PUBLICATIONS MATHÉMATIQUES DE L'I.H.É.S.

MICHEL TALAGRAND

Concentration of measure and isoperimetric inequalities in product spaces

Publications mathématiques de l'I.H.É.S., tome 81 (1995), p. 73-205
<http://www.numdam.org/item?id=PMIHES_1995__81__73_0>

CONCENTRATION OF MEASURE AND ISOPERIMETRIC INEQUALITIES IN PRODUCT SPACES

by Michel TALAGRAND

Abstract. The concentration of measure phenomenon in product spaces roughly states that, if a set A in a product Ω^N of probability spaces has measure at least one half, " most " of the points of Ω^N are " close " to A. We proceed to a systematic exploration of this phenomenon. The meaning of the word " most " is made rigorous by isoperimetric-type inequalities that bound the measure of the exceptional sets. The meaning of the work " close " is defined in three main ways, each of them giving rise to related, but different inequalities. The inequalities are all proved through a common scheme of proof. Remarkably, this simple approach not only yields qualitatively optimal results, but, in many cases, captures near optimal numerical constants. A large number of applications are given, in particular to Percolation, Geometric Probability, Probability in Banach Spaces, to demonstrate in concrete situations the extremely wide range of application of the abstract tools.

TABLE OF CONTENTS

AMS Classification numbers: Primary 60E15, 28A35, 60G99; Secondary 60G15, 68C15.

Dedicated to Vitali Milman

I. INEQUALITIES

1. Introduction

Upon reading the words " isoperimetric inequality " the average reader is likely to think of the classical statement:

(1.1) Among the bodies of a given volume in $\mathbf{R}^N$, the ball is the one with the smallest surface area.

This formulation, that needs the notion of surface area, is not very appropriate for generalization in abstract setting. A less known (equivalent) formulation is as follows:

(1.2) Among the bodies A of a given volume in $\mathbf{R}^N$, the one for which the set A_t of points within Euclidean distance t of A has minimum volume is the Euclidean ball.

It should be intuitive, taking $t \to 0$, that (1.2) implies (1.1). We will, however, rather be interested in large values of t. At first sight, this is uninteresting; but this first impression is created only by our deficient intuition, that functions correctly only for $N \leqslant 3$, and miserably fails for the large values of N that are of interest here.

For our point of view, the main feature of (1.2) is that it gives a lower bound on the volume of A_t that depends only on t and the volume of A.

From now on, all the measures considered will be probabilities (i.e. of total mass one). Following [G-M], [M-S], the basic ideas of concentration of measure may be described in the following way. Consider a (Polish) metric space (X, d). For a subset A of X, consider the d-ball A_t centered on A, i.e.

$$A_t = \{ x \in X : d(x, A) \leqslant t \}. \tag{1.3}$$

Consider now a Borel probability measure P on X. The concentration function $\alpha(P, t)$ is defined as

$$\alpha(P, t) = \sup \left\{ 1 - P(A_t) : P(A) \geqslant \frac{1}{2}, A \subset X, A \text{ Borel} \right\}.$$

In other words

$$P(A) \geqslant \frac{1}{2} \Rightarrow P(A_t) \geqslant 1 - \alpha(P, t). \tag{1.4}$$

It turns out that in many situations the function $\alpha(P, t)$ becomes extremely small when t grows. In rough words, if one starts with any set A of measure $\geqslant 1/2$, A_t is almost the entire space. This is the concentration of measure phenomenon. This idea started with the work of V. Milman on Dvoretzky's theorem on almost Euclidean sections of convex bodies [Mi1]. Most importantly, Milman understood that concentration of measure occurs extremely often [Mi2], and most vigorously promoted the idea. (In particular we refer to his paper [Mi3] to supplement our sketchy discussion.) Concentration of measure plays an important role in local theory of Banach spaces, and has become the central concept of the area of probability known as Probability in Banach spaces. (See the book [L-T2], and subsequent work such as [T6], [T7].)

A prime example of space where concentration of measure holds is the Euclidean sphere S_N of $\mathbf{R}^{N+1}$ equipped with its geodesic distance d and normalized Haar measure P_N, for which it can be shown that

$$\alpha(P_N, t) \leqslant \left(\frac{\pi}{8}\right)^{1/2} \exp\left(-\frac{(N-1)}{2} t^2\right). \tag{1.5}$$

(The central fact in Milman's approach to Dvortzky's theorem.) Closely related, and more in line with the topic of the present paper is the case $X = \mathbf{R}^N$, equipped with the

Euclidean distance and the canonical Gaussian measure γ_N (whose covariance is the Euclidean dot product). In that case

$$\alpha(\gamma_N, t) \leqslant \int_t^\infty \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du \leqslant \frac{1}{2} e^{-t^2/2}. \tag{1.6}$$

It should be pointed out that more is known. The Gaussian isoperimetric inequality states that

$$\gamma_N(A) = \gamma_1((-\infty, a]) \Rightarrow \gamma_N(A_t) \geqslant \gamma_1((-\infty, a+t]) \tag{1.7}$$

which implies (1.6) when $a = 0$. However, it is sufficient for many applications to know (1.6) or even the weaker inequality

$$\alpha(\gamma_N, t) \leqslant K e^{-t^2/K} \tag{1.8}$$

where K is a universal constant.

In the present work we perform a systematic investigation of the concentration of measure phenomenon in product spaces. Thus with the terminology above, X will be a product of probability spaces, and P a product measure. The statements will have the form (1.4). However, the set A_t, which consists of points close in a certain sense to A (and that, for convenience, we will call the t-fattening of A), will not always have the form (1.3). This is the crucial difference between the present work and previous investigations, such as [A-M], [M-S]. Indeed, it turns out that it is extremely fruitful to consider various notions of fattening. We will define three rather distinct notions of fattening. These notions are studied respectively in Chapters 2 to 4. Each of these notions can be studied with various levels of sophistication, and they are at times closely connected. Discussing the whole theory in this introduction would require too much repetition and is inappropriate for an article of the present length. Thereby, we have decided to mention here only the main new theme (that did not appear in this author's previous work) as well as a simple result that appears to have a considerable potential for applications.

Assume that $X = \Omega^N$ is a product of probability spaces, and that $P = \mu^N$ is a product probability. We recall that the Hamming distance d on X is given by

$$d(x, y) = \operatorname{card}\{i \leqslant N : x_i \neq y_i\}. \tag{1.9}$$

When A_t is given by (1.3), where d is the Hamming distance, an important result, proved in a special case in [A-M] (see [M-S] with a proof that extends verbatim to the general situation) is that the concentration function $\alpha(P, t)$ satisfies

$$\alpha(P, t) \leqslant K \exp\left(-\frac{t^2}{KN}\right), \tag{1.10}$$

where K is a universal constant.

One could interpret (1.9) by saying that we put a penalty 1 for each coordinate i where $x_i \neq y_i$. One recurring theme of the present paper is the investigation of what happens when, instead, we put a penalty $h(x_i, y_i)$, where $h(x, y)$ is a non-negative function on Ω^2. A striking and unexpected finding is that in several instances there is a high disymmetry between the roles of x and y. For example, in one of the main results of the paper (Theorem 4.4.1) if one requires that $h(x,y)$ should depend on x only, it has to be bounded; but, if it depends on y only, weak integrability conditions suffice.

Suppose now that $(\alpha_i)_{i \leqslant N}$ are positive numbers, and let us replace the distance (1.9) by

$$d_\alpha(x,y) = \sum_{i \leqslant N} \alpha_i 1_{\{x_i \neq y_i\}}.$$

It is then shown in [M-S] that (1.10) can be extended into

$$\textbf{(1.11)} \qquad \alpha(P, t) \leqslant K \exp\left(-\frac{t^2}{K \sum_{i \leqslant N} \alpha_i^2}\right).$$

One way to spell out this result is as follows:

Given $A \subset \Omega^N$, with $P(A) \geqslant \frac{1}{2}$, then, for all numbers $(\alpha_i)_{i \leqslant N}$, $\alpha_i \geqslant 0$, $\sum_{i \leqslant N} \alpha_i^2 = 1$, we have

$$\textbf{(1.12)} \qquad P(A_{t,\alpha}) \geqslant 1 - K \exp\left(-\frac{t^2}{K}\right)$$

where $\quad A_{t,\alpha} = \{ x \in \Omega^N : \exists y \in A, \sum_{i \leqslant N} \alpha_i 1_{\{x_i \neq y_i\}} \leqslant t \}.$

The first result of Chapter 4 states that (1.12) can be improved into

$$\textbf{(1.13)} \qquad P\left(\bigcap_\alpha A_{t,\alpha}\right) \geqslant 1 - K \exp\left(-\frac{t^2}{K}\right)$$

where the intersection is over all families $\alpha = (\alpha_i)_{i \leqslant N}$ as above. The power of this principle (that will be considerably perfected in Chapter 4) is by no means obvious at first sight, but will be demonstrated repeatedly through Chapters 6 to 9 (the easiest applications being in Chapters 6 and 7).

We have explained in terms of sets what is the concentration of measure phenomenon. However, rather than sets, one is more often interested in *functions*. In that case, the concentration of measure phenomenon takes the following form: if a function f on X is sufficiently regular, it is very concentrated around its median (hence around its mean). If M_f is a median of f, this is expressed by a (fast decreasing) bound on $P(|f - M_f| > t)$. For a simple example, (1.4) implies that if f has a Lipschitz constant 1 with respect to the underlying distance, then

$$\textbf{(1.14)} \qquad P(|f - M_f| \geqslant t) \leqslant 2\alpha(P, t).$$

Despite the fact that functions are potentially more important than sets, all our concentration of measure results are stated in terms of sets. (This is done in Part I.)

The essential reason for this choice is that the power and the generality of these results largely arises from the fact that they require only minimal structure (a condition better achieved by considering sets only). A secondary reason is that much of the progress reported in the present paper (including on some rather concrete questions presented in Part II) has been permitted, or at least helped by the abstract point of view; and thereby, it seems worthwhile to promote this approach. Nevertheless, the natural domain of application of the tools of Part I is obtaining bounds on $P(|f - M_f| \geq t)$, when f is a function defined on a product of measure spaces. We will, however, give no abstract statement of this type. We prefer instead to analyze a number of specific situations, reducing each time to statements about sets (the great variety of situations encountered indicates that this is possibly a clever choice). This is the purpose of Part II, where we will demonstrate the efficiency of the tools of Part I. It must be said that these specific situations have been of considerable help in pointing out the directions in which the abstract theory should be developed. Most of the abstract results are indeed directly motivated by applications.

Certainly there is a considerable number of situations where functions that are defined on a product of many measure spaces naturally occur, or equivalently that depend on many independent random variables. The examples presented here are certainly influenced by the past interests of the author. Their boundary, however, is likely to reflect the limited knowledge of this author rather than the limit of the power of abstract tools of Part I. (Should a reader be aware of another potential domain of application, he is urged to mention it to this author.) Quite logically, several of the examples we present have an " applied " flavor. This is simply because stochastic models occur in physics (such as in Percolation and Statistical Mechanics) and Computer Science (bin packing, assignment problem, geometric probability). The reason for the later is that these stochastic models do shed some light on the behavior of computationally intractable problems, and, for this reason, are widely studied today; see e.g., [C-L]. No previous knowledge whatsoever of these problems is required for reading the material of Part II, that we briefly describe now.

Each of the examples of Part II studies the deviation of a specific function f of many independent random variables from its mean. In every example but one, the function f is obtained as the solution of an optimization problem. This is not a coincidence, but rather reflects the fact that such situations are well adapted to the use of our methods. In Chapter 6, we apply (4.1.3) to stochastic bin packing. This simple application is presented first since it is while considering this problem that the power of (4.1.3) beyond Probabilities in Banach spaces was first realized. The application is not really typical. More typical is the application of Chapter 7, to the length of the longest increasing subsequence of a random permutation. This application puts forward the fact that when one studies the size of substructures whose existence is determined by a comparatively small number of random variables, rather than by the whole collection of random variables, inequality (4.1.3) fully takes advantage of that feature. This characteristic occurs again in Chapter 8, where is presented a general result that allows, as a rather weak and special corollary, to improve upon H. Kesten's recent results on first time passage in Percolation [K2]. In Chapter 9, we show how (4.1.3) again provides a natural approach to questions on random graphs. The challenge of the assignment problem considered in Section 10 is that the objective function f considered there is very small; it is of order one, while depending on N^2 independent variables of order one, each of them with a potentially disastrous influence on the objective function. In Chapter 11, we consider situations where the objective function f is defined in a geometrical manner from a random set of N points in the unit square. The common objective is to prove that f has Gaussian-like tails. However, the richness of the situation is unsuspected beforehand; apparently similar definitions require rather different levels of sophistication. In Chapter 12, we provide a simple derivation of the free energy in the Sherrington-Kirpatrick model for spin glasses at high temperature. Finally, in Chapter 13, we discuss how the study of sums of vector-valued independent random variables motivated the approach of this paper, and we discuss a few new specific results.

We now comment on the methods of Part I, their history, and compare them with competing methods.

There is a general method, that is becoming increasingly popular, to prove deviation inequalities for $|f - \mathrm{E}f|$. (That the mean rather than the median is involved is very much irrelevant.) It is to decompose f as the sum of a martingale difference sequence $f = \Sigma\, d_i$, and to use martingale inequalities. The generality of the method stems from the fact that such a decomposition is easy, simply writing $d_i = \mathrm{E}(f \mid \mathscr{F}_i) - \mathrm{E}(f \mid \mathscr{F}_{i-1})$ for any increasing filtration $(\mathscr{F}_i)$. This method was used in Probability in Banach Spaces (under the name of " Yurinski's method ") for the study of $f = \| \sum_{i \leqslant N} X_i \|$, where X_i are independent Banach space random variables (r.v.). (After an important step by B. Maurey [Mau1], the generality of the method was understood by G. Schechtman [S]. It soon became apparent, however, that this method would not always yield optimal results; this is what prompted the invention of the isoperimetric inequality of [T2] (more details on history are given in Chapter 12). An inequality very similar to the inequality of [T2], but with a much simpler proof, appears in the present paper as Theorem 3.1.1. The phenomenon described by this inequality was completely new at that time, and had a major impact in Probability in Banach spaces (prompting, in particular, the writing of the book [L-T2]). One could reasonably hope that this inequality would find applications to other domains; but as of today, this has not been the case. Another inequality that was discovered in relation with Probability in a Banach space is a predecessor of (4.1.3) [T1]. The inequality of [T1] did not, however, play a crucial role in that theory, because, for most applications, it could be replaced by the Gaussian isoperimetric inequality (1.7) to which it is related. For this reason, the discovery that (4.1.3) was the direction to pursue for applications outside Probability in Banach spaces was delayed until very recently. It does not seem possible to prove either (4.1.3), or even some of its most interesting consequences we will present in Part II through the martingale method. This should not be so surprising, since the inequalities of the present paper have been developed precisely to achieve what martingales seem unable to attain. Among the results of Chapters 2 to 5, apparently only those of Sections 2.1, 2.2 can be obtained using martingales; and the only reason why these are included here is that they provide an excellent and very simple setting to introduce our basic scheme of proof. A major thesis of the present paper is that, while in principle the martingale method has a wider range of applications, in many situations the abstract inequalities of Part I are not only more powerful, but require considerably less ingenuity to apply. In all the examples we examined, only in some rare situations, where the martingale is close to a sum of independent r.v., and where the value of numerical constants is crucial (such as [M-H]), did our methods fail to supersede martingales.

We now comment on the method of proof of the inequalities of Part I. Isoperimetric inequalities such as (1.7) are proved via rearrangements. That is, one produces a (simple if possible) way to transform the set A in a set T(A), of the same measure, but more regular, so that the measure of $T(A)_t$ is not more than the measure of A_t. The procedure is then iterated, in a way that the iterates of A converge to the " extremal case ". Rearrangements are the only known technique to obtain perfect inequalities such as (1.5), (1.6). The inequality of [T2], that started the present line of work was proved using rearrangements. The difficult proof requires different types of transformations, some of which prevent from obtaining the extremal sets.

Despite considerable efforts, rearrangements did not yield a proof of the inequality of [T1]. (As pointed out to me by N. Alon, the reason could be the complicated nature of the extremal sets.) A completely new method was developed there. The main discovery was that of a formulation that allows an easy proof by induction upon the number of coordinates. The wide applicability of the method became apparent only gradually. This method and its variations provide a unified scheme of proof of all our inequalities, that, in its simplest occurrence, is described in great detail in Section 2.1. Ironically enough, this method is, in its principle, rather similar to the martingale method; the extra power is gained from the possibility of abstract manipulations in product spaces. A considerable advantage of the method is that, proving the induction hypothesis reduces to proving certain statements involving only functions on Ω. At times this is extremely easy; sometimes it is a bit harder. But certainly the nature of the statements that have to be decided is such that they are bound to yield to sufficient effort. What on the other hand, is not entirely clear, is why this simple procedure seems so miraculously sharp; in the situations where explicit computations of the best possible constants given by the method has been possible, these constants have proved very close to the optimal. In the cases where only less precise estimates have been possible, these estimates appear nonetheless to capture, up to a constant, the exact order of what really happens, and this, in every single situation that has been investigated.

The paper has been written to be read without any knowledge of this author's previous work or of the topic in general. For the sake of completeness, the only previous result of the author that has not been either vastly generalized or considerably simplified has been reproduced (as Theorem 4.2.4). Significant effort has been made in writing the paper in an easily accessible form. For example, it turns out in several situations that the simplest occurrence of a new principle is also the most frequently used. In these cases, we have made a point to give a separate proof for this most important case. These (short) proofs also serve as an introduction to the more complicated proofs of subsequent more specialized results.

During the preparation of this paper, I asked a number of people whether they were aware of recent or potential uses of the martingale method. I am pleased to thank D. Aldous, E. Bolthausen, A. Frieze, C. McDiarmid, B. Pittel, M. Steele, W. Szpankowski for their precious suggestions. Special thanks are due to H. Kesten, who communicated to me preprints of his recent work on percolation [K]. Analysis of his results pointed the way to several of the major developments that are presented in the present paper. The material of Chapter 5 was directly motivated by questions of G. Schechtman concerning the " correct form " of the concentration of measure on the symmetric group. A. Frieze, J. Wehr and particularly S. Janson most helpfully contributed to literally hundreds of improvements upon an earlier version of this work. I also followed several precious suggestions from M. Ledoux. Finally, it must be acknowledged that this paper would not have been written if Vitali Milman had not, over the years, convinced this author of the central importance of the concentration of measure phenomenon and if Wansoo Rhee had not introduced him to most of the topics considered in Part II.

2. Control by one point

2.1. *The basic principle*

Throughout the paper we will consider a probability space (Ω, Σ, μ) and the product (Ω^N, μ^N). The product probability μ^N will be denoted simply by P.

Consider a subset A of Ω^N. For $x \in \Omega^N$, we measure how far x is from A by

$$f(A, x) = \min\{\operatorname{card}\{i \leqslant N; x_i \neq y_i\}; y \in A\}. \tag{2.1.1}$$

This is simply the Hamming distance from x to A. The reason that we use a different notation is that at later stages, we will introduce different ways to measure how far x is from A. These ways will not necessarily arise from a distance.

It should be observed that the function $f(A, \cdot)$ need not be measurable even when A is measurable. This is the reason for the upper integral and outer probability in Proposition 2.1.1. below. On the other hand, measurability questions are well understood, and are irrelevant in the study of inequalities. Since it would be distracting to spend time and energy on routine considerations, we have felt that it would be better to simply ignore all measurability questions, and treat all sets and functions as if they were measurable. This is certainly the case if one should assume that Ω is Polish, μ is a

Borel measure, and that one studies only compact sets, which is the only situation that occurs in applications. The reader will keep in mind that in the sequel, when measurability problems do arise, certain integrals (resp. probabilities) have to be replaced by upper integrals (resp. outer probabilities) just as in the statement of Proposition 2.1.1. (The reader who desires to have a proof of our statements without measurability assumption should be warned that it does not work to try to extend the proofs we give by putting outer integrals rather than integrals—the reason being that Fubini theorem fails for outer integrals. Rather one has to derive the general result from the special case of well-behaved sets by approximating general sets from inside by well-behaved sets.)

Proposition **2.1.1.** — *For $t > 0$, we have*

$$\text{(2.1.2)} \qquad \int^* e^{tf(A,x)}\, dP(x) \leqslant \frac{1}{P(A)}\left(\frac{1}{2} + \frac{e^t + e^{-t}}{4}\right)^N \leqslant \frac{1}{P(A)}\, e^{t^2 N/4}.$$

In particular,

$$\text{(2.1.3)} \qquad P^*(\{f(A, \cdot) \geqslant k\}) \leqslant \frac{1}{P(A)}\, e^{-k^2/N}.$$

As was pointed out in the introduction, the power of our approach largely rests upon the fact that it reduces the proof of an inequality in Ω^N such as (2.1.2) to the proof of a much simpler fact about functions on Ω. In the present case, the meat of Proposition 2.1.1 is as follows.

Lemma **2.1.2.** — *Consider a (measurable) function g on Ω. Assume $0 \leqslant g \leqslant 1$. Then we have*

$$\text{(2.1.4)} \qquad \int_\Omega \min\left(e^t, \frac{1}{g(\omega)}\right) d\mu(\omega) \int_\Omega g(\omega)\, d\mu(\omega) \leqslant a(t)$$

where we have set $a(t) = \left(\frac{1}{2} + \frac{e^t + e^{-t}}{4}\right)$.

Proof. — If we replace g by $\max(g, e^{-t})$, this does not change the first integral, but increases the second. Thus it suffices to prove that if $e^{-t} \leqslant g \leqslant 1$, we have

$$\int_\Omega \frac{1}{g}\, d\mu \int_\Omega g\, d\mu \leqslant a(t).$$

Consider the convex set $\mathscr{C}$ of measurable functions g on Ω for which $e^{-t} \leqslant g \leqslant 1$. On $\mathscr{C}$, the functional $g \mapsto \int_\Omega g^{-1}\, d\mu$ is convex. On the subset $\mathscr{C}_b$ of $\mathscr{C}$ that consists of the functions with integral b, this functional attains its maximum on an extreme point. There is no loss of generality to assume that μ has no atoms; then it is well known that an extreme point of $\mathscr{C}$ takes only the values e^{-t} and 1. Thereby it suffices to show that for $0 \leqslant u \leqslant 1$ we have

$$(1 - u + ue^{t})\,(1 - u + ue^{-t}) \leqslant a(t).$$

But the left hand side is invariant by changing u into $1 - u$, so that the maximum is obtained at $u = 1/2$ by concavity of the left-hand side, and is $a(t)$. □

The proof of Proposition 2.1.1 goes by induction over N. The case $\mathrm{N} = 1$ follows from the application of (2.1.4) to $g = 1_{\mathrm{A}}$.

Suppose now that the result has been proved for N, and let us prove it for $\mathrm{N} + 1$. Consider $\mathrm{A} \subset \Omega^{\mathrm{N}+1} = \Omega^{\mathrm{N}} \times \Omega$. For $\omega \in \Omega$, we set

(2.1.5) $$\mathrm{A}(\omega) = \{ x \in \Omega^{\mathrm{N}};\ (x, \omega) \in \mathrm{A} \}$$

and $$\mathrm{B} = \{ x \in \Omega^{\mathrm{N}};\ \exists\, \omega \in \Omega,\ (x, \omega) \in \mathrm{A} \}.$$

With obvious notation, we have

$$f(\mathrm{A}, (x, \omega)) \leqslant f(\mathrm{A}(\omega), x).$$

Indeed, if $y \in \mathrm{A}(\omega)$, then $(y, \omega) \in \mathrm{A}$, and the number of coordinates where (y, ω) and (x, ω) differ is the number of coordinates where x and y differ. Thus, by induction hypothesis, we have

(2.1.6) $$\int_{\Omega^{\mathrm{N}}} \exp(tf(\mathrm{A}, (x, \omega)))\, d\mathrm{P}(x) \leqslant \frac{a(t)^{\mathrm{N}}}{\mathrm{P}(\mathrm{A}(\omega))}.$$

We also observe that

$$f(\mathrm{A}, (x, \omega)) \leqslant f(\mathrm{B}, x) + 1$$

so that, by induction hypothesis, we have

$$\int_{\Omega^{\mathrm{N}}} e^{tf(\mathrm{A},(x,\omega))}\, d\mathrm{P}(x) \leqslant \frac{e^{t}\, a(t)^{\mathrm{N}}}{\mathrm{P}(\mathrm{B})},$$

and combining with (2.1.6) we get

$$\int_{\Omega^{\mathrm{N}}} e^{tf(\mathrm{A},(x,\omega))}\, d\mathrm{P}(x) \leqslant a(t)^{\mathrm{N}} \min\left(\frac{e^{t}}{\mathrm{P}(\mathrm{B})}, \frac{1}{\mathrm{P}(\mathrm{A}(\omega))}\right).$$

Integrating in ω, we have

$$\int_{\Omega^{\mathrm{N}+1}} e^{tf(\mathrm{A},(x,\omega))}\, d\mathrm{P}(x)\, d\mu(\omega) \leqslant a(t)^{\mathrm{N}} \int_{\Omega} \min\left(\frac{e^{t}}{\mathrm{P}(\mathrm{B})}, \frac{1}{\mathrm{P}(\mathrm{A}(\omega))}\right) d\mu(\omega).$$

To complete the induction, it suffices to show, by Fubini theorem, that

$$\int_\Omega \min\left(\frac{e^t}{P(B)}, \frac{1}{P(A(\omega))}\right) d\mu(\omega) \leqslant \frac{a(t)}{P \otimes \mu(A)} = \frac{a(t)}{\int_\Omega P(A(\omega))\, d\mu(\omega)}.$$

But this follows from (2.1.4) applied to the function $g(\omega) = P(A(\omega))/P(B)$.

We now finish the proof of Proposition 2.1.1. We note that

$$a(t) = 1 + \sum_{n \geqslant 1} \frac{t^{2n}}{2(2n)!}.$$

Now $2(2n)! \geqslant 4^n n!$. Indeed, this holds for $n = 1$, $n = 2$, while if $n + 1 \geqslant 4$, we have

$$\frac{(2n)!}{n!} = (n+1) \ldots (2n) \geqslant 4^n.$$

Thus

$$a(t) \leqslant 1 + \sum_{n \geqslant 1} t^{2n}/4^n\, n! = \exp(t^2/4).$$

Finally, (2.1.3) follows from Chebyshev inequality

$$\begin{aligned} P(\{f(A, \cdot) \geqslant k\}) &\leqslant e^{-tk} \int e^{tf(A,x)}\, dP(x) \\ &\leqslant \frac{1}{P(A)} e^{-tk + Nt^2/4} \end{aligned}$$

for $t = 2k/N$. □

Remark **2.1.3.** — Consider a sequence $(a_i)_{i \leqslant N}$ of positive numbers. If we now replace (2.1.1) by

(2.1.7) $$f(A, x) = \inf\{ \Sigma\{ a_i : i \leqslant N;\ x_i \neq y_i \} : y \in A\}$$

the proof of Proposition 2.1.1 shows that

(2.1.8) $$\int e^{tf(A,x)}\, dP(x) \leqslant \frac{1}{P(A)} e^{t^2 \Sigma_{i \leqslant N} a_i^2/4}$$

and, by Chebyshev inequality,

(2.1.9) $$P(\{f(A, \cdot) \geqslant u\}) \leqslant \frac{1}{P(A)} e^{-u^2/\Sigma_{i \leqslant N} a_i^2}.$$

A number of inequalities presented in Chapters 2 to 5 allow extensions that parallel the way Remark 2.1.3 expands Proposition 2.1.1. These extensions are immediate, and will not be stated. It should be pointed out, on the other hand, that no gain of generality would be obtained in Proposition 2.1.1 by replacing the product Ω^N, μ^N by a product $\prod_{i \leqslant N} \Omega_i$, $\bigotimes_{i \leqslant N} \mu_i$. This comment also applies to many inequalities that we will subsequently prove.

2.2. *Sharpening*

Having proved (2.1.2), it is natural to wonder whether this could be improved by allowing another type of dependence of the right-hand side as a function of P(A). The most obvious choice is to replace $P(A)^{-1}$ by $P(A)^{-\alpha}$ for some $\alpha > 0$.

Proposition **2.2.1.** — *For* $t \geqslant 0$, *we have*

$$\int e^{tf(A,x)}\, dP(x) \leqslant \frac{a(\alpha, t)^N}{P(A)^\alpha} \tag{2.2.1}$$

where

$$a(\alpha, t) = \frac{\alpha^\alpha}{(\alpha+1)^{\alpha+1}} \frac{(e^t - e^{-t/\alpha})^{1+\alpha}}{(1-e^{-t/\alpha})(e^t - 1)^\alpha}. \tag{2.2.2}$$

Proof. — Following the scheme of proof of Proposition 2.1.1, (2.2.1) holds provided that, for each function $0 \leqslant g \leqslant 1$ on Ω, we have

$$\int_\Omega \min\left(e^t, \frac{1}{g^\alpha}\right) d\mu \left(\int_\Omega g\, d\mu\right)^\alpha \leqslant a(\alpha, t).$$

Following the proof of Lemma 2.1.2, we see that we can take

$$a(\alpha, t) = \sup_{0 \leqslant u \leqslant 1} (1 + u(e^t - 1))\,(1 - u(1 - e^{-t/\alpha}))^\alpha, \tag{2.2.3}$$

from which (2.2.2) follows by calculus. □

Certainly neither the author nor the reader are enthusiastic about the prospect of using (2.2.1) and optimizing in Chebyshev inequality. The purpose of the next result is to obtain a more manageable bound, that also makes clearer the gain obtained by taking large values of α.

Lemma **2.2.2.**

$$a(\alpha, t) \leqslant \exp \frac{t^2}{8}\left(1 + \frac{1}{\alpha}\right).$$

Proof. — Interestingly, rather than using (2.2.2), it seems simpler to go back to (2.2.3) and to show that, whenever $0 \leqslant u \leqslant 1$, we have

$$(1 + u(e^t - 1))\,(1 - u(1 - e^{-t/\alpha}))^\alpha \leqslant \exp \frac{t^2}{8}\left(1 + \frac{1}{\alpha}\right),$$

or, equivalently (after removing 1 from each term of the left-hand side)

$$\log(1 + u(e^t - 1)) + \alpha \log(1 - u(1 - e^{-t/\alpha})) \leqslant \frac{t^2}{8}\left(1 + \frac{1}{\alpha}\right). \tag{2.2.4}$$

Since (2.2.4) holds for $t = 0$, it suffices to show that the derivative of the left-hand side is bounded by the derivative of the right-hand side for $t \geqslant 0$, i.e.,

$$t \geqslant 0 \Rightarrow \frac{ue^t}{1 + u(e^t - 1)} - \frac{ue^{-t/\alpha}}{1 - u(1 - e^{-t/\alpha})} \leqslant \frac{t}{4}\left(1 + \frac{1}{\alpha}\right),$$

or, equivalently

$$t \geqslant 0 \Rightarrow \frac{u - 1}{1 + u(e^t - 1)} - \frac{u - 1}{1 - u(1 - e^{-t/\alpha})} \leqslant \frac{t}{4}\left(1 + \frac{1}{\alpha}\right). \tag{2.2.5}$$

Again (2.2.5) holds for $t = 0$. So it suffices to show that for $t \geqslant 0$, the derivative of the left-hand side of (2.2.5) is bounded by the derivative of the right-hand side, or, equivalently, that

$$u(1 - u)\left[\frac{e^t}{(1 - u + ue^t)^2} + \frac{1}{\alpha}\,\frac{e^{-t/\alpha}}{(1 - u + ue^{-t/\alpha})^2}\right] \leqslant \frac{1}{4} + \frac{1}{4\alpha}.$$

Now, using the inequality $4ab \leqslant (a + b)^2$, we see that

$$\frac{u(1 - u)\, e^t}{(1 - u + ue^t)^2} \leqslant \frac{1}{4}; \qquad \frac{u(1 - u)\, e^{-t/\alpha}}{(1 - u + ue^{-t/\alpha})^2} \leqslant \frac{1}{4}. \qquad \square$$

Corollary **2.2.3.** — *For* $t \geqslant 0$, *we have*

$$\int e^{tf(A,x)}\, dP(x) \leqslant \frac{1}{P(A)^\alpha} \exp N\frac{t^2}{8}\left(1 + \frac{1}{\alpha}\right). \tag{2.2.6}$$

In particular, for $k \geqslant \sqrt{\frac{N}{2} \log \frac{1}{P(A)}}$, *we have*

$$P(\{f(A, \cdot) \geqslant k\}) \leqslant \exp\left(-\frac{2}{N}\left(k - \sqrt{\frac{N}{2} \log \frac{1}{P(A)}}\right)^2\right). \tag{2.2.7}$$

Proof. — Certainly (2.2.6) follows from (2.2.1) and Lemma 2.2.2. Optimization over t in Chebyshev inequality yields

$$P(\{f(A, \cdot) \geqslant k\}) \leqslant \frac{1}{P(A)^\alpha} \exp\left(-\frac{2k^2}{N}\,\frac{\alpha}{\alpha + 1}\right).$$

For $k \geqslant \sqrt{\frac{N}{2} \log \frac{1}{P(A)}}$, making the (optimal) choice

$$\alpha = -1 + \sqrt{\frac{2k^2}{N \log \frac{1}{P(A)}}}$$

yields (2.2.7). □

It is an interesting fact that (2.2.7) is exactly the best bound that has been proved on $P(\{f(A, \cdot) \geqslant k\})$ using martingales (see [McD]). It is a natural question to wonder whether, when $P(A) \geqslant 1/2$, one indeed has

$$P(\{f(A, \cdot) \geqslant k\}) \leqslant K \exp\left(-\frac{2k^2}{N}\right)$$

for some universal constant K. More or less standard arguments (e.g., those contained in [T2]) show that it suffices to consider the case where $\Omega = \{0, 1\}$, where P is the product of measures $(\mu_i)_{i \leqslant N}$ on Ω, and where A is even " hereditary ". The case where $\mu_i(\{1\}) = 1/2$ for each $i \leqslant N$ is known, as a consequence of more precise results, such as Harper's inequality. Intuitively, this is the worst case.

Having obtained (2.2.6), one must wonder whether further improvements upon (2.2.6) are possible by considering yet other general dependencies of the right-hand side as a function of P(A). The reader who wishes to truly penetrate this paper will convince himself that this is not the case.

2.3. *Two point space*

Let us now consider the case where $\Omega = \{0, 1\}$, and set $p = \mu(\{1\})$, so that $\mu(\{0\}) = 1 - p$.

Proposition **2.3.1.** — *For* $t \geqslant 0$, $\alpha \geqslant 1$, *we have*

$$\textbf{(2.3.1)} \qquad \int e^{tf(A, x)} \, dP(x) \leqslant \frac{b(\alpha, t, p)^N}{P(A)^\alpha},$$

where, for $p \geqslant 1/2$, *we have set*

$$\textbf{(2.3.2)} \qquad b(\alpha, t, p) = ((1-p)\, e^t + p)\, (p + (1-p)\, e^{-t/\alpha})^\alpha,$$

and, for $p \leqslant 1/2$,

$$\textbf{(2.3.3)} \qquad b(\alpha, t, p) = b(\alpha, t, 1-p) = ((1-p)\, e^{-t} + p)\, (p + (1-p)\, e^{t/\alpha})^\alpha.$$

Proof. — Following the proofs of Propositions 2.1.1 and 2.2.1 it suffices to show that for any function $0 \leqslant g \leqslant 1$ on Ω we have

$$\int_\Omega \min\left(e^t, \frac{1}{g^\alpha}\right) d\mu \left(\int g \, d\mu\right)^\alpha \leqslant b(\alpha, t, p).$$

As in the proof of Lemma 2.1.2, we reduce to the case where $g \geqslant e^{-t/\alpha}$. Setting $a = g(0)$, $b = g(1)$, it suffices to show that, for $e^{-t/\alpha} \leqslant a, b \leqslant 1$ we have

$$\left((1-p)\frac{1}{a^\alpha} + \frac{p}{b^\alpha}\right)((1-p)\,a + pb)^\alpha \leqslant b(\alpha, t, p).$$

Setting $x = b/a$, it suffices to show that

$$e^{-t/\alpha} \leqslant x \leqslant e^{t/\alpha} \Rightarrow \varphi(x) \leqslant b(\alpha, t, p)$$

where we have set

$$\varphi(x) = ((1-p)\,x^\alpha + p)\left(\frac{1-p}{x} + p\right)^\alpha.$$

Now,

$$\varphi'(x) = \alpha p(1-p)\left(x^{\alpha-1} - \frac{1}{x^2}\right)\left(\frac{1-p}{x} + p\right)^{\alpha-1}$$

so that φ decreases for $x \leqslant 1$, increases for $x \geqslant 1$.

Also, we have

$$\varphi'(x) - \left(\varphi\left(\frac{1}{x}\right)\right)'$$
$$= \alpha p(1-p)\left(1 - \frac{1}{x^{\alpha+1}}\right)(((1-p) + px)^{\alpha-1} - ((1-p)\,x + p)^{\alpha-1}),$$

so that, for $x \geqslant 1$, this has the sign of $2p - 1$. Thus for $p \leqslant 1/2$, φ attains its maximum on the interval $[e^{-t/\alpha}, e^{t/\alpha}]$ at the right end of this interval, while for $p \geqslant 1/2$ it attains its maximum at the left end. (One should observe that changing x in $1/x$ and p in $1-p$ leave φ invariant.) □

A particularly important example is when

$$\mathrm{A} = \{x = (x_i) \in \{0, 1\}^{\mathrm{N}};\ \sum_{i \leqslant \mathrm{N}} x_i \leqslant k\}.$$

The use of (2.3.1) for this set and of Chebyshev inequality will in particular produce bounds for the tails of the binomial law. Thereby, it is not surprising that the computations involved in the use of (2.3.1) do run into the same type of difficulties as those involving the tails of the binomial law. We now show how, nonetheless, some simple and reasonably sharp results can be deduced (for general sets A) from (2.3.1). The reader will observe that the bound (2.3.1) is (of course) invariant when p is replaced by $1-p$, so that there is no loss of generality in assuming $p \geqslant 1/2$. Let us fix p, $\alpha \geqslant 1$, and consider

$$f(t) = \log b(\alpha, t, p) = \log((1-p)\,e^t + p) + \alpha \log(p + (1-p)\,e^{-t/\alpha}).$$

Thus $f(0) = 0$, and

$$f'(t) = (1-p)\left(\frac{1}{(1-p)+pe^{-t}} - \frac{1}{pe^{t/\alpha}+(1-p)}\right).$$

Thus $f'(0) = 0$, and

$$f''(t) = p(1-p)\,(h(e^{-t}) + \frac{1}{\alpha}\,h(e^{t/\alpha}))$$

where $$h(x) = \frac{x}{(px+1-p)^2} = \frac{x}{(1-(1-x)\,p)^2}.$$

Simple computations show that when $x \geqslant 1/e$, we have $|\, h(x) - 1 \,| \leqslant \mathrm{K}\,|\, x - 1 \,|$ for some universal constant K. It follows that

$$t \leqslant 1 \Rightarrow f''(t) \leqslant p(1-p)\left(\left(1+\frac{1}{\alpha}\right) + 4\mathrm{K}t\right)$$

and, by integration, that

$$t \leqslant 1 \Rightarrow f(t) \leqslant p(1-p)\left(\left(1+\frac{1}{\alpha}\right)\frac{t^2}{2} + \mathrm{K}t^3\right).$$

Thus, we have shown the first half of the following.

Corollary **2.3.2.** — *For* $\alpha \geqslant 1$, $0 \leqslant t \leqslant 1$, *we have*

$$\text{(2.3.4)} \qquad \int e^{tf(\mathrm{A},x)}\,d\mathrm{P}(x) \leqslant \frac{1}{\mathrm{P}(\mathrm{A})^{\alpha}} \exp \mathrm{N}\left[p(1-p)\left(1+\frac{1}{\alpha}\right)\frac{t^2}{2} + \mathrm{K}t^3\right].$$

In particular, for

$$\left(4p(1-p)\,\mathrm{N}\log\frac{1}{\mathrm{P}(\mathrm{A})}\right)^{1/2} \leqslant k \leqslant p(1-p)\,\mathrm{N}$$

we have

$$\text{(2.3.5)} \qquad \mathrm{P}(\{f(\mathrm{A},x) \geqslant k\})$$

$$\leqslant \exp\left(-\frac{1}{2p(1-p)\,\mathrm{N}}\left(k - \sqrt{2p(1-p)\,\mathrm{N}\log\frac{1}{\mathrm{P}(\mathrm{A})}}\right)^2 + \frac{\mathrm{K}k^3}{(p(1-p))^3\,\mathrm{N}^2}\right).$$

To obtain (2.3.5), one proceeds as in the proof of (2.2.7), using first Chebyshev inequality for $t = \dfrac{k}{p(1-p)}\,\dfrac{\alpha}{(1+\alpha)\,\mathrm{N}}$, then taking

$$\alpha = -1 + \sqrt{\frac{k^2}{2p(1-p)\,\mathrm{N}\log\dfrac{1}{\mathrm{P}(\mathrm{A})}}}.$$

It is of interest to compare the bound (2.3.5) with the isoperimetric inequalities obtained in [Lea]; these isoperimetric inequalities are optimal, but apply only to special sets (the so-called hereditary sets). The bound (2.3.5) is more general, and provides estimates of essentially the same quality.

We now turn to a rather different situation. Beside the measure μ, we consider on Ω another probability μ_1, with $p_1 = \mu_1(\{1\}) > p$, and we set $P_1 = \mu_1^N$.

Theorem **2.3.4.** — *For a subset* A *of* Ω^N, *and* $x \in \Omega^N$, *we consider*

$$f(A, x) = \min\{\operatorname{card}\{i \leqslant N;\, x_i = 1, y_i = 0\};\, y \in A\}.$$

Then, for $t \geqslant 0$,

$$\textbf{(2.3.6)} \qquad \int e^{tf(A,x)}\, dP(x) \leqslant \frac{a(\alpha, t)^N}{P_1(A)^\alpha}$$

where $\quad a(\alpha, t) = \max(1, (1 - p + pe^t)\,(p_1 e^{-t/\alpha} + 1 - p_1)^\alpha).$

Comment. — The really new phenomenon here is that for small t, one has $a(\alpha, t) = 1$. In particular, if $\alpha = 1$, this occurs whenever $e^t \leqslant p_1(1-p)/p(1-p_1)$ so that one has

$$\textbf{(2.3.7)} \qquad \int \left(\frac{p_1(1-p)}{p(1-p_1)}\right)^{f(A,x)} dP(x) \leqslant \frac{1}{P_1(A)}.$$

The remarkable feature about this statement is that it is independent of N (and so is in essence an infinite dimensional statement). This is the first of the results we present that apparently cannot be obtained via martingales (so it deserves to be called a theorem rather than a proposition). The reader that would like to gain intuition about the phenomenon captured by Theorem 2.3.4 should consider the case where $A = \{x \in \{0,1\}^N;\, \sum_{i \leqslant N} x_i \leqslant n\}$. In order to have $P_1(A)$ of order 1/2, one takes n equal to Np_1, assuming for simplicity that this number is an integer. Observing that

$$f(A, x) > k \Leftrightarrow \sum_{i \leqslant N} x_i > n + k = Np_1 + k = Np + (k + N(p_1 - p))$$

the quantity $P(\{f(A, x) > k\})$ can be estimated through the tails of the binomial law; the most interesting values of N are such that $N(p_1 - p) \sim k$.

The induction scheme of Proposition 2.1.1 will reduce Theorem 2.3.4 to an elementary two-point inequality, that is the object of the next lemma.

Lemma **2.3.5.** — *If* $a \leqslant b \leqslant 1$, *we have*

$$\textbf{(2.3.8)} \qquad \frac{1-p}{b^\alpha} + p \min\left(\frac{1}{a^\alpha}, \frac{e^t}{b^\alpha}\right) \leqslant \frac{a(\alpha, t)}{(ap_1 + b(1-p_1))^\alpha}.$$

Proof. — If we set $x = \min(b/a, e^{t/\alpha})$, we are reduced to show that

$$1 \leqslant x \leqslant e^{t/\alpha} \Rightarrow \varphi(x) \leqslant a(\alpha, t)$$

where $$\varphi(x) = (1 - p + px^{\alpha})\left(\frac{p_1}{x} + (1 - p_1)\right)^{\alpha} \leqslant a(\alpha, t).$$

But $\varphi'(x)$ has the sign of $p(1 - p_1)\, x^{\alpha+1} - p_1(1 - p)$, so it is negative for values of x close to one, and then, possibly, becomes positive. Thus φ attains its maximum on the interval $[1, e^{t/\alpha}]$ at one of the endpoints. □

Proof of Theorem 2.3.4. — We proceed by induction over N. For $N = 1$, since $f(A, \omega) \equiv 0$ when $1 \in A$, it suffices to consider the case $A = \{0\}$, in which case the result follows from (2.3.8) with $a = 0$, $b = 1$.

Assuming now that the theorem has been proved for N, we prove it for $N + 1$. Consider $A \subset \Omega^{N+1}$, and set $A_1 = \{ x \in \Omega^N ; (x, 1) \in A \}$. Consider the projection B of A on Ω^N. We observe that

$$f(A, (x, \omega)) \leqslant 1 + f(B, x)$$

$$f(A, (x, \omega)) \leqslant f(A_1, x)$$

so that setting $a = P_1(A_1)$, $b = P_1(B)$ and using the induction hypothesis, the result follows from (2.3.8). □

2.4. *Penalties, I*

A (somewhat imprecise) way to reformulate (2.1.1) is that we measure how far x is from A by simply counting the smallest number of coordinates of x that cannot be captured by a point of A. Rather than just giving a penalty of 1 for each coordinate we miss, it is natural to consider, given a non-negative function h on $\Omega \times \Omega$, the quantity

(2.4.1) $$f_h(A, x) = \inf\{ \sum_{i \leqslant N} h(x_i, y_i)\, 1_{\{x_i \neq y_i\}} ; y \in A \}.$$

To simplify the notation, we will assume

(2.4.2) $$\forall x \in \Omega, \quad h(x, x) = 0$$

so that (2.4.1) becomes

(2.4.3) $$f_h(A, x) = \inf\{ \sum_{i \leqslant N} h(x_i, y_i) ; y \in A \}.$$

Concerning (2.4.2), we should point out that we will let x, y denote points in Ω^N as well as points in Ω; when there is too much danger of confusion, however, points of Ω will be denoted by ω, ω'.

The function h will always be assumed to be measurable. The following simple result is already useful, as will be demonstrated in Chapter 11.

Theorem **2.4.1.** — *For each measurable subset* A *of* Ω^N, *and each* $t > 0$ *for which* $\iint \exp th(x,y)\, d\mu(x)\, d\mu(y) < \infty$, *we have, setting* $v(\omega, \omega') = \max(h(\omega, \omega'), h(\omega', \omega))$, *that*

$$\int_{\Omega^N} e^{t f_h(A,x)}\, dP(x) \leqslant \frac{1}{P(A)} \left(\frac{1}{2}\int_{\Omega^2} (e^{tv(w,w')} + e^{-tv(w,w')})\, d\mu(\omega)\, d\mu(\omega')\right)^N. \tag{2.4.4}$$

The crucial point of Theorem 2.4.1 is as follows.

Proposition **2.4.2.** — *Consider a function* $g \geqslant 0$ *on* Ω, *and set*

$$\hat{g}(x) = \inf_{y \in \Omega} (g(y) + th(x,y)). \tag{2.4.5}$$

Then

$$\int e^{\hat{g}}\, d\mu \int e^{-g}\, d\mu \leqslant \frac{1}{2}\int_{\Omega^2} (e^{tv(w,w')} + e^{-tv(w,w')})\, d\mu(\omega)\, d\mu(\omega'). \tag{2.4.6}$$

Indeed, a simple truncation argument shows that Proposition 2.4.2 remains true if one allows g to take values in $\mathbf{R}^+ \cup \{\infty\}$ (using obvious conventions). To prove Theorem 2.4.1 by induction over N, considering a subset A of Ω^{N+1}, we set

$$A(\omega) = \{ x \in \Omega^N ; (x, \omega) \in A \}$$

for $\omega \in \Omega$, and we define g by $P(A(\omega)) = e^{-g(\omega)}$. It should then be clear that (2.4.6) is exactly what is needed to make the induction work.

