ISRAEL JOURNAL OF MATHEMATICS 111 (1999), 275–284

CONCENTRATION AND INFLUENCES

BY

MICHEL TALAGRAND

Equipe d'Analyse, Université Paris VI
Boîte 186, 4, place Jussieu, 75252 Paris Cedex 5, France
e-mail: mit@ccr.jussieu.fr

ABSTRACT

Consider the discrete cube $\Omega = \{0,1\}^N$, provided with the uniform probability P. We denote by $d(x, A)$ the Hamming distance of a point x of Ω and a subset A of Ω. We define the influence $I(A)$ of the ith coordinate on A as follows. For x in Ω, consider the point $T_i(x)$ obtained by changing the value of the ith coordinate. Then

$$I_i(A) = P(\{x \in A; T_i(x) \notin A\}).$$

We prove that we always have

$$P(A) \int_\Omega d(x, A) dP(x) \le \frac{1}{2} \sum_{i \le N} I_i(A).$$

Since it is easy to see that $\sum_{i \le N} I_i(A)^2 \le \frac{1}{4}$, this recovers the well known fact that $\int_\Omega d(x, A) dP(x)$ is at most of order $\sqrt{N}$ when $P(A) \ge 1/2$. The new information is that $\int_\Omega d(x, A) dP(x)$ can be of order $\sqrt{N}$ only if A reassembles the Hamming ball $\{x; \sum_{i \le N} x_i \ge N/2\}$.

1. Introduction

The theory of concentration of measure in the product of probability spaces (an introduction to which can be found in [T2]) expresses in various ways that a generic point of a product is always "close" to a subset of probability $\ge 1/2$. A large part of the success of this theory in its applications stems from the fact that no assumption is made on the factors. On the other hand, the crucial case always seems to be the case of the two point space $\{0, 1\}$, provided with the probability that gives mass p to 1 $(0 < p < 1)$. Denoting by P the product probability on

Received February 10, 1997

$\{0,1\}^N$, the "extremal case" for several key results seems to be the Hamming ball

$$A = \Big\{(x_i)_{i\le N}; \sum_{i\le N} x_i \ge pN\Big\}.$$

It is thereby natural to expect that the concentration phenomenon improves when A is very different from a Hamming ball. Let us define the influence $I_i(A)$ of the ith coordinate upon a subset A of $\Omega = \{0,1\}^N$ as

$$I_i(A) = P(\{x \in A; T_i(x) \notin A\}),$$

where $T_i(x)$ is the point obtained from x by changing the ith coordinate. We recall that the Hamming distance is given by

$$d(x,y) = \sum_{i\le N} |x_i - y_i| = \operatorname{card}\{i \le N; x_i \ne y_i\}.$$

THEOREM 1: *For each subset A of $\{0,1\}^N$, we have*

$$\int_\Omega d(x,A)dP(x) \le \frac{p^*}{P(A)} \sum_{i\le N} I_i(A) \tag{1.1}$$

where $p^* = \max(p, 1-p)$.

Let us say that a function f on Ω is monotone if it is increasing for the natural (product) order on Ω.

Let us recall that the **transportation cost** $T(\nu,P)$ of a probability measure ν to P is given as

$$\inf \int_{\Omega^2} d(x,y)d\theta(x,y)$$

where the infimum is taken over all probability measures θ on Ω^2 of marginals ν, P.

Let us define the function r_i by $r_i(x) = x_i - p$.

THEOREM 2: *If ν is a probability on Ω such that $f = d\nu/dP$ is monotone, then*

$$T(\nu,P) = \sum_{i\le N} \int_\Omega r_i d\nu.$$

When the set A is monotone, considering $f = P(A)^{-1}1_A, \quad \nu = fP$, and observing that in this case

$$\int_\Omega d(x,A)dP(x) \le \int_{\Omega^2} d(x,y)d\theta(x,y)$$

whenever θ has P as a first marginal and a second marginal supported by A, we recover Theorem 1 in that case (with a better constant), since $\int_A r_i dP = (1-p)I_i(A)$. We will now discuss Theorem 1 in the case of monotone sets and (for simplicity) when $p = \frac{1}{2}$. In that case we have $I_i(A) = 2\int_A r_i dP$, and it is easy to show [T2, Proposition 2.2] that

