Atti Accademia Nazionale Lincei Classe Scienze Fisiche Matematiche Naturali

Rendiconti Lincei Matematica e Applicazioni

Michel Talagrand

Spin Glasses: A New Direction for Probability Theory?

Atti della Accademia Nazionale dei Lincei. Classe di Scienze Fisiche, Matematiche e Naturali. Rendiconti Lincei. Matematica e Applicazioni, Serie 9, Vol. **11** (2000), n.S1 (Mathematics Towards The Third Millenium), p. 127–146.

Accademia Nazionale dei Lincei

<http://www.bdim.eu/item?id=RLIN_2000_9_11_S1_127_0>

Atti della Accademia Nazionale dei Lincei. Classe di Scienze Fisiche, Matematiche e Naturali. Rendiconti Lincei. Matematica e Applicazioni, Accademia Nazionale dei Lincei, 2000.

Rend. Mat. Acc. Lincei
s. 9, fasc. spec. 2000: 127-146

MICHEL TALAGRAND

SPIN GLASSES: A NEW DIRECTION FOR PROBABILITY THEORY?

ABSTRACT. — Physicists have proposed a wealth of new ideas concerning the collective behavior of large families of random variables, in which they have discovered the emergence of new types of order. The quest for mathematically rigorous proofs of their conjectures is likely to challenge and stimulate probabilists for years to come.

KEY WORDS: Extreme values; Replica-symmetry; Mean field.

1. INTRODUCTION

Independence is a fundamental concept of probability theory. A random event is independent of a collection of other random events if its outcome is not influenced by the outcome of the events of this collection. We all agree that the temperature in Paris has no measurable influence on the outcome of the Ohio lottery. Probabilists (but possibly not all gamblers) also believe that one does not increase ones chances of winning this lottery by buying a ticket from a store that recently sold a winning ticket. (Actually probabilists recommend overall not to invest more than a tiny fraction of one's assets in lottery tickets). Independence, however important, is the exception rather than the rule: the outcome of future events is often influenced by the outcome of past events. The closing value of the Dow Jones index is not independent of its closing value the day before. In many situations, events occur at different times, and we consider them with the natural order of their occurrences. This gives rise to the idea of stochastic processes, the mainstay of Probability Theory. Although many different classes of stochastic processes have been studied, their mathematical structure almost always makes essential use of the order in which events occur. For example, particularly successful is the idea of Markov processes, where all the dependence of the future from the past is only through the present. There are however important cases where the events are not naturally ordered, and it is much less clear what are the important structures in such cases. Somewhat separate from the issue of a natural order on the events, is the situation of very complicated dependence. Consider for example a complex system (such as the space shuttle) that can be in a first approximation considered as being made from many subsystems. Each of the subsystems has a certain probability of failure, and these probabilities are not independent. We are interested in the probability of failure of the entire system, failure that occurs when certain subsets of subsystems fail. How do we model such a situation and estimate the probability of failure of the entire system? The unfortunate truth is that this is very difficult, and that we are not very good at it. The estimates of a major failure in a space shuttle launch were revised from $1/300000$ to $1/300$ after the Challenger accident, figures that

reflect the fact that these estimates contain more guesswork than hard science. (It is a sobering thought that the estimates of a major accident in a nuclear plant are not likely to be more reliable. Yet these estimates greatly influence decisions concerning these plants). A natural first step towards the understanding of the behavior of large families of random variables is the study of the case where the correlation structure of the family is in a sense, very regular, yet is distinctively different from the sort of patterns naturally associated with an ordering. This paper pertains to that direction of research.

A central theme of Probability Theory is the emergence of simple patterns (collective behavior) from disorder, a theme that will also be prominent here. Best known is the Law of Large Numbers. It is very unlikely that by throwing a large number N of fair coins, the number X of heads that come up is very different from $N/2$. In fact much more is true. One has to be *really* unlucky to have $|X - N/2| > 100\sqrt{N}$ (the odds are less than 1 in 10^{8000}). This example is a very special case of «the concentration of measure phenomenon». Although the concentration of measure phenomenon is not the topic of the present paper, we must at least mention this idea (which will be ubiquitous here). It has been a driving force of part of analysis and probability over the last decade, in particular under the influence of Vitali Milman [17, 18]. In probability [22, 23], this idea asserts (in the form of certain inequalities), that any well-behaved function of many, say N, probabilistically independent random variables, has small fluctuations provided its values do not depend too much upon any of these variables. This is a particularly strong type of simple pattern. Concentration of measure has the advantage that it applies even to situations where the function is very complicated, such as the solution of a stochastic optimization problem, and where it is otherwise very hard to say anything at all. Of course, in the statement above, what really matters is what is not said, the precise definition of «well-behaved» and «small fluctuations». Indeed, and even though this could look surprising at first sight, any random variable, at least in principle, can be approximated well by a function of the outcome of flipping N fair coins (with arbitrarily good accuracy as N becomes large), and certainly not every random variable has small fluctuations. Remarkably, the information brought by concentration of measure is still of interest even when N is very large, and in fact sometimes does not depend at all of the value of N, a very precious quality in our area of increasing computer power that makes really large problems relevant. Another very noticeable feature of this theory is that all that is required to prove and formulate these inequalities was probably known around 1930; yet it took another sixty years to elaborate the underlying concepts. Abstractions do require a very long gestation time.

Now we formulate more precisely the theme of this paper: the emergence of collective behavior in large correlated families of random variables. We will consider cases where the correlation structure is very regular, but is in some (rather imprecise) sense «high dimensional». Two distinct points of view of probability theory are the study of «typical outcomes» and of «rare outcomes». Our topic will include a bit of both. We will consider a typical realization of our large family of random variables, giving us

a large family of numbers. We will be interested in the largest (or the «few largest») element(s) of this family.

After these general considerations that attempted to provide some broad perspective for our topic, as seen from the point of view of probability theory, let us turn to a more specific discussion. The words «spin glasses» refer to a class of alloys that display unconventional magnetic behavior at low temperature. A typical example is a mixture of 95% gold and 5% iron. In this alloy, only the iron atoms have magnetic properties. The distances between an iron atom and its closest neighbors are somewhat random, due to the dilution among the atoms of gold. It is believed that this randomness is at the root of the strange magnetic behavior. This explains the link with probability theory. The positions of the iron atoms are disordered, but this disorder does not evolve, it is «frozen» or, more accurately, «quenched» as physicists say. This situation of «quenched disorder» is typical of glasses. Theoretical physicists desired to find models that would explain the strange behavior of these alloys, at least qualitatively, and would help to understand its causes. Any realistic model being too complicated for theoretical analysis, they turned towards simplified models. These models are rather simple mathematical objects. Yet the physicists predict that they have a very rich behavior, a situation of considerable interest. The understanding of this behavior, even by the mathematically non-rigorous methods of the physicists, required considerable ingenuity and the invention of a number of new concepts. These new concepts seem relevant to a number of purely mathematical questions. We will describe in detail in Section 2 some of these questions, which are particularly natural and simple. Having realized the importance of their discoveries, M. Mézard, G. Parisi and M. Virasiro wrote the brilliant book *Spin Glass Theory and Beyond* [16]. A striking feature of this book is that it studies purely mathematical objects and problems in the language and with the methods of theoretical physics. It contains no mathematical proofs, and it is not obvious at all how even to formulate mathematically many of the statements made there. Both physicists and mathematicians have now written hundreds of pages about models for spin glasses, but the questions treated with great efforts by the mathematicians are at the very most the object of a few introductory comments in the physicist's work. Physicists consider this topic as well understood. Mathematicians wonder whether they will ever be able to provide rigorous proofs of the main predictions. This paradoxical situation is the object of Section 4, where we will comment on the different meaning mathematicians and physicists give to the word «understanding». In the subsequent sections, we will try to explain a few of the main ideas while attempting to avoid anything technical. A more detailed account can be found in [33, 34].

