

Introduction

Let us denote by \mathbb{S}_N the sphere of \mathbb{R}^N of center 0 and radius \sqrt{N} , and by μ_N the uniform measure on \mathbb{S}_N . For $i, k \geq 1$, consider independent standard Gaussian random variables (r.v.s) $g_{i,k}$ and the subset U_k of \mathbb{R}^N given by

$$U_k = \left\{ (x_1, \dots, x_N) \in \mathbb{R}^N ; \sum_{i \leq N} g_{i,k} x_i \geq 0 \right\} .$$

The direction of the vector $(g_{i,k})_{i \leq N}$ is random (with uniform distribution over all possible directions) so that U_k is simply a half-space through the origin of random direction. (It might not be obvious now why we use Gaussian r.v.s to define a space of random direction, but this will become gradually clear.) Consider the set $\mathbb{S}_N \cap_{k \leq M} U_k$, the intersection of \mathbb{S}_N with many such half-spaces. Denoting by \mathbb{E} mathematical expectation, it should be obvious that

$$\mathbb{E} \left(\mu_N \left(\mathbb{S}_N \bigcap_{k \leq M} U_k \right) \right) = 2^{-M} , \quad (0.1)$$

because every point of \mathbb{S}_N has a probability 2^{-M} to belong to all the sets U_k , $k \leq M$. This however is not really interesting. The fascinating fact is that when N is large and $M/N \simeq \alpha$, if $\alpha > 2$ the set $\mathbb{S}_N \cap_{k \leq M} U_k$ is typically empty (a classical result), while if $\alpha < 2$, with probability very close to 1, we have

$$\frac{1}{N} \log \mu_N \left(\mathbb{S}_N \bigcap_{k \leq M} U_k \right) \simeq \text{RS}(\alpha) . \quad (0.2)$$

Here,

$$\text{RS}(\alpha) = \min_{0 < q < 1} \left(\alpha \mathbb{E} \log \mathcal{N} \left(\frac{z\sqrt{q}}{\sqrt{1-q}} \right) + \frac{1}{2} \frac{q}{1-q} + \frac{1}{2} \log(1-q) \right) ,$$

where $\mathcal{N}(x)$ denotes the probability that a standard Gaussian r.v. g is $\geq x$, and where $\log x$ denotes (as everywhere through the book) the natural logarithm of x . Of course you should rush to require medical attention if this formula seems transparent to you. We simply give it now to demonstrate that we deal with a situation whose depth cannot be guessed beforehand. The wonderful fact (0.2) was not discovered by a mathematician, but by

a physicist, E. Gardner. More generally theoretical physicists have discovered wonderful new areas of mathematics, which they have explored by their methods. This book is an attempt to correct this anomaly by exploring these areas using mathematical methods, and an attempt to bring these marvelous questions to the attention of the mathematical community. This is a book of *mathematics*. No knowledge of physics or statistical mechanics whatsoever is required or probably even useful to read it. If you read enough of this volume and the next, then in Volume II you will be able to understand (0.2).

More specifically, this is a book of probability theory (mostly). Attempting first a description at a “philosophical” level, a fundamental problem is as follows. Consider a large finite collection $(X_k)_{k \leq M}$ of random variables. What can we say about the largest of them? More generally, what can we say about the “few largest” of them? When the variables X_k are probabilistically independent, everything is rather easy. This is no longer the case when the variables are correlated. Even when the variables are identically distributed, the answer depends very much on their correlation structure. What are the correlation structures of interest? Most of the familiar correlation structures in Probability are low-dimensional, or even “one-dimensional”. This is because they model random phenomena indexed by time, or, equivalently, by the real line, a one-dimensional object. In contrast with these familiar situations, the correlation structures considered here will be “high-dimensional” – in a sense that will soon become clear – and will create new and truly remarkable phenomena. This is a direction of probability theory that has not yet received the attention it deserves.