$$\sum_{i \le N} I_i(A)^2 \le eP(A)^2 \log \frac{e}{P(A)}.$$

Thus (1.1) implies, using Cauchy-Schwarz, that

$$\int_\Omega d(x,A)dP(x) \le \sqrt{Ne}\sqrt{\log \frac{e}{P(A)}} \tag{1.2}$$

which is the optimal order as shown by the case $A = \{x; \sum_{i \le N} x_i \ge k\}$. But of course (1.1) is far superior to (1.2) because the bound does not depend upon N, and also because the influences can be small. In particular Kahn, Kalai and Linial [K-K-L] constructed an example with $P(A) = 1/2$ and where each number $I_i(A)$ is of order $\log N/N$. (They also proved the deep fact that this is the smallest possible order.) In this case, the bound provided by (1.1) is of order $\log N$ only. (On the other hand, in that case the left-hand side is of order 1.)

The weakness of Theorem 1 is that it controls only the first moment of $d(.,A)$, while (1.2) can be reinforced into an exponential inequality, e.g. [T1]

$$P(d(.,A) \ge t) \le \frac{1}{P(A)} \exp\Big(-\frac{t^2}{N}\Big). \tag{1.3}$$

In order to bound higher moments of $d(.,A)$, we can interpolate between (1.2) and (1.3). To give a specific example, assuming $P(A) \ge 1/2$ for simplicity we then have, for each $\Delta > 0$,

$$\begin{aligned}\int d(x,A)^2 dP(x) &= 2\int_0^\infty tP(d(.,A) \ge t)dt \\ &\le 2\int_0^\Delta \Delta P(d(.,A) \ge t)dt + \int_\Delta^\infty 4t \exp\Big(-\frac{t^2}{N}\Big)dt \\ &\le 2\Delta\Big(\sum_{i \le N} I_i(A)\Big) + 2N\exp\Big(-\frac{\Delta^2}{N}\Big)\end{aligned}$$

so that, by a suitable choice of Δ,

$$\int d(x,A)^2 dP(x) \le 4\sqrt{N} \sum_{i \le N} I_i(A) \sqrt{\log \frac{\sqrt{N}}{\sum_{i \le N} I_i(A)}}.$$

We will show that such an inequality appears optimal within logarithmic terms. In particular, it is unfortunately not true that

$$\int d(x,A)^2 dP(x) \le K\Big(\sum_{i\le N} I_i(A)\Big)^2. \tag{1.4}$$

This is shown in Example 3.3 below. The optimal bound that one can find for the left-hand side of (1.4) at the level of logarithmic terms is better left for future research. Another direction worthy of investigation would be to prove that the concentration inequalites involving the "convex distance" of [T2] improve subject to the condition $\sum I_i(A)^2 \ll 1$.

2. Proofs

We prove Theorem 1 by induction upon N and we perform the induction step from N to $N+1$. The quantities related to $N+1$ variables will be denoted with a $'$ to distinguish them from quantities related to N variables. For x in $\Omega = \{0,1\}^N$, we denote by $x^\frown j$ the sequence obtained by adding a last coordinate equal to j. Given a subset A of $\Omega' = \{0,1\}^{N+1}$ we write, for $j = 0,1$,

$$A_j = \{x \in \Omega^N;\ x^\frown j \in A\}$$

and we observe that

$$a = P'(A) = pa_1 + (1-p)a_0 \tag{2.1}$$

where $a_j = P(A_j)$. Also,

$$\forall j \le N, \quad I_j(A) = pI_j(A_1) + (1-p)I_j(A_0) \tag{2.2}$$

and

$$\begin{aligned} I_{N+1}(A) &= pP(A_1 \setminus A_0) + (1-p)P(A_0 \setminus A_1) \\ &\ge q|a_1 - a_0| \end{aligned} \tag{2.3}$$

where $q = \min(p, 1-p)$. Using the trivial bounds

$$d'(x^\frown j, A) \le d(x, A_j)$$
$$d'(x^\frown j, A) \le 1 + d(x, A_{j'}) \qquad \forall j, j' \in \{0,1\}$$

we see that

$$\int d'(y,A')dP'(y) \le \min\Big(p\int d(x,A_1)dP(x) + (1-p)\int d(x,A_0)dP(x);$$
$$1-p+\int d(x,A_1)dP(x); p+\int d(x,A_0)dP(x)\Big)$$