2. Three good problems

In this section we state three innocent looking but very difficult problems. These are examples of important questions which deserve more attention than they are currently getting. We have striven to describe these problems in simple terms. We have added a few observations which could help the mathematically inclined reader to get a better

focus, but which might puzzle others. Some of our comments might be lost on a reader who has forgotten the mathematical definition of covariance, or what is the normal law (ever heard of the bell-shaped curve?). But these are just that, comments, and are not essential for following the main story.

2.1. *The Dean's problem.*

This somewhat mischievous name refers to the ordeal of the Dean of a College trying to split a rowdy department in two new departments to decrease internal fighting. In this model, we consider a (large) population of individuals, numbered 1 to N. They all know each other. The feelings of the individual i toward the individual j are measured by a number g_{ij} which can be positive, or unfortunately, negative. For simplicity we will assume symmetry, $g_{ij}=g_{ji}$. This means that the feelings are reciprocal (the colleagues you can't stand can't stand you either). We want to model the situation where these respective feelings are random. We will make the assumption that the numbers (g_{ij}), for $1\leq i<j\leq N$ are independent identically distributed random variables. This choice aims at simplicity, and has no claim to be realistic. For example, it ignores the fact that (at least in a human population) some individuals are definitely more likeable than others. There are several reasons for making such an assumption. First, it is unclear what more realistic assumptions one should use. Second, even the present very simplified model is hard enough to analyze. Third, as surprising as it may seem, even such an extreme simplification apparently preserves many important features arising from randomness. Since we try to write a simple model, we will make a simple choice for the variables g_{ij}. One obvious choice would be $P(g_{ij}=1)=P(g_{ij}=-1)=1/2$. This simply means that, given any two individuals i, j, we flip a fair coin. If heads come up they are friends. Otherwise they are enemies and there is only one possible «intensity» in this feeling. It is believed that the particular choice of the (g_{ij}) does not matter much and we will always assume that g_{ij} is a standard normal random variable. This will make the discussion slightly easier than the «coin flipping» case. The variables g_{ij} will be called «the disorder» because they model just that. A very important feature of the model, called frustration, is that for many triples i, j, k, i and j, i and k are friends $(g_{ij}, g_{ik}>0)$ but j and k are enemies $(g_{jk}<0)$. These are called frustrated triples. The friends of your friends are not necessarily your friends; they are equally likely to be your enemies. There are many social tensions inside the population. The Dean will try to decrease the social tensions by splitting the population in two parts, putting as much as possible friends together and enemies apart. The existence of frustrated triples shows that this cannot be achieved perfectly. Either some enemies will still be together, or some friends will be separated, and most likely both of these undesirable features will be present. Still, what is the best that can be done? It is convenient to assign to each individual i a number σ_i that is either 1 or -1 (its «spin»). A sequence $\boldsymbol{\sigma}=(\sigma_1,\cdots,\sigma_N)$ of such numbers thus defines a partition of the population in two sets. It is natural to call such a sequence $\boldsymbol{\sigma}$ a *configuration.* It describes the configuration of a population of N individuals, each of which having the choice between two values of its spin, $+1$ and -1. How good is

the partition associated to σ at regrouping friends and at separating enemies? A natural way to measure this is by the quantity

$$A_N(\sigma) = \sum_{i<j} g_{ij}\sigma_i\sigma_j \tag{2.1}$$

that adds the interactions between each pair of individuals in the same class, and subtracts the interactions between pairs in different classes. Thus putting a pair of friends together increases $A_N(\sigma)$, while separating them decreases this quantity. Similarly putting a pair of enemies together decreases $A_N(\sigma)$, while separating them increases it. The Dean's problem is to find the maximum of $A_N(\sigma)$ over all possible choices of the configuration σ. Since we cannot exactly separate the enemies and regroup the friends, any choice of σ involves a lot of compromises. To choose the best among them is going to be awfully difficult. Before we discuss further the issues of this problem, let us now consider the case where our N individuals, rather than being persons, are atoms, a situation that prompted two physicists, Sherrington and Kirkpatrick, to introduce this model (that will henceforth be called the SK model) in 1978 [21], the starting point of an ebullient period of research culminating in [16]. One striking feature in this model is that any two atoms interact, in a way that does not depend upon their relative positions. This is very different from any realistic situation where atoms can have significant interactions only with their close neighbors. This assumption is a huge simplification. Even the simplest models which take into account the «geometry» of the situation are so difficult to study that consensus has not yet been reached even concerning the results of numerical simulations. The decision to forget the «geometry» of the situation is called the mean field approximation, and the SK-model is called a mean-field model. One should again wonder at the fact that such a drastic simplification seems to preserve at least some of the main features one tries to model.

Now let us discuss the quantity we are interested in, namely $\max_\sigma A_N(\sigma)$. For each choice of σ, $A_N(\sigma)$ is a random variable. It depends upon the «disorder», *i.e.*, the interactions g_{ij}. Since we assumed these to be standard normal and independent, $A_N(\sigma)$ is again normal, with $EA_N(\sigma) = 0$, and $EA_N(\sigma)^2 = N(N-1)/2$, where E denotes the expected value (average) with respect to the disorder. The important feature is that if $\rho \neq \sigma$, then the random variables $A_N(\sigma)$ and $A_N(\rho)$ are *not* independent. In fact we have

$$E(A_N(\sigma)A_N(\rho)) = \sum_{i<j} \sigma_i\sigma_j\rho_i\rho_j = \frac{1}{2}\left(\sum_{i\leq N}\sigma_i\rho_i\right)^2 - N. \tag{2.2}$$

The very large family of random variables $A_N(\sigma)$, consisting of 2^N elements, where σ varies over all possible choices, has thus a simple correlation structure. The correlation between $A_N(\sigma)$ and $A_N(\rho)$ is determined by the single parameter $\sum_{i\leq N}\sigma_i\rho_i$. It is useful to note that even though our basic randomness (the g_{ij}'s) was very simple (independence) we reach, through the simple formula (2.1), a situation that turns out to be extremely non trivial. What can we say about $\max_\sigma A_N(\sigma)$?