A natural idea to study the few largest elements of a given realization of the family $(X_k)_{k \leq M}$ is to assign weights to these elements, giving large weights to the large elements. Ideas from statistical mechanics suggest that, considering a parameter β , weights proportional to $\exp \beta X_k$ are particularly appropriate. That is, the (random) weight of the k -th element is

$$\frac{\exp \beta X_k}{\sum_{i \leq M} \exp \beta X_i} . \quad (0.3)$$

These weights define a random probability measure on the index set. Under an appropriate normalization, one can expect that this probability measure will be essentially supported by the indices k for which X_k is approximately a certain value $x(\beta)$. This is because the number of variables X_k close to a given large value x should decrease as x increases, while the corresponding weights increase, so that an optimal compromise should be reached at a certain level. The number $x(\beta)$ will increase with β . Thus we have a kind of “scanner” that enables us to look at the values of the family $(X_k)_{k \leq M}$ close to the (large) number $x(\beta)$, and this scanner is tuned with the parameter β .

We must stress an essential point. We are interested in what happens for a *typical* realization of the family (X_k) . This can be very different (and much

harder to understand) than what happens in average of all realizations. To understand the difference between typical and average, consider the situation of the Spinland State Lottery. It sells 10^{23} tickets at a unit price of one spin each. One ticket wins the single prize of 10^{23} spins. The average gain of a ticket is 1 spin, but the typical gain is zero. The average value is very different from the typical value because there is a large contribution coming from a set of very small probability. This is exactly the difference between (0.1) and (0.2). If $M/N \simeq \alpha < 2$, in average, $\mu_N(\mathbb{S}_N \cap_{k \leq M} U_k) = 2^{-\alpha N}$, but typically $N^{-1} \log \mu_N(\mathbb{S}_N \cap_{k \leq M} U_k) \simeq \text{RS}(\alpha)$.

In an apparently unrelated direction, let us consider a physical system that can be in a (finite) number of possible configurations. In each configuration, the system has a given energy. It is maintained by the outside world at a given temperature, and is subject to thermal fluctuations. If we observe the system after it has been left undisturbed for a long time, what is the probability to observe it in a given configuration?

The system we will mostly consider is $\Sigma_N = \{-1, 1\}^N$, where N is a (large) integer. A *configuration* $\sigma = \{\sigma_1, \dots, \sigma_N\}$ is an element of Σ_N . It tells us the values of the N “spins” σ_i , each of which can take the values ± 1 . When in the configuration σ , the system has an energy $H_N(\sigma)$. Thus H_N is simply a real-valued function on Σ_N . It is called the Hamiltonian of the system. We consider a parameter β (that physically represents the inverse of the temperature). We weigh each configuration proportionally to its so-called Boltzmann factor $\exp(-\beta H_N(\sigma))$. This defines Gibbs’ measure, a probability measure on Σ_N given by

$$G_N(\{\sigma\}) = \frac{\exp(-\beta H_N(\sigma))}{Z_N} \quad (0.4)$$

where the normalizing factor Z_N is given by

$$Z_N = Z_N(\beta) = \sum_{\sigma} \exp(-\beta H_N(\sigma)) . \quad (0.5)$$

The summation is of course over σ in Σ_N . The factor Z_N is called the *partition function*. Statistical mechanics asserts that Gibbs’ measure represents the probability of observing a configuration σ after the system has reached equilibrium with an infinite heat bath at temperature $1/\beta$. (Thus the expression “high temperature” will mean “ β small” while the expression “low temperature” will mean “ β large”.) Of course the reader might wonder why (0.4) is the “correct definition”. This is explained in physics books such as [102], [126], [125], and is not of real concern to us. That this definition is very fruitful will soon become self-evident.