To simplify notation, we set $J_j = \sum_{i\le N} I_i(A_j)$; $J = pJ_1 + (1-p)J_0$. Thus, using (2.2), (2.3), it suffices to show that

$$\min\Big(p\frac{J_1}{a_1} + (1-p)\frac{J_0}{a_0}; \frac{1-p}{p^*} + \frac{J_1}{a_1}; \frac{p}{p^*} + \frac{J_0}{a_0}\Big) \le \frac{J + q|a_1 - a_0|}{a}. \tag{2.4}$$

We observe that

$$p\frac{J_1}{a_1} + (1-p)\frac{J_0}{a_0} - \frac{J}{a} = \frac{p(1-p)(a_0-a_1)(a_0J_1 - a_1J_0)}{aa_0a_1}.$$

Without loss of generality, we can assume $a_0 \ge a_1$. Thus (2.4) holds if

$$p(1-p)(a_0J_1 - a_1J_0) \le qa_0a_1.$$

On the other hand, if

$$p(1-p)(a_0J_1 - a_1J_0) > qa_0a_1$$

then

$$\begin{aligned}\frac{p}{p^*} + \frac{J_0}{a_0} - \frac{J}{a} &= \frac{p}{p^*} - \frac{p(a_1J_0 - a_0J_0)}{aa_0}\\ &\le \frac{p}{p^*} - \frac{qa_1}{(1-p)a}\\ &\le \frac{q}{1-p}\Big(1 - \frac{a_1}{a}\Big) = \frac{q(a_0-a_1)}{a}. \quad \blacksquare\end{aligned}$$

We now prove Theorem 2. In the case $N = 1$, it is well known (and very easy here) that

$$\begin{aligned}T(\nu,P) = \frac{1}{2}\int|f-1|dP &= \frac{1}{2}(p(f(1)-1) + (1-p)(1-f(0)))\\ &= p(1-p)(f(1)-f(0))\\ &= \int r_1 d\nu\end{aligned} \tag{2.5}$$

where we have used the fact that $pf(1) + (1-p)f(0) = 1$. To get a lower bound for $T(\nu, p)$, we use that for any Lipshitz function g for the Hamming distance,

$$\int g d\nu - \int g dP \le T(\nu, P)$$

for the function $g = \sum_{i \le N} r_i$. To prove the upper bound, we proceed by induction upon N. To perform the induction step from N to $N+1$, we consider the projection ν_1 of ν upon the first N coordinates, and set $f_1 = d\nu_1/dP$. For x in $\{0,1\}^N$, consider ν_x on $\{0,1\}$ given by

$$\nu_x(\{1\}) = \frac{pf(x^\frown 1)}{pf(x^\frown 1) + (1-p)f(x^\frown 0)}.$$

We then use the inequality (that is intuitively obvious, and is the basis of many uses of the transportation method [M])

$$T(\nu, P') \le T(\nu_1, P) + \int T(\nu_x, \mu) d\nu_1(x)$$

(where P' is the measure on $\{0,1\}^{N+1}$ and P on $\{0,1\}^N$). This makes the induction obvious.

3. Three examples and a method

In this section we assume $p = 1/2$.

Example 3.1: Consider $a > 0$ (small) and

$$\nu = (1-a)P + a\delta_{\mathbf{1}}$$

where $\mathbf{1}$ denotes the sequence with ones only. Then

$$T(\nu, P) = Na/2 \tag{3.1}$$

since $\int r_i d\nu = ar_i(\mathbf{1}) = \frac{a}{2}$ for each i. Consider $A = \{x; \sum_{i \le N} x_i \ge 3N/4\}$ and $\alpha = P(A)$, that is very small. If $a = 2\alpha$, at least half of the mass of ν at $\mathbf{1}$ must go outside A when one transports ν to P; so it must travel a distance $\ge N/4$. Thus if θ is any measure on Ω^2 of marginals P, ν, we must have

$$\int d(x,y)^2 d\theta(x,y) \ge \left(\frac{N}{4}\right)^2 \frac{a}{2}$$

(the formal argument is left to the reader) and this is much bigger than $(Na/2)^2$.

Example 3.2: We consider the sets

$$A = \{x \in \{0,1\}^N; x_1 = 1 \text{ or } x_1 = 0 \text{ and } \sum_{2 \le i \le N} x_i \ge 3N/4\},$$

$$A_0 = \{x \in A; x_1 = 0\}.$$

If we apply the scheme of Theorem 2 to the uniform probability on A, we see that the resulting transport first spreads the mass on A_0 uniformly on the set of points x with first coordinate zero, creating the same situation as in Example 1 (a mass of order $|A_0|$ is transported a distance at least $N/3$), that is, the probability θ on Ω^2 corresponding to this transport is such that $\int d(x,y)^2 d\theta(x,y)$ is much bigger than $(\int d(x,y)d\theta(x,y))^2$. There are, however, transports for which both these quantities are of optimal order. One such transport can be obtained through the scheme of Theorem 2 by reversing the order of the coordinates. What this shows is that if we try to use the scheme of Theorem 2 to control the second moment of $d(.,A)$ in an efficient way, we must the choose very carefully the **order** of the coordinates (if this is at all possible) and we do not know how to do this.