To try to guess what could happen, the mathematically inclined reader could compute what would happen if these random variables $A_N(\sigma)$ were independent. Considering independent normal random variable $B_N(\sigma)$ with $EB_N(\sigma) = 0$, $EB_N^2(\sigma) = N(N-1)/2$, the probability that $\max B_N(\sigma)$ be less than a number t is exactly

$$\prod_{\sigma} P(B_N(\sigma) \leq t) = \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp\left(-\frac{u^2}{N(N-1)}\right) du \right)^{2N} \tag{2.3}$$

from which simple considerations imply that $N^{-3/2} \max_\sigma B_N(\sigma)$ must be (for N large) very close to $\sqrt{2 \log 2}$ with overwhelming probability.

But what about $\max_\sigma A_N(\sigma)$? Remarkably $N^{-3/2} \max_\sigma A_N(\sigma)$ is essentially independent of the disorder, a consequence of the ubiquitous concentration of measure phenomenon. Thus its value is well captured by its average $E \max_\sigma A_N(\sigma)$. It is of order $N^{3/2}$. The second statement is of interest, because there are $N(N-1)/2$ terms in the sum (2.1). This statement can be reformulated by saying that splitting a Department of size N can yield a relative improvement in the personal relationships of at most about $N^{-1/2}$, hardly worth the trouble. This surprising fact results from the randomness of the interactions. Enemies and friends are so completely intertwined that they can hardly be separated at all. It also becomes less surprising if one believes that $\max_\sigma A_N(\sigma)$ should not behave too differently from $\max_\sigma B_N(\sigma)$. According to computer simulation (for N of order 100) the average $N^{-3/2} E \max_\sigma A_N(\sigma)$ should be close to .7633 for large N. Unfortunately the only argument towards the existence of a limiting value of this average as N becomes very large is that one sees no reason why it should not exist.

2.2. *An integer programming problem, and the lottery phenomenon.*

The goal of integer programming is to find integer solutions to systems of linear inequalities. In the more restricted version considered here we allow only values that are $+1$ or -1. We look for numbers (called again «spins») $\sigma_i; i = 1, \cdots, N$ that satisfy M linear constraints,

$$\sum_{i \leq N} \eta_{i,k} \sigma_i \geq 0 \tag{2.4}$$

for $k = 1, \cdots, M$ and we require $\sigma_i = 1$ or $\sigma_i = -1$. We try again to model disorder (or maybe the «generic» situation) by assuming that the constraints (2.4) are random. For reasons similar to those in the Dean's problem, we make the simplest possible choice, and we assume that the coefficients $\eta_{i,k}$, $i = 1, \cdots, N; j = 1, \cdots, M$, are independent random variables, with the same distribution. The specific form of this distribution is not expected to be very important. We assume for specificity $P(\eta_{i,k} = 1) = P(\eta_{i,k} = -1) = 1/2$. It of course depends upon the realization of the disorder whether the system (2.4) has solutions. But does it typically have solutions, and if so, how many, out of 2^N possible choices for the numbers σ_i? A first analysis of the problem is as follows. Assuming N odd to prevent equality in (2.4), given a particular choice

of $\boldsymbol{\sigma} = (\sigma_1, \cdots, \sigma_N)$, by symmetry there is exactly a 50% chance that any given constraint (2.4) is satisfied. By independence, there is a probability 2^{-M} that $\boldsymbol{\sigma}$ is a solution; The expected number of solutions is thus 2^{N-M}. The reason why this is not the interesting part of the story is that there is a subtle difference between the «expected» and the «typical» number of solutions. Much of the difficulty is linguistic. The word «expected» in probability does not mean «that one can realistically expect» but means «in average» over repeating the same experiment many times. The Ohio lottery can help us again to understand the difference. If in a particular drawing, the jackpot is \$10 million, and you have a chance of 1 in 20 million of winning it, then your expected (= average) gain for buying a \$1 ticket is $50c$, but your typical gain is zero, a hard fact that motivated our earlier investment advice. (On the other hand, should there be such a drawing once a week, and should you buy a ticket at each drawing, after only about 554992 years you would have a fifty-fifty chance of having won at least once). In the rest of the paper we will invoke the «lottery phenomenon» in situations where the typical value of a random quantity is much smaller than its average value.

The interesting case is when M is large, of order N. The natural measure of the size of M is the ratio $\alpha = M/N$. Then the expected number of solutions 2^{N-M} is a very small proportion of 2^N. There are comparatively very few solutions. It is hard to find them, and to decide whether they exist at all.

The physicists conjecture [14] that when N is large, there are solutions provided $\alpha \leq .83$ (about). Thus, for $\alpha = .9$, there should typically be no solution, while the average number of solutions is $\gg 1$. This is an occurrence of the «lottery phenomenon» described above. There are (extremely rare) occurrences of the random coefficients $\eta_{i,k}$ for which the system (2.4) has so many solutions that this does influence the average number of solutions; but certainly we should not count on these to occur any more than we should count on winning the Ohio lottery. When there are solutions, a good way to «rescale» their number S, to obtain a quantity of order 1 is to consider the quantity

$$\frac{1}{N} \log(2^{-N} S). \tag{2.5}$$

This quantity is known to be essentially independent from the randomness for large N, and the problem (to which physicists propose an explicit solution [12]) is to compute its limiting values as $N \to \infty$, for given α. The problem considered in this section is rather important in the theory of neural networks, where it is called the «perceptron capacity problem».

2.3. *The assignment problem and the cost of monogamy.*

In this problem we are given N boys and N girls, numbered 1 to N, and we have to create happy couples. Creating couples means that we have to find a one to one correspondence ρ between the set $\{1, \cdots, N\}$ of girls and the set $\{1, \cdots, N\}$ of boys. We assign a partner to each boy or girl, (so that the problem is usually called the *assignment problem*). The reader will observe that this is a perfectly politically correct problem. Neither sex has any advantage over the other. Not all couples get

along equally well. There is a cost a_{ij} in assigning j to i as a partner. The assignment problem is to minimize the total cost $\sum_{i\le N} a_{i\rho(i)}$ (that is, the sum of the costs of creating each couple) over all possible choices of the correspondence ρ. It would be mathematically equivalent (but much more uplifting) to «maximize happiness» rather than to «minimize costs» when creating couples, but we have not felt appropriate to change the (gloomy ...) traditional formulation. What makes the problem interesting is the requirement that ρ be one to one. The girl i would like to pick for partner the boy j she likes the most; but some boys are elected as best choices by several girls, and these conflicts must be resolved. Some girls will have to settle for their second best choices. A few might even have to accept their third best, or worse. We will consider the stochastic version where the costs a_{ij} are independent random variables uniformly distributed over $[0, 1]$. The comments made in Section 2.1 about the meaningfulness of the choices of the variables (g_{ij}) there apply equally well here, and this choice has no pretense of being realistic. It is known that $\min_\rho \sum_{i\le N} a_{i\rho(i)}$ is essentially independent of the randomness (concentration of measure again), so that this random quantity is well represented by its expectation. The physicists conjecture that

$$\lim_{N\to\infty} E\min_\rho \sum_{i\le N} a_{i\rho(i)} = \frac{\pi^2}{6}, \tag{2.6}$$

or in words, that for N large, the cost of the best assignment is very close to $\pi^2/6$. In my opinion this is one of the prettiest conjectures of the entire probability theory. To understand why this sum of N terms is of order 1, we must keep in mind that, for each girl i there are typically several boys j for which a_{ij} is of size of order $1/N$. If we were not having the requirement that different girls must choose different boys we would have 1 in (2.6) rather than $\pi^2/6 \simeq 1.64$; the extra cost arises from the monogamy requirement. In short, the cost of monogamy is believed to be $\pi^2/6 - 1 \simeq 0.64$! (On the other hand, it was firmly pointed out to me that any attempt at polygamy would incur very high costs not taken into account by this model).