The reason for the minus sign in Boltzmann’s factor $\exp(-\beta H_N(\sigma))$ is that the system favors low (and not high) energy configurations. It should be stressed that the (considerable...) depth of the innocuous looking definition (0.4) stems from the normalizing factor Z_N . This factor, the partition function, is the sum of many terms of widely different orders of magnitude, and it

is unclear how to estimate it. The (few) large values become more important as β increases, and predominate over the more numerous smaller values. Thus the problem of understanding Gibbs' measure gets typically harder for large β (low temperature) than for small β (high temperature).

At this stage, **the reader has already learned all the statistical mechanics** (s)he needs to know to read this work.

The energy levels $H_N(\sigma)$ are closely related to the “interactions” between the spins. When we try to model a situation of “disordered interactions” these energy levels will become random variables, or, equivalently, the Hamiltonian, and hence Gibbs' measure, will become random. There are two levels of randomness (a probabilist's paradise). The “disorder”, that is, the randomness of the Hamiltonian H_N , is given with our sample system. It does not evolve with the thermal fluctuations. It is frozen, or “quenched” as the physicists say. The word “glass” of the expression “spin glasses” conveys (among many others) this idea of frozen disorder.

Probably the reader has met with skepticism the statement that no further knowledge of statistical mechanics is required to read this book. She might think that this could be formally true, but that nonetheless it would be very helpful for her intuition to understand some of the classical models of statistical mechanics. This is not the case. When one studies systems at “high temperature” the fundamental mental picture is that of the model with random Hamiltonian $H_N(\sigma) = -\sum_{i \leq N} h_i \sigma_i$ where h_i are i.i.d. Gaussian random variables (that are not necessarily centered). This particular model is completely trivial because there is no interaction between the sites, so it reduces to a collection of N models consisting each of one single spin, and each acting on their own. (All the work is of course to show that this is in some sense the way things happen in more complicated models.) When one studies systems at “low temperature”, matters are more complicated, but this is a completely new subject, and simply nothing of what had rigorously been proved before is of much help.

In modeling disordered interactions between the spins, the problem is to understand Gibbs' measure for a typical realization of the disorder. As we explained, this is closely related to the problem of understanding the large values among a typical realization of the family $(-H_N(\sigma))$. This family is correlated. One reason for the choice of the index set Σ_N is that it is suitable to create extremely interesting correlation structures with simple formulas.

At the beginning of the already long story of spin glasses are “real” spin glasses, alloys with strange magnetic properties, which are of considerable interest, both experimentally and theoretically. It is believed that their remarkable properties arise from a kind of disorder among the interactions of magnetic impurities. To explain (at least qualitatively) the behavior of real spin glasses, theoretical physicists have invented a number of models. They fall into two broad categories: the “realistic” models, where the interacting atoms are located at the vertices of a lattice, and where the strength of

the interaction between two atoms decreases when their distance increases; and the “mean-field” models, where the geometric location of the atoms in space is forgotten, and where each atom interacts with all the others. The mean-field models are of special interest to mathematicians because they are very basic mathematical objects and yet create extremely intricate structures. (As for the “realistic” models, they appear to be intractable at the moment.) Moreover, some physicists believe that these structures occur in a wide range of situations. The breadth, and the ambition, of these physicists’ work can in particular be admired in the book “Spin Glass Theory and Beyond” by M. Mézard, G. Parisi, M.A. Virasoro, and in the book “Field Theory, Disorder and Simulation” by G. Parisi. The methods used by the physicists are probably best described here as a combination of heuristic arguments and numerical simulation. They are probably reliable, but they have no claim to rigor, and it is often not even clear how to give a precise mathematical formulation to some of the central predictions. The recent book [102] by M. Mézard and A. Montanari is much more friendly to the mathematically minded reader. It covers a wide range of topics, and succeeds well at conveying the depth of the physicists’ ideas.

It was rather paradoxical for a mathematician like the author to see simple, basic mathematical objects being studied by the methods of theoretical physics. It was also very surprising, given the obvious importance of what the physicists have done, and the breadth of the paths they have opened, that mathematicians had not succeeded yet in proving any of their conjectures.