Proof of Proposition 2.4.2. — For simplicity we assume $\hat{g}$ measurable. Then the left-hand side of (2.4.6) coincides with

$$\iint_{\Omega^2} e^{\hat{g}(x) - g(y)}\, d\mu(x)\, d\mu(y).$$

We set $u(x,y) = \hat{g}(x) - g(y)$. By definition of $\hat{g}$, we have $\hat{g}(x) \leqslant g(y) + th(x,y)$. Since $h(x,x) = 0$, we also have $\hat{g}(x) \leqslant g(x)$. Hence

$$u(x,y) \leqslant th(x,y)\,; \quad u(x,y) \leqslant g(x) - g(y). \tag{2.4.7}$$

We now observe that for two numbers a, b, if $a + b \leqslant 0$, then

$$e^a + e^b \leqslant e^{\max(a,b,0)} + e^{-\max(a,b,0)}.$$

Thereby, by (2.4.7), we have

$$e^{u(x,y)} + e^{u(y,x)} \leqslant e^{tv(x,y)} + e^{-tv(x,y)}. \tag{2.4.8}$$

The result follows by integration. □

It is of interest to get simpler bounds. Let us observe the following elementary fact (that is obvious on power series expansions). For $t \leqslant 1$,

$$\frac{e^{tv} + e^{-tv}}{2} \leqslant 1 + \frac{t^2}{2}(e^v + e^{-v} - 2). \tag{2.4.9}$$

We note that, for an increasing function φ,

$$\varphi(\max(a, b)) \leqslant \max(\varphi(a), \varphi(b)) \leqslant \varphi(a) + \varphi(b).$$

Using this for $\varphi(x) = e^x + e^{-x} - 2$, $a = h(\omega, \omega')$, $b = h(\omega', \omega)$, using then (2.4.9) and integrating, we get the following from (2.4.4).

Theorem **2.4.3.** — *If*

$$\iint_{\Omega^2} \exp h(x, y)\, d\mu(x)\, d\mu(y) < \infty, \tag{2.4.10}$$

we have for $t \leqslant 1$,

$$\begin{aligned} \int_{\Omega^N} e^{tf_h(A, x)}\, dP(x) \\ \leqslant \frac{1}{P(A)} \exp\left(Nt^2 \iint_{\Omega^2} (e^{h(\omega, \omega')} + e^{-h(\omega, \omega')} - 2)\, d\mu(\omega)\, d\mu(\omega')\right). \end{aligned} \tag{2.4.11}$$

Corollary **2.4.4.** — *Assume*

$$\iint_{\Omega^2} \exp h(x, y)\, d\mu(x)\, d\mu(y) \leqslant 2. \tag{2.4.12}$$

Then for all $u \leqslant 2N$ *we have*

$$P(\{f_h(A, \cdot) \geqslant u\}) \leqslant \frac{1}{P(A)} e^{-u^2/4N}. \tag{2.4.13}$$

Proof. — Since $e^{-h} \leqslant 1$, under (2.4.12), the right-hand side of (2.4.11) becomes bounded by $P(A)^{-1} \exp Nt^2$, from which (2.4.13) follows by Chebyshev inequality. □

The following resembles Bernstein's inequality.

Corollary **2.4.5.** — *Assume that* $\| h \|_\infty = \sup_{x, y \in \Omega} h(x, y)$ *is finite. Then*

$$P(\{f_h(A, \cdot) \geqslant u\}) \leqslant \frac{1}{P(A)} \exp\left(-\min\left(\frac{u^2}{8N \| h \|_2^2}, \frac{u}{2 \| h \|_\infty}\right)\right) \tag{2.4.14}$$

where we have set $\| h \|_2 = \left(\iint_{\Omega^2} h^2(\omega, \omega')\, d\mu(\omega)\, d\mu(\omega')\right)^{1/2}$.

Proof. — By homogeneity, we can replace h by $h' = h/\| h \|_\infty$. For $x \leqslant 1$, by (2.4.9) taking $v = 1$, $t = x$, we have $e^x + e^{-x} - 2 \leqslant x^2(e + e^{-1} - 2) \leqslant 2x^2$. Thereby the right-hand side of (2.4.11) becomes bounded by $\mathrm{P(A)}^{-1} \exp 2\mathrm{N}t^2 \| h \|_2^2$, from which the result follows by Chebyshev inequality.

Remark. — The reader has possibly observed that we have made no special efforts to get sharp numerical constants (in contrast with the previous sections) and we have used the simplest estimates, however crude. This feature will occur repeatedly. For a number of the results we will present, the proofs do not seem adapted to obtaining sharp constants. Thereby, there is actually no point to track explicit values of the numerical constants involved. Throughout the paper, K will denote a universal constant, that may vary at each occurrence.

2.5. *Penalties, II*

It should be apparent from (2.4.1) that f_h depends on h only through the properties of the following functional, defined for subsets B of Ω

(2.5.1) $$h(\omega, \mathrm{B}) = \inf\{ h(\omega, \omega') ; \omega' \in \mathrm{B} \}.$$

(The reader should carefully compare this definition with (2.4.3) and note that in both cases the infimum is taken over the second variable.)

Thereby, one should expect that the exponential integrability of h can be replaced in Theorem 2.4.1 by a weaker condition on the functional $h(x, \mathrm{B})$. This is indeed the case.

Theorem **2.5.1.** — *Assume that for each subset* B *of* Ω *we have*

(2.5.2) $$\int_\Omega \exp 2h(x, \mathrm{B})\, d\mu(x) \leqslant \frac{e}{\mu(\mathrm{B})}.$$

Then, for each subset A *of* Ω^{N}, *and each* $0 \leqslant t \leqslant 1$, *we have*

(2.5.3) $$\int_{\Omega^{\mathrm{N}}} e^{t f_h(\mathrm{A}, x)}\, d\mathrm{P}(x) \leqslant \frac{e^{3t^2\mathrm{N}}}{\mathrm{P(A)}}.$$

Discussion. — 1) It is good to observe and keep in mind that by Hölder's inequality, we have for $a \leqslant 1$

(2.5.4) $$\int e^{ah}\, d\mu \leqslant \left(\int e^h\, d\mu\right)^a.$$

Thus, the precise value of constants such as the constants 2, e that occur in (2.5.2) is completely irrelevant. Actually we will use the following consequence of (2.5.2):

(2.5.5) $$\int_\Omega \exp h(x, \mathrm{B})\, d\mu(x) \leqslant \frac{\sqrt{e}}{\sqrt{\mu(\mathrm{B})}} \leqslant \frac{2}{\sqrt{\mu(\mathrm{B})}}.$$

2) Is it very instructive to compare (2.5.2) with a condition such as (2.4.10). Indeed, under (2.4.10), we have for all x

$$(2.5.6) \qquad \mu(B) \exp h(x, B) = \mu(B) \inf_{y \in B} \exp h(x,y) \leqslant \int_\Omega \exp h(x,y)\, d\mu(y).$$

Integrating in x gives

$$\int \exp h(x, B)\, d\mu(x) \leqslant \frac{1}{\mu(B)} \iint_{\Omega^2} \exp h(x,y)\, d\mu(x)\, d\mu(y).$$

Thus (with the exception of the largely irrelevant factor 2), (2.4.10) appear stronger than (2.5.2). It is indeed much stronger, a fact that is not surprising in view of the crudeness of (2.5.6). To get a concrete example, consider the case where Ω is itself a product of m spaces (and μ a product measure), and denote by $f(x,y)$ the Hamming distance in Ω. Then Proposition 2.1.1 asserts that the function $h = m^{-1/2} f$ satisfies (2.5.2). On the other hand (except in trivial cases) the function f/a will fail (2.4.12) unless a is of order m.

To prove Theorem 2.5.1, the induction method reduces to the proof of the following.

Proposition **2.5.2.** — *Consider* $0 \leqslant t \leqslant 1$, *and a function* $g \geqslant 0$ *on* Ω. *For* $s \geqslant 0$, *we set* $B_s = \{ g \leqslant s \}$, *and we consider*

$$(2.5.7) \qquad \hat{g}(x) = \inf_{s > 0} s + th(x, B_s).$$

Then under (2.5.2) *we have*

$$(2.5.8) \qquad \int e^{\hat{g}}\, d\mu \int e^{-g}\, d\mu \leqslant e^{3t^2}.$$

Proof. — We observe that

$$(2.5.9) \qquad \hat{g}(x) - g(y) \leqslant th(x, B_{g(y)}).$$

We then follow the argument of Proposition 2.4.2, using (2.5.9) rather than the first part of (2.4.7). Combining with the argument of Theorem 2.4.3, we are led to show that

$$\iint_{\Omega^2} e^{h(x, B_{g(y)})}\, d\mu(x)\, d\mu(y) \leqslant 4.$$

Using (2.5.5) and Fubini theorem, it suffices to show that

$$(2.5.10) \qquad \int_\Omega \frac{1}{\sqrt{\mu(B_{g(y)})}}\, d\mu(y) \leqslant 2.$$

The best way to prove this inequality is to observe that the left-hand side depends only on the function $s \to \mu(B_s)$. Thus there is no loss of generality in assuming that $\Omega = [0, 1]$, that μ is the Lebesgue measure, and that g is nondecreasing. But then $\mu(B_{g(y)}) \geqslant y$, and $\int_0^1 y^{-1/2}\, dy = 2$. □

As pointed out in the discussion, a natural application of Theorem 2.5.1 is to the case where Ω is already a product space. This will be used implicitly, but crucially in Section 11.5. To formulate in words what happens, Proposition 2.1.1 states that if A is a subset of a product Ω^N of N spaces, of measure 1/2, all but exceptional points x of Ω^N are such that there is a point in A that captures all but about $\sqrt{N}$ of their coordinates. Suppose now that $N = N_1 N_2$, and we think of the N coordinates as N_1 blocks of N_2 coordinates. Then, using Theorem 2.5.1, we know that (but for exceptional points x) not only we will find a point in A that misses only about $\sqrt{N}$ coordinates of x, but these coordinates will be concentrated in only about $\sqrt{N_1}$ blocks. An interesting question would be to quantify precisely what can be said when, rather than considering only two " levels ", one considers a large number of levels.

2.6. *Penalties, III*

In this section, we explore a new phenomenon, that will also be met in Sections 3.3.3 and 4.4.4. The notation of the present section will be used throughout the paper. Roughly speaking, what happens is that if, in (2.5.3), one allows a more general type of dependence in P(A) of the right-hand side, then a weaker condition than (2.5.2) will suffice; this will mean in practice weaker integrability requirements on h.

The dependence in P(A) we will consider will be of the type $e^{\theta(P(A))}$. Throughout the paper, θ will denote a convex decreasing function from $]0, 1]$ to $\mathbf{R}^+$, such that $\theta(1) = 0$, $\lim_{x \to 0} \theta(x) = \infty$. The most important example is $\theta(x) = -\log x$, in which case $e^{\theta(P(A))}$ is the familiar quantity $1/P(A)$. We will always denote by ξ the inverse function of θ, so that ξ is a convex function from $\mathbf{R}^+$ to $]0, 1]$, with $\xi(0) = 1$. We will always assume the following

$$\textbf{(2.6.1)} \qquad \xi'' \text{ decreases}; \quad \forall\, b > 0, \quad \xi''(b) \leqslant |\, \xi'(b) \,|.$$

For $x \in \mathbf{R}$, we set $x^+ = \max(x, 0)$, and we will keep the following notation, for $x \in \mathbf{R}$, $b \in \mathbf{R}^+$

$$\textbf{(2.6.2)} \qquad \Xi(x, b) = \xi(x^+) - \xi(b) - (x^+ - b)\, \xi'(b).$$

We denote by λ the Lebesgue measure on $[0, 1]$. The measure of a Borel set B is simply denoted by $|\, B \,|$.

Central to this section is the following technical condition, that relates ξ and a function $w \geqslant 0$ defined on $[0, 1]$.

Condition H(ξ, w).

(2.6.3) $\forall\, b \geqslant 0, \quad \forall\, t, \quad 0 \leqslant t \leqslant 1, \qquad \int_0^1 \Xi(b - tw(u), b)\, d\lambda(u) \leqslant t^2 \mid \xi'(b) \mid.$

First we will investigate conditions that imply (2.6.3) under two simple choices of ξ. Then we will look at a rather general situation where the meaning of (2.6.3) can be considerably clarified; and before stating the main result (Theorem 2.6.5) we will prove a technical lemma that will explain the precise purpose of condition H(ξ, w).

Proposition **2.6.1.** — *When* $\xi(x) = e^{-x}$, *condition* H(ξ, w) *holds provided* $\int e^{w}\, d\lambda \leqslant 2$.

Proof. — Indeed, we have

$$\begin{aligned}\Xi(x, b) &= e^{-x^+} - e^{-b} + (x^+ - b)\, e^{-b}\\ &\leqslant e^{-x} - e^{-b} + (x - b)\, e^{-b}\\ &= e^{-b}(e^{-(x-b)} + (x - b) - 1).\end{aligned}$$

Thus (2.6.3) holds provided

$$t \leqslant 1 \Rightarrow \int (e^{tw} - tw - 1)\, d\lambda \leqslant t^2.$$

But, since the function $x^{-2}(e^x - x - 1)$ increases for $x \geqslant 0$, we have

$$(e^{tw} - tw - 1) \leqslant t^2(e^w - w - 1). \ \square$$

Proposition **2.6.2.** — *If* $\xi(x) = \dfrac{1}{x+2}$, *then* (2.6.1) *holds and condition* H(ξ, ω) *holds provided* $\int w^2\, d\lambda \leqslant 2$.

Proof. — Setting $y = x^+$, we have

$$\begin{aligned}\Xi(x, b) &= \frac{1}{y+2} - \frac{1}{b+2} + \frac{y-b}{(b+2)^2}\\ &= \frac{(y-b)^2}{(y+2)\,(b+2)^2} \leqslant \frac{1}{2}\,(y-b)^2 \mid \xi'(b) \mid. \ \square\end{aligned}$$

One obvious consequence of (2.6.3), taking $t = 1$, is that

(2.6.4) $\mid \{ w \geqslant b \} \mid \Xi(0, b) \leqslant \mid \xi'(b) \mid.$

In practice for b large $\Xi(0, b)$ is of order 1; so (2.6.4) is really a tail condition. The next result shows that a condition of a similar nature is indeed sufficient, provided ξ' varies smoothly (i.e., satisfies the Δ_2-condition; which is not the case when $\xi(x) = e^{-x}$).

Proposition **2.6.3.** — *Assume that for a certain number* $L > 0$, *we have*

$$\forall b > 0, \quad \forall t \leqslant 1, \qquad \left| \xi'\left(\frac{b}{2t}\right) \right| \leqslant Lt^2 \,|\, \xi'(b) \,|. \tag{2.6.5}$$

Then (2.6.3) *holds provided the following two conditions hold:*

$$\int w^2 \, d\lambda \leqslant \frac{1}{L} \tag{2.6.6}$$

$$\forall b > 0, \quad |\{ w \geqslant b \}| \leqslant \frac{1}{2L} \,|\, \xi'(b) \,|. \tag{2.6.7}$$

Proof. — We write

$$\int \Xi(b - tw, b) \, d\lambda \leqslant \int_{\{tw \leqslant b/2\}} \Xi(b - tw, b) \, d\lambda + |\{ tw \geqslant b/2 \}|. \tag{2.6.8}$$

By Taylor's formula, since ξ'' decreases, and $\xi''(b) \leqslant |\, \xi'(b) \,|$, we have, by (2.6.5)

$$\begin{aligned} x \geqslant b/2 \Rightarrow \Xi(x, b) &\leqslant \frac{(x-b)^2}{2} \xi''\left(\frac{b}{2}\right) \\ &\leqslant \frac{(x-b)^2}{2} \left| \xi'\left(\frac{b}{2}\right) \right| \\ &\leqslant L \frac{(x-b)^2}{2} \,|\, \xi'(b) \,|. \end{aligned}$$

Thus

$$\int_{\{tw \leqslant b/2\}} \Xi(b - tw, b) \, d\lambda \leqslant \frac{Lt^2}{2} \,|\, \xi'(b) \,|\, \int w^2 \, d\lambda.$$

Also, by (2.6.7), (2.6.5)

$$|\{ tw \geqslant b/2 \}| \leqslant \frac{1}{2L} \left| \xi'\left(\frac{b}{2t}\right) \right| \leqslant \frac{t^2}{2} \,|\, \xi'(b) \,|.$$

The result follows, combining with (2.6.8). □

The reader should observe that the functions $\xi(x) = (1 + x)^{-\alpha}$ $(\alpha \leqslant 1)$ satisfy (2.6.5).

The following lemma explains the purpose of condition $H(\xi, w)$.

Lemma **2.6.4.** — *Consider a function* $f \geqslant 0$ *on* Ω. *Assume that for a certain* t, $0 \leqslant t \leqslant 1$ *and all* $s \leqslant b$ *we have*

$$\mu(\{f \leqslant s\}) \leqslant |\{ tw \geqslant b - s \}| = |\{ b - tw \leqslant s \}|. \tag{2.6.9}$$

Then, under condition H(ξ, w), *we have*

$$\textbf{(2.6.10)}\qquad \int_C \xi(f)\, d\mu$$
$$\leqslant \xi(b)\,\mu(C) + \xi'(b)\int_C (f-b)\,d\mu + t^2\,|\,\xi'(b)\,| + \frac{1}{2}\xi''(b)\int_{C\cap\{f\geqslant b\}} (f-b)^2\,d\mu$$

for each set C.

Proof. — By definition of Ξ, (2.6.10) is equivalent to

$$\int_C \Xi(f,b)\,d\mu \leqslant t^2\,|\,\xi'(b)\,| + \frac{1}{2}\xi''(b)\int_{C\cap\{f\geqslant b\}} (f-b)^2\,d\mu.$$

By Taylor's formula, and since ξ'' decreases, for $x > b$ we have

$$\Xi(x,b) \leqslant \frac{1}{2}(x-b)^2\,\xi''(b)$$

and thus

$$\int_{C\cap\{f\geqslant b\}} \Xi(f,b)\,d\mu \leqslant \frac{1}{2}\xi''(b)\int_{C\cap\{f\geqslant b\}} (f-b)^2\,d\mu.$$

If we remember that $\Xi \geqslant 0$, and if we use condition H(ξ, w), we then see that it suffices to show that

$$\textbf{(2.6.11)}\qquad \int_{\{f\leqslant b\}} \Xi(f,b)\,d\mu \leqslant \int \Xi(b-tw,b)\,d\lambda.$$

Now, (2.6.9) implies that for all $s < b$ we have

$$\mu(\{f\leqslant s\}) \leqslant |\{\,b-tw\leqslant s\,\}|.$$

Thus, since $\Xi(x,b)$ decreases for $x \leqslant b$, we have, for all $z > 0$,

$$\mu(1_{\{f\leqslant b\}}\,\Xi(f,b)\geqslant z) \leqslant |\{\,\Xi(b-tw,b)\geqslant z\,\}|,$$

from which (2.6.11) follows. □

Theorem **2.6.5.** — *Consider a function* h *on* $\Omega\times\Omega$, *and a nonincreasing function* w *on* [0, 1] *such that* $\int w^2\,d\lambda \leqslant 1$. *Assume that for each subset* B *of* Ω, *we have*

$$\textbf{(2.6.12)}\qquad \int_\Omega \exp h(x,B)\,d\mu(x) \leqslant \exp(w(\mu(B))),$$

where we keep the usual notation $h(x, \mathrm{B}) = \inf\{ h(x,y)\,; y \in \mathrm{B}\}$. *Consider a function* θ *as usual, and assume that the condition* $\mathrm{H}(\zeta, w)$ *holds. Then, for each subset* A *of* Ω^{N} *and all* $t \leqslant 1$, *we have, for all* $t \leqslant 1$,

$$\int_{\Omega^{\mathrm{N}}} e^{tf_h(\mathrm{A},x)}\, d\mathrm{P}(x) \leqslant \exp(4\mathrm{N}t^2 + \theta(\mathrm{P}(\mathrm{A}))).$$

To understand better (2.6.12) it is of interest to specialize to the case where h depends only on x (resp. y). If h depends on x only, (2.6.12) means that $\int_\Omega \exp h(x)\, d\mu(x) \leqslant \exp w(0)$. If h depends on y only, then (2.6.12) becomes

$$\inf\{ h(y)\,; y \in \mathrm{B}\} \leqslant w(\mu(\mathrm{B})).$$

Taking $\mathrm{B} = \{ h \geqslant s\}$, we get $s \leqslant w(\mu(\{ h \leqslant s\}))$ and, since w in nonincreasing, this implies

$$\mu(\{ h \geqslant s\}) \leqslant |\{ w \geqslant s\}|.$$

It is easy to see that, conversely, this implies (2.6.12) when h depends upon y only and when w is left continuous.

To prove Theorem 2.6.5, it suffices, using induction over N, to prove the following.

Proposition **2.6.6.** — *Consider a function* g *on* Ω, $0 < g \leqslant 1$, *and set*

$$\widehat{\theta g}(x) = \inf_{y \in \Omega}\{ \theta(g(y)) + th(x,y)\}.$$

Then, under the conditions of Theorem 2.6.5, *for* $t \leqslant 1$, *we have*

$$\int e^{\widehat{\theta g}}\, d\mu \leqslant \exp\left(4t^2 + \theta\left(\int g\, d\mu\right)\right).$$

Clearly, this is equivalent to the following.

Proposition **2.6.7.** — *Consider a function* f *on* Ω, $f \geqslant 0$, *and set*

$$\hat{f}(x) = \inf_{y \in \Omega}\{ f(y) + th(x,y)\}.$$

Then, under the conditions of Theorem 2.6.5, *for* $t \leqslant 1$, *we have*

$$\textbf{(2.6.13)} \qquad \int e^{\hat{f}}\, d\mu \leqslant \exp\left(4t^2 + \theta\left(\int \xi(f)\, d\mu\right)\right).$$

Proof. — The problem is that we have on the right of (2.6.13) the quantity $\theta\left(\int \xi(f)\, d\mu\right)$ rather than the larger quantity $\int f\, d\mu$. We consider t as fixed through the proof.

Step 1. — We set $B_s = \{f \leqslant s\}$ for $s \geqslant 0$, and

$$b = \inf_s \{ s + tw(\mu(B_s)) \}.$$

We note that $\hat{f}(x) \leqslant f(x)$. We consider the function f' given by

$$\begin{aligned} f'(x) &= \hat{f}(x) && \text{if } \hat{f}(x) > b, \\ f'(x) &= b && \text{if } \hat{f}(x) \leqslant b < f(x), \\ f'(x) &= f(x) && \text{if } f(x) < b. \end{aligned}$$

Since $\hat{f} \leqslant f$, we have $\hat{f} \leqslant f' \leqslant f$. Thus $\int e^{\hat{f}}\, d\mu \leqslant \int e^{f'}\, d\mu$, and $\xi(f) \leqslant \xi(f')$, so $\int \xi(f)\, d\mu \leqslant \int \xi(f')\, d\mu$ and $\theta\left(\int \xi(f')\, d\mu\right) \leqslant \theta\left(\int \xi(f)\, d\mu\right)$. Thereby, it suffices to prove that

$$\int e^{f'}\, d\mu \leqslant \exp\left(4t^2 + \theta\left(\int \xi(f')\, d\mu\right)\right). \tag{2.6.14}$$

Step 2. — By definition of b, we have, for $s < b$,

$$tw(\mu(B_s)) \geqslant b - s.$$

Since w is nonincreasing,

$$|\{ tw \geqslant b - s \}| \geqslant \mu(B_s). \tag{2.6.15}$$

Since $f'(x) = f(x)$ when $f(x) < b$, we see that (2.6.9) holds (for f' rather than f). Since $f' = \hat{f}$ when $f'(x) > b$, by (2.6.10) used for $C = \Omega$, we get, since $\xi''(b) \leqslant |\xi'(b)|$,

$$\int \xi(f')\, d\mu \leqslant \xi(b) + \xi'(b) \int (f' - b)\, d\mu + |\xi'(b)| \left(t^2 + \frac{1}{2} \int_{\{\hat{f} > b\}} (\hat{f} - b)^2\, d\mu\right). \tag{2.6.16}$$

Step 3. — If $y \in B_s$, we have

$$\hat{f}(x) \leqslant f(y) + th(x,y) \leqslant s + th(x,y),$$

so that $$\hat{f}(x) \leqslant s + th(x, B_s).$$

Thus, by (2.6.12) we have

$$\int e^{t^{-1}\hat{f}}\, d\mu \leqslant \exp(t^{-1}(s + tw(\mu(B_s)))).$$

Taking the infimum over s yields

$$\int e^{t^{-1}(\hat{f}-b)}\, d\mu \leqslant 1. \tag{2.6.17}$$

Since $e^{x^+} \leqslant 1 + e^x$, we get

$$\int e^{t^{-1}(\hat{f}-b)^+}\, d\mu \leqslant 2. \tag{2.6.18}$$

Step 4. — The inequality $e^x \geqslant 1 + x^2/2$ for $x \geqslant 0$, and (2.6.18), show that $\int ((\hat{f} - b)^+)^2\, d\mu \leqslant 2t^2$. Combining with (2.6.16), we get

$$\int \xi(f')\, d\mu \leqslant \xi(b) + \xi'(b) \int (f' - b)\, d\mu + 2t^2 \,|\, \xi'(b) \,|. \tag{2.6.19}$$

The convexity of θ implies that $\theta(x) \geqslant \theta(y) + (x - y)\,\theta'(y)$. Also, since $\theta(\xi(x)) = 1$, we have $\theta'(\xi(b)) = 1/\xi'(b)$. Thus (2.6.19) implies

$$\begin{aligned} \theta\left(\int \xi(f')\, d\mu\right) &\geqslant b + \int (f' - b)\, d\mu - 2t^2 \\ &= \int f'\, d\mu - 2t^2. \end{aligned} \tag{2.6.20}$$

Step 5. — To finish the proof, it is thereby sufficient to show that

$$\int e^{f'}\, d\mu \leqslant \exp\left(2t^2 + \int f'\, d\mu\right). \tag{2.6.21}$$

Consider the function $r(x) = e^x - x - 1$, so that

$$\begin{aligned} \int e^{f'-b}\, d\mu &= 1 + \int (f' - b)\, d\mu + \int r(f' - b)\, d\mu \\ &\leqslant \exp\left(\int (f' - b)\, d\mu + \int r(f' - b)\, d\mu\right) \end{aligned}$$

and thus it suffices to show that $\int r(f' - b)\, d\mu \leqslant 2t^2$. We observe by (2.6.18) that

$$\int_{\{f' > b\}} r(t^{-1}(f' - b))\, d\mu \leqslant 1$$

and, since as already observed, the function $x^{-2} r(x)$ increases for $x > 0$, this implies

$$\int_{\{f' > b\}} r(f' - b)\, d\mu \leqslant t^2.$$

Also, it is elementary to see that $r(x) \leqslant x^2/2$ for $x < 0$. Now, by (2.6.15), we have

$$\int_{\{f' < b\}} (f' - b)^2\, d\mu \leqslant t^2 \int w^2\, d\mu \leqslant t^2. \quad \square$$

2.7. *Penalties, IV*

This section is devoted to the remarkable fact that if (2.5.2) is suitably reinforced, the term $\exp t^2 \mathrm{N}$ can be removed in (2.5.3).

To express conveniently the conditions we need, we introduce the function $c(a, t)$, defined for $0 < a < 1$, $t > 0$, as follows (c stands for concentration): if ν_1 is the measure on $\mathbf{R}$ of density $\frac{1}{2} e^{-|x|}$ with respect to the Lebesgue measure, we have $c(a, t) = \nu_1((-\infty, b + t])$, where b is given by $a = \nu_1((-\infty, b])$. Simple considerations show that

$$a \geqslant \frac{1}{2} \Rightarrow c(a, t) = 1 - e^{-t}(1 - a)$$

$$a \leqslant \frac{1}{2}, \quad e^t a \leqslant \frac{1}{2} \Rightarrow c(a, t) = e^t a$$

$$a \leqslant \frac{1}{2}, \quad e^t a \geqslant \frac{1}{2} \Rightarrow c(a, t) = 1 - \frac{1}{e^t a}.$$

Theorem **2.7.1.** — *Assume that for each subset* B *of* Ω *we have*

(2.7.1) $\quad t \leqslant 1 \Rightarrow \mu(\{h(\cdot, \mathrm{B}) \leqslant t^2\}) \geqslant c(\mu(\mathrm{B}), t)$

(2.7.2) $\quad t \geqslant 1 \Rightarrow \mu(\{h(\cdot, \mathrm{B}) \leqslant t\}) \geqslant c(\mu(\mathrm{B}), t).$

Then, for each subset A *of* Ω^{N}, *we have*

(2.7.3) $$\int_{\Omega^{\mathrm{N}}} e^{\mathrm{K}^{-1} f_h(\mathrm{A}, x)}\, d\mathrm{P}(x) \leqslant \frac{1}{\mathrm{P}(\mathrm{A})}$$

where f_h *is given by* (2.4.3) *and where* K *is universal.*

Our first task should be to give natural examples of situations where (2.7.1), (2.7.2) occur.

Proposition **2.7.2.** — *Consider the probability* ν_1 *on* $\mathbf{R}$, *of density* $\frac{1}{2}e^{-|x|}$ *with respect to the Lebesgue measure. Then the function* $h(x,y) = \min(|x-y|, |x-y|^2)$ *satisfies* (2.7.1), (2.7.2) *(for* ν_1, *rather than* μ*).*

Proof. — For a subset C of $\mathbf{R}$, and $t > 0$, let us set $C_t = \{x \in \mathbf{R}; d(x, C) \leqslant t\}$. To prove (2.7.1), (2.7.2), it suffices to show that

$$\nu_1(C_t) \geqslant c(\nu_1(C), t).$$

This is proved in [T4] using rearrangements.

We sketch below a simpler alternative argument to prove the weaker result

$$\nu_1(C_t) \geqslant c(\nu_1(C), t/2). \tag{2.7.4}$$

(The reader should observe that this suffices to prove that $h/4$ satisfies (2.7.1), (2.7.2).)

First, we reduce to the case where C is a finite union of intervals. Setting

$$u(t) = \inf\{|x|; x \in \overline{C}_t/C_t\},$$

it should be clear that

$$\frac{d\nu_1}{dt}(C_t) \geqslant \frac{1}{2}\exp(-u(t)). \tag{2.7.5}$$

By definition of $u(t)$, we see that the interval $[-u(t), u(t)]$ is either contained in the closure of C_t, or else it does not meet C_t. Thereby, we have either

$$\nu_1(C_t) \geqslant 1 - 2\nu_1([u(t), \infty)) = 1 - e^{-u(t)}$$

or else $$\nu_1(C_t) \leqslant 2\nu_1([u(t), \infty)) = e^{-u(t)}$$

so that, in any case

$$e^{-u(t)} \geqslant \min(\nu_1(C_t), 1 - \nu_1(C_t)).$$

Combining with (2.7.5) shows that as long as $\nu_1(C_t) \leqslant 1/2$, we have $\frac{d}{dt}(\log \nu_1(C_t)) \geqslant 1/2$, so that $\nu_1(C_t) \geqslant e^{t/2}\,\nu_1(C)$. Similar considerations complete the proof. □

Other examples can be generated using Proposition 2.7.2 and the following simple observation.

Proposition **2.7.3.** — *Consider a probability space* (Ω, μ) *and a function h on* Ω^2, *that satisfies* (2.7.1), (2.7.2). *Consider a measurable map* η *from* Ω *to a measured space* Ω', *and the measure* $\mu' = \eta(\mu)$ *on* Ω'. *Consider a function* h' *on* Ω'^2 *such that*

$$\forall x, y \in \Omega, \quad h'(\eta(x), \eta(y)) \leqslant h(x,y). \tag{2.7.6}$$

Then h', μ' *satisfy* (2.7.1), (2.7.2).

Proof. — This is obvious using the relations

$$\mu(\eta^{-1}(B)) = \mu'(B), \qquad h(x, \eta^{-1}(B)) \geqslant h'(\eta(x), B). \quad \square$$

The use of Propositions 2.7.2 and 2.7.3 will allow the construction of a wide class of examples.

Proposition **2.7.4.** — *Consider a convex symmetric function* $\psi \geqslant 0$ *on* $\mathbf{R}$, *with* $\lim_{x\to\infty} \psi'(x) = \infty$, *and the probability* ν_ψ *of density* $a_\psi\, e^{-\psi(x)}$ *with respect to the Lebesgue measure, where* a_ψ *is the normalizing constant. Then there is a constant* $K(\psi)$ *depending on* ψ *only such that the function* $h(x,y)$ *on* $\mathbf{R}^2$ *given by*

$$(2.7.7) \qquad |x-y| \leqslant 1 \Rightarrow h(x,y) = \frac{1}{K(\psi)}\,|x-y|^2$$

$$(2.7.8) \qquad |x-y| \geqslant 1 \Rightarrow h(x,y) = \frac{1}{K(\psi)}\,\psi\left(\frac{1}{K(\psi)}\,|x-y|\right)$$

satisfies (2.7.1), (2.7.2) *with respect to* ν_ψ.

Proof of Proposition 2.7.4. — Consider the nondecreasing map η from $\mathbf{R}$ to $\mathbf{R}$ that transports ν_1 to ν_ψ. Thus

$$(2.7.9) \qquad \int_{\eta(x)}^{\infty} a_\psi\, e^{-\psi(t)}\, d\lambda(t) = \int_x^{\infty} \frac{1}{2} e^{-|t|}\, d\lambda(t).$$

By Proposition 2.7.2, 2.7.3, it suffices to show that

$$(2.7.10) \qquad h(\eta(x), \eta(y)) \leqslant \min(|x-y|, |x-y|^2).$$

It is simple to see that (2.7.10) will follow from (2.7.7), (2.7.8) (with a suitable choice of the constant there) provided we can show that

$$(2.7.11) \qquad |\eta(x) - \eta(y)| \leqslant K(\psi)\,|x-y|$$

$$(2.7.12) \qquad \psi\left(\frac{1}{K(\psi)}\,|\eta(x) - \eta(y)|\right) \leqslant |x-y|.$$

There, as in the rest of this proof, $K(\psi)$ denotes a constant depending on ψ only, that may vary at each occurrence.

To prove (2.7.11), it suffices to prove that $\eta'(x)$ is bounded when $x > 0$. Differentiating (2.7.9), we get $a_\psi\, \eta'(x)\, e^{-\psi(\eta(x))} = e^{-x}/2$, and plugging back in (2.7.9), we get

$$\eta'(x) = \int_{\eta(x)}^{\infty} e^{\psi(\eta(x)) - \psi(t)}\, dt.$$

Thereby, it suffices to show that

$$\sup_{u \geqslant 0} \int_u^\infty e^{-\psi(t)+\psi(u)}\, dt < \infty.$$

Given $u_0 > 0$, the supremum for $u \leqslant u_0$ is certainly bounded. On the other hand, for $u \geqslant u_0$, by convexity of ψ we have $\psi(t) - \psi(u) \geqslant (t-u)\,\psi'(u_0)$, so it suffices to choose u_0 with $\psi'(u_0) > 0$.

We now turn to the proof of (2.7.12). It suffices to prove that, for $y > x \geqslant 0$, we have $\psi((\eta(y) - \eta(x))/\mathrm{K}(\psi)) \leqslant y - x$. Setting $a = \psi^{-1}(y-x)$, it suffices to show that $\eta(x + \psi(a)) \leqslant \eta(x) + \mathrm{K}(\psi)\, a$, i.e., that

$$a_\psi \int_{\eta(x)+\mathrm{K}(\psi)a}^\infty e^{-\psi(t)}\, dt \leqslant \frac{1}{2} e^{-x-\psi(a)}. \tag{2.7.13}$$

First, we note that, since $\psi(t) \geqslant \psi(y) + (t-y)\,\psi'(y)$, we have, for $y > 0$,

$$\int_y^\infty e^{-\psi(t)}\, dt \leqslant \frac{1}{\psi'(y)} e^{-\psi(y)},$$

so that

$$a_\psi \int_{\eta(x)+2a} e^{-\psi(t)}\, dt \leqslant \frac{a_\psi}{\psi'(\eta(x)+2a)} e^{-\psi(\eta(x)+2a)}. \tag{2.7.14}$$

Also,

$$\frac{1}{2} e^{-x} = a_\psi \int_{\eta(x)}^\infty e^{-\psi(t)}\, dt \geqslant a_\psi\, a e^{-\psi(\eta(x)+a)}. \tag{2.7.15}$$

Since $\psi'(y)$ increases for $y > 0$, we have $\psi(\eta(x)+2a) \geqslant \psi(\eta(x)+a) + \psi(a)$. Thus, from (2.7.14), (2.7.15) we see that (2.7.13) holds provided $\mathrm{K}(\psi) \geqslant 2$, $a\psi'(\eta(x)+2a) \geqslant a_\psi$.

On the other hand, using again convexity, we see that

$$\int_{\eta(x)+a}^\infty e^{-\psi(t)}\, dt = \int_{\eta(x)}^\infty e^{-\psi(v+a)}\, dv \leqslant e^{-a\psi'(\eta(x))} \int_{\eta(x)}^\infty e^{-\psi(v)}\, dv$$
$$\leqslant \frac{1}{2a_\psi} e^{-a\psi'(\eta(x))-x}.$$

Thereby, if $\mathrm{K}(\psi) \geqslant 1$, (2.7.13) will hold provided $a\psi'(\eta(x)) \geqslant \psi(a)$, and in particular if $\eta(x) \geqslant a$.

Thus we can assume $\eta(x) \leqslant a$, $a\psi'(\eta(x)+2a) \leqslant a_\psi$. This means that a and x remain bounded; but the conclusion is then obvious. □

It is of particular interest to consider the case where $\psi(x) = x^2$, so that ν_ψ is Gaussian. In this case, Proposition 2.7.4 shows that one can take $h(x,y) = \mathrm{K}^{-1}(x-y)^2$. This recovers the concentration of measure for the Gauss space, as expressed by (1.7). There is, however, a big loss of information in (2.7.10); and the result obtained by taking

$$h(x,y) = \min(|\eta^{-1}(x) - \eta^{-1}(y)|, (\eta^{-1}(x) - \eta^{-1}(y))^2)$$

is rather more precise than (1.7).

The induction step of the proof of Theorem 2.7.1 reduces to the following.

Proposition **2.7.5.** — *There exists a universal constant* L *with the following property. Consider a function g on* Ω, *and define*

$$\hat{g}(x) = \inf_{y\in\Omega} g(y) + \frac{1}{\mathrm{L}} h(x,y). \qquad \textbf{(2.7.16)}$$

Then, under (2.7.1), (2.7.2), *we have*

$$\int_\Omega e^{\hat{g}}\, d\mu \int_\Omega e^{-g}\, d\mu \leqslant 1. \qquad \textbf{(2.7.17)}$$

Let us recall that we denote by ν_1 the probability measure on $\mathbf{R}$ of density $e^{-|x|}/2$ with respect to the Lebesgue measure. During the end of this section, for $x \in \mathbf{R}$ we set $\varphi(x) = \min(|x|, x^2)$.

The proof of Proposition 2.7.5 is considerably simplified by the following observation.

Proposition **2.7.6.** — *Consider a function g on* Ω, *and* $\hat{g}$ *given by* (2.7.16). *Then we can find two nonincreasing functions* g_1, g_2 *on* $\mathbf{R}$ *with the following properties*

$$\int_\Omega e^{\hat{g}}\, d\mu = \int_{\mathbf{R}} e^{g_2}\, d\nu_1; \quad \int_\Omega e^{-g}\, d\mu = \int_{\mathbf{R}} e^{-g_1}\, d\nu_1 \qquad \textbf{(2.7.18)}$$

$$\forall\, x \in \mathbf{R}, \quad g_2(x) \leqslant \inf_{y\in\mathbf{R}} g_1(y) + \frac{1}{\mathrm{L}} \varphi(|x-y|). \qquad \textbf{(2.7.19)}$$

In particular, this implies that we have reduced to the case $\mu = \nu_1$, g nonincreasing, $h(x,y) = \varphi(|x-y|)$.

Proof. — We define, for $y \in \mathbf{R}$,

$$g_1(y) = \inf\{\, t;\ \mu(\{g \leqslant t\}) \geqslant \nu_1([y,\infty))\,\}, \qquad \textbf{(2.7.20)}$$

$$g_2(x) = \sup\{\, u;\ \mu(\{\hat{g} \geqslant u\}) \geqslant \nu_1((-\infty, x])\,\}. \qquad \textbf{(2.7.21)}$$

Thereby both g_1, g_2 are nonincreasing; it should be obvious that (2.7.18) holds. We prove (2.7.19). Consider $x \leqslant y$. By (2.7.20), we have $\mu(\mathrm{B}) \geqslant \nu_1([y, \infty))$, where $\mathrm{B} = \{ g \leqslant g_1(y) \}$. By (2.7.1), (2.7.2), we have

$$\begin{aligned}\mu(\{ h(\cdot, \mathrm{B}) \leqslant \varphi(t) \}) &\geqslant c(\mu(\mathrm{B}), t)\\ &\geqslant c(\nu_1([y, \infty)), t)\\ &= \nu_1([y - t, \infty)).\end{aligned}$$

Since $\hat{g} \leqslant g_1(y) + \varphi(t)/\mathrm{L}$ on the set $\{ h(\cdot, \mathrm{B}) \leqslant \varphi(t) \}$, we get

$$\mu\left(\left\{\hat{g} \leqslant g_1(y) + \frac{\varphi(t)}{\mathrm{L}}\right\}\right) \geqslant \nu_1([y - t, \infty)).$$

On the other hand, by (2.7.21) we have

$$\mu(\{ \hat{g} \geqslant g_2(x) \}) \geqslant \nu_1((-\infty, x]).$$

Thus, if $t > y - x$, we have $g_2(x) < g_1(y) + \varphi(t)/\mathrm{L}$. Thus $g(x) \leqslant g(y) + \varphi(y - x)/\mathrm{L}$, and (2.7.19) follows. □

We next show that we have reduced the proof of Proposition 2.7.5 to the following.

Proposition **2.7.7.** — *There exists a universal constant* L *with the following property. Consider a nonincreasing function f on* **R**, *with $f(0) = 0$. Define*

$$\hat{f}(x) = \inf_{y \in \mathbf{R}} f(y) + \frac{1}{\mathrm{L}} \varphi(|x - y|).$$

Then, if f has a Lipschitz constant $\leqslant 2/\mathrm{L}$, we have

$$\int e^{\hat{f}}\, d\nu_1 \int e^{-f}\, d\nu_1 \leqslant 1.$$

We prove the claim stated before Proposition 2.7.7. In view of Proposition 2.7.6 and (2.7.19), it suffices to prove that $\int e^{\hat{g}_1}\, d\nu_1 \int e^{-g_1}\, d\nu_1 \leqslant 1$, where $\hat{g}_1$ is given by the right-hand side of (2.7.19). Define now

$$\textbf{(2.7.22)} \qquad f(y) = \sup_{x \in \mathbf{R}} \hat{g}_1(x) - \frac{1}{\mathrm{L}} \varphi(|x - y|).$$

Since for all x and y we have $\hat{g}_1(x) \leqslant g_1(y) + \frac{1}{\mathrm{L}} \varphi(|x - y|)$, we see that $f(y) \leqslant g_1(y)$. Thus, $\int e^{-f}\, d\nu_1 \geqslant \int e^{-g_1}\, d\nu_1$. Also, by (2.7.22), we have $\hat{g}_1(x) \leqslant f(y) + \frac{1}{\mathrm{L}} \varphi(|x - y|)$

for all x, y, so that $\hat{g}_1 \leqslant \hat{f}$. Thereby it suffices to prove that $\int e^{\hat{f}} d\nu_1 \int e^{-f} d\nu_1 \leqslant 1$. The condition $f(0) = 0$ is certainly not restrictive, and f has a Lipschitz constant $\leqslant 2/L$ by (2.7.22) since φ has a Lipschitz constant $\leqslant 2$.

Upon seeing the result of [T4] exposed in a seminar, B. Maurey produced a rather magic proof of Proposition 2.7.7 [Mau2]. The proof we will give is more in the spirit of the arguments of the present paper, and is likely to be more instructive as it prepares for the considerably more delicate results to be presented in Chapter 4. We start by a simple lemma.

Lemma **2.7.8.** — *Consider a nonincreasing function u on* **R**, *such that $u(0) = 0$. Then*

$$\int_{\mathbf{R}^-} u^2 \, d\nu_1 \leqslant K \sum_{k \geqslant 1} (u(-k) - u(-k+1))^2 e^{-k}.$$

Proof. — For simplicity we set $u_k = u(-k)$. Thus

$$\int_{\mathbf{R}^-} u^2 \, d\nu_1 \leqslant S =: \sum_{k \geqslant 1} u_k^2 e^{-k+1}.$$

Since $u_k^2 \leqslant 2u_{k-1}^2 + 2(u_k - u_{k-1})^2$, we have

$$S \leqslant 2 \sum_{k \geqslant 1} u_{k-1}^2 e^{-k+1} + 2 \sum_{k \geqslant 1} (u_k - u_{k-1})^2 e^{-k+1}.$$

But since $u_0 = 0$, the first sum is exactly $2S/e$, so that

$$S\left(1 - \frac{2}{e}\right) \leqslant 2e \sum_{k \geqslant 1} (u_k - u_{k-1})^2 e^{-k}. \quad \square$$

During the proof of Proposition 2.7.7, we will consider another number $1 \leqslant M \leqslant L$. The numbers M, L will be chosen later. The crucial part of the proof of Proposition 2.7.7 is as follows.

Proposition **2.7.9.** — *Consider a non-increasing function u on* **R**, *with $u(0) = 0$. Assume that $|u| \leqslant 1/M$, and set $\hat{u}(x) = \inf_y u(y) + \varphi(|x - y|)/L$. Then, if $L \geqslant KM$, we have*

$$\text{(2.7.23)} \qquad \int_{\mathbf{R}} (u - \hat{u}) \, d\nu_1 \geqslant \frac{M}{K} \int_{\mathbf{R}} u^2 \, d\nu_1.$$

Moreover, if $M \geqslant K$, *we have*

$$\text{(2.7.24)} \qquad \int_{\mathbf{R}} e^{\hat{u}} \, d\nu_1 + \int_{\mathbf{R}} e^{-u} \, d\nu_1 \leqslant 2 - \frac{M}{K} \int_{\mathbf{R}} u^2 \, d\nu_1.$$

Proof. — To prove (2.7.23), it suffices to prove it when the right-hand side is replaced by $\int_{\{u\geqslant 0\}} u^2\,d\mu$ $\left(\text{resp.} \int_{\{u\leqslant 0\}} u^2\,d\mu\right)$. The arguments for these two cases are similar so we treat the first case only. We set $u_k = u(-k)$, so that $u_k \leqslant M^{-1}$, and $M(u_k - u_{k-1}) \leqslant 1$. We set $N_k = [2/M(u_k - u_{k-1})]$. Thus we have $N_k \geqslant 2$ and

$$\textbf{(2.7.25)} \qquad \frac{1}{2MN_k} \leqslant u_k - u_{k-1} \leqslant \frac{1}{MN_k}.$$

For $k \geqslant 1$, $\ell \geqslant 0$ we set $a_{k,\ell} = -k + 1 - \ell/N_k$, and $u_{k,\ell} = u(a_{k,\ell})$. Thus $u_{k,0} = u_{k-1}$, $u_{k,N_k} = u_k$. For $k \geqslant 1$, $1 \leqslant \ell \leqslant N_k$, we consider the subset $R_{k,\ell}$ of $\mathbf{R}^2$ given by

$$R_{k,\ell} =]a_{k,\ell+1}, a_{k,\ell}[\times \left]\min\left(u_{k,\ell}, u_{k,\ell-1} + \frac{4}{LN_k^2}\right), u_{k,\ell}\right[.$$

We observe that no point belongs to more than two intervals $]a_{k,\ell+1}, a_{k,\ell}[$, for $1 \leqslant \ell \leqslant N_k$, $k \geqslant 1$, so that the rectangles $R_{k,\ell}$ have the same property. Since $u(x) \geqslant u_{k,\ell}$ for $x \leqslant a_{k,\ell}$, $R_{k,\ell}$ is below the graph of u; but, since $u(a_{k,\ell-1}) = u_{k,\ell-1}$, we have $\hat{u}(x) \leqslant u_{k,\ell-1} + 4/LN_k^2$ on $[a_{k,\ell+1}, a_{k,\ell}]$. Thus $R_{k,\ell}$ is above the graph of $\hat{u}$, and hence

$$\int_{\mathbf{R}} (u - \hat{u})\,d\nu_1 \geqslant \frac{1}{2} \sum_{k\geqslant 1} \sum_{1\leqslant \ell\leqslant N_k} \nu_1([a_{k,\ell+1}, a_{k,\ell}]) \left(u_{k,\ell} - \min\left(u_{k,\ell}, u_{k,\ell-1} + \frac{4}{LN_k^2}\right)\right).$$

Since $\nu_1([a_{k,\ell+1}, a_{k,\ell}]) \geqslant e^{-k}/KN_k$, we have

$$\begin{aligned}\int_{\mathbf{R}} (u - \hat{u})\,d\nu_1 &\geqslant \frac{1}{K} \sum_{k\geqslant 1} \frac{e^{-k}}{N_k} \sum_{1\leqslant \ell\leqslant N_k} \left(u_{k,\ell} - u_{k,\ell-1} - \frac{4}{LN_k^2}\right) \\ &= \frac{1}{K} \sum_{k\geqslant 1} \frac{e^{-k}}{N_k} \left(u_{k,N_k} - u_{k,0} - \frac{4}{LN_k}\right) \\ &= \frac{1}{K} \sum_{k\geqslant 1} \frac{e^{-k}}{N_k} \left(u_k - u_{k-1} - \frac{4}{LN_k}\right) \\ &\geqslant \frac{1}{K} \sum_{k\geqslant 1} e^{-k}(u_k - u_{k-1})^2\end{aligned}$$

by (2.7.25), and provided $L \geqslant 16M$. Thus, (2.7.23) follows from Lemma 2.7.8.