THE CANONICAL TRANSPORT. Since the previous example shows the relevence of the order of the coordinates, one would like to define a transportation method that is more intrinsic. There is a seemingly canonical method to do that. The intuitive idea is to think of $f(x)$ as the level of a fluid, that will flow toward equilibrium. The rules governing the fluid are as follows. The fluid at x can flow only to a node y where $d(x,y) = 1$ and $f_t(y) < f_t(x)$, and the rate at which it flows is proportional to $f_t(x) - f_t(y)$ (where $f_t(x)$ denotes the level of fluid at time t). Moreover, the new fluid reaching a given node instantly mixes with the fluid already there. Letting the fluid reach equilibrium defines (by tracking the path of the fluid) a transport from ν to P. More formally, this transport can be defined as follows. Consider the probability measure μ_t on $\{0,1\}$ that gives mass $\frac{1}{2}(1-e^{-t})$ to 1, and its product P_t on $\Omega = \{0,1\}^N$. Define $f_t = P_t * f$ (the canonical semi-group). Let us say that a matrix $(a_t(x,y))_{x,y\in\Omega}$ transports fP to f_tP and satisfies the following:

$$\forall x, y \in \Omega, \quad a_t(x,y) \geq 0, \tag{3.2}$$

$$\forall x \in \Omega, \quad \sum_{y\in\Omega} a_t(x,y) = 1, \tag{3.3}$$

$$\forall y \in \Omega, \quad \sum_{x\in\Omega} a_t(x,y) f(x) = f_t(y). \tag{3.4}$$

We define

$$J(y) = \{i \leq N, y_i = 1\}$$

and we recall that $T_i(y)$ is obtained from y by changing the ith coordinate of y.

We consider the system of differential equations

$$\tag{3.5}\begin{aligned}\frac{d}{dt}a_t(x,y) = &-\frac{1}{2}\sum_{i\in J(y)}\frac{f_t(y)-f_t(T_i(y))}{f_t(y)}a_t(x,y)\\ &+\frac{1}{2}\sum_{i\notin J(y)}\frac{f_t(T_i(y))-f_t(y)}{f_t(T_i(y))}a_t(x,T_i(y)),\end{aligned}$$

(with the convention $\frac{0}{0}=0$). If we use the initial condition $a_0(x,y)=\delta_{xy}$, then the system of solutions to (3.5) satisfies (3.2) to (3.4). To prove (3.3) and (3.4), we simply differentiate and use (3.5). To prove (3.2), we use in an essential way that f is monotone, so that, by (3.5),

$$\frac{d}{dt}\,a_t(x,y)\geq -\frac{N}{2}|a_t(x,y)|,$$

from which (3.2) follows (the details are left to the reader). Writing $a(x,y)=\lim_{t\to\infty}a_t(x,y)$, we then have $a(x,y)\geq 0$,

$$\forall x\in\Omega,\quad \sum_{y\in\Omega}a(x,y)=1,$$

$$\forall y\in\Omega,\quad \sum_{x\in\Omega}a(x,y)f(x)=1,$$

so that if we define $\theta(\{x,y\})=2^{-N}f(x)a(x,y)$ the probability θ has ν and P as marginals.

Let us write $h_i(x,y)=1$ if $x_i\neq y_i$, and $h_i(x,y)=0$ if $x=y$. Thus $d(x,y)=\sum_{i\leq N}h_i(x,y)$. It turns out that

$$\tag{3.6}\sum_{x,y}h_j(x,y)\theta(\{x,y\})=\int r_j d\nu.$$

To see this, it is enough to show (by differentiation) that for each $t>0$

$$\tag{3.7}2^{-N}\sum_{x,y}h_j(x,y)a_t(x,y)f(x)=\int r_j f_t dP.$$

We substitute in the left-hand side the value of $\frac{d}{dt}a_t(x,y)$ given by (3.5), and we observe that the terms corresponding to $i\neq j$ cancel out. We also observe that $a_t(x,y)=0$ unless $x_i\geq y_i$ for each i. This means that the "fluid can flow only downwards" and can be proved in the spirit of the proof of (3.2). Thus the summation of (3.7) reduces to the case $x_j = 1, y_j = 0$, and use of (3.4) then completes the proof. Thus the previous approach does provide an alternate proof to Theorem 2 (in the case $p = 1/2$).