Despite considerable efforts in recent years, the program of giving a sound mathematical basis to the physicists’ work is still in its infancy. We already have tried to make the case that in essence this program represents a new direction of probability theory. It is hence not surprising that, as of today, one has not yet been able to find anywhere in mathematics an already developed set of tools that would bear on these questions. Most of the methods used in this book belong in spirit to the area loosely known as “high-dimensional probability”, but they are developed here from first principles. In fact, for much of the book, the most advanced tool that is not proved in complete detail is Hölder’s inequality. The book is long because it attempts to fulfill several goals (that will be described below) but reading the first two chapters should be sufficient to get a very good idea of what spin glasses are about, as far as rigorous results are concerned.

The author believes that the present area has a tremendous long-term potential to yield incredibly beautiful results. There is of course no way of telling when progress will be made on the really difficult questions, but to provide an immediate incitement to seriously learn this topic, the author has stated as research problems a number of interesting questions (the solution of which would likely deserve to be published) that he believes are within the reach of the already established methods, but that he purposely did not, and

will not, try to solve. (On the other hand, there is ample warning about the potentially truly difficult problems.)

This book, together with a forthcoming second volume, forms a second edition of our previous work [157], “Spin Glasses, a Challenge for Mathematicians”. One of the goals in writing [157] was to increase the chance of significant progress by making sure that no stone was left unturned. This strategy greatly helped the author to obtain the solution of what was arguably at the time the most important problem about mean-field spin glasses, the validity of the “Parisi Formula”. This advance occurred a few weeks before [157] appeared, and therefore could not be included there. Explaining this result in the appropriate context is a main motivation for this second edition, which also provides an opportunity to reorganize and rewrite with considerably more details all the material of the first edition.

The programs of conferences on spin glasses include many topics that are not touched here. This book is not an encyclopedia, but represents the coherent development of a line of thought. The author feels that the real challenge is the study of spin systems, and, among those, considers only pure mean-field models from the “statics” point of view. A popular topic is the study of “dynamics”. In principle this topic also bears on mean-field models for spin glasses, but in practice it is as of today entirely disjoint from what we do here.

This work is divided in two volumes, that total a rather large number of pages. How is a reader supposed to attack this? The beginning of an answer is that many of the chapters are largely independent of each other, so that in practice these two volumes contain several “sub-books” that can be read somewhat independently of each other. For example, there is the “perceptron book” (Chapters 2, 3, 8, 9). On the other hand, we must stress that we progressively learn how to handle technical details. Unless the reader is already an expert, we highly recommend that she studies most of the first four chapters before attempting to read anything else in detail.

We now proceed to a more detailed description of the contents of the present volume. In Chapter 1 we study in great detail the Sherrington-Kirkpatrick model (SK), the “original” spin glass, at sufficiently high temperature. This model serves as an introduction to the basic ideas and methods. In the remainder of the present volume we introduce six more models. In this manner we try to demonstrate that the theory of spin glasses does not deal only with such and such very specific model, but that the basic phenomena occur again and again, as a kind of new law of nature (or at least of probability theory). We present enough material to provide a solid understanding of these models, but without including any of the really difficult results. In Chapters 2 and 3, we study the so-called “perceptron capacity model”. This model is fundamental in the theory of neural networks, but the underlying mathematical problem is the rather attractive question of computing the “proportion” of the discrete cube (resp. the unit sphere) that is