To prove (2.7.24), we use the inequality $e^x \leqslant 1 + x + x^2$ for $|x| \leqslant 1$. Thus

$$\int_{\mathbf{R}} e^{\hat{u}}\,d\nu_1 + \int_{\mathbf{R}} e^{-u}\,d\nu_1 \leqslant 2 + \int_{\mathbf{R}} (\hat{u} - u)\,d\nu_1 + \int_{\mathbf{R}} u^2\,d\nu_1 + \int_{\mathbf{R}} \hat{u}^2\,d\nu_1.$$

Now,

$$\hat{u}^2 \leqslant 2u^2 + 2(u - \hat{u})^2 \leqslant 2u^2 + \frac{2}{M}(u - \hat{u}) \leqslant 2u^2 + \frac{1}{2}(u - \hat{u})$$

provided $M \geqslant 4$, and thus

$$\int_{\mathbf{R}} e^{\hat{u}}\, d\nu_1 + \int_{\mathbf{R}} e^{-u}\, d\nu_1 \leqslant 2 + 3\int_{\mathbf{R}} u^2\, d\nu_1 - \frac{1}{2}\int_{\mathbf{R}} (u - \hat{u})\, d\nu_1$$

and the result follows from (2.7.23). □

Proof of Proposition 2.7.7. — We observe that, for $a \in \mathbf{R}$, we have $a(2 - a) \leqslant 1$. Thus it suffices to show that

$$\int_{\mathbf{R}} e^{\hat{f}}\, d\nu_1 + \int_{\mathbf{R}} e^{-f}\, d\nu_1 \leqslant 2. \tag{2.7.26}$$

We set $u = \min(1/M, \max(f, -1/M))$. Thus

$$\int_{\mathbf{R}} e^{-f}\, d\nu_1 \leqslant \int_{\mathbf{R}} e^{-u}\, d\nu_1 + \int_b^{\infty} (e^{-f} - e^{1/M})\, d\nu_1 \tag{2.7.27}$$

where $f(b) = -1/M$. We observe that if $\hat{u}(x) < 1/M$, then $\hat{f}(x) \leqslant \hat{u}(x)$. Indeed if $\hat{u}(x) < 1/M$, then given ε with $\hat{u}(x) < \varepsilon < 1/M$, there exists y with

$$u(y) + L^{-1}\varphi(|x - y|) < \varepsilon.$$

Thus $u(y) < 1/M$, and $f(y) \leqslant u(y)$, so that $\hat{f}(x) < \varepsilon$. Then, if c is the largest constant so that $\hat{f}(c) = 1/M$, we have

$$\int_{\mathbf{R}} e^{\hat{f}}\, d\nu_1 \leqslant \int_{\mathbf{R}} e^{\hat{u}}\, d\nu_1 + \int_{-\infty}^{c} (e^{\hat{f}} - e^{1/M})\, d\nu_1. \tag{2.7.28}$$

Since $f(0) = 0$, we have $c < 0 < b$. Since f has a Lipschitz constant $\leqslant 2/L$, we have, for $x \geqslant b$,

$$-f(x) \leqslant \frac{1}{M} + \frac{2}{L}(x - b),$$

and thus

$$\int_b^{\infty} e^{-f}\, d\nu_1 \leqslant \int_b^{\infty} \frac{1}{2} e^{1/M + 2(x-b)/L} e^{-x}\, dx$$

$$= \frac{e^{1/M}}{1 - 2/L}\, \nu_1([b, \infty)).$$

Hence,

$$\int_b^\infty (e^{-f} - e^{1/M})\, d\nu_1 \leqslant e^{1/M}\left(\frac{1}{1-2/L} - 1\right)\nu_1([b, \infty])$$
$$\leqslant \frac{K}{L}\nu_1([b, \infty[)$$
$$\leqslant \frac{KM^2}{L}\int_{\mathbf{R}} u^2\, d\nu_1,$$

since $u(x) = -1/M$ for $x \geqslant b$. Using (2.7.27), (2.7.28), and making a similar computation for $\int_{-\infty}^{c} (e^{\hat{f}} - e^{1/M})\, d\nu_1$ yields

$$\int_{\mathbf{R}} e^{\hat{f}}\, d\nu_1 + \int_{\mathbf{R}} e^{-f}\, d\nu_1 \leqslant \int_{\mathbf{R}} e^{\hat{u}}\, d\nu_1 + \int_{\mathbf{R}} e^{-u}\, d\nu_1 + \frac{KM^2}{L}\int_{\mathbf{R}} u^2\, d\nu_1.$$

It then follows from (2.7.24) that (2.7.26) holds provided $M \geqslant K$, $L \geqslant KM^2$. □

It would be of interest to understand exactly which are the functions φ such that, if one sets

$$\hat{f}(x) = \inf_{y \in \mathbf{R}} f(y) + \varphi(x - y),$$

then $\int e^{\hat{f}}\, d\nu_1 \int e^{-f}\, d\nu_1 \leqslant 1$. On the other hand, the situation is considerably clearer if one considers the standard Gaussian density γ_1 rather than ν_1. In that case, the obvious adaptation of Maurey's argument shows that if $\alpha \geqslant 1$, and if $\hat{f}(x) = \inf_{y \in \mathbf{R}} \alpha f(y) + \frac{\alpha}{2(\alpha+1)}(x-y)^2$, then $\int e^{\hat{f}}\, d\gamma_1 \left(\int e^{-f}\, d\gamma_1\right)^\alpha \leqslant 1$. Thereby, by induction, and with the notation of (1.7), we get

$$\gamma_N(A_t) \geqslant 1 - \frac{1}{\gamma_N(A)^\alpha} e^{-\frac{\alpha t^2}{2(\alpha+1)}},$$

hence, by optimization over α and for $t \geqslant \sqrt{2\log(1/\gamma_N(A))}$,

$$\gamma_N(A_t) \geqslant 1 - \exp -\frac{1}{2}\left(t - \sqrt{2\log(1/\gamma_N(A))}\right)^2$$

which is not so far from (1.7).

3. Control by q points

In Section 2 the basic theme was that the " distance " from a point x to a set A was measured by how many coordinates of x can be " captured " by a single point of A. The theme of the present section is that we allow several points of A to capture as many coordinates of x as possible.

3.1. *Basic result*

Consider an integer $q \geq 2$. For subsets $A_1, \ldots, A_q$ of Ω^N, and $x \in \Omega^N$, we set

$$f(A_1, \ldots, A_q, x) = \inf\{\operatorname{card}\{i \leq N : x_i \notin \{y_i^1, \ldots, y_i^q\}\}; y^1 \in A_1, \ldots, y^q \in A_q\}. \tag{3.1.1}$$

Theorem **3.1.1.** — *We have*

$$\int q^{f(A_1, \ldots, A_q, x)}\, dP(x) \leq \frac{1}{\prod_{i \leq q} P(A_i)}. \tag{3.1.2}$$

In particular,

$$P(\{f(A, \ldots, A, x) \geq k\}) \leq \frac{1}{q^k\, P(A)^q}. \tag{3.1.3}$$

The induction method will reduce this statement to a simple fact about functions.

Lemma **3.1.2.** — *Consider a function g on Ω, such that $1/q \leq g \leq 1$. Then*

$$\int_\Omega \frac{1}{g}\, d\mu \left(\int_\Omega g\, d\mu\right)^q \leq 1. \tag{3.1.4}$$

Proof. — We could use the extreme point argument of Lemma 2.1.2. One alternative method is as follows. Observing that $\log x \leq x - 1$, to prove that $ab^q \leq 1$ it suffices to show that $a + qb \leq q + 1$. Thus, it suffices to show that

$$\int_\Omega \frac{1}{g}\, d\mu + q \int_\Omega g\, d\mu \leq q + 1.$$

But this is obvious since $x^{-1} + qx \leq q + 1$ for $q^{-1} \leq x \leq 1$. □

Corollary **3.1.3.** — *Consider functions g_i on Ω, $g_i \leq 1$. Then*

$$\int_\Omega \min_{i \leq q}\left(q, \frac{1}{g_i}\right) d\mu \prod_{i \leq q} \int g_i\, d\mu \leq 1. \tag{3.1.5}$$

Proof. — Set $g = (\min_{i \leq q}(q, g_i^{-1}))^{-1}$, observe that $g_i \leq g$, and use (3.1.4).

We now prove Theorem 3.1.1 by induction over N. For $N = 1$, the result follows from (3.1.5), taking $g_i = 1_{A_i}$.

We now assume that Theorem 3.1.1 has been proved for N, and we prove it for $N + 1$. Consider sets $A_1, \ldots, A_q$ of Ω^{N+1}. For $\omega \in \Omega$, we define the sets $A_i(\omega)$ as in (2.1.5) and we consider the projection B_i of A_i on Ω^N. The basic observation is that

$$f(A_1, \ldots, A_q, (x, \omega)) \leq 1 + f(B_1, \ldots, B_q, x) \tag{3.1.6}$$

and that, if $j \leqslant q$,

$$f(A_1, \ldots, A_q, (x, \omega)) \leqslant f(C_1, \ldots, C_q, x),$$

where $C_i = B_i$ for $i \neq j$, $C_j = A_j(\omega)$.

If we set $g_i(\omega) = P(A_i(\omega))/P(B_i)$, using Fubini theorem and the induction hypothesis, we are reduced to show that

$$\int_\Omega \min\left(q, \min_{i \leqslant q} \frac{1}{g_i(\omega)}\right) \leqslant \frac{1}{\prod_{i \leqslant q} \int_\Omega g_i \, d\mu},$$

which is (3.1.5). □

3.2. *Sharpening*

Given $\alpha > 1$, we can now, in the spirit of Proposition 2.2.1, look for the largest number $a = a(q, \alpha)$ for which we can prove that

$$\textbf{(3.2.1)} \qquad \int^* a(q, \alpha)^{f(A_1, \ldots, A_q, x)} \, dP(x) \leqslant \frac{1}{\prod_{i \leqslant q} P(A_i)^\alpha}.$$

Following the proof of Theorem 3.1.1, we see that we can take for $a(q, \alpha)$ the unique number $x > 1$ such that

$$\textbf{(3.2.2)} \qquad x + q\alpha x^{-1/\alpha} = 1 + q\alpha.$$

It then follows from (3.2.1) that

$$\textbf{(3.2.3)} \qquad P(\{f(A, \ldots, A, x) \geqslant k\}) \leqslant \inf_{\alpha \geqslant 1} \frac{a(q, \alpha)^{-k}}{P(A)^{q\alpha}}.$$

There is no obvious way to compute the right-hand side of (3.2.3). However, for large q, we have the following, that improves upon (3.1.2) for large values of k ($k \gg q \log q$).

Proposition **3.2.1.** — *There exists a universal constant q_0 such that, if $q \geqslant q_0$, we have*

$$\textbf{(3.2.4)} \qquad P(\{f(A, \ldots, A, x) \geqslant k\}) \leqslant \left(\frac{e}{(e-1)\, q \log q}\right)^k \left(\frac{1}{P(A)}\right)^{q \log q}.$$

Proof. — We take $\alpha = \log q$, and we show that, for q large enough,

$$a(q, \alpha) \geqslant a := 1 + \left(1 - \frac{1}{e}\right) q \log q.$$

For large q, we have $a \geqslant q$, so that $a^{1/\alpha} \geqslant e$, hence

$$a - 1 \leqslant q\alpha \left(1 - \frac{1}{a^{1/\alpha}}\right)$$

and thus $a + q\alpha a^{-1/\alpha} \leqslant 1 + q\alpha$. □

It is interesting to note that Proposition 3.2.1 is rather sharp. Consider the case where $\Omega = \{0, 1\}$, and where μ gives weight p to 1 ($p \leqslant 1/2$). Assume for simplicity that $r = p\mathrm{N}$ is an integer. Consider the set $\mathrm{A} = \{x \in \Omega^{\mathrm{N}};\ \sum_{i \leqslant \mathrm{N}} x_i \leqslant r\}$. Then $\mathrm{P(A)}$ is of order $1/2$. Considering $s = rq + k$, we clearly have that $\sum_{i \leqslant \mathrm{N}} x_i = s$ implies $f(\mathrm{A}, \ldots, \mathrm{A}, x) \geqslant k$. Thus $\mathrm{P}(\{f(\mathrm{A}, \ldots, \mathrm{A}, x) \geqslant k\}) \geqslant p^s(1-p)^{\mathrm{N}-s}\binom{\mathrm{N}}{s}$.

When $s \leqslant \mathrm{N}/2$, we have $\binom{\mathrm{N}}{s} \geqslant (\mathrm{N}/2s)^s$, so that

$$\mathrm{P}(\{f(\mathrm{A}, \ldots, \mathrm{A}, x) \geqslant k\}) \geqslant \left(\frac{p\mathrm{N}}{2s}\right)^s e^{-2p\mathrm{N}} \geqslant \left(\frac{r}{2es}\right)^s \geqslant \left(\frac{1}{2e(q + k/r)}\right)^{rq+k}.$$

If we take $k \geqslant q \log q$, fixed, and then r of order $k/q \log q$, we get a lower bound of order $(1/\mathrm{K}q \log q)^k$.

3.3. *Penalties*

The result of this section is the one single major theorem of Part I that has not been motivated by direct applications. Rather, it has been motivated by a desire of symmetry with Sections 2.7 and 4.4.

We consider a " penalty function " $h(\omega, \omega^1, \ldots, \omega^q)$ on Ω^{q+1}. We assume $h \geqslant 0$ and

(3.3.1) $$\omega \in \{\omega^1, \ldots, \omega^q\} \Rightarrow h(\omega, \omega^1, \ldots, \omega^q) = 0.$$

For subsets $\mathrm{A}_1, \ldots, \mathrm{A}_q$ of Ω^{N}, we consider

(3.3.2) $$f_h(\mathrm{A}_1, \ldots, \mathrm{A}_q, x) = \inf\{\sum_{i \leqslant \mathrm{N}} h(x_i, y_i^1, \ldots, y_i^q);\ y^1 \in \mathrm{A}_1, \ldots, y^q \in \mathrm{A}_q\}.$$

The case considered in Section 3.1 is where $h(\omega, \omega^1, \ldots, \omega^q) = 1$, unless $\omega \in \{\omega^1, \ldots, \omega^q\}$, in which case it is zero.

Given subsets $\mathrm{B}_1, \ldots, \mathrm{B}_q$ of Ω, we set

(3.3.3) $$h(\omega, \mathrm{B}_1, \ldots, \mathrm{B}_q) = \inf\{h(\omega, \omega^1, \ldots, \omega^q);\ \omega^1 \in \mathrm{B}_1, \ldots, \omega^q \in \mathrm{B}_q\}.$$

To control how large h is, we will consider a nonincreasing function γ from $]0, 1]$ to $\mathbf{R}^+$, and assume that

(3.3.4) $$\forall\, \omega \in \Omega,\quad \forall\, \mathrm{B}_1, \ldots, \mathrm{B}_q \subset \Omega,\quad h(\omega, \mathrm{B}_1, \ldots, \mathrm{B}_q) \leqslant \sum_{i \leqslant q} \gamma(\mu(\mathrm{B}_i)).$$

A typical case where this condition is satisfied is when

$$h(\omega, \omega^1, \ldots, \omega^q) = \sum_{i \leqslant q} h_i(\omega^i)$$

for functions h_i that satisfy the tail condition $\mu(\{h_i \geqslant \gamma(t)\}) \leqslant t$ and when γ is left continuous. Indeed, if $t < \mu(B_i)$, then B_i contains a point y_i with $h_i(y_i) < \gamma(t)$.

We consider a convex function $\theta :]0, 1] \to \mathbf{R}^+$, and we make the mild technical assumption that the inverse function ξ satisfies

$$\text{(3.3.5)} \qquad |\xi'(x+1)| \geqslant \frac{1}{3} |\xi'(x)|.$$

$\left(\text{We put } \frac{1}{3} \text{ rather than } \frac{1}{2} \text{ simply to allow the case } \xi(x) = e^{-x}.\right)$

Theorem **3.3.1.** — *There exists a universal constant* K *such that for* $q \geqslant$ K, *under* (3.3.1), (3.3.4), (3.3.5), *if, for each* $s \leqslant 1$, *we have*

$$\text{(3.3.6)} \qquad \int_s^1 \gamma^{-1}(\theta(s) - \theta(t))\, d\lambda(t) \leqslant \frac{\log(q/\mathrm{K})}{q\,|\theta'(s)|},$$

then, for each subsets $A_1, \ldots, A_q$ *of* Ω^N, *we have*

$$\text{(3.3.7)} \qquad \int e^{f h(A_1, \ldots, A_q, x)}\, dP(x) \leqslant \exp\big(\sum_{i \leqslant q} \theta(P(A_i))\big).$$

To understand (3.3.6) better, we observe that the term $\theta'(s)$ arises simply because $\theta(s) - \theta(t)$ resembles $(s-t)\,\theta'(s)$ for t close to s. Actually, since $\theta(s) - \theta(t) \leqslant (s-t)\,\theta'(s)$, change of variable and Lebesgue's theorem show that (3.3.6) implies that $\int_0^\infty \gamma^{-1}(u)\, du \leqslant q^{-1}\log(q/\mathrm{K})$. In the case where γ is constant, one can take $h(\omega, \omega^1, \ldots, \omega^q) = q\gamma$ whenever $\omega \notin \{\omega^1, \ldots, \omega^q\}$ (and otherwise $h = 0$). Then the integral in (3.3.6) has to be interpreted as $|\{t : s \leqslant t;\ \theta(t) \geqslant \theta(s) - \gamma\}|$. When $\theta(x) = -\log x$, this is $s(e^\gamma - 1)$, and (3.3.6) holds whenever $\gamma \leqslant q^{-1}\log(q/\mathrm{K})$. We then almost recover Theorem 3.1.1.

To prove Theorem 3.3.1, it suffices, by the induction method, to prove the following.

Proposition **3.3.2.** — *There exists a universal* K *such that, under conditions* (3.3.1), (3.3.4), (3.3.5), (3.3.6), *if we consider functions* $(u_i)_{i \leqslant q}$ *on* Ω, $0 \leqslant u_i \leqslant 1$, *and define*

$$\text{(3.3.8)} \qquad v(\omega) = \inf_{\omega^1, \ldots, \omega^q} \sum_{i \leqslant q} \theta(u_i(\omega^i)) + h(\omega, \omega^1, \ldots, \omega^q),$$

then we have

$$\text{(3.3.9)} \qquad \int_\Omega e^v\, d\mu \leqslant \exp\left(\sum_{i \leqslant q} \theta\left(\int_\Omega u_i\, d\mu\right)\right).$$

Proof. — For clarity, we will replace (3.3.6) by

$$\int_s^1 \gamma^{-1}(\theta(s) - \theta(t))\, d\lambda(t) \leqslant \frac{\tau}{|\theta'(s)|} \tag{3.3.10}$$

and we will determine in due time a good choice for τ. We already assume $\tau \leqslant 1$. The two main parts of the proof are the search of upper bounds for $\int_\Omega e^v\, d\mu$, and of lower bounds for $\sum_{i \leqslant q} \theta\left(\int_\Omega u_i\, d\mu\right)$.

Step 1. — For $i \leqslant q$, we set $S_i = \inf_\omega \theta(u_i(\omega)) = \theta(\sup_\omega u_i(\omega))$. By (3.3.8) and (3.3.1), taking $\omega^i = \omega$, we see that if we set $S = \sum_{i \leqslant q} S_i$, we have

$$v(\omega) \leqslant \theta(u_i(\omega)) + \sum_{j \neq i} S_j = \theta(u_i(\omega)) + S - S_i. \tag{3.3.11}$$

Step 2. — We make the convention that $\gamma(0) = \infty$. For $i \leqslant q$, we define s_i by

$$\theta(s_i) = \inf_{t \geqslant 0} \{ \gamma(\mu(\{ u_i \geqslant t \})) + \theta(t) \}.$$

Thus we have $\theta(s_i) \geqslant S_i$ and, for $t > s_i$,

$$\mu(\{ u_i \geqslant t \}) \leqslant \gamma^{-1}(\theta(s_i) - \theta(t)). \tag{3.3.12}$$

Step 3. — We show that, for any subset C of Ω,

$$\int_C e^v\, d\mu \leqslant \mu(C) \exp\Big(\sum_{i \leqslant q} \theta(s_i)\Big). \tag{3.3.13}$$

By definition of s_i, given $\varepsilon > 0$, we can find t_i such that $\gamma(\mu(B_i)) + \theta(t_i) \leqslant \theta(s_i) + \varepsilon$, where $B_i = \{ u_i \geqslant t_i \}$. Since $\theta(u_i(\omega^i)) \leqslant \theta(t_i)$ for $\omega^i \in B_i$, we have, by (3.3.8),

$$v(\omega) \leqslant \sum_{i \leqslant q} \theta(t_i) + h(\omega, B_1, \ldots, B_q),$$

so that (3.3.13) follows by (3.3.4), since ε is arbitrary.

Step 4. — Consider now a number m. We set

$$z = \int \min((v - m)^+, 1)\, d\mu \quad \text{and} \quad C = \{ v \geqslant m + 1 \}. \tag{3.3.14}$$

Thus, in particular $\mu(C) \leqslant z$.

Using the inequality $e^x \leqslant 1 + 2x^+$ for $x \leqslant 1$, we get, using (3.3.13),

$$\int_\Omega e^{v-m}\, d\mu \leqslant \int_{\Omega \setminus C} + \int_C \leqslant 1 + 2z + \mu(C) \exp\Big(\sum_{i \leqslant q} \theta(s_i) - m\Big),$$

so that

$$\int_\Omega e^{v-m}\,d\mu \leqslant 1 + z(2 + \exp(\sum_{i\leqslant q}\theta(s_i) - m)). \tag{3.3.15}$$

Step 5. — We now turn to lower bounds for $\sum_{i\leqslant q}\theta\left(\int_\Omega u_i\,d\mu\right)$. For each $i \leqslant q$, consider a number m_i, and set

$$w_i(\omega) = v(\omega) - S + S_i - \theta(m_i), \tag{3.3.16}$$

$$W_i = \int_\Omega \min(w_i^+, 1)\,d\mu. \tag{3.3.17}$$

We show that

$$\int_{\{w_i\geqslant 0\}} (u_i - m_i)\,d\mu \leqslant \frac{W_i}{3\theta'(m_i)}\left(= -\frac{W_i}{3\,|\,\theta'(m_i)\,|}\right). \tag{3.3.18}$$

By(3.3.11), we have

$$\theta(u_i(\omega)) \geqslant v(\omega) - S + S_i = w_i(\omega) + \theta(m_i). \tag{3.3.19}$$

Now, by (3.3.5), we have, for $y \geqslant x$,

$$\xi(y) \leqslant \xi(x) + \frac{1}{3}\xi'(x)\min(1, y - x).$$

Taking $x = \theta(m_i)$, $y = x + w_i(\omega)$, combining with (3.3.19) and recalling that $\xi'(\theta(m_i)) = \theta'(m_i)^{-1}$, yields, when $w_i(\omega) \geqslant 0$, that

$$u_i(\omega) \leqslant m_i + \frac{1}{3\theta'(m_i)}\min(1, w_i(\omega)),$$

from which (3.3.18) follows by integration.

Step 6. — We take $m_i = s_i$. It follows from (3.3.12), (3.3.10) that

$$\int_{\{u_i\geqslant s_i\}} (u_i - s_i)\,d\mu \leqslant \frac{\tau}{|\,\theta'(s_i)\,|}.$$

Combining with (3.3.18), observing that $w_i(\omega) > 0$ implies $u_i(\omega) < m_i$ by (3.3.19), and using convexity of θ yield

$$\theta\left(\int_\Omega u_i\,d\mu\right) \geqslant \theta(s_i) - \tau + \frac{W_i}{3}. \tag{3.3.20}$$

We choose the number m of Step 4 as the smallest for which

$$\mathrm{card}\{\, i \leqslant q;\, S - S_i + \theta(s_i) \leqslant m \,\} \geqslant \frac{q}{2}.$$

We observe that if $S - S_i + \theta(s_i) \leqslant m$, then $W_i \geqslant z$, where W_i is given by (3.3.17) and z by (3.3.14). Thus (3.3.20) shows that if we set $R = \sum_{i \leqslant q} \theta\left(\int_\Omega u_i \, d\mu\right)$, we have

$$\sum_{i \leqslant q} \theta(s_i) \leqslant R + q\tau - \frac{q}{6} z. \tag{3.3.21}$$

Combining with (3.3.15) gives

$$\begin{aligned}\int_\Omega e^{v-m} \, d\mu &\leqslant 1 + z(2 + e^{R + q\tau - \frac{q}{6}z - m}) \\ &\leqslant 3 + ze^{q\tau - \frac{q}{6}z} e^{R-m}.\end{aligned}$$

Calculus show that $\sup_z ze^{-qz/6} = 6/qe$. Thus, if we assume

$$e^{q\tau} \leqslant \frac{qe}{12}, \tag{3.3.22}$$

we have

$$\int e^{v-m} \, d\mu \leqslant 3 + \frac{1}{2} e^{R-m}.$$

For $R - m \geqslant 2$, this is $\leqslant e^{R-m}$, so the proof is finished.

Step 7. — Thus, we only have to consider the case $R \leqslant m + 2$. By definition of m, the set

$$I = \{ i \leqslant q ; m \leqslant S - S_i + \theta(s_i) \}$$

has cardinality $\geqslant q/2$. For i in I, we have

$$R \leqslant m + 2 \leqslant 2 + S + \theta(s_i) - S_i$$

and summation over $i \in I$ yields

$$R - S \leqslant 2 + \frac{1}{\operatorname{card} I} \sum_{i \in I} (\theta(s_i) - S_i) \leqslant 2 + \frac{2}{q} \sum_{i \leqslant q} (\theta(s_i) - S_i) \tag{3.3.23}$$

since $\theta(s_i) - S_i \geqslant 0$ for all $i \leqslant q$. On the other hand, (3.3.21) implies that

$$\sum_{i \leqslant q} (\theta(s_i) - S_i) \leqslant R - S + q\tau$$

which, combined with (3.3.23), yields (for $q \geqslant 3$, $\tau \leqslant 1$) that

$$\sum_{i \leqslant q} \theta(s_i) - S \leqslant \left(1 - \frac{2}{q}\right)^{-1} (2 + q\tau) \leqslant K + q\tau. \tag{3.3.24}$$

Step 8. — We assume that $q \geqslant 3$, $\tau \leqslant 1$, so that (3.3.24) holds, and we finish the proof. In Step 5, we take $m_i = \sup u_i(\omega)$, so that $\theta(m_i) = S_i$, and $w_i = v - S$ does not depend on i. From (3.3.18) and convexity, we get

$$\theta\left(\int_\Omega u_i \, d\mu\right) \geqslant S_i + \frac{W}{3},$$

where $W = W_i = \int_\Omega \min(1, w^+) \, d\mu$.

We now have by summation that

$$\frac{qW}{3} \leqslant R - S. \tag{3.3.25}$$

In Step 4, we take $m = S$, so that $z = W$. From (3.3.15), (3.3.24) we get

$$\int_\Omega e^{v-S} \, d\mu \leqslant 1 + 3W \exp(K + q\tau) \leqslant \exp(3W \exp(K + q\tau)). \tag{3.3.26}$$

According to (3.3.25), this is less than $\exp(R - S)$ provided $\exp(K + q\tau) \leqslant q/9$, i.e. $\tau \leqslant q^{-1} \log(q/K)$. Moreover, this requirement implies (3.3.22).

The proof is now complete. □

3.4. *Interpolation*

One can express Proposition 2.1.1 as the fact that, if $P(A) > 1/2$, then for most of the elements x of Ω^N, all but of order $\sqrt{N}$ coordinates can be copied by an element of A. On the other hand, Theorem 3.1.1 asserts that for most of the elements x of Ω^N, all but a bounded number of coordinates of x can be copied by one of two elements of A. A rather natural question is whether both phenomena can be achieved simultaneously (using the *same* elements of A). In this section, we will show that this is indeed the case.

This fact seems to be a special case of a rather general phenomenon that can be informally formulated as follows: Suppose we have defined two notion of the idea " the points x and y are within « distance» t "; we call these I and II respectively. Assume that there is good concentration of measure when the fattening A_t of A is defined as the collection of points x that are within distance t of A, when the meaning of this is defined with respect to notion I (resp. II). Then, in all the cases we have considered, it remains true that we have good concentration of measure when A_t is now defined as the collection of points x for which there exists a point y which is within distance t of x with respect of the two notions *simultaneously*. Two specific examples are presented, one in this section, the other in Section 4.5. In both sections, we present an inequality, that quantitatively contains two rather separate inequalities presented before. Considerably more difficult (if at all possible) would be the task of finding a formulation that would allow to recover sharp forms of these two inequalities. This direction of finding inequalities that " merge " several other inequalities is very natural. It remains at an embryonic stage. The reason is partly the intrinsic difficulty, partly the lack of concrete applications that would help to formulate precise needs.

We now go back to question of finding an inequality encompassing at the same time the essence of Proposition 2.1.1 and Theorem 3.1.1. For simplicity, we consider only the case $q = 2$ in Theorem 3.1.1. For two subsets A_1, A_2 of Ω^N, $x \in \Omega^N$, $a, t > 0$, we set

$$(3.4.1) \qquad f(A_1, A_2, a, t, x) = \inf\{f(y^1, y^2, a, t, x); y^1 \in A_1, y^2 \in A_2\},$$

where

$$f(y^1, y^2, a, t, x) = a \operatorname{card}\{i \leqslant N; x_i \neq y_i^1; x_i \neq y_i^2\} + t \operatorname{card}\{i \leqslant N; x_i \neq y_i^1 \text{ or } x_i \neq y_i^2\}.$$

Theorem **3.4.1.** — *For each* $a < \log 2$, *there exists* $t_0 > 0$ *such that*

$$(3.4.2) \qquad t < t_0 \Rightarrow \int e^{f(A_1, A_2, a, t, x)} \, dP(x) \leqslant \frac{e^{4Nt^2}}{P(A_1)\, P(A_2)}.$$

In particular, by Chebyshev inequality, this implies that for $u \leqslant 8Nt_0^2$, we have

$$P\left(\left\{f\left(A_1, A_2, a, \sqrt{\frac{u}{8N}}, x\right) \geqslant u\right\}\right) \leqslant \frac{e^{-u/2}}{P(A_1)\, P(A_2)}.$$

When $f\left(A_1, A_2, a, \sqrt{\frac{u}{8N}}, x\right) \leqslant u$, we can, by definition, find $y^1 \in A_1$, $y^2 \in A_2$ such that

$$a \operatorname{card}\{i \leqslant N; x_i \notin \{y_i^1, y_i^2\}\} + \sqrt{\frac{u}{8N}} \operatorname{card}\{i \leqslant N; x_i \neq y_i^1 \text{ or } x_i \neq y_i^2\} \leqslant u,$$

so that, in particular,

$$\operatorname{card}\{i \leqslant N; x_i \notin \{y_i^1, y_i^2\}\} \leqslant \frac{u}{a},$$

$$\operatorname{card}\{i \leqslant N; x_i \neq y_i^1 \text{ or } x_i \neq y_i^2\} \leqslant \sqrt{8Nu}.$$

We would like to point out that the factor e^{4Nt^2} in (3.4.2) is not optimal. This factor can be improved, in particular, with greater effort on the calculus computations of the proof we will present. Further improvement would be possible as in Section 1.2, but we have not pursued that direction since it is not clear at the present time what would be an optimal quantitative form of the phenomenon described by Theorem 3.4.1.

The key to Theorem 3.4.1 is the following.

Proposition **3.4.2.** — *Given* $b < \log 2$, *there exists* $t_0 > 0$ *such that, if* $t < t_0$, *for any two functions* $g_1, g_2 \leqslant 1$ *on* Ω, *we have*

(3.4.3)
$$\int_\Omega \min\left(e^b, \frac{e^t}{g_1(\omega)}, \frac{e^t}{g_2(\omega)}, \frac{1}{g_1(\omega)\, g_2(\omega)}\right) d\mu(\omega) \leqslant \frac{e^{4t^2}}{\int g_1\, d\mu \int g_2\, d\mu}.$$

Proof. — The relatively simple method we present does not yield the optimal dependence in t in the right hand side of (3.4.3), but it avoids lengthy unpleasant computations. Arguing as in the proof of Lemma 3.3.2, we see that

$$\int h\, d\mu \int g_1\, d\mu \int g_2\, d\mu \leqslant \exp \int (h + g_1 + g_2 - 3)\, d\mu.$$

Thus, if we set

$$h(g_1, g_2) = \min\left(e^b, \frac{e^t}{g_1}, \frac{e^t}{g_2}, \frac{1}{g_1 g_2}\right),$$

it suffices to show that, for t small enough and all numbers $g_1, g_2 \leqslant 1$, we have

(3.4.4)
$$h(g_1, g_2) + g_1 + g_2 \leqslant 3 + 4t^2.$$

Certainly, we can assume $g_1 \geqslant g_2$ and $2t \leqslant b$.

Case 1: $g_2 \leqslant g_1 \leqslant e^{t-a}$. In that case

$$h(g_1, g_2) + g_1 + g_2 - 3 \leqslant e^b + 2e^{t-b} - 3.$$

Since $e^b < 2$, we have $e^b + 2e^{-b} - 3 < 0$, so that we can find t_0 such that $e^b + 2e^{t-b} - 3 \leqslant 0$ if $t \leqslant t_0$.

Case 2: $g_2 \leqslant e^{t-b} \leqslant g_1$. In that case

$$\begin{aligned} h(g_1, g_2) + g_1 + g_2 - 3 &\leqslant \frac{e^t}{g_1} + g_1 + g_2 - 3 \\ &\leqslant e^b + 2e^{t-b} - 3, \end{aligned}$$

since the function $x + e^t/x$ decreases for $x \leqslant 1$, and we conclude as above.

Case 3: $e^{t-b} \leqslant g_2 \leqslant e^{-t}$. In that case, using again that the function $x + e^t/x$ decreases for $x \leqslant 1$, and the inequality $g_1 \geqslant g_2$, we have

$$\begin{aligned} h(g_1, g_2) + g_1 + g_2 - 3 &\leqslant \frac{e^t}{g_1} + g_1 + g_2 - 3 \\ &\leqslant \frac{e^t}{g_2} + 2g_2 - 3 \leqslant e^{2t} + 2e^{-t} - 3 \end{aligned}$$

since the function $2x + e^t/x$ is convex, and thus is bounded on the interval $[e^{t-b}, e^{-t}]$ by the maximum of its values at the endpoints. Also, we note that $e^{2t} + 2e^{-t} - 3 \leqslant 4t^2$.

Case 4: $g_2 \geqslant e^{-t}$. Then

$$h(g_1, g_2) + g_1 + g_2 - 3 \leqslant \frac{1}{g_1 g_2} + g_1 + g_2 - 3 \leqslant e^{2t} + 2e^{-t} - 3$$

since, when $c > 1$, the function $x + c/x$ decreases for $x \leqslant 1$. We then conclude as above. □

We will let the reader complete the proof of Theorem 3.4.1 using the induction method and Proposition 3.4.2. The basic observation is that, if B_i denotes the projection of A_i on Ω^N, we have for $x \in \Omega^N$, $\omega \in \Omega$,

$$f(A_1, A_2, a, t, (x, \omega)) \leqslant a + t + f(B_1, B_2, a, t, x),$$
$$f(A_1, A_2, a, t, (x, \omega)) \leqslant t + f(B_1, A_2(\omega), a, t, x),$$
$$f(A_1, A_2, a, t, (x, \omega)) \leqslant t + f(A_1(\omega), B_2, a, t, x),$$
$$f(A_1, A_2, a, t, (x, \omega)) \leqslant f(A_1(\omega), A_2(\omega), a, t, \omega).$$

For the induction hypothesis, one then fixes $a < b < \log 2$, and takes t_0 small enough so that $a + t_0 \leqslant b$.

4. Convex Hull

4.1. *The basic result*

The main idea of this section is the introduction of a rather different way of measuring how far a point x is from a subset A of Ω^N. We introduce the set

$$U_A(x) = \{ (s_i)_{i \leqslant N} \in \{ 0, 1 \}^N ; \exists y \in A, s_i = 0 \Rightarrow x_i = y_i \}.$$

We denote by $V_A(x)$ the convex hull of $U_A(x)$, when $U_A(x)$ is seen as a subset of $\mathbf{R}^N$. Thus $V_A(x)$ contains zero if and only if x belongs to A. We denote by $f_c(A, x)$ the ℓ^2-distance from zero to $V_A(x)$ (the letter c refers to " convexity "). The corresponding notion of " enlargement " of A is as follows:

(4.1.1) $$A_t^c = \{ x \in \Omega^N ; f_c(A, x) \leqslant t \}.$$

This notation will be kept throughout the paper.

Theorem **4.1.1.** — *For every subset* A *of* Ω^N, *we have*

(4.1.2) $$\int \exp \frac{1}{4} f_c^2(A, x) \, dP(x) \leqslant \frac{1}{P(A)}.$$

In particular

(4.1.3) $$P(A_t^c) \geqslant 1 - \frac{1}{P(A)} e^{-t^2/4}.$$

In order to understand better (4.1.1) it is worthwhile to note the following simple result.

Lemma **4.1.2.** — *The following are equivalent*:

(4.1.4) $x \in A_t^c$,

(4.1.5) $\forall\, (\alpha_i)_{i \leqslant N}, \quad \exists y \in A, \quad \sum_{i \leqslant N} \{ \alpha_i ; x_i \neq y_i \} \leqslant t \sqrt{\sum_{i \leqslant N} \alpha_i^2}.$

Proof. — The linear functional $\bar{\alpha} : x \to \sum_{i \leqslant N} \alpha_i x_i$ on $\mathbf{R}^N$, provided with the Euclidean norm, has a norm $\| \bar{\alpha} \| = \sqrt{\sum_{i \leqslant N} \alpha_i^2}$. Since $V_A(x)$ contains a point of norm $\leqslant f_c(A, x)$, the infimum of $\bar{\alpha}$ on $V_A(x)$ is $\leqslant f_c(A, x) \| \bar{\alpha} \|$; but since $V_A(x)$ is the convex hull of $U_A(x)$, the infimum of $\bar{\alpha}$ on $U_A(x)$ is the same as the infimum on $V_A(x)$. Thus (4.1.4) implies (4.1.5). The converse (that is not needed in the paper) follows from the Hahn-Banach theorem. □

It is very instructive to compare (4.1.3) with (2.1.3). If one takes $t = k/\sqrt{N}$, $\alpha_i = 1$, one sees that (4.1.3) implies

$$P(f(A, x) \geqslant k) \leqslant \frac{1}{P(A)} e^{-k^2/4N}.$$

The only difference with (2.1.3) is the worse numerical coefficient in the exponential. But the strength of (4.1.3) is, of course, that *all* choices of α_i are possible. This makes Theorem 4.1.1 a principle of considerable power, as will be demonstrated at length in Part II. It does, however, take some effort to fully understand the potential of Theorem 4.1.1. To illustrate one use of Theorem 4.1.1, let us consider the case where $\Omega = \{0, 1\}$, and where the probability μ gives mass p to 1 (and mass $1 - p$ to zero), where $p \leqslant 1/2$. Consider a subset A of $\{0, 1\}^N$, and assume that A is hereditary, i.e., that if $y = (y_i)_{i \leqslant N} \in A$, and if $(z_i)_{i \leqslant N}$ is such that $z_i \leqslant y_i$ for all i, then $z \in A$. Consider $x \in \{0, 1\}^N$, and $J = \{ i \leqslant N ; x_i = 1 \}$. Set $m(x) = \operatorname{card} J$. Define $\alpha_i = 1$ if $i \in J$, $\alpha_i = 0$ otherwise. Then Lemma 4.1.2 shows that we can find $y \in A$ such that

$$\operatorname{card} \{ i \in J ; x_i \neq y_i \} \leqslant f_c(A, x) \sqrt{m(x)}.$$

Since A is hereditary, we have $f(A, x) \leqslant f_c(A, x) \sqrt{m(x)}$.

Thus we have, for all m',

$$\begin{aligned} P(\{f(A, \cdot) \geqslant t\}) &\leqslant P\left(\left\{f_c(A, \cdot) \geqslant \frac{t}{\sqrt{m'}}\right\}\right) + P(m(y) > m') \\ &\leqslant \frac{1}{P(A)} \exp\left(-\frac{t^2}{4m'}\right) + P(m(y) > m'). \end{aligned}$$

Since the last term becomes very small for $m' > pN$, we recover the correct order $1/Np$ of the coefficient of t^2 in (2.3.5).

The key to Theorem 4.1.1 is the following simple lemma.

Lemma **4.1.3.** — *Consider* $0 \leqslant r \leqslant 1$. *Then*

(4.1.6) $$\inf_{0 \leqslant \lambda \leqslant 1} r^{-\lambda} \exp \frac{(1 - \lambda)^2}{4} \leqslant 2 - r.$$

This lemma is taken from [J-S]. This paper played an important role in the development of Theorem 4.1.1 and of the present paper. Roughly speaking, the author had proved Theorem 13.2 below in the case where $P(X_i = 1) = 1/2 = P(X_i = -1)$. Johnson and Schechman extended this to the present formulation of Theorem 13.2. The author's desire to have the last word prompted the discovery of the abstract setting of Theorem 4.1.1. It is this abstract setting that is largely responsible for the great range of applications of Theorem 4.1.1, and that lead to the present systematic investigation.

Proof. — Taking $\lambda = 1 + 2\log r$ if $r \geqslant e^{-1/2}$, and $\lambda = 0$ otherwise, and taking logarithms, it suffices to show that

$$f(r) = \log(2 - r) + \log r + (\log r)^2 \geqslant 0.$$

Now $f(1) = 0$, so it suffices to show that $f'(r) \leqslant 0$. Since $f'(1) = 0$, it suffices to show that $(rf'(r))' \geqslant 0$, or, equivalently, by calculation, that $(2 - r)^{-2} - r^{-1} \leqslant 0$. But $(2 - r)^{-2} \leqslant 1 \leqslant r^{-1}$. $\square$

We now prove Theorem 4.1.1, by induction upon N. We leave to the reader the easy case $N = 1$. For the induction step from N to $N + 1$, consider a subset A of Ω^{N+1} and its projection B on Ω^N. For $\omega \in \Omega$, we set as usual

$$A(\omega) = \{ x \in \Omega^N ; (x, \omega) \in A \}.$$

Consider $x \in \Omega^N$, $\omega \in \Omega$, $z = (x, \omega)$. The basic observation is that

$$s \in U_{A(\omega)}(x) \Rightarrow (s, 0) \in U_A(z),$$
$$t \in U_B(x) \Rightarrow (t, 1) \in U_A(z).$$

Thus, for $s \in V_{A(\omega)}(x)$, $t \in V_B(x)$, $0 \leqslant \lambda \leqslant 1$, we have $(\lambda s + (1 - \lambda)\, t, 1 - \lambda) \in V_A(z)$. The convexity of the function $u \mapsto u^2$ shows that

$$f_c^2(A, z) \leqslant (1 - \lambda)^2 + \lambda f_c^2(A(\omega), x) + (1 - \lambda) f_c^2(B, x). \tag{4.1.7}$$

The main trick of the proof is to resist the temptation to optimize now over λ. By Holder's inequality and induction hypothesis, we have

$$\begin{aligned}
&\int \exp \frac{1}{4} f_c^2(A, (x, \omega))\, dP(x) \\
&\quad \leqslant \exp \frac{1}{4}(1 - \lambda)^2 \left(\int_{\Omega^N} \exp \frac{1}{4} f_c^2(A(\omega), x)\, dP(x)\right)^{\lambda} \left(\int_{\Omega^N} \exp \frac{1}{4} f_c^2(B, x)\, dP(x)\right)^{1-\lambda} \\
&\quad \leqslant \exp \frac{1}{4}(1 - \lambda)^2 \left(\frac{1}{P(A(\omega))}\right)^{\lambda} \left(\frac{1}{P(B)}\right)^{1-\lambda} \\
&\quad = \frac{1}{P(B)} \exp \frac{1}{4}(1 - \lambda)^2 \left(\frac{P(A(\omega))}{P(B)}\right)^{-\lambda}.
\end{aligned}$$

This inequality holds for all $0 \leqslant \lambda \leqslant 1$. Using (4.1.6) with $r = \mathrm{P}(\mathrm{A}(\omega))/\mathrm{P}(\mathrm{B}) \leqslant 1$, we get

$$\int_{\Omega^{\mathrm{N}}} \exp \frac{1}{4} f_c^2(\mathrm{A}, (x, \omega))\, d\mathrm{P}(x) \leqslant \frac{1}{\mathrm{P}(\mathrm{B})}\left(2 - \frac{\mathrm{P}(\mathrm{A}(\omega))}{\mathrm{P}(\mathrm{B})}\right).$$

Integrating with respect to ω and using Fubini theorem yields

$$\int \exp \frac{1}{4} f_c^2(\mathrm{A}, \cdot)\, d(\mathrm{P} \otimes \mu) \leqslant \frac{1}{\mathrm{P}(\mathrm{B})}\left(2 - \frac{\mathrm{P} \otimes \mu(\mathrm{A})}{\mathrm{P}(\mathrm{B})}\right)$$

$$\leqslant \frac{1}{\mathrm{P} \otimes \mu(\mathrm{A})}, \quad \square$$

since $x(2 - x) \leqslant 1$ for all real x.

4.2. *Sharpening*

We now try to improve (4.1.2) by allowing a right-hand side $\mathrm{P}(\mathrm{A})^{-\alpha}$ for some $\alpha \geqslant 0$. In that case, it will be advantageous to measure the " distance " of s to $\mathrm{V}_{\mathrm{A}}(x)$ by the function

$$f_\alpha(\mathrm{A}, x) = \inf\{ \sum_{i \leqslant \mathrm{N}} \xi(\alpha, s_i)\, ; s \in \mathrm{V}_{\mathrm{A}}(x)\}$$

where

(4.2.1) $$\xi(\alpha, u) = \alpha(1 - u) \log(1 - u) - (\alpha + 1 - \alpha u) \log\left(\frac{1 + \alpha - \alpha u}{1 + \alpha}\right).$$

The reader should observe right away that $f_\alpha(\mathrm{A}, x)$ corresponds (with the notation of Section 4.1) to $f_c^2(\mathrm{A}, x)$ rather than to $f_c(\mathrm{A}, x)$. This will be the case for all the extensions of Theorem 4.1.1 we will consider.

As pointed out, Lemma 4.1.3 is the key to Theorem 4.1.1. It is a somewhat magic fact that when one tries to improve upon Lemma 4.1.3, the best possible function that can be used instead of the function $(1 - \lambda)^2/4$ can be computed exactly, leading to the formula (4.2.1).