It seems natural to conjecture that the previous method constructs a transport that is not only canonical, but in some sense optimal. Unfortunately it is very difficult to analyse.

Example 3.3: For $t, u \geq 1$, we consider

$$A_{t,N} = \Big\{x; \sum_{i\leq N} x_i \geq \frac{N}{2} - t\sqrt{N}\Big\} \quad \text{and} \quad B_{u,N} = \Big\{x; \sum_{i\leq N} x_i \leq \frac{N}{2} - u\sqrt{N}\Big\}.$$

Thus, by the central limit theorem,

$$\lim_{N\to\infty} P(B_{u,N}) \geq \frac{1}{Ku}\exp(-2u^2).$$

There, as well as in the rest of the proof, K denotes a universal constant, not necessarily the same at each occurrence. Since $d(x, A_{t,N}) \geq (u-t)\sqrt{N}$ on $B_{u,N}$, we have

$$\begin{aligned}\int d^2(x, A_{t,N})dP(x) &= 2\int vP(d(.,A_{t,N}) \geq v)dv \\ &\geq 2N\int vP(B_{t+v,N})dv\end{aligned}$$

and thus

$$\begin{aligned}\lim\inf_{N\to\infty} \frac{1}{N}\int d^2(x, A_{t,N})dP(x) &\geq \frac{1}{K}\int_0^\infty \frac{v}{t+v}\exp(-2(t+v)^2)dv \\ &\geq \frac{1}{Kt^3}\exp(-2t^2).\end{aligned} \tag{3.8}$$

On the other hand, it is easily seen through normal approximation that

$$\limsup N^{-1/2}\sum_{i\leq N} I_i(A_{t,N}) \leq K\exp(-2t^2). \tag{3.9}$$

Relations (3.8) and (3.9) represent the main features of the example; the rest of the proof consists in transforming this example in a set of probability of order 1/2 while presuming these features. Consider now (as provided by [K-K-L]) a monotone set $C_N \subset \{0,1\}^N$ such that $P(C_N)$ is of order 1/2, while $\sum_{i\leq N} I_i(C_N) \leq K\log N$.

Consider the set $D_{N,t} = C_N \cap A_{t,N}$, which is also such that $P(D_{N,t})$ is of order $\frac{1}{2}$. Now, given two monotone sets C, A, it is simple to see that $I_i(A\cap C) \leq$

$I_i(A) + I_i(C)$. Indeed, we recall that if $T_i(x)$ denotes the point obtained from x by changing the ith coordinate,

$$I_i(A) = P(\{x \in A;\, T_i(x) \notin A\})$$

and, if $x \in A \cap C$, $T_i(x) \notin A \cap C$, then either $T_i(x) \notin A$ or $T_i(x) \notin C$. Thus, fixing t, we have from (3.9) that

$$\lim \sup_{N \to \infty} N^{-1/2} \sum_{i \le N} I_i(D_{N,t}) \le K \exp(-2t^2).$$

Combining with (3.8) we see that for large N

$$\begin{aligned}\int d^2(x, D_{N,t}) dP(x) &\ge \frac{1}{Kt^3} \sqrt{N} \Big(\sum_{i \in N} I_i(D_{N,t}) \Big) \\ &\ge \frac{e^{2t^2}}{Kt^3} \Big(\sum_{i \le N} I_i(D_{N,t}) \Big)^2 .\end{aligned}$$

ACKNOWLEDGEMENT: This research was motivated by questions of G. Kalai (who had first proved a weaker result in the same direction using very different methods).

References

[K-K-L] J. Kahn, G. Kalai and N. Linial, *The influence of variables on Boolean functions,* Proc. 29th IEEE, FOCS 58–80, IEEE, New York, 1988.

[M] K. Marton, *A concentration of measure inequality for contracting Markov chains,* Geometric and Functional Analysis **6** (1996), 553–571.

[T1] M. Talagrand, *How much are increasing sets positively correlated,* Combinatorica **16** (1996), 243–258.

[T2] M. Talagrand, *A new look at independence,* The Annals of Probability **24** (1996), 1–34.