typically contained in the intersection of many random half-spaces, the question to which (0.2) answers in a special case. Despite the fact that the case of the cube and of the sphere are formally similar, the case of the sphere is substantially easier, because one can use there fundamental tools from convexity theory. In Chapter 4 we study the Hopfield model, using an approach invented by A. Bovier and V. Gayrard, that relies on the same tools from convexity as Chapter 3. This approach is substantially simpler than the approach first invented by the author, although it yields less complete results, and in particular does not seem to be able to produce either the correct rates of convergence or even to control a region of parameters of the correct shape. Chapter 5 introduces a new model, based on V -statistics. It is connected to the Perceptron model of Chapter 2, but with a remarkable twist. The last two chapters present models that are much more different from the previous ones than those are from each other. They require somewhat different methods, but illustrate well the great generality of the underlying phenomena. Chapter 6 studies a common generalization of the diluted SK model, and of the K -sat problem (a fundamental question of computer science). It is essentially different from the models of the previous chapters, since it is a model with “finite connectivity”, that is, a spin interacts in average only with a number of spins that remains bounded as the size of the system increases (so we can kiss goodbye to the Central Limit Theorem). Chapter 7 is motivated by the random assignment problem. It is the least understood of all the models presented here, but must be included because of all the challenges it provides. An appendix recalls many basic facts of probability theory.

Let us now give a preview of the contents of the forthcoming Volume II. We shall first develop advanced results about the high-temperature behavior of some of the models that we introduce in the present volume. This work is heartfully dedicated to all the physicists who think that the expressions “high-temperature” and “advanced results” are contradictory. We shall demonstrate the depth of the theory even in this supposedly easier situation, and we shall present some of its most spectacular results. We shall return to the Perceptron model, to prove the celebrated “Gardner formula” that gives the proportion of the discrete cube (resp. the sphere, of which (0.2) is a special case) that lies in the intersection of many random half spaces. We shall return to the Hopfield model to present the approach through the cavity method that yields the correct rates of convergence, as well as a region of parameters of the correct shape. And we shall return to the SK model to study in depth the high-temperature phase in the absence of an external field.

In the rest of Volume II, we shall present low-temperature results. Besides the celebrated Ghirlanda-Guerra identities that hold very generally, essentially nothing is known outside the case of the SK model and some of its obvious generalizations, such as the p -spin interaction model for p even. For these models we shall present the basic ideas that underline the proof of the

validity of the Parisi formula, as well as the complete proof itself. We shall bring attention to the apparently deep mysteries that remain, even for the SK model, the problem of ultrametricity and the problem of chaos. A final chapter will be devoted to the case of the p -spin interaction model, for p odd, for which the validity of the Parisi formula will be proved in a large region of parameters using mostly the cavity method.

At some point I must apologize for the countless typos, inaccuracies, or downright mistakes that this book is bound to contain. I have corrected many of each from the first edition, but doubtlessly I have missed some and created others. This is unavoidable. I am greatly indebted to Sourav Chatterjee, Albert Hanen and Marc Yor for laboring through this entire volume and suggesting literally hundreds of corrections and improvements. Their input was really invaluable, both at the technical level, and by the moral support it provided to the author. Special thanks are also due to Tim Austin, David Fremlin and Frédérique Watbled. Of course, all remaining mistakes are my sole responsibility.

This book owes its very existence to Gilles Godefroy. While Director of the Jussieu Institute of Mathematics he went out of his way to secure what has been in practice unlimited typing support for the author. Without such support this work would not even have been started.

While writing this book (and, more generally, while devoting a large part of my life to mathematics) it was very helpful to hold a research position without any other duties whatsoever. So it is only appropriate that I express here a life-time of gratitude to three colleagues, who, at crucial junctures, went far beyond their mere duties to give me a chance to get or to keep this position: Jean Braconnier, Jean-Pierre Kahane, Paul-André Meyer.

It is customary for authors, at the end of an introduction, to warmly thank their spouse for having granted them the peaceful time needed to complete their work. I find that these thanks are far too universal and overly enthusiastic to be believable. Yet, I must say simply that I have been privileged with a life-time of unconditional support. Be jealous, reader, for I yet have to hear the words I dread the most: “Now is not the time to work”.

Mean Field Models for Spin Glasses

Volume I: Basic Examples

Talagrand, M.

2011, XVIII, 485 p., Hardcover

ISBN: 978-3-642-15201-6