Lemma **4.2.1.** — *Consider* $0 < r < 1$. *Then*

(4.2.2) $$\inf_{0 \leqslant \lambda \leqslant 1} r^{-\lambda\alpha} \exp \xi(\alpha, 1 - \lambda) = 1 + \alpha - \alpha r.$$

Proof. — We will not give the shortest possible proof (that consists in checking by computation that for $\lambda = r(\alpha + 1 - \alpha r)^{-1}$, we have $r^{-\lambda\alpha} \exp \xi(\alpha, 1 - \lambda) = 1 + \alpha - \alpha r$). Rather, we will explain how (4.2.2) was discovered. We fix α, and we set $f(x) = \alpha^{-1} \xi(\alpha, x)$. The best choice for λ is such that $\alpha \log r + \alpha f'(1 - \lambda) = 0$, i.e., $r = \exp(-f'(1 - \lambda))$. So we would like to have, for $0 \leqslant \lambda \leqslant 1$, the identity

$$\exp(\alpha f(1 - \lambda) + \alpha\lambda f'(1 - \lambda)) = 1 + \alpha - \alpha \exp(-f'(1 - \lambda)).$$

Setting $\nu = 1 - \lambda$, and taking logarithms, we want

$$\alpha f(\nu) + \alpha(1-\nu) f'(\nu) = \log(1 + \alpha - \alpha \exp(-f'(\nu))).$$

Differentiating in ν and setting $g(\nu) = \exp(-f'(\nu))$, we get

$$\alpha(1-\nu) f''(\nu) = \frac{\alpha f''(\nu)\, g(\nu)}{1 + \alpha - \alpha g(\nu)},$$

so that $g(\nu) = \dfrac{(\alpha+1)(1-\nu)}{\alpha + 1 - \alpha\nu}$. Taking logarithms and integrating yields (4.2.1). □

Lemma **4.2.2.** — *The function* $\xi(\alpha, \cdot)$ *is increasing and convex on* $[0, 1]$ *and* $\xi(\alpha, u) \geqslant \dfrac{\alpha}{2(\alpha+1)} u^2$.

Proof. — Computation shows that $\xi(\alpha, 0) = \dfrac{d\xi}{du}(\alpha, 0) = 0$, and

$$\frac{d^2 \xi}{du^2} = \frac{\alpha}{(\alpha + 1 - \alpha u)(1-u)} \geqslant \frac{\alpha}{\alpha+1}$$

since $u \geqslant 0$. □

Lemma **4.2.3.** — *For* $\alpha, a > 0$, *we have*

$$1 + \alpha - \alpha a \leqslant a^{-\alpha}, \tag{4.2.3}$$

$$a + (1-a) \exp \xi(\alpha, 1) \leqslant a^{-\alpha}. \tag{4.2.4}$$

Proof. — To prove (4.2.3), we observe that the graph of the convex function $x^{-\alpha}$ is above its tangent at the point $x = 1$. To prove (4.2.4), we observe that $\xi(\alpha, 1) = \log(1 + \alpha)$, so that the left-hand side is

$$a + (1-a)(1+\alpha) = 1 + \alpha - \alpha a,$$

and the result follows from (4.2.3). □

Theorem **4.2.4.** — *For a subset* A *of* Ω^N, *we have*

$$\int \exp f_\alpha(A, x)\, dP(x) \leqslant \frac{1}{P(A)^\alpha}. \tag{4.2.5}$$

Proof. — It is an obvious adaptation of the proof of Theorem 4.1.1. The case $N = 1$ follows from (4.2.4), and (4.2.3) is used as a substitute for the last inequality of (4.1.8). □

If we use Lemma 4.2.2, we see that (4.1.3) can be generalized into

$$\textbf{(4.2.6)} \qquad P(A_t^c) \geqslant 1 - \frac{1}{P(A)^\alpha} \exp\left(-\frac{\alpha t^2}{2(\alpha+1)}\right).$$

Optimization over α as in Corollary 2.2.3 yields:

Corollary **4.2.5.** — *For each subset* A *of* Ω^N,

$$\textbf{(4.2.7)} \qquad t \geqslant \sqrt{2 \log \frac{1}{P(A)}} \Rightarrow P(A_t^c) \geqslant 1 - \exp\left(-\frac{1}{2}\left(t - \sqrt{2\log\frac{1}{P(A)}}\right)^2\right).$$

It is an interesting question whether the term $\sqrt{2 \log 1/P(A)}$ can be removed in (4.2.7). We will, however, see in Section 4.3 that the coefficient 1/2 cannot be improved. It must be pointed out that Theorem 4.2.4 brings considerably more than a simple improvement of the coefficient of t^2 in (4.1.3). The reason is that $\xi(\alpha, 1) = \log(\alpha + 1)$ becomes very large when α is large. In that case, (4.2.5) recovers certain features of (3.1.2) and, in some ways, improves simultaneously upon Theorem 3.1.1 and Proposition 2.1.1. To see this, consider $q \geqslant 1$. We fix $A \subset \Omega^N$, and for $x \in \Omega^N$, we consider

$$k(x) = \inf\left\{k; \exists s \in V_A(x); \operatorname{card}\left\{i \leqslant N; s_i \geqslant 1 - \frac{1}{q}\right\} \leqslant k\right\}.$$

Then, certainly, we have

$$k(x)\, \xi\left(q, 1 - \frac{1}{q}\right) \leqslant f_q(A, x).$$

Now,

$$\xi\left(q, 1 - \frac{1}{q}\right) = \log\frac{1}{q} - 2\log\frac{2}{1+q}$$
$$= \log\frac{(1+q)^2}{4q} \geqslant \log\frac{q}{4}$$

so that

$$\textbf{(4.2.8)} \qquad k(x) \log\frac{q}{4} \leqslant f_q(A, x).$$

On the other hand, by (4.2.5), we have

$$P(f_q(A, x) \geqslant t) \leqslant \frac{e^{-t}}{P(A)^q}$$

so that, by (4.2.8),

$$\text{(4.2.9)} \qquad P(k(x) \geqslant k) \leqslant \frac{e^{-k \log \frac{q}{4}}}{P(A)^q} = \left(\frac{4}{q}\right)^k \frac{1}{P(A)^q}.$$

The relationship with (3.1.2) is as follows.

If $k(x) \leqslant k$, we can find a family $(y^j)_{j \leqslant m}$ of points of A, and coefficients $(\alpha_j)_{j \leqslant m}$, $0 \leqslant \alpha_j \leqslant 1$, $\sum_{j \leqslant m} \alpha_j = 1$, such that

$$\text{(4.2.10)} \qquad \operatorname{card}\left\{ i \leqslant N;\ \sum_{j \leqslant m} \alpha_j\, 1_{\{x_i \neq y_i^j\}} \geqslant 1 - \frac{1}{q} \right\} \leqslant k.$$

On the other hand, if $f(A, \ldots, A, x) \leqslant k$, we can find $y^1, \ldots, y^q$ in A such that

$$\text{(4.2.11)} \qquad \operatorname{card}\left\{ i \leqslant N;\ \sum_{j \leqslant q} \frac{1}{q}\, 1_{\{x_i \neq y_i^j\}} > 1 - \frac{1}{q} \right\} \leqslant k.$$

Certainly (4.2.11) is more precise than (4.2.10); however, for some important applications (see [T3]) (4.2.10) is just as powerful as (4.2.11).

4.3. *Two-point space*

In this section, we consider the case where $\Omega = \{0, 1\}$ and where μ gives weights $1 - p$ to zero and p to 1. The miracle of Lemma 4.2.1 does not seem to happen again, so we will only consider statements of the type

$$\text{(4.3.1)} \qquad \int \exp f_u(A, x)\, dP(x) \leqslant \frac{1}{P(A)^\alpha}$$

where, for a couple $u = (u_0, u_1)$ of positive numbers, we set

$$f_u(A, x) = \inf\{ u_0 \Sigma \{ s_i^2; x_i = 0 \} + u_1 \Sigma \{ s_i^2; x_i = 1 \} : s \in V_A(x) \}.$$

In other words, we take into account the fact that the points 0 and 1 do not play the same role.

If one analyzes the arguments of Sections 2.3, 4.1, 4.2, one sees that the best value the induction method allows to take for u_0 is the largest number s such that, whenever $a < b$, we have

$$(1 - p) \inf_{0 \leqslant \lambda \leqslant 1} \frac{1}{a^{\alpha(1-\lambda)}} \frac{1}{b^{\alpha\lambda}} e^{\lambda^2 s} + \frac{p}{b^\alpha} \leqslant \frac{1}{((1 - p)\, a + pb)^\alpha}$$

or, equivalently,

$$\text{(4.3.2)} \qquad (1 - p) \frac{1}{x^\alpha} \inf_{0 \leqslant \lambda \leqslant 1} x^{\alpha\lambda} e^{\lambda^2 s} \leqslant \frac{1}{((1 - p)\, x + p)^\alpha} - p$$

for all $0 \leqslant x \leqslant 1$.

(The best possible value of u_1 is obtained in a similar way, changing p in $1 - p$, and will not be considered.)

The infimum in (4.3.2) is obtained for

$$\lambda = \max\left(0, -\frac{\alpha \log x}{2s}\right).$$

The left-hand side of (4.3.2) is constant for $x \leqslant \exp(-2s/\alpha)$; thereby (4.3.2) holds provided we have, for $x \geqslant \exp(-2s/\alpha)$,

$$(1-p)\,e^{-\frac{\alpha^2}{4s}(\log x)^2} \leqslant \frac{1}{((1-p)\,x+p)^\alpha} - p. \tag{4.3.3}$$

Determining the best value of s for which this holds is an unpleasant task, so we will content ourselves with finding good values of s. Taking logarithms and differentiating, one sees that (4.3.3) will hold provided we have, for $x > 0$,

$$-\frac{\alpha}{2s}\log x \geqslant 1 - \frac{(1-p)\,x}{A - pA^{\alpha+1}} = \frac{p(1-A^{\alpha+1})}{A - pA^{\alpha+1}}, \tag{4.3.4}$$

where we have set $A = (1-p)\,x + p$.

It suffices that for $x \geqslant 0$ we have

$$-\frac{\alpha}{2s}\log x \geqslant \frac{p}{1-p}\,\frac{1-A^{\alpha+1}}{A}. \tag{4.3.5}$$

We first consider the case $p = \frac{1}{2}$, and we show that in this case we can take $s = \frac{\alpha}{\alpha+1}$. Since

$$1 - A^{\alpha+1} \leqslant (\alpha+1)\,(1-A) = (\alpha+1)\,(1-p)\,(1-x), \tag{4.3.6}$$

it suffices to see that

$$0 < x \leqslant 1 \Rightarrow \log x \geqslant \frac{2(1-x)}{1+x}.$$

But the function

$$f(x) = \log x - \frac{2(1-x)}{1+x}$$

satisfies $f(1) = 0$, $f'(x) = -(1-x)^2/(1+x)^2 \leqslant 0$. Using the notation $f_c(A, x)$ of Section 4.1.1, we then have proved the following.

Theorem **4.3.1.** — *When* $\Omega = \{0, 1\}$ *and* μ *is uniform, for each* $\alpha \geqslant 1$ *and each subset* A *of* Ω^N, *we have*

$$\int \exp\left(\frac{\alpha}{\alpha+1} f_c^2(A, x)\right) dP(x) \leqslant \frac{1}{P(A)^\alpha}. \tag{4.3.7}$$

Compared with (4.2.6), we have gained a factor 2 in the exponent in the special case of the two-point space.

Corollary **4.3.2.** — *When* $\Omega = \{0, 1\}$ *and* μ *is uniform, for each* $\alpha \geqslant 1$, *and each subset of* Ω^N, *we have*

$$(4.3.8) \qquad t \geqslant \sqrt{\log \frac{1}{P(A)}} \Rightarrow P(A_t^c) \geqslant 1 - \exp\left(-\left(t - \sqrt{\log\frac{1}{P(A)}}\right)^2\right).$$

Proof. — From (4.3.7) and Chebyshev inequality, we get

$$P(A_t^c) \geqslant 1 - \frac{1}{P(A)^\alpha} \exp\left(-\frac{\alpha}{\alpha+1} t^2\right)$$

and we optimize over α as in the proof of Corollary 2.2.3.

It is a natural question whether (4.3.8) can be improved into

$$(4.3.9) \qquad P(A_t^c) \geqslant 1 - K \exp(-t^2).$$

It should, however, be pointed out that the coefficient of t^2 is optimal. We will now show this, and at the same time, the optimality of the coefficient 1/2 in (4.2.7). Provide $\Omega = \{0, 1\}$ with the probability μ that gives mass p to 1. Set

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

(Thus, for N large, P(A) is about 1/2.) Consider $y \in \{0, 1\}^N$, such that card $J = m$, where $J = \{i \leqslant N; y_i = 1\}$. Assume $m > pN$. Then any element x of A differs of y in at least $m - pN$ of the coordinates indexed by J. Using Lemma 4.1.2 for $\alpha_i = 1/\sqrt{m}$ when $i \in J$, $\alpha_i = 0$ otherwise, we see that

$$(4.3.10) \qquad f_c(A, y) \geqslant \frac{(m - pN)}{\sqrt{m}} = \frac{m - pN}{\sqrt{N}} \sqrt{\frac{N}{m}}.$$

If we think of $m = m(y)$ as a r.v., the central limit theorem shows that, as $n \to \infty$, $(m - pN)/\sqrt{N}$ is asymptotically normal, with standard deviation $\sqrt{p(1-p)}$. On the other hand, $\sqrt{N/m}$ converges to $\sqrt{1/p}$ in probability. Thus

$$\lim_{N\to\infty} P(f_c(A, \cdot) \geqslant t) \geqslant \frac{1}{\sqrt{2\pi}} \int_{t/\sqrt{(1-p)}}^{\infty} \exp\left(-\frac{u}{2}\right)^2 du$$
$$\geqslant \frac{1}{Kt} \exp\left(-\frac{t^2}{2(1-p)}\right).$$

If $p = 1/2$, the coefficient of t^2 is -1, and if we let p arbitrary, we cannot do better than the coefficient $-1/2$ of (4.2.7).

We now go back to our main line of discussion, and we consider the case $p \leqslant 1/2$; we will show that in this case we can take

$$(4.3.11) \qquad s = \min\left(\frac{\alpha}{K} \log \frac{1}{p}, \frac{\alpha}{4(\alpha+1)p}\right).$$

In particular, for α large, this is of order $1/p$, rather than order $\log(1/p)$. This remarkable fact is closely connected to Theorem 4.4.1 below. To prove (4.3.11), we prove (4.3.5), depending on the value of x.

Case 1: $x \geqslant 1/2$. Then $-\log x \geqslant 1-x$, $1-A^{\alpha+1} \leqslant (\alpha+1)(1-p)(1-x)$, so that it suffices that

$$-\frac{\alpha}{2s} \geqslant \frac{p(1+\alpha)}{A}.$$

Now $A \geqslant \frac{1}{2}$, so that it suffices that $s \leqslant \frac{\alpha}{4p(1+\alpha)}$.

Case 2: $x \leqslant \sqrt{p}$. Then $-\log x \geqslant -\frac{1}{2}\log 1/p$, so that it suffices that

$$\frac{\alpha}{4s}\log\left(\frac{1}{p}\right) \geqslant \frac{p}{(1-p)\,A}.$$

Since $A \geqslant p$, it suffices that

$$s \leqslant \frac{\alpha(1-p)}{4}\log\frac{1}{p}.$$

Case 3: $\sqrt{p} \leqslant x \leqslant \frac{1}{2}$. It then suffices, since $-\log x \geqslant \log 2$ and $A \geqslant (1-p)\sqrt{p}$, that

$$s \leqslant \frac{(1-p)^2}{\sqrt{p}}\,\frac{\alpha}{2}\log 2,$$

which holds when $s \leqslant \frac{\alpha}{K}\log\frac{1}{p}$.

4.4. *Penalties*

We now consider a function h on $\Omega \times \Omega$, such that $h \geqslant 0$ and $h(\omega,\omega) = 0$ for $\omega \in \Omega$. For a subset A of Ω^N, and $x \in \Omega^N$, we set

$$U_A(x) = \{(s_i) \in \mathbf{R}_+^N;\ \exists y \in A;\ \forall i \leqslant N,\ s_i \geqslant h(x_i, y_i)\}.$$

We denote by $V_A(x)$ the convex hull of $U_A(x)$. The situation of Sections 4.1, 4.2 corresponds to the case where $h(\omega,\omega') = 1$ if $\omega \neq \omega'$.

In order to measure the " distance " of zero to $V_A(x)$, we consider a convex function ψ on $\mathbf{R}$, with $\psi(0) = 0$. We will assume

(4.4.1) $\quad x \leqslant 1 \Rightarrow \psi(x) \leqslant x^2; \quad x \geqslant 1 \Rightarrow \psi(x) \geqslant x.$

We set

$$f_{h,\psi}(A, x) = \inf\{\sum_{i\leqslant N} \psi(s_i);\ s = (s_i)_{i\leqslant N} \in V_A(x)\}.$$

(Thus, the situation of Section 4.1 corresponds to the case $\psi(s) = s^2$ and the situation of Section 4.2 corresponds to the case $\psi(s) = \xi(\alpha, s)$.) The material of this section is connected to that of Section 2.6, and the notation of Section 2.6 is in force in the present section. Thus θ denotes a convex function from $]0, 1]$ to $\mathbf{R}^+$, with $\theta(1) = 0$, $\lim_{x \to 0} \theta(x) = \infty$, and ξ denotes the inverse function. We assume that (2.6.1) holds, and assume moreover that for a certain number $\gamma > 0$, we have

$$b \geqslant 0 \Rightarrow |\xi'(b+1)| \geqslant \gamma |\xi'(b)|, \tag{4.4.2}$$

$$|\theta'(1)| \geqslant \gamma, \quad w(1/2) \geqslant \gamma. \tag{4.4.3}$$

We recall the function Ξ of (2.6.2), as well as condition $\mathrm{H}(\xi, w)$ of (2.6.3).

Theorem **4.4.1.** — *Consider a nonincreasing function w on $]0, 1]$, $w \leqslant \theta$. Assume that $\int_0^1 w^2 \, d\lambda \leqslant 1$, and that condition $\mathrm{H}(\xi, w)$ holds. Assume that for each subset B of Ω, we have*

$$0 < \mu(\mathrm{B}) \leqslant \frac{1}{2} \Rightarrow \int_\Omega \exp \psi(h(x, \mathrm{B})) \, d\mu(x) \leqslant \exp w(\mu(\mathrm{B})), \tag{4.4.4}$$

$$\mu(\mathrm{B}) \geqslant \frac{1}{2}, \quad t \geqslant 1 \Rightarrow \mu(\{x; \psi(h(x, \mathrm{B})) \geqslant t\}) \leqslant e^{-t}(1 - \mu(\mathrm{B})). \tag{4.4.5}$$

Then, for each subset A of Ω^{N}, we have

$$\int_{\Omega^{\mathrm{N}}} \exp \frac{1}{\mathrm{K}} f_{h,\psi}(\mathrm{A}, x) \, d\mathrm{P}(x) \leqslant \exp \theta(\mathrm{P}(\mathrm{A})), \tag{4.4.6}$$

where K depends on γ only.

We should observe first that only the values of $w(x)$ for $x \leqslant 1/2$ matter.

In order to compare Theorem 4.4.1 with Theorems 2.6.5 and 2.7.1, we first have to keep in mind that it is the function $\psi \circ h$ here that plays the role of h in these theorems. The conclusion of Theorem 4.4.1 is stronger than that of Theorem 2.6.5 (the way Theorem 4.1.1 improves on Proposition 2.1.1) but weaker than the conclusion of Theorem 2.7.1 (since one takes convex hulls). Condition (4.4.5) strongly resembles (2.7.2). Condition (4.4.4) coincides with Condition (2.6.12) when $\mu(\mathrm{B}) \leqslant 1/2$. A simple calculation using (4.4.5) shows that for $\mu(\mathrm{B}) \geqslant 1/2$, condition (4.4.5) is of a somewhat stronger nature than (2.6.12).

An interesting case where it is worth to spell out (4.4.4) and (4.4.5) is when $h(x, y) = h(y)$ depends on y only. Denoting by m a median of h, (4.4.5) will hold if $\psi(m) < 1$. And, as seen after Theorem 2.6.5, (4.4.4) holds provided $w(\mu(\{h \geqslant t\})) \geqslant \psi(t)$ (a tail condition of h).

To prove Theorem 4.4.1 when $N = 1$, we observe that, since $w \leqslant \theta$, (4.4.6) follows from (4.4.4) when $\mu(B) \leqslant 1/2$. When $\mu(B) \geqslant 1/2$, a simple computation using (4.4.5) shows that given γ, if K is large enough, then

$$\int_\Omega \exp \frac{1}{K} \psi(h(x, B))\, d\mu(x) \leqslant 1 + \gamma(1 - \mu(B)) \leqslant \theta(\mu(B)) \leqslant \exp \theta(\mu(B)),$$

since $\theta'(1) \geqslant \gamma$.

For the induction step, comparison with the proof of Theorem 4.1.1 shows that it suffices to prove the following (used for $g = \xi(f)$).

Proposition **4.4.2.** — *There exists a constant* L, *depending on* γ *only, with the following property. Under the conditions of Theorem* 4.4.1, *consider a function* $f \geqslant 0$ *on* Ω. *Set*

$$\hat{f}(x) = \inf_{y \in \Omega,\, 0 \leqslant \lambda \leqslant 1} (\lambda f(x) + (1 - \lambda) f(y) + \frac{1}{L} \psi((1 - \lambda)\, h(x, y))). \tag{4.4.7}$$

Then we have

$$\int e^{\hat{f}}\, d\mu \leqslant e^{\theta(\int \xi(f)\, d\mu)}. \tag{4.4.8}$$

Understandably, with the level of generality considered here, the proof cannot be very short. The reason why we have opted for great generality is that Theorem 4.1.1 is a principle of considerable power (as will be demonstrated in Chapter 8) and that thereby it seems worthwhile to prove extensions of it under weak hypotheses on the function h. The proof will incorporate in particular ideas from Theorems 4.1.1, 2.6.5, 2.7.1.

A simple idea is that we will need to control $\theta\left(\int \xi(f)\, d\mu\right)$ from below. This means controlling the lower tail of f. Set $B_s = \{f \leqslant s\}$, and denote by m a median of f, so that $\mu(B_m) \geqslant 1/2$. We set

$$b = \inf_{s \leqslant m} \left\{ s + \frac{1}{L} w(\mu(B_s)) \right\}. \tag{4.4.9}$$

The first step of the proof will be to show that $\mu(B_s)$ is not too big, i.e. that b is not too small.

Proposition **4.4.3.** — *To prove Proposition* 4.4.2, *if* $L > 4/\gamma$, *we can assume*

$$m \leqslant b + \frac{4}{L\gamma}. \tag{4.4.10}$$

Proof. — We assume $m > b$, for otherwise there is nothing to prove. Using (4.4.7) with $\lambda = 0$, we see that for each s we have $\hat{f}(x) \leqslant s + L^{-1} \psi \circ h(x, B_s)$. Using (4.4.4) together with Hölder's inequality, it follows that

$$\int_\Omega e^{\hat{f}}\, d\mu \leqslant \exp\left(s + \frac{1}{L} w(\mu(B_s))\right),$$

so that $\int_\Omega e^{\hat{f}}\, d\mu \leqslant e^b$ by taking the infimum over $s \leqslant m$. On the other hand, (4.4.9) implies

$$s \leqslant m \Rightarrow b - s \leqslant \frac{1}{L}\, w(\mu(B_s)) \tag{4.4.11}$$

i.e.

$$\left| \left\{ \frac{1}{L}\, w \geqslant b - s \right\} \right| \geqslant \mu(B_s). \tag{4.4.12}$$

We can hence appeal to Lemma 2.6.4 with $C = \{f < b\}$ and $t = 1/L$ to see that

$$\int_C \xi(f)\, d\mu \leqslant \mu(C)\, \xi(b) + \xi'(b) \int_C (f - b)\, d\mu + \frac{1}{L^2}\, |\, \xi'(b)\, |.$$

But, by (4.4.12) we have

$$\int_C |\, f - b\, |\, d\mu \leqslant \frac{1}{L} \int w\, d\mu \leqslant \frac{1}{L} \left(\int w^2\, d\mu \right)^{1/2} \leqslant \frac{1}{L}$$

and thus

$$\int_C \xi(f)\, d\mu \leqslant \mu(C)\, \xi(b) + \frac{2}{L}\, |\, \xi'(b)\, |. \tag{4.4.13}$$

On the other hand, when $f(\omega) > b$, we have

$$\xi(f(\omega)) \leqslant \xi(b) - (\xi(b) - \xi(m))\, 1_{\{f \geqslant m\}}(\omega)$$

and, by integration, since $\mu(\{f \geqslant m\}) \geqslant 1/2$ (and $m > b$),

$$\int_{\Omega \setminus C} \xi(f)\, d\mu \leqslant (1 - \mu(C))\, \xi(b) - \frac{1}{2}\, (\xi(b) - \xi(m)).$$

Combining with (4.4.13) we get

$$\int_\Omega \xi(f)\, d\mu \leqslant \xi(b) + \frac{2}{L}\, |\, \xi'(b)\, | - \frac{1}{2}\, (\xi(b) - \xi(m)).$$

Since we have shown that $\int_\Omega e^{\hat{f}}\, d\mu \leqslant e^b$, there is nothing to prove unless $\int_\Omega \xi(f)\, d\mu \geqslant \xi(b)$ $\left(\text{for otherwise } \theta\left(\int \xi(f)\, d\mu\right) \geqslant b\right)$. Thus we can assume

$$\frac{1}{2}\, (\xi(b) - \xi(m)) \leqslant \frac{2}{L}\, |\, \xi'(b)\, |. \tag{4.4.14}$$

Now, since $m > b$, from (4.4.2) follows that $\xi(m) \leqslant \xi(b) - \gamma \mid \xi'(b) \mid \min((m - b), 1)$. Comparing with (4.4.14), we see that $\min(m - b, 1) \leqslant 4/(L\gamma)$, so that if $L > 4/\gamma$, we must have $m - b \leqslant 4/L\gamma$. □

We consider the smallest number α for which

$$\forall\, s \leqslant m, \qquad m - s \leqslant \alpha w(\mu(B_s))$$

so that

(4.4.15) $$\forall\, s \leqslant m, \qquad |\{ \alpha w \geqslant m - s \}| \geqslant \mu(B_s).$$

It is rather important to note that

(4.4.16) $$\alpha \leqslant \frac{8}{L\gamma^2}.$$

Indeed, if $m - s \leqslant 8/L\gamma$, then

$$\frac{m - s}{w(\mu(B_s))} \leqslant \frac{m - s}{w(1/2)} \leqslant \frac{8}{L\gamma^2}.$$

On the other hand, if $m - s \geqslant 8/L\gamma$, then, by (4.4.10), we have $m - s \leqslant 2(b - s)$, so that

$$\frac{m - s}{w(\mu(B_s))} \leqslant 2\,\frac{b - s}{w(\mu(B_s))} \leqslant \frac{2}{L}.$$

We consider a second parameter $M \leqslant L$. Throughout the rest of this section, we will have to put conditions on L, M, L/M. For simplicity we make the convention that the expression " if L is large enough "... means " there exists a constant $K(\gamma)$, depending on γ only, such that, if $L \geqslant K(\gamma)$... " and similarly for M, L/M.

We set $m' = m - 16/L\gamma^2$. We consider the function

$$f' = \min\left(f, m + \frac{1}{M}\right)$$

and the function g defined as

$$g(\omega) = f(\omega) \qquad \text{if } f(\omega) \leqslant m',$$

$$g(\omega) = \max\left(m', \min\left(\hat{f}(\omega), m + \frac{1}{M}\right)\right) \qquad \text{if } f(\omega) > m'.$$

Since $\hat{f} \leqslant f$, it is simple to see that $g \leqslant f'$. It is also simple to see that

(4.4.17) $$g(\omega) \neq f'(\omega) \Rightarrow g(\omega), \quad f'(\omega) \in \left[m', m + \frac{1}{M}\right].$$

Indeed, the right-hand side does not occur only when $f(\omega) < m'$, and then $f(\omega) = f'(\omega) = g(\omega)$. We set

$$C = \{f \geqslant m\}; \quad D = \left\{f \geqslant m + \frac{1}{M}\right\}.$$

Lemma **4.4.4.** — *We have*

$$\text{(4.4.18)} \qquad \theta\left(\int_\Omega \xi(f)\, d\mu\right) \geqslant \int_\Omega f'\, d\mu - \alpha^2 - \int_C (f' - m)^2\, d\mu.$$

Proof. — Since $f' \leqslant f$, we have $\xi(f') \geqslant \xi(f)$ and

$$\theta\left(\int_\Omega \xi(f)\, d\mu\right) \geqslant \theta\left(\int_\Omega \xi(f')\, d\mu\right).$$

We now appeal to Lemma 2.6.4 with $t = \alpha$. We have

$$\int_\Omega \xi(f') \leqslant \xi(m) + \xi'(m) \int_\Omega (f' - m)\, d\mu + \alpha^2 \mid \xi'(m) \mid + \xi''(m) \int_C (f' - m)^2\, d\mu.$$

By convexity of θ and since $\xi''(m) \leqslant \mid \xi'(m) \mid$ this implies

$$\theta\left(\int_\Omega \xi(f)\right) \geqslant m + \int_\Omega (f - m)\, d\mu - \alpha^2 - \int_C (f' - m)^2\, d\mu. \quad \square$$

Lemma **4.4.5.** — *If* L *and* M *are large enough, we have*

$$\text{(4.4.19)} \qquad \int_\Omega e^{\hat{f}}\, d\mu \leqslant \exp\left(\frac{1}{2}\int_\Omega (g + f')\, d\mu + 2\alpha^2 + 2\int_C (f' - m)^2\, d\mu + \int_\Omega (e^{\hat{f} - m} - e^{1/M})^+\, d\mu\right).$$

Proof. — First, we observe that

$$\text{(4.4.20)} \qquad \int_\Omega e^{\hat{f} - m}\, d\mu \leqslant \int_\Omega \exp\left(\min\left(\hat{f} - m, \frac{1}{M}\right)\right) d\mu + \int_\Omega (e^{\hat{f} - m} - e^{1/M})^+\, d\mu.$$

We observe that $\min\left(\hat{f} - m, \frac{1}{M}\right) \leqslant g - m$. Since $e^x \leqslant 1 + x + x^2$ for $x \leqslant 1/M \leqslant 1$, we have

$$\text{(4.4.21)} \qquad \int_\Omega \exp\left(\min\left(\hat{f} - m, \frac{1}{M}\right)\right) d\mu \leqslant \int_\Omega e^{g - m}\, d\mu \leqslant 1 + \int_\Omega (g - m)\, d\mu + \int_\Omega (g - m)^2\, d\mu.$$

Now, by (4.4.17), and provided L, M are large enough,

$$(g - m)^2 \leqslant 2(f' - m)^2 + 2(g - f')^2 \leqslant 2(f' - m)^2 + \frac{1}{2}(f' - g). \tag{4.4.22}$$

We recall also that

$$\int_{\Omega \setminus C} (f' - m)^2 \, d\mu \leqslant \int_{\Omega \setminus C} (f - m)^2 \, d\mu \leqslant \alpha^2.$$

The result follows by combining these inequalities, and using that $1 + x \leqslant e^x$. □

It follows from Lemmas 4.4.4 and 4.4.5 that to prove Proposition 4.4.2, it suffices to prove the following when M, L/M are large enough.

$$\int_\Omega (f' - g) \, d\mu \geqslant 6\alpha^2 + 6 \int_C (f' - m)^2 \, d\mu + 2 \int_\Omega (e^{\hat{f} - m} - e^{1/M})^+ \, d\mu. \tag{4.4.23}$$

This follows from the next three lemmas.

Lemma **4.4.6.** — *We have*

$$\int_\Omega (e^{\hat{f} - m} - e^{1/M})^+ \, d\mu \leqslant \frac{K}{L} \mu(D) \leqslant \frac{KM^2}{L} \int_C (f' - m)^2 \, d\mu. \tag{4.4.24}$$

Lemma **4.4.7.** — *If* L/M *is large enough, we have*

$$\int_\Omega (f' - g) \, d\mu \geqslant \frac{M}{K} \int_C (f' - m)^2 \, d\mu.$$

Lemma **4.4.8.** — *If* L/M *is large enough, we have*

$$\int_\Omega (f' - g) \, d\mu \geqslant \frac{L\gamma^4 \alpha^2}{K}.$$

Proof of Lemma 4.4.6. — The definition of $\hat{f}$ (with $\lambda = 1$) shows that $\hat{f}(\omega) \leqslant m + 1/M + L^{-1} \psi(h(\omega, \Omega \setminus D))$. Thus by (4.4.5) we have

$$\mu\left(\left\{\hat{f} \geqslant m + \frac{1}{M} + \frac{k}{L}\right\}\right) \leqslant e^{-k} \mu(D)$$

and thus

$$\int_\Omega (e^{\hat{f} - m} - e^{1/M})^+ \, d\mu \leqslant \sum_{k \geqslant 1} e^{\frac{1}{M}} (e^{\frac{k}{L}} - 1) \, e^{-k+1} \mu(D),$$

from which the first inequality of (4.4.24) follows by elementary estimates. (The second inequality of (4.4.24) is obvious.) □

Proof of Lemma 4.4.7. — *Step 1.* — For $k \geqslant 0$, we define

$$a_k = \sup \left\{ t;\ \mu(\{f' \geqslant t\}) \geqslant \frac{1}{2e^k} \right\}.$$

Thus $m \leqslant a_k \leqslant a_{k+1} \leqslant m + 1/M$. We consider a set $Z_k \subset \{f \leqslant a_k\}$ such that

$$\mu(Z_k) = 1 - \frac{1}{2e^k}.$$

We set $Z'_k = \{\omega;\ h(\omega, Z_k) \leqslant 2\}$. Since $\psi(x) \geqslant x$ for $x \geqslant 1$, we have

$$Z'_k \supset \{\omega;\ \psi(h(\omega, Z_k)) \leqslant 2\}$$

so that, by (4.4.5), $\mu(Z'_k) \geqslant 1 - 1/2e^{k+2}$. We set, for $k \geqslant 0$,

$$W_k = Z'_k \cap (Z_{k+2} \backslash Z_{k+1}).$$

We observe that the sets $(W_k)_{k \geqslant 0}$ are disjoint, and that

$$\mu(W_k) \geqslant \frac{1}{2e^k}\left(\frac{1}{e} - \frac{2}{e^2}\right) \geqslant \frac{1}{2e^{k+3}}. \tag{4.4.25}$$

Step 2. — We show that

$$\int_{W_k} (f' - g)\, d\mu \geqslant \frac{M}{K}(a_{k+1} - a_k)^2\, \mu(W_k \backslash D). \tag{4.4.26}$$

Consider $\omega \in W_k \backslash D$. Then $f'(\omega) = f(\omega)$, so that, given $\lambda \in [0, 1]$ and $\omega' \in \Omega$

$$\begin{aligned} f'(\omega) - \hat{f}(\omega) &= f(\omega) - \hat{f}(\omega) \\ &\geqslant (1 - \lambda)(f(\omega) - f(\omega')) - \frac{1}{L}\psi((1 - \lambda)\, h(\omega, \omega')). \end{aligned} \tag{4.4.27}$$

We can find $\omega' \in Z_k$ such that $h(\omega, \omega') \leqslant 3$. Then $f(\omega) - f(\omega') \geqslant a_{k+1} - a_k$. We can take $0 \leqslant \lambda \leqslant 1$ such that $1 - \lambda = M(a_{k+1} - a_k)/3$. Then (4.4.27) yields, since $\psi(x) \leqslant x^2$ for $x \leqslant 1$, that

$$f'(\omega) - \hat{f}(\omega) \geqslant \frac{M}{3}(a_{k+1} - a_k)^2 - \frac{9M^2}{L}(a_{k+1} - a_k)^2.$$

Thus, if L/M is large enough,

$$f'(\omega) - \hat{f}(\omega) \geqslant \frac{M}{4}(a_{k+1} - a_k)^2,$$

that is

$$\hat{f}(\omega) \leqslant f'(\omega) - \frac{M}{4}(a_{k+1} - a_k)^2.$$

Since $a_{k+1} - a_k \leqslant 1/M$, and $f'(\omega) \geqslant a_{k+1}$, the right-hand side is $\geqslant a_k \geqslant m$, so that

$$g(\omega) \leqslant f'(\omega) - \frac{M}{4}(a_{k+1} - a_k)^2$$

and thus
$$f'(\omega) - g(\omega) \geqslant \frac{M}{4}(a_{k+1} - a_k)^2$$

from which (4.4.26) follows by integration.

Step 3. — Denote by k_0 the largest integer such that $1/4e^{k_0+3} \geqslant \mu(D)$. Thus $\mu(W_k) \geqslant 2\mu(D)$ for $k \leqslant k_0$, and, by (4.4.26) and summation, we get, since $\mu(W_k \backslash D) \geqslant \mu(W_k)/2$,

$$\text{(4.4.28)} \qquad \int (f' - g)\, d\mu \geqslant \frac{M}{K} \sum_{k \leqslant k_0} (a_{k+1} - a_k)^2 e^{-k}.$$

By the argument of Lemma 2.7.8, we have

$$\sum_{k \leqslant k_0} (a_{k+1} - a_k)^2 e^{-k} \geqslant \frac{1}{K} \int_C (\min(f', m + a_{k_0+1}) - m)^2\, d\mu.$$

Thus the proof is completed if $a_{k_0+1} \geqslant 1/2M$.

Step 4. — Assuming now $a_{k_0+1} \leqslant 1/2M$, we shall show that

$$\text{(4.4.29)} \qquad \int (f' - g)\, d\mu \geqslant \frac{1}{KM} e^{-k_0}.$$

Since

$$e^{-k_0} \geqslant \frac{M^2}{K} \sum_{k > k_0} (a_{k+1} - a_k)^2 e^{-k},$$

combining with (4.4.28), we get

$$\int (f' - g)\, d\mu \geqslant \frac{M}{K} \sum_{k \geqslant 0} (a_{k+1} - a_k)^2 e^{-k} \geqslant \frac{M}{K} \int_C (f - m)^2\, d\mu$$

by (the argument of) Lemma 2.7.8, completing the proof of Lemma 4.4.7.

To prove (4.4.29), we observe that, by definition of k_0, we have $e^{-k_0-6} \leqslant \mu(D)$. Consider the set

$$Z = \{ \psi(h(\cdot, Z_{k_0+1})) \leqslant 6 \}.$$

Then, by (4.4.5), we have $\mu(Z) \geqslant 1 - e^{-k_0-7}$, so that $\mu(Z \cap D) \geqslant e^{-k_0}/K$. Now, if $\omega \in D$, we have $f'(\omega) \geqslant m + 1/M$ while if $\omega \notin D$, we have

$$\begin{aligned}\hat{f}(\omega) &\leqslant a_{k_0+1} + m + \frac{1}{L}\,\psi(h(\omega, Z_{k_0+1}))\\ &\leqslant m + \frac{1}{2M} + \frac{6}{L}.\end{aligned}$$

Thus $g(\omega) \leqslant m + \dfrac{3}{4M}$ if L/M is large enough. Hence, $f - g \geqslant 1/4M$ on $Z \cap D$. □

Proof of Lemma 4.4.8. — *Step 1.* — We show that we can assume $\mu(D) \leqslant 1/8$. Indeed otherwise by Lemma 4.4.7 we have $\int (f' - g)\, d\mu \geqslant 1/KM$ and, since $\alpha \leqslant 8/\gamma^2 L$, this is $\geqslant L\alpha^2$ when L/M is large enough.

Step 2. — By definition of α, there exists $s < m$ with $m - s > \alpha w(\mu(B_s))/2$. By (4.4.5) and Chebyshev inequality, the set

$$H = \{\, \psi(h(\cdot, B_s)) \leqslant 2 + w(\mu(B_s)) \,\}$$

has measure $\geqslant 3/4$. Thus if we set $G = H \cap (C \backslash D)$, we have $\mu(G) \geqslant 1/8$.

Step 3. — Set

$$\beta = \frac{m - s}{3 + w(\mu(B_s))}.$$

Since $w(\mu(B_s)) \geqslant w(1/2) \geqslant \gamma$, and $m - s \geqslant \alpha w(\mu(B_s))/2$, we have

$$\frac{\gamma\alpha}{K} \leqslant \beta \leqslant \alpha.$$

Since $\mu(G) \geqslant 1/8$, it suffices to show that

$$\forall\, \omega \in G,\quad f'(\omega) - g(\omega) \geqslant \frac{L\gamma^2}{8}\,\beta^2. \tag{4.4.30}$$

Step 4. — We prove (4.4.30). Consider $\omega \in G$. Then $f'(\omega) = f(\omega) \geqslant m$. Consider $\omega' \in B_s$ with $h(\omega, \omega') \leqslant 3 + w(\mu(B_s))$. Then

$$f(\omega) - \hat{f}(\omega) \geqslant \sup_{0 \leqslant \lambda \leqslant 1} \left((1 - \lambda)(m - s) - \frac{1}{L}\,\psi((1 - \lambda)(3 + w(\mu(B_s))))\right). \tag{4.4.31}$$

We choose $0 \leqslant \lambda \leqslant 1$ such that

$$1 - \lambda = \frac{L\gamma^2}{4}\,\frac{\beta}{2 + w(\mu(B_s))}.$$

This is possible since $\beta \leqslant \alpha \leqslant 8/L\gamma^2$. Then (4.4.31) yields, since $\psi(x) \leqslant x^2$ for $x \leqslant 1$, that

$$f(\omega) - \hat{f}(\omega) \geqslant \frac{L\gamma^2 \beta^2}{8}.$$

Thus $\hat{f}(\omega) \leqslant f(\omega) - L\gamma^2 \beta^2/8$. Since the right-hand side is $\geqslant m'$, we have

$$g(\omega) \leqslant f(\omega) - L\gamma^2 \beta^2/8.$$

The proof is complete.

4.5. *Interpolation*

The result of this section will interpolate between (a weak form of) Theorem 3.1.1, for $q = 2$, and (a weak form of) Theorem 4.1.1. Consider three points $x = (x_i)_{i \leqslant N}$, $y^1 = (y_i^1)_{i \leqslant N}$, $y^2 = (y_i^2)_{i \leqslant N}$ of Ω^N. Set

$$r_i(x, y^1, y^2) = (1_{\{x_i \neq y_i^1\}}, 1_{\{x_i \neq y_i^2\}}, 1_{\{x_i \notin \{y_i^1, y_i^2\}\}}).$$

Thus $r_i(x, y^1, y^2) \in \{0, 1\}^3$. Set

$$r(x, y^1, y^2) = (r_i(x, y^1, y^2))_{i \leqslant N} \in (\{0, 1\}^3)^N.$$

Given two subsets A_1, A_2 of Ω^N, let

$$U_{A_1, A_2}(x) = \{ r(x, y^1, y^2) ; y^1 \in A_1, y^2 \in A_2 \},$$

and consider the convex hull $V_{A_1, A_2}(x)$ of $U_{A_1, A_2}(x)$, when $U_{A_1, A_2}(x)$ is seen as a subset of $(\mathbf{R}^3)^N$.

Throughout this section, we define $b > 0$ by $e^{6b} = 3 - 2e^{-2b}$, so that $b < 1/6$. We make the convention to write a point $r \in (\mathbf{R}^3)^N$ as $(r_{1,i}, r_{2,i}, r_{3,i})_{i \leqslant N}$. We set

$$f(A_1, A_2, x) = \inf\{ \sum_{i \leqslant N} r_{1,i}^2 + r_{2,i}^2 + r_{3,i} ; r \in V_{A_1, A_2}(x) \}.$$

Theorem **4.5.1.** — *We have*

$$\int_{\Omega^N} \exp b f(A_1, A_2, x)\, dP(x) \leqslant \frac{1}{P(A_1)\, P(A_2)}.$$

To understand this statement better, set $u = f(x, A_1, A_2)$. Consider $r \in V_{A_1, A_2}(x)$ such that

$$\sum_{i \leqslant N} r_{1,i}^2 + r_{2,i}^2 + r_{3,i} \leqslant u.$$

Consider numbers $(c_{1,i})_{i \leqslant N}$, $(c_{2,i})_{i \leqslant N}$. Then, for $j = 1, 2$

$$\sum_{i \leqslant N} c_{j,i}\, r_{j,i} \leqslant (\sum_{i \leqslant N} c_{j,i}^2)^{1/2} (\sum_{i \leqslant N} r_{j,i}^2)^{1/2}$$

$$\leqslant u^{1/2} (\sum_{i \leqslant N} c_{j,i}^2)^{1/2}.$$

Thus

$$\sum_{i\leqslant N} (c_{1,i}\, r_{1,i} + c_{2,i}\, r_{2,i} + br_{3,i}) \leqslant t =: u + u^{1/2}\Big(\sqrt{\sum_{i\leqslant N} c_{1,i}^2} + \sqrt{\sum_{i\leqslant N} c_{2,i}^2}\Big).$$

If we recall that $V_{A_1, A_2}(x)$ is the convex hull of $U_{A_1, A_2}(x)$, this implies that we can find $y^1 \in A_1$, $y^2 \in A_2$ such that

$$\Sigma\{ c_{1,i};\, x_i \neq y_i^1 \} + \Sigma\{ c_{2,i};\, x_i \neq y_i^2 \} + \operatorname{card}\{ i;\, x_i \notin \{y_i^1, y_i^2\}\} \leqslant t.$$

The proof of Theorem 4.5.1 goes by induction over N. The case $N = 1$ is left to the reader. For the induction from N to $N + 1$, one observes, with the usual notation, that, when $a_{0,0}, a_{1,0}, a_{0,1}, a_{1,1} \geqslant 0$ are of sum one, then

$$\begin{aligned} f(A_1, A_2, (x, \omega)) \leqslant a_{0,0} f(A_1(\omega), A_2(\omega), x) &+ a_{1,0} f(B_1, A_2(\omega), x) \\ &+ a_{0,1} f(A_1(\omega), B_2, x) + a_{1,1} f(B_1, B_2, x) \\ &+ (a_{1,0} + a_{1,1})^2 + (a_{0,1} + a_{1,1})^2 + a_{1,1}. \end{aligned}$$

Thereby, to perform the induction it suffices to show that, when g_1, g_2 are two functions on Ω, $g_1, g_2 \leqslant 1$, then

$$\text{(4.5.1)} \qquad \int \inf \exp(ba_{1,1} + b(a_{0,1} + a_{1,1})^2 + b(a_{1,0} + a_{1,1})^2)\, \frac{1}{(g_1 g_2)^{a_{0,0}}}\, \frac{1}{g_2^{a_{1,0}}}\, \frac{1}{g_1^{a_{0,1}}}\, d\mu \leqslant \frac{1}{\int g_1\, d\mu \int g_2\, d\mu}$$

where the infimum is taken over all the allowed choices of $a_{0,0}, a_{0,1}, a_{1,0}, a_{1,1}$.

Lemma **4.5.2.** — *We have*

$$\text{(4.5.2)} \qquad \inf \exp(ba_{1,1} + b(a_{1,0} + a_{1,1})^2 + b(a_{0,1} + a_{1,1})^2)\, \frac{1}{(g_1 g_2)^{a_{0,0}}}\, \frac{1}{g_2^{a_{1,0}}}\, \frac{1}{g_1^{a_{0,1}}} \leqslant \sqrt{(3 - 2g_1)(3 - 2g_2)}.$$

We first use (4.5.2) to prove (4.5.1). By (4.5.2) and Cauchy-Schwarz, the left-hand side of (4.5.1) is bounded by

$$\sqrt{\int (3 - 2g_1)\, d\mu \int (3 - 2g_2)\, d\mu} = \sqrt{\Big(3 - 2\int g_1\, d\mu\Big)\Big(3 - 2\int g_2\, d\mu\Big)}.$$

Thus it suffices to observe that for $0 \leqslant x \leqslant 1$, we have $3 - 2x \leqslant x^{-2}$, which expresses the fact that the convex function x^{-2} is above its tangent at $x = 1$.

Proof of Lemma 4.5.2. — We will actually restrict the infimum to the cases $a_{1,1} = 1$ or $a_{1,1} = 0$. We will prove

$$\text{(4.5.3)} \qquad \min\left(e^{3b}, \inf_{0 \leqslant a_1 + a_2 \leqslant 1} e^{b(a_1^2 + a_2^2)} \frac{1}{g_1^{1-a_1}} \frac{1}{g_2^{1-a_2}}\right) \leqslant \sqrt{(3 - 2g_1)(3 - 2g_2)}.$$

We distinguish cases.

Case 1: $g_1 g_2 \leqslant e^{-2b}$.

It suffices to see that

$$e^{3b} \leqslant \sqrt{(3 - 2g_1)(3 - 2g_2)}.$$

The right-hand side has minimum at $g_1 = 1$, $g_2 = e^{-2b}$, and our value of b has been chosen so that inequality holds in that case.

Case 2: $g_1 g_2 \geqslant e^{-2b}$.