Of course, the fascinating part of conjecture (2.6) is the occurrence of the exact value $\pi^2/6$ $(= 1 + 1/4 + 1/9 + \cdots)$. There is even a more fascinating conjecture by G. Parisi. Consider independent exponentially distributed random variables b_{ij}, that is, for $t \ge 0$,

$$P(b_{ij} \ge t) = e^{-t}.$$

It can be proved that (2.6) is equivalent to the corresponding statement where one replaces a_{ij} by b_{ij}. This is simply because only the small values among the numbers a_{ij}, b_{ij} matter for the assignment problem, and for small t,

$$P(a_{ij} \ge t) = 1 - t \simeq e^{-t} = P(b_{ij} \ge t).$$

The conjecture of Parisi is that for each N,

$$E\min_\rho \sum_{i\le N} b_{i\rho(i)} = 1 + \frac{1}{2^2} + \cdots + \frac{1}{N^2}. \tag{2.7}$$

This has been checked by M. Mézard up to $N = 5$. Such a miraculous formula for such a complicated problem, if true, almost certainly indicates a simple underlying structure yet to be discovered. This points out to the fact that the «statistical mechanics» approach at which we hint below is probably not needed. In fact, D. Aldous proved in [1] that the limit of the right-hand side of (2.6) exists, and, motivated by the conjectures of the physicists, recently succeeded in refining the arguments of [1] to prove that

$$\lim_{N\to\infty} E \min_{\rho} \sum_{i\leq N} a_{i\rho(i)} \leq \frac{\pi^2}{6},$$

a giant step towards (2.6) (private communication). The method of Aldous does not seem to be related to the method of the physicists. However, even if Aldous' method succeeds in proving (2.6), the very subtle structures predicted by the physicists will remain of considerable interest.

We chose to discuss the three previous problems in detail because they are particularly simple to state, and yet lie very deep. Lack of space prevents us from explaining other important questions, and in particular from discussing the Hopfield model, a very popular model of memory [13, 3, 25, 4, 30]. Also important are the random p-sat problem (a stochastic version of the problem of satisfiability of many logical clauses which is fundamental in computer science) [27] and stochastic versions of time-honored optimization problems such as the Traveling Salesman Problem.

3. Statistical mechanics and combinatorial optimization

The problem of finding the maximum of a quantity such as (2.1) over all possible configurations (choices of $\boldsymbol{\sigma}$) is usually called a combinatorial optimization problem. Arising from statistical mechanics is the idea of replacing the search of the largest value by a study of the «large values», with a suitable weight. This weight is provided by a parameter $\beta \geq 0$ that physically represents «the inverse of the temperature of the system». Combinatorial optimization problems then appear as the «zero temperature case» of a more general question. It is necessary to replace (2.1) by a quantity more adapted to the present purpose, namely

$$H_N(\boldsymbol{\sigma}) = -\frac{1}{\sqrt{N}} \sum_{i<j} g_{ij}\sigma_i\sigma_j. \tag{3.1}$$

The minus sign respects the physical idea, that $H_N(\boldsymbol{\sigma})$ represents the energy of the configuration $\boldsymbol{\sigma}$, (Hamiltonian) and that systems always tend to favor low energy levels. The factor $\sqrt{N}$ ensures that $\min_{\sigma} H_N(\boldsymbol{\sigma})$ is of order N, and, as we will see, this is the correct normalization. Given a parameter $\beta \geq 0$, we introduce Gibbs measure G_N, a probability measure on the set of all configurations, that gives to $\boldsymbol{\sigma}$ the weight

$$G_N(\boldsymbol{\sigma}) = \frac{1}{Z_N} \exp(-\beta H_N(\boldsymbol{\sigma})). \tag{3.2}$$

The number Z_N is the factor required to ensure that the sums of the weights is one, namely

$$Z_N = \sum \exp(-\beta H_N(\sigma)) \tag{3.3}$$

where the sum is over the 2^N possible configurations. A large part of the subtlety of what occurs here is contained in the fact that this sum is over a great many terms of very different order of magnitudes. Many terms are at most one (*e.g.* if $H_N(\sigma) \geq 0$) but a few terms are very large. The normalization that we have chosen, that $\max_\sigma H_N(\sigma)$ be of order N, is designed to ensure that in the sum (2.1) is not obviously dominated by the contribution from the many small terms or by the contribution from the rare large ones. The precise magnitude of the terms that contribute the most depends upon the value of β.

The relationship between the behavior of $G_N(\sigma)$ and the problem of the minimum value of $H_N(\sigma)$ is the following. If one fixes N, and lets β become large, the weights (3.2) concentrate on the configuration σ that minimizes H_N (and that is called the ground state in physics). One can then hope to get information about the energy of the ground state by studying Gibbs's measure for large values of β. There are several gains in this approach. First, even though we are interested in the case of β large, we can start by studying the case of β small, which should be easier. The parameter β can be viewed as interpolating the difficulty between a trivial situation ($\beta = 0$) and a very tough one (β very large). Second, it has historically been a very successful approach to find a wider framework in which to study a difficult problem. Rather than being stuck with a single narrow problem, we have new objects from statistical mechanics to study to make progress. And, last but not least, we can appeal to intuition from physics.

We have here introduced statistical mechanics as a way to approach hard problems of combinatorial optimization, because it is possibly the most attractive feature of this circle of problems. But of course statistical mechanics is very interesting in itself. Consider a physical system that can be in any of the 2^N configurations σ considered above, and assume that when it is in the configuration σ, its energy is $H_N(\sigma)$. Then Gibbs' measure (3.2) is the probability that the system is in configuration σ when it is in thermal equilibrium with a heat bath at temperature $1/\beta$. It is of course statistical mechanics that motivated Sherrington and Kirkpatrick to introduce their famous model (that we called here the Dean's problem).

The link between the Dean's problem and statistical mechanics is easy to see; but what about the other two problems of Section 2? What is the relevant way to define the energy of a configuration? In the case of our integer programming problem, a very natural choice is

$$H_N(\sigma) = -S(\sigma)$$

where $S(\sigma)$ is the number of constraints (2.4) that σ satisfies. Then there exists a configuration σ satisfying all constraints if and only if there is one for which $S(\sigma) = M$, that is $\min_\sigma H_N(\sigma) = -M$. Deciding the existence of solutions is naturally reduced in this manner to the problem of finding the minimum of a certain function H_N.

Concerning the assignment problem, the correct choice of the Hamiltonian is a notch more subtle, and the reader is better referred to [16] for this.

