the coordinate θ, we can then replace the the family X_i^k by n_k different families $X_i^{k,\ell}$, each of which is independent and distributed according to the conditional probability that it belongs to the sector. The cost of doing this is again at most $Cr_k^2 \le C\log N$ for each annulus. The points $X_i^{k,\ell}$ are then matched to the points $Y_i^{k,\ell}$. Using that the probability we consider in each sector is a Lipschitz image of a probability on a square of a comparable area, according to (2) the cost of the matching in one given sector is bounded by $C\log N$ times the area of the sector, and the sum of the areas of these sectors is about $\log N$.

References

[1] M. Ajtai, J. Komlós, G. Tusnády, On optimal matchings, Combinatorica 4 (4) (1984) 259–264.
[2] L. Ambrosio, F. Stra, D. Trevisan, A PDE approach to a 2-dimensional matching problem, Probab. Theory Relat. Fields (2016), in press.
[3] M. Ledoux, On optimal matching of Gaussian samples, Zap. Nauč. Semin. POMI 457 (2017), Veroyatnost' i Statistika 25 226–264.
[4] M. Talagrand, Upper and Lower Bounds for Stochastic Processes, new edition in preparation, available at http://michel.talagrand.net/ULB.pdf.
[5] J. Yukich, Some generalizations of the Euclidean two-sample matching problem, in: Probability in Banach Spaces, 8, in: Progress in Probability, vol. 30, Birkhäuser, 1992, pp. 55–66.