For $j = 1, 2$, we take $a_j = -\dfrac{\log g_j}{2b}$. The purpose of the condition $g_1 g_2 \geqslant e^{-2b}$ is to ensure that $a_1 + a_2 \leqslant 1$. It suffices to prove the inequality

$$\frac{1}{g_1} e^{-\frac{(\log g_1)^2}{4b}} \leqslant \sqrt{3 - 2g_1}.$$

We will show that, for $0 \leqslant x \leqslant 1$,

$$e^{-(\log x)^2/2b} \leqslant x^2(3 - 2x)$$

or, equivalently that

$$\varphi(x) = \frac{(\log x)^2}{2b} + 2\log x + \log(3 - 2x) \geqslant 0.$$

Since $\varphi'(0) = 0$, $\varphi(0) = 0$ it suffices to show that $(x\varphi'(x))' \geqslant 0$, i.e.

$$\frac{1}{b} - \frac{6x}{(3 - 2x)^2} \geqslant 0.$$

But, since $\dfrac{1}{b} \leqslant 6$, it suffices to show that $x \leqslant (3 - 2x)^2$, which is true since $x \leqslant 1$, $(3 - 2x)^2 \geqslant 1$. □

5. The Symmetric Group

We denote by S_N the group of permutations of $\{1, \ldots, N\}$. Our interest in the symmetric group stems from the fact that it is closely related to a product. To see this, let us denote by $t_{i,j}$ the transposition of i and j. Then, it is easily seen that every $\sigma \in S_N$ can be written in a unique way as

$$\sigma = t_{N,i(N)} \circ t_{N-1,i(N-1)} \cdots \circ t_{2,i(2)} \tag{5.1}$$

where, for $j \leqslant N$, we have $i(j) \leqslant j$. This decomposition allows to transfer (2.1.3) to S_N. The result thus obtained is the result of Maurey [M1]. The purpose of the present chapter is to prove a version of Theorem 4.1.1 for S_N that will improve upon Maurey's result the way Theorem 4.1.1 improves upon Proposition 2.1.1. The reason why proving this is not such an easy task is that the decomposition (5.1) is highly non-commutative.

For a subset A of S_N, and $\sigma \in S_N$, we set

$$U_A(\sigma) = \{ s \in \{0, 1\}^N;\ \exists \tau \in A;\ \forall \ell \leqslant N,\ s_\ell = 0 \Rightarrow \tau(\ell) = \sigma(\ell) \}$$

and we consider the convex hull $V_A(\sigma)$ of $U_A(\sigma)$ in $[0, 1]^N$. We set

$$f(A, \sigma) = \inf\{ \sum_{\ell \leqslant N} s_\ell^2;\ s = (s_\ell) \in V_A(\sigma) \}.$$

We denote by P_N the canonical (= homogenous) probability on S_N.

Theorem **5.1.** — *For every subset* A *of* S_N *we have*

$$\int_{S_N} \exp \frac{1}{16} f(A, \sigma)\, dP_N(\sigma) \leqslant \frac{1}{P_N(A)}. \tag{5.2}$$

In a natural way, S_N can be considered as a subset of $\{1, \ldots, N\}^N$ by the map $\sigma \mapsto (\sigma(i))_{i \leqslant N}$. If S_N were equal to all of $\{1, \ldots, N\}^N$, (5.2) would be a consequence of Theorem 4.1.1, but S_N is only a very small subset of $\{1, \ldots, N\}^N$.

The challenge of Theorem 5.1 is that it is apparently not possible to prove (5.2) by induction over N. Rather, we will use a stronger induction hypothesis. Given $p \leqslant N$, we set

$$f(A, \sigma, p) = \inf\{ s_p^2 + \sum_{\ell \leqslant N} s_\ell^2;\ s \in V_A(\sigma) \}.$$

Theorem 5.1 is obviously a consequence of the following.

Proposition **5.2.** — *For each subset* A *of* S_N *and each* $p \leqslant N$, *we have*

$$\int_{S_N} \exp \frac{1}{16} f(A, \sigma, p)\, dP_N(\sigma) \leqslant \frac{1}{P_N(A)}, \tag{5.3$_N$}$$

$$\int_{S_N} \exp \frac{1}{16} f(A, \sigma, \sigma^{-1}(p))\, dP_N(\sigma) \leqslant \frac{1}{P_N(A)}. \tag{5.4$_N$}$$

We leave to the reader to prove Proposition 5.2 when $N = 1$. We now assume that Proposition 5.2 has been proved for N and we prove it for $N + 1$. A noticeable feature of this proof is that the proof of $(5.3)_{N+1}$ (resp. $(5.4)_{N+1}$) will require the use of $(5.4)_N$ (resp. $(5.3)_N$). Before the proof starts, we need to introduce some notation. Given $p, m \leq N$, $p \neq m$, we set

$$f(A, \sigma, p, m) = \inf\{s_p^2 + \sum_{\ell \leq N} s_\ell^2 ; s \in V_A(\sigma), s_m = 0\}. \tag{5.5}$$

Given $i, j \leq N$, we set

$$g(A, \sigma, i, j) = \inf\{\sum_{\ell \neq i, j} s_\ell^2 ; s \in V_A(\sigma)\}. \tag{5.6}$$

We start the proof of $(5.4)_{N+1}$. Certainly there is no loss of generality to assume that $p = N + 1$.

Lemma **5.3.** — *Consider* $i, j \leq N + 1$, $i \neq j$, $\sigma \in S_{N+1}$, $0 \leq \lambda \leq 1$. *Then*

$$f(A, \sigma, i) \leq 4(1 - \lambda)^2 + (1 - \lambda)\, g(A, \sigma, i, j) + \lambda f(A, \sigma, j, i). \tag{5.7}$$

Proof. — Consider $s \in V_A(\sigma)$, $t \in V_A(\sigma)$, with $t_i = 0$. By convexity of $V_A(\sigma)$, we have

$$u = (1 - \lambda)\, s + \lambda t \in V_A(\sigma).$$

Thus

$$f(A, \sigma, i) \leq \sum_{\ell \neq i} u_\ell^2 + 2u_i^2.$$

Since $s_i \leq 1$ and since $t_i = 0$, we have

$$f(A, \sigma, i) \leq \sum_{\ell \neq i, j} u_\ell^2 + 2(1 - \lambda)^2 + ((1 - \lambda)\, s_j + \lambda t_j)^2.$$

Since $s_j \leq 1$, we have

$$((1 - \lambda)\, s_j + \lambda t_j)^2 \leq 2(1 - \lambda)^2 s_j^2 + 2\lambda^2 t_j^2 \leq 2(1 - \lambda)^2 + 2\lambda t_j^2.$$

Since the function $x \mapsto x^2$ is convex, we have

$$u_\ell^2 \leq (1 - \lambda)\, s_\ell^2 + \lambda t_\ell^2.$$

Thus

$$f(A, \sigma, i) \leq (1 - \lambda) \sum_{\ell \neq i, j} s_\ell^2 + \lambda(2t_j^2 + \sum_{\ell \leq N} t_\ell^2) + 4(1 - \lambda)^2.$$

The result follows by taking the infimum over s, t. □

Following the idea of Theorem 4.1.1, (5.7) will be used together with Holder's inequality. Some work is, however, needed to relate the resulting terms to the induction hypothesis. For $i \leq N + 1$, we set

$$G_i = \{\sigma \in S_{N+1} ; \sigma(i) = N + 1\}.$$

For simplicity, we denote by $t_i = t_{N+1,i}$ the transposition of $N+1$ and i. We consider the map $R : \rho \mapsto \rho \circ t_i$. We observe that, if $\rho \in G_i$, then

$$R(\rho)\,(N+1) = \rho \circ t_i(N+1) = \rho(i) = N+1.$$

Thereby, we can consider R as a map from G_i to S_N. We set $A_i = A \cap G_i$.

Lemma **5.4.** — *If* $\sigma \in G_i$, *we have*

$$f(A, \sigma, j, i) \leqslant f(R(A_i), R(\sigma), t_i(j)). \tag{5.8}$$

Proof. — We let the reader consider the essentially obvious case where $i = N+1$, and we assume $i \neq N+1$. Given a sequence $s \in \{0, 1\}^N$, we consider the sequence $\bar{s} = (\bar{s}_\ell) \in \{0, 1\}^{N+1}$ defined by $\bar{s}_i = 0$, $\bar{s}_{N+1} = s_i$, $\bar{s}_\ell = s_\ell$ if $\ell \neq i, N+1$. We note that $\bar{s}_\ell = s_{t_i(\ell)}$ for $\ell \neq i$. Thus it suffices to prove that $\bar{s} \in U_A(\sigma)$ whenever $s \in U_{R(A_i)}(R(\sigma))$. Consider $s \in U_{R(A_i)}(R(\sigma))$. By definition, there exists $\tau \in R(A_i)$ such that, for $\ell \leqslant N$

$$s_\ell = 0 \Rightarrow \tau(\ell) = R(\sigma)\,(\ell).$$

Since $\tau \in R(A_i)$, we have $\tau = R(\rho)$ for a certain $\rho \in A_i$. Thus

$$s_\ell = 0 \Rightarrow \rho(t_i(\ell)) = \sigma(t_i(\ell)). \tag{5.9}$$

We will show that, for $\ell \leqslant N+1$,

$$\bar{s}_\ell = 0 \Rightarrow \rho(\ell) = \sigma(\ell).$$

This holds for $\ell = i$, since $\rho(i) = \sigma(i) = N+1$. For $\ell \neq i$, this follows from (5.9), since $\bar{s}_\ell = s_{t_i(\ell)}$, and $t_i \circ t_i$ is the identity of S_N. □

We denote by Q_i the uniform probability on G_i.

Corollary **5.5.** —

$$\int \exp \frac{1}{16} f(A, \sigma, j, i)\, dQ_i(\sigma) \leqslant \frac{1}{Q_i(A_i)} = \frac{1}{Q_i(A)}. \tag{5.10}$$

Proof. — Using (5.8) and $(5.3)_N$, the left-hand side of (5.10) is bounded by

$$\int \exp \frac{1}{16} f(R(A_i), R(\sigma), t_i(j))\, dQ_i(\sigma) = \int \exp \frac{1}{16} f(R(A_i), \rho, t_i(j))\, dP_N(\rho)$$
$$\leqslant \frac{1}{P_N(R(A_i))} = \frac{1}{Q_i(A)}. \quad \square$$

Lemma **5.6.** — *Assume* $j \neq i$. *Then*

$$\int \exp \frac{1}{16} g(A, \sigma, i, j)\, dQ_i(\sigma) \leqslant \frac{1}{Q_j(A)}. \tag{5.11}$$

Proof. — The map $S : \rho \mapsto \rho \circ t_{ij}$ is one-to-one from G_j to G_i. We will prove that setting $B = R(S(A_j))$, we have

$$g(A, \sigma, i, j) \leqslant f(B, R(\sigma)) \tag{5.12}$$

where we recall that R is seen as a map from G_i to S_N. Since $P_N(B) = Q_j(A)$, (5.11) will follow from either $(5.3)_N$ or $(5.4)_N$ as in the proof of Corollary 5.5.

Given a sequence $s \in \{0, 1\}^N$, we consider the sequence $\bar{s} \in \{0, 1\}^{N+1}$ defined as follows. We set $\bar{s}_i = \bar{s}_j = 1$. If $N + 1 \neq i, j$, we set $\bar{s}_{N+1} = s_i$. If $\ell \notin \{i, j, N+1\}$, we set $\bar{s}_\ell = s_\ell$.

We will show that when $s \in U_B(R(\sigma))$, then $\bar{s} \in U_A(\sigma)$. By definition of $U_B(R(\sigma))$, there exists $\tau \in B$ such that

$$s_\ell = 0 \Rightarrow \tau(\ell) = R(\sigma)(\ell) = \sigma \circ t_i(\ell).$$

Since $\tau \in B$, we can write $\tau = \rho \circ t_{ij} \circ t_i$, where $\rho \in A_j$. Thus

$$s_\ell = 0 \Rightarrow \rho \circ t_{ij} \circ t_i(\ell) = \sigma \circ t_i(\ell).$$

We will show that for $\ell \leqslant N + 1$ we have

$$\bar{s}_\ell = 0 \Rightarrow \rho(\ell) = \sigma(\ell).$$

The only nontrivial case is $\ell = N + 1$, when $N + 1 \neq i, j$. In that case, when $\bar{s}_{N+1} = 0$, we have $s_i = 0$, so that $\tau(i) = R(\sigma)(i) = \sigma(N + 1)$. But

$$\tau(i) = \rho \circ t_{ij} \circ t_i(i) = \rho \circ t_{ij}(N + 1) = \rho(N + 1).$$

since $N + 1 \neq i, j$. □

We now complete the proof of $(5.4)_{N+1}$. We select j such that $Q_j(A)$ is maximum. If $i \leqslant N + 1$, $i \neq j$, for $0 \leqslant \lambda \leqslant 1$, we have, using Lemmas 5.3, 5.4, Corollary 5.5 and Holder's inequality

$$\begin{aligned}\int \exp \frac{1}{16} f(A, \sigma, i)\, Q_i(\sigma) &\leqslant \exp\left[\frac{1}{4}(\lambda - 1)^2\right] \frac{1}{Q_i(A)^\lambda} \frac{1}{Q_j(A)^{1-\lambda}} \\ &= \frac{1}{Q_j(A)} \left(\frac{Q_i(A)}{Q_j(A)}\right)^{-\lambda} \exp \frac{1}{4}(1 - \lambda)^2.\end{aligned}$$

If we appeal to Lemma 4.1.3, we have

$$\int \exp \frac{1}{16} f(A, \sigma, i)\, dQ_i(\sigma) \leqslant \frac{1}{Q_j(A)} \left(2 - \frac{Q_i(A)}{Q_j(A)}\right). \tag{5.13}$$

It should be obvious from the induction hypothesis that (5.13) still holds for $i = j$. Since $P_{N+1} = \Sigma_{i \leqslant N+1} \frac{1}{N+1} Q_i$, we have, from (5.13), and since $i = \sigma^{-1}(N+1)$ for $\sigma \in G_i$ that

$$\int \exp \frac{1}{16} f(A, \sigma, \sigma^{-1}(N+1))\, dP_{N+1}(\sigma) \leqslant \frac{1}{Q_j(A)}\left(2 - \frac{P_{N+1}(A)}{Q_j(A)}\right) \leqslant \frac{1}{P_{N+1}(A)}. \quad \square$$

Having proved $(5.4)_{N+1}$, we turn towards the proof of $(5.3)_{N+1}$. We can assume again $p = N + 1$. The proof is not identical to that of $(5.4)_{N+1}$, but is completely parallel.

Lemma **5.7.** — *For* $\sigma \in S_{N+1}$, $j \leqslant N+1$, $j \neq \sigma(N+1)$, $0 \leqslant \lambda \leqslant 1$, *we have*

$$\textbf{(5.14)} \qquad f(A, \sigma, N+1) \leqslant 4(1-\lambda)^2 + (1-\lambda)\, g(A, \sigma, N+1, \sigma^{-1}(j)) + \lambda f(A, \sigma, \sigma^{-1}(j), N+1).$$

Proof. — This is (5.7) if one replaces i by $N + 1$, j by $\sigma^{-1}(j)$. $\square$

We set

$$G'_i = \{\, \sigma \in S_{N+1};\ \sigma(N+1) = i \,\}.$$

We fix i, and we consider the map $R' : \rho \mapsto t_i \circ \rho$. Thus, for $\rho \in G'_i$, we have $R'(\rho)\,(N+1) = t_i(i) = N+1$, and we can view R' as a map from G'_i to S_N. We set $A'_i = A \cap G'_i$.

Lemma **5.8.** — *If* $\sigma \in G'_i$, $i \neq j$, *we have*

$$\textbf{(5.15)} \qquad f(A, \sigma, \sigma^{-1}(j), N+1) \leqslant f(R'(A'_i), R'(\sigma), R'(\sigma)^{-1}(t_i(j))).$$

Proof. — Given a sequence $s \in \{0, 1\}^N$, we consider the sequence $\bar{s} = (\bar{s}_\ell) \in \{0, 1\}^{N+1}$ defined by $\bar{s}_\ell = s_\ell$ if $\ell \neq N+1$, and $\bar{s}_{N+1} = 0$. Since $\sigma^{-1}(j) = R'(\sigma)^{-1}(t_i(j)) \neq N+1$, it suffices to prove that $\bar{s} \in U_A(\sigma)$ whenever $s \in U_{R'(A'_i)}(R'(\sigma))$. Thus consider s in this later set. By definition, there exists $\tau \in R'(A'_i)$ such that

$$\forall\, \ell \leqslant N, \quad s_\ell = 0 \Rightarrow \tau(\ell) = R'(\sigma)\,(\ell).$$

Since $\tau \in R'(A'_i)$, we have $\tau = R'(\rho)$, $\rho \in A'_i$. Thus,

$$\forall\, \ell \leqslant N, \quad s_\ell = 0 \Rightarrow t_i \circ \rho(\ell) = t_i \circ \sigma(\ell) \Rightarrow \rho(\ell) = \sigma(\ell).$$

Since $\rho(N+1) = \sigma(N+1) = i$, we then have

$$\forall\, \ell \leqslant N+1, \quad \bar{s}_\ell = 0 \Rightarrow \rho(\ell) = \sigma(\ell).$$

Thus $\bar{s} \in U_A(\sigma)$. $\square$

We denote by Q'_i the homogeneous probability on G'_i.

Corollary **5.9.** — *If* $j \neq i$,

$$\int \exp \frac{1}{16} f(A, \sigma, \sigma^{-1}(j), N+1) \, dQ'_i(\sigma) \leq \frac{1}{Q'_i(A)}. \tag{5.16}$$

Proof. — Using (5.15) and the fact that R' transports Q'_i to P_N, the left-hand side of (5.16) is bounded by

$$\int \exp \frac{1}{16} f(R'(A_i), \rho, \rho^{-1}(t_i(j))) \, dP_N(\rho) \leq \frac{1}{P_N(R'(A_i))} = \frac{1}{Q'_i(A)}$$

using $(5.4)_N$. □

Lemma **5.10.** — *If* $i \neq j$, *we have*

$$\int \exp \frac{1}{16} g(A, \sigma, N+1, \sigma^{-1}(j)) \, dQ'_i(\sigma) \leq \frac{1}{Q'_j(A)}. \tag{5.17}$$

Proof. — The map $S' : \rho \to t_{ij} \circ \rho$ is one-to-one from G'_j to G'_i. We will prove that, setting $B = R' \circ S'(A_j)$, we have, for σ in G'_i that

$$g(A, \sigma, N+1, \sigma^{-1}(j)) \leq f(B, R'(\sigma)) \tag{5.18}$$

where we recall that R' is seen as a map from G'_i to S_N. Since $P_N(B) = Q'_j(A)$, (5.17) will then follow from either $(5.3)_N$ or $(5.4)_N$.

Given a sequence $s \in \{0, 1\}^N$, we consider the sequence $\bar{s} \in \{0, 1\}^{N+1}$ defined as follows. We set $\bar{s}_{N+1} = \bar{s}_{\sigma^{-1}(j)} = 1$. We set $\bar{s}_\ell = s_\ell$ if $\ell \notin \{N+1, \sigma^{-1}(j)\}$. To prove (5.18) it suffices to prove that if $s \in U_B(R'(\sigma))$, then $\bar{s} \in U_A(\sigma)$. Thus, consider $s \in U_B(R'(\sigma))$. By definition, there exists $\tau \in B$ such that

$$s_\ell = 0 \Rightarrow \tau(\ell) = R'(\sigma)(\ell) = t_i \circ \sigma(\ell). \tag{5.19}$$

Since $\tau \in B$, we can write $\tau = t_i \circ t_{ij} \circ \rho$, where $\rho \in A'_j$. Thus, by (5.19)

$$s_\ell = 0 \Rightarrow t_{ij} \circ \rho(\ell) = \sigma(\ell) \Rightarrow \rho(\ell) = t_{ij} \circ \sigma(\ell).$$

Now, for $\ell \neq N+1, \sigma^{-1}(j)$, we have $\sigma(\ell) \neq i, j$; thus $t_{ij} \circ \sigma(\ell) = \sigma(\ell)$. Thus for these values of ℓ we have

$$\bar{s}_\ell = 0 \Rightarrow s_\ell = 0 \Rightarrow \rho(\ell) = \sigma(\ell). \quad \square$$

The end of the proof of $(5.3)_{N+1}$ is similar to the end of the proof of $(5.4)_{N+1}$, and is left to the reader.

II. APPLICATIONS

6. Bin Packing

Given a collection $x_1, \ldots, x_N$ of items, of sizes $\leqslant 1$, the bin packing problem requires finding the minimum number $B_N(x_1, \ldots, x_N)$ of unit size bins in which the items $x_1, \ldots, x_N$ can be packed, subject to the restriction that the sum of the sizes of items attributed to a given bin cannot exceed one. (For simplicity, we will denote items and item sizes by the same letters.) The bin packing problem is a fundamental question of computer science, and, accordingly, has received considerable attention. Much work has been done on stochastic models [C-L]. In the model we will consider, the items $X_1, \ldots, X_N$ are independently distributed according to a given distribution μ. One of the natural questions that arises is the study of the fluctuations of the random variable $B_N(X_1, \ldots, X_N)$. One early result, [R-T1], [McD1], using martingales, is that for all $t > 0$, one has

$$(6.1) \qquad P(|B_N(X_1, \ldots, X_N) - EB_N(X_1, \ldots, X_N)| \geqslant t) \leqslant 2 \exp\left(-\frac{2t^2}{N}\right).$$

However, especially when EX_1 is small, one expects that the behavior of $B_N(X_1, \ldots, X_N)$ resembles the behavior of $\sum_{i \leqslant N} X_i$. Thereby one should expect that the exponent in the right-hand side of (6.1) should be of order $t^2/N \operatorname{var}(X_1)$, or, at least, less ambitiously, $t^2/NE(X_1^2)$. This is apparently not so easy to prove, and despite several attempts, was established only recently using non-trivial bin-packing theory [R4]. The purpose of the present section is to prove this result as an application of Theorem 4.1.1. Several features of the proof will appear repeatedly in future applications. One advantage of our approach is that it uses only trivial facts about bin packing, such as the following observation.

Lemma **6.1.** — *We have*

$$B_N(x_1, \ldots, x_N) \leqslant 2 \sum_{i \leqslant N} x_i + 1.$$

Proof. — It suffices to construct a packing in which at most one bin is less than half full. Such a packing exists since bins that are less than half full can be merged. □

We take $\Omega = [0, 1]$. For a subset A of Ω^N, and $x \in \Omega^N$, we recall the notation $f_c(A, x)$ introduced in Section 4.1. For $x = (x_1, \ldots, x_N) \in \Omega^N$, we write simply $B_N(x)$ rather than $B_N(x_1, \ldots, x_N)$. For $x \in \Omega^N$, we set $\| x \|_2 = (\sum_{i \leqslant N} x_i^2)^{1/2}$. Finally, for $a > 0$, we set

$$A(a) = \{y \in \Omega^N; B_N(y) \leqslant a\}.$$

The crucial observation is as follows.

Lemma **6.2.** — *For all* $x \in \Omega^N$, *we have*

$$(6.2) \qquad B_N(x) \leqslant a + 2 \| x \|_2 f_c(A(a), x) + 1.$$

Proof. — As follows from Lemma 4.1.2 (taking α_i there equal to x_i) we can find $y \in A(a)$ such that, if I denotes the set of indices $i \leqslant N$ for which $x_i = y_i$, we have

$$\sum_{i \notin I} x_i \leqslant \| x \|_2 f_c(A(a), x).$$

By Lemma 6.1 the items $(x_i)_{i \notin I}$ can be packed using at most $2 \| x \|_2 f_c(A(a), x) + 1$ bins. The items $(x_i)_{i \in I}$ are exactly the items $(y_i)_{i \in I}$, so they can certainly be packed using at most a bins, since $y \in A(a)$. The result follows. □

We provide $[0, 1]$ with the measure μ, and we denote by P the product probability on $[0, 1]^N$. The term $\| x \|_2$ of (6.2) will be disposed of by the following simple observation.

Lemma **6.3.** — *We have*

$$\textbf{(6.3)} \qquad P(\| x \|_2 \geqslant 2\sqrt{N}(EX_1^2)^{1/2}) \leqslant \exp(-2NEX_1^2).$$

Proof. — Since $e^x \leqslant 1 + 2x$ for $x \leqslant 1$, we have

$$E \exp X_i^2 \leqslant 1 + 2EX_i^2 \leqslant \exp 2EX_1^2$$

so that

$$E \exp(\sum_{i \leqslant N} X_i^2) \leqslant \exp 2NEX_1^2$$

from which (6.3) follows by Chebyshev inequality. □

We can now prove the basic inequality.

Proposition **6.4.** — *We have, for all* $t > 0$ *and all* $a > 0$, *that*

$$\textbf{(6.4)} \qquad P(B_N(x) \leqslant a)\, P(B_N(x) \geqslant a + 4t\sqrt{N}(EX_1^2)^{1/2} + 1) \leqslant e^{-t^2/4} + e^{-2NEX_1^2}.$$

Proof. — Indeed, by (6.2), if $B_N(x) \geqslant a + 4t\sqrt{N}(EX_1^2)^{1/2} + 1$, we have either $f_c(A(a), x) \geqslant t$ or $\| x \|_2 \geqslant 2\sqrt{N}(EX_1^2)^{1/2}$. The result then follows from (4.1.2) and (6.2).

Theorem **6.5.** — *Denote by* M *a median of* $B_N(x)$. *Then for all* $u \leqslant 8\sqrt{2}NEX_1^2$ *we have*

$$P(| B_N(X_1, \ldots, X_N) - M | \geqslant 1 + u) \leqslant 8 \exp\left(-\frac{u^2}{64NEX_1^2}\right).$$

Proof. — First, we take $a = M$ to obtain from (6.4), setting $u = 4t\sqrt{N}(EX_1^2)^{1/2}$ and since $P(B_N \leqslant M) \geqslant 1/2$,

$$\begin{aligned} P(B_N \geqslant M + u + 1) &\leqslant 2(e^{-t^2/4} + e^{-2NEX_1^2}) \\ &\leqslant 4e^{-t^2/4}. \end{aligned}$$

The bound for $P(B_N \leqslant M - u - 1)$ follows similarly taking $a = M - u - 1$. □

Remarks. — 1) One can also get bounds for larger values of u, by adapting Lemma 6.3.

2) It is instructive to find an alternate proof of Theorem 6.5 using Corollary 2.2.4 rather than Theorem 4.1.1.

7. Subsequences

7.1. *The longest increasing subsequence*

Consider points $x_1, \ldots, x_N$ of $[0, 1]$. We denote by $L_N(x_1, \ldots, x_N)$ the length of the longest increasing subsequence of $x_1, \ldots, x_N$. That is, the largest integer p such that we can find $i_1 < \ldots < i_p$ for which $x_{i_1} \leqslant \ldots \leqslant x_{i_p}$. It is simple to see that when $X_1, \ldots, X_N$ are independent uniformly distributed over $[0, 1]$ (or, actually, distributed according to any non atomic probability), the r.v. $L_N(X_1, \ldots, X_N)$ is distributed like the longest increasing subsequence of a random permutation σ of $\{1, \ldots, N\}$ (where the symmetric group S_N is of course provided with the uniform probability). The concentration of $L_N(X_1, \ldots, X_N)$ around its mean has been studied in particular in [F] and [B-B]. Sharper results will be obtained here as a simple consequence of Theorem 4.1.1. We consider $\Omega = [0, 1]^N$. For $x = (x_i)_{i \leqslant N}$ in Ω, we set $L_N(x) = L_N(x_1, \ldots, x_N)$. For $a > 0$, we set

$$A(a) = \{ x \in \Omega ; L_N(x) \leqslant a \}.$$

The basic observation is as follows.

Lemma **7.1.1.** — *For all* $x \in \Omega^N$, *we have*

$$a \geqslant L_N(x) - f_c(A(a), x) \sqrt{L_N(x)}. \tag{7.1.1}$$

In particular,

$$L_N(x) \geqslant a + v \Rightarrow f_c(A(a), x) \geqslant \frac{v}{\sqrt{a + v}}. \tag{7.1.2}$$

Proof. — For simplicity, we write $b = L_N(x)$. By definition, we can find a subset I of $\{1, \ldots, N\}$ of cardinality b such that if $i, j \in I$, $i < j$, then $x_i < x_j$. By Lemma 4.1.2 (taking $\alpha_i = 1$ if $i \in I$ and $\alpha_i = 0$ otherwise), there exists $y \in A(a)$ such that $\operatorname{card} J \leqslant f_c(A(a), x) \sqrt{b}$, where $J = \{ i \in I ; y_i \neq x_i \}$. Thus $(x_i)_{i \in I \setminus J}$ is an increasing subsequence of y; since $y \in A(a)$, we have $\operatorname{card}(I \setminus J) \leqslant a$, which proves (7.1.1).

To prove (7.1.2), we observe that by (7.1.1) we have

$$f_c(A(x), x) \geqslant \frac{L_N(x) - a}{\sqrt{L_N(x)}}$$

and that the function $u \mapsto (u - a)/\sqrt{u}$ increases for $u \geqslant a$. □

We denote by M $(= M_N)$ a median of L_N.

Theorem **7.1.2.** — *For all* $u > 0$ *we have*

$$P(L_N \geqslant M + u) \leqslant 2 \exp - \frac{u^2}{4(M + u)}, \tag{7.1.3}$$

$$P(L_N \leqslant M - u) \leqslant 2 \exp - \frac{u^2}{4M}. \tag{7.1.4}$$

Proof. — To prove (7.1.3), we combine (7.1.2) with $M = a$ and (4.1.2). To prove (7.1.4), we use (7.1.2) with $a = M - u$, $v = u$ to see that

$$L_N(x) \geqslant M \Rightarrow f_c(A(M-u), x) \geqslant \frac{u}{\sqrt{M}}$$

so that

$$\text{(7.1.5)} \qquad P\left(f_c(A(M-u), x) \geqslant \frac{u}{\sqrt{M}}\right) \geqslant \frac{1}{2}.$$

On the other hand, by (4.4.2),

$$\text{(7.1.6)} \qquad P\left(f_c(A(M-u), x) \geqslant \frac{u}{\sqrt{M}}\right) \leqslant \frac{1}{P(A(M-u))} e^{-\frac{u^2}{4M}}.$$

Comparing (7.1.5), (7.1.6) gives the required bound on $P(A(M-u))$. □

It seems worthwhile to state an abstract version of Theorem 7.1.2. Let us say that a function $L_N : \Omega^N \to \mathbf{N}$ is a *configuration function* provided it has the following property.

(7.1.7) Given any $x = (x_i)_{i \leqslant N}$ in Ω^N, there exists a subset J of $\{1, \ldots, N\}$ with $\operatorname{card} J = L_N(x)$ such that, for each y in Ω^N, we have $L_N(y) \geqslant \operatorname{card}\{i \in J; y_i = x_i\}$.

The reason for this name is that, intuitively, L_N counts the size of the largest " configuration " formed by the points x_i.

The proof of the following is identical to that of Theorem 7.1.2.

Theorem **7.1.3.** — *If* L_N *is a configuration function, then* (7.1.3) *and* (7.1.4) *hold.*

7.2. *Longest common subsequence*

Consider two sequences $x = (x_1, \ldots, x_N)$, $y = (y_1, \ldots, y_N)$ of numbers. We define the length $L_{N,M}(x;y)$ of the longest common subsequence of x, y as the largest integer p for which there exists $1 \leqslant i_1 < \ldots < i_p \leqslant N$ and $1 \leqslant j_1 < \ldots < j_p \leqslant N'$ such that $x_{i_\ell} = y_{j_\ell}$ for each $\ell \leqslant p$. One interpretation of this is when $x_1, \ldots, x_N$ are chosen among a (small) finite number of possibilities (the letters of an alphabet) $L_{N,N'}(x;y)$ is then the length of the longest " subword " of the words x, y (and $N + N' - L_{N,N'}(x;y)$ is the so-called " edit distance " of the two words). These considerations arise in a number of situations, such as genetics, speech recognition, etc. Consider now a r.v. X, and two independent sequences $(X_i)_{i \leqslant N}$, $(Y_j)_{j \leqslant N'}$ independently distributed like X. We are interested in the random variable $L_{N,N'} = L_{N,N'}(X_1, \ldots, X_N; Y_1, \ldots, Y_{N'})$.

Theorem **7.2.1.** — *Consider a median* $M (= M_{N,N'})$ *of* $L_{N,N'}$. *Then, for all* $u > 0$, *we have*

$$\text{(7.2.1)} \qquad P(L_{N,N'} \geqslant M + u) \leqslant 2 \exp\left(-\frac{u^2}{32(M+u)}\right),$$

$$\text{(7.2.2)} \qquad P(L_{N,N'} \leqslant M - u) \leqslant 2 \exp\left(-\frac{u^2}{32M}\right).$$

Comments. — It is known that $\lim_{N\to\infty} E(L_{N,N})/N$ exists. However, this limit can be very small, in the case where X takes many possible values. In this case, we have $M \ll N$, and (7.2.1), (7.2.2) give a better result than Azuma's inequality.

Proof. — The proof is very similar to the proof of Theorem 7.1.2. Consider $\Omega = [0, 1]$, and for $x \in \Omega^{N+N'}$, consider

$$L(x) = L_{N,N'}(x_1, \ldots, x_N; x_{N+1}, \ldots, x_{N+N'}).$$

Consider the set

$$A(a) = \{ x; L(x) \leqslant a \}.$$

The basic inequality is that

$$a \geqslant L(x) - 2\sqrt{2} f_c(A(a), x)\sqrt{L(x)}. \tag{7.2.3}$$

To see this, we set $b = L(x)$; we can find indices

$$1 \leqslant i_1 < \ldots < i_b \leqslant N < i_{b+1} < \ldots < i_{2b} \leqslant N + N'$$

such that $x_{i_k} = x_{i_{k+b}}$ for $1 \leqslant k \leqslant b$. Consider the set $I = \{ i_k; 1 \leqslant k \leqslant 2b \}$. By Lemma 4.4.2, we can find $y \in A(a)$ such that

$$\operatorname{card}\{ i \in I; x_i \neq y_i \} \leqslant f_c(A(a), x)\sqrt{2b}. \tag{7.2.4}$$

Consider then

$$J = \{ k \leqslant b; x_{i_k} = y_{i_k}; x_{i_{k+b}} = y_{i_{k+b}} \}.$$

By (7.2.4) we see that

$$\operatorname{card} J \geqslant b - 2\operatorname{card}\{ i \in I; x_i \neq y_i \} \geqslant b - 2\sqrt{2b} f_c(A(a), x).$$

On the other hand, $L(y) \geqslant \operatorname{card} J$ since, for $k \in J$, we have $y_{i_k} = y_{i_{k+b}}$. Also, since $y \in A(a)$, we have $L(y) \leqslant a$. Condition (7.2.3) follows. The rest of the proof is identical to that of Theorem 7.2.2. □

Remark. — It is also possible to find a more general version of Theorem 7.1.3 that contains Theorem 7.2.1.

8. Infimum and Percolation

Consider an independent sequence $(X_i)_{i \leqslant N}$ of positive r.v. Consider a family $\mathcal{F}$ of N-tuples $\alpha = (\alpha_i)_{i \leqslant N}$ of positive numbers. Our prime topic of interest in the present section is the random variable

$$Z' = Z'_{\mathcal{F}} = \inf_{\alpha \in \mathcal{F}} \sum_{i \leqslant N} \alpha_i X_i. \tag{8.1}$$

It does matter a lot that we take an infimum rather than a supremum. The function of the ′ in Z′ is to indicate that we take such an infimum. Rather that (8.1) one can also write

$$Z' = - \sup_{\alpha \in \mathcal{F}} \sum_{i \leq N} (-\alpha_i) X_i$$

but the numbers $-\alpha_i$ are negative. In Section 13, we will have to study the r.v.

$$\textbf{(8.2)} \qquad Z = \sup_{\alpha \in \mathcal{F}} \sum_{i \leq N} \alpha_i X_i$$

where α_i and X_i can possibly have any signs. In order to avoid repetition, we will study the variables Z given by (8.2).

8.1. *The basic result*

Consider a family $\mathcal{F}$ of N-tuples $\alpha = (\alpha_i)_{i \leq N}$. We make no assumption on the sign of α_i. We set $\sigma = \sup_{\alpha \in \mathcal{F}} \| \alpha \|_2$, where $\| \alpha \|_2 = (\sum_{i \leq N} \alpha_i^2)^{1/2}$. We consider independent r.v. X_i, and we assume that for each i there is number r_i such that $r_i \leq X_i \leq r_i + 1$.

Theorem **8.1.1.** — *Consider the r.v.* Z *given by* (8.2), *and a median* M *of* Z. *Then, for all* $u > 0$, *we have*

$$\textbf{(8.1.1)} \qquad P(|Z - M| \geq u) \leq 4 \exp\left(-\frac{u^2}{4\sigma^2}\right).$$

Proof. — This will again follow from Theorem 4.1.1.

Step 1. — Set $\Omega = [0, 1]$, and for $x = (x_i)_{i \leq N} \in \Omega^N$, set

$$Z(x) = \sup_{\alpha \in \mathcal{F}} \sum_{i \leq N} \alpha_i (r_i + x_i).$$

Consider $a \in \mathbf{R}$, and $A(a) = \{ y \in \Omega^N ; Z(y) \leq a \}$. The basic observation is that

$$\textbf{(8.1.2)} \qquad \forall x \in \Omega^N, \quad Z(x) \leq a + \sigma f_c(A(a), x).$$

To prove this, consider $\alpha \in \mathcal{F}$. By Lemma 4.1.2, we can find $y \in A(x)$ such that, if $I = \{ i \leq N ; y_i \neq x_i \}$, then

$$\textbf{(8.1.3)} \qquad \sum_{i \in I} | \alpha_i | \leq \| \alpha \|_2 f_c(A(a), x) \leq \sigma f_c(A(a), x).$$

We then have

$$\Big| \sum_{i \leq N} \alpha_i (r_i + y_i) - \sum_{i \leq N} \alpha_i (r_i + x_i) \Big| \leq \sum_{i \in I} | \alpha_i | \, | y_i - x_i | \leq \sum_{i \in I} | \alpha_i |.$$

Thus, by (8.1.3)

$$\sum_{i \leq N} \alpha_i (r_i + x_i) \leq Z(y) + \sigma f_c(A(a), x) \leq a + \sigma f_c(A(a), x),$$

and taking the supremum over α proves (8.1.2).

Step 2. — We provide the i-th factor [0, 1] with the law μ_i of $X_i - r_i$. We denote by P the product probability. Thus by (8.1.2) and (4.1.2)

$$P(Z(x) \geq b) \leq P\left(f_c(A(a), x) \geq \frac{b-a}{\sigma}\right) \leq \frac{1}{P(A(a))} \exp\left(-\frac{(b-a)^2}{4\sigma^2}\right),$$

i.e.
$$P(Z(x) \geq b)\, P(Z(x) \leq a) \leq \exp\left(-\frac{(b-a)^2}{4\sigma^2}\right)$$

from which (8.1.1) follows as in Chapter 7, since the law of $Z(x)$ under P coincides with the law of Z. □

8.2. *General moments*

In the present section we rely on the theory of Section 4.4. We start with some preliminaries. Consider a convex function ψ on $\mathbf{R}^+$ that satisfies (4.4.1) and $\psi(0) = 0$. Consider a family $\mathscr{F}$ of N-tuples as in Section 8.1. For $u > 0$, we define

$$\psi_{\mathscr{F}}(u) = \inf\{ \sum_{i \leq N} \psi(s_i);\ \exists\, \alpha \in \mathscr{F},\ \sum_{i \leq N} s_i \mid \alpha_i \mid \geq u \}.$$

The simplest case is when $\psi(x) = x^2$. In that case it is easily seen that $\psi_{\mathscr{F}}(u) = u^2/\sigma^2$, where $\sigma^2 = \sup\{ \| \alpha \|_2^2;\ \alpha \in \mathscr{F} \}$. The most interesting case is arguably the case where $\psi = \psi_0$ is given by

$$\psi_0(x) = x^2 \quad \text{if } x \leq 1; \qquad \psi_0(x) = 2x - 1 \quad \text{if } x \geq 1.$$

If we set

$$\tau = \sup\{ \mid \alpha_i \mid;\ i \leq N;\ \alpha \in \mathscr{F} \},$$

we note that, for given $\alpha \in \mathscr{F}$, for each $s = (s_i)_{i \leq N}$, setting $J = \{ i \leq N;\ s_i \leq 1 \}$, we have

$$\begin{aligned} \sum_{i \leq N} s_i \mid \alpha_i \mid &= \sum_{i \in J} s_i \mid \alpha_i \mid + \sum_{i \notin J} s_i \mid \alpha_i \mid \\ &\leq \sigma\Big(\sum_{i \in J} s_i^2\Big)^{1/2} + \tau \sum_{i \notin J} s_i \\ &\leq \sigma\Big(\sum_{i \leq N} \psi(s_i)\Big)^{1/2} + \tau \sum_{i \leq N} \psi(s_i). \end{aligned}$$

Thus, if $\sum_{i \leq N} s_i \mid \alpha_i \mid \geq u$, then either $\sum_{i \leq N} \psi(s_i) \geq u^2/4\sigma^2$, or else $\sum_{i \leq N} \psi(s_i) \geq \tau/2$, and thus

(8.2.1)
$$\Psi_{\mathscr{F}}(u) \geq \min\left(\frac{u^2}{4\sigma^2}, \frac{u}{2\tau}\right).$$

The basic observation is as follows.

Proposition **8.2.1.** — *Consider* $\mathcal{F}$, ψ *as above. Set* $\Omega = \mathbf{R}$, *and consider the function* $Z(x) = \sup_{\alpha \in \mathcal{F}} \sum_{i \leq N} \alpha_i x_i$. *Consider* $a \in \mathbf{R}$, *and* $A(a) = \{y; Z(y) \leq a\}$. *Then*

$$\forall x \in \Omega^N, \quad f_{h,\psi}(A(a), x) \geq \Psi_{\mathcal{F}}(Z(x) - a) \tag{8.2.2}$$

when the function h *is defined on* $\mathbf{R} \times \mathbf{R}$ *by*

$$h(\omega, \omega') = |\omega - \omega'|. \tag{8.2.3}$$

Moreover, when $\alpha_i \leq 0$ *for each* $i \leq N$ *and each* $\alpha \in \mathcal{F}$, *we can take*

$$h(\omega, \omega') = (\omega' - \omega)^+. \tag{8.2.4}$$

Proof. — By definition of $f_{h,\psi}$, given $\varepsilon > 0$, we can find $s \in V_{A(a)}(x)$ such that

$$\sum_{i \leq N} \psi(s_i) \leq f_{h,\psi}(A(a), x) + \varepsilon.$$

Consider $\alpha = (\alpha_i) \in \mathcal{F}$. Then there exists $s' \in U_{A(a)}(x)$ such that $\sum_{i \leq N} |\alpha_i| s'_i \leq \sum_{i \leq N} |\alpha_i| s_i$. This means that there is $y \in A(a)$ for which $\sum_{i \in I} |\alpha_i| h(x_i, y_i) \leq \sum_{i \leq N} |\alpha_i| s_i$, where $I = \{i \leq N; x_i \neq y_i\}$. Now

$$\sum_{i \leq N} \alpha_i x_i = \sum_{i \leq N} \alpha_i y_i + \sum_{i \in I} \alpha_i (x_i - y_i). \tag{8.2.5}$$

We have $\alpha_i(x_i - y_i) \leq |\alpha_i| \, |x_i - y_i|$. If α_i is ≤ 0 we have $\alpha_i(x_i - y_i) \leq |\alpha_i| (y_i - x_i)^+$. Thus in all cases under consideration, we have

$$\sum_{i \in I} \alpha_i (x_i - y_i) \leq \sum_{i \in I} |\alpha_i| h(x_i, y_i) \leq \sum_{i \leq N} |\alpha_i| s_i$$

so that, by (8.2.5),

$$\sum_{i \leq N} \alpha_i x_i \leq \sum_{i \leq N} \alpha_i y_i + \sum_{i \leq N} |\alpha_i| s_i$$

$$\leq a + \sum_{i \leq N} |\alpha_i| s_i.$$

Taking the sup over α yields

$$\sup_{\alpha \in \mathcal{F}} \sum_{i \leq N} |\alpha_i| s_i \geq Z(x) - a$$

and the result follows by definition of $\Psi_{\mathcal{F}}$. □

Corollary **8.2.2.** — *Consider a family* $\mathcal{F}$ *of* N-*tuples* $\alpha = (\alpha_i)_{i \leq N}$. *Consider a sequence of independent r.v.* $(X_i)_{i \leq N}$ *with common law* μ. *Assume that* (4.4.6) *holds (for a certain* *function* θ) *when* $P = \mu^{\otimes N}$, *and where* h *is the function determined in Proposition* 8.2.1. *Then the r.v.* $Z = \sup_{\alpha \in \mathcal{F}} \sum_{i \leqslant N} \alpha_i X_i$ *satisfies*

$$\textbf{(8.2.6)} \qquad u \geqslant 0 \Rightarrow P(Z \geqslant M + u) \leqslant \exp\left(\theta\left(\frac{1}{2}\right) - \frac{1}{K}\Psi_{\mathcal{F}}(u)\right)$$

$$\textbf{(8.2.7)} \qquad u \geqslant 0 \Rightarrow P(Z \leqslant M - u) \leqslant \xi\left(\frac{1}{K}\Psi_{\mathcal{F}}(u) - \log 2\right)$$

where M *is a median of* Z.

Proof. — Using (4.4.6) and Chebyshev inequality, we have

$$P(f_{h,\psi}(A(a), x) \geqslant t) \leqslant \exp\left(\theta(P(A(x))) - \frac{t}{K}\right)$$

where $A(a)$ is the set of Proposition 8.2.1; thus, by (8.2.2), setting $t = \Psi_{\mathcal{F}}(b - a)$, for $Z(x) \geqslant b$ we have $f_{h,\psi}(A(a), x) \geqslant t$, so that

$$P(Z \geqslant b) \leqslant \exp\left(\theta(P(Z \leqslant a)) - \frac{1}{K}\Psi_{\mathcal{F}}(b - a)\right).$$

Taking $a = M$, $b = M + u$ imply (8.2.6). Taking $b = M$, $a = M - u$ imply

$$\frac{1}{2} \leqslant \exp\left(\theta(P(Z \leqslant M - u)) - \frac{1}{K}\Psi_{\mathcal{F}}(u)\right)$$

from which (8.2.7) follows. □

We now go back to our main line of study, that of the r.v. $Z' = \sup_{\alpha \in \mathcal{F}} \sum_{i \leqslant N} (-\alpha_i) X_i$. In order to apply Corollary 8.2.2, we need (4.4.6) for the penalty function $h(x, y) = (y - x)^+$. Since X_i is positive, its law μ is supported by $\mathbf{R}^+$. Thereby, only the properties of h on $\mathbf{R}^+ \times \mathbf{R}^+$ matter; but then $(y - x)^+ \leqslant y$. Thus, to have (4.4.6) it suffices that the function $h(x, y) = y$ satisfies the conditions of Theorem 4.4.1. The case where the function $h(x, y)$ depends on y only has been discussed after Theorem 4.4.1. Thus, we have proved the following.

Theorem **8.2.3.** — *Consider a family* $\mathcal{F}$ *on* N-*tuples of positive numbers, and independent identically distributed nonnegative r.v. variables* $(X_i)_{i \leqslant N}$. *Consider functions* θ, ξ, w *as in Theorem* 4.4.1. *Assume that* (2.6.1), (4.4.2), (4.4.3) *hold, that condition* $H(\xi, w)$ *holds, that the median* m *of* X_1 *is* $\leqslant 1$, *and that for* $t \geqslant m$, *we have*

$$\textbf{(8.2.8)} \qquad w(P(X_1 \geqslant t)) \geqslant \psi(t).$$

Then if M *is a median of* $Z' = \inf_{\mathcal{F}} \sum_{i \leqslant N} \alpha_i X_i$, *the following holds (where the constant* K *depends only on the parameter* γ *of Theorem* 4.4.1*)*:

(8.2.9) $$u \geqslant 0 \Rightarrow P(Z' \leqslant M - u) \leqslant \exp\left(\theta\left(\frac{1}{2}\right) - \frac{1}{K} \Psi_{\mathcal{F}}(u)\right)$$

(8.2.10) $$u \geqslant 0 \Rightarrow P(Z' \geqslant M + u) \leqslant \xi\left(\frac{1}{K} \Psi_{\mathcal{F}}(u) - \log 2\right).$$

Comment. — A striking feature of this result is the different forms of (8.2.9) and (8.2.10). This phenomenon is well-known in the case where $\mathcal{F}$ consists of a single point α. In that case, $\mathcal{F}$ is a sum of positive independent r.v. Y_i. The lower tails of Z have a tendency to be " subgaussian " ([H]) while the upper tails of Z certainly depend much on the upper tails of the variables Y_i.