4. Mathematical and physical understanding

The author must first point out that he has no particular qualification to discuss this difficult question other than being a modest practitioner in his field. The comments of this section are probably naive and have no claim whatsoever to apply beyond the specific area of spin glasses. They are directed especially to non-mathematicians, for whom the nature of mathematical work is not necessarily clear. The physicists feel they understand almost everything about spin glasses, while mathematicians feel they understand close to nothing. We are not aware of any authoritative statistical study that rules out a lower I.Q. level among mathematicians; yet at least part of the answer seems to lie in different meanings of the word «understanding».

The physicists do not demand irrefutable arguments, but rather try to understand the behavior of the objects they study with a high degree of certainty. Their natural tendency is always to assume that «everything is fine». They do not pay attention to pathologies that are possible in principle but are very unlikely. They always propose as solution «the most natural behavior». «The most natural behavior» needs by no means be simple; it can even be extremely intricate as in many parts of [16]. Once this behavior has been guessed, it is checked as much as possible. It is first checked by theoretical computations. A mathematician would likely call these «heuristic» (as opposed to rigorous) since they often rely on natural, sensible, but unproven approximations. They also sometimes rely upon procedures to which it is currently very unclear how to assign a mathematical meaning. The proposed behavior is finally checked through numerical simulation. Numerical simulation for spin glasses is not easy. Often it can be carried out only for rather small values of the parameters (number of atoms of order 100); but this is compensated by great ingenuity, in trying to check behaviors that cannot occur by mere accident.

On the other hand, the mathematician does require irrefutable arguments, and will spend unlimited energy to rule out pathologies, regardless of how unlikely they might be. One basic reason why the results from physicists and mathematicians are so much apart is quite simple. The physicists make extensive use of the intuitive properties of the material objects they have acquired over literally hundreds of years, and this despite the fact that the SK model is a mean-field model (that is, does not take into account the geometric locations of the atoms) and gives little reason to use this intuition. For example, it is obvious that the addition of a single atom does not change the properties of a sample of matter. It is obvious, and can be measured very precisely, that the properties of a sample of matter do not depend upon its size (modulo an obvious scaling), a principle known as the existence of the thermodynamical limit. These principles are not so obvious for the SK and related models, if only because they are at best true for most, but not all, the realizations of the random interactions. Many of the arguments of the physicists rely from the start on the existence of the thermodynamical

limit, leaving the mathematician looking for an approach. Another principle, intuitively obvious for a physicist, is that at high temperature, a system is «in a pure equilibrium state» (an expression we will try to explain later). This is the «default situation» that is abandoned only when untenable. The validity of this principle is unquestionable in the physical world, but does it remain valid for mathematical objects rather far from this world? In this particular case, the answer turns out to be yes, because rigorous justification has recently been given, but at the expense of considerable work.

Mathematicians feel that this work is necessary. This of course raises the larger issue of what is the value of trying to prove results mathematically when their validity is already widely accepted. Are we talking only of hair splitting here? There are of course several levels of answer. One obvious reason is that mathematical results build upon previous results. This has been going on for many generations, and hopefully will keep going on for many more. It would compromise the whole construction to allow at any step a possibility of error, however remote, since errors could accumulate in a disastrous way. The main motivation (at least for this author) is however rather different. In order to be able to find complete arguments, one has to work much harder than if one allows «reasonable» unproven assumptions. In working much harder, one (sometimes) gains a deeper level of understanding, and (more rarely) one discovers connections and abstract principles that greatly enhance our knowledge, and would have remained undiscovered otherwise. One could give here the example of concentration of measure, which was briefly explained in the introduction. It can be learned in a few days, but took sixty years and the work of many to emerge. The rate at which important abstract principles are discovered is very slow, but there is seemingly no other method to find them than to go through a huge amount of unrewarding nitty-gritty work, just as is the case for most worthy discoveries in other areas of science. Situations where simple objects lead to very complicated behavior as is the case here seem the best potential mining ground. And in conclusion, to reassure the reader not yet convinced that this is a worthwhile endeavor, let us point out that the resources invested by mankind in this particular project (and even in mathematics in general) are really minuscule by any measure.

5. First results on the SK model

Having decided to study problems such as these of Section 2 through the statistical mechanics approach of Section 3, what questions should one ask? Obviously the Gibbs' measure (3.2) is a central object, but for pedagogical reasons we postpone its discussion to the next section. Another quantity of interest is the number Z_N of (3.3), called in physics the partition function. As we previously explained, it is not at all obvious a priori how large Z_N is, so we should try to calculate it. This is not just a technical question. It can be shown that one can recover many of the properties of Gibbs' measure from the value of Z_N. (Therefore, it will not come as a surprise that the computation of Z_N requires in turn a detailed understanding of Gibbs measure). It is in general not difficult to compute the average of Z_N over the disorder; but we must beware of the «lottery phenomenon». We are not interested in the average value of Z_N,

but in its «typical value», which can be much smaller. The typical value of Z_N increases exponentially fast with N, so it is convenient to rescale it and to consider instead

$$\frac{1}{N} \ln Z_N. \tag{5.1}$$

This quantity depends upon the disorder. However it turns out that for many models the fluctuations of this quantity due to the disorder are of size about $N^{-1/2}$. This is again a consequence of concentration of measure. Thus, at this order of accuracy, (almost ...) all the information contained in this random variable is in fact contained in the number

$$E \frac{1}{N} \ln Z_N \tag{5.2}$$

and we would like to compute this for large N. The quantity (5.2) measures, on a logarithmic scale, the typical value of Z_N; if the quantity (5.2) is about a, this means simply that Z_N is typically of size e^{aN}. On the same logarithmic scale, the average value of Z_N is measured by

$$\frac{1}{N} \ln EZ_N, \tag{5.3}$$

a quantity that is always larger than or equal to (5.2) (as follows from the concavity of the logarithm). Thus, when the quantities (5.2) and (5.3) are nearly equal, this means that the «lottery phenomenon» does not occur.

We will now center the rest of our discussion on the SK model, not because it is necessarily the most interesting, but because it does have some simple features that involve a number of general phenomena. The best known rigorous result about the SK model is the fact that for $\beta \leq 1$, we have [2]

$$\lim_{N \to \infty} E \frac{1}{N} \ln Z_N = \ln 2 + \frac{\beta^2}{4}. \tag{5.4}$$

On the other hand, for each $\beta \geq 0$, it is very simple to show that

$$\lim_{N \to \infty} \frac{1}{N} \ln EZ_N = \ln 2 + \frac{\beta^2}{4}, \tag{5.5}$$

so that the content of (5.4) is that the lottery phenomenon does not occur for $0 \leq \beta \leq 1$. While this is certainly not obvious a priori, it can be interpreted to mean that nothing of real interest happens in that case. The proof of (5.4) is actually not very complicated. On the other hand, the range $\beta > 1$ is the domain of the famous predictions of G. Parisi, that are considerably more interesting, and that will be discussed further in Section 7.