Corollary **8.2.4.** — *There exists a universal constant* K *with the following property. Assume that* ψ *satisfies* (4.4.1). *Assume that*

(8.2.11) $$\forall\, t \geqslant 1, \quad P(X_1 \geqslant t) \leqslant \exp(-2\psi(t)).$$

Then we have

(8.2.12) $$u \geqslant 0 \Rightarrow P(|Z - M| \geqslant u) \leqslant 3 \exp\left(-\frac{1}{K} \Psi_{\mathcal{F}}(u)\right).$$

Proof. — We take $\xi(x) = e^{-x}$, $\theta(x) = -\log x$. According to Proposition 2.6.1, condition $H(\xi, w)$ holds if $\int e^{w}\, d\lambda \leqslant 2$, so, in particular, if $w(t) = -\frac{1}{2}\log t$. Also, by (4.4.1), $\psi(1) = 1$, so that (8.2.11) implies that the median of X_1 is $\leqslant 1$. Thus Corollary 8.2.4 follows from Theorem 8.2.3.

Corollary **8.2.5.** — *Assume that* (2.6.5) *holds for a certain number* L. *Then, for some constant* K *depending on* ξ *only, if for all* $t \geqslant 1$ *we have*

(8.2.13) $$P(X_1 \geqslant t) \leqslant \frac{1}{K} |\xi'(\psi(t))|$$

then (8.2.9), (8.2.10) *hold (for a constant* K *depending on* ξ *only).*

Proof. — We simply have to find a function w that satisfies (8.2.8) and such that condition $H(\xi, w)$ holds. It follows from Proposition 2.6.3 that if we take R large enough $\left(\text{R can actually be taken depending on L and} \int_0^\infty \xi\, d\lambda \text{ only}\right)$ then the function w such that

$$\forall\, b \geqslant c \quad |\{w \geqslant b\}| = \frac{1}{R} |\xi'(b)|$$

satisfies condition $H(\xi, w)$. Now, if we take $K = R$ in (8.2.13), then

$$|\{ w \geqslant \psi(t)\}| = \frac{1}{R} |\xi'(\psi(t))| \geqslant P(X_1 \leqslant t),$$

so that $w(P(X_1 \leqslant t)) \geqslant \psi(t)$ since w is non-decreasing. □

We now explain why Corollary 8.2.5 is sharp. Consider the case where $\mathscr{F}$ consists of the single element $\alpha = (\alpha_i)$, where $\alpha_i = 1/\sqrt{N}$. Consider ψ such that $\psi(x) = x^2$ if $x \leqslant 1$ and $\psi(x) = 2x - 1$ for $x \geqslant 1$. Then, for $u = \sqrt{N}$, $\Psi_{\mathscr{F}}(u) \geqslant N/4$ by (8.2.1). Consider a r.v. $X_i \in \{0, N\}$, with

$$P(X_i = N) = p =: \frac{1}{K} |\xi'(2N - 1)|.$$

Under condition (2.6.5), we have $\lim_{x\to\infty} x\xi'(x) = 0$, and it is not a restriction to assume $Np \leqslant 1/2$. Thus the median of $Z = N^{-1/2} \sum_{i\leqslant N} X_i$ is zero.

Now

$$P(Z \geqslant u) = P\left(\frac{1}{\sqrt{N}} \sum_{i\leqslant N} X_i \geqslant u\right) \sim Np = \frac{N}{R} \xi'(2N - 1)$$

and the bound $\xi\left(\frac{N}{K}\right)$ of (8.2.10) is indeed reasonably good, as $x\xi'(x)$ is of order $\xi(x)$ for many choices of ξ.

8.3. *First time passage in percolation*

Consider a graph (V, E) where V is the set of vertices, E the set of edges. Assume that we have a family $(X_e)_{e\in E}$ of positive r.v. distributed like a given r.v. X (X_e represents the passage time through the edge e). Consider a family $\mathscr{S}$ of sets of edges, and for $S \in \mathscr{S}$, consider $X_S = \sum_{e\in S} X_e$. In the case where S is a path, i.e., consists of the edges $e_{v_1 v_2}$, $e_{v_2 v_3}$, ..., e_{v_{k-1}, v_k} linking vertices $v_1, \ldots, v_k$, X_S represents the "passage time through S". Let us set $Z'_{\mathscr{S}} = \inf_{S\in\mathscr{S}} X_S$ and $r = \sup_{S\in\mathscr{S}} \operatorname{card} S$. Denote M a median of $Z_{\mathscr{S}}$. The following is a consequence of (8.2.1) and Corollary 8.2.4.

Proposition **8.3.** — *There exists a universal constant* K *such that if* $E \exp KX \leqslant 2$, *we have*

$$\textbf{(8.3.1)} \qquad \forall u > 0, \quad P(|Z_{\mathscr{S}} - M| \geqslant u) \leqslant 4 \exp\left(-\frac{1}{K} \min\left(\frac{u^2}{r}, u\right)\right).$$

Consider the case where $V = \mathbf{Z}^2$, E consists of the edges that link any two adjacent vertices. Denote by $\mathscr{S}$ the sets of self-avoiding paths linking the origin to the point $(0, n)$, and by $\mathscr{S}(C)$ the subset of $\mathscr{S}$ consisting of paths of length $\leqslant Cn$. It was proved by H. Kesten [K1] that if $P(X = 0) < \frac{1}{2}$, then, for some constant C independent of n, we have

$$P(Z'_{\mathscr{S}} = Z'_{\mathscr{S}(C)}) \geqslant 1 - Ce^{-n/C}.$$

It then follows from (8.3.1) that for some constant C' independent of n, we have

(8.3.2) $$u \leqslant \frac{n}{C'} \Rightarrow P(|Z'_{\mathscr{S}} - M| \geqslant u) \leqslant 5 \exp\left(-\frac{u^2}{C'n}\right).$$

This improves recent results of H. Kesten [K2], based on the use of martingales, who proves (8.3.2) with an exponent $u/C'\sqrt{n}$. It should, however, be pointed out that the reason why martingales allow some success on this problem is because we consider only sums of the type $\Sigma\, \alpha_e X_e$ for very special families $\alpha = (\alpha_e)$. Martingales are apparently powerless to approach Corollary 8.2.5.

It is pointed out in the literature that (in the case $V = \mathbf{Z}^2$) (8.3.2) apparently does not give the correct rate. In view of Corollary 8.2.5, the obvious approach to improve (8.3.2) would be to show that $Z_{\mathscr{S}}$ is very close to $Z_{\mathscr{F}}$, where the family $\mathscr{F}$ of sequences $(\alpha_e)_{e\in E}$ satisfies $\sigma = \sup_{\alpha\in\mathscr{F}} \|\alpha\|_2 \ll n$. There is an obvious candidate for $\mathscr{F}$. Indeed, consider the family $\mathscr{F}'$ defined as follows: $\mathscr{F}'$, seen as a subset of $(\mathbf{R}^+)^E$, is the convex hull of the family of points a_S given by $a_S(e) = 1$ if $e \in S$ and $a_S(e) = 0$ if $e \notin S$, for all $S \in \mathscr{S}$. Then, obviously, $Z'_{\mathscr{S}} = Z'_{\mathscr{F}'}$. Then consider the family $\mathscr{F}(\sigma)$ of sequences $(\alpha_e)_{e\in V}$ of $\mathscr{F}'$ for which $\|\alpha\|_2 \leqslant \sigma$. Then $Z'_{\mathscr{S}} \leqslant Z'_{\mathscr{F}(\sigma)}$. Thus if one could show that, for some $\sigma = o(n)$, one still has $Z'_{\mathscr{F}(\sigma)} \leqslant Z'_{\mathscr{S}} + o(\sqrt{n})$, with probability $1 - o(n^{-1})$, one would obtain that the likely fluctuations of $Z'_{\mathscr{S}}$ from M are $o(\sqrt{n})$. Roughly speaking, this means that the shortest passage time from (0, 0) to (0, n) is (within $o(\sqrt{n})$) obtained through a number of rather disjoint paths. Proving such a statement is apparently a long range program in Percolation theory.

9. Chromatic Number of Random Graphs

The use of martingales has allowed several important progresses in the understanding of the chromatic number of random graphs. Use of martingales does require ingenuity. This chapter will demonstrate that Theorem 4.1.1 achieves somewhat better results than martingales in a completely straightforward manner.

For simplicity we call a graph G with vertice set $V = \{1, \ldots, n\}$ a subset of $E_0 = \{(i,j) ; i < j\}$. If (i,j) belongs to G, we say that i, j are linked by an edge.

A subset I of V is called independent if no two points of I are linked by an edge (the word independent here should not be confused with its probabilistic meaning). The chromatic number $\chi(G, A)$ of a subset A of V is the smallest number of independent sets that can cover A; that is, the vertices of A can be given $\chi(G, A)$ colors so that no two points with the same color are linked by an edge. We set

$$\chi(G, m) = \inf\{\chi(G, A) ; \text{card } A = m\}.$$

Given p, $0 < p < 1$, the random graph $G = G(n, p)$ is defined by putting each possible edge (i, j) in G with probability p, independently of what is done for the other edges.

The chromatic number is remarkably concentrated, as the following shows.

Theorem **9.1.** — *Consider $k \in \mathbf{N}$ and $t > 0$. Then there exists an integer a such that*

$$\textbf{(9.1)} \qquad P(\chi(G(n, p), m) \in [a - k, a])$$
$$\geq 1 - 2e^{-t^2/8} - P(\sup\{\chi(G(n, p), F) ; F \subset V, \text{card } F \leq t\sqrt{m}\} > k).$$

Comments. — 1) The last term is always zero for $k > t\sqrt{m}$. But when $p = n^{-\alpha}$ $(\alpha > 0)$, it is still small for smaller values of k. See [S-S], [A-S, p. 88].

2) Another version of this Theorem could be proved, in the spirit of Theorem 7.1.3, concerning the concentration property of the number

$$\max\{\text{card } F ; \chi(G(n, p), F) \leq m\}.$$

3) With a bit of care, we can replace m by $m - 1$ in the right-hand side of (9.1), and improve the coefficient 1/8.

Proof. — We set

$$b = P(\sup\{\chi(G(n, p), F) ; \text{card } F \leq t\sqrt{m}\} > k).$$

We then define a as the largest integer for which

$$\textbf{(9.2)} \qquad P(\chi(G(n, p), m) \geq a) \geq e^{-t^2/8} + b.$$

Thus

$$P(\chi(G(n, p), m) > a) < e^{-t^2/8} + b.$$

In order to apply Theorem 4.1.1, we must represent the underlying probability space as a product space. The first idea that comes to mind would be to use $\{0, 1\}^{E_0}$; this is not a good choice. For $2 \leq j \leq n$, set $\Omega_j = \{0, 1\}^{j-1}$. Set $\Omega' = \prod_{2 \leq j \leq n} \Omega_j$. We write $\omega \in \Omega'$ as $(\omega_j)_{j \leq n}$, where $\omega_j = (\omega_{i,j})_{i \leq j-1} \in \Omega_j$. To ω we associate the graph $G(\omega)$ such that, for $i < j$, $(i, j) \in G(\omega)$ if and only if $\omega_{i,j} = 1$. The only property of $G(n, p)$ we need is that it is distributed as $G(\omega)$ for a certain product measure P on $\prod_{j \leq N} \Omega_j$.

Define $A \subset \Omega'$ as the set of ω for which

$$\chi(G(\omega), m) \geq a ; \quad \sup\{\chi(G(\omega), F) : \text{card } F \leq t\sqrt{m}\} \leq k.$$

Thus by (9.2) we have $P(A) \geqslant e^{-t^2/8}$. Combining Theorem 4.1.1 and Lemma 4.1.2, we see that $P(B) \geqslant 1 - e^{-t^2/8}$, where we have set

$$\text{(9.3)} \qquad B = \left\{ \omega ;\ \forall\ (\alpha_j)_{2 \leqslant j \leqslant n},\ \exists\ \omega' \in A;\ \Sigma \alpha_j 1_{\{\omega_j \neq \omega'_j\}} \leqslant t \sqrt{\sum_j \alpha_j^2} \right\}.$$

To finish the proof, it suffices to show that

$$\omega \in B \Rightarrow \chi(G(\omega), m) \geqslant a - k.$$

So, consider $\omega \in B$, and set $r = \chi(G(\omega), m)$. Consider a subset F of V, of cardinal m, such that $\chi(G(\omega), F) = r$. We use (9.3) with $\alpha_j = 1$ if $j \in F$ and zero otherwise. Thus there is $\omega \in A$ such that, if $J = \{ j \in F;\ \omega_j \neq \omega'_j \}$, then card $J \leqslant t\sqrt{m}$. But obviously,

$$\chi(G(\omega'), F \backslash J) = \chi(G(\omega), F \backslash I) \leqslant r$$

and thus

$$\begin{aligned} a \leqslant \chi(G(\omega'), F) &\leqslant r + \chi(G(\omega'), I) \\ &\leqslant r + k. \quad \square \end{aligned}$$

In order to obtain an upper bound for χ_G, the most obvious approach is the " greedy " one: one chooses an independent set W_1 of maximal size, and removes its vertices and all edges adjacent. One is then left with a graph on fewer vertices, and one iterates the process until exhaustion. To make this approach work one needs a competent bound on the probability that a random graph contains at least one independent set of size r. Such bounds were first obtained by B. Bollobas [B], using martingales. A recent powerful correlation inequality of Janson [J] is both simpler and more powerful than the martingale approach (compare [A-S] p. 87 and p. 148). It is of some interest to note that Theorem 4.1.1 does as well as Janson's inequality. We fix an integer r. For $e = (i, j) \in E_0$, we denote by $N(G, e)$ the number of independent sets of size r that contain i, j.

Proposition **9.2.** — *Consider a number u, and assume that*

$$\text{(9.4)} \qquad P\left(u \sqrt{\sum_{e \in E_0} N(G(n,p), e)^2} \leqslant \sum_{e \in E_0} N(G(n,p), e) \right) > \frac{1}{2}.$$

Then

$$P(G(n,p) \text{ contains no independent set of size } r) \leqslant 2 \exp\left(- \frac{u^2}{r^2(r-1)^2} \right).$$

Proof. — We set $\Omega = \{0, 1\}$, provided with the probability that gives weight p to 1 (and $1 - p$ to 0). Consider the product probability P on Ω^{E_0}. For $x \in (x_e)_{e \in E_0}$

we define $G(x)$ by $e(i,j) \in G(x)$ if and only if $x_e = 1$. The graph $G(x)$ is distributed like $G(n,p)$.

Consider the set $A \subset \Omega^{E_0}$, given by

$$A = \{y;\ G(y) \text{ contains no independent set of size } r\}.$$

Consider $t_0 = 2\sqrt{\log(2/P(A))}$. If we combine (9.4), Theorem 4.1.1 and Lemma 4.1.2, we see that there exists x such that

$$u\sqrt{\sum_{e\in E_0} N(G(x),e)^2} \leqslant \sum_{e\in E_0} N(G(x),e) \tag{9.5}$$

with the property that

$$\forall\ (\alpha_e)_{e\in E_0},\ \exists y \in A,\ \sum_{x_e \neq y_e} \alpha_e \leqslant t_0\sqrt{\sum_{e\in E_0} \alpha_e^2}.$$

In particular, there exists $y \in A$, such that if

$$C = \{e \in E_0,\ x_e \neq y_e\}$$

we have

$$\begin{aligned} \sum_{e\in C} N(G(x),e) &\leqslant t_0\sqrt{\sum_{e\in E_0} N(G(x),e)^2} \\ &\leqslant \frac{t_0}{u}\sum_{e\in E_0} N(G(x),e) \end{aligned} \tag{9.6}$$

where the last inequality follows from (9.5). The total number N of independent sets of $G(x)$ of size r is

$$N = \binom{r(r-1)}{2}^{-1} \sum_{e\in E_0} N(G(x),e). \tag{9.7}$$

We must have

$$N \leqslant \sum_{e\in C} N(G(x),e)$$

for otherwise there would be an independent set of size r of $G(x)$ that would contain no edge of C, and thus would be an independent set of $G(y)$, which is impossible. Combining with (9.6), (9.7), we get $t_0 \geqslant \frac{2u}{r(r-1)}$, so that

$$P(A) \leqslant 2\exp -\frac{u^2}{r^2(r-1)^2}. \quad \square$$

In order to take advantage of Proposition 9.2, one must find competent (= large) values of u for which (9.4) holds. For example, one can take $u = u_1/u_2$, where

$$\text{(9.8)} \qquad P\Big(\sum_{e\in E_0} N((G,p),e)^2 \leqslant u_2^2\Big) > \frac{3}{4},$$

$$\text{(9.9)} \qquad P\Big(\sum_{e\in E_0} N((G,p),e) \geqslant u_1\Big) > \frac{3}{4}.$$

We then find values of u_2 (resp. u_1) using Chebyshev inequality (resp. the second moment method). Not surprisingly that leads to unpleasant computations (as seems unavoidable in this topic). These are better not reproduced here, and left to the specialist that wants to evaluate the strength of Proposition 9.2.

10. The Assignment Problem

Consider a number N, and two disjoint sets I, J of cardinal N. An assignment is a one-to-one map τ from I to J. Consider a matrix $a = (a_{i,j})_{i\in I, j\in J}$, such that $a_{i,j}$ represents the cost of assigning j to i. The cost of the assignment τ is $\sum_{i\in I} a_{i,\tau(i)}$ and the problem is to find the assignment of minimal cost.

Assume now that the costs $a_{i,j}$ are taken equal to $X_{i,j}$, where the r.v. $(X_{i,j})_{i\in I, j\in J}$ are independent uniformly distributed over [0, 1]. Consider the r.v.

$$L_N = \inf\{\sum_{i\in I} X_{i,\tau(i)};\ \tau \text{ assignment}\}.$$

It is a remarkable fact [W] that $E(L_N)$ is bounded independently of L_N. (Actually $E(L_N) \leqslant 2$ [Ka].)

In this section we try to bound the fluctuations of L_N; the challenge is that the average value of L_N is of the same order as the average value of the costs $X_{i,j}$, and that N^2 of these costs are involved.

We will first show that we can replace the costs $X_{i,j}$ by $Y_{i,j} = \min(X_{i,j}, v)$ for v of order $N^{-1}(\log N)^2$; then we will appeal to Theorem 4.1.1.

A *digraph* D will be a subset of I × J. (If $(i,j) \in D$, we think of i, j as being linked by an edge.) The digraphs of use will mostly consist of those couples (i,j) for which $X_{i,j}$ is small. Consider a digraph D, and S ⊂ I. We set

$$D(S) = \{j \in J;\ \exists\, i \in S,\ (i,j) \in D\}.$$

We will say that a digraph D is α-expanding ($\alpha \geqslant 2$) if the following occurs, for all subsets S of I:

$$\text{(10.1)} \qquad \operatorname{card} S \leqslant \frac{N}{2} \Rightarrow \operatorname{card} D(S) \geqslant \min\left(\alpha \operatorname{card} S, \frac{N}{2}\right),$$

$$\text{(10.2)} \qquad \operatorname{card} S \geqslant \frac{N}{2} \Rightarrow \operatorname{card} D(S) \geqslant N - \frac{1}{\alpha}(N - \operatorname{card} S).$$

Our first lemma mimics an argument of Steele and Karp [S-K].

Lemma **10.1.** — *Consider an α-expanding digraph* D *and an integer m such that $\alpha^m \geqslant N/2$. Consider a one-to-one map τ from* I *to* J. *Then, given any $i \in I$, we can find $n \leqslant 2m$ and disjoint points $i_1 = i, i_2, \ldots, i_{n+1} = i$ such that for $1 \leqslant \ell \leqslant n$, we have $(i, \tau(i_{\ell+1})) \in D$.*

Proof. — We fix $i \in I$. Consider the set S_p of points of $i_p \in I$ with the property that we can find $i_2, \ldots, i_p$ in I, for which $(i_\ell, \tau(i_{\ell+1})) \in D$ for $1 \leqslant \ell < p$. We observe that, obviously, $S_{p+1} \supset \tau^{-1}(D(S_p))$. Since we can assume without loss of generality that $\alpha^{m-1} \leqslant N/2$, we see from (10.1) and by induction that for $p \leqslant m$, we have card $S_p \geqslant \alpha^{p-1}$. Then (10.1) shows that card $S_{m+1} \geqslant N/2$, and (10.2) shows that for $p \geqslant 1$, $N - \text{card}\, S_{m+p+1} \leqslant \alpha^{-p} N/2$. Thus $N - \text{card}\, S_{2m+1} \leqslant \alpha^{-m} N/2 < 1$, which means $S_{2m+1} = I$. Thus $i \in S_{2m+1}$. Consider then the smallest n for which $i \in S_{n+1}$; thus $n \leqslant 2m$. Then one can find $i_1 = i, i_2, i_3, \ldots, i_{n+1} = i$ such that, for $1 \leqslant \ell \leqslant n$, we have $(i_\ell, \tau(i_{\ell+1})) \in D$. The minimality of n implies that the points i_ℓ are all disjoint. □

Consider $u > 0$ and consider the digraph D_u given by

$$(i, j) \in D_u \Leftrightarrow X_{i,j} \leqslant 2uN^{-1} \log N.$$

Corollary **10.2.** — *Assume that the digraph D_u is α-expanding, and consider an integer m such that $\alpha^m \geqslant N/2$. Then for an optimal assignment τ we have $X_{i,\tau(i)} \leqslant 4muN^{-1} \log N$ for all $i \leqslant N$.*

Proof. — Consider any $i \in I$, and consider $i = i_1, \ldots, i_{n+1} = i$ as in Lemma 10.1, used for $D = D_u$. Define $\sigma(i_\ell) = \tau(i_{\ell+1})$ for $1 \leqslant \ell \leqslant n$, and $\sigma(i') = \tau(i')$ if $i' \notin \{i_1, \ldots, i_n\}$. Since τ is optimal, we have

$$\sum_{i' \leqslant N} X_{i', \tau(i')} \leqslant \sum_{i' \leqslant N} X_{i', \sigma(i')},$$

so that

$$X_{i, \tau(i)} \leqslant \sum_{1 \leqslant \ell \leqslant n} X_{i_\ell, \sigma(i_\ell)} \leqslant 2nuN^{-1} \log N. \quad \square$$

It remains to do computations.

Proposition **10.3.** — *For some constant* K *and all $u > K$ with $u \log N \leqslant N$, the random digraph D_u is $u \log N$-expanding with probability $\geqslant 1 - N^{-u/K}$.*

Proof. — We explain why (10.1) is satisfied with probability $\geqslant 1 - N^{-u/K}$. The case of (10.2) is similar and is left to the reader. For simplicity, we set $\theta = uN^{-1} \log N$.

Consider a subset S of $\{1, \ldots, N\}$, and set $s = \text{card}\, S$. For $j \in J$, we have

$$P(j \notin D(S)) = \left(1 - \frac{2u \log N}{N}\right)^s = (1 - 2\theta)^s \leqslant \exp(-2\theta s)$$

and thus

$$P(j \in D(S)) \geqslant 1 - \exp(-2\theta s).$$

We observe that

$$0 \leq x \leq 1 \Rightarrow 1 - e^{-x} \geq (1 - e^{-1})\, x.$$

Thus, if we assume

$$\text{(10.3)} \qquad s\theta \leq \frac{1}{2}$$

we have

$$\text{(10.4)} \qquad \mathrm{P}(j \in \mathrm{D(S)}) \geq \gamma s\theta$$

where we have set $\gamma = 2(1 - e^{-1}) > 1$.

Consider $\gamma' = (1 + \gamma)/2$. We claim that, under (10.3),

$$\text{(10.5)} \qquad \mathrm{P}(\operatorname{card} \mathrm{D(S)} < \gamma' s\theta \mathrm{N}) \leq \exp\left(-\frac{s\theta\mathrm{N}}{\mathrm{K}}\right).$$

This follows from (10.4) and the following general fact:

Lemma **10.4.** — *Consider independent events* $(\mathrm{A}_i)_{i \leq \mathrm{N}}$ *with* $\mathrm{P}(\mathrm{A}_i) = p$, *and consider* $\delta < 1$. *Then, the probability that less than* $\delta p\mathrm{N}$ *events occur is at most* $\exp(-\mathrm{N}p/\mathrm{K}(\delta))$, *where* $\mathrm{K}(\delta)$ *depends on* δ *only.*

Proof. — Set $\mathrm{Y}_i = 1_{\mathrm{A}_i}$, so that

$$\mathrm{E} \exp(-\lambda \mathrm{Y}_i) = 1 - p(1 - e^{-\lambda}) \leq \exp(-p(1 - e^{-\lambda})).$$

Thus

$$\mathrm{E} \exp\left(-\lambda \sum_{i \leq \mathrm{N}} \mathrm{Y}_i\right) \leq \exp(-\mathrm{N}p(1 - e^{-\lambda})).$$

By Chebyshev inequality we get

$$\mathrm{P}\left(\sum_{i \leq \mathrm{N}} \mathrm{Y}_i \leq \delta p \mathrm{N}\right) \leq \exp \mathrm{N}p(\lambda\,\delta - (1 - e^{-\lambda}))$$

and the result follows by taking λ small enough so that $\lambda\,\delta - (1 - e^{-\lambda}) < 0$. □

The number of subsets S of I of cardinal s is at most N^s. For $u \geq \mathrm{K}$, we have

$$\mathrm{N}^s \exp\left(-\frac{s\theta\mathrm{N}}{\mathrm{K}}\right) \leq \exp\left(-\frac{s\theta\mathrm{N}}{\mathrm{K}}\right)$$

and

$$\sum_{s \geq 1} \exp\left(-\frac{s\theta\mathrm{N}}{\mathrm{K}}\right) \leq \mathrm{N}^{-u/\mathrm{K}}.$$

Thus, it follows that with probability $\geqslant 1 - N^{-u/K}$, for all subsets S of I such that $s = \text{card } S$ satisfies $\theta s \leqslant 1/2$, we have $\text{card } D(S) \geqslant \gamma' \theta N s$. Equivalently, we have

$$u \log N \text{ card } S \leqslant \frac{N}{2} \Rightarrow \text{card } D(S) \geqslant \gamma' u \log N \text{ card } S. \tag{10.6}$$

To complete the proof that (10.1) holds for $\alpha = u \log N$, it suffices to show that $\text{card } D(S) \geqslant N/2$ whenever $\alpha \text{ card } S \geqslant N/2$. This follows by applying (10.7) to a subset S' of S for which card S' satisfies $\alpha \text{ card } S' \leqslant N/2$ and is as large as possible. □

We can now prove the main result.

Theorem **10.5.** — *Denote by* M *a median of* L_N. *Then (for* $N \geqslant 3$*),*

$$t \leqslant \sqrt{\log N} \Rightarrow P\left(|L_N - M| \geqslant \frac{Kt(\log N)^2}{\sqrt{N} \log \log N}\right) \leqslant 2 \exp(-t^2), \tag{10.7}$$

$$t \geqslant \sqrt{\log N} \Rightarrow P\left(|L_N - M| \geqslant \frac{Kt^3 \log N}{\sqrt{N} \log t^2}\right) \leqslant 2 \exp(-t^2). \tag{10.8}$$

Proof. — *Step 1.* — Consider $u \leqslant N/(2 \log N)$, $\alpha = u \log N$ and the smallest m such that $\alpha^m \geqslant N/2$. Set $v = 4muN^{-1} \log N$, and $Y_{i,j} = \min(X_{i,j}, v)$. Consider the r.v. L_N^u defined as L_N but using the costs $Y_{i,j}$ rather that $X_{i,j}$. It follows from Corollary 10.2 that $L_N^u = L_N$ whenever D_u is α-expanding, so that by Proposition 10.3

$$P(L_N = L_N^u) \geqslant 1 - N^{-u/K}. \tag{10.9}$$

Step 2. — When $N^{-u/K} \leqslant 1/2$, it follows from (10.9) that M is also a median of L_N^u. It then follows from (8.1.1) (and scaling) that, for all $w > 0$,

$$P(|L_N^u - M| \geqslant w) \leqslant 2 \exp\left(-\frac{w^2}{4Nv^2}\right)$$

and, combining with (10.9), we get

$$P(|L_N - M| \geqslant w) \leqslant 2 \exp\left(-\frac{w^2}{4Nv^2}\right) + N^{-u/K}.$$

Step 3. — We choose the parameters. We take $w = 3\sqrt{N} t v$. If $t^2 \leqslant \log N$, we take $u = K$; if $t^2 \geqslant \log N$, we take $u = Kt^2/\log N$.

Theorem 10.6 follows easily. □

Remark. — A simple computation using Theorem 10.6 shows that the standard deviation of L_N is not more than $K(\log N)^2/\sqrt{N} \log \log N$.

11. Geometric Probability

11.1. *Irregularities of the Poisson Point Process*

In this Chapter we will consider N points $X_1, \ldots, X_N$ that are independent uniformly distributed in $[0, 1]^d$, where, except in Section 13.5, $d = 2$, and we will study certain functionals $L(X_1, \ldots, X_N)$ of this configuration $X_1, \ldots, X_N$ (that is L will depend only on $\{X_1, \ldots, X_N\}$ rather than on the order in which the points are taken).

One would like to think that the sample $X_1, \ldots, X_N$ is rather uniform on $[0, 1]^2$; say, that it meets every subsquare of side $K/\sqrt{N}$. This is not the case; there are empty squares of side of order $(N^{-1} \log N)^{1/2}$ (an empty square will informally be called a hole). More importantly, in exceptional situations there are larger empty squares. Several of the functionals we will study have the property that, if one deletes or adds a point to a finite set F, the amount by which L(F) can vary depends on whether F has a " large " hole close to x. Thereby the first task is to study the size and number of holes.

It is not convenient to work with the sample $X_1, \ldots, X_N$. The difficulty is that what happens, say, in the left half of $[0, 1]^2$ (for example, if there is an excess of points here), affects what happens in the right half (there must then be a deficit of points there). Rather, one will work with a Poisson point process of constant intensity μ. This process generates a random subset Π $(= \Pi_\mu)$ of $[0, 1]^2$ with the following properties:

(11.1.1) If A and B are disjoint (Borel) subsets of $[0, 1]^2$, $\Pi \cap A$ and $\Pi \cap B$ are independent.

(11.1.2) If A is a (Borel) subset of $[0, 1]^2$, the r.v. $\operatorname{card}(\Pi \cap A)$ is Poisson of parameter $\mu \mid A \mid$, where $\mid A \mid$ denotes the area of A.

Let us recall that a r.v. Y is Poisson of parameter λ if $P(Y = k) = e^{-\lambda} \lambda^k/k!$ for $k \geq 0$. Thus

$$E(e^{uY}) = \sum_{k \geq 0} e^{uk} e^{-\lambda} \frac{\lambda^k}{k!} = \exp(\lambda(e^u - 1)).$$

For the convenience of the reader, we recall some simple facts.

Lemma **11.1.1.** — *If a r.v.* Y *satisfies*

$$\textbf{(11.1.3)} \qquad E(e^{uY}) \leq \exp(\lambda(e^u - 1))$$

for $u \geq 0$, *then*

$$\textbf{(11.1.4)} \qquad P(Y \geq t) \leq \exp\left(- t \log \frac{t}{e\lambda}\right).$$

Proof. — One can assume $t \geqslant \lambda$. Write

$$P(Y \geqslant t) \leqslant e^{-tu} E(e^{uY}),$$

use (11.1.3) and take $u = \log(t/\lambda)$. □

Lemma **11.1.2.** — *If the r.v.* Y *is Poisson of parameter* λ, *then*

$$P\left(Y \leqslant \frac{\lambda}{8}\right) \leqslant \exp -\frac{\lambda}{2}.$$

Proof. — Write, for all $u \geqslant 0$,

$$P\left(Y \leqslant \frac{\lambda}{8}\right) \leqslant \exp \frac{\lambda u}{8} Ee^{-uY} = \exp\left(\frac{\lambda u}{8} + \lambda(e^{-u} - 1)\right)$$

and take $u = 2$. □

For $k \geqslant 1$, we denote by $\mathscr{C}_k$ the family of the 2^{2k} " dyadic squares " of side 2^{-} . So the vertices of these squares are of the type $(\ell_1 2^{-k}, \ell_2 2^{-k})$, $0 \leqslant \ell_1, \ell_2 \leqslant 2^k$, $\ell_1, \ell_2 \in \mathbf{N}$.

For $C \in \mathscr{C}_k$, we set

$$Z_C = 1 \quad \text{if } \operatorname{card}(C \cap \Pi) \leqslant \mu 2^{-2k-3},$$
$$Z_C = 0 \quad \text{otherwise.}$$

From (11.1.2) and Lemma 11.1.2, it follows that $\delta_k = P(Z_C = 1)$ satisfies

$$\textbf{(11.1.5)} \qquad \delta_k \leqslant \exp(-\mu 2^{-2k-1}).$$

Now, for $u > 0$,

$$\textbf{(11.1.6)} \qquad \begin{aligned} Ee^{uZ_C} &= 1 - \delta_k + \delta_k e^u \\ &= 1 + \delta_k(e^u - 1) \leqslant \exp \delta_k(e^u - 1). \end{aligned}$$

By (11.1.1) the variables $(Z_C)_{C \in C_k}$ are independent; so that, by (11.1.6),

$$Ee^{u\Sigma_{C \in \mathscr{C}_k} Z_C} \leqslant \exp 2^{2k} \delta_k(e^u - 1),$$

and by Lemma 11.1.1 we have

$$\textbf{(11.1.7)} \qquad P\left(\sum_{C \in \mathscr{C}_k} Z_C \geqslant v\right) \leqslant \exp\left(-v \log \frac{v}{e2^{2k} \delta_k}\right).$$

Observe that $n_k = \sum_{C \in \mathscr{C}_k} Z_C$ is simply the number of squares of $\mathscr{C}_k$ that contain no more than 1/8 of the expected number of points of Π they should contain. Combining (11.1.5) and (11.1.7) we see that

$$\textbf{(11.1.8)} \qquad \forall\, k, \quad P(n_k \geqslant 2e^2\, 2^{2k} \exp(-\mu 2^{-2k-1})) \leqslant \exp(-2e^2\, 2^{2k} \exp(-\mu 2^{-2k-1})).$$

We now fix a number t, and we study how the number n_k can be controlled if one rules out an exceptional set of probability $\leqslant e^{-t^2}$. We assume $t \geqslant 1$, $\mu \geqslant 4$.

We denote by k_1 the largest integer such that

$$(11.1.9) \qquad e^2\, 2^{2k_1} \exp(-\,\mu 2^{-2k_1-1}) \leqslant t^2.$$

Thus, $k_1 \geqslant 0$ and for $k > k_1$ we have

$$e^2\, 2^{2k} \exp(-\,\mu 2^{-2k-1}) \geqslant t^2.$$

We now observe that if $a > 1$, we have $\sum_{\ell \geqslant 0} \exp(-\,2^{2\ell} a) \leqslant 2\exp(-\,a)$, so that, combining with (11.1.8),

$$(11.1.10) \qquad \mathrm{P}(\forall\, k > k_1,\ n_k \leqslant 2e^2\, 2^{2k} \exp(-\,\mu 2^{-2k-1})) \geqslant 1 - 2e^{-2t^2}.$$

Lemma **11.1.3.** — *If* $t \leqslant \sqrt{\mu}/\mathrm{K}$, *we have*

$$\frac{t^2}{\mu\, \delta_{k_1-1}} \geqslant \frac{1}{\sqrt{\delta_{k_1-1}}}.$$

Proof. — It suffices to show that $\sqrt{\delta_{k_1-1}} \leqslant t^2/\mu$. Now, by (11.1.5) and (11.1.9),

$$\sqrt{\delta_{k_1-1}} \leqslant \exp(-\,\mu 2^{-2k_1}) \leqslant \left(\frac{t^2}{e^2\, 2^{2k_1}}\right)^2.$$

Thus it suffices to show that $2^{2k_1} \geqslant t\sqrt{\mu}/e^2$, i.e. $2^{2(k_1+1)} \geqslant 4t\sqrt{\mu}/e^2$. The function $f(x) = e^2 x \exp(-\,\mu/2x)$ is increasing for $x > 0$. Thereby, since $f(2^{2(k_1+1)}) \geqslant t^2$ by definition of k_1, it suffices to show that $f(4t\sqrt{\mu}/e^2) < t^2$, which is equivalent to $\exp(-\,ae^2/2) < 1/16a$ for $a = \sqrt{\mu}/4t$. □

We now apply (11.1.7), taking $k = k_1 - 1$ and $v = e2^{2k_1} t^2/\mu$. We observe that, by Lemma 11.1.3 and (11.1.5), we have, for $t \leqslant \sqrt{\mu}/\mathrm{K}$,

$$\log \frac{v}{e2^{2(k_1-1)}\, \delta_{k_1-1}} \geqslant \log \frac{1}{\sqrt{\delta_{k_1-1}}} \geqslant \mu 2^{-2k_1},$$

so that

$$(11.1.11) \qquad \mathrm{P}\left(n_{k_1-1} \geqslant \frac{e2^{2k_1} t^2}{\mu}\right) \leqslant \exp(-\,2t^2).$$

We now go back to the sample $\mathrm{X}_1, \ldots, \mathrm{X}_\mathrm{N}$ and state our conclusions.

Proposition **11.1.4.** — *Consider* $t \leqslant \sqrt{\mathrm{N}}/\mathrm{K}$. *Denote by* k_0 *the largest integer for which* $2^{2k_0} \leqslant \mathrm{N}$. *There exists an integer* $k_1 \leqslant k_0$ *such that*

$$(11.1.12) \qquad \frac{1}{\mathrm{K}} \log \frac{\mathrm{N}}{t^2} \leqslant 2^{2(k_0-k_1)} \leqslant \mathrm{K} \log \frac{\mathrm{N}}{t^2}$$

and such that, with probability $\geqslant 1 - Ke^{-t^2}$, *we have the following properties, where* m_k *denotes the number of squares* C *of* $\mathscr{C}_k$ *satisfying*

$$\text{(11.1.13)} \qquad \operatorname{card}(C \cap \{X_1, \ldots, X_N\}) \leqslant N2^{-2k-6}.$$

For each $k_1 \leqslant k \leqslant k_0$, *we have*

$$\text{(11.1.14)} \qquad m_k \leqslant K 2^{2k} \exp(-N2^{-2k-6})$$

and

$$\text{(11.1.15)} \qquad m_{k_1-1} \leqslant K \frac{2^{2k_1} t^2}{N}.$$

Proof. — *Step 1.* — Consider the process $\Pi = \Pi_\mu$, for $\mu = N/8$. It follows from (11.1.4) that with probability $\geqslant 1 - \exp(-N/K)$, we have card $\Pi \leqslant N$. It is obvious that, conditionally on the event $\{\operatorname{card} \Pi \leqslant N\}$, the number n_k of squares C of $\mathscr{C}_k$ for which $\operatorname{card}(C \cap \Pi) \leqslant N2^{-2k-6} = \mu/8$ stochastically dominates the number m_k. Thus it suffices to prove (11.1.13) to (11.1.15) for n_k rather than m_k, since, as we consider only $t \leqslant \sqrt{N}/K$, the term $\exp(-N/K)$ is swallowed by the term $K\exp(-t^2)$.

Step 2. — We define k_1 as in (11.1.9). We observe that, since $1 \leqslant N2^{-2k_0} \leqslant 4$ and $t \leqslant \sqrt{N}/K$, we can assume $t^2 \leqslant e^2 2^{2k_0} \exp(-\mu 2^{-2k_0-1})$, so that $k_1 \leqslant k_0$. By (11.1.9) and the definition of k_1 we have

$$\exp \mu 2^{-2(k_1+1)-1} \leqslant \frac{e^2 2^{2k_1+2}}{t^2} \leqslant \frac{KN}{t^2},$$

so that $\mu 2^{-2k_1} \leqslant K \log \frac{KN}{t^2}$, and thus $2^{2(k_0-k_1)} \leqslant K \log \frac{KN}{t^2}$. By (11.1.9),

$$\exp(\mu 2^{-2k_1-1}) \geqslant \frac{e^2 2^{2k_1}}{t^2} \geqslant \frac{2^{2k_0}}{t^2 2^{2(k_0-k_1)}} \geqslant \frac{N}{t^2}\left(K \log \frac{KN}{t^2}\right)^{-1}$$

and this finishes the proof of (11.1.12).

Step 3. — By (11.1.10), we have

$$\text{(11.1.16)} \qquad n_k \leqslant 2e^2 2^{2k} \exp(-N2^{-2k-4})$$

with probability $\geqslant 1 - 2e^{-2t^2}$, for each $k > k_1$. Now we observe that $n_{k_1} \leqslant n_{k_1+1}$. This is obvious, since, if $C \in \mathscr{C}_{k_1}$, one of the 4 squares of $\mathscr{C}_{k_1+1}$ contained in C must contain at most $\operatorname{card}(\Pi \cap C)/4$ points. Therefore, by (11.1.16), we must have

$$n_k \leqslant 8e^2 2^{2k} \exp(-N2^{-2k-6})$$

for each $k \geqslant k_1$.

Also, (11.1.11) shows that, with probability $\geqslant 1 - e^{-2t^2}$, we have

$$(11.1.17) \qquad n_{k_1 - 1} \leqslant \frac{K 2^{2k_1} t^2}{N}.$$

The events described above occur simultaneously with probability $\geqslant 1 - 3e^{-2t^2}$. □

Having studied when and how the sample $X_1, \ldots, X_N$ can have a " deficit " of points, we study how it can have excesses of points. While Proposition 11.1.4 is central to this chapter, the following result will be used only in Section 11.4.

Proposition **11.1.5.** — *Consider the integer* k_0 *of Proposition* 11.1.4, *and consider* $k_2 \leqslant k_0$. *For* $k_2 \leqslant k \leqslant k_0$ *consider a number* r_k *such that* $2^{2k} \geqslant r_k \geqslant 2^{2k} t^2/N$. *Then, with probability* $\geqslant 1 - Ke^{-t^2}$, *the following occurs:*

(11.1.18) *Given* $k_2 \leqslant k \leqslant k_0$, *and given a set* $S \subset \mathscr{C}_k$ *with* card $S \leqslant r_k$, *then*

$$\operatorname{card}\{ i \leqslant N; X_i \in \bigcup \{ C : C \in S \} \} \leqslant KN2^{-2k} r_k + r_k \log \frac{e2^{2k}}{r_k}.$$

Proof. — For a subset U of $[0, 1]^2$, we have

$$(11.1.19) \qquad P(\operatorname{card}\{ i \leqslant N; X_i \in U \} \geqslant u) \leqslant \exp\left(- u \log \frac{u}{eN \mid U \mid}\right).$$

This follows from (11.1.3) and (the argument of) (11.1.6).

For a subset S of $\mathscr{C}_k$, denote U_S the union of the elements of S. It suffices to consider the sets S with card $S = r_k$. For these we get from (11.1.19)

$$P(\operatorname{card}\{ i \leqslant N; X_i \in U_S \} \geqslant u) \leqslant \exp\left(- u \log \frac{u}{eN2^{-2k} r_k}\right).$$

There are at most $\binom{2^{2k}}{r_k} \leqslant \exp(r_k \log(e2^{2k}/r_k))$ choices for S. We take

$$u = r_k \log(e2^{2k}/r_k) + e^3 N2^{-2k} r_k.$$

Thus we see that

$$\binom{2^{2k}}{r_k} \exp\left(- u \log \frac{u}{eN2^{-2k} r_k}\right) \leqslant \binom{2^{2k}}{r_k} \exp(- 2u) \leqslant \left(\frac{r_k}{e2^{2k}}\right)^{r_k} \exp(- t^2).$$

Since $\left(\frac{x}{e2^{2k}}\right)^x \leqslant 2^{-2k}$ for $1 \leqslant x \leqslant 2^{2k}$, we see that (11.1.18) occurs with probability at least $1 - Ke^{-t^2}$. □

11.2. *The Traveling Salesman Problem*

The Traveling Salesman Problem (TSP) requires, given N points $x_1, \ldots, x_N$ in the plane, to find the shortest tour through these points; in other words, to minimize

$$\| x_{\sigma(N)} - x_{\sigma(1)} \| + \sum_{i=1}^{N-1} \| x_{\sigma(i)} - x_{\sigma(i+1)} \|$$

over all permutations $\sigma \in S_N$. The charm of the TSP is that it is the archetype of an untractable question. In this section, we denote by L(F) the length of the shortest tour through F, and we study the r.v. $L_N = L(X_1, \ldots, X_N)$ where $X_1, \ldots, X_N$ are independent uniformly distributed over $[0, 1]^2$.

While the TSP is usually very hard, somewhat surprisingly, it turns out that as far as the concentration of L_N is concerned, it is the easiest problem we will consider. The reason for this is its good regularity properties. The only fact we will use about the TSP is as follows.

Lemma **11.2.1.** — *Consider* $F \subset [0, 1]^2$, $C \in \mathscr{C}_k$, $G \subset C$, *and assume that there is a point of* F *within distance* 2^{-k+2} *of* C. *Then*

$$L(F) \leqslant L(F \cup G) \leqslant L(F) + K2^{-k} \sqrt{\operatorname{card} G}. \tag{11.2.1}$$

Proof. — An essential property of the TSP is its monotonicity: $L(F) \leqslant L(F \cup \{ x \})$, as is seen by bypassing x in a tour through $F \cup \{ x \}$. This implies the left-hand side inequality in (11.2.1). To prove the right-hand side inequality, one first uses the (well-known, elementary) fact that there is a tour through G of length $\leqslant K2^{-k} \sqrt{\operatorname{card} G}$, and one connects this tour to a tour of F.

Theorem **11.2.2.** — *Assume that the functional* L *satisfies the regularity condition of Lemma* 11.2.1. *Then, if* $X_1, \ldots, X_N$ *are independent uniformly distributed over* $[0, 1]^2$, *for each* $t \geqslant 0$ *the r.v.* $L_N = L(X_1, \ldots, X_N)$ *satisfies* $P(| L_N - M | \geqslant t) \leqslant Ke^{-t^2/K}$, *where* M *is a median of* L_N.

Since the TSP is the simplest case we will consider, we will give the shortest proof we can, which is considerably simpler than the original proof. The idea of this proof is, however, a bit tricky; a more straightforward, but somewhat longer proof will be given in Section 11.3.

The basic idea of the whole chapter is as follows: consider $\Omega = [0, 1]^2$, and the subset $A(a)$ of Ω^N that consists of the N-tuples $y_1, \ldots, y_N$ for which $L(y_1, \ldots, y_N) \leqslant a$. When $a = M$ is the median of L, Proposition 2.1.1 shows that, except for a set of probability $2e^{-t^2}$, given $X_1, \ldots, X_N$, we can find $(y_1, \ldots, y_N) \in A(a)$ such that $\operatorname{card} J \leqslant Kt\sqrt{N}$, where $J = \{ i \leqslant N; X_i \neq y_i \}$. Thus we have a tour through $\{ X_i; i \notin J \}$ of length $\leqslant M$. The points X_i, $i \in J$, should be in average at distance $\leqslant K/\sqrt{N}$ of the set $\{ X_i; i \notin J \}$, so each of them can be inserted in the tour by lengthening the tour of at most $K/\sqrt{N}$, for a total lengthening $\leqslant Kt$. This would prove that $P(L_N \geqslant M + Kt) \leqslant e^{-t^2}$. The problem with this argument is that the points X_i, $i \in J$, could be precisely chosen among those which are much further than $K/\sqrt{N}$ from their closest neighbour. So we have to find a way to show that this does not happen, or at least that the effect of this phenomenon does not affect the final result. The idea of this section is to give appropriate weights $\alpha(X_i)$ to each point X_i (the more isolated the point is, the higher its weight) and then to use Theorem 4.1.1 to minimize the influence of points with large weights.

For $x \in [0, 1]^2$, throughout this chapter, $C_k(x)$ denotes the square $C \in \mathscr{C}_k$ containing x. Throughout this section, we will set $F = \{X_1, \ldots, X_N\}$,

$$\mathscr{H}_k = \{ C \in \mathscr{C}_k ; \operatorname{card}(F \cap C) \leqslant N2^{-2k-6} \},$$

and $m_k = \operatorname{card} \mathscr{H}_k$.

We fix $t \leqslant \sqrt{N}/K$, and we recall the integers k_0, k_1 of Proposition 11.1.4.

For $x \in [0, 1]^2$, we define

$$\alpha(x) = \sup\{ 2^{-k} ; k_1 \leqslant k \leqslant k_0 ; \operatorname{card}(\{X_1, \ldots, X_N\} \cap C_k(x) \}) \leqslant N2^{-2k-7} \}$$

when the set on the right is non-empty, and we set $\alpha(x) = 2^{-k_0}$ otherwise.