It is rather unfortunate that the well known result (5.4) bears on a somewhat uninteresting situation, because it gives a false idea of the «high temperature» problems. (Since β represents the inverse of the temperature, a «high temperature» problem is a situation where β is not too large. A precise definition will be given in the next section). It is also very unfortunate that the best known model (the SK model) is either

easy ($\beta \leq 1$) or very hard ($\beta > 1$). This is not conducive to progress. To get out of this situation, it is very fruitful to consider a generalization of (3.1), namely

$$H_N(\sigma) = -\frac{1}{\sqrt{N}} \sum_{i<j} g_{ij}\sigma_i\sigma_j - h\sum_{i\leq N} \sigma_i. \tag{5.6}$$

The new term $h\sum_{i\leq N}\sigma_i$ is physically very natural. It represents the action of an external magnetic field that tends to align the spins in the $+1$ direction. The presence of this term makes the situation much more interesting, because it is expected that when $h \neq 0$ the «lottery phenomenon» always takes place. The prediction is that in a certain domain of the parameters (β, h), called the «high temperature» region (and which consists of the left of a certain analytically defined curve), we have

$$\lim_{N\to\infty} \frac{1}{N} E \ln Z_N = \min_q \varphi(q, \beta, h) \tag{5.7}$$

where the minimum is over $0 \leq q \leq 1$, and where

$$\varphi(q, \beta, h) = \frac{\beta^2}{4}(1-q)^2 + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \ln(2\cosh\beta(t\sqrt{q} + h)) e^{-t^2/2} dt. \tag{5.8}$$

On the other hand, it is easy to show that

$$\frac{1}{N} \ln EZ_N = \varphi(0, \beta, h) = \frac{\beta^2}{4} + \ln(2\cosh\beta h) \tag{5.9}$$

and, for $h > 0$, this is larger than the quantity (5.7). The remarkable character of a formula such as (5.7) should be self apparent. Of course, such a remarkable formula does not occur by chance, but reflects an underlying structure that is the object of the next section.

6. Systems in a pure equilibrium state

In the previous section we postponed the study of Gibbs' measure in order to state the striking formula (5.7). We return to this study now. It will be done at a given β, as $N \to \infty$. The basic idea is that if β is not too large (that is, the temperature $1/\beta$ is high enough), Gibbs' measure has a remarkable structure that is the reason behind formulas such as (5.7). This remarkable structure also occurs when the temperature is high enough in many other systems, and it is the purpose of this section to describe it, even though it would take us too far to explain in detail the derivation of (5.7). This structure appears to be the correct mathematical formulation of the physical idea that the system «is in a pure equilibrium state» which means roughly that it cannot be decomposed in simpler systems.

First we should keep in mind that we have two levels of randomness. There is the randomness associated to the thermal fluctuations. This is what the Gibbs measure describes. Then there is the randomness associated to the disorder of the system, the «randomness of the glass», which is called quenched (= frozen in) in physics because it is not affected by the thermal fluctuations. In line with the spirit of the previous sections,

we will study Gibbs' measure for «typical» realizations of the disorder. What kind of structure should one expect? The Gibbs' measure is a priori a very complex object: it involves the weights given to the 2^N possible configurations $\boldsymbol{\sigma} = (\sigma_1, \cdots, \sigma_N)$, $\sigma_i = \pm 1$. Could it be possible that one can describe Gibbs' measure with much fewer parameters? Obviously of importance are the averages $\langle \sigma_i \rangle$ where the brackets denote average with respect to Gibbs' measure (called *thermal averages*), that is

$$\langle \sigma_i \rangle = \int \sigma_i dG_N(\boldsymbol{\sigma}) = \frac{1}{Z_N} \sum \sigma_i \exp(-\beta H_N(\boldsymbol{\sigma})) \tag{6.1}$$

when the sum is over all configurations. These numbers $\langle \sigma_i \rangle$ do not a priori determine Gibbs' measure. To know Gibbs measure, we need to know not only the quantities (6.1), but also we need to know the values of all the averages $\langle \sigma_{i_1} \sigma_{i_2} \cdots \sigma_{i_n} \rangle$ for all possible choices of $1 \leq n \leq N$, $i_1 < \cdots < i_n$. This simple fact should be intuitively clear if one observes that the number of such averages (*i.e.* $2^N - 1$) is exactly the number of «parameters» needed to determine a probability measure on the configuration space. There are 2^N configurations, the weights of which must add to one. The simplest structure one could have is that

$$\langle \sigma_{i_1} \cdots \sigma_{i_n} \rangle = \langle \sigma_{i_1} \rangle \cdots \langle \sigma_{i_n} \rangle \tag{6.2}$$

in which case the numbers (6.1) determine Gibbs's measure. However, condition (6.2) is too much to ask. It is never true for our class of models (unless $\beta = 0$). What will happen instead (at high enough temperature) is that (roughly speaking) for N large, the two sides of (6.2) will be nearly equal for «most» (but *not* all) choices of $i_1, \cdots, i_n$. Quite remarkably, requiring this asymptotic near equality for any choice of n turns out to be the same as requiring it for $n = 2$. We will say (by definition) that the system is in a *pure equilibrium state* (at a given value of the parameters) if for large N we have $\langle \sigma_1 \sigma_2 \rangle \sim \langle \sigma_1 \rangle \langle \sigma_2 \rangle$ for the typical disorder, or more formally, if

$$\lim_{N \to \infty} E|\langle \sigma_1 \sigma_2 \rangle - \langle \sigma_1 \rangle \langle \sigma_2 \rangle| = 0. \tag{6.3}$$

All the sites play the same role, so in (6.3) we can replace the indices 1, 2 by any other. Even though (6.3) involves only 2 spins, it implies more generally that given any number n of spins, we have

$$\lim_{N \to \infty} E|\langle \sigma_1 \sigma_2 \cdots, \sigma_n \rangle - \langle \sigma_1 \rangle \cdots \langle \sigma_n \rangle| = 0. \tag{6.4}$$

The link between (6.4) and the physical idea of «pure equilibrium state» is unfortunately not intuitive, but it is easier to explain the mathematical usefulness of (6.4). If (6.2) was true, this would mean that the maps $\boldsymbol{\sigma} \to \sigma_i$ are independent random variables when the configuration space is provided with probability G_N, and computations involving independent random variables tend to be easy. Condition (6.2) is not true, but its substitute (6.4) means that any fixed collection of spins is asymptotically independent under Gibbs' measure. It can then be shown (with work) that in a number of crucial situations computations can be carried out as if the spins were independent under G_N. This is in particular the case for the use of the *cavity method*. This method is of great

importance, both in the heuristic arguments of the physicists and in mathematically rigorous arguments, and we explain its principle now. The basic idea is to examine what happens when one adds spins one at a time. Or, equivalently, one can remove one spin (atom), creating a «cavity», to study the $(N+1)$-spin system in function of a smaller N-spin system. It is a simple algebraic manipulation to express the Gibbs' measure G_{N+1} for the $(N+1)$-spin system in function of the Gibbs' measure G_N for the N-spin system. The basic quantities in this computation are averages

$$\left\langle \exp t \sum_{i\le N} g_i\sigma_i \right\rangle = \left\langle \prod_{i\le N} \exp tg_i\sigma_i \right\rangle \tag{6.5}$$

where $t = \pm\beta/\sqrt{N}$, and where $(g_i)_{i\le N}$ is an independent sequence of standard normal random variables, which is independent of the disorder in G_N. Condition (6.3) can be shown to imply that the quantity (6.5) can be replaced in the calculations by the quantity

$$\prod_{i\le N} \langle \exp tg_i\sigma_i \rangle \tag{6.6}$$

just as if the spins where independent under Gibbs' measure. The averages involved in (6.5) are easy to compute, because σ_i takes only the values 1 and -1.

In summary of this discussion, under condition (6.3), explicit computations are possible with Gibbs' measure (and it is these conditions that lead to (5.7)). The difficulty of the mathematical approach (in contrast with the physicist's approach that assumes a priori that (6.3) holds) is that before (6.3) has been proved, it is very difficult to say anything at all about Gibbs' measure. It has now been proved that condition (6.3) holds for the SK model when β is small enough (and h is arbitrary), and has also been proved (again at high enough temperature) for several other typical models [25-28].

The basic property of a system in a pure state is that for many purposes Gibbs' measure is determined by the N averages (6.1), so it is natural to wonder how these behave. For the SK model and several other important models it has now been proved that at high enough temperature the behavior is somehow as simple as possible. The averages $\langle\sigma_i\rangle$ are random variables, (since they depend upon the disorder). Any fixed number of them, as N increases at fixed temperature become asymptotically independent with a limiting distribution that can be explicitly described.

7. Low temperature

Since condition (6.3), the asymptotic vanishing of spin correlations, is true for many models at high temperature, it is quite natural to define the «high temperature» behavior of a system by the validity of (6.3). This terminology however is somewhat misleading. In certain cases (and in particular in the case of the assignment problem), it is conjectured that high temperature behavior takes place at *any* temperature. This is of importance, because this means that it would be sufficient to understand a certain case of high temperature behavior to prove (2.6), and the prospects of mathematically rig-

orous results in cases of high temperature behavior are currently much brighter than in the case of «low temperature behavior» (which we define by the failure of (6.3)). The only case of low temperature behavior that is completely understood is the «random energy model», a «toy model», invented by B. Derrida [8]. The idea is the same as in (2.3). Since the difficulty of the SK model stems from the correlations (2.2), let us throw them overboard! Let us consider a model where the energy $H_N(\sigma)$ of the configuration σ is a normal r.v, with

$$EH_N(\sigma) = 0\,, \quad EH_N^2(\sigma) = \frac{N}{2} \tag{7.1}$$

(in harmony with (3.11)), but where the correlations are gone, for two configurations $\sigma\,, \rho$

$$\sigma \neq \rho \Rightarrow E(H_N(\sigma)H_N(\rho)) = 0. \tag{7.2}$$

It can be shown that the system is in a pure state if $\beta < 2\sqrt{\ln 2}$, but not if $\beta > 2\sqrt{\ln 2}$.

The analysis of the later case is not difficult, yet is very instructive. The largest term of the sum (3.3) is of the same order than the whole sum, and the few largest terms contribute essentially all the sum. (More precisely, there is a number $n(\beta)$, depending upon β but *not* upon N, such that for 99% (say) of the realizations of the randomness, the $n(\beta)$ largest terms of the sum contribute 99% of the value of this sum). To express what happens in terms of Gibbs' measure, it is convenient to rank the configurations in decreasing order of their Gibbs weights. These weights do fluctuate in an essential way with the disorder (and the probability distribution of these weights is a natural and interesting object). Any given proportion (however close to one) of the mass of Gibbs' measure is (with probability close to one) carried by a number of configurations independent of N; but on the other hand no given number of configurations will carry all the mass as N large. We will describe this situation in a pictorial manner by saying that (in the limit $N \to \infty$) Gibbs' measure is «carried by a sequence of configurations» (an expression assuming quite incorrectly that it is possible to take this limit). This is a highly non-trivial situation, specially in view of the simplicity of the model.

The random energy model is the «building block» for the behavior of the SK model at low temperature proposed by G. Parisi. The breadth (and the ambition) of the ideas proposed by Parisi can be appreciated in [19, pp. 298-346], which is one of the most accessible introductions to his theory.

The very basis of his prediction, that the Gibbs' measure «is carried by a sequence of pure equilibrium states» is motivated by analogy with the behavior of much simpler models (with no disorder). There are no reasons to seriously doubt that this is the correct picture, but rigorous arguments, and even maybe precise definitions are still lacking. Intuitively a pure equilibrium state is a part of the configuration space such that if we forbid the configurations not in this part, the corresponding system is in a pure equilibrium state as defined in the previous section. The «organization of the pure equilibrium states», described by Parisi is so fascinating that one might forget that the *existence* of these states is in itself quite remarkable. We will explain a very striking rigorous consequence of this existence to make this important point. Let us define

the overlap of two configurations σ, ρ by $N^{-1}\sum_{i\leq N}\sigma_i\rho_i$. Then, if Gibbs measure is carried by a sequence of pure equilibrium states there is a number n independent of N such that, however large N, given the typically realization of the disorder, if one chooses independently two configurations σ, ρ (weighted according to their Gibbs' weights), there is a 99% chance that their overlap is (extremely close to) one of *only* n possible values, (and these permitted values depend upon the disorder).

Since the SK model is so difficult at $\beta > 1$, it is not the best starting point of a rigorous investigation. We will discuss a model (also invented by B. Derrida) which is conjectured to behave somewhat like the random energy model, but in a much less trivial way: the p-spin interaction model (p integer ≥ 2). The idea is to replace the 2-spin interaction $\sigma_i\sigma_j$ of (3.1) by a p-spin interaction $\sigma_{i_1}\cdots\sigma_{i_p}$. The Hamiltonian (3.1) is replaced by

$$H_N(\sigma) = \left(\frac{p!}{2N^{p-1}}\right)^{1/2}\sum_{i<j} g_{i_1\cdots i_p}\sigma_{i_1}\cdots\sigma_{i_p}. \tag{7.3}$$

The sum is over all choices of indices $1 = i_1 < \cdots < i_p \leq N$; the coefficients $g_{i_1\cdots i_p}$ are independent standard normal random variables. The normalization coefficient $(p!/2N^{p-1})^{1/2}$ in (7.2) is chosen so that

$$EH_N^2(\sigma) = \frac{p!}{2N^{p-1}}\binom{N}{p} \sim \frac{N}{2} \tag{7.4}$$

in harmony with (3.1), (7.2). For $p = 2$ this is the SK model; but the behavior is believed to be very different if $p \geq 3$. The correlation between the variables $H_N(\sigma)$ is now, at the first order,

$$E(H_N(\sigma)H_N(\rho)) \simeq \frac{N}{2}\left(\frac{1}{N}\sum_{i\leq N}\sigma_i\rho_i\right)^p.$$

Since $N^{-1}|\sum_{i\leq N}\sigma_i\rho_i| \leq 1$, the larger p, the smaller the correlations. The basic idea of Derrida was that if we fix N and have $p \to \infty$, this model «converges» to the random energy model; but we are instead interested in fixing p and having $N \to \infty$.