Proposition **11.2.3.** — *With probability* $\geqslant 1 - K\exp(-t^2)$, *we have*

$$\sum_{i \leqslant N} \alpha^2(X_i) \leqslant K. \tag{11.2.2}$$

Proof. — It should be obvious that

$$\sum_{i \leqslant N} \alpha^2(X_i) \leqslant K + \sum_{k_1 \leqslant k \leqslant k_0} 2^{-2k} \operatorname{card}(F \cap \cup \{ C ; C \in \mathscr{H}_k \})$$
$$\leqslant K + \sum_{k_1 \leqslant k \leqslant k_0} 2^{-2k} \times N2^{-2k-6} \operatorname{card} \mathscr{H}_k .$$

By Proposition 11.1.4, we have

$$m_k = \operatorname{card} \mathscr{H}_k \leqslant K2^{2k} \exp(-N2^{-2k-6}) \tag{11.2.3}$$

with probability $\geqslant 1 - Ke^{-t^2}$, for all $k_1 \leqslant k \leqslant k_0$. The result then follows from the elementary fact that $\sum_{k \leqslant k_0} 2^{-2k} \exp(-N2^{-2k-6}) \leqslant K/N$. □

Proposition **11.2.4.** — *In order to prove Theorem* 11.2.2, *it suffices to prove Proposition* 11.2.5 *below.*

Proposition **11.2.5.** — *Consider* $X_1, \ldots, X_N$, *and a subset* J *of* $\{1, \ldots, N\}$. *Assume that*

$$\sum_{i \notin J} \alpha(X_i) \leqslant Kt, \tag{11.2.4}$$

$$\operatorname{card} \mathscr{H}_{k_1 - 1} \leqslant \frac{K2^{2k_1} t^2}{N}. \tag{11.2.5}$$

Then

(11.2.6) $L(X_1, \ldots, X_N) \leqslant L(\{X_i; i \in J\}) + K' t,$

where K′ *depends on the constants in* (11.2.4) *and* (11.2.5) *only.*

Proof of Proposition 11.2.4. — To prove Theorem 11.2.1, since $L_N \leqslant K\sqrt{N}$, it suffices to consider the case $t \leqslant \sqrt{N}/K$. We fix such a t, and we consider a such that $P(L_N \leqslant a) \geqslant e^{-t^2}$. We will prove that

(11.2.7) $P(L_N \geqslant a + Kt) \leqslant Ke^{-t^2}$

and this clearly implies the result. The condition $P(L_N \leqslant a) \geqslant e^{-t^2}$ means $P(A(a)) \geqslant e^{-t^2}$ (where P denotes now the product measure on Ω^N). If we combine Lemma 4.1.2 and Theorem 4.1.1, we see that with probability $\geqslant 1 - e^{-t^2}$, the set $\{X_1, \ldots, X_N\}$ has the property that we can find $(y_1, \ldots, y_N) \in A(a)$ for which

$$\sum_{i \notin J} \alpha(X_i) \leqslant Kt \sqrt{\sum_{i \leqslant N} \alpha(X_i)^2},$$

where $J = \{i \leqslant N; X_i = y_i\}$. Now, by Proposition 11.2.3 and Proposition 11.1.4, we can moreover assume, with probability $\geqslant 1 - Ke^{-t^2}$, that $\sum_{i \leqslant N} \alpha(X_i)^2 \leqslant K$ and that (11.2.5) holds. By Proposition 11.2.5, we than have

$$L(X_1, \ldots, X_N) \leqslant L(\{y_i; i \in J\}) + Kt \leqslant a + Kt. \quad \square$$

Proof of Proposition 11.2.5. — We set $F' = \{X_i; i \in J\}$, $G = \{X_i; i \notin J\}$. We have to incorporate the points of G into a tour through F′ without lengthening too much the tour.

Step 1. — For $0 \leqslant k < k_0$, we denote by U'_k the collection of those $C \in \mathscr{C}_k$ that satisfy $C \cap F' = \emptyset$; we set $U'_{k_0} = \mathscr{C}_{k_0}$, and, for $0 \leqslant k \leqslant k_0$, we denote by U_k the collection of those $C \in U'_k$ that are not included in any $C' \in U_{k-1}$. Thus, if $C \in U_k$, its distance to F′ is $\leqslant 2^{-k+2}$.

By repeated applications of Lemma 11.2.1, we see that

$$L(X_1, \ldots, X_N) \leqslant L(\{X_i; i \in J\}) + K \sum_{0 \leqslant k \leqslant k_0} \sum_{C \in U_k} 2^{-k} \sqrt{\operatorname{card}(G \cap C)}.$$

Thereby, it suffices to show that this double sum is $\leqslant Kt$.

Step 2. — We consider three types of terms:

Type 1: $\operatorname{card}(G \cap C) \geqslant N 2^{-2k-7}$.

In that case, since $\alpha(X_i) \geqslant N^{-1/2}$, we have

(11.2.8) $$2^{-k}\sqrt{\operatorname{card}(G \cap C)} \leqslant \frac{K}{\sqrt{N}} \operatorname{card}(G \cap C) \leqslant K \Sigma\{\alpha(X_i); X_i \in C \cap G\}.$$

Type 2: $k \geqslant k_1$, $\mathrm{card}(G \cap C) < N2^{-2k-7}$.

In that case, the definition of $\alpha(X_i)$ shows that $\alpha(X_i) \geqslant 2^{-k}$ for $X_i \in C$. Thus

$$2^{-k}\sqrt{\mathrm{card}(G \cap C)} \leqslant 2^{-k}\,\mathrm{card}(G \cap C) \leqslant \Sigma\{\alpha(X_i)\,;\, X_i \in C \cap G\}.$$

We observe that the total contribution of the terms of Types 1 and 2 is $< Kt$ by (11.2.4), since for different values of k, the unions of the sets in U_k are disjoint by construction.

Type 3: $k < k_1$, $\mathrm{card}(G \cap C) < N2^{-2k-7}$.

Step 3. — We control the contribution of the terms of Type 3. We denote by V_k the union of the sets $C \in U_k$ for which $\mathrm{card}(G \cap C) < N2^{-2k-7}$. Denoting by $|V|$ the area of V, the key observation is that, under (11.2.5) we have

$$\textbf{(11.2.9)} \qquad \Big|\bigcup_{k<k_1} V_k\Big| \leqslant \frac{Kt^2}{N}.$$

The reason is simply that if $C \in \mathscr{C}_k$ satisfies $\mathrm{card}(G \cap C) < N2^{-2k-7}$, when $C \in U_k$, $\mathrm{card}(G \cap C) = \mathrm{card}(F \cap C) < N2^{-2k-7}$, so that, among the $2^{2(k_1-k-1)}$ squares C' of $\mathscr{C}_{k_1-1}$ that are contained in C, at least half must satisfy $\mathrm{card}(C' \cap F) < N2^{-2(k_1-1)-6}$, so belong to $\mathscr{H}_{k_1-1}$. Thereby the area of $\bigcup_{k<k_1} V_k$ can be at most twice the area of the union of $\mathscr{H}_{k_1-1}$.

There are $2^{2k}\,|V_k|$ sets C of $\mathscr{C}_k$ included in V_k. Thus, by Cauchy-Schwarz, we have

$$\sum_{C\in\mathscr{C}_k,\, C\subset V_k} 2^{-k}\sqrt{\mathrm{card}(G \cap C)} \leqslant 2^{-k}\sqrt{\mathrm{card}(G \cap V_k)\, 2^{2k}\,|V_k|}$$
$$= \sqrt{\mathrm{card}(G \cap V_k)\,|V_k|}.$$

Using Cauchy-Schwarz again, the sum of these terms over $k < k_1$ is at most $\sqrt{|V|\,\mathrm{card}(G \cup V)} \leqslant \sqrt{N|V|}$ where $V = \bigcup_{k<k_1} V_k$. This is less than Kt by (11.2.8). □

11.3. *The Minimum Spanning Tree*

A spanning tree of a finite subset F of $\mathbf{R}^2$ is a connected set that is a union of segments (called *edges*) each of which joins two points of F. Its length is the sum of the lengths of these segments. We denote by L(F) the length of the shortest (= minimum) spanning tree of F. An interesting difference with the TSP is that it can happen that $L(F \cup \{x\}) < L(F)$. This is e.g. the case if F consists of the three vertices of an equilateral triangle and x is its center.

The regularity property of L that we will use is as follows.

Lemma **11.3.1.** — *Consider* $C \in \mathscr{C}_k$ $(k \geqslant 1)$ *and a subset* F *of* $[0, 1]^2$. *Assume that each* $C' \in \mathscr{C}_{k-1}$ *that is within distance* 2^{-k+5} *of* C *meets* F. *Consider a subset* G *of* C. *Then*

$$\textbf{(11.3.1)} \qquad |L(F \cup G) - L(F)| \leqslant K2^{-k}\sqrt{\mathrm{card}\, G}.$$

Proof. — *Step 1.* — The inequality

$$L(F \cup G) \leqslant L(F) + K2^{-k}(\text{card } G)^{1/2}$$

is proved as in the case of the TSP. The problem is the reverse inequality.

Consider a minimum spanning tree of $F \cup G$. We remove all the edges adjacent to G. This breaks the spanning tree in a number of pieces, and we have to add edges to connect it again. We will prove two facts.

Fact 1. — There is at most 6 card G pieces;

Fact 2. — Each of the pieces contains a point within distance $K2^{-k}$ of C.

Once this is known, we simply take a point in each of these pieces within distance $K2^{-k}$ of C. We build a tour of length $\leqslant K2^{-k}(\text{card } G)^{1/2}$ through these points to reconnect the pieces.

Step 2. Proof of Fact 1. — Consider three points x, a, b of $F \cup G$, such that the segments $[x, a]$, $[x, b]$ both belong to a minimum spanning tree of $F \cup G$. Then we must have $\| a - b \| \geqslant \| x - a \|$ for otherwise we could remove the edge $[x, a]$ and replace it by $[a, b]$ to get a shorter spanning tree. Similarly, we have $\| a - b \| \geqslant \|x - b \|$. Thus the angle between the lines xa, xb is at least $\pi/3$. Thereby the spanning tree must contain at most 6 edges adjacent to each point. Thus removing k points and the edges adjacent creates at most $6k$ connected components.

Step 3. Proof of Fact 2. — Consider a finite set H of $[0, 1]^2$. Consider a, b in H, and assume that $[a, b]$ belongs to a minimum spanning tree of H. We show that the " lens "

$$L_{a,b} = \{ x ; \| a - x \| < \| a - b \|, \| b - x \| < \| a - b \| \} \tag{11.3.2}$$

does not meet H. Indeed if we remove $[a, b]$ from the minimum spanning tree, we split H into the component H_a containing a and the component H_b containing b. If there existed $c \in L_{a,b} \cap H_a$, we could remove the edge $[a, b]$ from the minimum spanning tree, and replace it by $[c, b]$ to get a shorter spanning tree. Similarly, $L_{a,b} \cap H_b = \emptyset$.

We apply the above result to $H = F \cup G$. An edge $[a, b]$ from a minimal spanning tree of H is such that $L_{a,b}$ does not contain a square C′ in $\mathscr{C}_{k-1}$ within distance 2^{-k+5} of C, because it is assumed that all such squares meet F, hence H. Thus, if $a \in C$, then, clearly, $\| b - a \| \leqslant K2^{-k}$. □

The main result of this section is as follows.

Theorem **11.3.2.** — *Assume that the functional* L *satisfies the regularity condition of Lemma* 11.3.1. *Then, if* $X_1, \ldots, X_N$ *are independent uniformly distributed over* $[0, 1]^2$, *the r.v.* $L_N = L(X_1, \ldots, X_N)$ *satisfies*

$$\forall\, t \geqslant 0, \quad P(| L_N - M | \geqslant t) \leqslant K e^{-t^2/K}$$

where M *is a median of* L_N.

One central idea of the approach will be to condition with respect to $X_1, \dots, X_m$, where $m = [N/2]$. The size of the holes of $\{X_1, \dots, X_N\}$ are then controlled by the sizes of the holes of $\{X_1, \dots, X_m\}$, independently of $X_{m+1}, \dots, X_N$. The main part of the proof of Theorem 11.3.2 is to obtain the following statement. We set $\Omega = [0, 1]^2$.

Proposition **11.3.3.** — *Consider an integer n with* $\left|\frac{N}{2} - n\right| \leqslant 1$. *We write* $\Omega_1 = \Omega^n$, $\Omega_2 = \Omega^{N-n}$; *we denote by* P_1, P_2 *the product measures on* Ω_1, Ω_2 *respectively. Given* $0 < t < \sqrt{N}/K$, *there exists a subset* H_t *of* Ω_1 *such that* $P_1(H_t) \leqslant K_1 e^{-t^2}$, *and that, whenever* $(x_1, \dots, x_n) \notin H_t$, *the r.v.*

$$L' = L'(X_{n+1}, \dots, X_N) = L_N(x_1, \dots, x_n, X_{n+1}, \dots, X_N)$$

defined on Ω_2 *has the following property*

(11.3.3) *If* $P_2(L' \leqslant a) \geqslant e^{-t^2}$, $P_2(L' \geqslant b) \geqslant e^{-t^2}$, *then* $b - a \leqslant Kt$.

First, we prove that Proposition 11.3.3 implies Theorem 11.3.2. To prove that theorem, it suffices to prove the following statement:

$$\text{if } P(L_N \leqslant a) \geqslant 2e^{-t^2/2},\ P(L_N \geqslant b) \geqslant 2e^{-t^2/2}, \text{ then } b - a \leqslant Kt.$$

Consider the set $A = \{L_N \leqslant a\}$ in Ω^N. We will write $\Omega^N = \Omega_1 \times \Omega_2$ ($\Omega_1 = \Omega^n$; $\Omega_2 = \Omega^{N-n}$) and $P = P_1 \otimes P_2$. Thus, given $\omega_1 \in \Omega_1$, we define L' on Ω_2 by $L'(\omega_2) = L_N(\omega_1, \omega_2)$. For $\omega_1 \in \Omega_1$, we write

$$A(\omega_1) = \{\omega_2 \in \Omega_2;\ (\omega_1, \omega_2) \in A\}.$$

Since $P(A) \geqslant 2e^{-t^2/2}$, the set

$$C_1 = \{\omega_1 \in \Omega_1;\ P_2(A(\omega_1)) \geqslant e^{-t^2/2}\}$$

satisfies $P_1(C_1) \geqslant e^{-t^2/2}$. Consider $C_2 = C_1 \backslash H_t$, so that $P_1(C_2) \geqslant e^{-t^2/2} - K_1 e^{-t^2}$. When $\omega_1 \in C_2$, we have $P_2(L' \leqslant a) \geqslant e^{-t^2/2}$, so that by (11.3.3) we have

$$P_2(L' \leqslant a + Kt) \geqslant 1 - e^{-t^2}.$$

By Fubini theorem, we get

$$P(W_1) \geqslant (1 - e^{-t^2})\, P(C_2 \times \Omega_2), \tag{11.3.4}$$

where $W_1 = \{L_N \leqslant a + Kt\} \cap (C_2 \times \Omega_2)$.

We observe that (11.3.3) implies

$$P_2(L' \geqslant b) \geqslant e^{-t^2} \Rightarrow P_2(L' \geqslant b - Kt) \geqslant 1 - e^{-t^2}.$$

Thus, we can apply the same argument as above to show that

$$P(W_2) \geqslant (1 - e^{-t^2})\, P(\Omega_1 \times D_2), \tag{11.3.5}$$

where $W_2 = \{ L_N \geqslant b - Kt \} \cap (\Omega_1 \times D_2)$ and $P_2(D_2) \geqslant e^{-t^2/2} - K_1 e^{-t^2}$. For t large enough,

$$P((C_2 \times \Omega_2)\backslash W_1) + P((\Omega_1 \times D_2)\backslash W_2) < P(C_2 \times D_2)$$

so that $W_1 \cap W_2 \neq \emptyset$. □

We now start the proof of Proposition 11.3.3. Consider $x_1, \ldots, x_n \in \Omega = [0, 1]^2$, and set $F' = \{ x_1, \ldots, x_n \}$. Denote by m'_k the number of squares of $\mathscr{C}_k$ that do not meet F'. We consider the integers k_1, k_0 of Proposition 11.1.4 (defined using n rather than N). We define H_t as the set of n-tuples $(x_1, \ldots, x_n)$ for which

(11.3.6) For each k, $k_1 \leqslant k \leqslant k_0$, we have $m'_k \leqslant K2^{2k} \exp(-n2^{-2k-6})$,

(11.3.7) $$m'_{k_1-1} \leqslant \frac{K2^{2k_1} t^2}{n}.$$

Thereby, $P_1(H_t) \geqslant 1 - Ke^{-t^2}$ by Proposition 11.1.4.

We now fix $(x_1, \ldots, x_n)$ such that (11.3.6), (11.3.7) hold and we start the proof of (11.3.3). For $x \in [0, 1]^2$, we denote by $\ell(x)$ the smallest integer ℓ such that there is $C \in \mathscr{C}_\ell$, C within distance $2^{-\ell+4}$ of $C_\ell(x)$, with the property that $F' \cap C = \emptyset$. Thus, by definition, we observe

(11.3.8) if $\ell = \ell(x)$, any square $C' \in \mathscr{C}_{\ell-1}$ that is within distance $2^{-\ell+5}$ of $C_\ell(x)$ meets F'.

We also observe that if $y \in C_{\ell(x)}(x)$, then $\ell(y) = \ell(x)$, so that $V_\ell = \{ x ; \ell(x) = \ell \}$ is a union of squares of $\mathscr{C}_\ell$.

Lemma **11.3.4.** — *a) We have, for each* $k_1 \leqslant k \leqslant k_0$,

(11.3.9) $$| V_k | \leqslant K \exp(-n2^{-2k-6}).$$

b) $$\Big| \bigcup_{\ell < k_1} V_\ell \Big| \leqslant \frac{Kt^2}{n} \leqslant \frac{Kt^2}{N}.$$

Proof. — Let us denote by U'_ℓ the union of the elements of $\mathscr{C}_\ell$ that do not meet F', and set $U_\ell = U'_\ell \backslash \bigcup_{k<\ell} U'_k$. It suffices to observe that if $x \in V_\ell$, then $C_{\ell(x)}$ is within distance $2^{-\ell+4}$ of U_ℓ, so that $| V_\ell | \leqslant K | U_\ell |$, and the result follows from (11.3.6), (11.3.7). □

We consider the function $g(x) = 2^{-\max(k_1, \ell(x))}$. By (11.1.12), we have

(11.3.10) $$\| g \|_\infty \leqslant 2^{-k_1} \leqslant \frac{K}{\sqrt{n}} \left(\log \frac{n}{t^2}\right)^{1/2} \leqslant \frac{K}{t}.$$

By (11.3.9) and an obvious computation, we have

(11.3.11) $$\| g \|_2 \leqslant K/\sqrt{n}.$$

To prove (11.3.3), we have to prove that if a, b are such that $P_2(L' \leqslant a) \geqslant e^{-t^2}$, $P_2(L' \geqslant b) \geqslant e^{-t^2}$, then $b - a \leqslant Kt$. We now appeal to Corollary 2.4.5 with $u = Kt$, for the function $h(x,y) = g(x) + g(y)$. From (11.3.10), (11.3.11), we see that we can find $y_{n+1}, \ldots, y_N, z_{n+1}, \ldots, z_N$ such that

(11.3.12) $$L'(y_{n+1}, \ldots, y_N) \leqslant a; L'(z_{n+1}, \ldots, z_N) \geqslant b,$$

and

(11.3.13) $$\sum_{i \in J} (g(y_i) + g(z_i)) \leqslant Kt,$$

where $J = \{ n + 1 \leqslant i \leqslant N; y_i \neq z_i \}$.

Consider the set F that consists of the points $x_1, \ldots, x_n$, as well as the points y_i, $i \notin J$. We will prove:

(11.3.14) $$| L'(y_{n+1}, \ldots, y_N) - L(F) | \leqslant Kt.$$

The same argument will show that

$$| L'(z_{n+1}, \ldots, z_N) - L(F) | \leqslant Kt$$

and this will finish the proof.

First we observe from (11.3.1) that if $F_1 \supset F$, and if $\ell(x) \geqslant k_1$, then

$$| L(F_1 \cup \{ x \}) - L(F) | \leqslant Kg(x).$$

Thereby, it follows from (11.3.13) that we can add to F all the points y_i, $i \in J$, for which $\ell(y_i) \geqslant k_1$, without changing the value of f by more than Kt. Denote by G the set of the other points y_i. We observe that G is contained in $\bigcup_{\ell < k_1} V_\ell$. Consider $C \in \mathscr{C}_\ell$, $C \subset V_\ell$. By (11.3.1), we have, for any set F_1 containing F, that

$$| L(F_1 \cup (G \cap C)) - L(F_1) | \leqslant K2^{-\ell}(\text{card } G \cap C)^{1/2}.$$

Therefore it suffices to show that

$$\sum_{\ell < k_1} 2^{-\ell} \sum_{C \subset V_\ell} (\text{card } G \cap C)^{1/2} \leqslant Kt.$$

But this is shown as in Step 3 of the proof of Proposition 11.2.5. □

11.4. *Gabriel Graph and Voronoi Polygons*

Given a subset F of $[0, 1]^2$, its Gabriel graph is the set of edges $[a, b]$ such that the closure $\overline{L}_{a,b}$ of the set $L_{a,b}$ of (11.3.2) meets F only in a and b. When the set F has the property that it does not contain points x, y, z such that $\| x - y \| = \| x - z \|$, (a property that is satisfied with probability one for random sets) this is equivalent to saying that F contains the edge $[a, b]$ if and only if $L_{a,b}$ does not meet F. In that case, the Gabriel graph contains the minimum spanning tree, as is shown in the course of the proof of Lemma 11.3.1. As in the case of the MST, at most 6 edges are adjacent to each point of F.

We denote by L(F) the length of the Gabriel graph. An interesting feature of this functional is that, in certain special configurations, adding a single point creates a big decrease of L(F). A typical such configuration consists of the points $(0, k/n)$, $(1, k/n)$, $0 \leqslant k \leqslant n$. The Gabriel graph contains all the edges between $(0, k/n)$ and $(1, k/n)$. All these edges will disappear when one adds the middle of the unit square to F. The following lemma shows that the previous example is close to be the worst possible behavior.

Lemma **11.4.1.** — *Consider* $C \in \mathscr{C}_k$, F *a subset of* $[0, 1]^2$, *and assume:*

(11.4.1) *every element* C' *of* $\mathscr{C}_{k-1}$ *that is within distance of* 2^{-k+3} *of* C *meets* F.

Then, if $G \subset C$, *we have*

(11.4.2) $$| L(F) - L(F \cup G) | \leqslant K2^{-k} \operatorname{card}\{(F \cup G) \cap B(C, K2^{-k})\},$$

where $B(C, r)$ *denotes the set of points within distance* r *of* C.

Comment. — The difference with Lemma 11.3.1 is that the bound now depends upon $F \cup G$ rather than G alone.

Proof. — As already seen, a point is adjacent to at most 6 edges, and, as in the case of the MST, edges adjacent to G have a length $\leqslant K2^{-k}$. Thus

$$L(F \cup G) \leqslant L(F) + K2^{-k} \operatorname{card} G.$$

To prove the reverse inequality, we observe that the edges $[a, b]$ that belong to the Gabriel graph of F but not to the Gabriel graph of $F \cup G$ are exactly these for which $\overline{L}_{a,b} \setminus \{a, b\}$ meets G but not F. Then $\| a - b \| \leqslant K2^{-k}$, for otherwise there would exist $C' \in \mathscr{C}_{k-1}$ within distance 2^{-k+3} of C that would not meet F. This implies, since $\overline{L}_{a,b}$ meets G, that $a, b \in B(C, K2^{-k})$. In the Gabriel graph of F, there are at most $6 \cdot \operatorname{card}(F \cap B(C, K2^{-k}))$ edges adjacent to points in $B(C, K2^{-k})$, so at most that many edges can be removed. □

Another natural example of functional that satisfies Lemma 11.4.1 is the total length of the Voronoi polygons. If F is a subset of $[0, 1]^2$, and $x \in F$, let us define the Voronoi polygon V_x of x as the set of all points y of $[0, 1]^2$ for which $d(x, y) = d(y, F \setminus \{x\})$. (This name is a bit abusive since when x is close to the boundary of $[0, 1]^2$ this set is not a polygon.) Denote by L(F) the sum of the lengths of the Voronoi polygons of all points of F. We sketch a proof that L(F) satisfies the condition of Lemma 11.4.1. First, we observe that if $y \in V_x$, there is no point of F within distance less than $\| x - y \|$ of y. Thus, if $x \in G$, the Voronoi polygon of x (with respect to $F \cup G$) is under (11.4.1) entirely contained in $B(x, K2^{-k})$, so is of length $\leqslant K2^{-k}$. Thus

$$L(F \cup G) \leqslant L(F) + K2^{-k} \operatorname{card} G.$$

To prove the reverse inequality, consider a point a belonging to the Voronoi polygon of $x \in \mathrm{F}$, with respect to F, but not with respect to $\mathrm{F} \cup \mathrm{G}$. Then there is no point of F within distance less than $\| x - a \|$ of a, but there is at least a point of G. Under (11.4.1) we have $a, x \in \mathrm{B}(\mathrm{C}, \mathrm{K}2^{-k})$; but the total length of the part of the Voronoi polygons of F contained in $\mathrm{B}(\mathrm{C}, \mathrm{K}2^{-k})$ is easily seen to be $\leqslant \mathrm{K}2^{-k}\,\mathrm{card}(\mathrm{F} \cap \mathrm{B}(\mathrm{C}, \mathrm{K}2^{-k}))$.

Theorem **11.4.2.** — *Consider a functional that satisfies the condition of Lemma* 11.4.1. *Set, as usual,* $\mathrm{L} = \mathrm{L}_\mathrm{N} = \mathrm{L}(\mathrm{X}_1, \ldots, \mathrm{X}_\mathrm{N})$, *and consider the median* M *of* L_N. *Then*

$$\textbf{(11.4.3)} \qquad \forall\, t > 0, \quad \mathrm{P}(|\,\mathrm{L} - \mathrm{M}\,| \geqslant t) \leqslant \mathrm{K} \exp\left(-\frac{1}{\mathrm{K}} \min\left(t^2, \left(t\sqrt{\mathrm{N}}\right)^{2/3}\right)\right).$$

In particular, the tails of L_N are subgaussian for values of t up to $\mathrm{N}^{1/4}$. We now sketch, in the case of the Gabriel graph, why, within logarithmic terms, the exponent in (11.4.3) is correct for $t \geqslant \mathrm{N}^{1/4}$. We give an informal argument, that could be made rigorous. For simplicity, let us argue about $\mathrm{L}(\Pi)$, where Π is a Poisson point process of intensity N. Consider $u \leqslant \sqrt{\mathrm{N}}$, and let $a = u/\sqrt{\mathrm{N}} \leqslant 1$. Denote by k the cardinality of $\Pi \cap [0, a]^2$. When k is even, conditionally on k, with probability $\geqslant (1/\mathrm{K}k^4)^k$, the k points of $\Pi \cap [0, a]^2$ are such that each of the discs of center $(\eta, 2\ell a/k)$, for $\eta \in \{0, a\}$, $1 \leqslant \ell \leqslant k/2$, and of radius $a/4k^2$ contains exactly one of these points. Then the Gabriel graph of Π contains the edge from the point in the disc of center $(0, 2\ell a/k)$ to the point in the disc of center $(a, 2\ell a/k)$, for a total length of order ka. Now with overwhelming probability k is of order u^2; so, with probability $\geqslant (1/\mathrm{K}u^8)^{u^2}$ we get the exceptional configuration described above that creates an abnormal length of order $t = u^2 a = u^3/\sqrt{\mathrm{N}}$. Now $u = \left(t\sqrt{\mathrm{N}}\right)^{1/3}$, and

$$\left(\frac{1}{\mathrm{K}u^8}\right)^{u^2} \geqslant \exp\left(-\frac{1}{\mathrm{K}}\left(t\sqrt{\mathrm{N}}\right)^{2/3} \log t\sqrt{\mathrm{N}}\right).$$

So this later quantity is a lower bound on the probability that we get an abnormal length of order t that will have L exceed the median by t.

To prove Theorem 11.4.2, we observe that, since $|\,\mathrm{L}_\mathrm{N}\,| \leqslant \mathrm{KN}$ by (11.4.2), it suffices to prove (11.4.3) for $t \leqslant \mathrm{N}/\mathrm{K}$. We follow the scheme of Section 11.3. It is enough to modify Proposition 11.3.3 so that, when $t \leqslant \sqrt{\mathrm{N}}/\mathrm{K}$, (11.3.3) can be replaced by

$$\textbf{(11.4.4)} \qquad \text{If } \mathrm{P}_2(\mathrm{L}' \leqslant a) \geqslant 2e^{-t^2},\ \mathrm{P}_2(\mathrm{L}' \geqslant b) \geqslant 2e^{-t^2}, \text{ then } b - a \leqslant \mathrm{K}\left(t + \frac{t^3}{\sqrt{\mathrm{N}}}\right).$$

Once this is known, as in Section 11.3, we prove that

$$\mathrm{P}(\mathrm{L} \leqslant a) \geqslant 2e^{-t^2/2},\ \mathrm{P}(\mathrm{L} \geqslant b) \geqslant 2e^{-t^2/2} \text{ imply } b - a \leqslant \mathrm{K}\left(t + \frac{t^3}{\sqrt{\mathrm{N}}}\right).$$

Theorem 11.4.2 follows since, if we set $u = t + t^3/\sqrt{\mathrm{N}}$, for $u \leqslant \mathrm{N}/\mathrm{K}$ we have $t \leqslant \sqrt{\mathrm{N}}/\mathrm{K}$; moreover, $t^2 \geqslant \mathrm{K}^{-1} \min\left(u^2, \left(u\sqrt{\mathrm{N}}\right)^{2/3}\right)$.

The construction of H_t and the proof of (11.4.4) will parallel the proof of Proposition 11.3.3. In order to avoid repetition, we will not reproduce the entire argument but simply explain the necessary modifications.

The construction of H_t is modified as follows. We require that for $k_1 - 1 \leqslant k \leqslant k_0$, and each subset S of $\mathscr{C}_k$, with card $S \leqslant r_k$, one has

(11.4.5) $$\operatorname{card}\{i \leqslant n;\ x_i \in \cup\{C;\ C \in S\}\} \leqslant KN2^{-2k} r_k + r_k \log \frac{e2^{2k}}{r_k},$$

where we set $r_{k_1-1} = 2^{2k_1} t^2/n$ and, for $k \geqslant k_1$,

$$r_k = Kt2^{4k-3k_0}.$$

We observe that, using (11.1.12),

$$2^{-2k} r_k \geqslant Kt2^{2k_1-3k_0} \geqslant \frac{Kt}{\sqrt{N}} 2^{2k_1-2k_0} \geqslant \frac{t^2}{n},$$

provided K is large enough. It then follows from Proposition 11.1.5 that imposing these extra conditions does not change the fact that $P_1(H_t) \geqslant 1 - Ke^{-t^2}$.

We change the definition of the function $g(x)$ to

$$g(x) = \frac{1}{2^{k_0}} (2^{k_0 - \max(k_1, \ell(x))})^4.$$

Thus

(11.4.6) $$\| g \|_\infty \leqslant \frac{1}{2^{k_0}} (2^{k_0-k_1})^4 \leqslant \frac{K}{\sqrt{n}} \left(\log \frac{Kn}{t^2}\right)^2 \leqslant \frac{K}{t}$$

and, obviously, (11.3.11) still holds.

Suppose now that we are given a, b with

$$P_2(L' \leqslant a) \geqslant 2e^{-t^2}, \qquad P_2(L' \geqslant b) \geqslant 2e^{-t^2}.$$

Using Proposition 11.1.5 again, we see that we can find a set $A \subset \{L' \leqslant a\}$, $P_2(A) \geqslant e^{-t^2}$, such that whenever $(y_{n+1}, \ldots, y_N) \in A$, we have

(11.4.7) for each $k_1 - 1 \leqslant k \leqslant k_0$ and each subset S of $\mathscr{C}_k$ such that card $S \leqslant r_k$, one has

$$\operatorname{card}\{n + 1 \leqslant i \leqslant N;\ y_i \in \cup\{C;\ C \in S\}\} \leqslant KN2^{-2k} r_k + r_k \log \frac{e2^{2k}}{r_k}.$$

We then consider, using Proposition 11.1.5 again, a subset B of $\{L' \geqslant b\}$ with $P_2(B) \geqslant e^{-t^2}$, such that when $(z_{n+1}, \ldots, z_N) \in B$, the property similar to (11.4.7) holds.

We now appeal to Corollary 2.4.5, to find $(y_{n+1}, \ldots, y_N) \in A$, $(z_{n+1}, \ldots, z_N) \in B$ such that if $J = \{i;\ n + 1 \leqslant i \leqslant N, y_i \neq z_i\}$, then

(11.4.8) $$\sum_{i \in J} g(y_i) + g(z_i) \leqslant Kt.$$

We denote by F the collection of points that consists of the points $(x_i)_{i \leqslant n}$, together with the points y_i, $i \notin J$. We denote by G the collection of points y_i, $i \in J$. We have to show that

(11.4.9) $$| L(F \cup G) - L(F) | \leqslant K(t + t^3/\sqrt{n}).$$

Let us denote by S_ℓ the collection of squares $C \in \mathscr{C}_\ell$ that contain at least one point y_i, $i \in J$, $\ell(y_i) = \ell$. It follows from (11.4.1) that, if $F \subset F_1 \subset F \cup G$, and if $C \in S_\ell$, we have

$$| L(F_1 \cup (G \cap C)) - L(F_1) | \leqslant K2^{-\ell} \operatorname{card}\{(F \cup G) \cap B(C, K2^{-\ell})\}.$$

Thereby, adding to F_1 all the points of $U_\ell \cap G$, where $U_\ell = \cup \{ C; C \in S_\ell \}$, we cannot change the value of L by more than

(11.4.10) $$2^{-\ell} \operatorname{card}\{(F \cup G) \cap B(U_\ell, K2^{-\ell})\}.$$

Since, for $\ell(y_i) = \ell \geqslant k_1$, we have by definition

$$g(y_i) \geqslant \frac{1}{2^{k_0}} (2^{k_0 - \ell})^4,$$

and since $\sum_{i \in J} g(y_i) \leqslant Kt$ by (11.4.8), we see that

$$\operatorname{card} S_\ell \leqslant Kt2^{4\ell - 3k_0}.$$

Now, $B(U_\ell, K2^{-\ell})$ is contained in a union of $\leqslant K \operatorname{card} S_\ell$ squares C of $\mathscr{C}_\ell$. Thereby, it follows from (11.4.5), (11.4.7) that the quantity (11.4.10) is bounded by

$$2^{-\ell} K \left(N2^{-2\ell} r_\ell + r_\ell \log \frac{e2^{2\ell}}{r_\ell} \right) \leqslant Kt2^{\ell - k_0},$$

and these quantities have a sum $\leqslant Kt$.

Now we have to control the influence of the points y_i for which $\ell(y_i) < k_1$.

We denote by V_ℓ the set $\{ \ell(x) = \ell \}$. We recall that by Lemma 11.3.4 we have $| \bigcup_{\ell < k_1} V_\ell | \leqslant Kt^2/N$. Since V_ℓ is union of squares of $\mathscr{C}_\ell$, we have in particular that $V_\ell = \emptyset$ for $\ell \leqslant k_3$, where $2^{-k_3} \leqslant Kt/\sqrt{N}$. Adding to a set F_1 such that $F \subset F_1 \subset F \cup G$, the points of $G \cap V_\ell$ can, by (11.4.1), change the value of L by at most

$$2^{-\ell} \operatorname{card}((F \cup G) \cap B(V_\ell, K2^{-\ell})).$$

Now we observe that $| B(V_\ell, K2^{-\ell}) | \leqslant K | V_\ell |$. Thus the total contribution of the points of $G \cap V_\ell$ is bounded by

(11.4.11) $$2^{-\ell} \operatorname{card}((F \cup G) \cap V),$$

where $|V| \leqslant Kt^2/N$ and V is a union of squares of $\mathscr{C}_{k_1-1}$. By (11.4.7), the summation of all these quantities over $\ell \geqslant k_3$ is at most

$$K2^{-k_3}\,\mathrm{card}((F \cup G) \cap V) \leqslant \frac{Kt}{\sqrt{N}}\left(N2^{-2k_1}r_{k_1-1} + r_{k_1-1}\log\frac{e2^{2k_1-2}}{r_{k_1-1}}\right)$$

$$\leqslant \frac{Kt^3}{\sqrt{N}}\left(1 + \frac{2^{2k_1}}{N}\log\frac{KN}{t^2}\right).$$

But, using the definition of k_1, the last term is easily seen to be bounded by a constant. □

11.5. *Simple matching*

In this section (for reasons that will become apparent later) we work in $[0, 1]^d$ for $d \geqslant 2$.

A *matching* of a set F is a decomposition of F as a union of disjoint pairs of points (points of the same pair are matched); we make the convention that when card F is odd, there is exactly one point that is unmatched (does not belong to any pair). A minimum matching is a matching that minimizes the sum of the distances of pairs of matched points. We denote by L(F) the length of a minimum matching of F. For simplicity, the point to which a given point is matched is called its partner.

Our interest in that functional stems from the fact that it apparently does not have good regularity properties. It is obvious that

$$L(F \cup \{x\}) - L(F \cup \{y\}) \leqslant \|x - y\|,$$

but in certain configurations this cannot be improved upon. The problem is that if one tries to match y to a point different from the partner of x, the partner of x has to find a new partner, etc., and there is no apparent way to control this chain reaction.

While the behavior of F is not good as far as the change of one point of F is concerned, the situation is somewhat better when a significant number of points of F are changed. We set $L'(F) = \sup\{L(F'); F' \subset F\}$.

Lemma **11.5.1.** — *One has* $|L(F) - L(G)| \leqslant L'(F \Delta G) + \sqrt{d}$.

Proof. — Consider $U = F \backslash G$, $V = G \backslash F$. Consider a minimal matching $\mathscr{M}$ of F, and, for $a \in F$, denote its partner by $\theta(a)$. Set

$$H = \{a \in F \backslash U;\ \theta(a) \in U\}.$$

When we remove U from F, the points of H lose their partners. Set $H' = \{\theta(a); a \in H\}$. Thus $H' \subset U$. To find partners for the points of $V \cup H$ we consider a minimum matching of $V \cup H'$. This matching induces a matching $\mathscr{M}'$ of $V \cup H$, using the bijection θ of H and H'. The union of the trace of $\mathscr{M}$ on $F \backslash (U \cup H)$ and $\mathscr{M}'$ is almost a matching of G, although it could happen that there remains an unmatched point in $V \cup H$ and one in $F \backslash (U \cup H)$. The two points are then matched together (creating the term $\sqrt{d}$). The matching we have constructed witnesses that

$$\begin{aligned} L(G) &\leqslant L(F) + L(V \cup H') + \sqrt{d} \\ &\leqslant L(F) + L'(V \cup U) + \sqrt{d}. \end{aligned}$$

To see it, it suffices to use the triangle inequality, and to observe that the edges $[a, \theta(a)]$ for $a \in U'$ do disappear from $\mathscr{M}$ when U is removed. □

Here is a simple observation.

Lemma **11.5.2.** — *Consider subsets* $F_1, \ldots, F_p$ *of* $[0, 1]^d$. *Then*

$$L'(\bigcup_{i \leqslant p} F_i) \leqslant \sum_{i \leqslant p} L'(F_i) + Kp^{1-1/d}$$

where, as in the rest of this section, K *denotes a constant that depends on d only.*

Proof. — It suffices to prove this for L rather than L'. The point is that if one considers an optimal matching of each F_i, their union fails to be a matching of $\bigcup_{i \leqslant p} F_i$ only because there could remain an unmatched point in each F_i, while we are permitted at most a single unmatched point. Thus, it suffices to match all but at most one of these points, using for example a shortest tour through them, and matching consecutive points on the tour. □

It seems an interesting question whether when $d = 2$ the inequality of Theorem 11.2.3 would hold, at least for smaller values of t. Possibly easier is the question whether the variance of L_N is bounded. The best results in that direction belong to Rhee. She proved that if $d = 2$, $\operatorname{Var} L_N \leqslant K(\log N)^2$ [R3], while if $d \geqslant 3$, $\operatorname{Var} L_N \leqslant KN^{1-1/d}$ [R2]. The arguments for these results are different. Our methods do not allow to improve on the result for $d = 2$, but allow significant improvement when $d \geqslant 3$ (and this is why we consider this case in this section). Although this has not been checked, it seems to be an exercise to show that $\operatorname{Var} L_N \geqslant \frac{1}{K} N^{1-2/d}$ using e.g., the method of [R1]. What we will prove is that $\operatorname{Var} L_N \leqslant (\log N)^K N^{1-2/d}$. The proof goes by first proving a Poissonized version of the result, and then using " dePoissonization ". The second part of the argument is standard (see e.g. [R1]) and will not be given here.

The Poissonized version of the problem is the study of the r.v. $L_\lambda = L(\Pi_\lambda)$, where Π_λ is the random subset of $[0, 1]^d$ that is generated by a Poisson point process of constant intensity λ. We consider the space Ω of all finite subsets of $[0, 1]^d$, and on Ω, we consider the probability P_λ induced by Π_λ. On Ω^2, we consider the function

$$f(F, G) = L'(F \Delta G). \tag{11.5.1}$$

For a subset B of Ω, we set

$$f(\mathrm{F}, \mathrm{B}) = \inf_{\mathrm{G} \in \mathrm{B}} f(\mathrm{F}, \mathrm{G}) = \inf_{\mathrm{G} \in \mathrm{B}} \mathrm{L}'(\mathrm{F} \,\Delta \mathrm{G}). \tag{11.5.2}$$

We set $\gamma = \frac{1}{2} - \frac{1}{d}$.

Theorem **11.5.3.** — *For all* $\lambda \geqslant 3$ *and all subsets* B *of* Ω, *we have*

$$\int_\Omega \exp \frac{f(\mathrm{F}, \mathrm{B})}{(\log \lambda)^{\mathrm{K}} \lambda^\gamma} \, d\mathrm{P}_\lambda(\mathrm{F}) \leqslant \frac{e}{\mathrm{P}_\lambda(\mathrm{B})}.$$

If we combine this result with Lemma 11.5.1 (and proceed as usual) we see that if M_λ denotes a median of L_λ, we have

$$\int \exp \left(\frac{1}{(\log \lambda)^{\mathrm{K}} \lambda^\gamma} \mid \mathrm{L}_\lambda - \mathrm{M}_\lambda \mid \right) d\mathrm{P} \leqslant \mathrm{K}$$

which certainly implies the previous claim about the variance of L_λ. To prove Theorem 11.5.3, we will prove the following statement, whose form is adapted to proof by induction.

Proposition **11.5.4.** — *There exists numbers* K_0, $\alpha > 1$ *depending on d only, such that for all* $q > 0$ *we have, for all* λ, $1 \leqslant \lambda \leqslant 2^{\alpha^q}$ *and all Borel subsets* B *of* Ω,

$$\int_\Omega \exp \frac{f(\mathrm{F}, \mathrm{B})}{\mathrm{K}_0^q \lambda^\gamma} \, d\mathrm{P}_\lambda(\mathrm{F}) \leqslant \frac{e}{\mathrm{P}_\lambda(\mathrm{B})}.$$

To see that this statement implies Theorem 11.5.3, we take the smallest q such that $\lambda \leqslant 2^{\alpha^q}$, so that α^q is of order $\log \lambda$, and K_0^q of order $(\log \lambda)^{\mathrm{K}}$.

The proof of Proposition 11.5.4 is by induction over q. For the case $q = 1$, one uses the brutal bound

$$f(\mathrm{F}, \mathrm{G}) \leqslant \mathrm{K}(\mathrm{card}\, \mathrm{F} + \mathrm{card}\, \mathrm{G})$$

and the exponential integrability of Poisson random variables. The easy details are left to the reader.

We will determine, in due time, suitable values for K_0 and α and we now start the proof of the induction step from q to $q + 1$. Consider λ such that $2^{\alpha^q} \leqslant \lambda \leqslant 2^{\alpha^{q+1}}$. Consider the smallest integer n such that $\lambda' = \lambda/n^d \leqslant 2^{\alpha^q}$. (Thus, we can apply the induction hypothesis to λ'.) By definition of n, we have $\lambda/(n-1)^d \geqslant 2^{\alpha^q}$, so that, since $\lambda \leqslant 2^{\alpha^{q+1}}$, we have $(n-1)^d \leqslant 2^{(\alpha-1)\alpha^q}$, and thus

$$n^d \leqslant \frac{n^d}{(n-1)^d} 2^{(\alpha-1)\alpha^q} \leqslant 2^d . 2^{(\alpha-1)\alpha^q}. \tag{11.5.3}$$

Also,

(11.5.4) $$\lambda' = \frac{\lambda}{n^d} \geqslant \left(\frac{n-1}{n}\right)^d \frac{\lambda}{(n-1)^d} \geqslant \left(\frac{n-1}{n}\right)^d 2^{\alpha^q} \geqslant 2^{\alpha^q - d}.$$

Consider a partition of $[0, 1]^d$ in n^d congruent cubes $(C_i)_{i \leqslant n^d}$. From Lemma 11.5.2, we observe that

(11.5.5) $$L'(F \Delta G) \leqslant \sum_{i \leqslant n^d} L'((F \Delta G) \cap C_i) + Kn^{d-1}.$$

We set

$$f_i(F, G) = L'((F \Delta G) \cap C_i).$$

Thus we have, from (11.5.5)

$$L'(F \Delta G) \leqslant \sum_{i \leqslant n^d} f_i(F, G) + Kn^{d-1}.$$

If we set

$$g(F, G) = \inf_{G \in B} \sum_{i \leqslant n^d} f_i(F, G),$$

we get by (11.5.2) that

(11.5.6) $$f(F, B) \leqslant g(F, B) + Kn^{d-1}.$$

The crucial point is that (Ω, P_λ) is naturally isomorphic to the product of n^d copies of $(\Omega, P_{\lambda'})$. To see this, let us denote by R_i an affine map from C_i to $[0, 1]^d$, for $i \leqslant n^d$. Then the isomorphism simply associates $(R_i(F \cap C_i))_{i \leqslant n^d}$ to F. We observe that

$$f_i(F, G) = \frac{1}{n} L'(R_i(F \cap C_i) \Delta R_i(G \cap C_i))$$

so that, under this isomorphism, each function f_i is distributed like the function h' on Ω^2 (provided with $P_{\lambda'} \otimes P_{\lambda'}$), where $h'(F, G) = \frac{1}{n} L'(F \Delta G)$. Moreover, with the notation of Definition 2.4.1, we have $f_{h'} = g$. By induction hypothesis, and taking the scaling factor n into account, we have for each Borel set $B \subset \Omega$,

$$\int_\Omega \exp(2h(F, B))\, dP_{\lambda'}(F) \leqslant \frac{e}{P_{\lambda'}(B)},$$

where $h = ah'$, $a = n(2K_0^q \lambda'^{\gamma})^{-1}$. It then follows from Theorem 2.5.1 and the definition of g that

$$\forall\, t \leqslant 1, \quad \int_\Omega \exp(atg(F, B))\, dP_\lambda(F) \leqslant \frac{1}{P_\lambda(B)} \exp(3n^d t^2)$$

for each Borel set $B \subset \Omega$. From (11.5.6), it follows that

$$\int_\Omega \exp(atf(F, B))\, dP_\lambda(F) \leqslant \frac{1}{P_\lambda(B)} \exp(3n^d t^2 + Kn^{d-1} at).$$

We see that if

$$n^{d/2-1} a \leqslant K, \tag{11.5.7}$$

then, taking $t = n^{-d/2}/K$, we get

$$\int_\Omega \exp\left(\frac{a}{Kn^{d/2}} f(F, B)\right) dP_\lambda(F) \leqslant \frac{e}{P_\lambda(B)}. \tag{11.5.8}$$

Now,

$$\frac{a}{Kn^{d/2}} = \frac{1}{2KK_0^q \lambda'^\gamma n^{d/2-1}} = \frac{1}{2KK_0^q \lambda^\gamma}$$

since $d\gamma = d/2 - 1$. Thus, provided $K_0 = 2K$, (11.5.8) is exactly what we need to complete the induction.