The rigorous results on the p-spin interaction model recover almost all the predictions of the physicists, and validate their most spectacular aspects. There is a critical value β_p such that if $\beta < \beta_p$ the system is in a pure equilibrium state. On the other hand, for $\beta > \beta_p$, but not too large, it has been proved that Gibbs measure is «carried by a sequence of small pieces of the configuration space». Small here means small for the natural distance on the configuration space. The distance between two configurations is simply the proportion of indices where they differ. For suitable (= large) p, one can achieve that any two configurations in the same «piece» differ in at most 1% of their indices, so that two configurations in different pieces will differ in at least 49% of these indices (almost as much as two «generic» configurations). These small pieces were very recently proved to be the «pure equilibrium states» predicted by the physicists. The remarkable fact is that these pure equilibrium states seem to form spontaneously, and that their existence is by no means obvious from the Hamiltonian. The reason why the

decomposition of the configuration space is much easier to obtain here than in the SK model is simply that the pieces are «as far from each other as they can possibly be» in contrast with the SK model where they «huddle together».

8. Conclusion

Physicists propose a rich collective behavior for large correlated families of random variables, both in the «high temperature» and the «low temperature» regime. It is a long range program to prove their conjectures mathematically. Significant progress has been achieved at «high enough temperature», where for several typical models the predicted behavior has been established with great precision. The progress on the much harder low temperature regime remains more modest, but the fresh ideas proposed by the physicists will hopefully become the focus of significant efforts.

Acknowledgements

I am grateful to several faculties at The Ohio State University for very helpful comments.

References

The reader will find in [5-7], [9-11], [15], [20], [24], [29], [31-32] material of related interest for further reading.

[1] D.J. Aldous, *Asymptotics in the random assignment problem.* Probab. Th. Rel. Fields, 93, 1992, 507-534.
[2] M. Aizenman - J. L. Lebowitz - D. Ruelle, *Some rigorous results on the Sherrington-Kirkpatrick model.* Commun Math. Phys., 112, 1987, 3-20.
[3] D.J. Amit - H. Gutfreund - H. Sompolinsky, *Statistical mechanics of neural networks near saturation.* Annals of Physics, 173, 1987, 30-67.
[4] A. Bovier - V. Gayrard, *Hopfield models as generalized random mean field models.* In: A. Bovier - P. Picco (eds.), *Mathematical aspects of spin glasses and neural networks.* Progress in Probability, vol. 41, Birkhäuser, Boston 1997.
[5] J.-P. Bouchaud - M. Mézard, *Universality classes for extreme values statistics.* J. Phys. A. Math. Gen., 30, 1997, 7997-8015.
[6] A. Bovier - P. Picco (eds.), *Mathematical aspects of spin glasses and neural networks.* Progress in Probability, vol. 41, Birkhäuser, Boston 1997.
[7] F. Comets - J. Neveu, *Sherrington-Kirkpatrick model of spin glasses and stochastic calculus: the high temperature case.* Commun Math. Phys., 166, 1995, 549-564.
[8] B. Derrida, *Random energy model: An exactly solvable model of disordered systems.* Phys. Rev., B 24 # 5, 1981, 2613-2626.
[9] B. Derrida - E. Gardner, *Optimal storage properties of neural network models.* J. Phys., A 21, 1988, 271-284.
[10] J. Fröhlich - B. Zegarlinski, *Some comments on the Sherrington-Kirkpatrick model of spin glasses.* Commun. Math. Phys., 112, 1987, 553-566.
[11] E. Gardner, *Spin glasses with p-spin interactions.* Nuclear Phys., B 257, #6, 1985, 747-765.
[12] E. Gardner, *The space of interactions in neural network models.* J. Phys., A 21, 1988, 257-270.
[13] J.J. Hopfield, *Neural networks and physical systems with emergent collective computational abilities.* Proc. Natl. Acad. Sci. USA, 79, 1982, 1554-2558.
[14] W. Krauth - M. Mézard, *Storage capacity of memory network with binary couplings.* J. Phys., 50, 1989, 3057-3066.

[15] M. MÉZARD, *The space of interactions in neural networks: Gardner's computation with the cavity method*. J. Phys., A 22, 1988, 2181-2190.
[16] M. MÉZARD - G. PARISI - M. VIRASIRO, *Spin Glass Theory and Beyond*. World Scientific, Singapore 1987.
[17] V. MILMAN, *The heritage of P. Lévy in functional analysis*. Colloque P. Lévy sur les processus stochastiques, Astérisque, 157-158, 1988, 273-301.
[18] V. MILMAN - G. SCHECHTMAN, *Asymptotic Theory of Finite Dimensional Normed Spaces*. Lecture Notes in Math., 1200, Springer Verlag, 1986.
[19] G. PARISI, *Field Theory, Disorder, Simulation*. World Scientific Lecture Notes in Physics, 45, World Scientific, Singapore 1992.
[20] M. SCHCHERBINA, *On the replica-symmetric solution for the Sherrington-Kirkpatrick Model*. Helv. Phys. Acta., 70, 1997, 838-853.
[21] D. SHERRINGTON - S. KIRKPATRICK, *Solvable model of a spin glass*. Phys. Rev. Lett., 35, 1972, 1792-1796.
[22] M. TALAGRAND, *Concentration of measure and isoperimetric inequalities in product spaces*. Publ. Math. I.H.E.S., 81, 1995, 73-205.
[23] M. TALAGRAND, *A new look at independence*. Ann. Probab., 24, 1996, 1-34.
[24] M. TALAGRAND, *The Sherrington-Kirkpatrick model: a challenge to mathematicians*. Probab. Th. Rel. Fields, 110, 1998, 109-176.
[25] M. TALAGRAND, *Rigorous results for the Hopfield model with many patterns*. Probab. Th. Rel. Fields, 110, 1998, 177-276.
[26] M. TALAGRAND, *Rigorous low temperature results for the p-spin interaction model*. Probab. Th. Rel. Fields, 117, 2000, 303-360.
[27] M. TALAGRAND, *The high temperature case of the K-sat problem*. Probab. Th. Rel. Fields, to appear.
[28] M. TALAGRAND, *Self averaging and the space of interactions in neural networks*. Random Structures and Algorithms, 14, 1998, 199-213.
[29] M. TALAGRAND, *Exponential inequalities and replica-symmetry breaking for the Sherrington-Kirkpatrick model*. Ann. Probab., to appear.
[30] M. TALAGRAND, *Exponential inequalities and convergence of moments in the replica-symmetric regime of the Hopfield model*. Ann. Probab., to appear.
[31] M. TALAGRAND, *Intersecting random half spaces: towards the Gardner-Derrida formula*. Ann. Probab., to appear.
[32] M. TALAGRAND, *Intersecting random half cubes*. Random Structures and Algorithms, 15, 1999, 436-449.
[33] M. TALAGRAND, *Huge random structures and mean field models for spin glasses*. Proceedings of the International Congress of mathematicians, Documenta Math. Extra, vol. I, 1998.
[34] M. TALAGRAND, *Verres de spin et optimisation combinatoire*. Séminaire Bourbaki, Vol. 1998/99, Exposés 850-864, Astérisque, 266, 2000, 287-317.

Equipe d'Analyse-Tour 46
E.S.A. au C.N.R.S. no. 7064
Université Paris VI
4 Pl. Jussieu -75230 PARIS Cedex (Francia)