It remains to check that (11.5.7) holds, but by (11.5.3), (11.5.4),

$$n^d \leqslant 2^d\, 2^{(\alpha-1)\alpha^q},$$

$$a/n \leqslant \lambda'^{-\gamma} \leqslant 2^{\gamma d}\, 2^{-\gamma\alpha^q},$$

so that (11.5.7) holds for $\alpha = 1 + 2\gamma$. □

12. The free energy in the Sherrington-Kirpatrick model at high temperature

Consider a sequence $(\varepsilon_i)_{i \leqslant N}$ with $\varepsilon_i \in \{-1, 1\}$. Each ε_i represents the two possible values of the spin of particule i. Consider numbers $(h_{ij})_{1 \leqslant i < j \leqslant N}$ that represent the interaction between spins. The energy of a given configuration is given by $\sum_{1 \leqslant i < j \leqslant N} h_{ij}\, \varepsilon_i\, \varepsilon_j$. Consider a parameter $\beta > 0$ (that plays the role of the inverse of the temperature). The so-called " partition function " is given by

$$Z_N = Z_N(h_{ij}) = 2^{-N} \sum_{(\varepsilon_i) \in \{-1, 1\}^N} \exp\left(\frac{\beta}{\sqrt{N}} \sum_{1 \leqslant i < j \leqslant N} h_{ij}\, \varepsilon_i\, \varepsilon_j\right). \tag{12.1}$$

The role of the factor $\sqrt{N}$ is for normalization purposes that will become apparent later.

If we think of ε_i as being a Bernoulli r.v., it is natural to write

$$Z_N(h_{ij}) = E_\varepsilon \exp\left(\frac{\beta}{\sqrt{N}} \sum_{1 \leqslant i < j \leqslant N} h_{ij}\, \varepsilon_i\, \varepsilon_j\right). \tag{12.2}$$

In the model we study, the numbers h_{ij} are random, and the sequence $(h_{ij})_{1 \leqslant i < j \leqslant N}$ is i.i.d. We assume $\mathrm{E}h_{ij} = \mathrm{E}h_{ij}^3 = 0$, and we assume for normalization purposes that $\mathrm{E}h_{ij}^2 = 1$. We will also assume that $\mathrm{E} \exp \alpha \mid h_{ij} \mid < \infty$ for α small enough. Then EZ_N is well-defined for N large enough. We are interested in the quantity $N^{-1} \mathrm{E} \log Z_N$ (mean free energy per site), whose study relies ultimately on the study of Z_N. It is proved in [A-L-R], and in [C-N] in the case where h_{ij} is gaussian, that for $\beta < 1$ the random variable $\log Z_N - \beta^2 N/4$ converges in law to a (non-standard) normal r.v. Equally interesting, but of a rather different nature is the research of tail estimates for $\log Z_N - \beta^2 N/4$ that are valid for all N.

Theorem **12.1.** — *There exists a universal constant* K *with the following property. Assume that* $\mathrm{E} \exp \pm h_{ij} < 2$. *Then, for* $0 < t < N/K$, $\beta < 1$,

$$\text{(12.3)} \qquad P\left(\left|\log Z_N - \frac{\beta^2 N}{4}\right| \geqslant K\left(t + \sqrt{\log \frac{K}{1-\beta^2}}\right)\sqrt{N}\right) \leqslant 2e^{-t^2}.$$

In particular

$$\text{(12.4)} \qquad -\frac{K}{\sqrt{N}}\sqrt{\log \frac{2}{1-\beta^2}} \leqslant \frac{1}{N} \mathrm{E} \log Z_N - \frac{\beta^2}{4} \leqslant \frac{K}{N}.$$

Comments. — 1) The reader might like to start with the significantly simpler gaussian case. In that case, the key deviation inequality (12.5) below can be replaced by

$$\forall\, t > 0, \quad P(|\log Z_N - M_N| \geqslant t) \leqslant 2 \exp\left(-\frac{t^2}{N-1}\right)$$

as a direct consequence of (1.6), (1.15). 2) In the condition $\mathrm{E} \exp \pm h_{ij} \leqslant 2$, the number 2 can be replaced by any other (with a different constant K). It seems reasonable to conjecture that (12.3) is not sharp in the gaussian case, and that, for a given $\beta < 1$,

$$\lim_{t \to \infty} \sup_N P\left(\left|\log Z_N - \frac{\beta^2 N}{4}\right| \geqslant t\right) = 0.$$

It should however be pointed out that (12.3) does not hold when the factor $\sqrt{N}$ is removed from (12.3). Indeed it would follow otherwise that for each n, $\sup_N \mathrm{E}(4Z_N/\beta^2 N)^n < \infty$, and it is pointed out in [A-L-R], p. 6, that this is not the case.

The key to Theorem 12.1 will be the following deviation inequality

$$\text{(12.5)} \qquad 0 < t \leqslant 4\sqrt{N}(N-1) \Rightarrow P(|\log Z_N - M_N| \geqslant t) \leqslant 2 \exp\left(\frac{-t^2}{32(N-1)}\right)$$

where M_N denotes a median of $\log Z_N$. We first show how to deduce this from Corollary 2.4.4. The second crucial step will then be to relate M_N and $\beta^2 N/4$ ($\approx \log \mathrm{EZ}_N$).

To prove (12.5), we observe that

$$\text{(12.6)} \qquad | \log Z_N(h_{ij}) - \log Z_N(h'_{ij}) | \leqslant \frac{\beta}{\sqrt{N}} \sum_{1 \leqslant i < j \leqslant N} | h_{ij} - h'_{ij} |$$

as follows from the fact that

$$\Big| \sum_{1 \leqslant i < j \leqslant N} a_{ij}\, \varepsilon_i\, \varepsilon_j \Big| \leqslant \sum_{1 \leqslant i < j \leqslant N} | a_{ij} |.$$

We now view $\log Z_N$ as a function on $\mathbf{R}^{N(N-1)/2}$. We wish to apply Corollary 2.4.4 in the case $\Omega = \mathbf{R}$, $h(x,y) = \frac{1}{4} | x - y |$, μ the law of h_{ij}. We note that (2.4.12) holds, since

$$\iint_{\mathbf{R}^2} \exp \frac{1}{4} | x - y |\, d\mu(x)\, d\mu(y) \leqslant \left(\int \exp \frac{1}{4} | x |\, d\mu(x) \right)^2$$
$$\leqslant (\mathrm{E} \exp | h_{ij} |)^{1/2}$$
$$\leqslant (\mathrm{E}(\exp h_{ij} + \exp - h_{ij}))^{1/2} \leqslant 2.$$

Consider now v and the set $A = \{ \log Z_N < v \}$. Combining (12.6) and (2.4.13) (used for $N(N-1)/2$ rather than N) we see that, for $u > v$, we have

$$u - v \leqslant 4\beta \sqrt{N}(N-1)$$
$$\Rightarrow P(\{ \log Z_N > u \})\, P(\{ \log Z_N < v \}) \leqslant \exp\left(- \frac{(u-v)^2}{32\beta^2(N-1)} \right).$$

Taking successively $u = M_N$ and $v = M_N$, (12.5) follows as usual.

In order to relate M_N and $\beta^2 N/4$, the key step is the elementary estimates

$$\text{(12.7)} \qquad \frac{1}{K} \exp \frac{\beta^2 N}{4} \leqslant \mathrm{E}Z_N \leqslant K \exp \frac{\beta^2 N}{4},$$

$$\text{(12.8)} \qquad \mathrm{E}Z_N^2 \leqslant \frac{K}{1 - \beta^2} (\mathrm{E}Z_N)^2.$$

These will be proved later. First, we conclude the main argument. Consider the set $A = \left\{ Z_N \geqslant \frac{1}{2} \mathrm{E}Z_N \right\}$. Then

$$\mathrm{E}Z_N = \mathrm{E}(Z_N 1_{A^c}) + \mathrm{E}(Z_N 1_A)$$
$$\leqslant \frac{1}{2} \mathrm{E}Z_N + \mathrm{E}(Z_N^2)^{1/2} P(A)^{1/2},$$

so that

$$P(A) \geqslant \frac{1}{4} \frac{(\mathrm{E}Z_N)^2}{\mathrm{E}Z_N^2}$$

(a fact going back to Paley and Zigmund). Combining with (12.8), we get $P(A) \geqslant (1 - \beta^2)/K$. To get a lower bound for M_N, we can assume $M_N \leqslant \log \frac{1}{2} \mathrm{E}Z_N$.

We set $t = \log\left(\frac{1}{2}\,\mathrm{EZ_N}\right) - \mathrm{M_N}$. Since $\log \mathrm{Z_N} \geqslant 0$, we have $\mathrm{M_N} \geqslant 0$ and hence $t \leqslant \mathrm{K} + \mathrm{N}/4$, by (12.7).

We certainly have

$$\mathrm{A} \subset \{\log \mathrm{Z_N} \geqslant \mathrm{M_N} + t\}.$$

Thus, by (12.5), we have

$$\frac{(1-\beta^2)}{\mathrm{K}} \leqslant \mathrm{P(A)} \leqslant 2\exp\left(-\frac{t^2}{32(\mathrm{N}-1)}\right),$$

so that

$$t \leqslant \mathrm{K}\sqrt{\mathrm{N}}\left(\log\frac{\mathrm{K}}{1-\beta^2}\right)^{1/2}$$

and thus

$$\mathrm{M_N} \geqslant \log\left(\frac{1}{2}\,\mathrm{EZ_N}\right) - \mathrm{K}\sqrt{\mathrm{N}}\left(\log\frac{\mathrm{K}}{1-\beta^2}\right)^{1/2}.$$

We also have $\mathrm{M_N} \leqslant \log(2\mathrm{EZ_N})$. Combining with (12.7) we get

$$\left|\mathrm{M_N} - \frac{\beta^2\,\mathrm{N}}{4}\right| \leqslant \mathrm{K}\sqrt{\mathrm{N}}\left(\log\frac{\mathrm{K}}{1-\beta^2}\right)^{1/2}$$

so that (12.3) now follows from (12.5).

To prove (12.4), we first observe that the lower bound follows from (12.3) and a routine computation. The upper bound follows from the concavity of log, which implies $\mathrm{E}\log \mathrm{Z_N} \leqslant \log \mathrm{EZ_N}$, and (12.7).

It remains to prove (12.7), (12.8). We start with the elementary inequality

$$\left| e^x - 1 - x - \frac{x^2}{2} - \frac{x^3}{3!}\right| \leqslant \frac{x^4}{4!}\,e^{|x|}$$

that is obvious on power series expansions. Thus, for $|u| \leqslant \frac{1}{2}$, we have (since $\mathrm{E}h_{ij}^2 = 1$, $\mathrm{E}h_{ij} = \mathrm{E}h_{ij}^3 = 0$)

$$\textbf{(12.9)} \qquad 1 + \frac{u^2}{2} - \mathrm{K}u^4 \leqslant \mathrm{E}\exp(uh_{ij}) \leqslant 1 + \frac{u^2}{2} + \mathrm{K}u^4,$$

and for $\varepsilon = \pm 1$, $\beta \leqslant 1$,

$$\exp\left(\frac{\beta^2}{2\mathrm{N}} - \frac{\mathrm{K}\beta^4}{\mathrm{N}^2}\right) \leqslant \mathrm{E}\exp\frac{\varepsilon\beta h_{ij}}{\sqrt{\mathrm{N}}} \leqslant \exp\left(\frac{\beta^2}{2\mathrm{N}} - \frac{\mathrm{K}\beta^4}{\mathrm{N}^2}\right).$$

Since

$$\mathrm{EZ_N} = \mathrm{E}_\varepsilon \prod_{ij} \mathrm{E}\exp \varepsilon_i\,\varepsilon_j\,\frac{\beta h_{ij}}{\sqrt{\mathrm{N}}},$$

(12.7) follows. Turning to the study of EZ_N^2, we have, using (12.9), and for $N \geq 8$, that, with obvious notation,

$$EZ_N^2 = EE_\varepsilon E_{\varepsilon'} \exp\Big(\sum_{1 \leq i < j \leq N} \frac{\beta h_{ij}}{\sqrt{N}} (\varepsilon_i \varepsilon_j + \varepsilon'_i \varepsilon'_j)\Big)$$

$$\leq KE_\varepsilon E_{\varepsilon'} \exp\Big(\sum_{1 \leq i < j \leq N} \frac{\beta^2}{2N} (\varepsilon_i \varepsilon_j + \varepsilon'_i \varepsilon'_j)^2\Big).$$

Now, $(\varepsilon_i \varepsilon_j + \varepsilon'_i \varepsilon'_j)^2 = 2 + 2\varepsilon_i \varepsilon_j \varepsilon'_i \varepsilon'_j$. Also, $\varepsilon_i \varepsilon_j \varepsilon'_i \varepsilon'_j$ is distributed like $\varepsilon_i \varepsilon_j$, so that

$$EZ_N^2 \leq K \exp\Big(\frac{\beta^2 N}{2}\Big) E_\varepsilon \Big(\frac{\beta^2}{2N} \big(\sum_{1 \leq i < j \leq N} 2\varepsilon_i \varepsilon_j\big)\Big).$$

Further,

$$\sum_{1 \leq i < j \leq N} 2\varepsilon_i \varepsilon_j = \big(\sum_{1 \leq i \leq N} \varepsilon_i\big)^2 - N.$$

Using the subgaussian inequality

$$P_\varepsilon\big(\big|\sum_{i=1}^{N} \varepsilon_i\big| \geq t\big) \leq 2 \exp\Big(-\frac{t^2}{2N}\Big),$$

we have

$$E_\varepsilon \exp\Big(\frac{\beta^2}{2N} \big(\sum_{i=1}^{N} \varepsilon_i\big)^2\Big) \leq 1 + 2\int_0^\infty \frac{d}{dt}\Big(\exp\frac{\beta^2 t^2}{2N}\Big) \exp\Big(-\frac{t^2}{2N}\Big)\, dt$$

$$= \frac{1 + \beta^2}{1 - \beta^2}$$

and (12.8) follows. □

13. Sums of (vector valued) independent random variables

The first objective of this section is to discuss the genesis of the key ideas of the isoperimetric approach as developed in the present paper, and to explain how these ideas have permitted the solution of the main problems of Probability in Banach spaces. In the second part of this section we will discuss, in detail, a situation that parallels the situation of Chapter 8, but where the infimum over $\alpha \in \mathscr{F}$ is replaced by a supremum. There are unexpected and subtle differences; this is closely connected to the fact that the conditions on the function $h(x, y)$ in Theorem 4.4.1 are (and must be) highly disymmetric in x and y.

Consider a sequence $(X_i)_{i \leq N}$ of r.v. valued in a Banach space W. A number of classical problems of probability (in particular, laws of large numbers and laws of the iterated logarithm) depend crucially on sharp estimates of the tail probability

$P(\|\sum_{i\leq N} X_i\| \geq t)$. For many years these estimates were found using martingales, and the results were not optimal. One big obstacle is that there is no obvious substitute for the positivity arguments that are central to Chapter 8. Although its importance became clear only later, a crucial contribution was made by M. Ledoux [L]. It was known at the time that in many situations, the tails of $\|\sum_{i\leq N} X_i\|$ resemble the tails of $\|\sum_{i\leq N} g_i X_i\|$, where $(g_i)_{i\leq N}$ is an independent sequence of standard normal r.v. that is independent of the sequence X_i. To study $\|\sum_{i\leq N} g_i X_i\|$, Ledoux wrote

$$\|\sum_{i\leq N} g_i X_i\| = E_g \|\sum_{i\leq N} g_i X_i\| + (\|\sum_{i\leq N} g_i X_i\| - E_g \|\sum_{i\leq N} g_i X_i\|), \tag{13.1}$$

where E_g denotes conditional expectation, given $(X_i)_{i\leq N}$. The idea was that either term of the right-hand side should be easier to study than the term of the left-hand side. This is particularly apparent for the second term, where, arguing conditionally on X_i, one can take advantage of the properties of Gaussian processes.

It turns out that the first term in the right of (13.7) has the exact property needed to replace positivity; namely, if $J \subset \{1, \ldots, N\}$, we have

$$E_g \|\sum_{i\in J} g_i X_i\| \leq E_g \|\sum_{i\leq N} g_i X_i\|. \tag{13.2}$$

The realization of the importance of positivity-like properties led first to the characterization of the Banach-space valued r.v. that satisfy the law of the iterated logarithm [L-T1]. Perhaps more importantly, (13.2) lead this author to the belief that some isoperimetric principle should be relevant, and hence to the theorem of [T2] (that is now superceeded by the comparable, but much easier to prove Theorem 3.1.1), and started the line of investigation that culminates in the present paper.

The author also understood that Bernoulli r.v. have regularity properties that almost match those of Gaussian r.v. (a crucial step is the comparison theorem of [T5]). They offer the extra advantage that the tails of $\|\sum_{i\leq N} \varepsilon_i X_i\|$ (where $P(\varepsilon_i = -1) = P(\varepsilon_i = 1) = 1/2$) always resemble the tails of $\|\sum_{i\leq N} X_i\|$. Thus, rather than (13.2) one should write

$$\|\sum_{i\leq N} \varepsilon_i X_i\| = E_\varepsilon \|\sum_{i\leq N} \varepsilon_i X_i\| + (\|\sum_{i\leq N} \varepsilon_i X_i\| - E_\varepsilon \|\sum_{i\leq N} \varepsilon_i X_i\|). \tag{13.3}$$

Our first task is the study of the r.v.

$$Z = E_\varepsilon \|\sum_{i\leq N} \varepsilon_i X_i\|. \tag{13.4}$$

We denote by $\|X_i\|^*$ the non-decreasing rearrangement of the sequence $(\|X_i\|)_{i\leq N}$. Thus

$$\|X_i\|^* = \sup\{t; \operatorname{card}\{j \leq N; \|X_j\| \geq t\} \geq i\}.$$

Proposition **13.1.** — *Consider* $a > 0$, $q, k \in \mathbf{N}$. *Then*

$$\text{(13.5)} \qquad P(Z \geqslant qa + t) \leqslant \frac{1}{q^{k+1}\, P(Z \leqslant a)^q} + P(\sum_{i \leqslant k} \| X_i \|^* \geqslant t).$$

Comment. — To obtain a useful bound, one estimates the last term using classical methods; one then optimizes over k, q.

Proof. — We set $\Omega = W$. Consider the function Z on Ω^N given for $x = (x_i)_{i \leqslant N} \in \Omega^N$ by

$$Z(x) = E_\varepsilon \| \sum_{i \leqslant n} \varepsilon_i x_i \|.$$

Consider the product probability P on Ω, when the i-th factor is provided with the law of X_i. Consider the set $A = \{ x ; Z(x) \leqslant a \}$. Setting, for simplicity

$$k(x) = f(A, \ldots, A, x)$$

(where A occurs q times), it suffices by Theorem 3.1.1 to prove that

$$\text{(13.6)} \qquad Z(x) \leqslant qa + \sum_{i \leqslant k(x)} \| x_i \|^*.$$

Indeed, we then have, for each k,

$$P(Z \geqslant qa + t) \leqslant P(k(x) > k) + P(\sum_{i \leqslant k} \| x_i \|^* \geqslant t).$$

To prove (13.6), we consider $y^1, \ldots, y^q$ in A such that if we set

$$I = \{ i \leqslant N ; x_i \notin \{ y_i^1, \ldots, y_i^q \} \},$$

then card $I \leqslant k(x)$. Denote by J the complement of I. Then, by the triangle inequality

$$\text{(13.7)} \qquad Z(x) = E_\varepsilon \| \sum_{i \leqslant N} \varepsilon_i x_i \| \leqslant E_\varepsilon \| \sum_{i \in J} \varepsilon_i x_i \| + \sum_{i \in I} \| x_i \|.$$

Now

$$\text{(13.8)} \qquad \sum_{i \in I} \| x_i \| \leqslant \sum_{i \leqslant k(x)} \| x_i \|^*,$$

since these last $k(x)$ terms are the $k(x)$ largest terms of the sequence $(\| x_i \|)_{i \leqslant n}$. By definition of J, we can find a partition $J_1, \ldots, J_q$ of this set such that, for $\ell \leqslant q$,

$$\forall\, i \in J_\ell, \quad x_i = y_i^\ell.$$

Thus

$$\text{(13.9)} \qquad E_\varepsilon \| \sum_{i \in J_\ell} \varepsilon_i x_i \| = E_\varepsilon \| \sum_{i \in J_\ell} \varepsilon_i y_i^\ell \|.$$

The essential fact is now that

$$\mathrm{E}_\varepsilon \| \sum_{i \in \mathrm{J}_\ell} \varepsilon_i y_i \| \leqslant \mathrm{E}_\varepsilon \| \sum_{i \leqslant \mathrm{N}} \varepsilon_i y_i^\ell \|.$$

To see this, simply observe that, in the left-hand side, the expectation in ε_i, $i \notin \mathrm{J}_q$ is taken inside rather than outside the norm. Since $y^\ell \in \mathrm{A}$, combining with (13.9) we get

$$\mathrm{E}_\varepsilon \| \sum_{i \in \mathrm{J}_\ell} \varepsilon_i x_i \| \leqslant a$$

and thus, by the triangle inequality

$$\mathrm{E}_\varepsilon \| \sum_{i \in \mathrm{J}} \varepsilon_i x_i \| \leqslant qa.$$

Combining with (13.7), (13.8) yield the result. □

To study the last term of (13.3) conditionally on $(\mathrm{X}_i)_{i \leqslant \mathrm{N}}$, one can rely, in particular, upon the following result.

Theorem **13.2.** — *Consider vectors* $(v_i)_{1 \leqslant i \leqslant \mathrm{N}}$ *in a Banach space* W, *and set*

(13.10) $$\sigma = (\sup\{ \sum_{i \leqslant \mathrm{N}} w^*(v_i)^2 : w^* \in \mathrm{W}^*, \| w^* \| \leqslant 1 \})^{1/2}.$$

Consider a sequence $(\mathrm{Y}_i)_{i \leqslant \mathrm{N}}$ *of independent real valued r.v. such that* $| \mathrm{Y}_i | \leqslant 1$. *Denote by* M *a median of the r.v.* $\| \sum_{i \leqslant \mathrm{N}} \mathrm{Y}_i v_i \|$. *Then for* $t > 0$ *we have*

(13.11) $$\mathrm{P}(| \| \sum_{i \leqslant \mathrm{N}} \mathrm{Y}_i v_i \| - \mathrm{M} | \geqslant t\sigma) \leqslant 4 \exp\left(-\frac{t^2}{16}\right).$$

Proof. — We observe that if we set

$$\mathscr{F} = \{ (w^*(v_i)) ; w^* \in \mathrm{W}^*, \| w^* \| \leqslant 1 \}$$

then

$$\mathrm{Z} = \| \sum_{i \leqslant \mathrm{N}} \mathrm{Y}_i v_i \| = \sup_{\alpha \in \mathscr{F}} \sum_{i \leqslant \mathrm{N}} \alpha_i \mathrm{Y}_i.$$

Thus Theorem 13.2 is a special case of Theorem 8.1.1 (using scaling). □

Remarks. — Certainly the constant in the exponent is not sharp, and could be improved using (4.2.7) rather than (4.1.3), especially in the case of Bernoulli r.v., where the use of (4.3.8) would yield a bound of $2 \exp -\frac{1}{4}(t - \sqrt{\log 2})^2$ for $t \geqslant \sqrt{\log 2}$.

Before we pursue the study of (13.3), we digress on an interesting sharpening of Theorem 13.2. There is another bound on the tails of $\| \sum_{i \in \mathrm{N}} \mathrm{Y}_i v_i \|$ namely the trivial bound $\| \sum_{i \leqslant \mathrm{N}} \mathrm{Y}_i v_i \| \leqslant \sup_{\| w^* \| \leqslant 1} \sum_{i \leqslant \mathrm{N}} | w^*(v_i) |$, and it is, of course, possible to interpolate between this bound and (13.11). This can be done as follows. For a sequence $(r_i)_{i\leq N}$ of real numbers, and $t > 0$, we write

$$K_{1,2}((r_i), t) = \inf\{ \sum_{i\leq N} | u_i | + t(\sum_{i\leq N} w_i^2)^{1/2};\ r_i = u_i + w_i \},$$

where the infimum is taken over all possible decompositions $r_i = u_i + w_i$. We set

$$\kappa(t) = \sup\{ K_{1,2}((w^*(v_i)), t);\ w^* \in W^*, \| w^* \| = 1 \}.$$

We observe that $\kappa(t) \leqslant t\sigma$. Only rather trivial modifications to the proof of Theorem 8.1.1 are needed to see that one can improve (13.11) into

$$\text{(13.12)} \qquad P(| \| \sum_{i\leq N} Y_i v_i \| - M | \geqslant \kappa(t)) \leqslant 4 \exp\left(-\frac{t^2}{16}\right).$$

This inequality streamlines a result of [D-MS].

If one observes that $\kappa(2t) \leqslant 2\kappa(t)$ one obtains, through a routine computation, that for all $p \geqslant 1$,

$$\| \sum_{i\leq N} Y_i v_i \|_p \leqslant M + K\kappa(\sqrt{p}),$$

a rather precise form of the so-called Kintchin-Kahane inequalities. It should also be pointed out that, by (13.12), $\| \sum_{i\leq N} Y_i v_i \|_p \geqslant M - K\kappa(\sqrt{p})$, that $\| \sum_{i\leq N} Y_i v_i \|_p \geqslant M2^{-1/p}$ (obviously) and that $\| \sum_{i\leq N} Y_i v_i \|_p \geqslant K\kappa(\sqrt{p}/K)$. To prove this last inequality, one reduces to the real-valued case; it is simple to see that this follows from [L-T1], Lemma 4.9 (see also [M-S]).

After this digression, we go back to the study of (13.3). We will apply Theorem 13.2 to the last term, conditionally on $(X_i)_{i\leq N}$, for $v_i = X_i$, $y_i = \varepsilon_i$. Thus we need control of the random quantity $\sigma(X)$ given by

$$\sigma^2(X) = \sup\{ \sum_{i\leq N} w^*(X_i)^2;\ w^* \in W^*, \| w^* \| \leqslant 1 \}.$$

Let us define $\Omega = W$, P as in Proposition 13.1, and consider the function σ on Ω^N given by

$$\sigma^2(x) = \sup\{ \sum_{i\leq N} w^*(x_i)^2;\ w^* \in W^*, \| w^* \| \leqslant 1 \}.$$

The basic idea is to control σ through the use of Theorem 3.1 again. Consider $b > 0$, and the set $B = \{ \sigma \leqslant b \}$. We set

$$k(x) = f(B, \ldots, B, x)$$

where B occurs q times.

Proposition **13.3.** —

$$\text{(13.13)} \qquad P_\varepsilon(\| \sum_{i\leq N} \varepsilon_i x_i \| \geqslant 2E_\varepsilon \| \sum_{i\leq N} \varepsilon_i x_i \| + u + \sum_{i\leqslant k(x)} \| x_i \|^*) \leqslant 4 \exp\left(-\frac{u^2}{16qb^2}\right).$$

Proof. — Consider $y^1, \ldots, y^q$ in B such that card $\mathrm{I} \leqslant k(x)$, where

$$\mathrm{I} = \{ i \leqslant \mathrm{N};\ x_i \notin \{y_i^1, \ldots, y_i^q\}\}.$$

Denote by J the complement of I. We have little control over the elements x_i, $i \in \mathrm{I}$, so we write

$$\begin{aligned}\| \sum_{i \leqslant \mathrm{N}} \varepsilon_i x_i \| &\leqslant \| \sum_{i \in \mathrm{J}} \varepsilon_i x_i \| + \sum_{i \in \mathrm{I}} \| x_i \| \\ &\leqslant \| \sum_{i \in \mathrm{J}} \varepsilon_i x_i \| + \sum_{i \leqslant k(x)} \| x_i \|^*.\end{aligned}$$

Now, it should be clear that

$$\sup \{ \sum_{i \in \mathrm{J}} w^*(x_i)^2;\ w^* \in \mathrm{W}^*, \| w^* \| \leqslant 1 \} \leqslant qb^2.$$

Denoting by M the median of $\| \sum_{i \in \mathrm{J}} \varepsilon_i x_i \|$, Theorem 13.2 gives

$$\mathrm{P}_\varepsilon(\| \sum_{i \in \mathrm{J}} \varepsilon_i x_i \| \geqslant \mathrm{M} + u) \leqslant 4 \exp - \frac{u^2}{16qb^2}.$$

Now

$$\mathrm{M} \leqslant 2\mathrm{E}_\varepsilon \| \sum_{i \in \mathrm{J}} \varepsilon_i x_i \| \leqslant 2\mathrm{E}_\varepsilon \| \sum_{i \leqslant \mathrm{N}} \varepsilon_i x_i \|$$

and (13.13) follows. □

If we combine Proposition 13.3 with Theorem 3.1.1, we have the following relation, for any $u, v, t > 0$, and any $k \in \mathbf{N}$:

(13.14)
$$\begin{aligned}&\mathrm{P}'(\| \sum_{i \leqslant \mathrm{N}} \varepsilon_i x_i \| \geqslant 2v + u + t) \\ &\leqslant \frac{1}{q^{k+1}\,\mathrm{P(B)}^q} + \mathrm{P}(\mathrm{E}_\varepsilon \| \textstyle\sum_{i \leqslant \mathrm{N}} \varepsilon_i x_i \| \geqslant v) + 4 \exp\left(- \dfrac{u^2}{16qb^2}\right) \\ &\quad + \mathrm{P}(\sum_{i \leqslant k} \| x_i \|^* > t),\end{aligned}$$

where P′ refers to the fact that we now consider the joint probability in (ε_i) and (x_i).

Theorem **13.4.** — *Set*

$$a = 2\mathrm{E} \| \sum_{i \leqslant \mathrm{N}} \varepsilon_i \mathrm{X}_i \|, \quad s^2 = \mathrm{E} \sup \{ \sum_{i \leqslant \mathrm{N}} w^*(\mathrm{X}_i)^2;\ w^* \in \mathrm{W}^*, \| w^* \| \leqslant 1 \}.$$

Then, for $q, k \in \mathbf{N}$, $u, t > 0$ *we have*

$$\begin{aligned}&\mathrm{P}(\| \sum_{i \leqslant \mathrm{N}} \varepsilon_i \mathrm{X}_i \| \geqslant 2aq + 3t + u) \\ &\leqslant \frac{2^{q+1}}{q^{k+1}} + 4 \exp\left(- \frac{u^2}{32qs^2}\right) + 2\mathrm{P}(\sum_{i \leqslant k} \| \mathrm{X}_i \|^* > t).\end{aligned}$$

Proof. — We use (13.14) with $v = aq + t$ and $b^2 = 2s^2$, so that $P(B) \geqslant 1/2$. We then control the term $P(E_\varepsilon \| \sum_{i \leqslant N} \varepsilon_i x_i \| \geqslant v)$ using Proposition 13.1. □

A slightly more general bound (that allows truncation of the variables X_i) is proved in [L-T2] Theorem 6.17, and (when combined with techniques to control the quantity b above) is at the basis of numerous results. An alternative approach, that relies rather on Theorem 4.2.4 is developed in [T3]. The bounds developed by isoperimetric methods are sharp in most situations (see however [Ro] for a beautiful example where other ingredients are needed).

We now turn to a more specialized topic and we continue the investigation of r.v. of the type $Z = \sup_{\alpha \in \mathscr{F}} \sum_{i \leqslant N} \alpha_i X_i$ that was started in Chapter 8. In order to apply Corollary 8.2.2, we need to have (4.4.6), where $h(x, y) = | x - y |$, or, if α_i is always positive, $h(x, y) = (x - y)^+$. When the variable X_1 is positive (i.e., its law μ is supported by $\mathbf{R}^+$), inspection of Theorem 4.4.1 shows that (whatever choice of θ, ξ) no integrability condition on X_1 except boundedness, will insure that the conditions of this theorem hold for this choice of h. We will now give an example that shows that this is not an artifact of our approach. We will show that (13.11) cannot be essentially improved, even if $P(| X_i | \neq 0)$ is arbitrary small. This implies (by scaling) that, given any *finite* function $\Phi : \mathbf{R}^+ \to \mathbf{R}^+$, with $\Phi(0) = 0$, one can find a real r.v. X_1 with $\int \Phi(| X_1 |) \, d\lambda \leqslant 1$, and vectors $(v_i)_{i \leqslant N}$ such that (13.11) is violated.

Example **13.5.** — This example is essentially a re-interpretation of the example presented at the end of Section 4.3. Consider an independent sequence $(X_i)_{i \leqslant N}$ of Bernoulli variables such that $P(X_i = 1) = p$ is small. Consider the family $\mathscr{F}$ of N-tuples of form $\alpha_i = 1/\sqrt{2pN}$ if $i \in I$, $\alpha_i = 0$ otherwise, where I varies over all subsets of $\{1, \ldots, N\}$ of cardinality $\leqslant 2pN$. Then $\sigma = 1$. Consider

$$Z = \sup_{\alpha \in \mathscr{F}} \sum_{i \leqslant N} \alpha_i X_i = \frac{1}{\sqrt{2Np}} \sup \{ \sum_{i \in I} X_i ; \text{ card } I \leqslant 2Np \}.$$

We can also view Z as $\| \sum_{i \leqslant N} X_i e_i \|_{\mathscr{F}}$, where e_i is the canonical basis of $\mathbf{R}^N$, and where the norm $\| . \|_{\mathscr{F}}$ is given by $\| x \|_{\mathscr{F}} = \sup_{\alpha \in \mathscr{F}} \sum_{i \leqslant N} \alpha_i | x_i |$.

The main observation is that

$$\sum_{i \leqslant N} X_i \leqslant 2Np \Rightarrow Z = \frac{1}{\sqrt{2Np}} \sum_{i \leqslant N} X_i .$$

Since the probability of the event on the left goes to 1, as $N \to \infty$, the r.v. Z is asymptotically normal, of mean $\sqrt{Np/2}$ and variance $\sqrt{(1 - p)/2}$; so its deviation from its median does not decay faster than $\exp - Kt^2$.

The conclusion to be drawn from Example 13.3 is that, in order to extend Theorem 13.2 to the case where X_1 is unbounded, we must require conditions of a different nature than integrability.

Theorem **13.6.** — *There exists a universal constant* L *with the following property. Consider a convex function* ψ *on* $\mathbf{R}^+$ *such that* $\psi(x) \leqslant x^2$ *if* $x \leqslant 1$ *and* $\psi(x) \geqslant x$ *if* $x \geqslant 1$. *Consider a probability measure* μ *on* $\mathbf{R}$. *Assume the following*:

$$\textbf{(13.13)} \qquad \forall\, t > 0, \quad \mu(\{x;\ |x| \geqslant t\}) \leqslant 2\exp(-\mathrm{L}\psi(2t)).$$

Given any subset B *of* $\mathbf{R}$, *with* $\mu(\mathrm{B}) \geqslant 1/2$, *and any* $t \geqslant 1$, *we have*

$$\textbf{(13.14)} \qquad \mu(\{x;\ \psi(\inf_{y \in \mathrm{B}} |x-y|) \geqslant t\}) \leqslant e^{-t}(1-\mu(\mathrm{B})).$$

Consider independent real valued r.v. $(\mathrm{X}_i)_{i \leqslant \mathrm{N}}$ *distributed like* μ, *and vectors* $(v_i)_{i \leqslant \mathrm{N}}$ *in a Banach space* W.

Then, for all $t > 0$, *we have*

$$\textbf{(13.15)} \qquad \mathrm{P}(|\|\sum_{i \leqslant \mathrm{N}} \mathrm{X}_i v_i\| - \mathrm{M}| \geqslant t) \leqslant 2\exp\left(-\frac{1}{\mathrm{L}}\Psi_{\mathscr{F}}(t)\right),$$

where M *is a median of* $\|\sum_{i \leqslant \mathrm{N}} \mathrm{X}_i v_i\|$, *where*

$$\mathscr{F} = \{(w^*(v_i))_{i \leqslant \mathrm{N}};\ w^* \in \mathrm{W}^*,\ \|w^*\| \leqslant 1\}$$

and where $\Psi_{\mathscr{F}}$ *is defined in Section* 8.2.

Proof. — According to Corollary 8.2.2, it suffices to prove that the hypothesis of Theorem 4.4.1 holds when $h(x,y) = |x-y|$, in the case $\theta(x) = -\log x$, $w(x) = -\frac{1}{2}\log x$ (so that $\mathrm{H}(\xi, w)$ holds by Proposition 2.6.1). Only (4.4.4) has to be checked, since (13.14) is a rewriting of (4.4.5).

Consider $\mathrm{B} \subset \mathbf{R}$ with $\mu(\mathrm{B}) \leqslant 1/2$. Set

$$\textbf{(13.16)} \qquad s = \inf\{|y|;\ y \in \mathrm{B}\}.$$

Clearly, $h(x, \mathrm{B}) \leqslant |x| + s$. Thus, by convexity of ψ we have

$$\int_{\mathbf{R}} \exp\psi(h(x,\mathrm{B}))\,d\mu(x) \leqslant \exp\frac{1}{2}\psi(2s)\int_{\mathbf{R}}\exp\frac{1}{2}\psi(2x)\,d\mu(x).$$

On the other hand, by (13.16) we have $\mathrm{B} \cap\,]-s, s[\, = \emptyset$, so that $\mathrm{B} \subset \{x;\ |x| \geqslant s\}$ and hence by (13.13) we have $\exp\psi(2s) \leqslant (2/\mu(\mathrm{B}))^{1/\mathrm{L}}$. Thus it suffices to show that for L large enough we have

$$x \leqslant \frac{1}{2} \Rightarrow \mathrm{I}\left(\frac{2}{x}\right)^{1/2\mathrm{L}} \leqslant \frac{1}{\sqrt{x}}\ (= \exp - w(x)),$$

where $\mathrm{I} = \int_{\mathbf{R}} \exp\frac{1}{2}\psi(2x)\,d\mu(x)$. It remains to show that, under (13.13), $\lim_{\mathrm{L}\to\infty} \mathrm{I} = 0$, uniformly in ψ, an easy exercise left to the reader. □

Theorem 13.6 can be applied to the case where μ is a measure ν_ψ of the type considered in Proposition 2.7.4, although in that case the simpler Theorem 2.7.1 will yield the same conclusion. There are however, situations covered by Theorem 13.6 that are not covered by Theorem 2.7.1, because in (13.14) we require only $t \geqslant 1$. In particular, if the law of X satisfies (13.14), and if $\| Z \|_\infty \leqslant 1$, the law of $\frac{1}{3}(X + Z)$ satisfies (13.14) (it is *not* required that Z be independent of X). (The corresponding statement for (13.13) is also true, under mild conditions on ψ, replacing if needed 1/3 by a smaller number.)

In conclusion of this section, we want to discuss a question that apparently is not fully clarified by the results of the present paper. Consider numbers $(a_i)_{i \leqslant N}$, and vectors $(v_i)_{i \leqslant N}$ in a Banach space. Of which order are the fluctuations of the r.v. $Z = \| \sum_{i \leqslant N} a_{\rho(i)} v_i \|$ around its median M, when ρ is seen as a random element of the symmetric group S_N, provided with the uniform probability P?

Proposition **13.7.** — *a) Assuming* $|a_i| \leqslant 1$ *for each* i, *we have*

$$t \geqslant 0 \Rightarrow P(|Z - M| \geqslant t) \leqslant 4 \exp - \frac{t^2}{16\sigma^2} \tag{13.17}$$

where as usual

$$\sigma^2 = \sup\{ \sum_{i \leqslant N} w^*(v_i)^2;\ w^* \in W^*,\ \| w^* \| \leqslant 1 \}.$$

b) Assuming $\| v_i \| \leqslant 1$ *for each* i, *we have*

$$t \geqslant 0 \Rightarrow P(|Z - M| \geqslant t) \leqslant 4 \exp - \frac{t^2}{16 \sum_{i \leqslant N} a_i^2}. \tag{13.18}$$

Remark. — A first problem is to find a bound that contains simultaneously (13.17) and (13.18).

Proof. — The proof follows that of Theorem 8.1.1, using now Theorem 5.1 rather than Theorem 4.1.1. Thus, we indicate only the key points.

To prove *a)*, one notes that if $\rho, \tau \in S_N$ and $I = \{ i \leqslant N;\ \rho(i) \neq \tau(i)\}$, then

$$\Big| \sum_{i \leqslant N} w^*(v_i)\, a_{\rho(i)} - \sum_{i \leqslant N} w^*(v_i)\, a_{\tau(i)} \Big| \leqslant \sum_{i \in I} | w^*(v_i) |.$$

To prove *b)*, one observes that Z has the same distribution as $\| \sum_{i \leqslant N} a_i v_{\rho(i)} \|$, and, with the notation above, one now has

$$\Big| \sum_{i \leqslant N} w^*(v_{\rho(i)})\, a_i - \sum_{i \leqslant N} w^*(v_{\tau(i)})\, a_i \Big| \leqslant \sum_{i \in I} | a_i |. \quad \square$$

It should be pointed out that it seems likely that a phenomenon similar to that of Example 13.5 occurs in case *a*, and that (13.17) cannot be improved even if a large majority of the numbers a_i are equal to zero.

Note added in proof. — After this work was completed, several new extensions of theorem 4.1.1 have emerged, with applications in particular to statistics [T7].

REFERENCES

[A-L-R] M. Aizenman, J. L. Lebowitz, D. Ruelle, Some rigorous results on the Sherrington-Kirkpatrick spin glass model, *Commun. Math. Phys.* **112** (1987), 3-20.

[A-S] N. Alon, J. Spencer, *The Probabilistic Method*, Wiley, 1991.

[A-M] D. Amir, V. D. Milman, Unconditional and symmetric sets in n-dimensional normed spaces, *Israel J. Math.* **37** (1980), 3-20.

[B1] B. Bollobás, The chromatic number of random graphs, *Combinatorica* **8** (1988), 49-55.

[B2] B. Bollabás, Random graphs revisited, *Proceedings of Symposia on Applied Mathematics*, Vol. 44, 1991, 81-98.

[B-B] B. Bollobás, G. Brightwell, The height of a random partial order: Concentration of Measure, *Annals of Applied Probab.* **2** (1992), 1009-1018.

[C-L] E. G. Coffman, Jr., G. S. Lucker, *Probabilistic Analysis of Packing and Partitioning Algorithms*, Wiley, 1991.

[C-N] F. Comets, J. Neveu, The Sherrington-Kirkpatrick Model of Spin Classes and Stochastic Calculus: the high temperature case, *Comm. Math. Phys.* **166** (1995), 549-564.

[D-MS] S. Dilworth, S. Montgomery-Smith, The distribution of vector-valued Rademacher series, *Ann. Probab.* **21** (1993), 2046-2052.

[F] A. M. Frieze, On the length of the longest monotone subsequence in a random permutation, *Ann. Appl. Prob.* **1** (1991), 301-305.

[G-M] M. Gromov, V. D. Milman, A topological application of the isoperimetric inequality, *Amer. J. Math.* **105** (1983), 843-854.

[Har] L. H. Harper, Optimal numbering and isoperimetric problems on graphs, *J. Comb. Theory* (1966), 385-395.

[H] W. Hoeffding, Probability inequalities for sums of bounded random variables, *J. Amer. Statist. Assoc.* **58** (1963), 13-30.

[J] S. Janson, Poisson approximation for large deviations, *Random Structures and Algorithms* **1** (1990), 221-290.

[J-S] W. Johnson, G. Schechtman, Remarks on Talagrand's deviation inequality for Rademacher's functions, *Lecture Notes in Math.* **1470**, Springer Verlag, 1991, 72-77.

[Ka] R. M. Karp, An upper bound on the expected cost of an optimal assignment, in *Discrete Algorithm and Complexity: Proceedings of the Japan-US joint Seminar*, Academic Press, 1987, 1-4.

[K1] H. Kesten, Aspects of first-passage percolation, Ecole d'Eté de Probabilité de Saint-Flour XIV, *Lecture Notes in Math.* **1180**, 125-264, Springer Verlag, 1986, 125-264.

[K2] H. Kesten, On the speed of convergence in first passage percolation, *Ann. Applied Probab.* **3** (1993), 296-338.

[K-S] R. M. Karp, J. M. Steele, Probabilistic analysis of heuristics, in *The Traveling Salesman Problem*, John Wiley and Sons, 1985, 181-205.

[Lea] J. Leader, Discrete isoperimetric inequalities, *Proceedings of Symposia on Applied Mathematics*, Vol. 44, 1991, 57-80.

[L] M. Ledoux, *Gaussian randomization and the law of the iterated logarithm in type 2 Banach spaces*, Unpublished manuscript, 1985.

[L-T1] M. Ledoux, M. Talagrand, Characterization of the law of the iterated logarithm in Banach spaces, *Ann. Probab.* **16** (1988), 1242-1264.

[L-T2] M. Ledoux, M. Talagrand, *Probability in Banach Spaces*, Springer Verlag, 1991.

[Lu] T. Luczak, The chromatic number of Random graphs, *Combinatorica* **11** (1991), 45-54.

[Mau1] B. Maurey, Construction de suites symétriques, *Comptes Rendus Acad. Sci. Paris* **288** (1979), 679-681.

[Mau2] B. Maurey, Some deviation inequalities, *Geometric and Functional Analysis* **1** (1991), 188-197.

[McD] C. McDiarmid, On the method of bounded differences, in *Survey in Combinatorics* (J. Simons, Ed.), London Mathematical Society Lecture Notes, Vol. 141, Cambridge Univ. Press, London/New York, 1989, 148-188.

[M-H] C. McDiarmid, Ryan Hayward, Strong concentration for Quicksort, *Proceedings of the Third Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 1992, 414-421.

[M-S] V. D. Milman, G. Schechtman, Asymptotic theory of finite dimensional normed spaces, *Lecture Notes in Math.* **1200**, Springer Verlag, 1986.

[Mi1] V. D. Milman, A new proof of the theorem of A. Dvoretzky on sections of convex bodies, *Func. Anal. Appl.* **5** (1971), 28-37.

[Mi2] V. D. MILMAN, Asymptotic properties of functions of several variables defined on homogenous spaces, *Soviet. Math Dokl.* **12** (1971), 1277-1491.

[Mi3] V. D. MILMAN, The heritage of P. Lévy in geometrical functional analysis, *Astérisque* **157/158** (1988), 273-301.

[P] G. PISIER, Probabilistic methods in the geometry of Banach spaces. Probability and Analysis, Varena (Italy) 1985, *Lecture Notes in Math.* **1206**, Springer Verlag, 1986, 167-241.

[R1] W. RHEE, On the fluctuations of the stochastic traveling salesperson problem, *Math. of Operation Research* **13** (1991), 482-489.

[R2] W. RHEE, A matching problem and subadditive Euclidean functionals, *Ann. Applied Probab.* **3** (1993), 794-801.

[R3] W. RHEE, On the fluctuations of simple matching, *Oper. Res. Letters* **16** (1994), 27-32.

[R4] W. RHEE, Inequalities for the Bin Packing Problem III, *Optimization* **29** (1994), 381-385.

[Ro] J. ROSINSKI, Remarks on a Strong Exponential Integrability of Vector Valued Random Series and Triangular Arrays, *Ann. Probab.*, to appear.

[R-T] W. RHEE, M. TALAGRAND, A sharp deviation inequality for the stochastic traveling salesman problem *Ann. Probab.* **17** (1989), 1-8.

[S] G. SCHECHTMAN, Levy type inequality for a class of metric spaces, Martingale Theory in Harmonic analysis and Banach spaces, Cleveland 1981, *Lecture Note in Math.* **939**, Springer Verlag, 1981, 211-215.

[S-S] E. SHAMIR, J. SPENCER, Sharp concentration of the chromatic number of random graphs $G_{n,p}$, *Combinatorica* **7** (1987), 121-129.

[T1] M. TALAGRAND, An isoperimetric theorem on the cube and the Kintchine Kahane inequalities, *Proc. Amer. Math. Soc.* **104** (1988), 905-909.

[T2] M. TALAGRAND, Isoperimetry and integrability of the sum of independent Banach space valued random variables, *Ann. Probab.* **17** (1989), 1546-1570.

[T3] M. TALAGRAND, A new isoperimetric inequality for product measure, and the tails of sums of independent random variables, *Geometric and Functional Analysis* **1** (1991), 211-223.

[T4] M. TALAGRAND, A new isoperimetric inequality for product measure, and the concentration of measure phenomenon, Israel Seminar (GAFA), *Lecture Notes in Math.* **1469**, Springer Verlag, 1991, 94-124.

[T5] M. TALAGRAND, Regularity of infinitely divisible processes, *Ann. Probab.* **21** (1993), 362-432.

[T6] M. TALAGRAND, Supremum of some canonical processes, *Amer. J. Math.* **116** (1994), 295-314.

[T7] M. TALAGRAND, New concentration inequalities, in preparation.

[W] D. W. WALKUP, On the expected value of a random assignment problem, *SIAM J. Comput.* **8** (1979), 440-442.

[Y] V. V. YURINSKII, Exponential bounds for large deviations, *Theor. Prob. Appl.* **19** (1974), 154-155.

Equipe d'Analyse — Tour 48
UA au CNRS nº 754
Université Paris VI
4 Pl. Jussieu
75230 Paris Cedex 05

Manuscrit reçu le 6 décembre 1993